И. Г. Венецкий

Математические методы в демографии

И. Г. ВЕНЕЦКИЙ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ДЕМОГРАФИИ

«СТАТИСТИКА» МОСКВА 1971

1-8-3
12-71

В книге рассматриваются математические методы и модели, применяемые при изучении демографических процессов. Характеризуется порядок доживания, приводятся графические конструкции, раскрывающие сущность статических соотношений и динамических процессов демографии.

Автор излагает разнообразные методологические проблемы, возникающие при применении математики в демографическом анализе, дает теоретическое обоснование ряда методов и показывает их практическое применение при обработке тех или иных конкретных статистических данных.

Излагая существующие методы, автор в ряде случаев совершенствует их и предлагает новые математические приемы демографического анализа.

Монография рассчитана на демографов, социологов, экономистов. Она может быть с успехом использована студентами экономических вузов в качестве учебного пособия при изучении курса демографии.

ИЛЬЯ ГРИГОРЬЕВИЧ ВЕНЕЦКИЙ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ДЕМОГРАФИИ

Редактор *Г. И. Чертова*

Техн. редактор *Р. Н. Феоктистова*
Корректор *Г. А. Башарина*
Худ. редактор *Т. В. Стихно*
Переплет художника *Л. С. Эрмана*

Сдано в набор 11/XII 1970 г. Подписано к печати 2/VII 1971 г. Формат бумаги $60 \times 90^1/_{16}$
Бумага № 1 Объем 18,5 печ. л. Уч.-изд. л. 19,73 Тираж 3 900 экз. А10061. Заказ 1603.

Цена 2 р. 16 к.

Тематич. план 1971 г. № 12

Изд-во «Статистика», Москва, ул. Кирова, 39.

Московская типография № 4 Главполиграфпрома
Комитета по печати при Совете Министров СССР
Б. Переяславская, 46

ВВЕДЕНИЕ

Население и факторы, определяющие изменение его численности, структуры, размещения и воспроизводства, изучаются одной из общественных наук — демографией.

Демография рассматривает население не только в статике, но и в динамике. Закономерности статических соотношений наилучшим образом можно осмыслить и воспринять на основе математических и математико-статистических уравнений или неравенств, связывающих одни факторы с другими. Закономерности же динамических процессов можно понять и изучить на основе математических уравнений, характеризующих развитие этих процессов во времени. Поэтому очевидно, что изучение демографических явлений и процессов требует хорошего знания математики и методов математической статистики. Легче научиться математике, чем научиться обходиться без нее.

Из решений XXIV съезда КПСС вытекает необходимость широкого применения математических методов в экономике и использования электронно-вычислительной техники для всемерного совершенствования статистики и управления социально-экономическими процессами страны.

Применение математических и математико-статистических методов в демографии дает возможность рассчитать продолжительность жизни, показатели рождаемости, плодовитости, смертности, естественного прироста, перспективные показатели воспроизводства населения, а также создать модели стационарного и стабильного населения, получить интегральные уравнения роста населения, исчислить брутто- и нетто-коэффициенты воспроизводства.

На Всесоюзном совещании статистиков в апреле 1968 г. начальник ЦСУ СССР член-корреспондент Академии наук СССР В. Н. Старовский, имея в виду применение математики, математической статистики, кибернетики и эвристических методов, говорил: «Сейчас новые методы должны найти широкое применение во многих отраслях экономической науки. Но положительных результатов можно достичь лишь при условии, что применение и совершенствование их будет всецело опираться на основные положения марксистско-ленинской экономической теории»[1].

[1] В. Н. Старовский. Важнейшие вопросы совершенствования государственной статистики.— В сб. «Всесоюзное совещание статистиков». М., «Статистика», 1969, стр. 19.

В каких же случаях и когда возникает, на наш взгляд, необходимость привлечения и использования математических и математико-статистических методов в демографии?

1. Когда попытки проникнуть в сущность демографических явлений затруднены сложностью самих явлений и многогранностью связей между ними. Тогда лишь графические конструкции и построения позволяют почти «на ощупь» рельефно и наглядно уловить качественное своеобразие этих явлений и дать им количественную оценку с помощью выведенных при этом математических формул.

2. Когда объектом исследования являются относительные величины, характеризующие интенсивность демографических процессов (коэффициенты), структуру, корреляционные связи, динамику (темпы роста и прироста) и др.

3. При выравнивании уже полученных демографических характеристик (численностей или показателей демографических таблиц).

4. Когда детализация приводит к необходимости использования групп такой численности, при которой нет уверенности в правильности показателей, а требуется обоснование достоверности получаемых при этом выводов.

5. Когда требуется доказать существенность или несущественность различий между характеристиками групп путем использования средних показателей и дисперсий.

6. При проверке той или иной гипотезы, подтверждение или опровержение которой должно быть основано на большом числе фактов.

7. Когда желают выявить объективные закономерности демографических процессов, освободиться от влияния случайностей на их развитие и построить модель этого развития.

8. При изучении динамических явлений, сезонных колебаний, изменения численности тех или иных демографических совокупностей.

9. При изучении формы и степени корреляционной связи между демографическими показателями и зависимостей между различными признаками.

10. Когда из ряда вариантов требуется выбрать такой, который наилучшим образом отвечает предъявляемым требованиям. Использование электронно-вычислительных машин при прогнозировании численности населения и моделировании процессов позволяет производить эти расчеты во многих вариантах и выбирать оптимальный из них.

11. При проектировании выборочных обследований в области демографии и для расчета степени репрезентативности этих обследований.

При этом на различных стадиях работы с демографическим материалом возникает необходимость привлечения тех или иных методов математики и математической статистики.

Мы не исчерпали, конечно, всех случаев привлечения математических и математико-статистических методов в демографии, и наша работа не претендует на раскрытие всех возможностей использования математики в демографии. Эта тема может быть раскрыта во многих

работах. В последнее время в демографии нашли применение и такие разделы математики, как теория стохастических процессов, теория информации, матричное исчисление и т. д. Мы поставили перед собой более скромную задачу: выявить возможности использования математики и математической статистики для изучения главным образом вопросов, связанных с порядком доживания и вымирания населения[1], его воспроизводства и интерполяцией демографических показателей.

Именно поэтому мы хотим в данной работе привлечь к анализу демографических явлений и процессов методы математики и математической статистики, которые мы считаем наиболее подходящими для этого, критически рассмотреть те из математических методов, которые использовались ранее, отобрать из них те, которые и в настоящее время могут быть использованы, отбросить явно непригодные или указать границы их возможного использования, изменить некоторые методы с учетом современных возможностей и т. д.

Так как в демографии отсутствует общепринятая, стандартная система обозначений (т. е. символика), иногда при изложении тех или иных идей приходится привлекать систему обозначений, использованную ее автором в оригинале, оговаривая и объясняя в каждом конкретном случае смысл и значение символов. Во многих случаях вместо подробного изложения теоретического материала в тексте дается ссылка на формулы, описанные в работах «Основы математической статистики», «Пособие по математической статистике», «Основы теории вероятностей и математической статистики»[2], изданных в соавторстве с Г. С. Кильдишевым.

Демография обязана высшей математике возникновением ряда методов исследования, важнейшим из которых является вычисление таблиц смертности (дожития). Построение таблиц смертности основывается на теории вероятностей, предполагающей знание дифференциального и интегрального исчисления. Теория вероятностей с самого начала своего возникновения была привлечена к разрешению вопросов, связанных с населением. Еще в Кодексе Юстиниана (528 г. до н. э.) имеется закон о продовольствии, который ясно показывает, что уже римляне занимались определением средней продолжительности предстоящей жизни в различных возрастах.

Ряд крупнейших ученых привлекал математику для решения интересующих их вопросов, относящихся к населению. Знаменитый французский математик Лаплас выдвинул идею и практически использо-

[1] Весьма ценные указания о возможностях использования математических методов, в частности теории вероятностей, при изучении вопросов, примыкающих к демографии — санитарной и медицинской статистики, можно найти в методическом пособии для врачей (см. А. М. Мерков. Общая теория и методика санитарно-статистического исследования. М., Медгиз, 1960, стр. 11—14).

[2] И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики. М., Госстатиздат, 1963; те же. Пособие по математической статистике. М., Госстатиздат, 1956; те же. Основы теории вероятностей и математической статистики. М., «Статистика», 1968.

вал теорию вероятностей для определения численности населения. Известно его исследование о брачности, рождаемости и смертности, где обосновывается определение численности населения страны (без переписи) путем деления общего числа рождений в стране на коэффициент рождаемости, исчисленный для какой-либо ее части.

Известный ученый Даниил Бернулли в своем трактате «О средней продолжительности браков при всяком возрасте супругов и о других смежных вопросах»[1] использует соотношения, выведенные им же в статье «Об употреблении алгоритма бесконечно малых в теории вероятностей»[2].

В заключение этой статьи Бернулли пишет: «Из этого моего опыта, какой он ни есть, становится ясным, что в роде человеческом происходит много вариаций и есть много взаимоотношений, которые можно определить обстоятельнее и лучше вычислениями, чем сделанными до сих пор бесчисленными наблюдениями»[3].

Много сделал французский математик Фурье для разработки математической модели, с помощью которой можно было бы изучать естественное и механическое движение населения.

Лексис, Цейнер, Кнапп создали теорию графических конструкций демографических показателей и привлекли математику для изучения порядка вымирания населения.

Американский демограф А. Лотка продолжил и углубил разработку методов Бека и сконструировал модели стабильного населения, позволяющие производить расчеты и оценивать степень замещения одних поколений другими. Большой известностью пользуются работы Граунта, Фарра, Борткевича и многих других.

Наши соотечественники академики В. Я. Буняковский, К. Ф. Герман и др. внесли большой вклад в изучение демографических явлений и процессов математическими методами. Обогатили советскую демографию своими работами С. А. Новосельский, В. В. Паевский, М. В. Птуха, Ю. А. Корчак-Чепурковский и др.

Б. С. Ястремский разработал дисперсионно-корреляционный метод изучения смертности населения и создал модель, связывающую показатели воспроизводства населения; А. Я. Боярский систематизировал демографические знания, создал один из эффективнейших методов построения таблиц смертности; Б. Ц. Урланис применил математические методы для доказательства ряда положений; А. М. Мерков и Г. А. Баткис разработали ряд методов для изучения здоровья населения.

---

[1] «De duratione media matrimoniorum, pro qualunque nonjugum aetate aliisque questionibus affinis, auctore Daniele Bernoulli. Conferatus specimen. De usu algorithmi infinitesimalis in arte conjectandi».— «Novi Commentarii Academiae Scientiarum Imperialis Petropolitani», tome XII, pro anno 1766—1768. Petropoli, 1768, p. 99—127.

[2] Там же.

[3] Цит. по М. Птуха. Очерки по истории статистики XVII—XVIII веков. Огиз. Государственное издательство политической литературы, 1945, стр. 301.

Что же позволило ввести смертность в область математических и математико-статистических исследований? Главная задача математики при изучении смертности — определение зависимости смертности от различных влияющих на нее причин, т. е. обоснование таких формул, в которых смертность фигурирует как функция влияющих на нее факторов. Конечно, при этом в качестве числовых данных, подлежащих обработке, для получения практически важных выводов привлекаются данные статистических наблюдений.

Статистические данные представляют собой результат совокупного воздействия ряда причин; действие каждой из них неизвестно. Путем группировок можно постепенно вводить в рассмотрение те или иные факторы.

Среди многих причин, имеющих наибольшее влияние на смертность, выделяется возраст человека. Именно поэтому внимание математиков и демографов было обращено в первую очередь на нахождение зависимости смертности от возраста при элиминировании влияния других факторов. Первым шагом к выявлению зависимости смертности от возраста является составление таблицы, в которой каждому конкретному значению возраста соответствует число смертных случаев.

Вторым шагом является нахождение закона изменения уровня смертности в зависимости от возраста, т. е. нахождение связи действующего фактора (возраста) с результативным (смертность).

Если найти такой показатель смертности, как вероятность умереть в определенном возрасте, то надо полагать задачу решенной.

Кроме возраста человека, имеется еще один фактор, связанный со временем, элиминировать влияние которого на смертность необходимо. Речь идет о времени (в данном случае моменте) наблюдения.

В статистических совокупностях, которыми оперирует исследователь, изучающий смертность, различаются две совокупности людей, остающихся в живых в различных возрастах: это, во-первых, *сверстники*, т. е. совокупность людей из данного поколения родившихся, переживших определенный возраст; во-вторых, *современники*, т. е. совокупность людей, достигших определенного возраста, но родившихся в разное время. Учитывая это, видим некоторую неопределенность при изучении указанных выше связей ввиду трудности устранения влияния времени. Именно поэтому одной из главных задач математики в приложении к смертности стало определение зависимости смертности одновременно от двух причин: возраста и времени, т. е. функции двух переменных.

Конечно, математические и математико-статистические построения и формулы не могут отразить всех аспектов каждого изучаемого демографического явления, и математическое понимание и описание явления не во всех деталях полностью совпадает с его объективной сущностью, но исследователь должен всегда стремиться максимально точно воспроизвести материальную природу описываемых явлений.

Глава 1 ПОРЯДОК ДОЖИВАНИЯ И ВЫМИРАНИЯ НАСЕЛЕНИЯ

ГРАФИЧЕСКИЕ КОНСТРУКЦИИ ИЗУЧЕНИЯ ДЕМОГРАФИЧЕСКИХ ПРОЦЕССОВ

ГРАФИКИ ДЛЯ ИЗУЧЕНИЯ СМЕРТНОСТИ И ДОЖИВАЕМОСТИ

Для детального измерения смертности, для придания демографическому материалу большей наглядности, для характеристики того, как единичные факты объединяются в группы, а из этих групп составляются статистические совокупности, служат геометрические построения, называемые демографическими сетками (или решетками) и представляющие собой первый координатный угол с осью времени ot и осью возраста ox. Часть этих построений основана на работах Лексиса, Вестергарда и теории графических построений Кнаппа. Элементы метода изохрон были в 1874 г. разработаны английским демографом Анселем, а позже теоретически обоснованы Лексисом.

Рассмотрим идеи такого рода графических конструкций и укажем современные возможности их практического использования.

Геометрические построения, использующие систему координат, позволяют четче проследить логическую необходимость определенных соотношений.

Изображение на плоскости (демографическая сетка)

Момент рождения человека фиксируется на оси ot, а длительность его жизни отмечается перпендикуляром — линией жизни. Конец линии жизни называют смертной точкой. На график линии жизни не наносят, а представляют мысленно. Смертные точки на плоскости указывают, когда родились умершие и в каком возрасте они умерли.

Разделим ось времени ot и ось возраста ox на одинаковые интервалы, соответствующие, например, одному году. Тогда график примет вид разграфленного на квадраты поля (см. рис. 1).

Возьмем линию жизни $\mu\gamma$. Она показывает, что человек родился в 1949 г. и умер в возрасте 6 лет. Горизонтальная линия, например $\varphi\gamma$, отмечает на линии жизни определенные моменты возраста и называется линией возраста.

Обозначим момент наблюдения — z, возраст человека — x, а момент рождения — t, тогда $t + x = z$.

Если установить момент наблюдения и принять z величиной постоянной, тогда $x = -t + z$, т. е. мы получаем уравнение прямой линии, образующей с осью ot угол 135°.

Такая прямая линия называется *изохроной* или *линией наблюдения*.

Уравнение $t+x=z$ может быть изображено при помощи поверхности Цейнера (см. рис. 2).

Рассмотрим 4 вида кривых на этом графике.

1. Кривая AB называется *кривой рождений*, выражается уравнениями $x = 0$ и $y = \varphi(t, x)$ и показывает изменение числа рождений от момента времени t_1, соответствующего на графике точке A_1, до момента времени t_2, соответствующего на графике точке B_1. Значит, кривая AB изображает линию, по которой поверхность смертности пересекается с плоскостью.

2. Кривая AC_2 называется *кривой переживания*, выражается уравнениями $t = t_1$ и $y = \varphi(t, x)$ и показывает изменение числа родившихся в момент времени t_1, соответствующий на графике точке A_1, с изменением возраста родившихся. Эта кривая показывает порядок переживания лиц данного момента рождения.

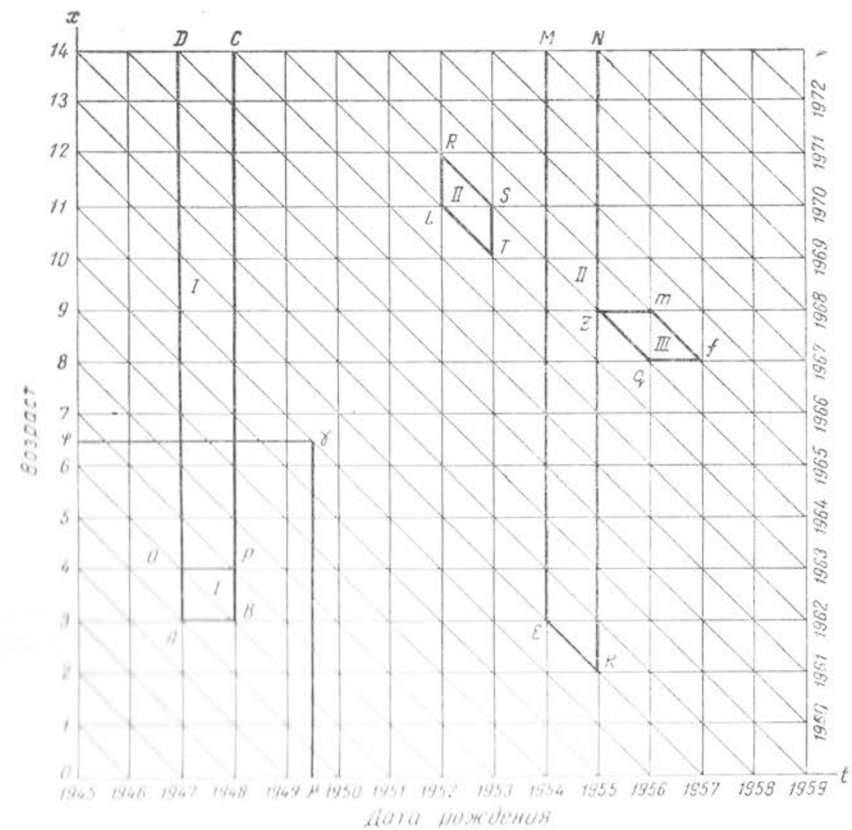

Рис. 1. Демографическая сетка.

3. Кривая C_2D — *кривая предельного возраста, или пределов долголетия*, выражается уравнениями $y = 0$ и $\varphi(t, x) = 0$ и показывает предельный возраст, до которого доживают лица разных моментов рождения.

4. Кривая EC — *кривая наличного населения* выражается уравнениями $t + x = 0$ и $y = \varphi(t, x)$ и показывает наличное население на данный момент.

Надо заметить, что такое усложненное представление разными кривыми рождений, переживания, предельного возраста и наличного населения не очень наглядно. Практически использование стереометрических моделей имеет большое преимущество для людей с хорошим пространственным воображением.

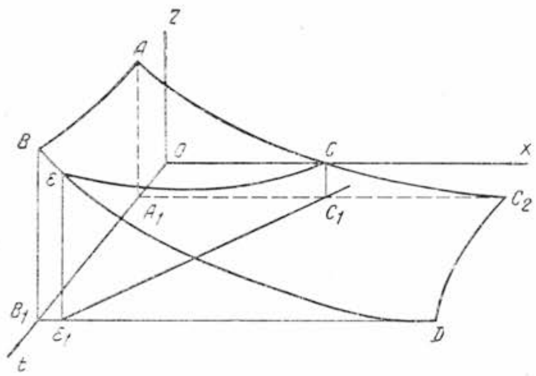

Рис. 2. Поверхность Цейнера.

Возвращаясь к нашему примеру, определяем, что человек, родившийся 1 июля 1949 г. и умерший в возрасте 6,5 года, умер 1 января 1956 г.

Демографическая сетка дает возможность с помощью трех систем параллелей (линий жизни, линий возраста, линий наблюдения) изучить различные группы людей как живущих, так и умерших. Эти группы людей называют совокупностями живущих и совокупностями умерших. Пересечение линий жизни с линиями возраста или с линиями времени дает определенные категории живых — *сверстников* и *современников*.

Под сверстниками понимают людей, находящихся в данном возрасте в определенный момент. Родившиеся одновременно — это тоже сверстники.

С помощью переписи населения мы получаем современников, которые по сути дела представляют собой совокупность живущих в один момент времени. Ни одна из этих совокупностей не может быть определена на основании другой, ибо очевидно, что число живущих в одно время не может быть определено числом живущих в разные времена. Если дано время наблюдения современников или возраст, которого достигают сверстники, то, чтобы вполне определить каждую совокупность, нужно еще знать промежуток времени, в течение которого родились эти лица. Значит, нужно знать начало периода, в течение которого наблюдались рождения, и конец его. Следовательно, для полной определенности каждой совокупности нужно знать три величины, из которых время наблюдения или возраст определяют принадлежность лиц к той или иной совокупности, а остальные две служат для определения времени рождения ее членов.

Таблица 1
Пределы возраста и времени рождения

Показатели	Пределы	
	низший	высший
Возраст	x_1	x_2
Время рождения	t_1	t_2
Момент наблюдения	z	

Для определения времени рождения достаточно две величины из четырех — x_1, x_2, t_1, t_2. Путем различных их сочетаний (t_1t_2, x_1x_2, t_1x_1, t_2x_1, t_1x_2 и t_2x_2) и присоединения к ним z мы получим шесть возможных совокупностей современников, живущих во время t.

Вообще говоря, при наличии шести пределов можно подсчитать число всех совокупностей людей, образованных этими пределами.

Получаем числа сочетаний $C_6^6=1$; $C_6^5=6$; $C_6^4=C_6^2=15$; $C_6^3=20$.

Прежде всего отметим, что не все полученные при этом совокупности имеют практическое применение. Некоторые геометрические фигуры по самой форме свидетельствуют, что им (т. е. данной совокупности людей) не может быть приписан статистический смысл. Иными словами, не существует ни теоретических, ни практических оснований для производства сводки и группировки первичного материала подобным образом.

Возьмем на демографической сетке (см. рис. 1) фигуру $ABCD$, представляющую собой вертикальную полосу. Эта совокупность, состоящая из лиц, родившихся в течение определенного промежутка времени t_1t_2 и переживших возраст x лет. Эта совокупность может быть определена числом линий жизни, пересекающих линию возраста в соответствующих точках. Такая совокупность называется *совокупностью живущих I рода*. В нашем примере это совокупность переживших возраст 3 года из родившихся между 1 января 1947 г. и 1 января 1948 г. Эта группа лиц одного поколения, доживших до определенного возраста, обозначается l_x. Некоторые демографы, например М. В. Птуха, в этой совокупности выделяют 3 группы: 1) все ровесники в возрасте x лет — l_x; 2) ровесники в возрасте x лет из поколения родившихся ${}_{n_t/n_x} - {}_{n_x/n_x}l_x$; 3) ровесники в возрасте x лет, наблюдаемые за время от t до $t+a$ — ${}_{t/t+a}l_x$.

Таким образом, совокупность живущих I рода, состоящая, следовательно, из числа лиц, доживающих до возраста x лет из числа родившихся от t_1 до t_2, равна:

$$l_x=\int_{t_1}^{t_2} f(x,t)\,dt.$$

При этом должно быть ясно, что достижение одного и того же возраста x лет лицами из совокупности l_x будет происходить не одновременно, а в промежутке времени от $t_1 + x$ до $t_2 + x$.

Если взять в полосе $ABCD$ один квадрат, например $ABPO$, и подсчитать в нем число смертных точек, то это будет группа умерших, определяемая границами поколения и границами возраста смерти. Такая группа людей называется *совокупностью умерших I рода*. Она состоит из лиц, родившихся в одном календарном году (t_1t_2) и умерших в одном возрасте (x_1x_2), и обозначается M^1. В нашем примере это совокупность умерших в возрасте трех лет из родившихся в 1947 г. Отношение числа умерших в возрасте от $x - 1$ до x лет к числу лиц, доживших до $x - 1$ лет, дает для рассматриваемого поколения эмпирическое значение теоретически правильно построенной вероятности умереть в течение x-го года жизни: $q_x = \frac{M_x^1}{l_{x-1}}$.

Совокупность умерших II рода определяется периодом вымирания и поколением. Если на демографической сетке взять фигуру $LRST$, то число смертных точек, заключенных в ней, будет состоять из умерших в течение одного календарного года z_1z_2 из родившихся в одном году t_1t_2. На сетке эта фигура представляет параллелограмм, у которого две стороны параллельны линиям жизни, а две другие — изохронам. Эта совокупность может быть обозначена M^2.

Совокупность умерших III рода представляет собой совокупность умерших в течение одного календарного года z_1z_2 на одном году жизни x_1x_2 (см. фигуру $qfmz$). В нашем примере она состоит из умерших в 1964 г. в 8-летнем возрасте. Из графика видно, что это параллелограмм, у которого две стороны параллельны линиям возраста, а две другие — изохронам. Следовательно, все умершие не принадлежат к одному поколению, а состоят из родившихся в 1955 и 1956 гг. Эта совокупность может быть обозначена M^3.

Большая часть статистических данных об умерших дается в виде какой-либо одной совокупности, чаще всего совокупности умерших III рода. Для вычисления совокупностей умерших одного рода путем использования совокупности умерших другого рода можно вывести целый ряд интерполяционных формул (некоторые из них в дальнейшем мы приведем и используем).

Имеется еще *совокупность живущих II рода*, которая состоит из людей, родившихся в одному году t_1t_2 и переживших один календарный момент.

В нашем примере эта совокупность изображена фигурой $MEKN$; в нее входят родившиеся в 1954 г. и пережившие 1 января 1957 г. Отличие этой совокупности от совокупности живущих I рода состоит в том, что если в совокупности живущих I рода речь идет о сверстниках, то в совокупности живущих II рода речь идет уже о современниках.

Эта совокупность может быть получена в итоге переписи (с распределением живущих по состоянию на критический момент по году рождения или по возрасту) и обозначена L_x. Ее можно расчленить на три совокупности:

1) совокупность современников в момент времени наблюдения вообще; 2) из поколения n_2/n_3; 3) в возрасте от x до $x + a$.

Все лица, входящие в состав данной совокупности, могут находится в различных возрастах в промежутке от $x_1 = z - t_2$ до $x_2 = z - t_1$.

Следовательно,

$$L_x = \int_{t_1}^{t_2} f(z - t_1, t)\, dt$$

или

$$L_x = \int_{x_1}^{x_2} f(x, z - x)\, dx.$$

Тогда это совокупность лиц, достигших момента z в промежутке возраста x_1x_2.

Рассмотрим соотношения, определяющие смертные случаи, входящие в состав различных совокупностей умерших. Нам известно, что $t + x = z$. Найдем продолжительность периода, в течение которого произошли смертные случаи. Для совокупностей умерших I рода имеем: $t_1 + x_1 = z_1$ и $t_2 + x_2 = z_2$. Тогда $z_2 - z_1 = (t_2 + x_2) - (t_1 + x_1) = (t_2 - t_1) + (x_2 - x_1)$. Получаем, что календарный период равен сумме двух интервалов — во времени и в возрастах, т. е. в нашем примере равен двум годам. Для совокупностей умерших II рода найдем границы возрастного интервала, в течение которого умирали люди, входящие в состав этой совокупности: $x_1 = z_1 - t_2$; $x_2 = z_2 - t_1$. Тогда $x_2 - x_1 = (z_2 - t_1) - (z_1 - t_2) = (z_2 - z_1) + (t_2 - t_1)$. Следовательно, возрастной интервал равен сумме интервала рождения и интервала смерти. Значит, имеем возрастной интервал в два года. Для совокупности умерших III рода соотношения при определении границ рождения могут быть найдены следующим образом: $t_1 = z_1 - x_2$; $t_2 = z_2 - x_1$. Откуда $t_2 - t_1 = (z_2 - x_1) - (z_1 - x_2) = (z_2 - z_1) + (x_2 - x_1)$, т. е. двум годам.

Сформулируем теперь общее правило, устанавливающее соотношения между годом рождения, возрастом и периодом наблюдения и основывающееся на определении каждой совокупности: *из трех периодов незаданный период каждой совокупности умерших равен сумме двух заданных периодов*.

Изохроны дают возможность выделить более мелкие совокупности. Дело в том, что каждая из трех совокупностей умерших состоит из двух прямоугольных треугольников. Смертные точки, расположенные в каждом прямоугольном треугольнике, составляют *элементарные совокупности умерших*.

Характерной особенностью элементарных совокупностей является то, что на каждый из трех признаков, определяющих совокупность (время рождения, возраст и период наблюдения), приходится по равному промежутку времени; во всех же трех главных совокупностях умерших на один какой-нибудь элемент приходится двойной промежуток.

Треугольники, охватывающие элементарные совокупности, расположены по разные стороны изохрон, образующих их гипотенузу. Так, часть изохроны BO создает две элементарные совокупности BAO и BOP (см. рис. 1).

В зависимости от того, как расположены треугольники — над изохроной или под ней, — различают *верхние* и *нижние* элементар-

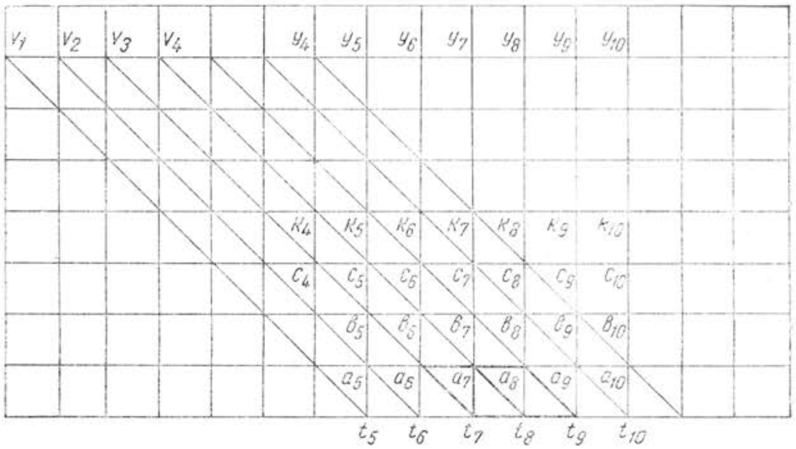

Рис. 3. Прямой учет совокупностей умерших II рода.

ные совокупности умерших: Δ^1 или $\Delta_{н}$ нижние и Δ^2 или $\Delta_{в}$ верхние. Общая формула при этом примет вид:

$$_tM_x^1 = {_t\Delta_x^1} + {_t\Delta_x^2},$$

где t — поколение, а x — возраст. Для нижней элементарной совокупности (Δ^1) год рождения равен разности года смерти и возраста смерти. Для верхней элементарной совокупности (Δ^2) год рождения на единицу меньше этой же разности. Элементарные совокупности связаны определенным образом с совокупностями живущих. Возьмем, например, совокупности живущих I рода (прямоугольник $ABCD$). Число лиц, входящих в эту совокупность, равно сумме числа линий, пересекающих OB, и числу смертных точек в пределах элементарной совокупности ABO:

$$_tl_x = {_tL_x} + {_t\Delta_x^1}.$$

Первое слагаемое — совокупность живущих II рода (L) — дается переписью населения на момент z. Следовательно, в конечном итоге вопрос о получении совокупностей живущих и умерших сводится к определению элементарных совокупностей. Эти элементарные сово-

купности могут быть учтены непосредственно или путем приближенных вычислений. Для непосредственного учета элементарных совокупностей достаточно группировать смертные случаи каждого календарного года по возрасту и по году рождения умерших. Для косвенного учета нужно группировать смертные случаи и по времени рождения и по возрасту смерти, но не в сочетании друг с другом, т. е. непосредственно учитывать совокупности умерших M^2 и M^3.

Для первой возрастной группы (рис. 3) дается численность умерших $M(t_7t_8a_6a_7)$, в которую входит элементарная совокупность $M(t_7a_6a_7)$. Разность между ними — элементарная совокупность $M(t_8t_7a_7)$. Статистика следующего календарного года дает непосредственно $M(t_8t_9a_7a_8)$ и $M(t_8t_9a_8)$, откуда вычитанием получаем элементарную совокупность $M(t_8a_7a_8)$. Далее могут быть разложены на элементарные совокупности $M(a_6a_7b_5b_6)$ и $M(a_7a_8b_6b_7)$, установленные путем прямого учета, и т. д.

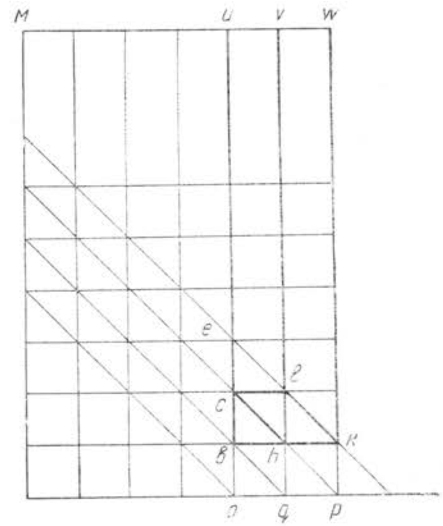

Рис. 4. Совокупности живущих и умерших.

Если же регистрация смертных случаев производится по элементарным совокупностям III рода, то, как показывает опыт, начиная с трехлетнего возраста, элементарные совокупности определяются достаточно точно путем простого деления совокупности M^3 пополам, ибо плотность распределения смертных точек для старших возрастов изменяется очень медленно. Для первых же двух лет жизни этот способ вообще не годится. Такое же неравенство элементарных совокупностей наблюдается и в преклонном возрасте старше 70—75 лет.

Для детей до двух лет приближенный подсчет может производиться методами, излагаемыми ниже, которые выводятся математически при разных предположениях.

Наибольшее практическое значение, очевидно, имеют те из выводимых методов, при которых интервалы по осям координат равны друг другу и приняты за один год.

Рассмотрим на графике соотношения между двумя совокупностями живущих и тремя совокупностями умерших (рис. 4).

Если MW — линия предельного возраста, то число смертных точек в четырехугольнике $oquv$ представляет полный итог поколения родившихся в течение времени oq и вымерших в естественных пределах человеческой жизни.

Тогда

$$bhc = bhuv - hcuv,$$
$$chl = chuv - cluv,$$
$$bhcl = bhuv - cluv,$$
$$chel = chuv - eluv,$$
$$hkcl = hkvw - klvw + hcuv - cluv.$$

Последнее соотношение показывает, что умершие в известном возрасте в течение известного времени составляют разницу между современниками, находящимися в данных возрастах у верхнего и нижнего предела времени наблюдения с добавлением разности сверстников, достигающих в течение времени наблюдения низшего и высшего пределов возраста.

Изображение в пространстве

Применяется также более строгое, но менее наглядное (мы уже это видели на примере с поверхностью Цейнера) изображение в трехосной системе координат, т. е. в пространстве.

Пусть совокупность лиц, родившихся в течение некоторого промежутка t и достигших возраста x лет, выражается непрерывной функ-

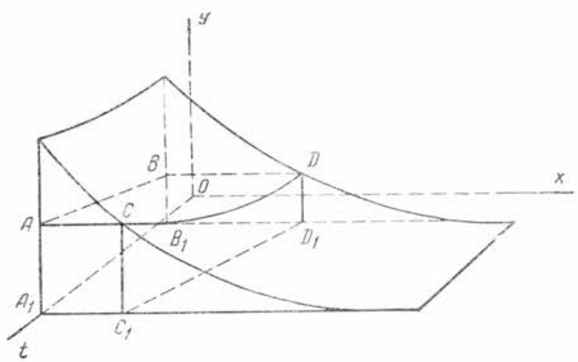

Рис. 5. Лица, достигшие возраста x лет из родившихся в период от t_1 до t_2.

цией $F = f(t;\ x)$. Данная функция имеет производную по t и по x. Совокупность родившихся в интервале от t до $t + dt$ и доживших до возраста x можно выразить следующим образом:

$$f(t + dt, x) - f(t, x) = f'(t, x)\,dx = \varphi(t, x)\,dx.$$

А совокупность родившихся в промежутке от t до $t + dt$ и умерших в возрасте от x до $x + dx$ составит

$$[\varphi(t, x + dx) - \varphi(t, x)] - \varphi'(t, x)\,dx = \psi(t, x).$$

Если $x = F(t)$ представляет собой уравнение кривой C_1D_1 (рис. 5), а точка D_1 этой линии удовлетворяет уравнению $t = t_1$, то интеграл

функции $\varphi(t;\ x)$, распространенной по дуге этой кривой, даст число лиц, достигших возраста $x = F(t)$ из родившихся в период времени от t_1 до t_2. Это число лиц выражается проекцией фигуры CDC_1D_1 на плоскость yot. Проекция примет вид фигуры ABA_1B_1.

На рис. 6 дано графическое изображение совокупности лиц, родившихся в интервале времени от t_1 до t_2 и умерших в возрасте $x = F(t)$.

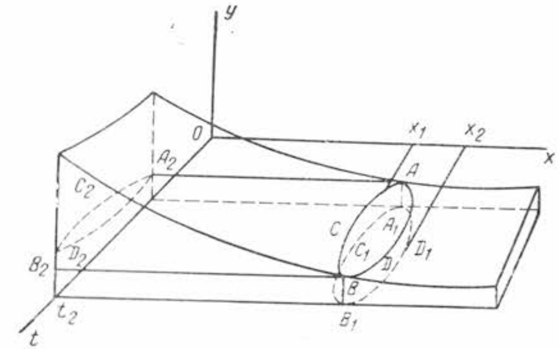

Рис. 6. Лица, родившиеся в интервале от t_1 до t_2 и умершие в возрасте x лет.

На графике эта совокупность выражается проекцией части поверхности Цейнера, ограниченной кривой $ACBD$, на плоскость yot. Эта проекция примет форму фигуры $A_2B_2C_2D_2$.

В заключение рассмотрим изображение совокупности лиц, родившихся в интервале от t_1 до t_2 и доживших до возраста x лет (рис. 7).

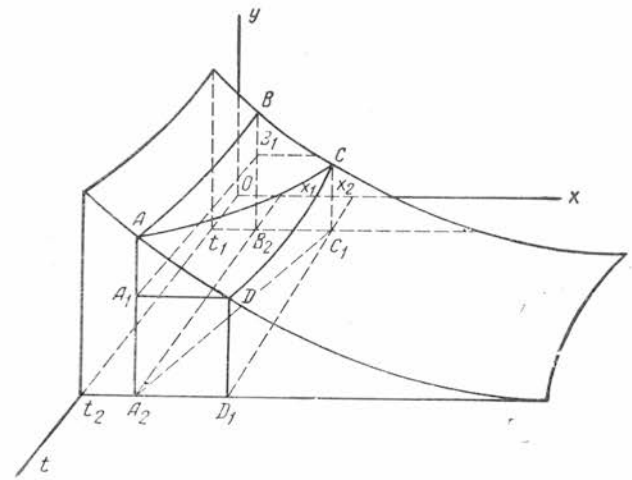

Рис. 7. Лица, родившиеся в интервале от t_1 до t_2 и дожившие до возраста x лет.

Проекция линии AB (первой главной совокупности доживающих до возраста x) на плоскость xot есть линия A_2B_2, описываемая уравнением $x = x_1$. Лица, соответствующие ординате точки B_2 этой проекции, относятся к $t_1 = z_1 - x_1$ году рождения. Аналитически эта совокупность лиц выражается уравнением:

$$V_x = \int_{t_1}^{t_2} \varphi(t, x)\,dx.$$

На этом же графике AC — вторая главная совокупность доживающих (наличное население), т. е. родившихся в интервале от t_1 до t_2. Проекция этой линии на плоскость xot, выражаемая уравнением $t + x = 0$, есть линия A_2C_1. Лица, соответствующие ординате точки C_1, родились в $t_1 = z_1 - x_2$ году. Их возраст $x_2 = z_1 - t_1$.

Лица, соответствующие ординате точки A_2, родились в $t_2 = z_2 - x_2$. Аналитически наличное население, получаемое при переписи, выражается уравнением:

$$V_z = \int_{t_1}^{t_2} \varphi(t, z - t)\,dt.$$

ИСПОЛЬЗОВАНИЕ ЛИНИЙ ЖИЗНИ ДЛЯ ИЗУЧЕНИЯ ВЛИЯНИЯ ДЕТСКОЙ СМЕРТНОСТИ НА РОЖДАЕМОСТЬ

При изучении сдвигов в рождаемости и детской смертности в связи с каким-нибудь фактором, например переселением в новые дома, могут быть использованы графики.

Изобразим обследуемый коллектив в виде множества линий жизни, составляющих прямоугольники (рис. 8).

Точка означает рождение ребенка по соответствующему возрасту матери. При такой графической конструкции вся совокупность родившихся распределяется на две группы: родившиеся до свершения какого-нибудь события и родившиеся после него.

Таблица 2

Распределение женщин по времени наступления события

Возраст наступления признака	Период до промышленного труда			Период промышленного труда		
	вступило в брак	прекратило брак	число проведенных брачных лет	вступило в брак	прекратило брак	число проведенных брачных лет
	1	2	3	4	5	6
16	41	—	41	14	—	14
17	55	2	94	14	5	23
18	58	10	142	25	5	43
и т. д.						

Коэффициент рождаемости для периода до свершения изучаемого явления и после него получается путем отнесения числа рождений (точек на линиях жизни) к числу прожитых лет соответственно до свершения события и после него.

Советский исследователь Г. А. Баткис проводил с помощью такого метода изучение влияния труда женщин в промышленности на их

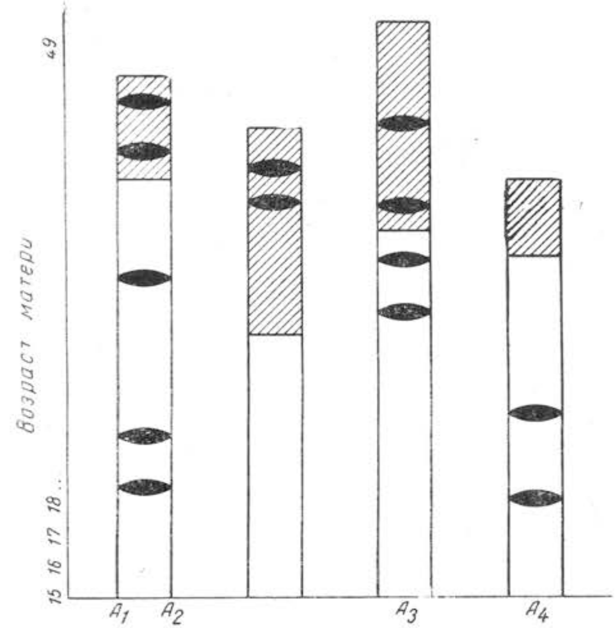

Рис. 8. Даты рождения, смерти и длительность изучаемого явления.

брачную плодовитость[1]. В качестве первого этапа была получена таблица, дающая распределение женщин, вступивших в брак и прекративших его, по их возрасту в момент наступления данного события для каждого периода труда в отдельности (табл. 2).

На втором этапе на основе этих данных и данных о числе родов производилось исчисление коэффициентов повозрастной брачной плодовитости по пятилетним группам (табл. 3).

Сопоставление коэффициентов брачной плодовитости позволяет делать определенные выводы о влиянии промышленного труда на брачную плодовитость.

Г. А. Баткис ввел в практику демографических исследований особый прием изохронных полигонов, основанный на том, что демографические явления рассматриваются в трех измерениях: так, умершие

[1] См. Г. А. Баткис. Анамнестический метод в демографической статистике.— В сб. «Проблемы демографической статистики». М., Госстатиздат, 1959.

Таблица 3

Влияние промышленного труда на брачную плодовитость

Возраст наступления признака	Период до промышленного труда			Период промышленного труда		
	число брачных лет	число родов	коэффициент брачной плодовитости (в %)	число брачных лет	число родов	коэффициент брачной плодовитости (в %)
16—19	463	160	34,5	130	43	33,0
20—24	1 048	466	44,5	393	154	39,3
25—29	786	325	41,4	310	138	44,5
и т. д.						

Таблица 4

Повозрастное распределение женщин, вступающих в брак

Возраст в момент обследования	Возраст в момент вступления в брак или прекращения брака*									Состоящие в браке	
	18	19	20	21	22	23	24	25	и т. д.	на 1945 г.	на 1955 г.
18(1955)	1	—									1
19	3										3
20	3	3	1								6
21	3	2	2/1	2							6
22	6	4/1	5	2/1	1						15
23	5	5	4	3							17
24	3	4	7	2	2/1	2/1					20
25	4	4	5	1	3/2		1				16
26	5	7	7	4	2	2	2				29
27	4	8	7	6	2	3/2	1				29
28(1945)	2	4	4	1	2	1	1				15
29	2	5	5	2	1	2/2				2	15
30	2	3	3	3/1	3	2	1/2			5	14
31	1	2/1	4	1	1	2/1	2			6	11
32	—	2	3	1	2	1	1/2			6	8
33	2	4	5	5/1	3	1				18	19
34	1	3	4	4/1	2	1/2	1			12	13
	и т. д.									и т. д.	и т. д.
В 1945—1954 гг. вступило в брак	38	48	49	25	18	16	10				
Прекратило	—	1	1	2	3	6	4				

* Числа вступивших в брак указаны в числителе, а числа прекративших брак— в знаменателе.

в определенный момент рассматриваются по датам рождения и возрасту смерти. Прием трехнаправленного изучения применен Г. А. Баткисом в анамнестических наблюдениях при изучении распределения рождений по возрасту женщин в момент обследования (подлежащее) и по возрасту в момент рождения ребенка (сказуемое). Эти два показателя, будучи расположены определенным образом в таблице, дают возможность диагонального изучения этих явлений по хронологическим датам.

Аналогично могут быть применены изохроны для вычисления по периодам прожитых брачных лет. В качестве примера приведем таблицу распределения женщин по возрасту в момент обследования (1955 г.) и по возрасту вступления в брак и его прекращения (табл. 4).

В этой таблице для каждого возраста подсчитывают число состоящих в браке к концу периода. Так, число 33-летних, состоящих в бра ке в 1945 г., получается по горизонтали возраста 33 года (на момент обследования) путем суммирования вступивших в брак до 1945 г. и вычитания случаев прекращения брака:

$$2+4+5+5+3-1=18 \text{ и т. д.}$$

На основании такой таблицы составляется другая, позволяющая вычислить число лет, прожитых в браке, для отдельных хронологических периодов (табл. 5)

Таблица 5

Исчисление брачного периода (число прожитых лет) для хронологического периода

Возраст	Число женщин данного возраста, состоящих в браке на начало периода (1945 г.)	Число женщин, вступивших в брак в периоде (1945—1954 гг.) в данном возрасте*	Число женщин, прекративших брак в период 1945—1954 гг. в данном возрасте	Число женщин данного возраста, состоящих в браке к концу периода (к 1955 г.)	Число лет прожитых в браке в данном возрасте в периоде (1945—1954 гг.)
1	2	3	4	5	6
18	—	38	—	—	38
19	2	48	1	3	84
20	5	49	1	6	131

* Данные этой колонки, представляющие собой изохронные полигоны, суммируются в пределах изохронной площади (см. табл. 4).

Данные этой таблицы определяются следующим образом: погодно для каждого возраста последовательно суммируются перешедшие в состоянии брака из предыдущего периода (графа 2) и вступившие в брак в данном возрасте и в данном периоде (графа 3), последовательно вычитаются все прекратившие брак (графа 4) и переходящие в состоянии брака в следующие периоды (графа 5).

Так, число лет прожитых в браке в возрасте 19 лет, для периода 1945—1954 гг. равно $38+2+48-1-3=84$ и т. д.

ИЗОБРАЖЕНИЕ ПОКОЛЕНИЙ

Большое значение в демографии имеют специально разработанные методы графических изображений численности каждого живущего поколения, числа одновременно живущих поколений и расслоения населения на поколения.

Один из таких методов был предложен В. Фарром, изучившим динамику показателей естественного движения населения Англии за столетие (1758—1857 гг.), и А. Лоткой[1].

Рассмотрим эти методы с некоторыми изменениями, которые нам кажется необходимым в них внести. Проследим последовательность поколений по женской линии, так как статистические данные обычно позволяют это делать (см. рис. 9).

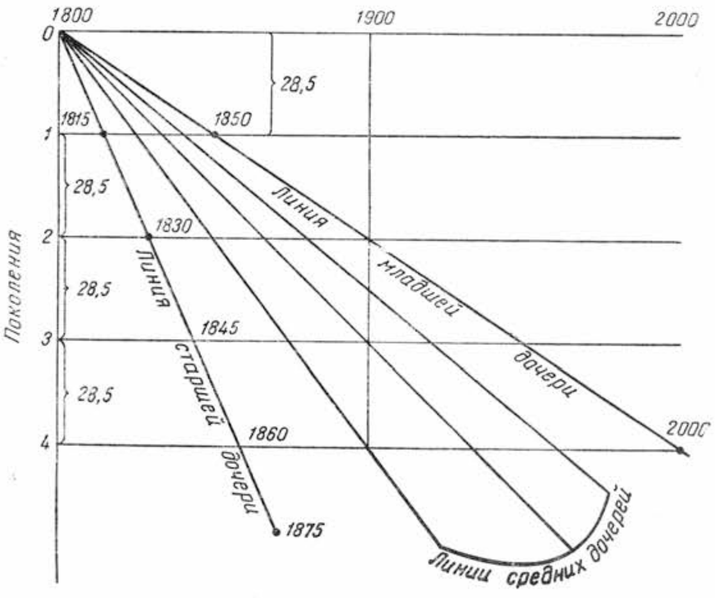

Рис. 9. Последовательность поколений по женской линии.

По оси абсцисс откладываются годы, по оси ординат — последовательные поколения, средняя длина которых принята равной 28,5 года. Нулевое поколение совпадает с осью x и представляет собой совокупность женщин, родившихся в 1800 г. Их дочери образуют первое поколение, внучки — второе и т. д.

От женщин нулевого поколения исходят дочерние линии. Имея в виду плодовитый возраст 15—50 лет, допустим (условно), что в каждой из дочерних линий дочь, внучка и правнучка впервые рожают в том же возрасте, в каком родила мать из нулевого поколения. Тогда по линии старшей дочери длина поколения равна минимуму, т. е. 15 лет, дочь появится в 1815 г., внучка — в 1830, правнучка — в 1845 и праправнучка — в 1860 г. Для младшей дочери таким периодом будет 50 лет и соответствующие поколения появятся в 1850, 1900, 1950 и 2000 гг.

[1] A. L o t k a. The spread of generations.—«Human biology», 1929, September.

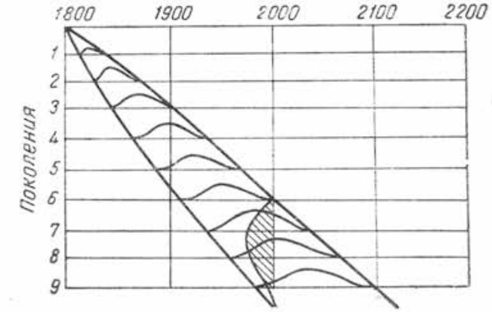

Рис. 10. Распределение рождений в десяти последовательных поколениях.

Возьмем теперь ординату 1900 г., которая пересекает разрез четырех разных поколений, живущих в 1900 г. На каждом отрезке времени нарождения любого поколения, например 1815—1850, 1830—1900, 1845—1950 и т. д., можно построить кривую распределения всех представителей поколения по годам рождения. Кривые распределения рождений в десяти последовательных поколениях даны на рис. 10.

Пользуясь этим графиком, можно выявить численность одновременно родившихся по поколениям. Возьмем 2000 г. за наблюдаемый. Его ордината пересечет отрезки времени, охватывающие рождение 6, 7, 8 и 9-го поколений. Возрастное распределение всего населения — потомства в 2000 г. отмечено заштрихованной площадью. Принимая численность населения в 2000 г. за 100% и выдвигая некоторые предположения, А. Лотка вычислил удельные веса численностей каждого поколения: 6-е — 12,5%; 7-е — 62,1%; 8-е — 24,3%; 9-е — 1,1%.

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И ИСХОДНЫЕ ФОРМУЛЫ ДЛЯ ХАРАКТЕРИСТИКИ ПОРЯДКА ВЫМИРАНИЯ. ПОКАЗАТЕЛИ ТАБЛИЦ СМЕРТНОСТИ (ДОЖИТИЯ)

В таблицах смертности (иногда называемых также таблицами дожития или средней продолжительности жизни) дается система показателей, связанных между собой определенными соотношениями. Например, в таблицах смертности 1926—1927 гг. вычислены восемь показателей: d_x; l_x; q_x; p_x; L_x; T_x; e_x^0; V_x. В таблице смертности 1958—1959 гг. вычислено только 7 показателей (без V_x).

ЧИСЛА УМЕРШИХ

Первым показателем является x, означающий точный возраст, до которого доживает та или иная часть наблюдаемого поколения. При $x = 0$ речь идет о новорожденных; при $x = 1$ — о находящихся в возрасте один год и т. д.

Начнем рассмотрение с чисел, характеризующих порядок вымирания по годам возраста, — d_x, представляющих частоты распределения людей по продолжительности жизни. Иными словами, d_x — это доля людей, умерших в возрасте x лет (на $x+1$ году жизни) из совокупности родившихся. Значит, речь идет о доле людей, умерших в возрастном интервале от x до $x+1$ лет. Если считать ω предельным возрастом и выражать d_x в долях единицы, то

$$\sum_{x=0}^{x=\omega-1} d_x = 1.$$

Можно рассматривать d_x так же, как порядок вымирания в стационарном населении.

ЧИСЛА ДОЖИВАЮЩИХ

Если Σd_x, равную единице, обозначить l_0 и понимать под этим начальное число родившихся, из которого исходит таблица смертности, то

$$l_0 = 1 = d_0 + d_1 + d_2 + \ldots + d_{x-1} + d_x + \ldots + d_{\omega-1}.$$

Отсюда

$$l_1 = l_0 - d_0 = 1 - d_0$$
$$l_2 = 1 - d_0 - d_1$$
$$\cdots\cdots\cdots$$
$$l_x = 1 - d_0 - d_1 - \ldots - d_{x-1}.$$

Примем начальное число родившихся $l_0 = 100\,000$. Как видно, l_x — убывающий ряд чисел. Если бы был известен закон убывания (т. е. формула), то по этому закону можно было бы найти число лиц, доживающих до любого возраста. Такая формула могла бы быть найдена путем наблюдения за порядком вымирания какого-либо поколения. Практическое осуществление такого наблюдения требует весьма продолжительного времени. Кроме того, практика показала, что порядок доживания человеческого населения не подчиняется какому-либо простому закону, который можно было бы выразить алгебраической формулой. Для получения убывающего ряда l_x прибегают к иным способам.

Так как числа доживающих выражают порядок вымирания известной совокупности людей, следовательно,

$$l_\omega = 0;\ l_x = d_x + d_{x+1} + \ldots + d_{\omega-1} = \sum_{i=x}^{i=\omega-1} d_i;$$
$$d_x = l_x - l_{x+1},\ \text{или}\ l_{x+1} = l_x - d_x,\ \text{или}\ l_x = l_{x+1} + d_x.$$

Если x рассматривать не как целое число, а как непрерывный признак, то $l(x)$ это уже непрерывная функция возраста, т. е. аргумента x. График $l(x)$ есть монотонно убывающая кривая дожития (см. рис. 11).

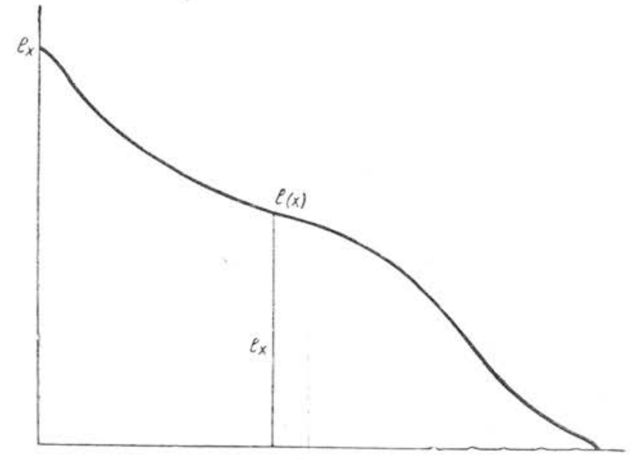

Рис. 11. Линия дожития.

Интересна попытка Фурье использовать кривую жизненности для понимания меры смертности. Так как кривая показывает число остающихся в живых в последующие возраста, он предположил, что с некоторого возраста $x = OE$ (см. рис. 12), числа остающихся в живых уменьшаются с увеличением возраста в арифметической прогрессии, т. е. числа умерших в течение равных промежутков равны между собой. Тогда кривая ab от точки c могла бы быть заменена прямой ch, являющейся касательной к кривой смертности в точке c. Через один год после c осталось бы в живых число лиц, соответствующих ординате $fd = EQ$.

Отрезок же cQ выразил бы число умерших в течение взятой единицы времени до достижения полного возраста $x = of$. Значит, отношение $\frac{cQ}{cE}$ можно считать мерой смертности в момент достижения возраста $x = OE$ остающимися в живых из данного числа новорожденных.

Таким образом, Фурье считал, что $\frac{cQ}{cE}$ есть мера смертности. А $cQ = Qd\ \mathrm{tg}\,Qdc = Ef\ \mathrm{tg}\,Ehc = \frac{dy}{dx} = y'$.

Принимая Ef равным единице времени, он получил, что мера смертности выражается отношением

$$\frac{cQ}{cE} = -\frac{y'}{y},$$

а это и есть сила смертности.

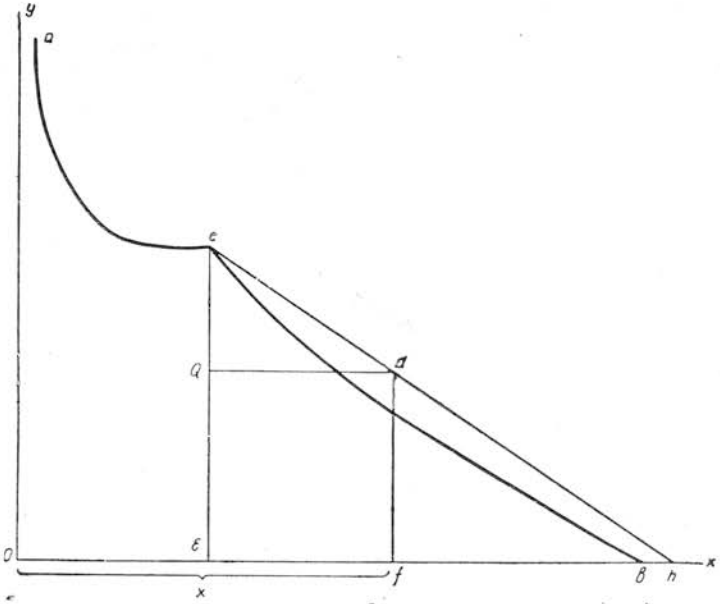

Рис. 12. Линия дожития и сила смертности $\left(\frac{cQ}{cE}\right)$.

ВЕРОЯТНОСТЬ ДОЖИТИЯ

Если в формуле $l_{x+1}+d_x=l_x$ обе части равенства разделить на l_x, то получим

$$\frac{l_{x+1}}{l_x}+\frac{d_x}{l_x}=1.$$

Первое слагаемое обозначим p_x, а второе — q_x.

Первое слагаемое $p_x=\frac{l_{x+1}}{l_x}$ представляет собой долю доживших до возраста $x+1$ год из совокупности доживших до x лет и называется вероятностью дожития.

Второе слагаемое $q_x=\frac{d_x}{l_x}$ представляет собой долю умерших в составе совокупности людей, достигших возраста x лет, и называется вероятностью умереть в возрасте x лет.

Практическое исчисление этих вероятностей основано на допущении, что вероятность умереть лиц, состоящих в одном возрасте, но принадлежащих смежным поколениям (рожденных в разных годах), одинакова. Это допущение не совсем правомерное, и результаты получаются приближенные. Дело в том, что вероятность умереть имеет тенденцию к снижению до 11—13 лет, а затем к возрастанию, значит, для младших возрастов это допущение плохо приложимо. Исполь-

зуется также предположение о пропорциональности числа случаев наблюдения времени, равносильное признанию возрастания интенсивности смертности. Эта гипотеза плохо применима к старшим возрастам. Практически гипотезе пропорциональности отдается предпочтение, ибо она облегчает расчеты, а неточности обычно незначительны по сравнению с ошибками, получаемыми при опросе населения, регистрации и т. д.

При этой гипотезе величина прожитого времени (L_x), среднее население за время $x/x+1$, число переживших возраст $x+\frac{1}{2}$ и число современников в конце календарного года равны.

ВЕРОЯТНОСТЬ УМЕРЕТЬ

Вероятность умереть q_x выражает относительное уменьшение l_x за годичный интервал от x до $x+1$. Но это уменьшение происходит не сразу, а постепенно на протяжении года. Раздробив годичный интервал на меньшие интервалы Δx, мы получим ряд чисел доживающих:

$$l_x=l(x),\ l(x+\Delta x),\ l(x+2\Delta x),\ l(x+3\Delta x),\ \dots,\ l(x+1-\Delta x),\ l(x+1).$$

Сила смертности

Найдем относительное изменение числа доживающих, приведенное к одному году:

$$\frac{l(x)-l(x+\Delta x)}{\Delta x\, l(x)}.$$

Эта величина характеризует смертность в возрасте от x лет до $x+\Delta x$. Для характеристики смертности в точном возрасте x лет нужно найти предел этого выражения при $\Delta x\to 0$. Тогда

$$\lim_{\Delta x\to 0}\left[\frac{l(x)-l(x+\Delta x)}{\Delta x}:l(x)\right]=\lim_{\Delta x\to 0}\left\{-\left[\frac{l(x+\Delta x)-l(x)}{\Delta x}:l(x)\right]\right\}=-\frac{l'(x)}{l(x)}.$$

Это есть приведенная к 1 году вероятность смерти в бесконечно малом возрастном интервале.

Этот показатель, введенный в 1825 г. Гомперцом, называется силой смертности и обозначается $\mu(x)$.

Сила смертности есть отношение производной функции к самой функции и, по сути дела, представляет собой в точном возрасте x лет логарифмическую производную функции дожития, взятую с отрицательным знаком.

Проинтегрируем силу смертности в пределах возраста от x до $x+1$

$$-\int_x^{x+1}\mu(x)\,dx=\int_x^{x+1}\frac{l'(x)}{l(x)}\,dx=[\ln l(x)]\Big|_x^{x+1}=$$

$$=\ln l(x+1)-\ln l(x)=\ln\frac{l(x+1)}{l(x)}=\ln p_x,$$

значит,

$$\ln p_x = -\int_x^{x+1} \mu(x)\,dx.$$

Отсюда

$$p_x = e^{-\int_x^{x+1} \mu(x)\,dx},$$

где e — нéперово число, основание натурального логарифма, равное 2,71828183,
и, следовательно,

$$q_x = 1 - e^{-\int_x^{x+1} \mu(x)dx}.$$

Эта формула вероятности умереть используется при построении таблиц смертности. Для бесконечно малого промежутка времени $\mu(x)dx$ есть вероятность умереть. Эта величина бесконечно мала, в то же время $\mu(x)$ величина положительная и конечная. В. И. Борткевич называл ее «густотой вероятности смерти». Так как числа доживающих $l(x)$ все время убывают, то их приращения — величины отрицательные, $\mu(x)$ можно построить аналогично тому, как строятся вообще вероятности, т. е. с помощью отношения числа случаев «благоприятствующих» смерти на общее число случаев (как «благоприятствующих», так и «неблагоприятствующих»). Конечно, при этом мы получаем «фиктивную» вероятность. Имеется бесконечно малый интервал dx, и число случаев смерти за этот интервал равно $-d[l(x)]$. Если полагать равномерное распределение смертных случаев внутри интервала $x + dx$, то из совокупности $l(x)$ умрет $\frac{d[l(x)]}{dx}$ и вероятность умереть выразится формулой

$$-\frac{d\,[l\,(x)]}{dx} : l\,(x),$$

т. е. то, что мы уже получали и называли приведенной к одному году вероятностью смерти в бесконечно малом возрастном интервале и обозначали $\mu(x)$.

В ряде случаев целесообразно произвести следующее преобразование. Учитывая, что пределы интегрирования равны одному году, $\int_x^{x+1} \mu(x)dx$ представляет собой среднюю величину силы смертности за год от x до $x + 1$, которую можно обозначить μ_x. Тогда $q_x = 1 - e^{-\mu_x}$. Разложим $e^{-\mu_x}$ в маклореновский ряд и возьмем только несколько первых членов. Получаем:

$$q_x = \mu_x - \frac{\mu_x^2}{2} + \frac{\mu_x^3}{6} - \frac{\mu_x^4}{24}.$$

Заменим этот ряд рядом геометрической прогрессии со знаменателем $\frac{\mu_x}{2}$ $\left(\text{при этом ошибка будет меньше } \frac{\mu_x^3}{12}\right)$:

$$q_x = \mu_x - \frac{\mu_x^2}{2} + \frac{\mu_x^3}{4} - \frac{\mu_x^4}{8}.$$

Произведя преобразование, получим:

$$q_x = \frac{2\mu_x}{2 + \mu_x}$$

(эту же формулу можно получить и другими способами). Теперь выразим μ_x через q_x:

$$\mu_x = \frac{2q_x}{2 - q_x} = \frac{q_x}{1 - \frac{q_x}{2}}.$$

Если теперь в полученную формулу вместо q_x поставить равную ей величину $\frac{l_x - l_{x+1}}{l_x}$, то можно записать

$$\mu_x = \frac{2\,(l_x - l_{x+1})}{l_x\left(2 - \frac{l_x - l_{x+1}}{l_x}\right)} = \frac{2(l_x - l_{x+1})}{l_x + l_{x+1}}.$$

Для большей точности в страховых расчетах используют другую формулу:

$$\mu_x = \frac{8\,(l_{x-1} - l_{x+1}) - (l_{x-2} - l_{x+2})}{12 l_x}.$$

Вероятности умереть и коэффициенты смертности

Введем в действие повозрастные коэффициенты смертности (m_x), представляющие собой средние арифметические из силы смертности в разных возрастах, взвешенные по численности населения:

$$m_x = \frac{\int_x^{x+1} \mu\,(x)\,l\,(x)\,dx}{\int_x^{x+1} l\,(x)\,dx}.$$

Числитель этой формулы — число смертных случаев в возрасте x лет, а знаменатель — среднее население в интервале от x до $x + 1$ лет (или прожитое время).

Чтобы выявить взаимосвязь между m_x и q_x, возьмем пример, приведенный В. И. Борткевичем, видоизменив его.

Пусть P и P_1 — совокупности живых в начале наблюдения; M и M_1 — числа смертных случаев; q и q_1 — вероятности смерти: $q = \frac{M}{P}$ и $q_1 = \frac{M_1}{P_1}$. Если $M = M_1$, а $P = P_1$, то $q = q_1$ при любом распределении смертей.

Означает ли это, что равны и коэффициенты смертности?

Пусть в первой совокупности P все умирают в начале периода t, а во второй P_1 все умирают в его конце. Тогда

$$m = \frac{M}{(P-M)t}; \quad m_1 = \frac{M}{Pt}.$$

Найдем формулу, устанавливающую связь m с m_1:

$$m = \frac{M}{P-M} = \frac{1}{\frac{P}{M}-1} = \frac{1}{\frac{1}{m_1}-1} = \frac{m_1}{1-m_1}.$$

Как видно из примера, хотя вероятности умереть, рассматриваемые в двух совокупностях, равны ($q = q_1$), коэффициенты смертности не совпадают ($m \neq m_1$). Однако теоретическая формула для определения коэффициента смертности не может быть практически реализована, так как точно найти среднее население или прожитое время невозможно. Поэтому практически исходят из того, что сила смертности (μ_x) в пределах одного года остается постоянной. Тогда значения силы смертности и коэффициента смертности совпадают ($\mu_x = m_x$). Поэтому для вычисления коэффициента смертности и перехода к вероятности смерти и наоборот действуют те же формулы, что и для силы смертности

$$q_x = \frac{2m_x}{2+m_x}; \qquad p_x = 1-q_x = \frac{2-m_x}{2+m_x}; \quad m_x = \frac{q_x}{1-\frac{q_x}{2}}.$$

Изложенное свидетельствует о твердо установленном факте: для одногодичного интервала $q_x < m_x$, потому что в начале периода в возрастную группу вливается больше людей, чем живет в ней.

Можно подойти к выяснению разницы между m_x и q_x иначе.

Коэффициент смертности показывает, сколько человек умирает из среднего числа живущих, вероятность же смерти — сколько умирает в течение года из лиц, живущих к началу года. Таким образом, если к 1 января 1971 г. в возрасте x лет живет 20 тыс. человек и ежедневно умирает 1 человек, то за год умрет 365 человек и вероятность смерти составит:

$$\frac{365}{20\,000} = 0{,}01825.$$

В то же время возрастной коэффициент смертности, определяемый для среднего населения $\left(\bar{S} = S_0 - \frac{\Delta}{2} = 20\,000 - \frac{365}{2} = 19\,818\right)$, составит:

$$m_x = \frac{365}{19\,818} = 0{,}01838.$$

Различие это сравнительно невелико. Чтобы получить q_x из m_x, используют формулу $q_x = \frac{M_x}{P+\frac{M_x}{2}}$, где M_x — число умерших за год, а P — число живущих.

Детская смертность не измеряется коэффициентом, а всегда исчисляется в виде вероятности. Обозначения, принятые в мировой литературе (в том числе и в изданиях ООН), m_0 и q_0 — это разные обозначения одной и той же величины.. Таким образом, принимается, что $m_0 \approx q_0$.

М. В. Птуха определял вероятность умереть в возрастах 5—7 лет по формуле, описанной А. Ньюсхолмом и Т. Э. Хейуордом.

$$q_x = \frac{2\left[{}_{1896}d_{x/x+1} + {}_{1897}d_{x/x+1}\right]}{P_{x+1/x+2} + 2P_{x/x+1} + P_{x-1/x} + {}_{1896}d_{x/x+1} + {}_{1897}d_{x/x+1}},$$

где $P_{x/x+1}$ — число живущих в возрасте x предварительно выровненное; $d_{x/x+1}$ — число умерших в интервале возраста от x до $x+1$ лет.

При определении вероятности смерти по формуле $q_x = \frac{d_x}{l_x}$ мы не учитывали миграцию, понимаемую в самом широком смысле слова. Однако при изучении некоторых явлений необходимо вносить поправку на миграцию. Например, при изучении смертности не состоящих в браке вступление в брак и разводы будем рассматривать как миграцию; вступление в брак — эмиграция, а развод — иммиграция. Предположим, что и смертность, и миграционные потоки распределяются по всему году равномерно. Тогда для исчисления вероятности умереть на практике используют формулу:

$$q_x = \frac{d_x}{l_x + \frac{a-b}{2}},$$

где a — число иммигрантов в промежутке от x до $x+1$ лет;
b — число эмигрантов в промежутке от x до $x+1$ лет.

Вывод этой формулы и ее конкретное использование см. стр. 89—91.

Вычисление показателей смертности для возрастов 0—1, 1—2, 2 и 3 года

Рассмотрение существующих методов

Существует несколько способов измерения смертности детей до одного года. Рассмотрим эти способы.

$$1) \quad q_0 = \frac{d_{0\,(t)} + d'_{0\,(t)}}{l_{0\,(t)}}.$$

В данном случае q_0 представляет собой отношение общего числа умерших детей до 1 года к числу живорожденных в том же году. Недостаток этого метода в том, что сравниваются неоднородные совокупности.

$$2) \quad q_0 = \frac{d_{0\,(t)} + d'_{0\,(t)}}{\frac{1}{2}\left(l_{0\,(t-1)} + l_{0(t)}\right)}.$$

Это вычисление по методу Раффмана. Оно устраняет неточность предыдущей формулы, так как в знаменателе берется средняя арифметическая живорождений за два примыкающих года. Недостатком этого метода является недооценка фактора живорождений за предыдущий год.

$$3) \quad q_0 = \frac{d_{0\,(t)} + d'_{0\,(t)}}{c_1 l_{0\,(t-1)} + c_2 l_{0(t)}}.$$

Это 1-я формула Рица. Она имеет в знаменателе среднюю взвешенную из числа живорожденных за два года; в качестве веса используется смертность детей по годам рождения.

$$4) \quad q_0 = \frac{d_{0(t)}}{l_{0\,(t)}} + \frac{d'_{0\,(t)}}{l_{0\,(t-1)}}.$$

Это 2-я формула Рица. Здесь используется группировка детей, умерших в календарном году, по годам их рождения. Недостаток 2-й формулы Рица заключается в том, что она не учитывает смертные случаи до начала текущего года среди родившихся в предыдущем году, а, кроме того, эта формула очень чувствительна к изменениям в распределении рождений.

$$5) \quad q_0 = 1 - \frac{l_{0\,(t+1)}}{l_{0\,(t)}} \cdot \frac{l_{1\,(t)}}{l_{0\,(t)}}$$

— формула Бёка.

Если известны элементарные совокупности умерших, полученные в результате непосредственного наблюдения или расчетным путем, то, складывая те из них, которые имеют общие гипотенузы, получим совокупности умерших первого рода, т. е. M^1. Знание M^1 необходимо (мы уже это видели) для установления теоретически точного порядка

вымирания. Если на рис. 3 взять полосу $t_7t_8y_7y_8$, то в ее одногодичных возрастных группах мы имеем M^1. Обозначив число родившихся S_0 и вычитая число лиц, умерших ранее достижения одного, двух, трех лет и т. д., получим числа доживающих до одного, двух, трех лет и т. д.:

$$S_0 - M_1^1 = l'_1,$$
$$S_0 - (M_1^1 + M_2^1) = l'_2,$$
$$S_0 - (M_1^1 + M_2^1 + M_3^1) = l'_3 \text{ и т. д.},$$

т. е. совокупности живущих I рода.

Такой порядок получения числителя и знаменателя вероятности умереть на x-м году жизни потребовал бы столетнего периода наблюдения и в конечном счете был бы неприменим из-за невозможности использовать результаты для других поколений. Поэтому практически применимым этот метод можно считать только для первых лет жизни:

$$q_0 = \frac{M_1^1}{S_0}; \quad q_1 = \frac{M_2^1}{l'_1}; \; q_2 = \frac{M_3^1}{l'_2}.$$

Чтобы получить вероятности на основании больших чисел, можно соединить несколько поколений родившихся ${}_tS_0$; ${}_{t+1}S_0$; ${}_{t+2}S_0$ и соответствующие им ${}_tM^1$; ${}_{t+1}M^1$; ${}_{t+2}M^1$ и рассчитать вероятности. Тогда для первой возрастной группы вероятность будет равна:

$$q = \frac{{}_tM^1 + {}_{t+1}M^1 + {}_{t+2}M^1 + \ldots}{{}_tS_0 + {}_{t+1}S_0 + {}_{t+2}S_0 + \ldots}.$$

Попытаемся рассчитать полученную вероятность приближенно. Возьмем совокупность умерших III рода M^3 (умерших в одном календарном году и в одном возрасте), т. е. вместо смертных точек в квадратах (см. рис. 3) $t_7t_8a_7a_8$; $a_7a_8b_7b_8$; $b_7b_8c_7c_8$ и т. д. возьмем смертные точки в параллелограммах $t_7t_8a_6a_7$; $a_7a_8b_6b_7$; $b_7b_8c_6c_7$ и т. д. Совокупности M^3 и M^1 имеют при этом всегда общую нижнюю элементарную совокупность Δ^1. Значит, ошибка, вызванная заменой M^1 на M^3, определяется разностью между верхними элементарными совокупностями, принадлежащими к поколениям разных лет, т. е. ${}_{x+1}\Delta^2 - {}_x\Delta^2$. Эта ошибка не очень велика.

Произведя замену, получаем:

$$q_0 = \frac{M_1^3}{S_0}.$$

Для большей точности при исчислении вероятностей соединим три годичных интервала и из поколения рождения t_7t_{10} возьмем совокупность умерших III рода $t_7t_{10}a_6a_9$ вместо совокупности умерших I рода $t_7t_{10}a_7a_{10}$. Тогда получим:

$$q_0 = \frac{{}_tM_1^3 + {}_{t+1}M_1^3 + {}_{t+2}M_1^3}{{}_tS_0 + {}_{t+1}S_0 + {}_{t+2}S_0}.$$

При этом абсолютная погрешность равна разности смертных точек треугольников $t_{10}a_9a_{10}$ — $t_7a_6a_7$.

Практически этот прием можно применять для изучения детской смертности.

Возьмем родившихся за два последующих года N_1 и N_0. Если из числа родившихся в прошлом году N_1 вычесть умерших из них до 1 января будущего года, т. е. $\Delta'_н$, то получим $N_1 - \Delta'_н$, т. е. число детей до одного года на 1 января. Учитывая, что, не достигнув возраста 1 года, умрет из той же совокупности N_1 еще часть детей ($\Delta'_в$), проведем последовательное вычитание и найдем число доживших до возраста 1 год из совокупности N_1 ($N_1 - \Delta'_н - \Delta'_в$). Теперь можно определить, какая доля детей, доживших до 1 января, доживет до возраста один год

$$l_1 = \frac{N_1 - \Delta'_н - \Delta'_в}{N_1 - \Delta'_н}.$$

Возьмем отношение, показывающее дожитие детей совокупности N_0 до конца календарного года своего рождения

$$L_0 = \frac{N_0 - \Delta'_н}{N_0}.$$

Перемножим обе дроби и получим вероятность дожития детей до возраста 1 год, а вычитая это произведение из единицы, найдем вероятность умереть на первом году жизни:

$$q_0 = 1 - \frac{N_1 - \Delta'_н - \Delta'_в}{N_1 - \Delta'_н} \cdot \frac{N_0 - \Delta'_н}{N_0}.$$

Бёком был предложен другой метод расчета, аналогичный этому (см. формулу 5).

Если в рассуждениях мы будем опираться на перепись в момент t_8 (см. рис. 3), т. е. в конце года t_7t_8, то в этом случае эмпирическое значение вероятности умереть, не достигнув возраста 1 год, для лиц, родившихся за время t_7t_8, измеряется частным от деления смертных точек в пределах квадрата $t_7t_8a_7a_8$ на число родившихся N_0.

Нужно иметь в виду, что не все случаи смерти детей приходятся на год t_7t_8; часть их, а именно $t_8a_7a_8$, имеет место только в следующем году. Значит, будем учитывать сначала только численность нижней элементарной совокупности $t_7t_8a_7$ или $\Delta'_н$; что же касается верхней элементарной совокупности $t_8a_7a_8$, то вместо ее действительной величины возьмем значение, вычисляемое путем умножения l_1^2 на δ'_1. А именно, по элементарной совокупности $t_7a_6a_7$, отвечающей совокупности живущих ${}_{x-1}l_2^1$ для момента t_7a_6, вычисляется путем сложной пропорции элементарная совокупность $t_8a_7a_8$, которая соответствует совокупности живущих l_1^2 для момента t_8a_7.

Если же элементарные совокупности неизвестны, то для определения вероятности умереть число M_0 нужно отнести не к N_0 и не к N_1, а к средней из них. Из рис. 3 видно, что $t_7t_8a_6a_7$ (M_0^3) состоит из двух элементарных совокупностей $\Delta'_н(t_7t_8a_7)$ и $\Delta'_в(t_7a_6a_7)$, т. е. $M_0^3 = \Delta'_н + \Delta'_в$.

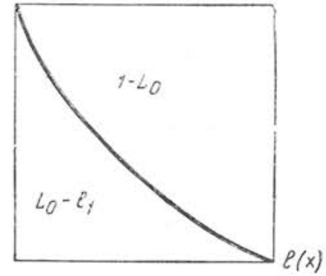

Рис. 13. Линия дожития в возрастном интервале 0—1 год.

Выразим элементарные совокупности следующим образом:

1) из формулы $L_0 = \frac{N_0 - \Delta'_н}{N_0}$ получим $\Delta'_н = N_0 - L_0 N_0 = N_0(1 - L_0)$;

2) аналогично выведем $\Delta'_в = N_1 L_0 - N_1 l_1 = N_1(L_0 - l_1)$.

Тогда

$$M_0^3 = N_0(1 - L_0) + N_1(L_0 - l_1).$$

Обозначая среднюю из N_0 и N_1 через $\overline{N}$, получаем

$$q_0 = \frac{M_0^3}{\overline{N}}.$$

С другой стороны $q_0 = 1 - p_0 = 1 - \frac{l_1}{l_0} = 1 - l_1$.

Тогда $1 - l_1 = \frac{M_0^3}{\overline{N}}$, откуда

$$\overline{N} = \frac{M_0^3}{1 - l_1} = \frac{N_0(1 - L_0) + N_1(L_0 - l_1)}{1 - l_1}.$$

Учитывая, что сумма $(1 - L_0) + (L_0 - l_1) = 1 - l_1$, можно заключить, что в данном случае из N_0 и N_1 определяется средняя арифметическая взвешенная, и весами являются $1 - L_0$ и $L_0 - l_1$.

Если рассмотреть кривизну линии l_x в возрастном интервале 0—1 год (рис. 13), можно увидеть, что отношение этих двух частей соответствует отношению площадей. Практические расчеты показали, что в условиях высокой детской смертности соотношение весов $\frac{1 - L_0}{L_0 - l_1}$ близко к отношению $\frac{2}{1}$ и в сумме равно 3[1].

Тогда

$$q_0 = \frac{M_0^3}{\frac{2}{3}N_0 + \frac{1}{3}N_1} \quad \text{(частный случай формулы 3—Рица)}$$

[1] В настоящее время это соотношение должно быть заменено другим, в большей степени соответствующим новым условиям значительно снизившейся детской смертности, т. е. отношением более высоким.

Формула Бёка (5) устраняет первый недостаток 2-й формулы Рица (4), но обладает ее вторым недостатком, т. е. чувствительна к распределению рождаемости.

Метод В. В. Паевского и А. М. Меркова

Наиболее правильным является метод, при котором умершие дети сгруппированы по месяцам рождений и смерти. Этот метод, предложенный Джини[1] в начале 20-х годов, был разработан в СССР В. В. Паевским и усовершенствован А. М. Мерковым. Особенности этого метода, имеющего большое практическое значение для измерения смертности детей, находящихся в детских яслях, исходят именно из специфики детской смертности.

Распределение помесячных показателей смертности детей на первом году жизни весьма неравномерно; при этом часть детей, умерших в течение определенного месяца в возрасте до одного года, состоит из родившихся в этом же месяце, а часть — из родившихся в предыдущие 12 месяцев. Исчисление коэффициентов смертности детей на первом месяце жизни путем отнесения числа умерших детей в определенном месяце к числу родившихся в этом же месяце даст, следовательно, искаженный результат. Значит, надо учитывать родившихся за 13 месяцев. Это обстоятельство предполагает группировку детей по месяцам жизни, при этом первый месяц целесообразно еще подразделить на две части: 0—14 дней и 15—30 дней.

Такая группировка наиболее удобна на 1 января каждого года с последующими исчислениями по указанной ниже схеме:

группы детей по интервалам;
число прибывших под наблюдение;
число выбывших;
число умерших;
число доживающих до начала возрастного интервала;
среднее число живущих;
вероятность умереть для 1000 детей данного возраста или число умирающих;
число остающихся в живых из 1000 детей данного возраста в этом же возрасте;
число доживающих и недоживающих детей из 1000 вступивших под наблюдение до следующего возраста.

В качестве знаменателя показателя смертности детей в определенном месяце следует брать помесячную среднюю из родившихся за последние 13 месяцев, включая и тот, за который исчисляется показатель смертности.

Таблицы детской смертности более детализированы и служат как бы приложением к общим таблицам смертности. Они раскрывают процесс вымирания в течение 1-го года жизни по месяцам. Для первого месяца жизни прослеживается вымирание реальной группы лиц,

[1] C. G i n i. Sulla mortalita infantile durante la guerra. «Atti della Societa italiana di Ostveticia e genicologia», vol. XIX, 1919.

родившихся в тот или иной период. Например, можно проследить вымирание в течение первого месяца жизни лиц, родившихся в 1970 г. Конечно, время вымирания нужно распространить не только на 1970 г., но и на январь 1971 г. Для исчисления смертности в следующие месяцы жизни, начиная со 2-го, можно воспользоваться методом Вестергарда, когда умершие в данном календарном периоде и в данном возрастном интервале считаются вышедшими из поколения, период рождений которого равен по величине календарному периоду смерти, но отстоит от него на среднюю арифметическую границ возрастного интервала.

Пусть, например, нужно определить смертность детей в возрасте 6—7 месяцев в 1970 г. Можно считать, что умершие в течение 12 месяцев 1970 г. в возрасте от 6 до 7 месяцев рождались также в течение 12 месяцев, но начало периода рождений отстоит от начала периода вымирания, т. е. от 1 января 1970 г., на $\frac{6+7}{2} = 6{,}5$ месяца. Следовательно, можно полагать, что взятое нами число умерших происходит из поколения родившихся в период от 15 июня 1969 г. до 15 июня 1970 г.

Теперь для получения числа доживших до возраста 6 месяцев мы должны из числа родившихся вычесть умерших в возрасте 0—1 месяц в течение периода с 1 июля 1969 г. по 1 июля 1970 г., 1—2 месяца в течение периода с 1 августа 1969 г. по 1 августа 1970 г. и т. д. до 5—6 месяцев в течение периода с 1 декабря 1969 г. по 1 декабря 1970 г. Для получения вероятности смерти q_6 (где 6 — число месяцев) нужно разделить число умерших в 1970 г. в возрасте 6—7 месяцев на число доживших до 6 месяцев (полученное вычитанием). Такой метод можно назвать косвенным, при котором сначала находят q_x, а затем l_x.

Построение показателей смертности детей в возрасте 0—1; 1—2; 2 и 3 года по материалам переписи населения СССР 1926 г.

При расчете q_0 (вероятности умереть в возрасте до 1-го года) использовались совокупности умерших III рода за 1926 и 1927 гг., т. е. число смертных точек в параллелограмме $a_0a_1c_1c_0$ (рис. 14). Задача состояла в подборе такого числа родившихся, при котором эта совокупность могла хотя бы приближенно заменить совокупность умерших I рода. Можно было использовать два приема.

1. Смертные точки в параллелограмме $a_0a_1c_1c_0$ состоят из смертных точек двух параллелограммов: $a_0a_1b_1b_0$ — совокупность умерших III рода в возрасте до одного года в 1926 г. ($M^3_{0(1926)}$) и $b_0b_1c_1c_0$ — совокупность умерших III рода в возрасте до одного года в 1927 г. ($M^3_{0(1927)}$).

Тогда

$$q_0 = \frac{M^3_{0(1926)} + M^3_{0(1927)}}{K_1 \cdot S_{1925} + S_{1926} + K_2 \cdot S_{1927}},$$

где S — число родившихся; K_1 и K_2 — соответствующие коэффициенты, определяемые методом интерполяции, предложенным П. В. Охочинским[1].

2. При отсутствии чисел родившихся в 1925 г. можно использовать и другую формулу:

$$q_0 = \frac{\Delta^1_{0(1926)} + \Delta^2_{0(1927)}}{S_{1926}}.$$

Для возрастов 1, 2, 3 года и т. д. эта формула также может быть использована, но уже в несколько измененном виде. Для нахождения q_1, например, надо число умерших в возрасте от 1 до 2 лет относить к числу доживших до возраста 1 год. Число умерших в 1926 г. в возрасте 1 год равно сумме двух элементарных совокупностей:

$$M^1_1 = \Delta^1_{1(1926)} + \Delta^2_{1\,(1927)}.$$

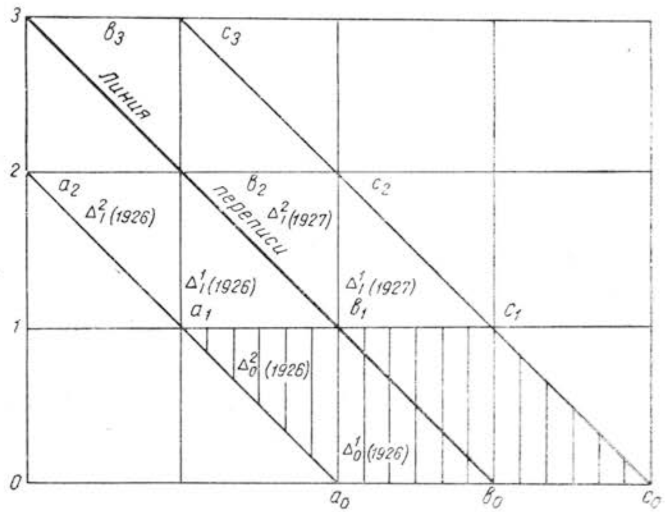

Рис. 14. Совокупности умерших первого и третьего родов.

Обозначим P_1 число живущих по переписи в возрасте 1 год. Тогда $L_1 = P_1 + \Delta^1_{1(1926)}$ даст совокупность живущих I рода, или число доживающих до возраста 1 год.

Значит,

$$q_1 = \frac{M^1_1}{P_1 + \Delta^1_{1(1926)}} = \frac{\Delta^1_{1(1926)} + \Delta^2_{1(1927)}}{P_1 + \Delta^1_{1(1926)}}.$$

[1] См. П. В. Охочинский. Опыт применения интерполирования к некоторым вопросам статистики движения населения. СПб., 1891.

Идея же метода интерполяции следующая: нужно разделить совокупность умерших III рода на элементарные совокупности. Возьмем интервал возраста равным интервалу вымирания, т. е. одному году.

Начнем с возраста 0 лет. Имеется совокупность умерших III рода (за 1 год), и задана интерполяционная формула, связывающая количество умерших z с возрастом x; $z = f(x)$. Используем формулу, выведенную русским демографом П. В. Охочинским. Предварительно распределим умерших в возрасте до одного года в соответствии со следующими периодами возраста и введем обозначения:

до 1 месяца	$-a$,
от 1 до 2 месяцев	$-b$,
от 3 до 5 месяцев	$-c$,
от 6 до 11 месяцев	$-d$.

Тогда

$$f\left(\frac{1}{12}\right) = a,$$
$$f\left(\frac{1}{4}\right) = a + b,$$
$$f\left(\frac{1}{2}\right) = a + b + c,$$
$$f(1) = a + b + c + d.$$

Интерполяционная формула в этом случае примет вид:

$$f(x) = -\frac{108}{55}a(8x^3 - 14x^2 + 7x - 1) + \frac{4}{3}(a+b) \times$$
$$\times (24x^3 - 38x^2 + 15x - 1) - \frac{2}{5}(a+b+c)(48x^3 - 64x^2 + 17x - 1) +$$
$$+ \frac{1}{33}(a+b+c+d)(96x^3 - 80x^2 + 18x - 1).$$

Представим совокупность умерших III рода графически (рис. 15).

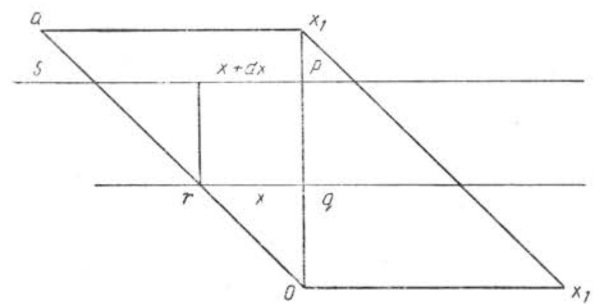

Рис. 15. Совокупность умерших третьего рода.

Производная $f'(x) = \frac{dz}{dx}$ будет представлять собой плотность смерти в точный момент возраста x лет. Отсюда $dz = f'(x)dx$ — число умерших за период от x до $x + dx$. Найдем число смертных точек в верхней элементарной совокупности $oax_1 = S_2$; сначала найдем площадь бесконечно малого участка $qrsp$ при $qp \to 0$. Этот участок может быть заменен параллелограммом с основанием rq. Треугольники ax_1o и rqo равнобедренные и площадь параллелограмма поэтому равна $dx\,dx$. Вся же площадь элементарной совокупности будет равна произведению площади прямоугольника на плотность смертей $f'(x)$

$$dx\,f'(x)\,dx.$$

Интегрируя в интервале от 0 до $x_1 = 1$ и подставляя значение функции, получаем:

$$\Delta^2_{x_1=1} = \int\limits_0^{x_1=1} xf'(x)\,dx = 0{,}327b + 0{,}105c + 0{,}838d.$$

Для более старших возрастов разложение совокупности III рода может быть произведено путем определения не верхней элементарной совокупности, а нижней. Рассмотрим отчленение нижней элементарной совокупности для возраста от x до $x + 1$ лет (рис. 16).

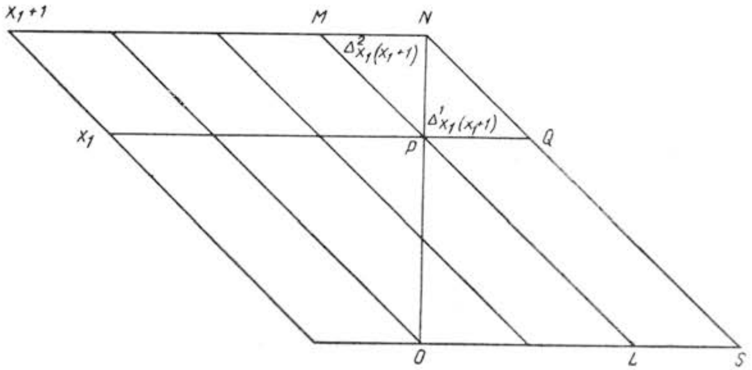

Рис. 16. Отчленение нижней элементарной совокупности от совокупности умерших III рода.

Имеется совокупность умерших III рода $M^3_{x_1/x_1+1}$ $(MNQP)$, выделяем число смертных точек в нижней элементарной совокупности PNQ. Имеем $PNQ = ONS - OPL - LPQS$, где $LPQS$ имеет основание, равное единице, и изображает $f(x)$. Тогда

$$\Delta^1_{x_1/x_1+1} = \Delta^1_{0/x_1+1} - \Delta^1_{0/x_1} - f(x_1).$$

Мы уже знаем, как находить Δ^1_{0/x_1+1} и Δ^1_{0/x_1}, а для $f(x)$ имеется интерполяционная формула. С другой стороны

$$\Delta^1_{0/x_1+1} = M^3_{0/x_1+1} - \Delta^2_{0/x_1+1} = (x_1+1)\,[f(x_1+1)] - \int\limits_0^{x_1+1} xf'(x)\,dx \text{ и } \Delta^1_{0/x_1} = M^3_{0/x_1} - \Delta^2_{0/x_1} = x_1 f(x_1) - \int\limits_0^{x_1} xf'(x)\,dx.$$

Следовательно,

$$\Delta^1_{x_1/x_1+1} = (x_1+1)\,[f(x_1+1)] - x_1 f(x_1) - f(x_1) - \left(\int\limits_0^{x_1+1} xf'(x)\,dx - \int\limits_0^{x_1} xf'(x)\,dx \right).$$

Интегрируя по частям, получаем окончательно:

$$\Delta^1_{x_1/x_1+1} = -f(x_1) + \int\limits_{x_1}^{x_1+1} f'(x)\,dx.$$

Обозначим число умерших в возрасте 0—1 года — a; 1—2 года — b; 2—3 года — c и 3—4 года — d, т. е.

$$f(1) = a;\ f(2) = a + b;\ f(3) = a + b + c;\ f(4) = a + b + c + d.$$

Тогда интерполяционая формула примет вид:

$$f(x) = a + (x-1)\,b + \frac{(x-1)(x-2)}{2}(c-b) + \frac{(x-1)(x-2)(x-3)}{6}(d - 2c + b).$$

Принимая $x_1 = 1$ и используя полученные выражения, найдем

$$\Delta^1_1 = \frac{5}{8}b - \frac{1}{6}c + \frac{1}{24}d.$$

При $x_1 = 2$

$$\Delta^1_2 = \frac{1}{24}b + \frac{1}{2}c - \frac{1}{24}d.$$

При $x_1 = 3$

$$\Delta^1_3 = \frac{1}{24}b + \frac{1}{6}c + \frac{3}{8}d \text{ и т. д.}$$

Для практического использования полученных формул привлечем данные переписи 1926 г. по женщинам сельской местности Сибирского края[1].

Зарегистрированное число родившихся в 1926 г. составляет 192 458 человек (см. табл. 6).

Используем формулу для определения верхней элементарной совокупности

$$\Delta^2_{0/1} = 0{,}327b + 0{,}105c + 0{,}838d$$

[1] См. «Смертность и продолжительность жизни населения СССР. 1926—1927. Таблицы смертности». М.—Л., Планхозгиз, 1930, стр. XXVII.

Таблица 6

Числа умерших в 1926 и 1927 гг.

Возраст (месяцы)	Годы	
	1926	1927
До 1	8 201 (*a*)	9 047
От 1 до 2	9 976 (*b*)	9 893
От 3 до 5	10 205 (*c*)	10 848
От 6 до 11	15 720 (*d*)	14 610
Итого до 1 года	44 102	44 398

и получим

$$\Delta^2_{0/1\,(1926)} = 0{,}327 \cdot 9\,976 + 0{,}105 \cdot 10\,205 + 0{,}838 \cdot 15\,720 \approx 17\,516.$$

Аналогично для 1927 г.

$$\Delta^2_{0/1\,(1927)} = 16\,626.$$

Путем вычитания находим:

$$\Delta^1_{0/1\,(1926)} = 44\,102 - 17\,516 = 26\,586 \text{ и}$$

$$\Delta^1_{0/1\,(1927)} = 44\,398 - 16\,626 = 27\,772.$$

Теперь можно определить вероятность умереть в возрасте до одного года.

$$q_0 = \frac{\Delta^1_{0/1\,(1926)} + \Delta^2_{0/1\,(1927)}}{S_0} = \frac{26\,586 + 16\,626}{192\,458} = 0{,}22453.$$

Для примера расчетов вероятности умереть в возрасте 1, 2 и 3 года привлечем данные о мужчинах сельской местности европейской части СССР.

Таблица 7

Возраст	Население по переписи 17/XII 1926 г.	Число умерших	
		1926 г.	1927 г.
0	1 713 577	407 503 (*a*)	457 088
1	1 444 534	96 548 (*b*)	105 532
2	1 351 771	48 872 (*c*)	47 023
3	1 412 004	28 832 (*d*)	27 283

Находим нижние элементарные совокупности:

для 1926 г. — для 1927 г.

$$\Delta^1_1 = 53\,398 \qquad \Delta^1_1 = 59\,257$$

$$\Delta^1_2 = 27\,258 \qquad \Delta^1_2 = 26\,772$$

$$\Delta^1_3 = 14\,934 \qquad \Delta^1_3 = 13\,671$$

Далее путем вычитания определяем верхние элементарные совокупности:

1926 г. — 1927 г.

$$\Delta^2_1 = 96\,548 - 53\,398 = 43\,150 \qquad \Delta^2_1 = 105\,532 - 59\,257 = 46\,275$$

$$\Delta^2_2 = 21\,614 \qquad \Delta^2_2 = 20\,251$$

$$\Delta^2_3 = 13\,898 \qquad \Delta^2_3 = 13\,612$$

И наконец, определяем вероятности умереть:

$$q_1 = \frac{\Delta^1_{1\,(1926)} + \Delta^2_{1\,(1927)}}{P_1 + \Delta^1_{1\,(1926)}} = \frac{53\,398 + 46\,275}{1\,444\,534 + 53\,398} = 0{,}06654;$$

$$q_2 = \frac{\Delta^1_{2\,(1926)} + \Delta^2_{2\,(1927)}}{P_2 + \Delta^1_{2\,(1926)}} = \frac{27\,258 + 20\,251}{1\,351\,771 + 27\,258} = 0{,}03445;$$

$$q_3 = 0{,}02001.$$

Построение вероятностей умереть в детском возрасте в Уральской области по данным переписи населения 1926 г. обладало некоторой спецификой.

Вероятность умереть в возрасте до года исчислялась, исходя из предположения, что умершие в этом возрасте в данном календарном году принадлежат к генерации родившихся во второй половине предыдущего года плюс родившиеся в первой половине данного года[1]. Поэтому формула имела вид:

$$q_0 = \frac{D_{0\,(1926)} + D_{0\,(1927)}}{N''_{1925} + N'_{1926} + N''_{1926} + N'_{1927}},$$

где D — число умерших (первый индекс означает возраст, второй — год смерти); N — число родившихся (внизу указывается год рождения, вверху — полугодие).

При исчислении вероятности умереть в возрасте 1 год предположили, что умершие в этом возрасте в 1926 и 1927 гг. принадлежат к совокупности родившихся с 1 июля 1924 г. до 1 июля 1926 г. за вычетом умерших в возрасте до 1 года в 1925 и 1926 гг.:

$$q_1 = \frac{D_{1\,(1926)} + D_{1\,(1927)}}{N''_{1924} + N'_{1925} + N''_{1925} + N'_{1926} - [D_{0\,(1925)} + D_{0\,(1926)}]}.$$

[1] См. «Смертность и продолжительность жизни населения Уральской области». Свердловск, Уральский Областной Статистический Отдел, 1929, стр. 55.

Для исчисления q_2, q_3, q_4 использовали формулу

$$q_x = \frac{D_x}{P_x + \frac{1}{2} D_x},$$

которая в данном случае, когда исчисление велось по двум годам, для которых дата переписи является центральной, принимала вид:

$$q_x = \frac{D_{x\,(1926)} + D_{x\,(1927)}}{2P_x + \frac{1}{2}\left[D_{x\,(1926)} + D_{x\,(1927)}\right]}.$$

В случае же резких колебаний смертности использовалась более усложненная формула. Например,

$$q_2 = \frac{D_{2\,(1926)} + D_{2\,(1927)}}{P_2 + P_3 + D_{2\,(1926)} + \frac{1}{2}\left[D_{2\,(1926)} + D_{2\,(1927)}\right]}.$$

Построение вероятности умереть по данным переписи 1959 г.

Построение вероятности умереть по данным переписи 1959 г. отражает успехи советской демографической статистики. В 1959 г. «... отчетные данные об умерших в возрасте до 5 лет позволяют получить показатели таблиц смертности для этих возрастов с использованием данных о поколениях умерших»[1].

Показатель q_0 был получен по формуле $q_0 = Q_0 + h_0$. В свою очередь

$$Q_0 = \frac{M_0^Q}{N_0 + N_{-1}}, \text{ а } h_0 = \frac{M_0^h}{N_1 + N_0},$$

где N_1, N_0 и N_{-1} — соответственно числа родившихся в 1957, 1958 и 1959 гг.;

M_0^Q — число умерших в возрасте до 1 года в 1958 г. из поколения 1958 г. плюс число умерших в возрасте до 1 года в 1959 г. из поколения 1959 г.;

M_0^h — число умерших в возрасте до 1 года в 1958 г. из поколения 1957 г. плюс число умерших в возрасте до 1 года в 1959 г. из поколения 1958 г.

Показатели q_1; q_2; q_3; q_4 были получены с использованием совокупностей умерших и живущих I рода. При получении совокупностей живущих I рода для возрастов 1, 2 и 3 года были использованы данные о числе родившихся в 1955, 1956, 1957 гг., а для возраста 4 года — данные переписи населения 1959 г. Числа живущих в возрасте 1, 2, 3 и 4 года (L_x) были получены по формуле $L_x = l_x - \alpha d_x$, где α — отношение нижней элементарной совокупности умерших ко всей совокупности умерших I рода.

[1] «Итоги Всесоюзной переписи населения 1959 года. СССР». М., Госстатиздат, 1962, стр. 257.

Интересен вывод этой формулы для возраста до 1 года. Возьмем $l_0 = 1$, $f(x)$ — плотность смертей, $M(x)$ — число умерших в возрасте до 1 года. Тогда $M'(x) = f(x)$. Значит,

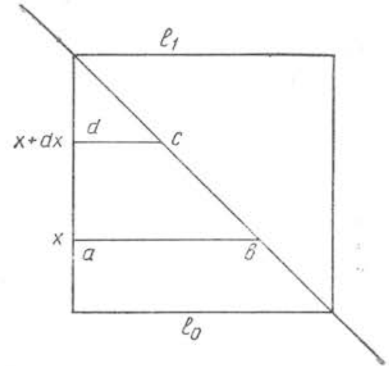

Рис. 17. Число умерших в возрасте до 1 года.

$$M(x) = \int_0^1 f(x)\,dx;$$

$$l_x = 1 - M(x) = 1 - \int_0^1 f(x)\,dx;$$

$$L_0 = \int_0^1 [1 - M(x)]\,dx = 1 - \int_0^1 M(x)\,dx.$$

На рис. 17 рассмотрим трапецию $abcd$ при $\Delta x \to 0$ как параллелограмм с основанием $1 - x$ и высотой dx. Число умерших в этом параллелограмме равно $f(x)\ [(1-x)]dx$. Умершие во всем треугольнике представляют нижнюю элементарную совокупность и рассчитываются по формуле

$$\Delta_0^1 = \int_0^1 f(x)[1 - x]\,dx = \int_0^1 f(x)\,dx - \int_0^1 xf(x)\,dx = M(1) - \int_0^1 xf(x)\,dx.$$

Интегрируя по частям, получаем:

$$\Delta_0^1 = M(1) - [xM(x)]\Big|_0^1 + \int_0^1 M(x)\,dx = \int_0^1 M(x)\,dx.$$

Мы вывели, что

$$L_0 = 1 - \int_0^1 M(x)\,dx \text{ и } \Delta_0^1 = \int_0^1 M(x)\,dx.$$

Отсюда $L_0 = 1 - \Delta_0^1$.

ЧИСЛА ЖИВУЩИХ

Найдем число человеко-лет, прожитых населением в один год. Раздробим год на малые интервалы Δx и предположим, что смерти, приходящиеся на каждый интервал, полностью сосредоточены в конце его.

При начальном числе родившихся l_0 и порядке вымирания $l(x)$ к возрасту x остается в живых l_x. Раздробив период от x до $x+1$ лет, получим интервалы $x+\Delta x$, $x+2\Delta x$ и т. д.

В первом интервале имеем живых $l_x = l(x)$, во втором — $l(x + \Delta x)$ и т. д. Общее число прожитых в данном году человеко-лет лицами в возрасте x лет составит:

$$l(x)\Delta x + l(x+\Delta x)\,\Delta x + l(x+2\Delta x)\,\Delta x + \ldots + l(x+1-\Delta x)\,\Delta x.$$

При $\Delta x \to 0$ получим предел:

$$\int\limits_x^{x+1} l(x)\,dx = L_x.$$

Это есть число живущих в возрасте x лет. Иначе это средняя величина $l(x)$ в возрастном интервале от x до $x+1$ лет, которая может быть записана так:

$$\overline{l(x)} = L_x.$$

Исходя из гипотезы пропорционального распределения смертей на протяжении короткого времени (одного года), получаем:

$$\overline{l(x)} = L_x = \frac{l_x + l_{x+1}}{2}.$$

Статистический и демографический смысл чисел живущих

Показатель L_x имеет два значения. С одной стороны, L_x показывает число человеко-лет, которое будет прожито совокупностью родившихся в течение любого интервала при данных условиях смертности. Пусть, например, до 3-летнего возраста никто не умрет, тогда совокупность родившихся 100 000 человек проживет 300 000 человеко-лет до этого возраста. Под влиянием смертности L_x функция убывающая.

С другой стороны, L_x показывает возрастной состав населения, которое постоянно пополняется за счет 100 000 рождающихся и сокращается за счет умирающих на такое же число человек. Следовательно, численность населения каждой возрастной группы будет неизменной.

Поэтому число живущих L_x иногда называют стационарным (постоянным) населением, в среде которого в течение единицы времени имеет место столько же рождений, сколько и смертей. Причем в течение этой единицы времени рождения и смерти распределены равномерно.

При расчете L_x по материалам переписи населения 1926 г. применялась упомянутая нами формула $L_x = l_x - \alpha d_x$, где α есть отношение нижней элементарной совокупности ко всей совокупности умерших I рода.

Интересный метод расчета L_x был использован при построении таблиц смертности 1926 г. по Уральской области. Сначала определялись

числа живущих по одногодичным интервалам по формулам:

$$L_0 = \frac{1}{3}\,l_0 + l_1,$$

$$L_1 = \frac{1}{2}\,(l_1 + l_2),$$

$$L_2 = \frac{1}{2}\,(l_2 + l_3),$$

$$L_3 = \frac{1}{2}\,(l_3 + l_4),$$

$$L_4 = \frac{1}{2}\,(l_4 + l_5).$$

Затем L_x дается по пятилетним интервалам до возраста 25 лет При расчетах использовалась формула квадратур

$$L_{5/10} = \frac{5\,(l_5 + l_{10})}{2} = 2{,}5\,(l_5 + l_{10}).$$

Для возраста свыше 25 до 75 лет L_x даны по десятилетним интервалам.

$$L_{25/35} = 2{,}5\,(l_{25} + 2l_{30} + l_{35}) \text{ и т. д.}$$

При этом промежуточные числа доживающих (l_{30}, l_{40}, l_{50}, l_{60} l_{70}) рассчитывались как средние геометрические из крайних. Например, $l_{30} = \sqrt{l_{25}\,l_{35}}$. Числа живущих в возрасте 75—85 и 85—95 лет исчислялись при помощи квадратуры 1,25 с тремя промежуточными точками. Например,

$$L_{75/85} = 1{,}25\,(l_{75} + 2l_{77,5} + 2l_{80} + 2l_{82,5} + l_{85}).$$

Все промежуточные члены являются средними геометрическими между их крайними:

$$l_{77,5} = \sqrt{l_{75}\,l_{80}};\quad l_{82,5} = \sqrt{l_{80}\,l_{85}}.$$

Для интервала возраста 95—105 лет число живущих рассчитано по формуле

$$L_{95/105} = l_{96} + l_{97} + \ldots + l_{104} + \frac{1}{2}\,(l_{95} + l_{105}).$$

Формулы квадратур дают возможность рассчитывать L_x по пятилетним группам без вычисления всех одногодичных L_x. Данный метод был разработан М. В. Птухой в работе «Смертность 11 народностей Е. России в конце XIX века». Сами формулы помещены в работе Ю. А. Корчак-Чепурковского «Природний рух населення Украïни в 1924 році».

Некоторые методы получения чисел живущих

При расчетах чисел живущих можно, как мы видели, брать среднюю арифметическую из чисел доживающих до двух смежных годов:

$$L_x = \frac{l_x + l_{x+1}}{2}.$$

Однако для первых лет жизни эта формула неверна, так как из родившихся в каком-нибудь конкретном году до 1 января доживает не половина, а меньше. Учитывая это, В. Я. Буняковский, например, пользовался формулой Симпсона:

$$\int f(x)\,dx = \frac{h}{3}(y_0 + y_1 + 2y_2 + 4y_3 + 2y_4 + \ldots + y_{n-1} + y_n).$$

В приложении к данному случаю формула видоизменяется. Пусть имеется группа лиц в возрасте от x до $x+2$ года (т. е. двух смежных лет от x до $x+1$ и от $x+1$ до $x+2$). Полагая l_x; l_{x+1}; l_{x+2} известными, получаем:

$$L_x + L_{x+1} = \frac{1}{3}(l_x + 4l_{x+1} + l_{x+2}).$$

К. М. Беккер счел формулу Симпсона неудовлетворительной и используя для исчисления параболу второго порядка, вывел формулу

$$L_x = \frac{l_x + l_{x+1}}{2} + \frac{a-c}{16},$$

где a — число умерших на x-м году жизни, предшествовавшем исчисляемой возрастной группе, а c — число умерших в последовавшем году.

В. И. Борткевич счел и формулу К. М. Беккера неточной. Предположив, что группы переживающих могут быть рассматриваемы как ряд одинаково отстоящих друг от друга ординат, заключенных между прямой линией и кривой 2-го порядка, он получил:

$$L_x = \frac{l_x + l_{x+1}}{2} + \frac{a-c}{24}.$$

Попробуем сопоставлять величины, найденные по трем формулам для трех возрастных периодов: 1—2 года; 2—3 года; 3—4 года. Привлечем данные по России (1867—1870 гг.).

Из 1000 родившихся лиц мужского пола дожили:

до одного года	(l_1)	— 699,79
до двух лет	(l_2)	— 626,19
до трех лет	(l_3)	— 586,64
до четырех лет	(l_4)	— 561,74
до пяти лет	(l_5)	— 544,12

Отсюда определим число умерших:

на втором году жизни — 73,60
на третьем году жизни — 39,55
на четвертом году жизни — 24,89
на пятом году жизни — 17,60
на шестом году жизни — 16,40

Следовательно, получаем:

1) по К. М. Беккеру

$$L_2 = \frac{1}{2}(626{,}19 + 586{,}64) + \frac{1}{16}(24{,}89 - 73{,}60) = 603{,}37,$$

$$L_3 = \frac{1}{2}(586{,}64 + 561{,}74) + \frac{1}{16}(17{,}60 - 39{,}55) = 572{,}82;$$

2) по В. И. Борткевичу

$$L_2 = \frac{1}{2}(626{,}19 + 586{,}64) + \frac{1}{24}(24{,}89 - 73{,}60) = 604{,}38,$$

$$L_3 = \frac{1}{2}(586{,}64 + 561{,}74) + \frac{1}{24}(17{,}60 - 39{,}55) = 573{,}28;$$

3) по В. Я. Буняковскому (с использованием формулы Симпсона). Найдем сначала число сверстников от одного до трех лет

$$L_1 + L_2 = \frac{1}{3}(699{,}79 + 4\cdot 626{,}19 + 586{,}64) = 1263{,}72.$$

Теперь вычтем из этой суммы числа живущих от двух до трех лет, полученные двумя методами (по К. М. Беккеру и В. И. Борткевичу), и получим:

пользуясь результатом К. М. Беккера,

$$L_1 = 1263{,}72 - L_2 = 1263{,}72 - 603{,}37 = 660{,}35;$$

используя результат В. И. Борткевича,

$$L_1 = 1263{,}72 - L_2 = 1263{,}72 - 604{,}38 = 659{,}34.$$

Возникает вопрос: нельзя ли применить формулу В. И. Борткевича непосредственно? Для этого нужно иметь одинаковое (хотя бы приблизительно) в двух каких-нибудь смежных годах число умерших, т. е. $l_x - l_{x+1} = l_{x+1} - l_{x+2}$.

Принимая числа умерших на пятом (17,60) и шестом (16,40) годах довольно близкими, получаем по К. М. Беккеру и В. И. Борткевичу:

$$L_4 = \frac{561{,}74 + 544{,}12}{2} = 552{,}93.$$

Для возрастной группы 3—5 лет получаем по В. Я. Буняковскому:

$$L_3 + L_4 = \frac{1}{3}(l_3 + 4l_4 + l_5) = \frac{1}{3}(586{,}64 + 4\cdot 561{,}74 + 544{,}12) = 1125{,}90.$$

Вычитая отсюда $L_4 = 552{,}93$, получаем:
$L_3 = 1125{,}90—552{,}93 = 572{,}97$.

Сравнивая, видим, что эта же величина по К. М. Беккеру была 572,82, а по В. И. Борткевичу — 573,28.

Для возрастной группы от двух до четырех лет находим по В. Я. Буняковскому:

$$L_2 + L_3 = \frac{1}{3}(l_2 + 4l_3 + l_4) = \frac{1}{3}(626{,}19 + 4 \cdot 586{,}64 + 561{,}74) = 1178{,}16.$$

Вычитая отсюда найденную уже величину $L_3 = 572{,}97$, получаем: $L_2 = 1178{,}16—572{,}97 = 605{,}19$. Сравнивая, видим, что эта величина по К. М. Беккеру — 603,37, по В. И. Борткевичу — 604,38.

Для возрастной группы 1—3 года мы уже нашли по В. Я. Буняковскому $L_1 + L_2 = 1263{,}72$. Откуда $L_1 = 1263{,}72 - L_2 = 1263{,}72—$ $— 605{,}19 = 658{,}53$ (по К. М. Беккеру — 660,35, а по В. И. Борткевичу — 659,34).

Теперь выпишем все полученные результаты.

Таблица 8

Сопоставление чисел живущих

L_x	По В. Я. Буняковскому	По К. М. Беккеру	По В. И. Борткевичу
L_1	658,53	660,35	659,34
L_2	605,19	603,37	604,38
L_3	572,97	572,82	573,28
L_4	552,93	552,93	552,93

Разность $L_x - l_{x+1}$ дает число живущих в возрасте x лет, которые должны умереть, не достигнув $x + 1$ года.

Так, по В. Я. Буняковскому $L_1 = 658{,}53$, а $l_2 = 626{,}19$. Тогда $L_1 - l_2 = 658{,}53—626{,}19 = 32{,}34$. Это число мужчин в возрасте 1—2 года, которые умрут, не достигнув трех лет.

Разность $l_x - L_x$ — число уже умерших в возрасте x лет. По В. Я. Буняковскому это число для возрастного периода 1—2 года равно:

$$l_1 - L_1 = 699{,}79 - 658{,}53 = 41{,}26.$$

Общее число умерших на втором году жизни равно $32{,}34 + 41{,}26 =$ $= 73{,}60$.

Теперь можно найти долю тех, кто умрет, не достигнув возраста $x + 1$, во всей совокупности умерших:

$$\frac{32{,}34}{73{,}60} \cdot 100 = 43{,}9\%.$$

Далее по В. Я. Буняковскому найдем $L_2 - l_3 = 605{,}19—586{,}64 =$ $= 18{,}55$, разделим на число умерших на третьем году жизни, т. е. 39,55, и получим: $\frac{18{,}55}{39{,}55} \cdot 100 \approx 46{,}8\%$ детей в возрасте двух лет, которые умрут до того, как им исполнится 3 года. Произведем расчеты по всем трем методам.

Таблица 9

Доля умерших, не достигнув возраста $x+1$ лет, во всей совокупности умерших (в %)

Метод расчета	Возрастные группы (в годах)		
	1—2	2—3	3—4
По В. Я. Буняковскому	$\frac{32{,}3}{73{,}6} \cdot 100 = 43{,}9$	$\frac{18{,}5}{39{,}5} \cdot 100 = 46{,}8$	$\frac{11{,}2}{24{,}9} \cdot 100 = 45{,}1$
По К. М. Беккеру	$\frac{34{,}1}{73{,}6} \cdot 100 = 46{,}4$	$\frac{16{,}7}{39{,}5} \cdot 100 = 42{,}5$	$\frac{11{,}1}{24{,}9} \cdot 100 = 44{,}8$
По В. И. Борткевичу	$\frac{33{,}1}{73{,}6} \cdot 100 = 45{,}1$	$\frac{17{,}7}{39{,}5} \cdot 100 = 45{,}1$	$\frac{11{,}54}{24{,}9} \cdot 100 = 46{,}7$

В возрастной группе 0—1 год это отношение составляет 33,5%

Использование совокупностей умерших первого и второго рода для исчисления чисел живущих

Рассматриваемые методы исчисления можно дополнить следующими, еще более точными методами, основанными на использовании элементарных совокупностей.

Пусть на демографической сетке имеются три совокупности умерших I рода, т. е. принадлежащих одному поколению a, b, и c (см. рис. 18), где $a = d_{x-1}$; $b = d_x$; $c = d_{x+1}$. Каждая из совокупностей распадается на две элементарные совокупности: a — на α_1, α_2; b — на β_1, β_2; c — на γ_1; γ_2. На рис. 18 l_x изображены линией B_1B_2, l_{x+1} — линией C_1C_2, искомое число L_x — диагональю C_1B_2. На рисунке видно, что $L_x = l_x - \beta_1$. Для определения β_1 используем элементарные совокупности и найдем разности первых трех порядков.

Таблица 10

Разности первых трех порядков

Элементарные совокупности	Первые разности	Вторые разности	Третьи разности
α_1			
α_2	$\alpha_2 - \alpha_1$		
β_1	$\beta_1 - \alpha_2$	$\beta_1 - 2\alpha_2 + \alpha_1$	
β_2	$\beta_2 - \beta_1$	$\beta_2 - 2\beta_1 + \alpha_2$	$\beta_2 - 3\beta_1 + 3\alpha_2 - \alpha_1$
γ_1	$\gamma_1 - \beta_2$	$\gamma_1 - 2\beta_2 + \beta_1$	$\gamma_1 - 3\beta_2 + 3\beta_1 - \alpha_2$
γ_2	$\gamma_2 - \gamma_1$	$\gamma_2 - 2\gamma_1 + \beta_2$	$\gamma_2 - 3\gamma_1 + 3\beta_2 - \beta_1$

3*

В соответствии с конечными разностями и исходя из предположения, что элементарные совокупности умерших образуют параболу 2-го порядка, вторые разности постоянны, а третьи разности должны равняться нулю. Таким образом, разности третьего порядка дают три уравнения. Присоединяем к ним еще три уравнения: $\alpha_1 + \alpha_2 = a$; $\beta_1 + \beta_2 = b$; $\gamma_1 + \gamma_2 = c$.

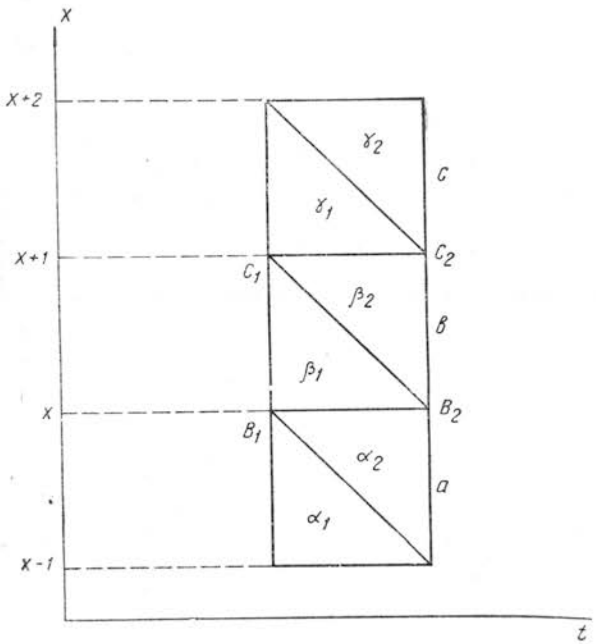

Рис. 18. Три совокупности умерших первого рода.

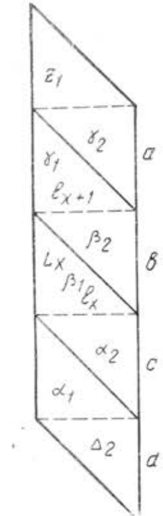

Рис. 19. Четыре совокупности умерших второго рода.

Получаем систему, состоящую из шести уравнений с шестью неизвестными.

$$\left\{\begin{array}{lr} \text{I} \quad \alpha_1 + \alpha_2 & = a \\ \text{II} \quad \beta_1 + \beta_2 & = b \\ \text{III} \quad \gamma_1 + \gamma_2 & = c \\ \text{IV} \quad -\alpha_1 + 3\alpha_2 - 3\beta_1 + \beta_2 & = 0 \\ \text{V} \quad -\alpha_2 + 3\beta_1 - 3\beta_2 + \gamma_1 & = 0 \\ \text{VI} \quad -\beta_1 + 3\beta_2 - 3\gamma_1 + \gamma_2 & = 0 \end{array}\right.$$

Определитель этой системы:

$$\Delta = \begin{vmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ -1 & 3 & -3 & 1 & 0 & 0 \\ 0 & -1 & 3 & -3 & 1 & 0 \\ 0 & 0 & -1 & 3 & -3 & 1 \end{vmatrix} = -64.$$

Заменяем третью колонку коэффициентов (при β_1) величинами a; b; c; 0; 0; 0 и получаем:

$$\Delta_1 = \begin{vmatrix} 1 & 1 & a & 0 & 0 & 0 \\ 0 & 0 & b & 1 & 0 & 0 \\ 0 & 0 & c & 0 & 1 & 1 \\ -1 & 3 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & -3 & 1 & 0 \\ 0 & 0 & 0 & 3 & 3 & 1 \end{vmatrix} = -4(8b + a - c),$$

$$\beta_1 = \frac{\Delta_1}{\Delta}.$$

Отсюда $\beta_1 = \frac{b}{2} + \frac{a-c}{16}$. Следовательно, $L_x = l_x - \frac{b}{2} + \frac{c-a}{16}$. Расчеты показывают, что для возрастов старше 5 лет последний член излишен.

Можно использовать также метод нахождения L_x по совокупности умерших II рода. Возьмем 4 совокупности, каждая из которых состоит из двух элементарных совокупностей (рис. 19). При этом предполагаем, что 8 элементарных совокупностей умерших образуют параболу 3-го порядка. Тогда третьи разности будут постоянными, а четвертые равны 0.

Таблица 11

Разности первых четырех порядков

Элементарные совокупности	Первые разности	Вторые разности	Третьи разности	Четвертые разности
z_1	$\gamma_2 - z_1$	$\gamma_1 - 2\gamma_2 + z_1$	$\beta_2 - 3\gamma_1 + 3\gamma_2 - z_1$	$\beta_1 - 4\beta_2 + 6\gamma_1 - 4\gamma_2 + z_1$
γ_2	$\gamma_1 - \gamma_2$	$\beta_2 - 2\gamma_1 + \gamma_2$	$\beta_1 - 3\beta_2 + 3\gamma_1 - \gamma_2$	$\alpha_2 - 4\beta_1 + 6\beta_2 - 4\gamma_1 + \gamma_2$
γ_1	$\beta_2 - \gamma_1$	$\beta_1 - 2\beta_2 + \gamma_1$	$\alpha_2 - 3\beta_1 + 3\beta_2 - \gamma_1$	$\alpha_1 - 4\alpha_2 + 6\beta_1 - 4\beta_2 + \gamma_1$
β_2	$\beta_1 - \beta_2$	$\alpha_2 - 2\beta_1 + \beta_2$	$\alpha_1 - 3\alpha_2 + 3\beta_1 - \beta_2$	$\Delta_2 - 4\alpha_1 + 6\alpha_2 - 4\beta_1 + \beta_2$
β_1	$\alpha_2 - \beta_1$	$\alpha_1 - 2\alpha_2 + \beta_1$	$\Delta_2 - 3\alpha_1 + 3\alpha_2 - \beta_1$	
α_2	$\alpha_1 - \alpha_2$	$\Delta_2 - 2\alpha_1 + \alpha_2$		
α_1	$\Delta_2 - \alpha_1$			
Δ_2.				

Получаем восемь уравнений с восемью неизвестными:

$$\begin{cases}
\text{I} & z_1+\gamma_2 & =a\\
\text{II} & \gamma_1+\beta_2 & =b\\
\text{III} & \beta_1+\alpha_2 & =c\\
\text{IV} & \alpha_1+\Delta_2 & =d\\
\text{V} & z_1-4\gamma_2+6\gamma_1-4\beta_2+\beta_1 & =0\\
\text{VI} & \gamma_2-4\gamma_1+6\beta_2-4\beta_1+\alpha_2 & =0\\
\text{VII} & \gamma_1-4\beta_2+6\beta_1-4\alpha_2+\alpha_1 & =0\\
\text{VIII} & \beta_2-4\beta_1+6\alpha_2-4\alpha_1+\Delta_2 & =0
\end{cases}$$

Определитель системы:

$$\Delta=\begin{vmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1\\
1 & -4 & 6 & -4 & 1 & 0 & 0 & 0\\
0 & 1 & -4 & 6 & -4 & 1 & 0 & 0\\
0 & 0 & 1 & -4 & 6 & -4 & 1 & 0\\
0 & 0 & 0 & 1 & -4 & 6 & -4 & 1
\end{vmatrix}=1024.$$

Заменяя пятую колонку коэффициентов (при β_1) величинами a, b, c, d, 0, 0, 0, 0, получаем:

$$\Delta_1=\begin{vmatrix}
1 & 1 & 0 & 0 & a & 0 & 0 & 0\\
0 & 0 & 1 & 1 & b & 0 & 0 & 0\\
0 & 0 & 0 & 0 & c & 1 & 0 & 0\\
0 & 0 & 0 & 0 & d & 0 & 1 & 1\\
1 & -4 & 6 & -4 & 0 & 0 & 0 & 0\\
0 & 1 & -4 & 6 & 0 & 1 & 0 & 0\\
0 & 0 & 1 & -4 & 0 & -4 & 1 & 0\\
0 & 0 & 0 & 1 & 0 & 6 & -4 & 1
\end{vmatrix}=520c+136b-40d-24a,$$

$$\beta_1=\frac{\Delta_1}{\Delta}.$$

Отсюда

$$L_x=l_x-\beta_1=l_x-\frac{\Delta_1}{\Delta}=l_x-\frac{520}{1\,024}c-\frac{136}{1\,024}b+\frac{40}{1\,024}d+\frac{24}{1\,024}a.$$

Пренебрегая малыми величинами, получим «грубую» формулу

$$L_x=l_x-\frac{520}{1\,024}c-\frac{136}{1\,024}b.$$

Требования, предъявляемые к математическим формулам, состоят в том, чтобы расчеты по ним соответствовали действительному положению дел. С этой точки зрения наилучшей является формула В. И. Борткевича и предлагаемые нами формулы. Конечно, более правильным является использование точного распределения статистических данных относительно умерших. Именно поэтому при учете умерших требуется фиксация не только возраста смерти, но и года рождения.

ОБЩЕЕ ЧИСЛО ЧЕЛОВЕКО-ЛЕТ, ПРОЖИТЫХ НАСЕЛЕНИЕМ

В качестве вспомогательного расчетного показателя используют T_x, т. е. общее число человеко-лет, которое предстоит прожить совокупности l_x от возраста x лет до предельного возраста $\omega-1$.

$$T_x=L_x+L_{x+1}+L_{x+2}+\ldots+L_{\omega-1}.$$

В частности, $T_0=\sum\limits_{i=0}^{i=\omega-1}L_i$ — это число человеко-лет, которое проживет совокупность родившихся l_0 до конца периода вымирания. Вообще говоря, T_x получается путем кумулятивного добавления функции L_x снизу вверх, т. е. начиная с L_{100}.

СРЕДНЯЯ ПРОДОЛЖИТЕЛЬНОСТЬ ПРЕДСТОЯЩЕЙ ЖИЗНИ И ЕЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЕ ЗНАЧЕНИЕ

Для каждого отдельного человека продолжительность жизни измеряется интервалом времени между его рождением и смертью. Для совокупности людей возникает необходимость вычисления средней продолжительности жизни этой совокупности.

Рассмотрим этот показатель с разных точек зрения и развеем некоторые заблуждения, связанные с ним, тем более, что, как пишет В. Н. Старовский, сведения «о продолжительности предстоящей жизни вызывают у некоторых скептицизм»[1].

Средняя продолжительность жизни для новорожденного

Если разделить число человеко-лет, прожитых родившимися, на число родившихся, то получим среднюю продолжительность предстоящей жизни для новорожденных:

$$e_0^{\circ}=\frac{T_0}{l_0}.$$

Если же разделить число человеко-лет, которое проживут дожившие до x лет после этого возраста, на число доживших до x лет, то

[1] См. В. Н. Старовский. Методика исследования элементов роста народонаселения.—В кн. «Социология в СССР», т. 1, М., «Мысль», 1966.

получим среднюю продолжительность предстоящей жизни для людей, достигших определенного возраста:

$$e_x^\circ = \frac{T_x}{l_x} = \frac{L_x + L_{x+1} + \ldots + L_{\omega-1}}{l_x}.$$

Раскроем геометрический смысл этой величины.

Пусть имеется функция $l(x)$. Приращение функции $l(x+\Delta x) - l(x)$ — это число умерших в точном возрасте x лет — можно приближенно приравнять дифференциалу. Значит,

$$l(x) - l(x+\Delta x) \approx -l'(x)\,dx.$$

Тогда среднюю продолжительность предстоящей жизни для доживающих до x лет получим путем вычитания x из величины средней продолжительности жизни. Аналогично обычной средней арифметической взвешенной, т. е. $\bar{x} = \frac{\Sigma xm}{\Sigma m}$, получим:

$$e_x^\circ = \frac{\int\limits_x^\omega -xl'(x)\,dx}{\int\limits_x^\omega -l'(x)\,dx} - x.$$

Выпишем числитель $\int\limits_x^\omega - xl'(x)dx$. Приняв $x = u$, а $l'(x)dx = dV$, произведем соответствующие подготовительные действия и, интегрируя по частям, получаем $dx = du$ и $V = l(x)$.
Тогда

$$-\int\limits_x^\omega xl'(x)dx = -xl(x)\Big|_x^\omega + \int\limits_x^\omega l(x)\,dx.$$

Следовательно, в числителе получаем:

$$xl(x) + \int\limits_x^\omega l(x)\,dx.$$

Знаменатель легко интегрируется и дает $l(x)$. Значит,

$$e_x^\circ = \frac{xl(x) + \int\limits_x^\omega l(x)\,dx}{l(x)} - x = \frac{xl(x) + \int\limits_x^\omega l(x)\,dx - xl(x)}{l(x)} = \frac{\int\limits_x^\omega l(x)\,dx}{l(x)}.$$

Вот теперь видно, что средняя продолжительность предстоящей жизни равна площади под кривой дожития, деленной на ординату l_x (см. рис. 11).

Средняя продолжительность жизни и средний возраст умерших

Совершенно ошибочным является приравнивание среднего возраста умерших средней продолжительности жизни.

Средний возраст умерших зависит не только от порядка вымирания, но и возрастно-половой структуры населения, а также от изменений в числах ежегодных рождений.

Если допустить, что эти числа постоянны, то при постоянном порядке вымирания мы получим, что в формуле средней продолжительности жизни $e = l_1 + l_2 + \ldots + l_{99}$ числа лиц, умерших в течение определенного промежутка времени в различных возрастах, будут пропорциональны величинам d_0, d_1, d_2, ... Если же учесть, что число родившихся из года в год не остается постоянным, а числа лиц, умерших в течение определенного периода времени, по возрастам не соответствуют d_0, d_1, d_2, ... и т. д., то в зависимости от того, увеличивается или уменьшается число ежегодно родившихся, несовпадение среднего возраста умерших со средней продолжительностью жизни будет склоняться в ту или иную сторону.

При увеличении ежегодной рождаемости распределение умерших в течение определенного периода времени по возрастам будет соответствовать ряду чисел d_0, d_1, d_2, ... и т. д. Это приведет к тому, что удельный вес младших возрастных групп будет увеличиваться. Учитывая, что смертность в этих младших группах очень высокая, а доля этих групп возрастает, средний возраст умерших окажется ниже средней продолжительности жизни. При уменьшении ежегодных чисел рождений все будет происходить наоборот, и средний возраст умерших окажется выше средней продолжительности жизни.

Однако практически реальное население всегда моложе, чем стационарное, и поэтому средний возраст умерших меньше, чем средняя длительность жизни.

Рассмотрим среднюю продолжительность жизни в зависимости от того, когда имеют место смертные случаи в том или ином возрастном интервале. Предположим, что все смертные случаи в одногодичном интервале происходят в начале интервала. Определим то число лет (S), которое в общей сложности проживут лица, входящие в состав совокупности l_0.

Имеем $S_1 = 0d_0 + 1d_1 + 2d_2 + \ldots + 99d_{99}$.

Учитывая, что $d_x = l_x - l_{x+1}$, делаем соответствующую подстановку и получаем:

$$S_1 = 0(l_0 - l_1) + 1(l_1 - l_2) + 2(l_2 - l_3) + \ldots + 99(l_{99} - l_{100}) = l_1 - l_2 + 2l_2 - 2l_3 + 3l_3 - 3l_4 + \ldots + 98l_{98} - 98l_{99} + 99l_{99} = l_1 + l_2 + \ldots + l_{99}.$$

Предположим, что все смертные случаи происходят в конце одногодичного интервала. Тогда

$$S_2 = 1d_0 + 2d_1 + 3d_2 + \ldots + 100d_{99} = l_0 + l_1 + l_2 + \ldots + l_{99}.$$

Следовательно, S_2 больше S_1 на l_0.

Если же исходить из предположения о распределении смертных случаев равномерно в каждом одногодичном интервале, то можно рассчитать среднюю арифметическую ($\overline{S}$).

Тогда

$$\overline{S}=\frac{l_0}{2}+l_1+l_2+\ldots+l_{99}.$$

Разделим найденное среднее число лет, которое в общей сложности проживут лица, входящие в состав родившихся, на численность начальной совокупности и получим величину, показывающую, сколько лет приходится в среднем на каждого, т. е. среднюю продолжительность жизни.

Получим:

$$e_0=\frac{\overline{S}}{l_0}=\frac{1}{2}+\frac{l_1+l_2+\ldots+l_{99}}{l_0}.$$

Если теперь взять возраст x лет и подсчитать, сколько лет остается прожить в среднем лицу, дожившему до этого возраста, то получим среднюю продолжительность предстоящей жизни:

$$e_x=\frac{l_{x+1}+l_{x+2}+\ldots+l_{99}}{l_x}+\frac{1}{2}.$$

Можно подойти к формуле средней продолжительности предстоящей жизни, используя аппарат математических ожиданий[1]. Известно, что вероятность смерти человека в возрасте x' лет в промежутке от $x+dx$ вычисляется по формуле силы смертности

$$\mu_x=\frac{dl_x}{l_{x'}}.$$

Так как это лицо пережило промежуток времени $x'-x$, то математическое ожидание его предстоящей жизни в возрасте x' лет будет:

$$e^0_{x'}=\int_{x'}^{\omega}\frac{dl_x}{l_{x'}}(x-x')\,dx=-\frac{1}{l_{x'}}\int_{x'}^{\omega}\frac{dl_x}{d_x}(x-x')\,dx.$$

Интегрируя по частям, получаем:

$$e^0_{x'}=-\frac{1}{l_{x'}}\int_{x'}^{\omega}\frac{dl_x}{d_x}(x-x')\,dx=\frac{1}{l_{x'}}\int_{x'}^{\omega}l_x\,dx=\frac{T_{x'}}{l_{x'}}.$$

[1] См. И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики, стр. 117 и далее.

Если предположить, что в одногодичных возрастных интервалах смертные случаи распределяются равномерно, получаем:

$$T_{x'}=\frac{1}{2}(d_x+3d_{x+1}+5d_{x+2}+\ldots)=\frac{1}{2}l_x+l_{x+1}+l_{x+2}+\ldots+l_{99}.$$

Следовательно,

$$e^{\circ}_x=\frac{T_{x'}}{l_{x'}}=\frac{1}{2}+\frac{l_{x+1}+l_{x+2}+\ldots+l_{99}}{l_x}$$

или

$$e^{\circ}_x=\frac{l_x+l_{x+1}+l_{x+2}+\ldots+l_{99}}{l_x}-\frac{1}{2}.$$

Факторы, влияющие на среднюю продолжительность жизни

Что же означает средняя продолжительность жизни в СССР, равная 70 годам? Чтобы ответить на этот вопрос, рассмотрим равенство, из которого вытекает возможность определения числа доживающих до возраста x лет.

$$l_x=l_0(1-q_0)(1-q_1)(1-q_2)\ldots(1-q_{x-1})=l_0p_0p_1p_2\ldots p_{x-1}.$$

Вычисленные таким образом значения l_1, l_2, l_3 отражают не действительный порядок вымирания определенной группы родившихся, а тот порядок, при котором постепенно вымирала бы известная группа родившихся, если бы она в одногодичных возрастах была подвержена той смертности, которая наблюдалась в этих годах за определенный период времени.

В этом смысле числа доживающих (l_x) и средняя продолжительность жизни (e^0_x) до некоторой степени условны.

Средняя продолжительность жизни в СССР, составляющая ныне 70 лет, не означает вовсе, что совокупность родившихся, например, в 1971 г. проживет в общей сложности такое число лет, что в среднем на каждого человека придется 70 лет. Эта цифра выведена на основании той смертности, которая существовала при вычислении средней продолжительности жизни различных возрастных групп, относящихся к различным периодам рождения. Только такая абстракция позволяет определить среднюю продолжительность жизни.

Влияют ли на среднюю продолжительность жизни социально-экономические факторы? Конечно, да. Смертность зависит от благоприятствующих и неблагоприятствующих обстоятельств на протяжении всей жизни человека. А эти обстоятельства в свою очередь диктуются социально-экономическими условиями жизни изучаемой социальной группы.

Однако достаточно ли знания средней продолжительности жизни для суждения о социально-экономическом характере данной социальной группы? Нужно иметь в виду недостатки этого показателя. В самом названии «средняя» подчеркивается, что величина эта абстрактная. Средняя продолжительность жизни всего населения может быть выше, чем в другой стране, а для отдельной группы, например работоспособных, — ниже и т. д.

В зависимости от структуры населения по возрасту и полу, занятости в отраслях, национальному составу, грамотности, удельному весу городского и сельского населения, трудоспособных во всем населении и т. д. величина средней продолжительности жизни может значительно меняться.

Это означает, что при определенных сопоставлениях величина показателя средней продолжительности жизни, весьма чутко реагирующего на изменение социально-экономических факторов в жизни населения, может быть завышена или занижена (искажена и затемнена) действием привходящих обстоятельств, т. е. различием в структуре сравниваемых совокупностей людей. Поэтому знания одной лишь общей средней продолжительности жизни недостаточно для суждения о социально-экономических условиях данного населения и порядке вымирания этого населения.

Очевидно, для сопоставления величин средней продолжительности жизни нужна подробная группировка, позволяющая исчислять порядок вымирания более узких групп населения. Однако это приводит к уменьшению численностей, а следовательно, снижает достоверность показателей.

Значит, любое исследование должно базироваться на оптимальном, т. е. наилучшем соотношении однородности населения по определенным признакам и достаточной численности, дающей возможность наиболее точно выявить закономерность показателей.

Так как влияние социально-экономических факторов затемняется различием структур населения, некоторые считают, будто коэффициенты рождаемости, коэффициенты смертности и возрастно-половая структура населения образуют систему, в которой значения двух факторов достаточно для определения третьего.

На таком предположении (при отсутствии миграции или ее незначительности) основана попытка статистиков ООН по значениям коэффициентов рождаемости и смертности определить возрастно-половой состав населения[1]. По этому поводу следует заметить, что, во-первых, возрастно-половая структура обладает большим показателем степеней свободы и может значительно варьировать при одинаковых коэффициентах рождаемости и смертности. Она зависит от рождаемости и смертности не в данный момент, а за многие предыдущие годы. Поэтому все попытки подобного рода основываются на тех или иных гипотезах (например, теории квазистабильного населения) и носят ограниченный характер.

Во-вторых, надо учесть, что статистики ООН одновременно пытаются преодолеть затруднения, вызванные отсутствием в некоторых

[1] Руководства по методам исчисления населения. Руководство III. Методы перспективного исчисления населения по возрасту и полу. ООН, Нью-Йорк, 1956, стр. 25.

странах фактических данных, позволяющих дать достаточно полную демографическую характеристику населения.

Разработка статистических приемов и методов элиминирования привходящих явлений позволит выявить действие социально-экономических факторов. Приемы элиминирования влияния структуры населения по одному и двум признакам на смертность населения являются предметом особого рассмотрения.

Вычисления по таблицам смертности населения СССР 1958—1959 гг. показали, что 100 000 родившихся проживут в общем 6 859 240 человеко-лет. Можно ожидать, что средняя продолжительность жизни новорожденных будет равна 68,59 лет. Интересно отметить, что e_x^0 с возрастом убывает, но не настолько, насколько увеличивается возраст. Так, при чествовании юбиляров-мужчин в 60-летнем возрасте можно ожидать по таблицам смертности 1958—1959 гг., что они проживут еще в среднем по 17,02 года, т. е. приблизительно до 77 лет, а для 77-летних ожидается средняя продолжительность предстоящей жизни — 8,07 года и т. д.

Следовательно, от одного года до 60 лет возраст увеличился на 59 лет, а средняя продолжительность жизни снизилась с 68,59 до 17,02, т. е. на 51,57 года, а дальнейшее увеличение возраста на 17 лет привело к снижению средней продолжительности жизни на 8,95 года.

Парадокс показателя средней продолжительности жизни

Итак, средняя продолжительность жизни (e_0^0), будучи функцией возраста, является функцией убывающей. Возникает вопрос: монотонная ли это функция? Нет ли участков, на которых она меняет свое направление? Поставим вопрос так: насколько больше продолжительность жизни достигших $x+1$ лет, чем достигших x-го года? Разумеется, что ответ: «На один год», — неправилен. Постараемся это доказать.

Всем дожившим до возраста x лет $\left(\sum\limits_x^{\omega} d_i\right)$ предстоит прожить в среднем $x+e_x^{\circ}$ лет. Следовательно, всего они проживут $(x+e_x^{\circ})\times \times\sum\limits_x^{\omega} d_i$ человеко-лет.

Допустим, что все умершие на $x+1$ году жизни (d_x) прожили по $x+\frac{1}{2}$ лет, т. е. $(x+\frac{1}{2})\,d_x$. Учитывая, что прожитая часть жизни при переходе от возраста x к возрасту $x+1$ увеличивается на один год, найдем, насколько уменьшится средняя продолжительность предстоящей жизни:

$$e_x^{\circ}-e_{x+1}^{\circ}=1-\left[\frac{(x+e_x^{\circ})\sum\limits_x^{\omega} d_i-\left(x+\frac{1}{2}\right)d_x}{\sum\limits_x^{\omega} d_i-d_x}-(x+e_x^{\circ})\right]=$$

$$
=1-\frac{d_x\left(e_x^\circ-\frac{1}{2}\right)}{\sum\limits_{x+1}^{\omega} d_i}=\frac{\sum\limits_{x+1}^{\omega} d_i-d_x\left(e_x^\circ-\frac{1}{2}\right)}{\sum\limits_{x+1}^{\omega} d_i}.
$$

Установим, при каких условиях (при каких значениях возраста x) $e_x^\circ-e_{x+1}^\circ$ будет величиной отрицательной, т. е. $e_x^\circ<e_{x+1}^\circ$. Тем самым мы зафиксируем, в каком возрасте по мере его увеличения средняя продолжительность предстоящей жизни будет возрастать. Числитель правой части равенства $\sum\limits_{x+1}^{\omega} d_i+\frac{1}{2}d_x-d_x e_x^\circ$ меньше нуля, если $\sum\limits_{x+1}^{\omega} d_i<d_x\left(e_x^\circ-\frac{1}{2}\right)$ или $\frac{\sum\limits_{x+1}^{\omega} d_i}{d_x}<e_x^\circ-\frac{1}{2}$.

Так как

$$
\sum_{x+1}^{\omega} d_i=\sum_{x}^{\omega} d_i-d_x,
$$

то

$$
\frac{\sum\limits_{x}^{\omega} d_i-d_x}{d_x}<e_x^\circ-\frac{1}{2},\qquad \frac{\sum\limits_{x}^{\omega} d_i}{d_x}<e_x^0+\frac{1}{2}.
$$

Это неравенство может иметь место при таких значениях возраста x, когда d_x очень велико, а e_x^0 — значительно. Оборачивая неравенство, получаем:

$$
\frac{d_x}{\sum\limits_{x}^{\omega} d_i}>\frac{1}{e_x^\circ+\frac{1}{2}}.
$$

Учитывая, что $\sum\limits_{x}^{\omega} d_i=l_x$, произведем подстановку:

$$
\frac{d_x}{l_x}>\frac{1}{e_x^\circ+\frac{1}{2}}.
$$

Следовательно,

$$
q_x>\frac{1}{e_x^\circ+\frac{1}{2}}.
$$

Полученный результат означает, что средняя продолжительность предстоящей жизни лиц в возрасте x лет меньше средней продолжительности предстоящей жизни лиц в возрасте $x+1$ год, когда *вероятность умереть в возрасте x лет (q_x) больше обратной величины средней продолжительности жизни, увеличенной на 0,5 года для возраста $x+1$ год.*

Когда же d_x и e_x^0 велики? Это бывает в детском возрасте. Правда, d_x в старческом возрасте тоже велико, но e_x^0 мало. В самом деле, посмотрим, например, на таблицы смертности в СССР.

Таблица 12

Извлечение из таблиц смертности (1926—1927 гг. и 1958—1959 гг.)

Возраст в годах x	1958—1959 гг. Оба пола*		1926—1927 гг. Оба пола**	
	Числа умерших d_x	Средняя продолжительность предстоящей жизни e_x^0	Число умерших d_x	Средняя продолжительность предстоящей жизни e_x^0
0	4 060	68,59	18 701	44,35
1	806	70,48***	5 193	53,48
2	354	70,08	2 429	56,10
3	212	69,34	1 376	56,94
4	152	68,49	1 033	57,01***
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
96	701	3,93	195	3,16
97	584	3,79	157	2,99
98	480	3,65	124	2,82
99	391	3,52	99	2,66
100	314	3,40	75	2,53

* «Итоги Всесоюзной переписи населения 1959 года. СССР.» М., Госстатиздат, 1962, стр. 262—263.

** «Смертность и продолжительность жизни населения СССР. 1926—1927. Таблицы смертности» М., Планхозгиз, 1930, стр. 2, 3.

*** Максимум, далее начинается снижение.

В 1926—1927 гг. увеличение продолжительности предстоящей жизни с увеличением возраста имело место у детей в возрасте от 0 до 4 лет. Снижение смертности привело к изменению этой тенденции к 1958—1959 гг. Наибольшая средняя продолжительность предстоящей жизни уже не у 4-летних детей, как в 1926—1927 гг., а у однолетних.

Рассмотрим соотношение между q_x и $\frac{1}{e_x^0+0{,}5}$ в таблицах смертности 1926—1927 гг. и 1958—1959 гг. (табл. 13).

Для 1926—1927 гг. имеем в возрасте 0 лет 0,18701>0,0186, поэтому e_0^0 при переходе к возрасту 1 год возрастает; в возрасте 1 год 0,06387>0,0177, поэтому e_1^0 при переходе к возрасту два года тоже возрастает. И так до возраста $x=4$, для которого вероятность умереть q_4 становится меньше $\frac{1}{e_4^0+0{,}5}$, т. е. 0,01429<0,0174. Значит, e_4^0 должна быть наибольшей.

Таблица 13

Извлечение из таблиц смертности 1926—1927 гг. и 1958—1959 гг.

Возраст x	1926—1927 гг.				1958—1959 гг.			
	q_x	$\overset{\circ}{e}_x+0,5$	$\frac{1}{\overset{\circ}{e}_x+0,5}$	изменение $\overset{\circ}{e}_x$	q_x	$\overset{\circ}{e}_x+0,5$	$\frac{1}{\overset{\circ}{e}_x+0,5}$	изменение $\overset{\circ}{e}_x+0,5$
0	0,18701	44,85	0,0222	возрастает	0,04060	69,09	0,01447	возрастает
1	0,06387	53,98	0,0186	»	0,00840	70,98	0,01410	max
2	0,03192	56,60	0,0177	»	0,00372	70,58	0,01417	убывает
3	0,01868	57,44	0,0175	»	0,00224	69,84	0,0143	»
4	0,01429	57,51	0,0174	max				и т. д.
5	0,01070	57,33	0,0174	убывает				
6	0,00827	56,94	0,0176	»				

Для 1958—1959 гг. имеем для возраста 0 лет $0{,}04060>0{,}01410$; поэтому при переходе к возрасту 1 год e_0^0 возрастает; в возрасте 1 год $q_1=0{,}00840$ уже становится меньше $\frac{1}{e_2^0+0{,}5}=0{,}01417$. Следовательно, в возрасте 1 год средняя продолжительность жизни должна быть наибольшей ($e_1^0=\max$). Таким образом, практические расчеты подтверждают наши теоретические выводы.

Можно подойти к парадоксу показателя средней продолжительности жизни иначе. Выразим среднюю продолжительность жизни как функцию возраста:

$$\overset{\circ}{e}_x=\frac{\int\limits_x^{\infty} l(x)\,dx}{l(x)}.$$

Исследуем функцию $\overset{\circ}{e}_x$ на возрастание. Для этого найдем первую производную $(\overset{\circ}{e}_x)'$ и решим неравенство $(\overset{\circ}{e}_x)'>0$.

$$(\overset{\circ}{e}_x)'=\frac{l(x)\,l(x)-l'(x)\int\limits_x^{\infty} l(x)\,dx}{[l(x)]^2}=1-\frac{l'(x)}{l(x)}\,\frac{\int\limits_x^{\infty} l(x)\,dx}{l(x)}=1-\mu(x)\,\overset{\circ}{e}_x.$$

Значит, $1-\mu(x)\,\overset{\circ}{e}_x>0$ для тех значений возраста x, при которых e_x^0 является возрастающей функцией. Отсюда следует:

$$\mu(x)=\frac{1}{\overset{\circ}{e}_x}.$$

Следовательно, парадокс показателя средней продолжительности жизни, состоящий в ее увеличении при увеличении возраста, имеет место в тех возрастах, где сила смертности больше обратной величины средней продолжительности жизни.

Развернем формулу средней продолжительности жизни и получим:

$$\overset{\circ}{e}_0=\frac{T_0}{l_0}=L_0+L_1+L_2+\ldots+L_{\omega-1}.$$

Учитывая, что $L_x=\frac{l_x+l_{x+1}}{2}$, получаем:

$$\overset{\circ}{e}_0=\frac{l_0+l_1}{2}+\frac{l_1+l_2}{2}+\frac{l_2+l_3}{2}+\ldots=\frac{1}{2}\,l_0+l_1+l_2+\ldots=$$

$$=\frac{1}{2}(d_0+d_1+d_2+\ldots)+(d_1+d_2+d_3+\ldots)+$$

$$+(d_2+d_3+d_4+\ldots)\ldots=\frac{1}{2}\,d_0+\frac{3}{2}\,d_1+\frac{5}{2}\,d_2+$$

$$+\frac{7}{2}\,d_3+\ldots+\frac{2\omega-1}{2}\,d_{\omega-1}=\sum_{i=0}^{i=\omega-1}\left(i+\frac{1}{2}\right)d_i.$$

Можно вычислить среднюю продолжительность жизни людей, достигающих возраста x лет, или среднюю продолжительность жизни людей за вычетом тех, которые до этого возраста не доживают

$$\frac{\left(x+\frac{1}{2}\right)d_x+\left(x+\frac{3}{2}\right)d_{x+1}+\ldots+\frac{2\omega-1}{2}\,d_{\omega-1}}{d_x+d_{x+1}+d_{x+2}+\ldots+d_{\omega-1}}=$$

$$=\frac{\sum\limits_{i=x}^{i=\omega-1}\left(i+\frac{1}{2}\right)d_i}{\sum\limits_{i=x}^{i=\omega-1}d_i}.$$

Если теперь из этой величины вычесть возраст x лет, то получим среднюю продолжительность предстоящей жизни для достигших возраста x лет:

$$\overset{\circ}{e}_x=\frac{\sum\limits_{i=x}^{i=\omega-1}\left(i+\frac{1}{2}\right)d_i}{\sum\limits_{i=x}^{i=\omega-1}d_i}-x.$$

После преобразования получаем:

$$\overset{\circ}{e}_x=\frac{\sum\limits_{i=0}^{i=\omega-x-1}\left(i+\frac{1}{2}\right)d_{x+i}}{\sum\limits_{i=0}^{i=\omega-x-1}d_{x+i}}.$$

При практических расчетах этой величины исходят из так называемой сокращенной средней продолжительности предстоящей жизни e_x, которая меньше полной на 0,5 года: $e_x = \overset{0}{e}_x - \frac{1}{2}$. После алгебраических преобразований получаем:

$$e_x = \frac{d_{x+1} + 2d_{x+2} + 3d_{x+3} + \dots}{d_x + d_{x+1} + d_{x+2} + \dots}.$$

Если мы приведем вариационный ряд[1],

Таблица 14

Возрастной интервал (в годах)	Центр возрастного интервала x	d_x	xd_x
0—1	1/2	d_0	1/2 d_0
1—2	3/2	d_1	3/2 d_1
2—3	5/2	d_2	5/2 d_2
и т. д.	и т. д.	и т. д.	и т. д.
Итого	—	1,00	$\Sigma\, xd_x$

то увидим, что $\overset{0}{e}_0$ есть средний арифметический возраст. Весами являются частоты интервалов по продолжительности жизни:

$$\overset{\circ}{e}_0 = \bar{x} = \frac{\Sigma\, xd_x}{\Sigma d_x} = \Sigma xd_x.$$

ВЕРОЯТНАЯ ПРОДОЛЖИТЕЛЬНОСТЬ ЖИЗНИ

Показатель V_x, представляющий собой на языке статистики медиану в распределении по продолжительности предстоящей жизни, называется вероятной продолжительностью жизни для лиц данного возраста. Вероятная продолжительность жизни — это число лет, которое проживет после возраста x лет ровно половина достигших этого возраста. Иначе говоря, V_x — это число лет, через которое число доживающих уменьшится вдвое.

Для исчисления V_x принимаем, что для лиц возраста x лет вероятность дожить до $x + n$ лет и не дожить равны:

$$\frac{l_{x+n}}{l_x} = \frac{l_x - l_{x+n}}{l_x}.$$

[1] И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики, стр. 7—13.

Расчет V_x может осуществляться по формуле

$$V_x = n + \frac{l_{x+n} - \frac{1}{2}\, l_x}{l_{x+n} - l_{x+n+1}},$$

где l_{x+n} и l_{x+n+1} — соседние табличные числа доживающих, из которых первое несколько больше, а второе несколько меньше $1/2\ l_x$.

Число n означает целую часть V_x или разность между возрастом x, для которого определяют величину вероятной продолжительности жизни, и тем возрастом $x + n$, в котором остается в живых несколько больше половины лиц возраста x лет. Например, из таблицы смертности 1958—1959 гг. для мужчин Советского Союза имеем $l_0 = 100\,000$, $\frac{l_0}{2} = 50\,000$, $n = 70$, $l_{0+70} = 50\,920$, $l_{0+n+1} = 48\,677$.

Тогда $$V_0 = 70 + \frac{50\,920 - 50\,000}{50\,920 - 48\,677} = 70{,}4 \text{ года}.$$

Вероятная продолжительность жизни так же, как и средняя продолжительность жизни, подводит итог условиям смертности.

Однако следует заметить, что в смысле обобщений данный показатель уступает средней продолжительности жизни. Дело в том, что средняя продолжительность жизни исчисляется с учетом смертности во всех возрастах, начиная с данного и до конца жизни, а вероятная продолжительность жизни зависит от смертности, существовавшей не во всем периоде, а лишь в его части, так как на V_x не влияет смертность в старших возрастах.

Таблица 15

Сопоставление средней и вероятной продолжительности жизни по данным таблицы смертности 1926—1927 гг.

Возраст в годах x	Мужчины		Женщины	
	средняя продолжительность жизни $\overset{\circ}{e}_x$	вероятная продолжительность жизни v_x	средняя продолжительность жизни $\overset{\circ}{e}_x$	вероятная продолжительность жизни v_x
0	41,93	50,18	46,79	58,69
10	51,65	55,37	55,72	60,82
20	43,24	46,18	47,36	51,46
30	35,65	37,53	39,75	42,47
40	28,02	29,02	32,12	33,57
50	20,99	21,10	24,41	24,82
52	19,70	19,62	—	—
55	—	—	20,66	20,62
60	14,85	14,11	17,07	16,65
70	9,65	8,62	10,96	10,16
80	6,05	4,90	6,77	5,69
90	3,97	3,10	4,37	3,43
100	2,44	1,85	2,59	1,93

Для сопоставления значений средней и вероятной продолжительности жизни приведем данные из таблиц смертности 1926—1927 гг. в европейской части СССР для мужчин и женщин (табл. 15).

Сравнение показывает, что вероятная продолжительность жизни (V_x) превышает среднюю продолжительность жизни (e_x^0) в детских и средних возрастах: у мужчин до 52 лет, а у женщин до 55 лет. Далее же соотношение меняется, т. е. V_x становится меньше. Анализ этих соотношений по данным таблицы смертности 1897 г. выявляет две особенности:

при рождении вероятная продолжительность не выше, как в 1926—1927 гг., а ниже средней продолжительности жизни; но уже после первого года жизни соотношение между этими двумя величинами такое же, как и в 1926—1927 гг.;

возраст, в котором e_x^0 становится больше V_x, для мужчин составляет 45 лет, а для женщин 43 года.

Ввиду отсутствия V_x в таблицах смертности 1958—1959 гг. произведем расчет V_x и сопоставим e_x^0 и V_x.

Таблица 16

Сопоставление средней и вероятной продолжительности жизни по расчетным данным таблицы смертности 1958—1959 гг.

Возраст x	Мужчины		Женщины	
	$\overset{\circ}{e}_x$	v_x	$\overset{\circ}{e}_x$	v_x
0	64,42	70,4	71,68	77,89
10	58,85	61,85	65,87	68,82
20	49,53	52,10	56,37	58,94
30	40,71	42,61	47,07	49,15
40	32,16	33,34	37,89	39,43
50	24,05	24,52	28,99	29,90
60	17,02	16,76	20,64	20,78
70	11,28	10,47	13,35	12,70
80	6,93	5,86	8,04	6,87
90	4,26	3,31	5,15	4,04
100	2,74	—	3,56	—

Замечаем, что превышение e_x^0 по сравнению с V_x начинается уже примерно с 60 лет (и у мужчин и у женщин).

Таблица 17

1897 г.		1926—1927 гг.		1958—1959 гг.	
Муж.	Жен.	Муж.	Жен.	Муж.	Жен.
45	43	52	55	60	60

Сопоставив данные разных лет, можно зафиксировать тенденцию к повышению возраста, в котором e_x^0 становится больше V_x (табл. 17).

Могут ли быть равны средняя и вероятная продолжительность жизни во всех возрастах?

Предположим, что смертные случаи распределены равномерно в течение всей жизни. Из такого предположения исходил в своих практических расчетах Муавр (правда, для населения с 12 лет). Тогда можно показать, что $V_x = e_x^0$. Нам известно, что

$$\overset{\circ}{e}_x = \frac{l_x + l_{x+1} + l_{x+2} + \dots l_{\omega-1}}{l_x} - \frac{1}{2}.$$

Если числа умерших в каждом возрасте равны друг другу и равны единице (для простоты), то

$$\overset{\circ}{e}_x = \frac{l_x + (l_x - 1) + (l_x - 2) + \dots}{l_x} - \frac{1}{2}.$$

Тогда, используя формулу убывающей арифметической прогрессии, получим:

$$\overset{\circ}{e}_x = 1 + \frac{l_x(l_x - 1)}{2l_x} - \frac{1}{2} = \frac{l_x - 1}{2} + \frac{1}{2} = \frac{l_x}{2}.$$

Но при гипотезе о равномерном распределении смертных случаев из l_x каждый год умирает один человек, вследствие чего половина их вымрет через 1/2 l_x лет, что и составляет вероятную продолжительность предстоящей жизни для лица из совокупности l_x. Значит, e_x^0 и V_x могут быть равны только в том случае, если *плотность смертей во всех возрастах одинакова.*

«НОРМАЛЬНАЯ ПРОДОЛЖИТЕЛЬНОСТЬ ЖИЗНИ»

Кроме показателей, публикуемых в таблицах смертности, имеется еще один, введенный в науку В. Лексисом, назвавшим его «нормальной продолжительностью жизни».

Этот показатель указывает на тот возраст в конце жизни человека, около которого числа умерших образуют купол.

Изучая смертность, В. Лексис произвел группировку населения на 3 группы, смертности которых резко различны. Первая группа — это «молодые» в возрасте до 14 лет, смерть которых является преждевременной, ненормальной, не обусловленной естественными причинами. Вторая группа — «преждевременно умершие» в возрасте 14—67 лет, смерть которых хотя и не является ненормальной, но так же, как и группы «молодых», еще не объясняется «естественным предрасположением людей». Третья группа — это население в возрасте 67—80 лет, смерть которых «нормальна».

Рассматривая кривую d_x, В. Лексис заметил два максимума: на первом году жизни и в интервале 67—80 лет. Второй максимум представляет собой моду или «уплотненную среднюю» и означает, по его мнению, «нормальную продолжительность жизни».

Таблица 18

Повозрастные числа умерших (в тыс.)

Возраст	30—35	35—40	40—45	45—50	50—55	55—60	60—65	65—70	70—75	75—80	80—85	85—90
Число умерших, d_x	2,1	2,5	2,9	3,5	4,5	5,8	7,5	9,2	10,3	9,6	6,7	3,1

Используя данные таблиц умерших по Германии с 1901 до 1910 г., В. Лексис нашел, что «нормальный возраст жизни» лежит в интервале 70—75 лет, а более точно

$$M_0 = 70 + \frac{10{,}3-9{,}2}{(10{,}3-9{,}2)+(10{,}3-9{,}6)} = 70 + 5\cdot\frac{1{,}1}{1{,}1+0{,}7} \approx 73 \text{ года.}$$

Предостережем от распространенного заблуждения при использовании моды, вычисленной по абсолютным данным. Пусть, например, известно, что модальная смертность от какой-нибудь причины по абсолютным данным о числе умерших приходится на возраст 20—30 лет (максимальное число умерших). Отсюда, казалось бы, следует вывод, что возраст 20—30 лет является наиболее угрожающим. Между тем при этом не учитывается соотношение разных возрастов в населении, т. е. удельный вес каждой группы.

Поэтому правильный вывод может быть сделан только по относительным величинам, вычисленным для каждого интервала и показывающим, какой удельный вес составляют умирающие в этом интервале от численности живущих. По максимуму удельного веса умирающих и определяется модальный возраст, т. е. возраст наибольшей угрозы.

Можно рекомендовать несколько способов исчисления «нормальной продолжительности жизни» в зависимости от резкости выражения второго максимума. Наиболее удачными, на наш взгляд, являются формулы, предложенные В. И. Борткевичем.

В случае резко выраженного второго максимума «нормальная продолжительность жизни» вычисляется по формуле

$$n = m + \frac{d_m\, d_{m-1}}{2d_m - d_{m-1} - d_{m+1}},$$

где d_m — максимальная величина d_x. При не очень резко выраженном втором максимуме, находящемся в каком-то более или менее явном возрастном периоде, формула имеет вид:

$$n = \frac{x'+x''}{2} + \frac{(x''-x')(c-a)}{2(2b-a-c)},$$

где x'' и x' — высший и низший пределы возрастного периода при $d_x = \max$; b — максимум, a и c смертные случаи за периоды, предшествующий максимуму и последующий за ним.

Для суждения о возможности использования «нормальной продолжительности жизни» при изучении смертности можно сравнить фактическое распределение смертных случаев с теоретическим. В. И. Борткевич предложил для этой цели особый метод, основанный на использовании трех величин: n — «нормальной продолжительности жизни»; $2l_n$ — возрастной группы, распределение которой соответствует нормальной кривой (т. е. укладывается в нее), и h — меры точности[1].

В данном случае $h = \frac{0{,}4769}{V_n}$, где V_n — вероятная продолжительность жизни.

Для полученных величин n, $2l_n$ и h строится нормальное теоретическое распределение.

$$F(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2}\, dt, \quad \text{где } u = hz.$$

$F(u)$ показывает вероятность того, что смертные случаи в «нормальной» группе находятся в пределах возраста от $n - z$ до $n + z$.

М. В. Птуха использовал эти построения для расчетов «нормальной продолжительности жизни» женщин на Украине по пятилетним интервалам[2].

На связь нормальной, средней и вероятной продолжительности жизни обратил внимание Б. Ц. Урланис в статье «Увеличение продолжительности жизни в СССР»[3] и предложил весьма удачное название «модальная продолжительность жизни».

На основании исследования изменяемости числа умерших, т. е. показателя d_x, можно установить показатель «нормальной продолжительности жизни» населения, для которого составлена таблица смертности. Рассмотрим изменение d_x по таблице смертности населения СССР 1958—1959 гг. и сравним с изменениями d_x по таблице смертности 1926—1927 гг. (рис. 20).

На рисунке видно, что d_x последовательно снижается от нулевого возраста до 12—15 лет, затем начинает возрастать сначала медленно примерно до 40 лет, а затем более быстро, достигает максимума в один из годов старческого возраста (между 75—85 годами) и снова начинает падать.

[1] См. И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики, стр. 121.

[2] См. М. В. Птуха. Смертность в России и на Украине. «Очерки по статистике населения». М., Госстатиздат, 1960.

[3] См. «Ученые записки по статистике», т. VI, М., изд-во АН СССР, 1961, стр. 107.

Надо помнить, что нормальная (или то, что мы называем модальная) продолжительность жизни характеризует смертность в периоды, для которых построены таблицы. Модальная продолжительность жизни зависит от распределения смертных случаев в старших возрастах таблицы смертности, а это распределение получается, как правило, экстраполяцией на основе тех или иных гипотез. Поэтому показатель модальной продолжительности жизни больше, чем другие показатели, зависит от субъективности автора таблицы смертности.

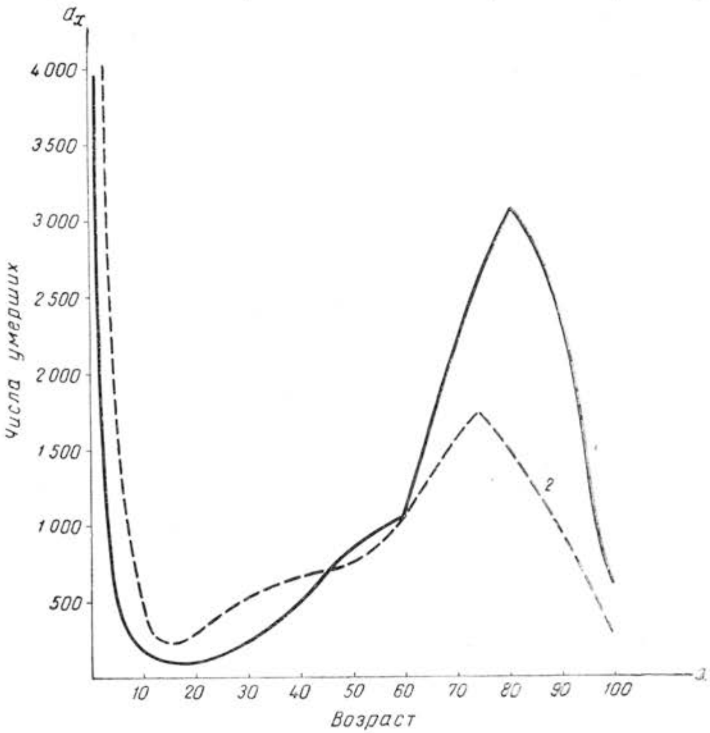

Рис. 20. Сопоставление чисел умерших из таблиц смертности (оба пола):
1—из таблиц 1958—1959 гг.; *2*—из таблиц 1926—1927 гг.

Корреляция между средней и нормальной продолжительностью жизни

Можно по данным нескольких таблиц найти корреляционное уравнение связи между вторым максимумом (II max) и e_0^0 (предполагая связь, например, линейной), а затем устанавливать значение средней продолжительности жизни по величине второго максимума.

По данным о средней продолжительности предстоящей жизни (e_0^0) и величине второго максимума, т. е. «нормальной продолжительности жизни», в республиках СССР по результатам переписи населения

1959 г. нами произведены вычисления и найдено корреляционное уравнение связи по прямой линии.

$$\overline{e}^0_{0(\mathrm{II\,max})} = 37{,}6 + 0{,}4 \cdot \mathrm{II\ max},$$

где $\overline{e}^0_{0(\mathrm{II\,max})}$ — выровненное значение средней продолжительности жизни по второму максимуму.

На рис. 21 представлена ломаная прогрессия и прямая линия, выражающая приведенное выше уравнение корреляционной связи между II max и e_0^0.

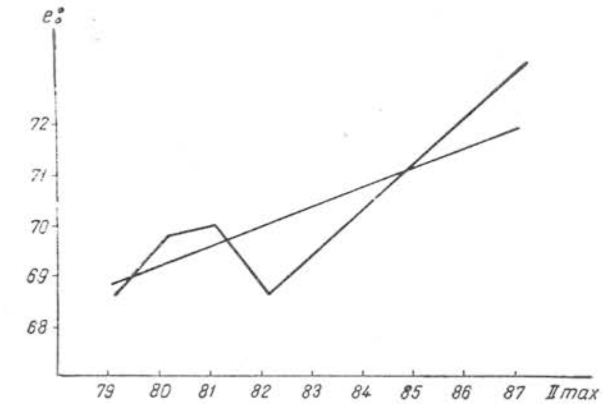

Рис. 21. Выравнивание средней продолжительности жизни по прямой линии.

Значит, по величине II max, определяемой из графы d_x таблицы смертности, можно с достаточной приближенностью и определенной вероятностью найти среднюю продолжительность предстоящей жизни.

Амплитуда демографических показателей

В качестве показателя, измеряющего жизненность населения, можно предложить амплитуду колебаний d_x от min до max, т. е. от возраста 10—12 лет до 75—85 лет; $A = x_{\max} - x_{\min}$. Вторым показателем, характеризующим жизненность, мы рекомендуем долю умерших в амплитудном интервале:

$$D_A = \sum_{x=\min}^{x=\max} d_x.$$

Учитывая, что $\sum_{x=0}^{x=\omega} d_x = 1$ и $\sum_{x=0}^{x=\omega} d_x = \sum_{x=0}^{x=\min} d_x + \sum_{x=\min}^{x=\max} d_x + \sum_{x=\max}^{x=\omega} d_x$, получим

$$D_A = \sum_{x=\min}^{x=\max} d_x = 1 - \left(\sum_{x=0}^{x=\min} d_x + \sum_{x=\max}^{x=\omega} d_x \right).$$

Доля умерших в амплитудном интервале может служить дополнением к средней продолжительности жизни. Чем ближе D_A к единице, тем, очевидно, меньшая доля людей умирает за пределами амплитудного интервала. Если смертность детей до 11—12 лет снижается, а D_A возрастает, то это должно служить показателем тех положительных сдвигов, которые происходят в смертности населения. Нами произведен расчет амплитуды и доли умирающих в амплитудном интервале по двум таблицам смертности.

Таблица 19

Таблица смертности	Амплитуда (в годах), A			Доля умирающих в амплитудном интервале D_A		
	Муж.	Жен.	Оба пола	Муж.	Жен.	Оба пола
1926—1927 гг,	60	62	61	0,45	0,45	0,45
1958—1959 гг.	65	70	68	0,65	0,59	0,61

Изменение величины D_A свидетельствует о существенном изменении режима вымирания.

Более точно характеризует сдвиги в смертности населения следующий коэффициент K:

$$K = \frac{\sum_{x=\min}^{x=\omega} d_x}{\sum_{x=0}^{x=\min} d_x}.$$

Чем больше K, тем успешнее борьба за снижение доли преждевременно умирающих.

В заключение заметим, что если из ряда d_x исключить умирающих до 12 лет, а по оставшимся d_x вычислить показатель скошенности, то он обнаружит любопытную устойчивость асимметрии. Найдем $\bar{x} = \overset{0}{e}_{12}$; $M_0 - 12 = M_0'$; $M_e = V_{12}$, где M_0 — мода, а M_e — медиана распределения. В случае симметричности распределения $M_0' = \bar{x} - 3(\bar{x} - M_e)$; при несимметричном распределении $\overset{0}{e}_{12} < V_{12} < M_0'$. Отклонение величины $z = \dfrac{\bar{x} - 3(\bar{x} - M_e)}{M_0'}$ от единицы служит мерой асимметрии.

Возьмем данные из таблиц смертности 1926—1927 гг. и 1958—1959 гг. для всего населения и произведем некоторые расчеты.

Таблица 20

Выявление симметричности чисел умерших

Пол	По данным переписи 1926—1927 г.			По данным переписи 1958—1959 гг.			Величина z по формуле $z = \dfrac{\bar{x} - 3(\bar{x} - M_e)}{M_0'}$	
	$\bar{x} = \overset{\circ}{e}_{12}$	$M_e = V_{12}$	$M_0' = M_0 - 12$	$\bar{x} = \overset{\circ}{e}_{12}$	$M_e = V_{12}$	$M_0' = M_0 - 12$	1926—1927 гг.	1958—1959 гг.
1	2	3	4	5	6	7	8	9
Мужчины .	50	54	60	57	60	65	1,01	1,02
Женщины .	54	58	62	64	67	69	1,06	1,06
Все население . . .	52	56	61	61	64	68	1,05	1,03

Значения z показывают небольшую, очень стабильную скошенность распределения чисел умирающих. С точки зрения статистики факт устойчивости характера асимметрии в распределениях чисел умирающих (d_x) в динамическом аспекте представляет собой закономерность, вскрытую эмпирическим путем.

СТАНДАРТИЗАЦИЯ ДЕМОГРАФИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Необходимость устранения влияния возрастной структуры

Сопоставляя общие демографические показатели двух или более совокупностей или нескольких групп одной совокупности, неоднородных по своему составу, можно сделать неправильный вывод. Так, в частности, результат сравнения коэффициентов естественного движения двух совокупностей может быть искажен влиянием различий в их возрастно-половой структуре. Общие коэффициенты рождаемости, смертности, брачности и другие представляет собой средние, взвешенные соответствующим образом.

Если мы имеем данные, например, о смертности отдельно мужчин и женщин, составляющих определенную совокупность, то общая их смертность будет средней из частных средних по мужчинам и женщинам, взвешенной по численности полов. Общая смертность представляет собой также среднюю из повозрастных показателей смертности, взвешенных по численности или по их удельным весам. Следовательно, средняя смертность зависит не только от показателей смертности отдельных возрастно-половых групп, но и от возрастно-половой структуры.

Учитывая это, при анализе демографических данных часто сравнивают не общие (средние) коэффициенты, а коэффициенты частных групп. Между тем анализ требует сравнения именно общих показателей. Возникает проблема элиминирования влияния различия в структурах сравниваемых совокупностей.

Применяемый при этом статистический прием с методической стороны представляет собой построение такой модели коллектива, в котором возможное влияние возрастных различий на смертность устранено.

Для дальнейшего рассуждения введем соответствующие обозначения:

q — варьирующий признак, коэффициенты смертности;

q', q'' — смертность соответственно в первой и второй совокупностях;

S' и S'' — численность населения соответственно первой и второй совокупностей;

w — удельные веса (частости), т. е. доли каждой возрастной группы в общей численности и соответственно w' и w'' — частости первой и второй совокупностей.

Средняя смертность этих совокупностей, взвешенная по численности населения, будет:

$$\bar{q}' = \frac{\sum q' S'}{\sum S'} \quad \text{и} \quad \bar{q}'' = \frac{\sum q'' S''}{\sum S''}.$$

Если взвешивать по частостям, то получим:

$$\bar{q}' = \frac{\sum q' w'}{\sum w'} = \sum q' w' \quad \text{и} \quad \bar{q}'' = \frac{\sum q'' w''}{\sum w''} = \sum q'' w''.$$

Сравнивая средние смертности, можно получить индекс смертности

$$I_{\text{смертн}} = \frac{\bar{q}''}{\bar{q}'} = \frac{\sum q'' w''}{\sum q' w'},$$

отвечающий на вопрос о том, насколько средняя смертность второй совокупности выше ($I > 1$) и ниже ($I < 1$) средней смертности первой совокупности, и отражающий действие двух факторов: 1) смертностей отдельных групп и 2) их удельных весов в каждой из совокупностей (т. е. структуры).

Методы элиминирования влияния различия в возрастных структурах

Задача состоит в том, чтобы устранить влияние различия в структурах. Это может быть произведено различными методами.

Прямой метод стандартизации смертности используется в предположении, что возрастной состав сравниваемых совокупностей одинаков, а повозрастная смертность этих совокупностей различна.

Если сравнивают только две совокупности, тогда взвешивание повозрастных показателей каждой из совокупностей производится по стандартным весам. В качестве стандартных весов можно принять структуру одной из совокупностей или среднюю структуру.

Результат сравнения стандартизованных показателей смертности (т. е. взвешенных) характеризует различие в средних при устранении влияния различных весов.

Полученное значение стандартизованных показателей смертности можно использовать только для сравнения, абсолютное же их значение принимать во внимание не следует.

Формула индекса смертности примет следующий вид:

$$I_{\text{смертн}} = \frac{\sum q'' w'}{\sum q' w'},$$ если структура первой совокупности принята за стандарт;

$$I_{\text{смертн}} = \frac{\sum q'' w''}{\sum q' w''},$$ если структура второй совокупности принята за стандарт;

$$I_{\text{смертн}} = \frac{\sum q'' w}{\sum q' w},$$ если за стандарт принята структура, отличная от структуры первой и второй совокупностей.

Рассмотрим применение прямого метода стандартизации показателя смертности на примере медицинской статистики.

Имеются данные о смертности (летальности) в детской больнице за 1967 и 1968 гг. Причем в 1968 г. летальность увеличилась по сравнению с 1967 г. от 0,042 (4,2%) до 0,046 (4,6%). Состав больных детей в эти годы был различен. В 1967 г. преобладали больные дети старше одного года, имеющие более низкие показатели летальности, а в 1968 г. удельный вес детей моложе одного года вырос. Требуется определить, имело ли место ухудшение лечения.

Таблица 21

Смертность в детской больнице

Возраст детей (в годах)	Число больных детей				Число умерших		Удельный вес умерших в числе больных (летальность)	
	1967 г.		1968 г.					
	абсолютное	в долях единицы w'	абсолютное	в долях единицы w''	1967 г.	1968 г.	1967 г. q'	1968 г. q''
До 1	200	0,2	300	0,6	16	18	0,08	0,06
1	600	0,6	100	0,2	18	2	0,03	0,02
2	200	0,2	100	0,2	8	3	0,04	0,03
Итого	1000	1,0	500	1,0	42	23	0,042	0,046

Получается парадоксальное положение, при котором в 1968 г. итоговая (общая средняя) летальность (0,046) выше, чем в 1967 г. (0,042), хотя по всем возрастам летальность в 1968 г. ниже. Причина парадокса в различии в структурах возрастного распределения больных детей в эти годы.

Устраним влияние различия в структурах и стандартизуем возрастное распределение больных детей.

В качестве стандарта возьмем сначала возрастное распределение больных детей в 1967 г., тогда в 1967 г. общая летальность ($\bar{q}' = 0{,}042$) останется без изменения. Для 1968 г. произведем пересчет и найдем $\bar{q}''$ стандартизованное.

Таблица 22

Стандартизация показателей смертности

Возраст детей (в годах)	Принятое за стандарт распределение больных детей в 1967 г.		Летальность в 1968 г. q''	$q''w'$
	абсолютное число	в долях единицы w'		
До 1	200	0,2	0,06	0,012
1	600	0,6	0,02	0,012
2	200	0,2	0,03	0.006
Итого . .	1 000	1,0	—	0,030

Отсюда $\bar{q}''_{\text{стандарт}} = \sum q''w' = 0{,}03$, или 3%.

Таким образом, после стандартизации летальность в 1968 г. оказалась значительно ниже, чем в 1967 г.:

$$I_{\text{летальности}} = \frac{\bar{q}''_{\text{стандарт}}}{\bar{q}'} = \frac{0{,}030}{0{,}042} \approx 0{,}71.$$

Если за стандарт принять возрастное распределение больных детей в 1968 г., то получим следующую стандартизованную летальность в 1967 г.:

$\bar{q}'_{\text{стандарт}} = \sum q'w'' = 0{,}08 \cdot 0{,}6 + 0{,}03 \cdot 0{,}2 + 0{,}04 \cdot 0{,}2 = 0{,}048 + 0{,}006 + 0{,}008 = 0{,}062$, а отношение стандартизованных средних мало изменится:

$$I_{\text{летальности}} = \frac{\bar{q}''}{\bar{q}'_{\text{стандарт}}} = \frac{0{,}046}{0{,}062} \approx 0{,}74.$$

При сравнении смертности двух совокупностей в качестве стандарта можно взять средний состав сравниваемых совокупностей (или лет). Так, продолжая предыдущий пример, получаем всего больных в 1967 и 1968 гг. 1000 + 500 = 1500, а в среднем 750; среднее число больных детей по возрастам:

до 1-го года	—250 (0,333)
1—2 года	—350 (0,467)
2—3 года	—150 (0,200)
Всего	750 (1,000)

Произведем стандартизацию летальности в 1967 г. и 1968 г. и найдем индекс летальности. В 1967 г. $\bar{q}'_{\text{стандарт}} = 0{,}08 \cdot 0{,}333 + 0{,}03 \times 0{,}467 + 0{,}04 \cdot 0{,}20 = 0{,}04865$. В 1968 г. $\bar{q}''_{\text{стандарт}} = 0{,}06 \cdot 0{,}333 + 0{,}02 \cdot 0{,}467 + 0{,}03 \cdot 0{,}200 = 0{,}03532$.

И в этом случае

$$I_{\text{летальности}} = \frac{\bar{q}''_{\text{стандарт}}}{\bar{q}'_{\text{стандарт}}} = \frac{0{,}03532}{0{,}04865} \approx 0{,}73,$$

т. е. соотношение стандартизованных средних мало отличается от полученных ранее. Несмотря на малые отличия индексов летальности, полученных с использованием различных стандартов, а именно 0,71; 0,74 и 0,73, можно сделать вывод о влиянии выбора стандарта на величину индекса летальности.

Прямой метод стандартизации смертности обладает тем достоинством, что для его применения достаточно знания повозрастных коэффициентов смертности сравниваемых совокупностей. Что же касается возрастного распределения, то оно берется стандартным. При этом не обязательно стандартное распределение должно совпадать с каким-либо из распределений сравниваемых совокупностей.

Недостатком же прямого метода является потеря специфичности изучаемых совокупностей.

В качестве стандарта возрастного распределения иногда рекомендовался «стандарт — население» — средний возрастной состав 17 европейских стран за период, близкий к 1900 г. Сейчас он, конечно, несколько устарел, но иногда очень удобен[1].

Косвенный метод стандартизации демографических показателей разработан в предположении, что возрастной состав сравниваемых совокупностей различен, а повозрастная смертность этих совокупностей одинакова (стандартна). Этот метод стандартизации используется в тех случаях, когда распределение умерших по возрасту неизвестно либо когда имеются основания предполагать, что повозрастные показатели смертности почему-либо недостоверны.

Сущность косвенного метода стандартизации состоит в предварительном сопоставлении средней смертности каждой совокупности со средней смертностью в стандарте, взвешенной по возрастному распределению каждой сравниваемой совокупности. Полученные результаты умножаются на среднюю из повозрастных смертностей стандарта. В качестве стандарта берут возрастные коэффициенты смертности какой-либо определенной группы населения.

К принятым ранее обозначениям добавим Q — коэффициент смертности в стандарте и P — удельные веса (доли) каждой возрастной группы в стандарте. Получаем формулы стандартизованных коэффициентов смертности:

[1] См. А. М. Мерков. Общая теория и методика санитарно-статистических исследований, стр. 90—91.

$$\bar{q}'_{\text{стандарт}} = \frac{\sum q' w'}{\sum Q w'} \sum QP = \bar{q}' \, \frac{\sum QP}{\sum Q w'},$$

$$\bar{q}''_{\text{стандарт}} = \frac{\sum q'' w''}{\sum Q w''} \sum QP = \bar{q}'' \, \frac{\sum QP}{\sum Q w''}.$$

Значит, стандартизованный коэффициент смертности каждой совокупности представляет собой произведение нестандартизованного коэффициента смертности этой совокупности на своеобразный (по аналогии с теорией статистики) агрегатный индекс доли, взвешенный по смертности стандарта. Сопоставив стандартизованные коэффициенты смертности двух совокупностей, получаем:

$$I_{\text{смертности}} \frac{\bar{q}''_{\text{стандарт}}}{\bar{q}'_{\text{стандарт}}} = \frac{\bar{q}''}{\bar{q}'} : \frac{\sum w'' Q}{\sum w' Q},$$

где делимое $\frac{\bar{q}''}{\bar{q}'}$ — индекс средних коэффициентов смертности, делитель — агрегатный индекс долей (или численностей) сравниваемых совокупностей, взвешенных по стандартной смертности.

Стандартизация коэффициентов косвенным методом облегчается, если в качестве стандарта принимается смертность одной из сравниваемых совокупностей.

Продолжая предыдущие примеры, из которых известно, что $\bar{q}' = 0{,}042$ и $\bar{q}'' = 0{,}046$, и вводя стандарт, представляющий собой среднюю из смертностей за 1967 и 1968 гг., составляем табл. 23.

Таблица 23

Возраст детей (в годах)	Стандарт: Общее число больных детей, абсолютное	в долях P	Число умерших	Летальность Q	w'	w''	Qw'	Qw''
До 1	500	0,333	34	0,068	0,2	0,6	0,0136	0,041
1	700	0,467	20	0,029	0,6	0,2	0,0174	0,005
2	300	0,200	11	0,037	0,2	0,2	0,0074	0,007
Итого	1 500	1,000	65	0,043	1,0	1,0	0,0384	0,053

$$\bar{q}'_{\text{стандарт}} = \bar{q}' \, \frac{\sum QP}{\sum Qw'} = 0{,}042 \cdot \frac{0{,}043}{0{,}0384} = 0{,}042 \cdot 1{,}12 = 0{,}047, \quad \text{или} \quad 4{,}7\%.$$

$$\bar{q}''_{\text{стандарт}} = \bar{q}'' \, \frac{\sum QP}{\sum Qw''} = 0{,}046 \cdot \frac{0{,}043}{0{,}0530} = 0{,}046 \cdot 0{,}81 \approx 0{,}037, \quad \text{или} \quad 3{,}7\%.$$

$$I_{\text{летальности}} = \frac{\bar{q}''_{\text{стандарт}}}{\bar{q}'_{\text{стандарт}}} = \frac{0{,}037}{0{,}047} \approx 0{,}79.$$

Индекс летальности мало отличается от полученных ранее.

Можно было найти индекс летальности и без предварительных расчетов стандартизованных коэффициентов:

$$I_{\text{летальности}} = \frac{\bar{q}''}{\bar{q}'} : \frac{\sum w'' Q}{\sum w' Q} = \frac{0{,}046}{0{,}042} : \frac{0{,}0530}{0{,}0384} = 1{,}095 : 1{,}380 = 0{,}79,$$

В качестве стандарта возьмем летальность 1967 г. Тогда $\bar{q}'_{\text{стандарт}} = 0{,}042$. Найдем $\bar{q}''_{\text{стандарт}}$.

Таблица 24

Возраст детей (в годах)	Стандарт: Общее число больных детей, абсолютное	в долях P	Летальность Q	w''	$w''Q$
До 1	200	0,2	0,08	0,6	0,048
1	600	0,6	0,03	0,2	0,006
2	200	0,2	0,04	0,2	0,008
Итого . .	1 000	1,0	0,042	1,0	0,062

$$\bar{q}''_{\text{стандарт}} = \bar{q}'' \, \frac{\sum QP}{\sum Qw''} = 0{,}046 \cdot \frac{0{,}042}{0{,}062} = 0{,}046 \cdot 0{,}68 = 0{,}031.$$

$$I_{\text{летальности}} = \frac{0{,}031}{0{,}042} \approx 0{,}74.$$

Разбирая отчет Боткинской больницы, из которого следует, что применение сыворотки повышает смертность больных дифтерией, и, используя косвенный метод стандартизации смертности, С. А. Новосельский доказал, что применение противодифтеритной сыворотки не только не повысило смертность, а снизило ее с 15,9 до 12,3%, т. е. почти на четверть. Такой результат С. А. Новосельский получил, принимая процент повозрастной смертности в период, когда сыворотка не применялась, за стандрат и подсчитывая ожидаемое число смертей по этому стандарту в период, когда стала применяться сыворотка.

«Обратный» метод стандартизации применяется в тех случаях, когда возрастной состав населения при изучении смертности или больных при изучении летальности неизвестен, а известен возрастной состав умерших. Стандартизация может быть произведена по способу, предложенному Д. Керриджем в 1958 г. и описанному А. М. Мерковым в работе «Общая теория и методика санитарно-статистических исследований»[1]. Этот метод называется «обратным». Возьмем, например,

[1] См. А. М. Мерков. Общая теория и методика санитарно-статистических исследований.

4 Зак. 1603

город-новостройку. Численность его возросла за несколько лет в четыре раза: с 70 тыс. в 1958 г. (S_{1958}) до 280 тыс. в 1970 г. (S_{1970}). Умерших в 1958 г. было 1400 человек, и коэффициент смертности составлял 20‰ (K_{1958}), а в 1970 г. умерших было 5320 человек, и коэффициент смертности составлял 19‰ (K_{1970}).

Индекс фактической смертности составляет:

$$I_{ф.с} = \frac{K_{1970}}{K_{1958}} = \frac{19{,}0}{20{,}0} = 0{,}95.$$

Однако простое сопоставление этих коэффициентов ничего не говорит о существе изменений. Имеются основания предположить, что в возрастном составе населения за 1958—1970 гг. произошли сдвиги. Для оценки влияния этих сдвигов на сравнительную величину показателей смертности необходимо произвести стандартизацию. Однако в данном случае ни прямой, ни косвенный методы не применимы. Можно, однако, в качестве стандарта взять общие и повозрастные показатели смертности какого-либо другого города и, исходя из известного нам количества умерших, вычислить для каждой повозрастной группы ту «ожидаемую» численность населения, при которой данное число умерших соответствовало бы возрастным показателям смертности, принятым за стандарт.

Возьмем за стандарт показатели смертности всего городского населения в 1970 г. Допустим, что из общего числа 5320 умерших в 1970 г. в городе A в возрасте до 5 лет умерших было 45. Пусть смертность в возрасте до 5 лет составляла 15‰. Тогда «ожидаемая» численность населения в городе A в возрасте до 5 лет составляла $\frac{45 \cdot 1\,000}{15} = 3000$ человек. Так, рассчитывая по всем возрастным интервалам, находим суммарную «ожидаемую» численность населения в 1970 г. — S'_{1970} и в 1958 г. — S'_{1958}.

Пусть «ожидаемые» численности оказались равны: $S'_{1970} = 168\,000$ человек; $S'_{1958} = 238\,000$ человек.

Находим отношения «ожидаемых» численностей (S') к фактическим (S):

$$z_{1970} = \frac{S'_{1970}}{S_{1970}} = \frac{168\,000}{280\,000} = 0{,}6$$

и

$$z_{1958} = \frac{S'_{1958}}{S_{1958}} = \frac{238\,000}{70\,000} = 3{,}4.$$

Теперь можно получить стандартизованный показатель смертности в каждом году. Для этого умножим общую смертность населения, принятого за стандарт, на отношение между «ожидаемой» смертностью и фактической. Сравнение двух уже стандартизованных коэффициентов смертности даст более правильный ответ на вопрос о том, насколько снизилась смертность в 1970 г. по сравнению с 1958 г.

Применение методов стандартизации

Метод стандартизации применялся первоначально к показателям смертности. Но рассматриваемые нами методы стандартизации пригодны не только для изучения смертности, но и для элиминирования влияния возрастной структуры всех женщин на рождаемость, фертильного контингента женщин на плодовитость.

При помощи стандартизации можно определить также масштабы изменения (роста или снижения) численности рабочей силы. Например, установить, каким бы при неизменности пропорций самодеятельного населения было число самодеятельных, если бы в течение периода сравнения численности данной категории лиц изменялись только под влиянием демографических факторов. Получив эти данные, можно косвенным путем выявить характер изменений, обусловленных влиянием недемографических факторов. Постепенно, проводя стандартизацию по отдельным факторам, можно измерить влияние каждого из них в отдельности.

Интересную закономерность при применении стандартизации показателей смертности установил Б. Ц. Урланис. Производя расчеты, он выявил, что в большинстве стран нестандартизованные коэффициенты смертности имеют гораздо большую тенденцию к снижению, чем стандартизованные. Это обстоятельство позволило ему прийти к целому ряду выводов.

Подводя итог сказанному, можно сделать вывод о целесообразности более широкого использования метода стандартизации, чем это практиковалось до сих пор. Этот метод помогает выявить и измерить влияние ряда факторов на какое-либо демографическое явление.

Исчисление стандартизованных показателей и индексов смертности

Приведем некоторые формулы, которыми можно пользоваться для исчисления стандартизованных показателей и индексов смертности.

1. «Грубый» показатель смертности

$$\frac{\Sigma K_{ф(x)} \bar{S}_{ф(x)}}{\bar{S}_{ф}},$$

где $\bar{S}_{ф(к)}$ — среднегодовая фактическая численность населения;
$\bar{S}_{ф}$ — общая фактическая численность населения;
$K_{ф(x)}$ — фактический повозрастной коэффициент смертности (центральный показатель смертности).

2. Показатель смертности при прямом методе стандартизации

$$\frac{\Sigma K_{ф(x)} \bar{S}_{S(x)}}{\Sigma \bar{S}_{S(x)}},$$

где $\bar{S}_{S(x)}$ — среднегодовая численность населения, принятого за стандарт.

4*

3. «Сопоставимый» показатель смертности

$$\frac{1}{2}\sum K_{\text{ф}(x)}\left(\frac{\bar{S}_{\text{ф}(x)}}{\sum \bar{S}_{\text{ф}(x)}}+\frac{\bar{S}_{S(x)}}{\sum \bar{S}_{S(x)}}\right).$$

4. Табличный показатель смертности

$$\frac{\sum K_{\text{ф}(x)}\, L_x}{\sum L_x}.$$

5. Показатель смертности при косвенном методе стандартизации

$$\frac{\dfrac{M_{\text{ф}}}{\bar{S}_{\text{ф}}}\cdot\dfrac{M_S}{\bar{S}_S}}{\dfrac{\sum K_{S(x)}\,\bar{S}_{\text{ф}(x)}}{\sum \bar{S}_{\text{ф}(x)}}},$$

где $M_{\text{ф}(x)}$ — фактическое число случаев смерти в возрасте x лет; $\sum M_{\text{ф}(x)} = M_{\text{ф}}$ — общее число смертей.

6. Индекс смертности

$$\frac{\sum \dfrac{K_{\text{ф}(x)}}{K_{S(x)}}\,n_x}{\sum n_x},$$

где n_x — относительная продолжительность возрастного интервала.

7. Относительный индекс смертности

$$\frac{\sum \dfrac{K_{\text{ф}(x)}}{K_{\bar{S}(x)}}\,\bar{S}_{\text{ф}}}{\sum \bar{S}_{\text{ф}}}.$$

Если рассматривать n_x как постоянные величины, а $K_{\text{ф}(x)}$, т. е. повозрастные коэффициенты смертности, как случайные переменные[1], можно получать выборочные дисперсии стандартизованных и повозрастных показателей смертности, а также квадратические ошибки стандартизованных показателей смертности. Эти ошибки представляют собой расхождение между неизвестной (генеральной) силой смертности и найденными нами выборочными стандартизованными показателями смертности. По сути дела такая постановка вопроса означает признание вероятностной основы приближения варьирующего показателя смертности к постоянной величине — силе смертности и предполагает выборочную изменчивость показателей смертности и необходимость средней квадратической ошибки, измеряющей эту изменчивость.

[1] Это делается потому, что стандартизация показателей смертности производится для сравнения смертности различных совокупностей с помощью стандарта, в котором пропорции могут быть представлены постоянными величинами.

Однако, на наш взгляд, исчисление и использование квадратических ошибок стандартизованных показателей смертности не всегда необходимо, ибо стандартизованные показатели смертности используются только для сравнения, а их абсолютная величина не должна приниматься во внимание.

Определение исправленного коэффициента естественного прироста

Использование метода стандартизации при изучении рождаемости и плодовитости встречается реже, чем при изучении смертности. В связи с этим при получении такой важной характеристики, как коэффициент естественного прироста, по формуле $K_{\text{ест.пр}} = n - m$ встречаются трудности. Для получения $K_{\text{ест.пр}}$ коэффициенты рождаемости и смертности должны быть однотипными. Если смертность стандартизована, а рождаемость оставлена без учета возрастной структуры, то разность между ними не будет однотипна. Поэтому возникает необходимость в стандартизации и коэффициента рождаемости. Разность между стандартизованными показателями рождаемости и смертности дает исправленный коэффициент естественного прироста.

Имеется и другой способ нахождения усовершенствованного коэффициента естественного прироста, предложенный Ландри. Этот способ, ориентированный на возрастную структуру стационарного населения, свободного от влияния всех других факторов, кроме повозрастной смертности, позволяет стандартизовать коэффициент рождаемости, смертности, а следовательно, и естественного прироста.

Для стандартизации используется таблица смертности (специфическая для изучаемого населения). В качестве стандарта берется повозрастное распределение женского населения. Стандартизуя коэффициенты плодовитости, можно рассчитать «ожидаемое» число рождений и путем последовательного их деления на фактическое число фертильных женщин и общую сумму стационарного населения получить коэффициенты плодовитости и рождаемости. После определения коэффициента смертности стационарного населения и нахождения соответствующей разности получаем исправленный коэффициент естественного прироста.

Расчет по данным УССР 1926—1927 гг. привел к следующим результатам.

Таблица 25

Обычные и исправленные коэффициенты естественного движения населения (в ‰)

Коэффициенты естественного движения населения	Обычные	Исправленные
Рождаемости	41,2	36,0
Смертности	17,9	21,2
Естественного прироста	23,3	14,8

Р. Кучинский показал, что уже в 20-х годах нашего века режим воспроизводства передовых капиталистических стран фактически привел к естественной убыли населения. Что же касается коэффициента естественного прироста, то он оставался положительным лишь благодаря благоприятной возрастной структуре (преобладанию возрастных групп с пониженной смертностью по сравнению с общей смертностью). Эта временная демографическая ситуация создавала иллюзию относительного благополучия. В 1930 г. в период мирового экономического кризиса Германское статистическое управление произвело расчет для ряда европейских стран, «очистив» (от влияния особенностей возрастного состава населения) коэффициенты рождаемости и смертности. Полученные на основании «очищенных» коэффициентов рождаемости и смертности коэффициенты естественного прироста приведены в табл. 26[1].

Таблица 26

Обычные (фактические) и «очищенные» коэффициенты естественного движения населения по данным 1930 г. (в ‰)

Страны	Коэффициенты рождаемости		Коэффициенты смертности		Коэффициенты естественного прироста	
	обычный	«очищенный»	обычный	«очищенный»	обычный	«очищенный»
1	2	3	4	5	6	7
Англия	16,3	14,6	11,4	17,4	+4,9	—2,8
Германия	17,6	15,8	11,0	17,4	+6,6	—1,6
Франция	18,0	17,2	15,6	18,5	+2,4	—1,3
Италия	26,7	23,6	14,0	20,0	+12,7	+3,6

Следует иметь в виду условность сравнения коэффициентов рождаемости, смертности и естественного прироста населения различных стран, стандартизованных по стационарному населению. Дело в том, что при исчислении всех этих показателей стандартом являлись таблицы смертности этих же стран. Налицо положение, когда прирост населения некоторых стран обеспечивается не интенсивностью воспроизводства, а лишь особенностью возрастной структуры; элиминирование ее влияния выявляет факт суженного воспроизводства.

Метод Н. А. Вигдорчика

Интересен метод, разработанный Н. А. Вигдорчиком. Суть его изложена применительно к исследованию инвалидности населения с привлечением «коэффициента изнашиваемости».

[1] «Bevölkerungsbewegung in Europäischen Ländern. «Statistik des Deutschen Reichs». Berlin, 1931.

Известно, что степень инвалидности различных профессиональных групп зависит от ряда причин: среднего возраста каждой категории инвалидности; количества инвалидов, выделяемых каждой группой за определенный период (скажем, за год); от распределения инвалидов по степени потери трудоспособности и др. Для объединения влияния всех вышеуказанных причин автор предложил показатель потерь рабочей силы каждой исследуемой группой. Для этого определяется крайний возраст наступления полной инвалидности и рассчитывается количество годовых единиц рабочей силы, утрачиваемых с момента наступления инвалидности. Результаты суммируются. Отнесение общей суммы «потерь» к численности группы дает «коэффициент изнашиваемости».

Допустим, имеются две совокупности: в первой — 1000, а во второй — 2000 рабочих. В течение года в каждой совокупности было зафиксировано число инвалидов различной категории (табл. 27).

Таблица 27

Фактические потери трудоспособности инвалидов

Категории инвалидности*	Первая совокупность		Вторая совокупность		Средняя потеря трудоспособности при данной категории инвалидности, в %
	число инвалидов	средний возраст	число инвалидов	средний возраст	
VI	10	40,0	15	35,0	10
V	—	—	6	40,0	22
IV	5	45,0	—	—	37
III	—	—	5	45,0	52
II	3	50,0	—	—	100
I	1	60,0	2	43,0	150**

* Категории инвалидности устанавливаются на основании принятого критерия.
** Имеется в виду 100% потери трудоспособности инвалидом и 50% затрат труда на уход за ним ввиду его беспомощности.

Исходя из условия, что физиологическая инвалидность наступает в 66 лет, делаем расчет (табл. 28).

Первая совокупность теряет в год 121,85 единиц рабочей силы, а на каждого рабочего из 1000 рабочих этой совокупности приходится потерь в год 0,122 единицы. Это и есть «коэффициент изнашиваемости» первой совокупности — $K_{\text{изнаш}(1)}$. Во второй совокупности, составляющей 2000 рабочих, $K_{\text{изнаш}(2)}$ составляет 0,102. Следовательно, условия в первой совокупности менее «благоприятны», чем во второй, ибо $K_{\text{изнаш}(1)} > K_{\text{изнаш}(2)}$.

Однако этот метод решает весьма ограниченную, узкую задачу и обладает рядом недостатков:

1) не учитываются возрастные структуры совокупностей, из которых происходят инвалиды; смертность инвалидов и вероятности их

Таблица 28

Потери трудоспособности инвалидов в годовых единицах

Категория инвалидности	Первая совокупность		Вторая совокупность	
	Остается лет до физиологической инвалидности	Потери в годовых единицах	Остается лет до физиологической инвалидности	Потери в годовых единицах
VI	26	26·10·10% = 26,00	31	31·15·10% = 46,50
V	—	—	26	26·6·22% = 34,32
IV	21	21·5·37% = 38,85	—	—
III	—	—	21	21·5·52% = 54,60
II	16	16·3·100%=48,00	—	—
I	6	6·1·150% = 9,00	23	23·2·150% = 69,00
Итого	—	121,85	—	204,42

перехода из одной категории инвалидности в другую; повозрастная производительность труда и т. д.;

2) допущение о наступлении физиологической инвалидности в 66 лет весьма условно. Очевидно, что для разных профессий эта величина различна. Лучше было бы ориентироваться на среднюю продолжительность жизни, как это делал Герш при подсчете потерь вследствие войн, а еще точнее ориентировка на среднюю продолжительность жизни в рабочем возрасте.

Учитывая остроту проблемы инвалидности и травматизма, более глубокое изучение этих вопросов должно основываться на таблицах инвалидности (см. стр. 142 и далее).

ИЗМЕРЕНИЕ СМЕРТНОСТИ (ЛЕТАЛЬНОСТИ) В УСЛОВИЯХ ПОДВИЖНОСТИ НАСЕЛЕНИЯ

Показатель летальности

Показатель летальности может быть построен аналогично коэффициенту смертности. Однако специфичность этого показателя дает возможность строить его и на другой основе.

Пусть L_x — число доживших до возраста x под наблюдением; A — число прибывших под наблюдение в промежутке возраста от x до $x+1$ (предполагается равномерное распределение прибывших во времени); B — число выбывших из-под наблюдения в возрасте от x до $x+1$ (также равномерно); M — число умерших под наблюдением.

Возьмем возраст от x до $x+z$, где $0<z<1$. При условии равномерности в бесконечно малом интервале времени от $x+z$ до $x+z+dz$ в совокупности произойдут следующие изменения:

поступит под наблюдение Adz лиц, которые дожили до момента возраста $x+z$, из числа $A\frac{l_x}{l_{x+z}}dz$, имевших возраст x (не входящих под наблюдение). Значит, до начала наблюдения умерло $A\left(\frac{l_x}{l_{x+z}}-1\right)dz$;

выбудет из-под наблюдения в бесконечно малый промежуток времени Bdz. Из них доживает до возраста $x+1$ (уже вне наблюдения) $B\frac{l_{x+1}}{l_{x+z}}dz$ и умрет после выхода из-под наблюдения $B\left(1-\frac{l_{x+1}}{l_{x+z}}\right)dz$.

Таким образом, общее число лиц, бывших под наблюдением в момент возраста x, может быть исчислено по формуле

$$L_x+Al_x\int_0^1\frac{dz}{l_{x+z}}.$$

Из этого числа умрут до достижения возраста $x+1$

$$M+A\left(l_x\int_0^1\frac{dz}{l_{x+z}}-1\right)-B\left(1-l_{x+1}\int_0^1\frac{dz}{l_{x+z}}\right).$$

Найдем вероятность смерти:

$$q_x=\frac{M+A\left(l_x\int_0^1\frac{dz}{l_{x+z}}-1\right)-B\left(1-l_{x+1}\int_0^1\frac{dz}{l_{x+z}}\right)}{L_x+Al_x\int_0^1\frac{dz}{l_{x+z}}}.$$

После преобразований получим:

$$L_xq_x+(A-B)\left(1-l_{x+1}\int_0^1\frac{dz}{l_{x+z}}\right)=M.$$

Предполагая, что в интервале от x до $x+1$ функция линейна, получаем:

$$l_{x+z}=l_x-z\,(l_x-l_{x+1}).$$

Тогда

$$\int_0^1\frac{dz}{l_{x+z}}=\frac{1}{l_x-l_{x+1}}\ln\frac{l_x}{l_{x+1}}=-\frac{1}{l_x-l_{x+1}}\ln(1-q_x).$$

Продолжая вывод, сделаем соответствующую подстановку

$$L_xq_x+(A-B)\left[1+\frac{l_{x+1}}{l_x-l_{x+1}}\ln(1-q_x)\right]=M$$

или

$$L_x q_x + (A-B)\left[1 - \frac{1-q_x}{q_x}\ln(1-q_x)\right] = M.$$

Используя разложение в ряд Маклорена и пренебрегая q_x более высоких степеней, чем первая, получим:

$$L_x q_x + (A-B)\frac{q_x}{2} = M,$$

откуда

$$q_x = \frac{M}{L_x + \frac{A-B}{2}}.$$

Значит, при наших предположениях вероятность умереть равна отношению числа умерших к сумме числа доживающих до возраста x под наблюдением и полуразности числа прибывших под наблюдение и числа выбывших из-под наблюдения в возрасте от x до $x+1$.

Более точно данная формула при тех же допущениях может быть записана так:

$$q_x = \frac{M}{L_x + \frac{A-B}{2}} - \frac{\frac{1}{6}(A-B)M^2}{L_x + \frac{1}{2}(A-B)^3}.$$

Если поступление и выбытие людей будет происходить не непрерывно, а распределяться поровну на m сроков в году, то можно вывести следующую формулу

$$q_x = \frac{M}{L_x + \frac{m+1}{2m}A - \frac{m-1}{2m}B}.$$

При $m=5$ получаем:

$$q_x = \frac{M}{L_x + 0{,}6A - 0{,}4B}.$$

Если допустить, что смертность в течение года постоянна, то вид формулы немного изменится:

$$q_x = \frac{M}{L_x + \frac{A-B}{2}} - \frac{\frac{1}{12}(A-B)M^2}{\left[L_x + \frac{1}{2}(A-B)\right]^3}.$$

В. В. Паевский, используя наиболее простую из этих формул, а именно:

$$q_x = \frac{M}{L_x + \frac{A-B}{2}},$$

установил, что она применима в тех случаях, когда разность $A-B$ незначительна.

С. А. Новосельский и П. А. Кувшинников также использовали эту формулу с различными модификациями.

Метод измерения смертности подвижного населения Г. А. Баткиса

Особого внимания заслуживает метод, предложенный Г. А. Баткисом для измерения смертности подвижного населения (для больничной летальности). Рассмотрим этот метод и укажем возможности дальнейшего его развития. В основе метода лежит предположение об общем порядке вымирания, не зависящем от условий подвижности населения. Иными словами, смертность принимается одинаковой в месте наблюдения и вне его.

Возьмем четыре частные совокупности, определяющие генеральную совокупность, и к каждой из них отнесем умерших за данный период наблюдения в данном месте и вне его.

Таблица 29

Частные совокупности	В данном месте		Прибывшие в данное место среди периода наблюдения	
	доживающие до конца наблюдения	выбывающие из него среди периода наблюдения	доживающие до конца наблюдения	выбывающие среди периода наблюдения
Первая	+			
Вторая		+		
Третья			+	
Четвертая				+

Частные совокупности могут быть определены при помощи статистической таблицы, в которой произведена группировка лиц по признакам:

присутствие в данном месте к началу наблюдения или отсутствие;

дожитие в данном месте до конца периода наблюдения или выбытие из него среди этого периода.

Обозначим L — показатель смертности (летальности). Далее берем величины из данных учета механического движения:

A — прибывшие в данное место среди периода наблюдения;

B — находившиеся в данном месте к началу периода наблюдения;

C — находившиеся в данном месте в конце периода наблюдения;

D — умершие в данном месте в период наблюдения;

E — выбывшие среди периода наблюдения.

Путем рассуждений можно установить, что показатель летальности (L) заключен в границах:

1) либо $\frac{D}{A} < L < \frac{D}{E+D}$;

2) либо $\frac{D}{A} > L > \frac{D}{E+D}$.

Если принять, что величина $\overline{L}$ — средняя арифметическая своих крайних значений, тогда окончательно:

$$\overline{L} = \frac{1}{2}\left(\frac{D}{A} + \frac{D}{E+D}\right) \pm \frac{1}{2}\left(\frac{D}{A} - \frac{D}{E+D}\right).$$

Рассмотрим применение указанных методов на примере движения больных.

Таблица 30

Движение больных в течение месяца

Состояло на начало отчетного периода B	Поступило A	Выписано E	Умерло D	Осталось на конец отчетного периода C
70	140	74	6	130

Вычислим:

$$L_1 = \frac{D}{A} = \frac{6}{140} = 0{,}043,$$

$$L_2 = \frac{D}{E+D} = \frac{6}{74+6} = \frac{6}{80} = 0{,}075,$$

$$L_3 = \frac{D}{\frac{A+E+D}{2}} = \frac{6}{\frac{140+74+6}{2}} = 0{,}055.$$

Первые два показателя представляют собой ранее рассмотренные нами границы смертности (летальности): $0{,}043 < L < 0{,}075$. Третий показатель пригоден для условия быстрой сменяемости больных. Установим величину этого показателя по методу Г. А. Баткиса:

$$\overline{L} = \frac{0{,}043+0{,}075}{2} \pm \frac{0{,}043-0{,}075}{2} = 0{,}059 \pm 0{,}016.$$

Как мы видим, величина $\overline{L} = 0{,}059$ при методе Г. А. Баткиса чуть выше L_3, полученной при использовании другого метода (0,055).

Перевод метода Г. А. Баткиса на вероятностную основу

Крайне интересно, что можно дополнить метод Г. А. Баткиса и поставить исчисление коэффициента смертности (летальности) на вероятностную основу.

Определив долю умерших в средней численности состоящих на начало периода и оставшихся на конец, т. е. $w = D : \frac{B+C}{2} = \frac{2D}{B+C}$, найдем дисперсию альтернативного признака:

$$\sigma^2 = w(1-w) = \frac{2D}{B+C}\left(1 - \frac{2D}{B+C}\right) =$$

$$= \frac{2D}{B+C}\left(\frac{B+C-2D}{B+C}\right) = \frac{2D(B+C-2D)}{(B+C)^2}.$$

Находим предельную ошибку выборки:

$$\Delta = t\sqrt{\frac{2D\cdot 2(B+C-2D)}{(B+C)^2(B+C)}},$$

$$\Delta = \frac{2}{B+C}\,t\sqrt{\frac{D(B+C-2D)}{B+C}},$$

откуда нетрудно определить t, а затем по таблице найти вероятность P.

Продолжая предыдущий пример и принимая разность $L_2 - L_1$ за величину ошибки без деления на два, как это делал Г. А. Баткис, получаем:

$$\frac{B+C}{2} = \frac{70+130}{2} = 100,$$

$$w = \frac{6}{100} = 0{,}06,$$

$$1 - w = 0{,}94,$$

$$w(1-w) = 0{,}06\cdot 0{,}94 = 0{,}0564,$$

$$\Delta = t\sqrt{\frac{0{,}0564}{100}} = t\sqrt{0{,}000564} = t0{,}024.$$

Итак, $\Delta = 0{,}024\ t$. Зная, что без деления на два $\Delta = 0{,}032$, находим $t = \frac{0{,}032}{0{,}024} = 1{,}34$. Значит, вероятность $P = 0{,}8$. Следовательно, вероятность того, что смертность (летальность) лежит в границах $0{,}059 \pm 0{,}016$ равна 0,8, т. е. не дает еще полной уверенности в невыходе летальности за указанные границы. Если бы в предыдущем примере число умерших составляло не 6 (как в примере у Г. А. Баткиса), а, например, 20 человек, то расчет привел бы к другим показателям и вывод соответственно изменился. Величина C получается равной:

$$C = B + A - E - D = 70 + 140 - 74 - 20 = 116.$$

Тогда

$$\frac{B+C}{2} = \frac{70+116}{2} = \frac{186}{2} = 93,$$

$$w = \frac{20}{93} = 0{,}215; \quad 1 - w = 0{,}785; \quad w(1-w) = 0{,}1688.$$

Следовательно, $\Delta = t\sqrt{\frac{0,1688}{100}} \approx 0,04\,t.$

Следует, кроме этого, учитывать изменение и смертности (L) с ее ошибкой (Δ).

Отношение умерших к поступившим:

$$L_1 = \frac{D}{A} = \frac{20}{140} = 0,142.$$

Отношение умерших к выбывшим дает:

$$L_2 = \frac{D}{E+D} = \frac{20}{74+20} = \frac{20}{94} = 0,213.$$

Тогда смертность принимает следующее значение:

$$\overline{L} = \frac{0,142+0,213}{2} = \frac{0,355}{2} = 0,178,$$

а ошибка $\Delta = 0,213 - 0,142 = 0,071$.

Следовательно, $t = \frac{0,071}{0,04} \approx 1,7$, откуда вероятность $P > 0,9$.

Таким образом, получив вероятность $P>0,9$, мы могли бы с практической достоверностью утверждать, что в данном случае, смертность лежит в границах $0,178 \pm 0,04$. В настоящее время гипотеза о независимости процессов смертности и миграций может быть принята только с существенными оговорками. Большинство смертных случаев происходит в среде мало мобильной, состоящей из хронических больных.

Следовательно, при использовании метода Г. А. Баткиса необходимо учитывать, что вероятность выбытия из-под наблюдения больных и здоровых неодинакова: у больного меньшая вероятность выезда. Поэтому метод Г. А. Баткиса завышает смертность в местах эмиграции и занижает ее в местах иммиграции.

Глава 2 ТАБЛИЦЫ СМЕРТНОСТИ (ДОЖИТИЯ) И СРЕДНЕЙ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ

ЗНАЧЕНИЕ ТАБЛИЦ СМЕРТНОСТИ. ТАБЛИЦЫ СМЕРТНОСТИ В РОССИИ

Смерть одного человека — явление, чаще всего объяснимое биологическими особенностями, но смертность всего населения определяется уже социально-экономическими условиями общества.

С демографической точки зрения таблицы смертности можно рассматривать как теоретическую модель населения, которое постоянно пополняется за счет родившихся и сокращается за счет умерших. Таблицы смертности, иначе называемые таблицами дожития, используются для углубленного и подробного анализа силы смертности и представляют собой развернутую систему показателей повозрастной смертности и доживаемости при переходе от одного возраста к другому. Таблицы смертности показывают, каким образом определенная совокупность родившихся постепенно уменьшается в своем составе под влиянием смертности. Со статистической точки зрения таблицы смертности представляют собой статистические таблицы. Подлежащим в них является возраст. При этом возраст может быть дан в одногодичных интервалах от 0 до 100 лет (полные таблицы смертности). Иногда строят таблицы смертности с пятилетними и десятилетними интервалами (сокращенные). В некоторых таблицах смертности возраст до 20—25 лет детализирован по одногодичным интервалам, а далее — по пятилетним; сказуемым же являются показатели, рассмотренные нами ниже. Первая таблица такого рода, фиксирующая среднюю продолжительность предстоящей жизни, была приведена в «Digestes» Ульпиана (около 170 г. н. э.). Источник не содержит указаний на методы получения результатов[1] (см. табл. 31).

Однако более или менее заслуживающей названия «таблица смертности» была таблица Джона Граунта, появившаяся в 1662 г., и Эдмунда Галлея (астронома), появившаяся в 1693 г. в периодическом издании «Philosophical Fransactions». Таблица Э. Галлея была составлена на основании материала, собранного из церковных ведомостей Каспаром Нейманом.

[1] Некоторые исследователи, например Б. Ц. Урланис (см. «Ученые записки по статистике», т. VII, М., изд-во АН СССР, 1963, стр. 154), считают, что нет уверенности в том, что приведенные результаты представляют собой среднюю продолжительность предстоящей жизни для разных возрастов.

Таблица 31

Повозрастная средняя продолжительность жизни

Возраст	до 20 лет	20—25	25—30	30—35	35—40	40—41	41—42	42—43	43—44	44—45	45—46	46—47	47—48	48—49	49—50	50—55	55—60	60 и более
Средняя продолжительность предстоящей жизни . . .	30	28	25	22	20	19	18	17	16	15	14	13	12	11	10	9	7	5

Э. Галлей предложил очень простой метод нахождения порядка вымирания, при разработке которого он ориентировался на извлечение статистических данных из смертных списков; поэтому его часто называют порядком вымирания по методу смертных списков.

При этом методе вместо совокупностей M^1 в пределах, например, вертикальной полосы $t_7 t_8 y_7 y_8$ (см. рис. 3) берутся совокупности M^3 из наклонной полосы $t_7 t_8\ v_3 v_4$ и число смертных точек в пределах наклонной полосы или ΣM^3 приравнивается числу родившихся.

Таким образом, за порядок вымирания принимаются числа:

$$\sum_t M^3; \quad \sum_t M^3 - {}_t M_1^3; \quad \sum_t M^3 - \left({}_t M_1^3 + {}_t M_2^3\right) \text{ и т. д.}$$

Данный метод основан на гипотезе стационарного состояния. В самом деле, если бы в течение столетия число рождений и смертей во всех возрастах не менялось, то полученные этим методом результаты можно было бы признать правильными.

Однако в действительности этого нет, и поэтому такой порядок вымирания не дает правильных результатов. Возникает вопрос: а нельзя ли, учитывая недостатки этого метода, улучшить его и приблизить к действительному положению дел. Оказывается, можно при помощи ряда допущений.

Во-первых, учтем, что если население возрастает, то рождаемость выше смертности (при элиминировании миграции). Значит, если мы хотим исправить полученный порядок вымирания, мы должны распределить разность (в большинстве случаев положительную) между рождаемостью и смертностью по возрастным группам. При этом числа умерших, начиная с M_2^3, M_3^3, M_4^3 и т. д., должны быть увеличены тем больше, чем моложе соответствующие поколения.

Во-вторых, допустим, что числа рождений, уходя в прошлое, уменьшаются в арифметической прогрессии. Значит, множитель, на который нужно умножить M_x^3, примет вид:

$$\frac{1}{1-(x-1)\beta}.$$

Раскладывая в ряд Маклорена и принимая в качестве приближения первые два члена, найдем искомый множитель:

$$\frac{1}{1-(x-1)\beta} \approx 1+(x-1)\beta.$$

Приведенные нами допущения позволяют вычислить неизвестные нам искомые числа лиц из поколения родившихся данного года, которые со временем действительно умрут в соответствующих возрастах, т. е. M_1^1, M_2^1, M_3^1 и т. д. Получаем:

$$(1+0\beta)\, M_1^3 \approx M_1^1$$
$$(1+1\beta)\, M_2^3 \approx M_2^1$$
$$(1+2\beta)\, M_3^3 \approx M_3^1$$
$$\cdots\cdots\cdots$$
$$[1+(\omega-1)\beta]\, M_\omega^3 \approx M_\omega^1.$$

Сложим обе части равенств. Сумма правой части равенств даст $\Sigma M^1 = N$ и будет, следовательно, равна числу рождений. Суммируя левую часть равенств и раскрывая скобки, получаем

$$\beta\left[M_2^3 + 2M_3^3 + 3M_4^3 + \ldots + (\omega-1)\, M_\omega^3\right] = N - \sum M^3 = D.$$

Рассмотрим выражение в квадратных скобках, но предварительно прибавим к каждому члену этого выражения еще его половину, т. е. увеличим все выражение на $\frac{1}{2}\ \Sigma M^3$. Тогда получим:

$$A = \frac{1}{2} M_1^3 + \frac{3}{2} M_2^3 + \frac{5}{2} M_3^3 + \frac{7}{2} M_4^3 + \ldots + \left[(\omega-1)+\frac{1}{2}\right] M_\omega^3,$$

т. е. общее число лет, прожитое всеми лицами, умершими в течение данного года, если предположить, что для каждого возраста умерших в одногодичной группе берется их средний, и иначе говоря, центральный возраст (т. е. $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$, $\frac{7}{2}$ и т. д.[1]). Обозначая выражение в квадратных скобках через B, получаем $B = A - {}^1/_2\, M^3$; $\beta = \frac{D}{B}$ и, следовательно

$$\beta = \frac{D}{A - \frac{1}{2} M^3}.$$

Метод смертных списков как средство изучения динамики смертности в зависимости от возраста оказался несостоятельным, так как в результате его применения смертность завышалась.

Построение таблиц смертности затруднено двумя обстоятельствами: во-первых, необходимостью иметь для этой цели точные данные о чис-

[1] Исключение составляет первый год жизни, где средний возраст правильнее взять от 1/4 до 1/3 года.

ленности населения и числе умерших, а, во-вторых, сложностью их вычислений.

Еще 15—20 лет назад одним из критериев при построении таблиц смертности было упрощение техники математических расчетов. Сейчас при наличии электронно-вычислительных машин и разработанных алгоритмов сложность вычислений уже не является затруднением. Речь может идти о математических методах любой сложности, лишь бы они отражали объективный характер доживания населения.

Таблицы смертности могут касаться как всего населения страны, так и отдельных районов. Назначение таблиц смертности — дать материал:

для теоретического анализа характера смертности, позволяющего выяснить возрастной состав населения в настоящее время и наметить возможности перспективного изменения этого состава на будущее;

для расчетов численности трудовых ресурсов;

для нахождения таких соотношений удельных весов возрастных групп (младенческой, детской, юношеской и старческой), с которыми можно сравнивать фактические, ибо возрастная структура населения определяет уровень целого ряда государственных расходов на социальное обеспечение, санитарные, культурно-просветительные мероприятия, меры здравоохранения и т. д.;

для определения участков (это могут быть и возрастные группы), на которых нужно сосредоточить все силы в борьбе за сохранение жизни;

для выявления факторов, влияющих на уровень смертности, сдвигов в изменении уровней смертности и причин снижения общего коэффициента смертности.

Таким образом, таблицы смертности являются тем чутким и точным инструментом, который улавливает изменение социально-экономических условий жизни населения и позволяет до повышения силы смертности говорить об отклонениях в условиях жизни той или иной совокупности.

При построении таблиц смертности исходят из гипотезы неподвижности (стационарности) населения. Предполагается, что в течение периода времени, равного целому поколению, ежегодно рождается одно и то же число детей при одинаковой для всего периода смертности. В этом случае исследователь как бы элиминирует возрастной состав сравниваемых масс и вычисляет, какой был бы показатель смертности, если бы возрастной состав данной массы определялся исключительно той смертностью, которая характерна для данного периода.

Для такого гипотетического населения определяются возрастной состав и общая численность, позволяющие найти такой коэффициент смертности, который обладает абсолютной сравнимостью, так как он зависит только от напряжения силы смертности. Обычно возрастной состав оказывает влияние на показатель смертности. В стационарном же, т. е. гипотетическом населении, все происходит наоборот: смертность формирует возрастной состав населения.

При оценке значения таблиц смертности для демографии следует исходить из того, что методы их построения имеют универсальный характер и могут быть применены и использованы не только для изучения смертности или доживаемости. Без преувеличения можно сказать, что по любому демографическому признаку можно построить таблицу, имеющую смысл таблицы смертности.

Возьмем брачность. Если в качестве причины прекращения брака рассматривать только смерть одного или обоих супругов, тогда, принимая за исходную совокупность число заключенных браков, можно найти число оставшихся налицо браков в конце первого, второго и т. д. года с их заключения. Вычисления таких таблиц продолжительности браков аналогичны расчетам обычной таблицы смертности.

При отсутствии переписей населения в России составлять таблицы смертности было трудно, тем не менее до проведения первой переписи населения таблицы смертности уже строились.

Первая таблица смертности в России была составлена в 1819 г. академиком К. Ф. Германом. В 1843 г. была составлена таблица смертности профессором Н. Е. Зерновым. К ней примыкает составленная в 1845 г. таблица смертности профессора В. К. Бруна. В 1854 г. была построена таблица смертности профессором М. Ф. Спасским.

Все эти таблицы строились методом смертных списков. Для страны с возрастающим населением этот метод является ошибочным, так как дает резкое преувеличение смертности. Смертность в России в те годы была действительно очень высокой. Ошибочный же метод построения таблиц смертности указанными авторами завысил показатели смертности до невероятных размеров.

Академик В. Я. Буняковский писал: «Исключительно неблагоприятное положение наше относительно законов смертности в сравнении с другими европейскими народами, до сих пор принимаемое за факт несомненный, по-моему мнению есть просто научное недоразумение, возникшее и державшееся единственно вследствие того, что занимавшиеся решением этого вопроса недостаточно вникали в его сущность»[1].

Сам же В. Я. Буняковский построил таблицы смертности для православного населения России за 1862 и 1863—1870 гг. Эти таблицы строились по новому методу, предложенному В. Я. Буняковским, состоящему в том, что повозрастные числа умерших относились к тем числам родившихся, к поколению которых принадлежат лица данного возраста.

В какой мере, по мнению В. Я. Буняковского, предыдущие таблицы смертности завышали уровень смертности и изображали смертность в России в очень неблагоприятном виде, можно судить по сравнительным данным о числе доживающих мужчин, взятым из двух таблиц смертности (см. табл. 32).

Следует признать, что несовершенство метода составления таблиц смертности В. Я. Буняковского, а главным образом, неточность дан-

[1] В. Я. Буняковский. Опыт о законах смертности в России и о распределении православного населения по возрастам. СПб., 1865, стр. IV.

Таблица 32

Сопоставление чисел доживающих мужчин

Возраст	0	10	20	30	40	50
Число доживающих по Н. Е. Зернову*	1 000	406	357	306	255	199
Число доживающих по В. Я. Буняковскому**	1 000	556	519	472	414	335

* В. Я. Буняковский. Опыт о законах смертности в России и о распределении православного населения по возрастам. СПб., 1865, стр. 13.

** Там же, стр. 71.

ных, которыми он располагал, несколько занижали уровень смертности.

Внимательное рассмотрение таблиц смертности, построенных В. Я. Буняковским для 1862 и 1870 гг., позволяет утверждать, что они отличаются приемами построения. Для перехода от совокупностей умерших III рода к совокупностям умерших I рода В. Я. Буняковский при построении таблицы смертности 1862 г. относил одногодичные группы умерших к средним арифметическим величинам из чисел родившихся в соответствующие 2 смежных календарных года. Например, умершие в 1862 г. в возрасте до 1 года относились к средней арифметической из родившихся в 1861 и 1862 гг. и т. д.

В таблицах же смертности 1870 г. одногодичные группы умерших относились к числам родившихся в одном соответствующем календарном году. Например, умершие в 1870 г. в возрасте до 1 года относились к числу родившихся лишь в том же 1870 г. Так как в России в тот период наблюдался непрерывный рост рождаемости[1], смертность 1870 г. должна была оказаться ниже, ибо в знаменателе рассматриваемой дроби в 1870 г. стояла бо́льшая величина, чем в 1862 г.

Отличие таблиц смертности 1862 и 1870 гг. привело к тому, что некоторые исследователи стали оспаривать факт чрезвычайно высокой смертности в России и даже делали вывод о происходящем в России процессе снижения смертности в период с 1862 по 1870 г. и улучшении условий жизни.

В. И. Гребенщиков вполне правильно понял разницу двух методов, примененных Буняковским, при составлении таблиц смертности для врачей на основании карточной регистрации, устанавливающей на 1 января в течение 1890—1896 гг. число всех врачей (мужчин) в России. Данная картотека давала распределение врачей по одногодичным возрастам с указанием их численности в каждом одногодичном возрасте и число умерших в течение календарного года.

[1] Здесь имеется в виду рост показателя рождаемости, который в значительной мере определялся улучшением регистрации рождений и снижением ранней детской смертности, так как регистрировались не рождения, а крещения.

Последователями В. Я. Буняковского являются К. А. Андреев (1871 г.), В. И. Борткевич (1890—1891 гг.), Л. Бессер и К. Баллод (1897 г.). Метод В. Я. Буняковского применил немецкий статистик Кнапп при построении таблицы смертности Ангальтского герцогства. В. И. Борткевич же ошибочно назвал этот метод ангальтским.

После переписи населения 1897 г. С. А. Новосельским была построена полная таблица смертности, охватывающая население 50 губерний европейской части России. По статистическим материалам Петербурга за 1910—1911 гг. также была построена таблица смертности.

Наибольшее научное и практическое значение имеют таблицы смертности, построенные ЦСУ СССР за 1926—1927 и 1958—1959 гг.

Имеется ли разница между таблицами смертности, построенными для сверстников и современников?

Обозначим число остающихся в живых в различных возрастах из рожденных в одном году $R_{1(x)}$ и аналогично остающихся в живых из рожденных в предшествовавшие годы $R_{2(x)}$, $R_{3(x)}$ и т. д., а число рождений в соответствующие годы N_1, N_2 и т. д. Выведем показатели доживаемости таблицы смертности для сверстников:

$$Q_0=\frac{R_{1\,(1)}}{N_1};\quad Q_1=\frac{R_{1(2)}}{N_1};\quad Q_2=\frac{R_{1\,(3)}}{N_1};\ \ldots$$

для современников:

$$Q_0'=\frac{R_{1\,(1)}}{N_1};\quad Q_1'=\frac{R_{2\,(2)}}{N_2};\quad Q_2'=\frac{R_{3\,(3)}}{N_3};\ \ldots$$

Для совпадения таблиц смертности нужно, чтобы $Q_x=Q_x'$, т. е. $\frac{R_{1\,(x)}}{N_1}=\frac{R_{n\,(x)}}{N_n}$.

Значит, требование состоит в том, что отношение оставшихся в живых в известном возрасте к числу рожденных в это время должно быть величиной постоянной. Однако в действительности этого нет и разница между показателями имеется.

СРАВНЕНИЕ ПОКАЗАТЕЛЕЙ ТАБЛИЦ СМЕРТНОСТИ

1896—1897 гг., 1926—1927 гг., 1958—1959 гг.

Сравнивать величины, характеризующие смертность населения, таблиц смертности, составленных или в различное время, или в одно и то же время, но в различных социально-экономических условиях, очень трудно. Во-первых, необходимо соблюдение единства территорий, которые охватываются таблицами, во-вторых — единство методов составления таблиц.

При сравнении показателей таблиц смертности 1896—1897 гг., 1926—1927 гг., 1958—1959 гг. ни одно из указанных требований не соблюдено. Поэтому такое сравнение следует считать во многих отношениях условным.

Сопоставление чисел живущих (L_x)

Начнем с сопоставления L_x — величины, дающей гипотетическое распределение наличного населения по одногодичным интервалам. L_x характеризует такое стационарное население, какое было бы, если бы в течение года рождалось столько же, сколько вымирало, а само вымирание происходило в полном соответствии с законом, описанным данной таблицей смертности.

В таблицах смертности 1896—1897 гг., составленных С. А. Новосельским, L_x не дано. Поэтому нами проведены расчеты по формуле $L_x = l_x - \frac{b}{2} - \frac{a-c}{16}$. Сопоставление величин L_x приведено в табл. 33.

Таблица 33

Возрастной состав стационарного населения (L_x)

Возраст	Мужчины 1896—1897 гг.	Мужчины 1926—1927 гг.	Мужчины 1958—1959 гг.	Женщины 1896—1897 гг.	Женщины 1926—1927 гг.	Женщины 1958—1959 гг.
0	837	860	970	857	883	976
10	520	671	934	555	704	943
20	492	648	921	524	681	935
30	457	610	897	485	644	923
40	416	565	861	439	603	904
50	358	497	802	383	555	874
60	277	397	687	300	485	814
70	168	260	498	180	358	683
80	66	109	254	73	177	418
90	16	20	61	18	39	129
100	2	1	5	2	3	17

Из этой таблицы вытекает, что во всех возрастах численность живущего женского населения больше мужского.

Укрупнение возрастных интервалов позволяет выявить динамику удельных весов наличного населения, разбитого на три группы[1]: до 20 лет, 20—60 лет, 60 лет и старше (см. табл. 34).

Таблица помогает установить:

1) снижение удельного веса детского и юношеского возрастов;

2) незначительное повышение удельного веса рабочих возрастов 20—60 лет;

4) непрерывное повышение удельного веса старческих возрастов.

Таким образом, можно сделать вывод о «постарении»[2] стационарного населения.

[1] Точный расчет ввиду отсутствия L_x в таблице С. А. Новосельского потребовал бы много времени и труда. Поэтому мы произвели приближенный расчет, позволивший, однако, выявить основные закономерности процесса.

[2] Математическая характеристика процесса постарения реального, а не стационарного населения рассматривается нами далее.

Таблица 34

Удельные веса различных возрастных групп в стационарном населении

Возрастные группы	Удельные веса в % 1896—1897 гг.	1926—1927 гг.	1958—1959 гг.
0; 10	37,8	31,3	26,4
20; 30; 40; 50; 60	56,4	57,0	59,4
70; 80; 90; 100	5,8	11,7	14,2

Сопоставление чисел доживающих (l_x)

Сопоставим числа доживающих (l_x) по данным разных лет (табл. 35).

Из таблицы видно, что во всех возрастах доживаемость женского населения выше, чем мужского.

Таблица 35

Числа доживающих (l_x)

Возраст	Мужчины 1896—1897 гг.	Мужчины 1926—1927 гг.	Мужчины 1958—1959 гг.	Женщины 1896—1897 гг.	Женщины 1926—1927 гг.	Женщины 1958—1959 гг.
0	1 000	1 000	1 000	1 000	1 000	1 000
10	521	673	934	557	705	944
20	493	650	922	526	683	936
30	459	612	898	487	646	923
40	418	567	863	441	605	905
50	362	502	806	387	558	876
60	282	403	694	305	489	818
70	173	268	509	186	366	692
80	70	116	266	77	186	435
90	18	22	67	19	43	140
100	3	1	5	4	4	19

Так как числа доживающих указывают на порядок вымирания населения, можно заключить, что в пределах одного периода скорость вымирания мужчин и женщин при переходе от одного возраста к другому не очень отличается; с течением времени скорость вымирания замедляется[1], наименьшая скорость вымирания у мужчин и женщин по всем таблицам наблюдается в возрастах 20—40, 20—50 лет.

[1] Для измерения интенсивности процесса замедления скорости вымирания населения можно применить выравнивание данных по прямой линии или параболе второго порядка и по параметру a_1 судить о том, насколько замедлилась скорость.

Сопоставление чисел умерших (d_x)

Сопоставление чисел умерших по данным разных лет приведено в табл. 36.

Таблица 36

Числа умирающих из 1000 родившихся (d_x)

Возраст	Мужчины			Женщины		
	1896—1897 гг.	1926—1927 гг.	1958—1959 гг.	1896—1897 гг.	1926—1927 гг.	1958—1959 гг.
0	298	201	44	259	172	37
10	4	2	1	4	2	1
20	3	3	2	4	3	1
30	4	4	3	4	4	1
40	5	5	4	5	4	2
50	7	8	8	6	5	4
60	9	11	15	10	9	8
70	12	16	22	12	17	19
80	8	13	25	9	17	33
90	3	4	12	3	8	20
100	0	0	1	0	1	4

Из данных таблицы видно, что: 1) в самом молодом возрасте число умирающих женщин меньше, чем мужчин (этот факт общеизвестен и причины его следует, очевидно, искать в биологии); 2) в десятилетнем возрасте наблюдается равновесие; 3) с 20 лет и старше опять превышение мужских смертей над женскими; 4) начиная с 70 лет и старше числа умирающих женщин больше, чем мужчин[1].

По таблице можно также проследить изменение чисел умирающих в динамике от 1896—1897 гг. до 1958—1959 гг.: 1) у мужчин до 40 лет, а у женщин до 60 лет наблюдается тенденция к снижению чисел умирающих, причем особенно резкое снижение в нулевом возрасте; 2) начиная с 50 лет у мужчин и с 70 лет у женщин число умирающих в последующих таблицах возрастает; 3) максимальное число[2] умирающих по таблице С. А. Новосельского и таблице 1926—1927 гг. приходится для мужчин и женщин на возраст 70 лет, а по таблице смертности 1958—1959 гг. — на возраст 80 лет. Возраст, в котором происходит наибольшее вымирание, за годы, отделяющие составление таблиц, повысился.

[1] Это превышение может быть объяснено тем, что до этого возраста мужчины вымирали интенсивнее и в предельных группах контингент мужчин поредел, вследствие чего умирающие женщины встречаются чаще, чем мужчины.

[2] Не учитывая, конечно, числа умирающих в нулевом возрасте.

Сопоставление вероятности смерти (q_x)

Рассмотрим данные о вероятности смерти (табл. 37).

Таблица 37

Вероятности умереть (q_x)

Возраст	Мужчины			Женщины		
	1896—1897 гг.	1926—1927 гг.	1958—1959 гг.	1896—1897 гг.	1926—1927 гг.	1958—1959 гг.
0	0,298	0,201	0,044	0,259	0,172	0,037
10	0,007	0,003	0,001	0,006	0,003	0,001
20	0,007	0,005	0,002	0,007	0,005	0,001
30	0,008	0,006	0,003	0,009	0,006	0,002
40	0,011	0,009	0,005	0,011	0,007	0,003
50	0,019	0,017	0,010	0,017	0,009	0,005
60	0,033	0,028	0,022	0,033	0,018	0,010
70	0,068	0,060	0,044	0,067	0,046	0,028
80	0,112	0,113	0,092	0,112	0,090	0,075
90	0,145	0,192	0,177	0,146	0,177	0,146
100	0,163	0,309	0,281	0,154	0,298	0,222

На основании этой таблицы можно сделать следующие выводы:

1) вероятность смерти женского населения по всем таблицам смертности меньше вероятности смерти мужского населения (исключение составляет возраст 30 лет; по таблице Новосельского в этом возрасте более высокая вероятность умереть у женщин объясняется очень высокой материнской смертностью);

2) динамика повозрастной смертности от 1896—1897 гг. до 1958—1959 гг. показывает, что вероятность умереть снижается (исключение составляют только мужчины в возрасте свыше 80 лет, у которых q_x в 1958—1959 гг. выше, чем в 1896—1897 гг., что, по-видимому, объясняется разными методами экстраполяции). Вероятность смерти для новорожденных снизилась в 1958—1959 гг. в 4,6 раза по сравнению с 1926—1927 гг. и в 6,8 раза по сравнению с 1896—1897 гг. у мальчиков и соответственно в 4,7 и 7,0 раза у девочек. Это свидетельствует о глубоких сдвигах в социально-экономических условиях жизни населения;

3) максимальная величина вероятности смерти во всех таблицах у мужчин и женщин относится к нулевому возрасту и старческим возрастам;

4) вероятность смерти у женщин снизилась в большей степени, чем у мужчин.

Сопоставление средней продолжительности жизни (e_x^0)

Сопоставим, наконец, среднюю продолжительность предстоящей жизни (табл. 38).

Таблица 38

Средняя продолжительность предстоящей жизни (e_x^0)

Возраст	Мужчины			Женщины		
	1896—1897 гг.	1926—1927 гг.	1958—1959 гг.	1896—1897 гг.	1926—1927 гг.	1958—1959 гг.
0	31	41,9	64,4	33	46,8	71,7
10	49	51,6	58,8	49	55,7	65,9
20	41	43,2	49,5	41	47,4	56,4
30	34	35,6	40,7	34	39,7	47,1
40	27	28,0	32,2	27	32,1	37,9
50	20	21,0	24,0	20	24,4	29,0
60	14	14,8	17,0	14	17,1	20,6
70	10	9,6	11,3	10	11,0	13,3
80	7	6,0	6,9	7	6,8	8,0
90	6	4,0	4,3	6	4,4	5,1
100	5	2,4	2,7	5	2,6	3,6

Из таблицы следует:

1) средняя продолжительность жизни женского населения во всех таблицах и во всех возрастах больше, чем у мужчин;

2) рассмотрение средней продолжительности жизни в динамике показывает, что за рассматриваемый период наивысшая продолжительность жизни была в 1958—1959 гг. Особенно возросла средняя продолжительность жизни новорожденных. Она составляла по сравнению с 1926—1927 гг. 154% для мужчин и 159% для женщин, а по сравнению с 1896—1897 гг. — 208% для мужчин и 218% для женщин.

Все вышеизложенное позволяет сделать интересные выводы, подтвержденные взаимно согласующимися фактами.

Вывод. 1. Смертность женского населения во всех таблицах меньше, чем мужского. Этот вывод подтверждается:

1) большей численностью женского населения в стационарном населении;

2) большей доживаемостью женского населения в любых возрастах;

3) меньшим числом умирающих женщин;

4) меньшей вероятностью умереть (большей вероятностью дожить) у женщин, чем у мужчин, в любом возрасте;

5) большей средней продолжительностью жизни у женского населения, чем у мужского.

Очевидно, объяснением может служить биологически обусловленная большая жизненность женского организма по сравнению с мужским, а также повышенный травматизм и токсикомания у мужчин.

Вывод 2. Понижение смертности в последних таблицах по сравнению с таблицами 1896—1897 гг. во всех возрастах, а в особенности в детском (до 1 года). Это положение подтверждают:

1) возрастающая во времени средняя продолжительность жизни у мужчин и женщин;

2) тот факт, что в 1967 г. детская смертность в СССР снизилась по сравнению с 1913 г. , с 273 до 26‰ т. е., более чем в 10 раз[1].

Здесь биологический фактор ни в коем случае не может служить объяснением. Снижение детской смертности является результатом изменения социально-экономических условий, улучшением медицинского и санитарного обслуживания и громаднейшим комплексом целенаправленных факторов.

КРАТКИЕ ТАБЛИЦЫ СМЕРТНОСТИ

Выбор метода построения таблиц смертности

Выбор метода построения таблиц смертности зависит от конкретных особенностей статистических данных, которыми располагает исследователь.

Если, используя первичный материал, производить вычисления показателей таблиц смертности не по одногодичным интервалам, а по укрупненным пятилетним или десятилетним, то полученные таблицы смертности называются сокращенными или краткими таблицами в отличие от полных таблиц смертности.

Ряд интересных методов построения таблиц смертности рассмотрен А. Я. Боярским в «Курсе демографии»[2]. Не ограничиваясь ими, рассмотрим и другие методы построения кратких таблиц смертности и укажем практические выводы, полученные при их реализации.

Начиная с 4 лет число сверстников в возрастных группах можно найти с достаточной приближенностью как среднюю арифметическую из доживающих до двух смежных возрастов.

Значит, группа в возрасте от x до $x+1$ равна $\frac{l_x+l_{x+1}}{2}$.

Если m — число умерших в возрасте от x до $x+1$, то $l_x-l_{x+1}=m$ и $l_{x+1}=l_x-m$.

Заменяя, получим: $\frac{l_x+l_x-m}{2}=l_x-\frac{m}{2}$.

Отношение m к полученной величине называется коэффициентом смертности (K_x в возрасте от x до $x+1$):

$$K_x=\frac{m}{l_x-\frac{m}{2}}=\frac{2m}{2l_x-m}.$$

Откуда

$$m=\frac{K_x\,l_x}{1+\frac{K_x}{2}}.$$

[1] «Народное хозяйство СССР в 1968 г.». М., «Статистика», 1969, стр. 36.

[2] См. А. Я. Боярский и др. Курс демографии. М., «Статистика», 1967.

По этой формуле можно исчислять числа умерших для последовательных возрастов.

Для упрощения произведем в формуле дальнейшие преобразования:

$$m = l_x - l_{x+1} = \frac{K_x l_x}{1+\frac{K_x}{2}}.$$

Откуда

$$l_{x+1} = l_x - \frac{K_x l_x}{1+\frac{K_x}{2}} = l_x\left(1 - \frac{K_x}{1+\frac{K_x}{2}}\right) =$$

$$= l_x\left(\frac{1+\frac{K_x}{2} - K_x}{1+\frac{K_x}{2}}\right) = l_x\left(\frac{1-\frac{K_x}{2}}{1+\frac{K_x}{2}}\right) = l_x\left(\frac{2-K_x}{2+K_x}\right).$$

Если мы хотим вычислить показатель смертности не одногодичной, а, например, пятилетней группы, то вместо l_{x+1} нужно взять l_{x+5} и соответственно этому K_x умножить на 5:

$$l_{x+5} = l_x\left(\frac{2-5K_x}{2+5K_x}\right) \quad \text{или} \quad l_{x+5} = l_x\left(\frac{1-\frac{5K_x}{2}}{1+\frac{5K_x}{2}}\right).$$

Выведенная формула используется при построении кратких таблиц смертности. Впервые она была получена Вильямом Фарром. Преобразуем ее с помощью тригонометрической подстановки в такое выражение, которое легко логарифмируется. Известно, что

$$\cos\alpha = \frac{1-\operatorname{tg}^2\frac{\alpha}{2}}{1+\operatorname{tg}^2\frac{\alpha}{2}}.$$

Принимая $\frac{5K_x}{2} = \operatorname{tg}^2\frac{\alpha}{2}$, получаем: $\operatorname{tg}\frac{\alpha}{2} = \sqrt{\frac{5K_x}{2}}$.

Находя по соответствующим таблицам угол α, следует удвоить эту величину и отыскать в таблице соответствующий этому удвоенному углу косинус, который можно подставить вместо выражения

$$\frac{1-\frac{5K_x}{2}}{1+\frac{5K_x}{2}}.$$

Рассмотрим пример. Требуется определить число мужчин, переживающих 10-й год жизни. Исчислим коэффициент смертности возрастной группы 5—10 лет. Допустим, что число умерших в течение 4 лет и 5 месяцев на 5—10 году жизни равно 49 496, среднее население — 1 708 523; число переживающих 5-й год (l_5) равно 729,86 из 1000 родившихся. Тогда годовой коэффициент смертности возрастной группы 5—10 лет составит:

$$K_{5-10} = \frac{49\,496}{4\frac{5}{12}\cdot 1\,708\,523} \approx 0{,}006559, \quad \text{или} \quad 6{,}559\ ^0/_{00};$$

$$l_{10} = l_{5-10}\left(\frac{1-\frac{5K_{5-10}}{2}}{1+\frac{5K_{5-10}}{2}}\right) = 729{,}86\cdot\frac{1-\frac{5}{2}\cdot 0{,}006559}{1+\frac{5}{2}\cdot 0{,}006559}.$$

Согласно принятому нами условию $\frac{5}{2}K_{5-10} = \operatorname{tg}^2\frac{\alpha}{2}$; $\frac{5}{2}\cdot 0{,}006559 = \operatorname{tg}^2\frac{\alpha}{2}$. Прологарифмируем: $\lg\operatorname{tg}^2\frac{\alpha}{2} = \lg\frac{5}{2} + \lg 0{,}006559 = 0{,}39794 + \bar{3}{,}81684 = \bar{2}{,}21478$. Отсюда $\lg\operatorname{tg}\frac{\alpha}{2} = \bar{1}{,}10739$. Для использования таблиц вычтем 10: $\lg\operatorname{tg}\frac{\alpha}{2} = 9{,}10739 - 10$. Находим $\frac{\alpha}{2} = 7°18'$ и $\alpha = 14°36'$, $\lg\cos\alpha = \lg\cos 14°36' = 9{,}98574 - 10 = \bar{1}{,}98574$. Теперь $\lg l_{10} = \lg 729{,}86 + \lg\cos\alpha = 2{,}86324 + \bar{1}{,}98574 = 2{,}84898$.

Откуда $l_{10} = 706{,}3$ и $m_{5-10} = l_5 - l_{10} = 729{,}86 - 706{,}3 = 23{,}56$.

Этот же ответ можно было получить, конечно, и без логарифмирования.

$$l_{10} = 729{,}86\left(\frac{1-\frac{5}{2}\cdot 0{,}006559}{1+\frac{5}{2}\cdot 0{,}006559}\right) = 729{,}86\left(\frac{1-0{,}0163975}{1+0{,}0163975}\right) =$$

$$= 729{,}86\cdot\frac{0{,}9836025}{1{,}0163975} = 729{,}86\cdot 0{,}967773 \approx 706{,}3.$$

Используя формулу Фарра, получаем среднюю вероятность дожития:

$$\frac{l_{x+5}}{l_x} = \frac{2-5K_x}{2+5K_x} = \bar{p}_x^5.$$

Откуда

$$l_{x+5} = l_x\bar{p}_x^5 \quad \text{и} \quad l_{x+10} = l_{x+5}\cdot\bar{p}_{x+5}^5. \text{ и т. д.}$$

Методы, ориентированные на предварительную математико-статистическую обработку данных переписи

Некоторые исследователи считают, что если для построения полной таблицы смертности переписные данные о численности населения, носящие на себе следы аккумуляции возрастов, нельзя использовать без надлежащей математико-статистической обработки, то для построения кратких таблиц смертности эти переписные данные, объединяемые в возрастные пятилетние или десятилетние интервалы, пригодны без выравнивания.

Академик М. В. Птуха проделал большую методологическую и вычислительную работу и по данным многих пробных вычислений доказал, что это мнение является несостоятельным.

В связи с этим приобретают значение те методы построения кратких таблиц смертности, которые ориентируются на предварительную математико-статистическую обработку фактических переписных данных.

Метод Кинга

Кингом[1] разработан способ, отличающийся от вышеизложенного получением средних коэффициентов K_x, где x — центр возрастного пятилетнего интервала.

Непременным условием применения метода Кинга является предварительное выравнивание численностей населения и умерших. Выравнивание производится по формуле

$$u_7 = 0{,}2w_5 - 0{,}008\,\Delta^2 w_0,$$

где w_0 — сумма живущих или умерших в данном пятилетнем интервале;

w_5 — сумма живущих или умерших в последующем пятилетнем интервале;

u_7 — искомое центральное число живущих или умерших.

При этом способе определяют величины вероятностей дожития, соответствующие центральным возрастам x:

$$p_x = \frac{2-K_x}{2+K_x}.$$

Для получения вероятностей дожития до конца пятилетнего интервала при этом способе предполагают, что промежуточные вероятности укладываются в параболу 3-го порядка и выводят формулу для 5-летней суммы логарифмов p_x центральных возрастов и их последовательных конечных разностей. После этого можно вывести вероятность дожития до конца пятилетнего периода ${}_5p_x$, т. е. рассчитать l_x через пятилетние интервалы и, интерполируя, получить Σl_x и из них сокращенную, а затем полную среднюю продолжительность предстоящей жизни.

При этом способе таблицы смертности могут начинаться лишь с одиннадцати лет. Зато значения их показателей почти совпадают с показателями полных таблиц.

Метод Сноу

Идея метода Сноу при построении кратких таблиц смертности состоит в признании функциональной связи между одногодичными коэффициентами смертности и вероятностями дожития и в установлении корреляционной связи между этими показателями, относящимися к пятилетним интервалам.

При этом применение уравнения $p_{x/x+5} = \dfrac{2-5K_{x/x+5}}{2+5K_{x/x+5}}$ для всех пятилетних групп не дает достаточного приближения. Поэтому, предположив, что в разных пятилетних интервалах корреляционная связь между вероятностями дожития и коэффициентами смертности различна, ее устанавливают эмпирически. Пользуясь методом наименьших квадратов и основываясь на нескольких английских таблицах смертности, Сноу, привлекая линейную и параболическую функции, находит корреляционные уравнения, связывающие коэффициенты смертности с вероятностями дожития.

Краткие таблицы смертности, построенные по этому способу, так же очень близки к полным таблицам смертности.

Метод Броунли

Метод Броунли примыкает к методу Сноу. Специфичность его состоит в том, что корреляционная связь устанавливается между табличными коэффициентами смертности и стандартизованными коэффицентами. Любопытно, что в качестве стандарта предлагается убывающий натуральный ряд чисел. Так, для возраста 0—5 лет последовательные численности живущих в стандарте равны 16, 15, 14, 13, 12. Линейные уравнения корреляционной связи устанавливаются автором на основании 10 английских таблиц смертности. Построение краткой таблицы смертности прозводится следующим образом.

Имея ряд пятилетних коэффициентов смертности и принимая за стандарт убывающий ряд чисел, находят ожидаемое число смертей в каждом интервале. Путем суммирования определяют ожидаемое число смертных случаев в возрастах выше возраста x лет. С помощью линейных корреляционных формул можно получить табличные коэффициенты смертности для стационарного населения во всех возрастах выше определяемого. Обратные величины этих коэффициентов дают среднюю продолжительность предстоящей жизни (e_x^0).

[1] «Supplement to the 75 Annual Report of the Registrar — General of births, deaths and marriages in England and Wales, p. I, Life — Tables, London, 1914.

Далее автор принимает $\overset{\circ}{e}_x = f(x)$; тогда $T_x = e^{K - \int \frac{dx}{f(x)}}$. Выбирая три значения $\overset{\circ}{e}_x$ и выравнивая $\overset{\circ}{e}_x$, находят значения параметров a, c, n из следующей формулы $\overset{\circ}{e}_x = \frac{(c-x)^n}{a}$.

Произведя подстановку в предыдущей формуле, получаем:

$$T_x = e^{K - \frac{a}{(n-1)(c-x)^{n-1}}}.$$

Разделив T_x на e_x^0, находят l_x. Далее расчет ведется обычным порядком.

Метод М. В. Птухи

Академик М. В. Птуха предложил в кратких таблицах смертности до возраста 5 лет сохранить тот же порядок построения показателей, как и для полных таблиц, т. е. находить l_0, l_1, l_2, l_3, l_4, l_5.

Для старших возрастов вычисляют коэффициенты смертности $m_{5/10}$, $m_{10/15}$ и т. д. путем деления числа умерших за два года в возрасте 5—10, 10—15 лет и т. д. на удвоенное число живущих в этих возрастах.

Вероятности дожития $p_{5/10}$, $p_{10/15}$ и т. д. определяют по формуле Фарра $p_x = \frac{2-5K_x}{2+5K_x}$ и используют для получения чисел доживающих

$$l_{10} = l_5 \, p_{5/10}; \quad l_{15} = l_{10} \, p_{10/15} \text{ и т. д.}$$

Полученные числа доживающих автор использовал для определения среднегодовых вероятностей дожития промежуточных возрастов путем деления последующих l_x на предыдущие с извлечением корня соответствующей степени (5-й или 10-й). Например:

$$\bar{l}_x = \sqrt[5]{\frac{l_{15}}{l_{10}}} \quad \text{или} \quad \bar{l}_x = \sqrt[10]{\frac{l_{15}}{l_5}}.$$

Время, прожитое населением в возрасте от 95 до 104 лет, М. В. Птуха исчислял по формуле средней хронологической для моментного ряда, т. е.

$$\frac{1}{2} l_{95} + l_{96} + l_{97} + \ldots + l_{104} + \frac{1}{2} l_{105}.$$

Кроме указанных особенностей, метод М. В. Птухи имеет и другие менее существенные.

Метод В. В. Паевского

Предположенный В. В. Паевским метод построения кратких таблиц смертности основан на вычислении вероятности дожития по формуле

$$p_{x/x+5} = e^{-5K_x}.$$

Можно доказать, что в порядке первого приближения величина $p_{x/x+5}$ в формуле В. В. Паевского равна значению, полученному из формулы Фарра, применявшейся при построении кратких таблиц смертности:

$$p_{x/x+5} = \frac{2-5K_x}{2+5K_x}.$$

Для этого разложим выражение e^{-5K_x} в маклореновский ряд:

$$p_{x/x+5} = 1 - 5K_x + \frac{5^2 K_x^2}{2!} - \frac{5^3 K_x^3}{3!} + \ldots =$$

$$= 1 - 5K_x \left(1 - \frac{5K_x}{2} + \frac{5^2 K_x^2}{6} - \ldots\right).$$

Если скобку заменить бесконечно убывающей прогрессией со знаменателем $q = -\frac{5K_x}{2}$, то по формуле прогрессии $S = \frac{a_1}{1-q}$ получим:

$$1 - \frac{5K_x}{2} + \frac{5^2 K_x^2}{4} - \ldots = \frac{1}{1 + \frac{5K_x}{2}}.$$

Следовательно $p_{x/x+5} = 1 - 5K_x \left(\frac{1}{1+\frac{5K_x}{2}}\right) = 1 - \frac{10K_x}{2+5K_x} = \frac{2-5K_x}{2+5K_x}$,

т. е. тот же результат.

Расхождение между расчетами по методу В. В. Паевского и методу Фарра не превышает $\frac{125K_x^2}{12}$. Практически получаются очень близкие результаты, но метод В. В. Паевского более прост. А мы как раз и нуждаемся в таком методе построения хотя бы кратких таблиц смертности, который был бы правилен, достаточно точен и позволял быстро находить соответствующие показатели таблиц смертности. Метод построения кратких таблиц смертности В. В. Паевского допускает ошибку в средней продолжительности предстоящей жизни в 0,5 года, т. е. дает вполне достаточную точность для практического использования. Результаты, полученные при помощи формул

$$q_{x/x+5} = \frac{2 \cdot 5m_{x/x+5}}{2+5m_{x/x+5}} = \frac{10m_{x/x+5}}{2+5m_{x/x+5}};$$

$$p_{x/x+5} = 1 - q_{x/x+5}$$

вполне удовлетворяли практическим нуждам и могли быть использованы для различных целей, в том числе для интерполяции промежуточных уровней.

Отдел статистики населения ЦСУ СССР ежегодно методом В. В. Паевского строит краткие таблицы смертности СССР по пятилетним возрастным группам.

5 Зак. 1603

При использовании способа Броунли краткая таблица смертности строится на основании предварительно найденной средней продолжительности жизни, затем находят T_x, а далее l_x.

При обычных же способах для получения средней продолжительности жизни в условиях стационарного населения по кратким таблицам смертности нужно найти число лет, которое проживут все лица l_x до полного вымирания.

Числа T_x можно получить суммированием $L_{x/x+5}$ в обратном порядке, т. е. начиная со старших возрастов. При пятилетних интервалах

$$L_{x/x+5}=\frac{5(l_x+l_{x+5})}{2}=2{,}5(l_x+l_{x+5}).$$

Для детских возрастов применяются другие формулы:

$$L_{1/4}=2(l_1+l_4) \quad \text{и} \quad L_{0/1}=0{,}33l_0+0{,}67l_1.$$

Для десятилетних интервалов $L_{x/x+10}$ можно найти, пользуясь соответствующей формулой.

Получив числа $L_{x/x+10}$, суммированием находят T_x. При этом исходят из предположения, что на всем пятилетнем интервале фактический коэффициент смертности (K) приближенно равен коэффициенту смертности стационарного населения (K'):

$$K_{i/i+5}=K'_{i/i+5}.$$

Так как $K_{i/i+5}=\dfrac{l_i-l_{i+5}}{T_i-T_{i+5}}$, то

$$T_i-T_{i+5}=\frac{l_i-l_{i+5}}{K'_{i/i+5}}=\frac{d_{i/i+5}}{K'_{i/i+5}}.$$

Следовательно, пятилетняя сумма L_i получается делением числа умерших за 5 лет на пятилетний возрастной коэффициент. Разделив T_i на l_i, получаем e_i^0.

Для решения многих вопросов, связанных с воспроизводством всего населения и различных его контингентов, в частности для расчета баланса рабочей силы, ведения страховых дел, перспективных расчетов, используются полные таблицы смертности. Могут ли для этих целей применяться краткие таблицы с интервалами 5 или 10 лет?

Как мы увидим в главе III, для устранения аккумуляции при построении полных таблиц смертности применяются довольно сложные (длительные и кропотливые) методы выравнивания, составляющие при этом основную массу работы. Преимущество кратких таблиц в том, что, отказываясь от одногодичных групп, можно подобрать такие укрупненные возрастные интервалы, которые сами по себе уменьшали и сглаживали неточности показания одногодичных возрастов. Однако трудно предугадать все те разнообразные возрастные пределы, которые могут понадобиться для решения различных задач. С этой точки зрения полные таблицы более универсальны.

Метод В. В. Паевского получения одногодичных вероятностей смерти

Может возникнуть любопытная ситуация, при которой нужны полные таблицы, но одногодичные данные об умерших отсутствуют, а имеются более или менее достоверные данные по пятилетним интервалам.

В. В. Паевский теоретически разработал метод получения одногодичных вероятностей смерти на основе фактических данных, сведенных в пятилетние интервалы, и практически его применил для построения таблиц смертности отдельных социальных групп населения по данным переписи населения 1926 г.

Допустим, что мы имеем данные по трем пятилетним интервалам о сумме умерших за каждые пять лет и пятилетние суммы живущих. Найдем средние пятилетние коэффициенты смертности:

$$m_1^{\text{ф}}=\frac{\Sigma_1 M}{\Sigma_1 l_x}; \quad m_2^{\text{ф}}=\frac{\Sigma_2 M}{\Sigma_2 l_x}; \quad m_3^{\text{ф}}=\frac{\Sigma_3 M}{\Sigma_3 l_x},$$

где $\Sigma_1 M$, $\Sigma_2 M$, $\Sigma_3 M$ — соответственно суммы умерших за первое, второе и третье пятилетия,
а $\Sigma_1 l_x$, $\Sigma_2 l_x$, $\Sigma_3 l_x$ — числа доживающих.

Перенеся начало координат в середину среднего пятилетия, получим средние пятилетние выровненные коэффициенты:

$$m_1=\frac{l_{-7,5}-l_{-2,5}}{\int\limits_{-7,5}^{-2,5} l(x)\,dx};$$

$$m_2=\frac{l_{-2,5}-l_{+2,5}}{\int\limits_{-2,5}^{+2,5} l(x)\,dx};$$

$$m_3=\frac{l_{+2,5}-l_{+7,5}}{\int\limits_{+2,5}^{+7,5} l(x)\,dx}.$$

В этом случае

$$m_1^{\text{ф}}=m_1; \quad m_2^{\text{ф}}=m_2; \quad m_3^{\text{ф}}=m_3.$$

Выразим одногодичный центральный коэффициент смертности среднего пятилетия:

$$m_c=\frac{l_{-0,5}-l_{+0,5}}{\int\limits_{-0,5}^{+0,5} l(x)\,dx}.$$

В качестве интерполяционной формулы за все три пятилетних интервала, т. е. за 15 лет, примем параболу 3-го порядка, дающую также и перегибы. Однако по некоторым соображениям, которые выясняются при выводе, В. В. Паевский берет формулу

$$l_x = \left(\frac{64}{125}ax^3 + \frac{24}{25}bx^2 + \frac{8}{5}cx + 1\right) r.$$

Тогда

$$m_{\text{с}} = \frac{l_{-0,5} - l_{+0,5}}{\int\limits_{-0,5}^{+0,5} l(x)dx} = \frac{\left[\frac{64}{125}ax^3 + \frac{24}{25}bx^2 + \frac{8}{5}cx + 1\right]_{+0,5}^{-0,5} r}{\left[\frac{64}{125}ax^3 + \frac{24}{25}bx^2 + \frac{8}{5}cx + 1\right]_{-0,5}^{+0,5} r}.$$

После алгебраических преобразований и сокращений получаем

$$m_{\text{с}} = -\frac{3,2a + 40c}{2b + 25}.$$

Подставляя выражения $l_x = \left(\frac{64}{125}ax^3 + \frac{24}{25}bx^2 + \frac{8}{5}cx + 1\right) \cdot r$ в формулы трех значений $m_1^{\text{ф}}$, $m_2^{\text{ф}}$ и $m_3^{\text{ф}}$, как мы выяснили, равных m_1, m_2 и m_3, и, интегрируя, получаем систему трех уравнений с тремя неизвестными (a, b c).

$$\begin{cases} 5m_1^{\text{ф}} = \left(400m_1^{\text{ф}} - 208\right) a - \left(130m_1^{\text{ф}} - 48\right) b + \left(40m_1^{\text{ф}} - 8\right) c, \\ 5m_2^{\text{ф}} = -16a - 10m_2^{\text{ф}} b - 8c, \\ 5m_3^{\text{ф}} = \left(-400m_3^{\text{ф}} - 208\right) a + \left(-130m_3^{\text{ф}} - 48\right) b + \left(-40m_3^{\text{ф}} - 8\right) c. \end{cases}$$

Используя определители (детерминанты), получаем знаменатель

$$\Delta = \left(315m_3^{\text{ф}} - 50m_2^{\text{ф}} m_3^{\text{ф}} - 315m_1^{\text{ф}} - 650m_1^{\text{ф}} m_3^{\text{ф}} - 50m_1^{\text{ф}} m_2^{\text{ф}} + 144\right).$$

Тогда

$$a = \frac{\Delta_1}{\Delta} = \frac{15\left(15m_2^{\text{ф}} m_3^{\text{ф}} - 15m_1^{\text{ф}} m_2^{\text{ф}} - 50m_1^{\text{ф}} m_2^{\text{ф}} m_3^{\text{ф}} - 2m_3^{\text{ф}} + 4m_2^{\text{ф}} - 2m_1^{\text{ф}}\right)}{16\Delta};$$

$$b = \frac{\Delta_2}{\Delta} = \frac{5\left(10m_1^{\text{ф}} m_3^{\text{ф}} - 3m_3^{\text{ф}} + 10m_2^{\text{ф}} m_3^{\text{ф}} + 3m_1^{\text{ф}} + 10m_1^{\text{ф}} m_2^{\text{ф}}\right)}{2\Delta};$$

$$c = \frac{\Delta_3}{\Delta} = \frac{15\left(250m_1^{\text{ф}} m_2^{\text{ф}} m_3^{\text{ф}} + 115m_1^{\text{ф}} m_2^{\text{ф}} - 115m_2^{\text{ф}} m_3^{\text{ф}} - 52m_2^{\text{ф}} + 2m_1^{\text{ф}} + 2m_3^{\text{ф}}\right)}{8\Delta}.$$

Подставляя найденные значения a, b, c в формулу центрального коэффициента смертности среднего пятилетия, получаем:

$$m_{\text{с}} = \frac{m_2^{\text{ф}}\left[715\left(m_3^{\text{ф}} - m_1^{\text{ф}}\right) - 1550m_1^{\text{ф}} m_3^{\text{ф}} + 324\right] - 12\left(m_3^{\text{ф}} + m_1^{\text{ф}}\right)}{655\left(m_3^{\text{ф}} - m_1^{\text{ф}}\right) - 100\left(m_1^{\text{ф}} + m_3^{\text{ф}}\right) m_2^{\text{ф}} - 1350m_1^{\text{ф}} m_3^{\text{ф}} + 300}.$$

Теперь можно определить все коэффициенты смертности центральных возрастов каждого из трех пятилетий. После этого интерполя-

цией находим все промежуточные значения коэффициентов смертности[1].

Затем В. В. Паевский использует метод соприкасательных парабол, описанный далее.

Через известные три точки $(-1, y_{-1})$, $(0, y_0)$ и $(1, y_1)$ проводится первая вспомогательная парабола 2-го порядка, а через три точки $(0, y_0)$, $(1, y_1)$, $(2, y_2)$ проводится вторая вспомогательная парабола 2-го порядка. Получаем параболу 3-го порядка (рис. 22).

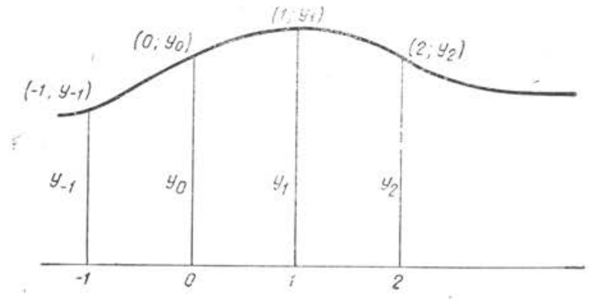

Рис. 22. Интерполяция промежуточных значений коэффициентов смертности.

Она должна проходить через точки $(0, y_0)$ и $(1, y_1)$ и в точке $(0, y_0)$ иметь общую касательную с первой вспомогательной кривой, а в точке $(1, y_1)$ — общую касательную со второй вспомогательной кривой. Если теперь обозначить приращение возраста S через h, коэффициент смертности центрального для первого пятилетия года y_{-1}, разности первого, второго и третьего порядков через Δy_{-1} $\Delta^2 y_{-1}$, $\Delta^3 y_{-1}$, то одногодичные разности возрастного интервала (0, 1) можно найти по следующим формулам:

$$\delta^3 m_0 = 0{,}024\Delta^3 y_{-1};$$

$$\delta^2 m_0 = 0{,}04\Delta^2 y_{-1} - 0{,}016\Delta^3 y_{-1};$$

$$\delta^1 m_0 = 0{,}2\Delta y_{-1} + 0{,}12\Delta^2 y_{-1} - 0{,}016\Delta^3 y_{-1}.$$

Если эти разности найдены, то коэффициенты смертности получаются суммированием.

Для примера расчетов возьмем данные 1926—1927 гг. по Ленинграду для рабочих-мужчин[2].

---

[1] См. С. А. Новосельский, В. В. Паевский. Таблицы смертности населения Ленинграда за 1910—1911, 1920 и 1923 гг. «Материалы по статистике Ленинграда и Ленинградской губернии». Вып. 6, Л., 1925.

[2] «Смертность и продолжительность жизни населения СССР 1926—1927. Таблицы смертности». М.—Л., Планхозгиз, стр. XXXIII.

Таблица 39

Коэффициенты смертности для центральных возрастов

Возрастные интервалы	Непосредственные средние коэффициенты смертности $m_i^{ф}$	Центральные возрасты	Коэффициенты смертности центральных возрастов* m_c
15—19	0,00413	17	0,00406
20—24	0,00598	22	0,00609
25—29	0,00521	27	0,00513
30—34	0,00650	32	0,00649

* Рассчитаны по рассмотренной выше формуле.

Для облегчения последующих расчетов будем интерполировать не сами коэффициенты смертности m_x, а логарифмы этих коэффициентов (табл. 40):

$$\lg(m_x+0{,}1)=u_x.$$

Таблица 40

Интерполяция логарифмов коэффициентов смертности

x	m_x	$m_x+0{,}1$	$\lg(m_x+0{,}1)$	Δu_0	$\Delta^2 u_0$	$\Delta^3 u_0$
17	0,00406	0,10406	1,01778	789	—1 183	2 135
22	0,00609	0,10609	1,02567	—394	952	—870
27	0,00513	0,10513	1,02173	558	82	—
32	0,00649	0,10649	1,02731	640	—	—
37	0,00807	0,10807	1,03371	—	—	—

Далее, произведя подготовительные действия, получим величины одногодичных разностей (табл. 41).

Таблица 41

Одногодичные разности

x	$\delta^3 u_1$	$\delta^2 u_1$	$\delta^1 u_1$
22	52,440	—84,280	—15,120
27	—20,880	52,000	49,360

Путем интерполирования найдем искомые годовые коэффициенты смертности m_x (табл. 42).

Таблица 42

Годовые коэффициенты смертности

x	$\delta^3 u_1$	$\delta^2 u_1$	$\delta^1 u_1$	$u=\lg(m_x+0{,}1)$	m_x
22	52,440	—84,280	—15,120	$\bar{1}$,02567	0,00609
23	—	—31,840	—99,400	$\bar{1}$,02552	0,00605
24	—	20,600	—131,240	$\bar{1}$,02452	0,00581
25	—	73,040	—110,640	$\bar{1}$,02321	0,00549
26	—	—	—37,600	$\bar{1}$,02211	0,00522
27	—20,880	52,000	49,360	$\bar{1}$,02173	0,00513
28	—	31,120	101,360	$\bar{1}$,02222	0,00525
29	—	10,240	132,480	$\bar{1}$,02324	0,00550
30	—	—10,640	142,720	$\bar{1}$,02456	0,00582
31	—	—	132,080	$\bar{1}$,02599	0,00617
32	—	—	—	$\bar{1}$,02731	0,00649

ТАБЛИЦЫ СМЕРТНОСТИ ПО ПРИЧИНАМ СМЕРТИ

Гипотеза устранения болезней как причин смерти (Д. Бернулли)

При изучении смертности очень важно уяснение роли и значения отдельных причин смерти. Еще академик Даниил Бернулли в 1766 г. в работе «Опыт нового анализа смертности, вызванный оспой, и тех преимуществ, которые возникают при ее прививке»[1] привел две таблицы смертности. Одна из них построена галлеевским методом для обычного населения, а другая — при гипотезе устранения оспы как причины смерти. При этом исходным являлось предположение о постоянстве меры заболеваемости оспой для всех возрастов и постоянстве средней меры смертности от оспы для заболевших ею.

Этот метод может быть применен для выявления уменьшения средней продолжительности жизни из-за любой болезни.

Обозначим: x — возраст; n — число людей, из которых заболевает данной болезнью один человек; m — число заболевших, из которых умирает один человек; l — число доживающих до возраста x лет; S — число не заболевших данной болезнью до возраста x лет.

В течение времени dx не заболевают dS_x людей. Число заболевших из S составляет $\frac{S}{n}$, а за время dx будет $dS=\frac{SdS}{n}$. Найдем число умерших от данной болезни: $\frac{SdS}{n}\cdot\frac{1}{m}=\frac{SdS}{nm}$.

[1] D. Bernoulli. Essei d'une nouvelle analyse de la mortalité causée par la petite verole et les avantages de l'inoculation pour la prévenir. «Histoire de l'Académie Royale des Sciences», année 1760. Paris, 1766, p. 1—45.

Значит от других болезней умирает $dl\left(-\frac{SdS}{mn}\right)$. Рассчитаем число тех из совокупности l, кто еще не болел данной болезнью, т. е. S:

$$-dS = \frac{Sdx}{n} - \frac{Sdl}{l} - \frac{Sdx}{mnl}.$$

Преобразуя и принимая $\frac{l}{S} = q$, получаем $dx = \frac{mndq}{mq-1}$. Интегрируя, получаем $x + c = n\ln(mq-1)$. Окончательно, после обратной подстановки имеем:

$$S = \frac{m}{l^{\frac{x+c}{n}} + 1}\, l.$$

Произвольная постоянная интегрирования c получается при $x = 0, S = l =$ числу новорожденных. После подстановки и преобразований получается:

$$S = \frac{m}{(m-1)\, l^{\frac{x}{n}} + 1}\, l.$$

Д. Бернулли, основываясь на эмпирических данных, принимает, что из восьми человек, не перенесших оспу, за год заболевает один человек и из восьми заболевших умирает тоже один человек. Тогда

$$S = \frac{8}{7 l^{\frac{7}{8}} + 1}.$$

На основании этой формулы Д. Бернулли составляет таблицу и приходит к выводу о том, насколько сокращается средняя продолжительность жизни под влиянием конкретной болезни — оспы. Например, при устранении влияния оспы длительность жизни для новорожденных в среднем увеличилась на 3 года и 2 месяца.

Краткие условные и краткие таблицы смертности

Изучением влияния отдельных причин на смертность в СССР занимались такие демографы, как В. В. Паевский, С. А. Новосельский, Ю. А. Корчак-Чепурковский, и врачи А. М. Мерков, Р. Н. Бирюкова, М. С. Бедный и др. Вычислялась вероятность смерти от отдельных причин: туберкулеза, злокачественных новообразований, воспаления легких, сердечно-сосудистых заболеваний, травм и насильственных причин; выяснялось влияние этих болезней на среднюю продолжительность жизни.

При этом по данным, например, 1926—1927 гг. вычислялись:

$$q_0 = \frac{M_0 \text{ (число умерших детей в возрасте до 1 года)}}{{}^1/_3 \text{ родившихся в 1925 г.} + \text{родившиеся в 1926 г.} + {}^2/_3 \text{ родившихся в 1927 г.}};$$

$$q_x = \frac{2m_x}{2+m_x}, \text{ где } m_x = \frac{M \text{ (число умерших в данном возрасте)}}{S \text{ (средняя численность населения в данном возрасте)}}$$

$$q_{1/1+4} = \frac{2 \cdot 4m_{1/1+4}}{2+4m_{1/1+4}}$$ для населения от 1 до 4 лет;

$$q_{x/x+5} = \frac{2 \cdot 5m_{x/x+5}}{2+5m_{x/x+5}}$$ для населения в пятилетнем интервале возраста;

$$q_{x/x+10} = \frac{2 \cdot 10m_{x/x+10}}{2+10m_{x/x+10}}$$ для населения в десятилетнем интервале возраста.

Расчет p_x осуществляется по формуле $p_x = 1 - q_x L_x$, где $L_x = \frac{l_x + l_{x+1}}{2}$ — численность стационарного населения в возрасте x лет.

Кроме указанных нами выше кратких таблиц смертности, можно вычислять *условные* таблицы смертности (тоже краткие).

Условные таблицы смертности получаются путем использования чисел умерших d_x, расчет которых производится поочередным исключением из числа всех умерших тех, кто умер от одной из изучаемых причин. Получаются различные ряды d_x. Количество рядов d_x зависит от числа изучаемых причин смерти.

По каждому из рядов d_x можно построить краткую условную таблицу смертности и затем сопоставить средние продолжительности жизни по данной таблице и фактической краткой таблице смертности. Таким образом, можно выявить влияние каждой из изучаемых причин на среднюю продолжительность жизни. Работа, проведенная А. М. Мерковым по определению влияния злокачественных новообразований на среднюю продолжительность жизни УССР в 1926—1927 гг. и Р. Н. Бирюковой для более позднего периода, позволила, например, для 1926—1927 гг. установить влияние следующих причин смерти на сокращение средней продолжительности жизни (табл. 43).

Таблица 43

Средняя продолжительность предстоящей жизни ($\overset{\circ}{e}_x$)*

	Фактическая средняя продолжительность жизни $\overset{\circ}{e}_x$	$\overset{\circ}{e}_x$ при исключении смертности от								
		воспаления легких	туберкулеза	болезней сердца	рака и прочих злокачеств. новообразований	травм и насильственных причин	скарлатины	дизентерии	брюшного тифа	дифтерии
Европейская часть СССР . .	46,12	51,60	51,03	48,68	48,04	48,04	47,57	46,71	46,52	46,33
БССР . . .	53,07	57,60	57,62	56,08	54,61	54,88	54,72	53,28	53,57	53,33
Москва . .	49,33	55,96	53,25	51,61	52,21	51,14	50,81	49,60	49,59	49,66

* Р. Н. Бирюкова. Таблица смертности по причинам смерти. В сб. «Проблемы демографической статистики». М., Госстатиздат, 1959, стр. 339.

Из данной таблицы можно определить, какой урон средней продолжительности жизни населения наносят те или иные болезни. Например, в Москве в 1926—1927 гг. урон от воспаления легких составлял 55,96—49,33 = 6,63 года средней продолжительности жизни и т. д. Этот метод является приближенным, так как при сокращении смертности нарушается требование $\Sigma d_x = 1$. Ошибка, как правило, мала, но может стать существенной, если одновременно исключается действие нескольких значительных причин смерти.

Можно рекомендовать методику составления таблиц смертности и для специального изучения эффективности отдаленных результатов лечения больных различными методами, т. е. в тех случаях, когда критерием эффективности лечения является не снижение летальности, а длительность жизни после лечения.

При этом сопоставление показателей выживаемости однородных больных после лечения различными методами[1] дает возможность определить наиболее целесообразные методы лечения.

Очень интересна задача, предложенная М. С. Бедным и состоящая в выявлении того, «за счет снижения смертности от каких причин произошло увеличение продолжительности жизни и как это отразилось на других показателях таблиц смертности?»[2]. Решение этой задачи позволит выявить корреляционную связь между причинами смерти и средней продолжительностью жизни, т. е. демографические последствия каждой причины в динамике средней продолжительности жизни.

М. С. Бедный предлагает решить эту задачу путем определения силы смертности от исследуемой причины на начало и конец изучаемого периода. Если затем из силы смертности от всех причин вычесть разность между этими двумя силами смертности, то получится уменьшенная сила смертности. В построенной на основе найденной уменьшенной силы смертности новой таблице смертности будет учтено снижение смертности от рассматриваемой причины за изучаемый период.

Отдел статистики населения и здравоохранения ЦСУ СССР применяет в своей работе метод построения кратких условных таблиц смертности, изложенный в учебнике А. Я. Боярского. При этом вероятность умереть при исключении смертности от той или иной болезни находят по формуле

$$q_x^1 = \frac{(1-w)\,q_x}{1-\frac{1}{2}\,wq_x},$$

где w — доля умерших от данной болезни в общем числе умерших в возрасте x лет; q_x — вероятность умереть по обычным таблицам смертности.

[1] Рекомендуемые для этой цели приемы измерения выживаемости, основанные на методике таблиц смертности, разработаны А. М. Мерковым и подробно изложены в его книге «Демографическая статистика» (М. Медгиз, 1959, стр. 123—129).

[2] М. С. Бедный. Продолжительность жизни (статистика, факторы, возможности увеличения). М., «Статистика», 1967, стр. 126.

В этих таблицах смертности производится элиминирование влияния следующих групп причин смерти:

всех болезней органов кровообращения (в том числе гипертонической болезни — отдельно);

злокачественных новообразований;

болезней нервной системы и органов чувств;

инфекционных и паразитарных болезней;

воспаления легких.

Изучение влияния каждой из указанных болезней на среднюю продолжительность жизни позволило сформулировать вывод, что наибольший урон в ее сокращении наносят болезни органов кровообращения. Для возрастов до 70 лет этот урон составляет около пяти лет. Далее по значимости следуют злокачественные новообразования, сокращающие для населения до возраста 40 лет среднюю продолжительность жизни на три года, и болезни нервной системы для всех возрастов до 70 лет причиняют потери в 2,4 года и т. д.

Условные таблицы смертности строятся ЦСУ СССР по следующим возрастным интервалам: 0, 1, 2, 3—4, 5—6, 7—13, 14—19, 20—24, 25—29 лет, далее с интервалом 10 лет.

С 1965 г. запроектировано и производится построение условных таблиц с группировкой всего населения по пятилетним интервалам, а для детей до 5 лет — по одногодичным.

МЕРА ТОЧНОСТИ ТАБЛИЦ СМЕРТНОСТИ

Для практического использования таблиц смертности нужно быть уверенным, что показатели в таблицах свободны от ошибок и являются достаточно надежными.

Если исключить из анализа арифметические ошибки, которые легко могут быть обнаружены при проверке расчетов, то причиной неточности может быть недостаточная массовость, т. е. малое число людей, охваченных наблюдением. В результате характеристики (средние и доли) не очень больших совокупностей будут носить следы случайности.

Достоверность вероятностей дожития и смерти

Допустим, что мы хотим оценить достоверность исчисленных в таблице смертности вероятностей дожить или умереть, т. е. p_x или q_x. Из математической статистики[1] известно, что средняя ошибка доли признака, т. е. относительной величины интенсивности, может быть вычислена по формуле

$$\mu_x = \sqrt{\frac{w\,(1-w)}{n}},$$

[1] См. И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики, стр. 128.

где μ_x — ошибка вероятности дожить или умереть в возрасте x лет; n — абсолютное число живущих в определенном возрасте; w и $1-w$ — соответственно p_x и q_x (т. е. вероятности). Предельная ошибка выборки равна утроенной средней ошибке. Используя соответствующие данные для городского населения, найдем, например, ошибки показателей p_0 и p_{80}. Численность населения в возрасте до 1 года примем в 1 млн. человек, а в возрасте 80 лет — 0,1 от численности населения до 1 года, т. е. 100 тыс. человек. Тогда

$$\mu_0 = \sqrt{\frac{0,95981 \cdot 0,04019}{1\,000\,000}} = 0,0002$$

и

$$\Delta_0 = 3 \cdot \mu_0 = 3 \cdot 0,0002 = 0,0006;$$

$$\mu_{80} = \sqrt{\frac{0,90983 \cdot 0,09017}{100\,000}} = 0,001$$

и

$$\Delta_{80} = 3 \cdot 0,001 = 0,003.$$

Можно использовать теорию вероятностей при проверке точности таблиц смертности и другим способом. Известно, что интеграл вероятностей можно записать и в следующем виде:

$$P = \frac{2}{\sqrt{\pi}} \int_0^{\gamma} e^{-t^2}\, dt = \Phi(\gamma).$$

Это соответствует вероятности того, что при n испытаниях число осуществлений события A будет находиться в пределах $np \pm a$. При этом

$$\gamma = \frac{a}{\sqrt{2\pi npq}} = \frac{a\sqrt{n}}{\sqrt{2m'\,n'}},$$

где m' и n' — наивероятнейшие исходы ($m' = np$; $n' = nq$). Примем отклонение вероятности $l = \frac{a}{n}$. Тогда

$$\gamma = l\,\frac{\sqrt{n^3}}{\sqrt{2m'\,n'}}.$$

Подставим вместо n число доживающих до возраста x лет n_x, вместо m' равное ему n_{x+1}, вместо n' число умерших из числа n_x при переходе от возраста x к возрасту $x+1$, т. е. d_x. Получаем:

$$\gamma = l\,\frac{\sqrt{n_x^3}}{\sqrt{2n_{x+1}\,d_x}}.$$

Зная вероятность наивероятнейшего исхода $P_\gamma = \frac{0,3989}{\sqrt{npq}}$ и подставляя в формулу γ, получим:

$$P_\gamma = 0,3989\sqrt{\frac{n_{x+1}\,d_x}{n_x}}.$$

Пусть известно, что вероятность умереть мужчине в возрасте, например, 50 лет в течение следующего года жизни по таблицам смертности какой-нибудь союзной республики определена на основе 116 984 наблюдений, причем смертных случаев в течение года в этом возрасте было 998. Определим вероятность того, что в полученном значении $q_{50} = \frac{998}{116\,984} = 0,00853$ верен:

1) второй знак после запятой;
2) третий знак после запятой;
3) четвертый знак после запятой.

Используем формулу

$$\gamma = l\,\frac{\sqrt{n_x^3}}{\sqrt{2n_{x+1}d_x}} = l\,\frac{\sqrt{116984^3}}{\sqrt{2 \cdot 115986 \cdot 998}} = l2629,703.$$

Придавая l значения $l_1 = 0,01$, $l_2 = 0,001$ и $l_3 = 0,0001$, получаем:

$$\gamma_1 = 0,01 \times 2629,703 = 26,297 \approx 26;$$

$$\gamma_2 = 0,001 \times 2629,703 = 2,6297 \approx 2,6;$$

$$\gamma_3 = 0,0001 \times 2629,703 = 0,26297 \approx 0,26.$$

Используя таблицу интеграла вероятностей Маркова, находим значения вероятностей

$$P_{\gamma_1} \approx 1; \quad P_{\gamma_2} \approx 0,9998; \quad P_{\gamma_3} \approx 0,2900.$$

Полученные значения можно рассматривать как ответы на интересующий нас вопрос и трактовать их следующим образом: ошибки во втором и третьем знаке после запятой в вероятности умереть, т. е. q_x, мало вероятны. Что же касается ошибки в четвертом знаке (которая в конечном итоге может коснуться и третьего знака), то она очень вероятна и, следовательно, на четвертый знак в вероятности умереть полагаться нельзя. Анализируя таблицу смертности, заключаем, что по мере увеличения возраста от 50 лет и старше необходимо располагать все большим и большим числом n_x, чтобы четвертый десятичный знак был точным.

Какова же минимальная численность возрастной группы, чтобы на основе теории вероятностей можно было бы быть уверенным в правильности четвертого или хотя бы третьего десятичного знака?

Возьмем, например, возраст $x=66$ годам. По таблице смертности 1958—1959 гг. для мужчин $q_{66} = 0,03300$. Чтобы при вероятности 0,9998 был точен третий десятичный знак, нужно взять n_{66} больше 400 000 человек.

Все вышеизложенное позволяет сделать вывод о том, что вероятностные показатели в таблицах смертности верны только до третьего десятичного знака. Поэтому уже на четвертый десятичный знак следует смотреть как на величину, которая может отклоняться от действительной на несколько единиц.

В заключение разбора ошибок данного вида следует заметить, что как бы ни были велики эти ошибки, их можно рассчитать и учитывать в практической деятельности.

Неточность вероятностных показателей из-за допущений, не соответствующих истине

Причиной неточности может быть несоответствующее истине допущение, что совокупности умерших III рода, т. е. умершие в определенном возрасте, по признаку года рождения делятся на две равные друг другу группы, т.е. в равной мере могут быть отнесены к двум соседним годам рождения. Оценим эту ошибку. Построим график (см. рис. 23).

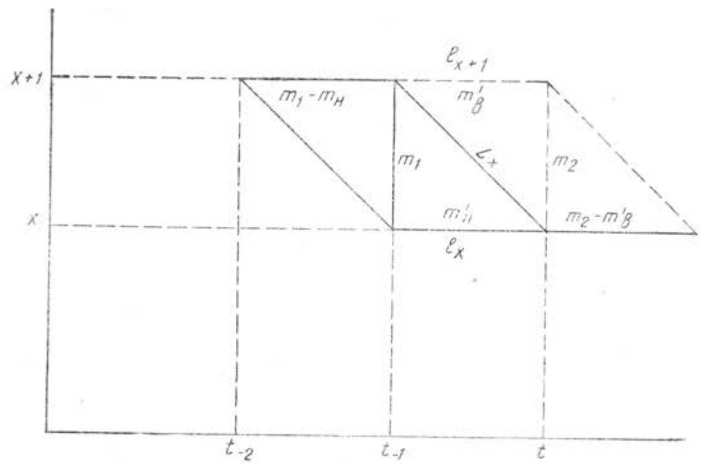

Рис. 23. Деление совокупностей умерших третьего рода по годам рождений.

Обозначим L_x — население, учтенное по переписи в возрасте x лет рождения t — 1 года; m_1 и m_2 — соответственно совокупности умерших III рода в возрасте x лет за год, предшествующий переписи и последовавший после переписи; $m'_н$ — нижняя элементарная совокупность умерших в возрасте x лет t — 1 года рождения за год, предшествующий переписи; $m'_в$ — верхняя элементарная совокупность умерших в возрасте x лет t — 1 года рождения за год, следующий за переписью.

Произведем расчет вероятности дожить до $x + 1$ года:

$$p_x = \frac{l_{x+1}}{l_x} = \frac{L_x - m'_в}{L_x + m'_н}.$$

По нашему предположению, умершие m_1 состоят из равных частей родившихся в t — 2 и t — 1 годах, а умершие m_2 состоят из равных частей родившихся в t — 1 и t-м годах. Следовательно,

$$m'_н = 0{,}5\ m_1, \quad m'_в = 0{,}5 m_2.$$

$$p_x = \frac{L_x - 0{,}5 m_2}{L_x + 0{,}5 m_1}.$$

Из возможных случаев рассмотрим самый неблагоприятный, соответствующий следующим предположениям: число умерших за год до переписи m_1 состоит только из родившихся в t — 2 году, т. е. $m'_н = 0$, а число умерших за год после переписи состоит только из родившихся в t — 1 году, т. е. $m_2 = m'_в$. Таким образом, из родившихся в t-м году в возрасте x лет никто не умирал. Возможен и другой самый неблагоприятный случай, когда $m'_в = 0$ и $m'_н = m_1$.

В первом, самом неблагоприятном случае (предположении) имеем:

$$m'_н = 0 \text{ и } m_2 = m'_в.$$

Значит, подставляя в формулу, получаем:

$$p_{x(\text{при первом предположении})} = \frac{L_x - m_2}{L_x + 0}.$$

Ошибка при этом получится:

$$\Delta_1 = p_{x\ (\text{при первом предположении})} - p_x;$$

$$\Delta_1 = \frac{L_x - m_2}{L_x} - \frac{L_x - m'_в}{L_x + m'_н} = \frac{L_x^2 - L_x m_2 + L_x m'_н - m'_н m_2 - L_x^2 + L_x m'_в}{L_x (L_x + m'_н)} =$$

$$= \frac{L_x (m'_н + m'_в) - m_2 (L_x + m'_н)}{L_x (L_x + m'_н)}.$$

После деления на L_x и сокращения получаем:

$$\Delta_1 = \frac{m'_н + m'_в - m_2 \left(1 + \frac{m'_н}{L_x}\right)}{L_x + m'_н}.$$

Проанализируем полученный результат и сделаем выводы. Во-первых, результаты наблюдений свидетельствуют, что совокупности умерших третьего рода m_2 и совокупности умерших первого рода $m'_н + m'_в$ отличаются друг от друга незначительно, т. е. весьма близки друг другу. Во-вторых, Δ_1 — ошибка, которая в одних случаях может получиться с плюсом, а в других случаях — с минусом. В-третьих, что $\frac{m'_н}{L_x}$ очень малая дробь, и поэтому сумма $1 + \frac{m'_н}{L_x}$ незначительно отличается от единицы, а весь числитель поэтому близок к нулю. Знаменатель же может, как, например, в таблицах смертности 1958—1959 гг., достигать нескольких миллионов. Поэтому ошибку Δ_1 можно считать равной десятитысячным или стотысячным долям.

Разбирая второй неблагоприятный случай ($m_в' = 0$ и $m_н' = m_1$), придем к такому же результату в отношении Δ_2.

Учитывая все вышеприведенное, можно сделать вывод, что неточностями, вызванными предположением о расчленении совокупности умерших III рода на две равные элементарные совокупности, можно пренебречь. На вероятность дожития или смерти эти ошибки оказывают очень слабое влияние, отражающееся только на четвертом или даже на пятом знаке после запятой. Тем более можно пренебречь этой ошибкой, что не во всякой возрастной группе числа умерших двух соседних лет рождения сочетаются так, что дают максимум ошибки. Очень большие ошибки представляют собой исключительное явление, т. е. встречаются крайне редко.

Неточность из-за непринятия в расчет механического движения населения

Если речь идет о таблицах смертности для населения всего Советского Союза, то ввиду крайне малой внешней миграции ею вообще можно пренебречь. Если же таблицы смертности составляются для населения какой-нибудь республики или еще меньшей территориальной единицы с последующей группировкой на городское и сельское население, где миграция может иметь ощутимые размеры, ошибка возрастет.

Вероятность дожития до возраста $x+1$ лет с учетом миграции может быть рассчитана по формуле

$$p_x' = \frac{L_x - m_в'}{L_x + m_н' \pm A},$$

где A — абсолютная величина механического прироста населения в возрасте x лет того же года рождения.

Оценим влияние величины A на p_x. Для этого найдем разность между вероятностями дожития с учетом и без учета миграции:

$$\Delta = p_x - p_x' = \frac{L_x - m_в'}{L_x + m_н'} - \frac{L_x - m_в'}{L_x + m_н' \pm A}.$$

Обозначая $L_x - m_в' = L_x'$ и $L_x + m_н' = L_x''$, получим

$$\Delta = \frac{L_x'}{L_x''} - \frac{L_x'}{L_x'' \pm A} = \frac{\pm L_x' A}{(L_x'')^2 \pm A L_x''}.$$

Принимая сальдо миграции A положительным, разделим на нее числитель и знаменатель:

$$\Delta = \frac{L_x'}{\dfrac{(L_x'')^2}{A} + L_x''}.$$

Рассмотрим случай, когда относительная миграция составляет 5% естественного прироста. Тогда $\frac{A}{L_x} = 0{,}05\ K_{ест.пр}$, а коэффициент естественного прироста равен $15^0/_{00}$. Следовательно, $A = L_x \cdot 0{,}05 \times \times 0{,}015 = 0{,}00075\ L_x$.

Принимая $L_x'' \approx L_x' = L_x$, получаем:

$$\Delta = \frac{L_x}{\dfrac{(L_x)^2}{0{,}00075 L_x} + L_x} \approx \frac{L_x}{\dfrac{L_x}{0{,}00075} + L_x} \approx 0{,}0001.$$

Значит ошибка составляет при этих предположениях одну десятитысячную и может, следовательно, влиять на четвертый знак после запятой в вероятности дожития или смерти.

При миграциях бо́льших, чем мы приняли, она может повлиять и сильнее. Так, в предположении, что миграция составляет уже не 5% естественного прироста, а больше, получаем неточным в вероятности дожития или смерти уже третий знак после запятой.

Неточности из-за неправильного выбора кривой для выравнивания показателей l_x или q_x

В качестве математико-статистического критерия можно предложить сравнение фактических данных с выровненными путем нахождения погодовых разностей. Из найденных погодовых разностей определяются средние по пятилетним или десятилетним интервалам. Если найденные средние разности по укрупненным интервалам не превышают утроенную или минимум удвоенную среднюю ошибку, то следует полагать, что выравнивание не привело к ошибкам, превышающим допустимые теорией вероятностей погрешности.

Рассмотрим еще одну погрешность, возникающую при составлении таблиц смертности для обоих полов, когда используют среднюю арифметическую (не взвешенную по числу рождений) из равновозрастных показателей мужского и женского населения.

Введем следующие обозначения: число доживающих до возраста x лет для мужчин — $l_x^м$, для женщин — $l_x^ж$, для обоих полов — l_x; число рождений мальчиков на 100 девочек $100+z$.

Тогда взвешенная средняя (общее число доживающих) будет:

$$l_x = \frac{(100+z)\,l_x^м + 100 l_x^ж}{200+z} = \frac{l_x^м + l_x^ж}{2}\left(\frac{1}{1+\dfrac{z}{100}}\right) +$$

$$+\frac{z \cdot l_x^м}{200+z} = \frac{l_x^м + l_x^ж}{2} - \frac{l_x^м + l_x^ж}{2}\left(\frac{z}{200} - \frac{z^2}{200^2} + \dots\right) + \frac{z l_x^м}{200+z}.$$

Следовательно, если взять только первый член, являющийся средней арифметической невзвешенной $\frac{(l_x^{м}+l_x^{ж})}{2}$, то при этом допускается какая-то погрешность. Принимая $z=5$ и в качестве примера возраст $x=5$, получим по таблице смертности 1958—1959 гг.:

$$l_5=\frac{0,94029+0,94823}{2}=0,94426.$$

Взвешенная же средняя равна:

$$l_5=\frac{105\cdot 0,94029+100\cdot 0,94823}{205}=0,94416.$$

Невзвешенная и взвешенная средние отличаются на одну сотую процента. Для дальнейших возрастов эта разница еще меньше. Однако величина средней продолжительности жизни, полученная как средняя из таблиц для мужчин и женщин, может существенно отличаться от показателя средней продолжительности жизни, полученного из единой таблицы для всего населения, из-за половых диспропорций в реальном населении.

Учитывая, что неточен только четвертый знак после запятой, нет надобности прибегать к взвешиванию, ибо погрешность выходит за пределы точности таблиц смертности.

На основании рассмотрения предполагаемых нами причин погрешностей при исчислении таблиц смертности можно сделать вывод, что большинство погрешностей лежит за пределами точности вычислений, а главная погрешность, влияющая уже на третий знак после запятой, — недостаточная массовость возрастных групп.

Отсюда следует, что *для практических целей полные таблицы смертности могут быть заменены неполными, где возрастной интервал в 5—10 лет приводит к увеличению численности групп и перенесению неточностей вероятностей уже на четвертый или пятый знаки после запятой.*

НЕКОТОРЫЕ ЭМПИРИЧЕСКИЕ ЗАКОНОМЕРНОСТИ ТАБЛИЦ СМЕРТНОСТИ

Отсроченная средняя продолжительность жизни как математическое ожидание

Если обозначить отсроченную на n лет временную, до возраста t лет, среднюю продолжительность жизни при рождении — ${}_{n/t}e_0^0$, то эта величина может быть представлена как математическое ожидание числа лет, которое проживет новорожденный в промежутке от n до t лет. Этот показатель используется для измерения эффекта снижения смертности. Для рассмотрения отсроченной временной средней продолжительности жизни возьмем число человеко-лет, которое прожито населением до возраста n лет (T_n) и до t лет (T_t), затем вычитанием находим число человеко-лет, прожитых населением в интервале от n до

t лет ($T_n - T_t$). Найденный результат разделим на число родившихся l_0, лежащее в основе таблицы смертности и равное 100 000:

$$_{n/t}\overset{\circ}{e}_0=\frac{T_n-T_t}{l_0}=\frac{T_n}{l_0}-\frac{T_t}{l_0}.$$

Подключим числа доживающих до возраста t и n лет. Тогда

$$_{n/t}\overset{\circ}{e}_0=\frac{l_n}{l_0}\cdot\frac{T_n}{l_n}-\frac{l_n l_t T_t}{l_0 l_n l_t}=\frac{l_n}{l_0}\left(\frac{T_n}{l_n}-\frac{l_t}{l_n}\cdot\frac{T_t}{l_t}\right).$$

Полагая $\frac{T_n}{l_n}$ и $\frac{T_t}{l_t}$ равными средней продолжительности предстоящей жизни в возрасте n и t лет, т. е. $\overset{\circ}{e}_n$ и $\overset{\circ}{e}_t$, получаем окончательно

$$_{n/t}\overset{\circ}{e}_0=\frac{l_n}{l_0}\left(\overset{\circ}{e}_n-\frac{l_t}{l_n}\overset{\circ}{e}_t\right).$$

Значит, *временная средняя продолжительность жизни* может означать число лет, прожитых в среднем одним лицом из числа l_x лиц в течение следующего определенного промежутка времени (от x до $x+n$ лет).

Отсроченная на n лет средняя продолжительность жизни означает число предстоящих в среднем лет жизни одному лицу из l_x лиц после достижения им возраста $x+n$ лет.

Таким образом, *отсроченная временная средняя продолжительность жизни* выражает среднее число лет жизни одного лица из числа l_x лиц в течение t лет, следующих после n лет, т. е. от $x+n$ до $x+n+$ $+t$ лет.

Установим три контингента лиц, чрезвычайно интересующих демографов:

фертильный контингент женщин в возрасте 15—49 лет;

военный контингент мужчин в возрасте 20—50 лет;

работоспособное население обоих полов в возрасте 15—60 лет.

Тогда ${}_{15/49}e_0^0$ для женского населения будет означать отсроченную на 15 лет временную до 49 лет среднюю продолжительность жизни девочек при рождении и сможет быть найдена по формуле

$$_{15/49}\overset{\circ}{e}_0=\frac{l_{15}}{l_0}\left(\overset{\circ}{e}_{15}-\frac{l_{49}}{l_{15}}\overset{\circ}{e}_{49}\right),$$

где все обозначения сохраняют указанные значения и берутся из таблиц смертности для женщин. Эта величина, т. е. отсроченная средняя продолжительность жизни женщин 15—49 лет, рассчитанная по таблицам смертности, принимает различные значения и колеблется в весьма широких границах. Так, по таблице смертности 1926—1927 гг. в СССР для женщин ${}_{15/49}e_0^0\approx 21,62$, а в 1958—1959 гг. равнялась 31,19 года. Аналогично можно установить, что отсроченная средняя продолжительность жизни общего населения в возрасте 15—60 лет, вычисленная по формуле

$$_{15/60}\overset{\circ}{e}_0 = \frac{l_{15}}{l_0}\left(\overset{\circ}{e}_{15} - \frac{l_{60}}{l_{15}}\overset{\circ}{e}_{60}\right),$$

по данным таблиц смертности для обоих полов колеблется в весьма широких границах. Так, по таблице смертности 1926—1927 гг. в СССР $_{15/60}e_0^0 = 26{,}42$, а по таблицам смертности 1958—1959 гг. этот показатель равен 39,63 года.

Удельные веса отсроченной средней продолжительности жизни разных контингентов в общей средней продолжительности жизни

Попробуем установить удельный вес ожидаемой при рождении средней продолжительности жизни рассматриваемых двух контингентов во всей средней продолжительности жизни новорожденных.

Обозначим эти удельные веса $K_а$ и $K_б$:

$$K_а = \frac{_{15/49}\overset{\circ}{e}_0}{\overset{\circ}{e}_0} \text{ и } K_б = \frac{_{15/60}\overset{\circ}{e}_0}{\overset{\circ}{e}_0}.$$

Очень важно, что эти коэффициенты представляют собой константы, совершенно не зависящие от значения коэффициента смертности, страны и времени их расчета. Замеченное соотношение представляет собой определенный статистический закон. Произведенные нами расчеты по таблицам смертности различных стран начиная с 1890 г. показали, что $K_а$ не выходит за границы $0{,}45 \pm 0{,}01$, а $K_б$ не выходит за границы $0{,}62 \pm 0{,}06$[1]. В качестве примера приведем расчеты этих коэффициентов по таблице смертности 1926—1927 гг. для европейской части:

$$K_а = \frac{\frac{69\,569}{100\,000}\left(51{,}45 - \frac{56\,259}{69\,569}\cdot 25{,}19\right)}{46{,}79} = \frac{21{,}62}{46{,}79} \approx 0{,}46;$$

$$K_б = \frac{\frac{67\,902}{100\,000}\left(49{,}48 - \frac{44\,702}{67\,902}\cdot 16{,}05\right)}{44{,}35} = \frac{26{,}42}{44{,}35} \approx 0{,}60.$$

По таблицам смертности 1958—1959 гг. расчеты дадут следующие результаты:

$$K_а = \frac{\frac{94\,044}{100\,000}\left(61{,}09 - \frac{87\,963}{94\,044}\cdot 29{,}86\right)}{71{,}68} = \frac{31{,}19}{71{,}68} \approx 0{,}44,$$

Для обоих полов:

$$K_б = \frac{\frac{93\,497}{100\,000}\left(58{,}22 - \frac{76\,693}{93\,497}\cdot 19{,}30\right)}{68{,}59} = \frac{39{,}63}{68{,}59} \approx 0{,}58.$$

[1] Величина 0,6, относящаяся к общему населению в возрасте 15—60 лет, впервые была получена и описана в работе M o r t a r a. Lezioni di Statistica Economica e Demografica. Roma, 1920.

Критика «теории нежелательности» снижения смертности

Полученная нами величина $K_б$ позволяет рассеять одно заблуждение буржуазной демографии, которая утверждает, что понижение общей смертности в стране является якобы фактором, действующим отрицательно на ее экономику, так как приводит к снижению доли производительной части населения. При этом приводят следующие далеко не оригинальные аргументы: понижение общей смертности происходит за счет понижения ее в детских, юношеских и старческих возрастах. Что же касается смертности в рабочих возрастах, то она снижается очень слабо, а иногда даже и повышается. Значит, увеличивается средняя продолжительность жизни экономически неактивных групп населения. Отсюда делают вывод об экономических потерях при снижении общего коэффициента смертности, а следовательно, о нежелательности этого снижения. Такой порочный человеконенавистнический вывод может быть опровергнут с помощью показателя $K_б$. В самом деле, если отношение $_{15/60}e_0^0$ к e_0^0 есть величина постоянная и бо̀льшая, чем 0,5, то снижение общего коэффициента смертности, увеличивая e_0^0, в большей степени увеличивает ожидаемую среднюю продолжительность жизни в рабочем, т. е. экономически активном, возрасте ($_{15/60}e_0^0$), чем в возрастах до 15 лет и старше 60 лет. Тем самым *увеличивается процент рабочих лет во времени, проживаемом всем населением.*

Т а б л и ц а 44

Сопоставление отсроченной и средней продолжительности жизни по республикам Советского Союза (1958—1959 гг.)

Республики	Удельный вес отсроченной продолжительности жизни во всей средней продолжительности жизни				
	Женщины в возрасте			Мужчины в возрасте 18—50 лет	Оба пола в возрасте 15—60 лет
	0—15	15—49	60—100		
РСФСР	0,199	0,438	0,230	0,446	0,583
Украинская ССР	0,198	0,437	0,233	0,435	0,579
Белорусская ССР	0,195	0,430	0,244	0,430	0,572
Узбекская ССР	0,191	0,410	0,268	0,424	0,540
Казахская ССР	0,199	0,420	0,257	0,423	0,558
Грузинская ССР	0,190	0,419	0,261	0,427	0,557
Азербайджанская ССР	0,188	0,406	0,273	0,409	0,543
Литовская ССР	0,199	0,440	0,227	0,434	0,582
Молдавская ССР	0,199	0,435	0,235	0,423	0,571
Латвийская ССР	0,200	0,441	0,225	0,440	0,585
Киргизская ССР	0,191	0,412	0,268	0,418	0,550
Таджикская ССР	0,188	0,410	0,277	0,405	0,528
Армянская ССР	0,191	0,414	0,263	0,495	0,552
Туркменская ССР	0,191	0,409	0,266	0,407	0,542
Эстонская ССР	0,199	0,444	0,221	0,445	0,589
Итого по СССР	0,198	0,435	0,234	0,437	0,578

Этот научный вывод опровергает теории о нежелательности снижения смертности из-за экономических потерь и убедительно свидетельствует о необходимости борьбы за снижение смертности. Произведенные нами расчеты по таблицам смертности 1958—1959 гг. по республикам Советского Союза дали очень устойчивые результаты (см. табл. 44).

ПРИЛОЖЕНИЕ ТЕОРИИ ВЕРОЯТНОСТЕЙ К ТАБЛИЦАМ СМЕРТНОСТИ

Рассмотрим применение некоторых теорем теории вероятностей в демографии.

Начнем с некоторых приложений теории вероятностей к вероятностным показателям таблиц смертности.

Вероятности простых событий

Известны вероятности мальчику, достигшему 12 лет, дожить до 13 лет ($p_{12} = 0{,}99905$), а достигшему 13 лет, дожить до 14 лет ($p_{13} = 0{,}99904$)[1]. Требуется определить вероятность мальчику, достигшему 12 лет, оставаться в живых до 14 лет.

Эти события *зависимые*. Чтобы прожить мальчику от 13 до 14 лет, нужно достичь возраста 13 лет. Поэтому применим *теорему умножения вероятностей* и получим:

$$p_{12-14} = p_{12}\,p_{13} = 0{,}99905 \cdot 0{,}99904 = 0{,}99809.$$

Имеются и противоположные задачи, требующие определения вероятности умереть. Определим вероятность мужчине, достигшему 30 лет, умереть на 31-м или 32-м году жизни. Имеем $q_{30} = 0{,}00328$ и $q_{31} = 0{,}00342$[2]. Использование в данном случае теоремы сложения вероятностей $Q_{31 \text{ или } 32} = q_{30} + q_{31} = 0{,}00670$ будет неправильным.

Данное событие (A), заключающееся в том, что мужчина, достигший возраста 30 лет, умрет на 31-м или 32-м году жизни, событие сложное, состоящее из двух событий (B и C): мужчина может умереть на 31-м году жизни (событие B) или на 32-м (событие C).

Вероятность события B известна и равна $q_{(30)}$. Что же касается события C, то оно предполагает, что прожит 31-й год жизни, т. е. мужчина дожил до 32 лет.

Поэтому для определения вероятности события C используется теорема умножения вероятностей. Умножим вероятность дожить до возраста 31 года ($p_{30} = 1 - q_{30}$) на вероятность умереть на 32-м году жизни (q_{31}):

$$Q_C = p_{30}\,q_{31} = 0{,}99672 \cdot 0{,}00342 = 0{,}00341.$$

[1] «Итоги Всесоюзной переписи населения 1959 года. СССР», стр. 264.

[2] Там же.

Окончательно получаем:

$$Q_A = Q_{B \text{ или } C} = q_{30} + p_{30}\,q_{31} = 0{,}00328 + 0{,}00341 = 0{,}00669.$$

Если теперь развернем полученную формулу, то подтвердим наше заключение о неправильном применении в данном случае теоремы сложения вероятностей и увидим как велика была допускаемая нами при этом ошибка.

$$Q_A = q_{30} + p_{30}\,q_{31} = q_{30} + (1 - q_{30})q_{31} = \boxed{q_{30} + q_{31}} - q_{30}\,q_{31}.$$

Чтобы получить правильный ответ, нужно было из суммы двух вероятностей умереть вычесть их произведение.

Вероятности сложных событий

Разберем два других случая.

1. Оценим вероятность сложного события. Известен возраст вступающих в брак: жениха (x) и невесты (y). Как велика вероятность того, что через n лет оба супруга будут живы, один из супругов будет жив, оба за период n лет умрут?

Возможные простые события:
муж через n лет жив;
жена через n лет жива;
муж через n лет умер;
жена через n лет умерла.

Возможные сложные события:
муж через n лет жив, и жена жива;
муж через n лет жив, а жена умерла;
муж через n лет умер, а жена жива;
муж через n лет умер, и жена умерла.

Придадим конкретные значения возрасту мужа и жены. Пусть возраст вступления в брак жениха — 25 лет, а невесты — 20 лет.

Определим вероятности различных событий. Вероятность первого сложного события, состоящего в том, что через 25 лет оба супруга будут живы, определяются по теореме умножения вероятностей.

Используя таблицу смертности 1958—1959 гг.[1], находим:

вероятность 25-летнему мужчине дожить до 50 лет, т. е. прожить еще 25 лет:

$$p^{\text{м}}_{25/50} = \frac{l^{\text{м}}_{25+25}}{l^{\text{м}}_{25}} = \frac{80604}{91170} = 0{,}88411;$$

вероятность 20-летней женщине дожить до 45 лет, т. е. прожить еще 25 лет:

$$p^{\text{ж}}_{20/45} = \frac{l^{\text{ж}}_{20+25}}{l^{\text{ж}}_{20}} = \frac{89256}{93601} = 0{,}95358.$$

[1] «Итоги Всесоюзной переписи населения 1959 года. СССР», стр. 264—267.

Вычитанием из единицы получаем соответствующие вероятности умереть:

$$q^{м}_{25/50}=1-p^{м}_{25/50}=1-0{,}88411=0{,}11589;$$
$$q^{ж}_{20/45}=1-p^{ж}_{20/45}=1-0{,}95358=0{,}04642.$$

Теперь для сложных событий определим:

1) вероятность того, что через 25 лет оба супруга будут живы:

$$p_{\text{оба живы}}=p^{м}_{25/50}\,p^{ж}_{20/45}=0{,}88411\cdot0{,}95358\approx0{,}843;$$

2) вероятность того, что через 25 лет муж будет жив, а жена умрет:

$$p_{\substack{\text{муж жив,}\\ \text{жена умерла}}}=p^{м}_{25/50}\,q^{ж}_{20/45}=0{,}88411\cdot0{,}04642\approx0{,}041;$$

3) вероятность того, что через 25 лет жена будет жива, а муж умрет:

$$p_{\substack{\text{жена жива,}\\ \text{муж умер}}}=p^{ж}_{20/45}\,q^{м}_{25/50}=0{,}95358\cdot0{,}11589\approx0{,}111.$$

Два последних сложных события несовместимы, и поэтому вероятность того, что в живых будет один из супругов, определится по теореме сложения вероятностей:

$$p_{\text{жив один из супругов}}=p_{\substack{\text{муж жив,}\\ \text{жена умерла}}}+p_{\substack{\text{жена жива,}\\ \text{муж умер}}}=0{,}041+0{,}111=0{,}152;$$

4) вероятность того, что через 25 лет обоих супругов не будет в живых:

$$p_{\text{оба умрут}}=q^{м}_{25/50}\,q^{ж}_{20/45}=0{,}11589\cdot0{,}04642\approx0{,}005.$$

Все рассматриваемые события образуют полную группу, и сумма их вероятностей должна равняться единице:

$$p_{\text{оба живы}}+p_{\text{жив один из супругов}}+p_{\text{оба умерли}}=$$
$$+0{,}843+0{,}152+0{,}005=1{,}000.$$

При помощи таких расчетов можно определить, сколько в среднем пар из вступающих в брак доживает, например, до серебряной свадьбы или золотой:

$$p^{м}_{25/75}=\frac{39125}{91170}\approx0{,}42914;$$
$$p^{ж}_{20/70}=\frac{69236}{93601}\approx0{,}73970.$$

Следовательно, вероятность того, что оба супруга будут живы через 50 лет, можно найти по теореме умножения вероятностей:

$$p_{\text{оба живы}}=0{,}42914\cdot0{,}73970\approx0.317.$$

Это говорит о том, что из 10 браков в 3 муж и жена доживут до золотой свадьбы (при данном уровне смертности в отдельных возрастах)[1].

Произведенный нами расчет ориентирован на будущее. А если мы хотим определить удельный вес пар, доживавших до золотой свадьбы из вступивших в брак 50 лет назад, мы должны использовать таблицы смертности 1920 г. Ввиду отсутствия таковых используем таблицы смертности 1907—1910 гг. (первый вариант)[2] и получим:

$$p^{м}_{25/75}=\frac{l^{м}_{25+50}}{l^{м}_{25}}=\frac{9\,147}{43\,707}=0{,}20928;$$
$$p^{ж}_{20/70}=\frac{l^{ж}_{20+50}}{l^{ж}_{20}}=\frac{15\,972}{47\,914}=0{,}33335;$$
$$p_{\text{оба живы}}=0{,}20928\cdot0{,}33335\approx0{,}070.$$

С учетом сделанных допущений только 7 из 100 пар вступивших в брак 50 лет назад имели шанс дожить до золотой свадьбы при существовавшем тогда уровне смертности. Для исчисления таких показателей для реального населения, т. е. для определения доли доживших сейчас до золотой свадьбы из вступивших в брак 50 лет назад, нужна была бы таблица смертности реального поколения, которая не соответствует ни уровню 1907—1910, ни современному, а является как бы переходной между ними. Такими таблицами мы не располагаем, но расчеты, проделанные нами, показывают границы этих показателей.

2. Известно, что вероятность двум близнецам быть одинакового пола вдвое больше, чем вероятность быть разнополыми. Вероятность рождения мальчика 0,51. Какова же вероятность, что второй близнец мальчик, если установлено, что первый из них мальчик? Введем обозначения:

$p_{М}=0{,}51$ — вероятность рождения мальчика;
$p_{Д}=0{,}49$ — вероятность рождения девочки;
$P_{ММ}$ — вероятность рождения двух мальчиков;
$P_{ДД}$ — вероятность рождения двух девочек;
$P_{МД}$ — вероятность рождения разнополых близнецов, когда неизвестен пол первого из родившихся.

Исходя из условия, запишем:

$$P_{ММ}+P_{ДД}=2P_{МД}.$$

Полагая событие рождения близнецов свершившимся, имеем:

$$P_{ММ}+P_{ДД}+P_{МД}=1;$$
$$2P_{МД}+P_{МД}=1;$$

[1] Этот вывод получен с учетом только одного фактора — смертности, другие же факторы, например разводы, не приняты во внимание.

[2] «Смертность и продолжительность жизни населения СССР 1926—1927. Таблицы смертности.» М.—Л., Планхозгиз, 1930, стр. 131.

$$P_{МД}=\frac{1}{3}.$$

Используя условные вероятности $P_{Д/М}$ и $P_{М/Д}$, а также теоремы сложения и умножения вероятностей, запишем:

$$P_{МД}=p_М P_{Д/М}+p_Д P_{М/Д}.$$

Допустим далее, что вероятности обоих последовательностей рождения близнецов равны. Тогда $\frac{1}{3}=2p_Д\ P_{М/Д}$ и, следовательно,

$$P_{М/Д}=\frac{1}{3\cdot 2\cdot 0,49}. \qquad (1)$$

Кроме того, мы имеем $p_М P_{М/М}+p_Д P_{Д/Д}=\frac{2}{3}$, т. е.

$$0,51\cdot P_{М/М}+0,49\cdot P_{Д/Д}=\frac{2}{3}, \qquad (2)$$

$$P_{Д/Д}+P_{М/Д}=1,$$
$$P_{Д/Д}=1-P_{М/Д}.$$

Из равенства (2) $\frac{2}{3}=0,51\,P_{М/М}+0,49(1-P_{М/Д})=0,51\,P_{М/М}+0,49-$

$-0,49P_{М/Д}.$

Следовательно,

$$P_{М/М}=\frac{\frac{2}{3}-0,49+0,49P_{М/Д}}{0,51}.$$

Используя равенство (1), получаем искомую вероятность:

$$P_{М/М}=\frac{\frac{2}{3}-0,49+0,49\cdot\frac{1}{6\cdot 0,49}}{0,51}=\frac{\frac{2}{3}-0,49+\frac{1}{6}}{0,51}\approx 0,673.$$

Использование теоремы Бейеса

Нам известна формула, выражающая теорему Бейеса[1]. Целый ряд преобразований этой формулы позволяет получить следующее выражение:

$$П=\frac{\int_0^1 x^{m+1}(1-x)^n\,dx}{\int_0^1 x^m(1-x)^n\,dx}.$$

[1] И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики, стр. 93.

где $П$ — вероятность события E после наблюдения события E';

E' — сложное событие, состоящее в m-кратном осуществлении простого события A (вероятность которого есть x), и n-кратном осуществлении противоположного события B (вероятность которого $1-x$);

E — будущее событие, т. е. случай еще одного осуществления события A с вероятностью, равной x.

Используем и применим эту формулу для таблиц смертности. Определим вероятность прожить еще b лет человеку, достигшему возраста a лет — $p_{a,b}$.

Число лиц, доживших до возраста a лет, равно l_a; число лиц, доживших до возраста $a+b$ лет из лиц l_a, равно l_{a+b}. Тогда $m=l_{a+b}$; $n=l_a-l_{a+b}$. При этом n — число умерших при переходе от возраста a к возрасту $a+b$. Получим:

$$p_{a,b}=\frac{\int_0^1 x^{l_{a+b}+1}(1-x)^{l_a-l_{a+b}}dx}{\int_0^1 x^{l_{a+b}}(1-x)^{l_a-l_{a+b}}\,dx}=\frac{l_{a+b}+1}{l_a+2}.$$

Если l_a и l_{a+b} достаточно большие числа, то

$$\lim_{n\to\infty} p_{a,b}=\lim\frac{l_{a+b}+1}{l_a+2}\approx\frac{l_{a+b}}{l_a}.$$

Знание точной формулы $p_{a,b}=\frac{l_{a+b}+1}{l_a+2}$ позволяет найти вероятную продолжительность предстоящей жизни, т. е. найти вероятное число b лет, которое проживет лицо, достигшее возраста a лет, при условии, что $p_{a,b}=0,5$ или $1/2$ (это соответствует медианному возрасту).

Извлекая из конкретной таблицы смертности число лиц, доживающих до возраста a лет (l_a), находим l_{a+b} из уравнения:

$$\frac{l_{a+b}+1}{l_a+2}=\frac{1}{2},$$

откуда

$$l_{a+b}=\frac{1}{2}\,l_a.$$

Определяя по этой же таблице половину l_a, находим тем самым l_{a+b} и, наконец, b.

Допустим, мы хотим определить вероятную продолжительность жизни мужчин города, достигших 50 лет (по данным таблицы смертности 1958—1959 гг. эта величина не вычислена).

Имеем:

$$l_{a=50}=81\,056;\quad l_{a+b}=\frac{1}{2}\,l_a=40\,528.$$

Эта величина приближенно соответствует возрасту $x = 73$ года. Находим вероятную продолжительность жизни мужчин, достигших 50 лет: 73—50 = 23 года. Для новорожденных мужчин эта величина приближенно равна 69 годам.

Наивероятнейшее число смертных случаев

Теория вероятностей позволяет определить наиболее вероятное число смертных случаев в течение года из числа n лиц данного возраста x лет. Имеется разложение бинома $(p_x + q_x)^n$, и требуется найти наибольший член. Предположим, что $m + 1$ член будет наибольшим. Тогда в соответствии с биномиальным законом распределения вероятностей[1] получаем:

$$\frac{n!\, p_x^{n-m-1}\, q_x^{m+1}}{(m+1)!\,(n-m-1)!} < \frac{n!}{m!\,(n-m)!} p_x^{n-m}\, q_x^m < \frac{n!\, p_x^{n-m+1}\, q_x^{m-1}}{(m-1)!\,(n-m+1)!}.$$

Откуда

$$\frac{p_x}{q_x} < \frac{n-m+1}{m} \text{ и } \frac{p_x}{q_x} > \frac{n-m}{m+1}.$$

Прибавляя к обеим частям неравенства по единице, находим:

$$\frac{1}{q_x} < \frac{n+1}{m} \text{ и } \frac{1}{q_x} > \frac{n+1}{m+1}.$$

Откуда $m < (n+1)\, q_x < m+1$.
Это неравенство показывает, что наивероятнейшее число смертей равно m', если оно есть ближайшее к $(n + 1)\ q_x$ целое число.

Считая n достаточно большим числом, можно использовать приближенное равенство

$$m' \approx nq_x.$$

Так, если известно, что число рождений в год в СССР составляет 4 млн. детей, а по таблице смертности населения СССР 1958—1959 гг. находим вероятность умереть на первом году жизни, равную 0,04060, то наивероятнейшее число смертей в возрасте до одного года будет:

$$m_0' \approx n_0 q_0 = 4\,000\,000 \cdot 0{,}4060 = 162\,400.$$

Аналогично может быть определено и наивероятнейшее число доживающих до одного года:

$$n_0' \approx n_0\, p_0 = 4\,000\,000 \cdot 0{,}95940 = 3\,837\,600.$$

Такой расчет m_x' по всем возрастам и последующее суммирование даст общее наивероятнейшее число умирающих $\Sigma m_x'$.

[1] И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики, стр. 95.

Использование интегральной формулы Лапласа

Используя интегральную формулу Лапласа, можно рассчитать вероятность того, что число смертей, например детских (т. е. до одного года), будет находиться в определенных заданных границах:

$$P\,(a < M < b) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{\frac{-t^2}{2}}\, dt = \frac{1}{2}\,[F\,(t=\beta) - F\,(t=\alpha)],$$

где a и b — заданные границы числа детских смертей, α и β — пределы интегрирования, связанные следующими соотношениями:

$$\alpha = \frac{a - m_0'}{\sqrt{n_0\, p_0\, q_0}}; \quad \beta = \frac{b - m_0'}{\sqrt{n_0\, p_0 q_0}}.$$

В качестве примера найдем вероятность того, что в СССР число детей, умирающих на первом году жизни, не выйдет за границы 161 400—163 000 человек. Зная, что наивероятнейшее число детей, умирающих на первом году жизни, равно 162 400, найдем

$$\alpha = \frac{161\,400 - 162\,400}{\sqrt{4\,000\,000 \cdot 0{,}04060 \cdot 0{,}95\,940}} = \frac{-1\,000}{395} \approx -2{,}53,$$

$$\beta = \frac{163\,000 - 162\,400}{\sqrt{4\,000\,000 \cdot 0{,}04060 \cdot 0{,}95\,940}} = \frac{600}{395} \approx +1{,}52.$$

Привлекая соответствующие таблицы[1], находим искомую вероятность:

$$P\,[161\,400 < M < 163\,000] = \frac{1}{2}\Big[F\,(t=1{,}52) - F\,(t=-2{,}53)\Big] =$$

$$= \frac{1}{2}\Big[F\,(t=1{,}52) + F\,(t=2{,}53)\Big] = \frac{1}{2}\,(0{,}87149 + 0{,}98859) = 0{,}93004.$$

Существенность различий вероятностных показателей, относящихся к мужчинам и женщинам

Вероятностные показатели у мужчин и женщин одного возраста в таблицах смертности отличаются друг от друга. Возникает вопрос: какие отличия вероятностей дожития или смерти у мужчин и женщин относятся к случайным колебаниям, а какие свидетельствуют о существенном различии?

Для ответа на этот вопрос можно использовать различные критерии. Воспользуемся критерием существенности В. И. Романовского. Так, с большой вероятностью в правильности утверждения о существенности различия между показателями дожития или смерти можно

[1] И. Г. Венецкий, Г. С. Кильдишев. Основы теории вероятностей и математической статистики, стр. 342—344.

утверждать, что если разность между вероятностью, например, умереть у мужчин и женщин в каком-нибудь возрасте, т. е. между $q_x^м$ и $q_x^ж$, превышает ошибку разности этих вероятностей ($\mu_{разности}$) втрое, то отличие существенно.

Таким образом, критерий Романовского выражается неравенством

$$\frac{q_x^м - q_x^ж}{\mu_{разн}} > 3,$$

где

$$\mu_{разн} = \sqrt{(\mu_x^м)^2 + (\mu_x^ж)^2}, \text{ а}$$

$$(\mu_x^м)^2 = \frac{q_x^м\, p_x^м}{n_x^м} \text{ и } (\mu_x^ж)^2 = \frac{q_x^ж\, p_x^ж}{n_x^ж}.$$

В качестве примера возьмем смертность на первом году мальчиков и девочек из таблиц смертности[1]:

$$q_0^м = 0{,}04424; \quad q_0^ж = 0{,}03677;$$
$$p_0^м = 0{,}95576; \quad p_0^ж = 0{,}96323.$$

Разность $q_0^м - q_0^ж = 0{,}04424 - 0{,}03677 = 0{,}00747$.

Ошибка показателей:

$$\mu_0^м = \sqrt{\frac{0{,}04424 \cdot 0{,}95576}{100\,000}} = \sqrt{\frac{0{,}04228}{100\,000}};$$

$$\mu_0^ж = \sqrt{\frac{0{,}03677 \cdot 0{,}96323}{100\,000}} = \sqrt{\frac{0{,}03542}{100\,000}};$$

$$\mu_{о\,(разн)} = \sqrt{\frac{0{,}04228 + 0{,}03542}{100\,000}} \approx 0{,}00090;$$

$$\frac{q_x^м - q_x^ж}{\mu_{разн}} = \frac{0{,}00747}{0{,}00090} \approx 8.$$

В соответствии с критерием Романовского можем утверждать, что разница между вероятностью умереть на первом году жизни мальчиков и девочек существенна, т. е. смертность мальчиков выше, чем девочек.

Построение вероятностных показателей таблиц инвалидности

Построение вероятностных показателей таблиц инвалидности производится аналогично показателям таблиц смертности. В настоящее время показатель смертности не может служить общей мерой оценки состояния здоровья населения, потому что при современных средствах и методах лечения смерть перестала быть неизбежным спутником многих болезней. Именно поэтому большое значение в демографии и здравоохранении должно быть придано изучению заболеваемости, физического развития и особенно инвалидности.

[1] «Итоги Всесоюзной переписи населения 1959 года. СССР», стр. 264—267

Понимая под инвалидностью частичную или полную потерю трудоспособности, можно построить таблицы инвалидности аналогично тому, как строятся таблицы смертности.

Рассмотрим специфику построения таблиц инвалидности, полагая при этом, что наиболее подходящей основой для научного изучения данной совокупности людей, расчлененной на две группы (работоспособных и инвалидов) и распределенной, кроме того, внутри каждой группы по возрасту, является вероятностная основа.

Пусть в результате единовременного статистического наблюдения по состоянию на критический момент, например начало календарного года, нам стали известны некоторые абсолютные данные, характеризующие изучаемую совокупность. Привлекая еще и дополнительные материалы текущего учета, строим сначала систему абсолютных показателей:

A — число работоспособных по состоянию на начало года;
B — число инвалидов по состоянию на начало года;
C — число новых инвалидов, появившихся за год из работоспособных;
D_1 — число смертей, зарегистрированных среди работающих;
D_2 — число смертей, зарегистрированных среди инвалидов.

На основании этих пяти показателей A, B, C, D_1, D_2 исчислим ряд вероятностей:

1) вероятность работоспособному в возрасте x лет, оставаясь работоспособным, умереть до наступления $x+1$ года:

$$q_x^{aa} = \frac{D_1}{A};$$

2) вероятность работоспособному в возрасте x лет стать инвалидом в ближайшем году (безразлично, доживет ли он до возраста $x+1$ год или умрет раньше):

$$w_x = \frac{C}{A};$$

3) вероятность работоспособному в возрасте x лет, оставаясь работоспособным, дожить до возраста $x+1$ год:

$$p_x^{aa} = \frac{A - D_1 - C}{A} = 1 - q_x^{aa} - w_x,$$

откуда вытекает, что

$$p_x^{aa} + q_x^{aa} + w_x = 1;$$

4) вероятность инвалиду возраста x лет, оставаясь инвалидом, умереть до наступления $x+1$ года (q_x^{ii}). Для исчисления этой вероятности следует иметь в виду, что умирающие инвалиды (D_2) относятся, во-первых, к совокупности инвалидов на начало года (B) и, во-вторых, к совокупности работоспособных, ставших инвалидами в течение всего

года (C). Так как новые инвалиды из числа C становились инвалидами в течение всего года, то для отнесения их к началу года (как мы это делаем при исчислении вероятностей) нужно взять только часть C.

Обозначим требуемую часть zC (где z — правильная дробь). Тогда, предполагая, что переход некоторой части здоровых в инвалиды в течение года совершается равномерно и также равномерно в течение года расщепляются и случаи смерти, получим $z=\frac{1}{2}$. Значит, из $B+zC$ инвалидов умирают в течение года D_2.

Следовательно, вероятность умереть может быть рассчитана по формуле

$$q_x^{ii}=\frac{D_2}{B+zC}=\frac{D_2}{B+\frac{1}{2}C}=\frac{2D_2}{2B+C};$$

5) вероятность инвалиду в возрасте x лет, оставаясь инвалидом, дожить до $x+1$ года:

$$p_x^{ii}=1-\frac{2D_2}{2B+C}=\frac{2B+C-2D_2}{2B+C}=\frac{2(B-D_2)+C}{2B+C};$$

6) вероятность работоспособному в возрасте x лет умереть инвалидом (q_x^{ai}) до возраста $x+1$ год. Используя формулу $q_x^{ii}=\frac{2D_2}{2B+C}$, получаем:

$$D_2=\frac{2q_x^{ii}B+q_x^{ii}C}{2}=q_x^{ii}B+\frac{1}{2}q_x^{ii}C.$$

Значит, D_2 состоит из двух частей: $q_x^{ii}B$ — число смертей инвалидов, бывших инвалидами в начале года, и $\frac{1}{2}q_x^{ii}C$ — число смертей среди инвалидов, ставших таковыми в течение этого года.

Очевидно, что это последнее число смертей будет равно числу работоспособных в начале года, умноженных на вероятность работоспособному умереть инвалидом в течение года, т. е. Aq_x^{ai}. Следовательно,

$$\frac{1}{2}q_x^{ii}C=Aq_x^{ai}.$$

Отсюда

$$q_x^{ai}=\frac{1}{2}\frac{q_x^{ii}C}{A}=\frac{1}{2}q_x^{ii}\frac{C}{A}.$$

Учитывая, что $\frac{C}{A}$ есть w_x, получаем окончательно:

$$q_x^{ai}=\frac{1}{2}q_x^{ii}w_x;$$

7) вероятность работоспособному в возрасте x лет дожить до возраста $x+1$ год, став инвалидом (p_x^{ai}). Для получения этой вероятности учтем, что из C инвалидов умирают в течение года $\frac{1}{2}Cq_x^{ii}$. Значит, доживут до конца года из новых инвалидов $C-\frac{1}{2}Cq_x^{ii}$.

С другой стороны, это же число доживающих до конца года из новых инвалидов равно Ap_x^{ai}. Получаем равенство:

$$Ap_x^{ai}=C-\frac{1}{2}Cq_x^{ii}.$$

Откуда

$$p_x^{ai}=\frac{C-\frac{1}{2}Cq_x^{ii}}{A}=\frac{C}{A}\left(1-\frac{1}{2}q_x^{ii}\right)=w_x\left(1-\frac{1}{2}q_x^{ii}\right).$$

Сложим две последние вероятности:

$$q_x^{ai}+p_x^{ai}=\frac{1}{2}q_x^{ii}w_x+w_x\left(1-\frac{1}{2}q_x^{ii}\right)=w_x\left(\frac{1}{2}q_x^{ii}+1-\frac{1}{2}q_x^{ii}\right)=w_x.$$

Следовательно,

$$q_x^{ai}+p_x^{ai}=w_x.$$

Эта формула иллюстрирует очевидную мысль, что лицо, ставшее инвалидом среди года, либо доживет инвалидом до конца года, либо умрет раньше окончания года. Сложим четыре вероятности — p_x^{aa}, q_x^{aa}, p_x^{ai}, q_x^{ai}. Сумма первых двух слагаемых равна $1-w_x$, сумма последних двух — w_x. Таким образом, общая сумма равна единице:

$$p_x^{aa}+q_x^{aa}+p_x^{ai}+q_x^{ai}=1;$$

8) вероятность работоспособному в возрасте x лет дожить до возраста $x+1$ год (без указания, в каком состоянии). Для получения этой вероятности нужно использовать теорему сложения вероятностей:

$$p_x^{a}=p_x^{aa}+p_x^{ai};$$

9) вероятность работоспособному в возрасте x лет умереть, не достигнув возраста $x+1$ год.

Также используем теорему сложения вероятностей и получим:

$$q_x^{a}=q_x^{aa}+q_x^{ai}.$$

Сумма полученных вероятностей дает:

$$p_x^{a}+q_x^{a}=p_x^{aa}+p_x^{ai}+q_x^{aa}+q_x^{ai};$$

10) вероятность лицу в возрасте x лет выбыть из числа работающих (a_x). Очевидно, что $a_x=q_x^{aa}+w_x$;

11) вероятность человеку в возрасте x лет (работоспособному или инвалиду) дожить до возраста $x+1$ (p_x). Теперь можно получить вероятности дожития, не зависящие от дополнительных условий:

$$p_x = \frac{Ap_x^a + Bp_x^{ii}}{A+B};$$

12) вероятность человеку в возрасте x лет умереть, не дожив до $x+1$ года, т. е. вероятность смерти, не зависящую от дополнительных условий:

$$q_x = \frac{Aq_x^a + Bq_x^{ii}}{A+B}.$$

Для построения таблиц инвалидности можно использовать l_x^a, т. е. число работоспособных в возрасте x лет (при этом $x > 18$).

Отличие этой таблицы от таблицы смертности состоит в том, что если в таблицах смертности единственной причиной выбытия является смерть, то в таблицах инвалидности причиной выбытия людей из числа работоспособных будет являться еще и наступление инвалидности. Поэтому более правильно называть такие таблицы таблицами смертности и инвалидности. Можно построить таблицу смертности для инвалидов, использовав l_x^i, т. е. число доживающих инвалидов.

УТОЧНЕНИЕ К ТАБЛИЦЕ СМЕРТНОСТИ ДЛЯ СТРАХОВЫХ РАСЧЕТОВ

При построении таблиц смертности для страховых целей может возникнуть вопрос, представляющий интерес и в теоретическом и практическом плане. Какой из двух совокупностей следует воспользоваться, чтобы построить таблицу смертности: совокупностью застрахованных или совокупностью медицинских осмотров, полагая, что каждый человек по крайней мере хотя бы один раз (при страховании) подвергался медицинскому осмотру?

Определим вероятность явления, в котором простые случаи не являются независимыми и одинаково возможными. Пусть число простых случаев равно n:

$$n = a_1\alpha_1 + a_2\alpha_2 + \ldots + a_r\alpha_r,$$

где $a_1\alpha_1$ распределены на a_1 групп по α_1 каждая так, что если случается один случай α_1, то случаются также и другие $\alpha_1 - 1$; $a_2\alpha_2$ случаев распределены на a_2 групп по α_2 каждая так, что если случается один случай α_2, то случаются также и другие $\alpha_2 - 1$ и т. д.

Предположим, что первые $a_1\alpha_1$ случаев одинаково возможны, последующие $a_2\alpha_2$ также одинаково возможны и т. д. Тогда вычисляются две вероятности P_1 и P_2, соответствующие $n = a_1 + a_2 + \ldots$

$+ a_r$ простых случаев и $n_r = a_1\alpha_1 + a_2\alpha_2 + \ldots + a_r\alpha_r$ простых случаев:

$$P_1 = \frac{S}{a_1 + a_2 + \ldots + a_r};$$

$$P_2 = \frac{t}{a_1\alpha_1 + a_2\alpha_2 + \ldots + a_r\alpha_r},$$

где S и t — число случаев, благоприятствующих данному явлению (в нашем примере число смертных случаев). Обозначим также через π_1 вероятности одного из случаев α_1, через π_2 вероятности одного из случаев α_2 и т. д. Тогда получим:

$$S = \pi_1 a_1 + \pi_2 a_2 + \ldots + \pi_r a_r;$$

$$t = \pi_1 a_1\alpha_1 + \pi_2 a_2\alpha_2 + \ldots + \pi_r a_r\alpha_r.$$

Откуда

$$P_1 = \frac{\pi_1 a_1 + \pi_2 a_2 + \ldots + \pi_r a_r}{a_1 + a_2 + \ldots + a_r};$$

$$P_2 = \frac{\pi_1 a_1\alpha_1 + \pi_2 a_2\alpha_2 + \ldots + \pi_r a_r\alpha_r}{a_1\alpha_1 + a_2\alpha_2 + \ldots + a_r\alpha_r}.$$

Допустим, что $\pi_1 = \pi_2 = \ldots = \pi_r$. Тогда $P_1 = P_2$, т. е. обе вероятности тождественны. В другом случае они отличаются друг от друга по распределению и составу групп. Обращаясь к случаю, который является предметом нашего рассмотрения, видим, что, так как лица, подвергающиеся многократному медицинскому осмотру, находятся в лучших условиях сопротивляемости смерти, то π будут различны, и поэтому $P_1 \neq P_2$. Если полагать, что действие закона больших чисел будет распространяться на распределение a и α в дальнейшем, то теоретически обоснованы и P_1 и P_2.

Однако практическое использование P_2 возможно только на основании предположения, что распределение лиц, подвергавшихся врачебному осмотру несколько раз, и на будущее время останется тем, какое было констатировано для материала, бывшего предметом нашего исследования. Но так как такое предположение не отвечает действительности, лучше всего для построения таблиц смертности выбрать в качестве статистической единицы человека, т. е. застрахованное лицо.

ПОСТРОЕНИЕ СОВРЕМЕННЫХ ТАБЛИЦ СМЕРТНОСТИ СИБИРИ И ДАЛЬНЕГО ВОСТОКА

В литературе последних лет встречаются описания конкретных методов и теоретических предпосылок построения таблиц смертности. Так, при построении таблиц смертности Сибири и Дальнего Востока исходным показателем был q_x, который для возрастов 0—4 года был получен путем использования данных о числе родившихся и умерших за 1954—1959 гг. Вычисление q_x в интервале 5—80 лет производилось двумя способами. Для населения крупных экономических районов:

Западной, Восточной Сибири и Дальнего Востока — показатель q_x вычислялся по формуле А. Я. Боярского

$$q_x=\frac{2(M_x+M'_x)}{S_{x-1}+2S_x+S_{x+1}+M_x+M'_x+\frac{1}{2}(M_x+M_{x+1})-\frac{1}{2}(M'_x+M'_{x-1})},$$

где M_x и M'_x — соответственно число умерших в возрасте x лет в 1958 и 1959 гг.; S_x — число лиц возраста x лет по переписи 1959 г. Для населения отдельных автономных республик, краев и областей Сибири и Дальнего Востока сначала вычислялись табличные коэффициенты смертности m_x делением чисел умерших за период таблицы на удвоенную численность населения по переписи. От табличных коэффициентов смертности был сделан переход к вероятности дожития p_x по формуле

$$\lg p_x=-m_x \lg e.$$

Таким образом, вероятности смерти q_x были определены по формуле $q_x=1-p_x$, где $p_x=e^{-m_x}$.

Значения p_x для возрастов старше 80 лет экстраполировались с помощью формулы $\lg p_x=a+bc^x$ в сочетании с графическим методом.

Отдельные показатели таблиц вычислялись по формулам:

$$d_x=l_x q_x;\quad l_{x+1}=l_x-d_x;\quad L_x=\frac{l_x+l_{x+1}}{2};$$

$$T_x=\sum L_x;\quad e_x^0=\frac{T_x}{l_x};\quad v_x=n+\frac{l_{x+n}-\frac{1}{2}l_x}{l_{x+n}-l_{x+n+1}}.$$

Глава 3 ПОКАЗАТЕЛИ ВОСПРОИЗВОДСТВА НАСЕЛЕНИЯ

ФЕРТИЛЬНОСТЬ И РОЖДАЕМОСТЬ

Режимом воспроизводства называется сочетание функции $f(x)$, измеряющей фертильность (плодовитость) каждого точного возраста, l_x — порядка вымирания и c — доли девочек среди рождающихся у матерей возраста x лет.

Начнем рассмотрение с плодовитости. Коэффициент плодовитости или, как его иногда называют, специальный коэффициент рождаемости $F=\frac{N}{S_{15-49}}$ представляет собой среднюю из повозрастных коэффициентов плодовитости. В России этот показатель был предложен академиком К. С. Веселовским, сообщившим об этом В. Я. Буняковскому[1].

Можно доказать, что между коэффициентом рождаемости и коэффициентом плодовитости нет строгой взаимной пропорциональности. Пусть в t-м году численность мужчин составит S_M, численность женщин — S_F, число родившихся — N, численность женщин в плодовитом возрасте — w; коэффициент рождаемости — n, коэффициент плодовитости — F.

Тогда

$$n=\frac{N}{S_M+S_F};\quad F=\frac{N}{w}.$$

Если бы n и F были строго пропорциональны, то результат их сравнения должен был бы оставаться постоянным в разное время и в разных странах, т. е.

$$\frac{n}{F}=\frac{N}{S_M+S_F};\ \frac{N}{w}=\frac{w}{S_M+S_F}=c_w.$$

c_w — доля женщин в плодовитом возрасте во всем населении — может быть расчленена и далее:

$$c_w=\frac{w}{S_M+S_F}\cdot\frac{S_F}{S_F}=\frac{w}{S_F}\cdot\frac{S_F}{S_M+S_F}=c_w^F c_F,$$

[1] См. В. Я. Буняковский. Исследования о возрастном составе женского православного населения России. Записки Императорской Академии Наук. СПб., 1866.

где c_w^F — доля женщин плодовитого возраста среди всех женщин, а c_F — доля женщин во всем населении.

Тогда $n = Fc_w^F c_F$.

Таким образом, величина отношения коэффициента рождаемости к коэффициенту плодовитости определяется двумя факторами: долей женщин плодовитого возраста среди всех женщин и долей женщин во всем населении.

Эти величины являются переменными. Поэтому строгой пропорциональности этих коэффициентов нет.

Если все же принять c_w^F равной $\frac{1}{2}$ и c_F тоже $\frac{1}{2}$, тогда $n \approx F \times \times \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} F$, или $F \approx 4n$.

Таким образом, коэффициент плодовитости приближенно равен учетверенному коэффициенту рождаемости.

Статистические данные позволяют думать, что в годы, когда убыль мужского населения особенно значительна (например, во время войн), процент мальчиков среди родившихся увеличивается. Если принять удельный вес мужчин во всем населении за x, а процент мальчиков среди родившихся за y и взять параллельное сопоставление x и y, а затем вычислить коэффициент корреляции между ними, то он укажет на связь между этими двумя факторами. Например, для населения Севера европейской части РСФСР за период 1897 — 1926 гг. $r_{x/y} = 0{,}5 \pm 0{,}1$, т. е. превышал свою вероятную ошибку в 5 раз и заслуживал внимания.

В чем же суть процесса повышения удельного веса мальчиков?

Пример великолепного анализа факторов, влияющих на долю мальчиков среди рождающихся, мы находим у А. А. Чупрова[1]. Сначала А. А. Чупров на основании большого статистического материала устанавливает тенденции:

с увеличением возраста зародыша внутриутробная смертность непрерывно убывает; сначала очень быстро, а затем медленнее;

смертность тех зародышей, которые рождаются преждевременно живыми или умирают в момент рождения, тоже убывает до 7-го месяца, после чего очень сильно возрастает.

Общая же картина из этих тенденций складывается так, что частость выкидышей с увеличением возраста зародышей постепенно снижается и достигнет минимума примерно на 8-м месяце беременности, после чего она снова поднимается.

Это происходит в силу того, что быстрое увеличение смертности преждевременно родившихся после 7-го месяца перекрывает уменьшение собственно внутриутробной смертности.

Рассмотрим механизм этого процесса математически.

Обозначим:

v_0 — число зачатых мальчиков на 100 зачатых девочек;
v_1 — число мальчиков на 100 девочек среди зародышей, достигших грани жизнеспособности;
$\overline{v}$ — среднее число мальчиков на 100 девочек среди выкидышей;
a — общая доля выкидышей из числа достигших грани жизнеспособности.

[1] Al. A. Tschuprow. Zur Frage des sinkenden Knabenüberschusses unter den endlich Geborenen. Institut International de Statistique, XIV Session, 1913, Raport № 22.

Тогда

$$v_0 = \frac{\dfrac{v_1}{100+v_1} + a\dfrac{\overline{v}}{100+\overline{v}}}{\dfrac{1}{100+v_1} + a\dfrac{1}{100+\overline{v}}}.$$

Используем в качестве границ величины a значения от 20 до 50% и для v_1 примем соотношение 106 : 100, это даст возможность подсчитать по указанной формуле значения v_0 при различных значениях a и $\overline{v}$ (см. табл. 45).

Таблица 45

a (в %)	v_0 при $\overline{v}=200$	v_0 при $\overline{v}=400$
20	117	128
25	120	134
33	123	141
50	130	156

Интересно отметить, что А. А. Чупров считал величину соотношения 125—130 ниже истинной. Последующие наблюдения подтвердили это мнение. Таким образом, можно считать доказанным зависимость доли мальчиков среди всех родившихся от процента выкидышей.

В тех случаях, когда при регистрации родившихся возраст матери не фиксируется (так было 30 лет назад в Италии, так и сейчас производится регистрация в ряде стран), повозрастные коэффициенты плодовитости могут быть приближенно определены по эмпирической формуле Тайта, который утверждал, что удельный вес рожающих в течение года женщин определенного возраста во всем женском контингенте данного возраста меняется пропорционально разности между данным возрастом и возрастом 50 лет:

$$n_x = ka_x(50 - t_x),$$

где n_x — число родившихся от матери возраста x лет;
a_x — число женщин данного возраста;
t_x — возраст женщин;
k — поправочный коэффициент, необходимый, чтобы вычисленное общее число родившихся равнялось действительному числу: $k = \frac{n}{a(50-\overline{x})}$ (n — число родившихся в течение года, a — общее число женщин плодовитого возраста, определяемое по переписи или другим способом, $\overline{x}$ — средний возраст женщин плодовитого периода).

БРУТТО-КОЭФФИЦИЕНТ ВОСПРОИЗВОДСТВА

Большое значение при изучении плодовитости имеет показатель F_x (аналогичен q_x из таблиц смертности), характеризующий интенсивность рождений в одногодичном интервале возраста. F_x представляет собой среднее число детей, рождаемых женщиной при переходе от возраста x до $x+1$. Можно F_x суммировать, получая среднее число детей, рождаемых женщиной в течение более длительного интервала возраста. Так, если найти сумму F_x за весь плодовитый период жизни (15—49 лет), то получим среднее число детей, рожденных женщиной за весь плодовитый период жизни. В одногодичном интервале имеем:

$$F_x = \int\limits_x^{x+1} f(x)\,dx,$$

где $f(x)$ — производная монотонной возрастающей непрерывной функции от x, указывающей на число детей, которое родит женщина, начиная с начала плодовитого возраста (15 лет) до его окончания (49 лет), dx — приращение этой функции при переходе аргумента от x до $x+1$; $f(x)$ — есть плодовитость точного возраста x, приведенная к одному году.

Сумма коэффициентов повозрастной плодовитости составит:

$$\Sigma F_x = \int\limits_{15}^{49} f(x)\,dx.$$

Эта величина называется показателем суммарной плодовитости и указывает число детей, которое воспроизводит женщина за свою жизнь при данном уровне плодовитости. Правильнее, однако, для характеристики воспроизводства населения учитывать, что женщина воспроизводит себя лишь в девочках. Поэтому нужно произвести корректировку ΣF_x путем умножения этого показателя на удельный вес девочек среди рожденных детей. Эта величина равна примерно 48,4%, или 0,484.

Следовательно, получаем «грубый» коэффициент воспроизводства, называемый брутто-коэффициентом, равный $0{,}484\Sigma\ F_x$. Обозначая этот коэффициент R_b, можно его представить следующим образом:

$$R_b \approx 0{,}49 \int\limits_{15}^{49} f(x)\,dx.$$

Допустим, что режим воспроизводства неизменен. Возникает вопрос: как будет воспроизводиться население?

При постоянном режиме воспроизводства установится определенный возрастно-половой состав населения, и при постоянной плодовитости каждого возраста рождаемость будет постоянной. Это положение впервые было отмечено Л. Эйлером, потом исследовано В. Борткевичем и в общей математической форме доказано А. Лоткой. В конечном счете это означало, что длительное повторение одних и тех же уровней

смертности и плодовитости (т. е. их постоянство) приводит к населению с неизменным составом и приростом. Полагая неизменными режим воспроизводства и плодовитость, можно сделать вывод, что порядок вымирания тоже будет постоянным.

ГИПОТЕЗЫ О НАСЕЛЕНИИ

Гипотеза Л. Эйлера

Гипотеза Эйлера (1760 г.) основана на двух предположениях, из которых второе вытекает из первого.

Первое предположение состоит в том, что число рождений из года в год возрастает по геометрической прогрессии. При постоянстве режима воспроизводства населения неизбежно и второе предположение — постоянное изменение смертности также в геометрической прогрессии. Хотя доказательства второго предположения Л. Эйлер не дает, попытаемся восполнить этот пробел.

Допустим, что первое предположение верно. Примем число рождений в календарном году равным l_0, а в предыдущем $t-1$ году $\frac{l_0}{q}$, где q — знаменатель прогрессии. Тогда в $t-2$ году число рождений было $\frac{l_0}{q^2}$, в $t-3$ году $\frac{l_0}{q^3}$ и в $t-n$ году — $\frac{l_0}{q^n}$. Примем во внимание, что из родившихся в t-м году l_0 доживает до возраста 1, 2, 3 лет и т. д. l_1, l_2, l_3 и т. д. Значит, из чисел родившихся в течение $t-1$, $t-2$ года и т. д. $\frac{l_0}{q}, \frac{l_0}{q^2}, \frac{l_0}{q^3}, \ldots, \frac{l_0}{q^n}$ достигнут возраста 1-го года $\frac{l_1}{q}, \frac{l_1}{q^2}, \frac{l_1}{q^3}, \ldots, \frac{l_1}{q^n}$; возраста 2 года — $\frac{l_2}{q}, \frac{l_2}{q^2}, \frac{l_2}{q^3}, \ldots, \frac{l_2}{q^n}$ и т. д. и возраста n лет — $\frac{l_n}{q}, \frac{l_n}{q^2}, \frac{l_n}{q^3}, \ldots, \frac{l_n}{q^n}$.

Допустим, что миграция незначительна и изменение численности населения происходит только из-за рождаемости и смертности. Общее число рождений в начале календарного года t будет:

$$L_0 = l_0 + \frac{l_1}{q} + \frac{l_2}{q^2} + \ldots + \frac{l_n}{q^n}.$$

В начале следующего года $(t+1)$ население будет состоять из родившихся в t-м году $l_0\,q$, достигших одного года из числа родившихся в t-м году l_1, оставшихся в живых из числа родившихся в $t-1$ году $\frac{l_2}{q}$, в $t-2$ году — $\frac{l_3}{q^2}$ и т. д.

Значит, в начале $t+1$ календарного года население составит:

$$L_1 = l_0\,q + l_1 + \frac{l_2}{q} + \frac{l_3}{q^2} + \ldots + \frac{l_n}{q^{n-1}}.$$

Отсюда $L_1 = L_0\,q$; $L_2 = L_1 q = L_0\,q^2$; $L_3 = L_2\,q = L_0\,q^3$ и т. д.

Числа же умерших составят: $d_0 = l_0 - l_1$ (умершие в течение t-го года в возрасте 0—1 год из числа l_0 родившихся в t-м году); $d_1 = (l_1 - l_2)\frac{1}{q}$ и т. д.

Значит, общее число умерших в t-м году составит:

$$D = l_0 - l_1 + (l_1 - l_2)\frac{1}{q} + (l_2 - l_3)\frac{1}{q^2} + \ldots = l_0 + \frac{l_1}{q} + \frac{l_2}{q^2} + \ldots - \left(l_1 + \frac{l_2}{q} + \frac{l_3}{q^2} + \ldots\right).$$

Следовательно, $D = L - (L_1 - l_0 q) = (1-q)L + l_0 q$.

В следующем $(t+1)$ году число умерших составит:

$$D_1 = (l_0 - l_1)q + (l_1 - l_2) + (l_2 - l_3)\frac{1}{q} + (l_3 - l_4)\frac{1}{q^2} + \ldots$$

Устанавливаем, что $D_1 = Dq$. Вывод: если $L_1 = Lq$, а $D_1 = Dq$, то по Эйлеру вся задача состоит в определении знаменателя прогрессии q, который может быть получен делением чисел рождений наблюдаемого года на число рождений в предыдущем году (т. е. нужны данные за два года)[1].

Зная q, можно найти d_0, d_1, d_2 и т. д., т. е. $l_1 = l_0 - d_0$; $l_2 = l_1 - d_1 q$ и т. д. Следует заметить, что недостатком гипотезы Эйлера является то, что для определения q мы берем только два года, между тем соблюдение элементарной статистической достоверности требует получения q как средней за длительный период. Но тогда гипотеза Эйлера и не нужна вовсе, ибо можно построить таблицу смертности методом смертных списков.

Значит, по Эйлеру $L_n = L_0 q^n$. Но можно предложить эмпирическую формулу $L_n = (1 + \alpha n)L_0$, где α означает коэффициент, определяемый на основании переписи.

Применение этой формулы дает лучший результат. В заключение сформулируем гипотезу Эйлера так: числа рождений, числа остающихся в живых, численность населения и т. д. возрастают в одной и той же геометрической прогрессии.

Гипотеза В. Фарра

Фарр — руководитель статистического бюро Лондона имел в своем распоряжении данные двух переписей 1841 и 1851 гг. На основании этих данных он построил таблицу смертности, предположив, что численность населения возрастала в геометрической прогрессии. При этом для каждого полученного возрастного интервала имелся свой знаменатель прогрессии. Далее Фарр распространил эту прогрессию на 3 года назад от переписи 1841 г. и на 3 года вперед от переписи

[1] q можно получить также и путем деления L_1 на L или D_1 на D. Если известны все три числа, т. е. l_0, L и D, то достаточно наблюдения в течение одного года: $D = L - (Lq - l_0 q)$; значит, $q = \frac{L-D}{L-l_0}$.

1851 г., т. е. для 1838 и 1854 гг. Имея эти числа, он определял среднюю вероятность умереть для каждой группы. Если численность населения в 1838 г. для какого-нибудь возрастного интервала принять L_0, то через x лет население будет составлять $L_x = L_0 q^x$ (где q — знаменатель геометрической прогрессии). Значит, за 17 лет (1838—1854) можно найти число лет, прожитое населением определенного возрастного интервала. Обозначив это число лет L_{17}, найдем:

$$L_{17} = \int_0^{17} L_0 q^x \, dx = \frac{L_0}{\ln q}(q^{17} - 1).$$

Вводя модуль перехода от натуральных логарифмов к десятичным, получим:

$$L_{17} = \frac{ML_0}{\lg q}(q^{17} - 1).$$

Если теперь итоги умерших за 17 лет в этой группе разделим на L_{17}, то получим среднюю вероятность умереть в данном возрастном интервале.

Далее Фарр воспользовался формулой Эдмонса для получения уже не средней вероятности для возрастного интервала, а вероятности для среднего возраста самого интервала.

Формула имеет вид для годичного интервала

$$p_a = e^{\frac{m_a}{\ln q}(1-q)},$$

где p_a — вероятность для лиц, достигших возраста a лет, прожить еще один год.

Придавая знаменателю прогрессии q и m_a соответствующие значения, Фарр получил искомые вероятности для центров возрастных интервалов. Для промежуточных же возрастов применялась интерполяция.

Оценивая этот метод определения порядка вымирания, заметим, что хотя он был точнее всех предыдущих, но сам по себе был бы еще точнее, если бы Фарр не прибегал к интерполяции, а определял вероятности, распределяя как данные переписи, так и итоги умерших по 5- и 10-летним интервалам.

Гипотеза А. Кетле

А. Кетле составил таблицу смертности на основании результатов переписи 1846 г. и итогов умерших за годы с 1848 по 1850. А. Кетле принимал число рождений в течение первого года равным N_0 (эта величина приурочена им к началу календарного года, данные по которому известны); $N_1, N_2, N_3, \ldots, N_n$ — числа лиц, достигших возраста 1, 2, 3, ..., n лет; $d_0, d_1, d_2, \ldots, d_n$ — числа умерших в соответствующем возрасте.

Полагая известными N_0, d_0, $d_1, \ldots, d_n$, можно определить вероятности умереть в течение следующего года: $\frac{d_0}{N_0}$, $\frac{d_1}{N_1}$, $\frac{d_2}{N_2}$ и т. д.

Предполагая, что в течение целого столетия закон смертности не меняется, и исходя из числа рожденных N_0, А. Кетле определяет число доживающих до возраста 1-го года:

$$l_1 = N_0\left(1-\frac{d_0}{N_0}\right) = N_0 - d_0, \text{ а далее } l_2 = l_1\left(1-\frac{d_1}{N_1}\right),$$

$$l_3 = l_2\left(1-\frac{d_2}{N_2}\right) \text{ и т. д.}$$

Умножив найденные числа на 10 000, он получил свою таблицу смертности. Из порядка вымирания А. Кетле следует, что число родившихся равно:

$$d_0 + \frac{l_1}{N_1}d_1 + \frac{l_2}{N_2}d_2 + \ldots + \frac{l_n}{N_n}d_n.$$

Если допустим, что $\frac{l_1}{N_1} = \frac{l_2}{N_2} = \ldots = \frac{l_n}{N_n}$ (это соответствует неподвижному населению), то получается галлеевская формула. Недостаток формулы вымирания А. Кетле в том, что она базируется на предположениях, не соответствующих действительности: 1) умершие в течение одного календарного года состоят из родившихся в одном календарном году; 2) родившиеся в одном календарном году умершие в одном возрасте состоят из умерших в одном календарном году; 3) смертность неизменна во времени.

Гипотеза стационарного населения и построение коэффициентов смертности стационарного населения

Гипотеза стационарного населения, необходимая для того, чтобы исключить влияние изменения числа родившихся и умерших, введенная в науку Э. Галлеем, является весьма плодотворной. Она легла в основу многих теорий научного обоснования демографических процессов и систематизации некоторых демографических показателей. С помощью этой гипотезы создается возможность отражения закономерностей, присущих населению, взятому в определенное время и в определенном месте.

Основные черты стационарного населения: 1) плотность рождений $N(t)$ постоянна. Это значит, что рождаемость распределена равномерно во времени; 2) порядок вымирания совокупности новорожденных (l_0) неизменен; 3) миграции не наблюдается.

На демографической сетке (см. рис. 24) видно, что при этих допущениях все элементарные совокупности умерших у разных поколений равны друг другу. Так, $t_{10}a_3a_4 = t_9a_2a_3$. Вследствие этого совокупность умерших I рода равна совокупности умерших III рода: $t_9t_{10}a_3a_4 = t_9t_{10}a_2a_3$ (имея в виду, что элементарная совокупность $t_9a_3t_{10}$ у них общая). Это означает, что при построении таблиц смертности можно ограничиться данными смертности за один год. Кроме того, числа современников при этой гипотезе означают также числа живущих.

В годичном интервале от x до $x+1$ живет население $N\int_x^{x+1} l(x)\,dx = NL_x$. А всего, следовательно, живет $N\Sigma L_x$. Это и есть число живущих. Используя эту гипотезу, найдем средний возраст умерших путем деления суммы лет, прожитых каждой возрастной группой умерших, на число умерших:

$$\frac{\frac{1}{2}d_0 + \frac{3}{2}d_1 + \ldots}{d_0 + d_1 + \ldots} = \frac{T_0}{l_0}.$$

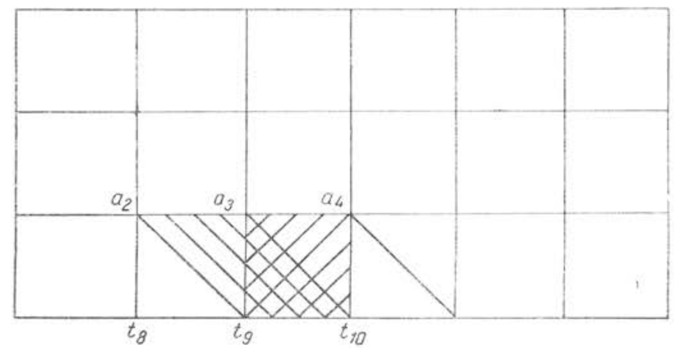

Рис. 24. Элементарные совокупности умерших разных поколений.

А это и есть средняя продолжительность жизни, т. е. e_0^0. При наших допущениях $e_0^0 = \frac{\Sigma L_x}{\Sigma d_x}$, т. е. средняя продолжительность жизни стационарного населения равна этому населению, деленному на число рождений (или число умерших).

Коэффициент смертности стационарного населения будет равен:

$$\frac{d_0 + d_1 + d_2 + \ldots}{\Sigma L_x} = \frac{\Sigma d_x}{\Sigma L_x} = \frac{1}{e_0^0}.$$

Следовательно, для получения общего коэффициента смертности стационарного населения нужно взять сумму умерших во всех возрастах и разделить на сумму живущих во всех возрастах.

Как видно, этот коэффициент зависит от силы смертности всех возрастных групп, но не зависит от конкретного возрастного состава совокупности, для которой исчислена таблица смертности. Поэтому его сравнительная ценность и практическое применение весьма велики.

Из вышеизложенного следует важный вывод, что вследствие равенства числа родившихся числу умерших при этой гипотезе коэффициент рождаемости равен коэффициенту смертности и, следовательно:

$$m_S = n_S = \frac{1}{e_0^0}.$$

Отсюда $e_0^0 = \frac{1}{n_S} = \frac{1}{m_S}$.

Это позволяет сделать следующие выводы о стационарном населении:

численность всего населения и в каждом возрастном интервале постоянна;

возрастной состав населения постоянен с долей каждого возраста, равной $c_x = \frac{L_x}{e_0^0 N}$;

коэффициенты рождаемости и смертности равны друг другу и равны обратному значению средней продолжительности жизни.

При этой гипотезе общая численность населения есть, следовательно, число лет, прожитых всем поколением от рождения до полного вымирания.

Полученная нами величина коэффициента смертности, равная $\frac{1}{e_0^0}$, есть мера общей интенсивности смертности населения. Термин «табличный коэффициент смертности» тождествен термину «коэффициент смертности стационарного населения».

Данный коэффициент смертности лучше и правильнее отражает действительный размер смертности, чем обычный коэффициент смертности, построенный по формуле $m = \frac{M}{S}$.

Для установления возрастного состава стационарного населения суммируют L_x в пределах возрастов 0—14; 15—49 и 50 и старше. Так как общая сумма $\Sigma L_x = e_0^0$, то получаем:

$$\sum_0^{14} L_x = L_0 + L_1 + L_2 + \ldots + L_{14};$$

$$\sum_{15}^{\omega-1} L_x = e_0^0 - \sum_0^{14} L_x = \sum_{15}^{49} L_x + \sum_{50}^{\omega-1} L_x = e_{15}^0\, l_{15};$$

$$\sum_{50}^{\omega-1} L_x = e_{50}^0\, l_{50}.$$

Отсюда

$$\sum_{15}^{49} L_x = \sum_{15}^{\omega-1} L_x - \sum_{50}^{\omega-1} L_x = e_{15}^0\, l_{15} - e_{50}^0\, l_{50}.$$

Возьмем таблицу смертности СССР 1958—1959 гг. оба пола[1]: $e_0^0 = 68{,}59$. Суммируем и получаем: $\sum_0^{14} L_x = 14{,}15$;

$$\sum_{15}^{49} L_x + \sum_{50}^{\omega-1} L_x = e_{15}^0\, l_{15} = 58{,}22 \cdot 0{,}93497 \approx 54{,}44;$$

$$\sum_{50}^{\omega-1} L_x = e_{50}^0\, l_{50} = 27{,}11 \cdot 0{,}84502 \approx 22{,}91.$$

Значит,

$$\sum_{15}^{49} L_x = 54{,}44 - 22{,}91 = 31{,}53.$$

Получим процентное соотношение трех возрастных групп:

Возрастные группы	0—14	15—49	50 и выше
Удельный вес в стационарном населении в %	$\frac{14{,}15}{68{,}59} \cdot 100 = 20{,}6$	$\frac{31{,}53}{68{,}59} \cdot 100 = 46{,}0$	$\frac{22{,}91}{68{,}59} \cdot 100 = 33{,}4$

С возрастным составом стационарного населения сравнивают структуру реального населения: если детей от 14 лет по данным переписи будет больше, чем 20,6%, то такой возрастной состав называют прогрессивным, а если меньше — регрессивным.

Как же строятся коэффициенты смертности стационарного населения?

Мы уже видели, что общий коэффициент смертности сильно зависит от возрастного состава населения. Проблема построения показателей смертности, которые *полностью* устраняют все различия возрастных структур сравниваемых совокупностей, решается, как мы уже видели, стандартизацией коэффициентов смертности.

Коэффициент смертности стационарного населения — это тоже стандартизованный коэффициент, но за стандарт принят возрастной состав стационарного населения, который, в свою очередь, определяется соотношением смертности в разных возрастах, поэтому можно считать, что в нем не отражено влияние никакой структуры. Коэффициенты смертности стационарного населения являются очень ценными с точки зрения возможностей их использования.

Пусть имеются повозрастные коэффициенты смертности $m_0, m_{x1}, m_{x2}, \ldots, m_{xn}$.

Найдем в условиях заданной смертности число доживающих до возраста i лет из $l_a = 100\,000$. Используя связь между силой смертности μ_x и числом доживающих l_x, имеем:

$$\mu_x\, dx = -\frac{d\,[l\,(x)]}{l\,(x)} \text{ или } \mu_x dx = -d \ln [l\,(x)].$$

[1] См. «Итоги Всесоюзной переписи населения 1959 года. СССР», стр. 262—263

Интегрируя в пределах от 0 до x_i лет, получаем:

$$-\int_0^{x_i} \mu(x)\,dx = \int_0^{x_i} \frac{l'(x)}{l(x)}\,dx = [\ln l(x)]_0^{x_i} =$$

$$= \ln l(x_i) - \ln(l_0) = \ln \frac{l(x_i)}{l_0}.$$

Отсюда

$$e^{-\int_0^{x_i} \mu(x)\,dx} = \frac{l(x_i)}{l_0}$$

и, следовательно,

$$l(x_i) = l_0 e^{-\int_0^{x_i} \mu(x)\,dx}.$$

Если теперь разбить интервал от 0 до i на x_1, x_2, x_3 и т. д., то полученный интеграл можно представить как сумму интегралов

$$\int_0^{x_i} \mu(x)\,dx = \int_0^{x_1} \mu(x)\,dx + \int_{x_1}^{x_2} \mu(x)\,dx + \ldots + \int_{x_{i-1}}^{x_i} \mu(x)\,dx.$$

Далее, следуя рекомендации В. В. Паевского[1], выберем такой очень малый промежуток интегрирования от κ до $\kappa+1$, внутри которого можно принять плотность смертей величиной постоянной. Тогда получим:

$$l(x_i) = l_0 e^{-\sum_0^{i-1} \mu(x_k) z_k},$$

где z_k — длина k-го промежутка в годах.

Принимая $m_{x_k} = \mu_{x_k}$ в силу постоянства смертности внутри интервала, установим, что это справедливо в случае, если m_{x_k} есть средняя взвешенная из всех значений, которые принимает сила смертности на этом интервале. Тогда

$$l(x_i) = l_0 e^{-\sum_0^{i-1} m_{x_k} z_k}.$$

Логарифмируя, получаем:

$$\ln [l(x_i)] = \ln l_0 - \sum_0^{i-1} m_{x_k} z_k.$$

[1] В. В. Паевский. О построении коэффициента смертности неподвижного населения. «Бюллетень Ленинградского областного отдела статистики», 1928, № 19.

Если $l_0 = 100\,000$, то $\ln l_0 = 9{,}21034$; тогда $\ln l(x_i) = 9{,}21034 - \sum_0^{i-1} m_{x_k} z_k$, где $\sum_0^{i-1} m_{x_k} z_k$ есть сумма последовательных повозрастных коэффициентов смертности. Каждый коэффициент смертности берется слагаемым столько раз, сколько лет включено в этот интервал. Так, одногодичные коэффициенты берутся слагаемым один раз, двухлетние два раза и т. д. Если возрастные интервалы берутся 0—1; 1—4, а далее по пятилетиям, то $\sum_0^{i-1} m_{x_k} z_k$ получается суммированием коэффициента смертности интервала 0—1, умноженного на единицу; коэффициента смертности интервала 1—4, умноженного на 4; коэффициентов смертности пятилетних интервалов, умноженных на 5.

После нахождения $l(x_i)$ можно найти коэффициент смертности стационарного населения по формуле

$$m_S = \frac{1}{\overset{\circ}{e}_0} = \frac{l_0}{T_0},$$

где T_0 — общая численность стационарного населения, или число человеко-лет, прожитых всем населением, $T_0 = \int_0^{\omega} l(x)dx$.

Дальнейшая задача состоит в определении T_0. Примем $L_0 = l_0 + \frac{2}{3} d_0$. С другой стороны, $T_0 - T_1 = L_0$.

Разобьем весь период 0 — ω на интервалы 0—1; 1—5; 5—10; 10—15 и т. д. до 90—95 лет и для каждого интервала заменим истинные значения $l(x)$ приближенным значением, полученным из параболы. Итак, имеем вычисленные $l_0, l_1, l_5, l_{10}, l_{15}, \ldots, l_{95}$.

В каждом десятилетнем интервале строим параболу 2-го порядка. Так, на интервале 0—5 лет, принимая уравнение параболы $y = a_0 + a_1 x + a_2 x^2$ и полагая, что она проходит через точки l_1, l_5, l_{10}, имеем:

$$T_1 - T_5 = \int_1^5 l(x)\,dx = \int_1^5 (a_0 + a_1 x + a_2 x^2)\,dx.$$

Интегрируя и производя алгебраические действия, получаем:

$$T_1 - T_5 = 2{,}533 l_5 - 1{,}704 l_1 - 0{,}237 l_{10}.$$

Запишем формулу в общем виде для нахождения разности любого десятилетнего интервала, если парабола проходит через точки l_i, l_{i+5}, l_{i+10}:

$$T_i - T_{i+10} = \int_i^{i+10} l(x)\,dx.$$

После преобразований получим:

$$T_i - T_{i+10} = \frac{5}{3}(l_i + 4l_{i+5} + l_{i+10}).$$

Для последнего значения T_{95} примем $T_{95} \approx \frac{2}{3}\, l_{95}$. Выразим T_0 как сумму всех разностей интервалов: $T_0 = T_0 - T_1 + T_1 - T_5 + \ldots + T_{85} - T_{95} + T_{95}$.
Обозначим:

$$l_5 + l_{15} + l_{25} + l_{35} + \ldots + l_{95} = A,$$
$$l_{10} + l_{20} + l_{30} + l_{40} + \ldots + l_{90} = B,$$
$$10l_0 + 26l_5 = C,$$
$$10l_1 - l_{10} = D.$$

Сделав подстановку и произведя алгебраические действия, получим $T_0 = 3{,}333\,(A + 2B - 0{,}1C + 0{,}071D)$. Теперь находим коэффициент смертности стационарного населения $m_S = \frac{100\,000}{T_0}$ и среднюю продолжительность жизни $e_0^0 = \frac{1}{m_S}$.

Таким образом, для практического использования данного метода при построении таблиц смертности следует:

1) использовать коэффициенты смертности возрастных групп 0 — 1; 1—5; 5—10; 10—15 и т. д. до 90—95 для нахождения их суммы $(\Sigma m_{x_k} z_k)$;

2) найти логарифмы чисел доживающих: $\ln[l(x_i)] = 9{,}21034 - \sum_{0}^{i-1} m_{x_k} z_k$;

3) найти числа доживающих (для 0 лет это 100 000) до возрастов 1; 5; 10 и т. д. до 95;

4) использовав указанные выше формулы, найти T_0, а затем коэффициент смертности стационарного населения m_S.

При построении показателей таблиц смертности, используя числа доживающих, получают числа умерших $(l_{x+5} - l_x = d_{x/x+5})$ и вероятность дожить $\left(p_{x/x+5} = \frac{l_{x+5}}{l_x}\right)$. Затем находят вероятности умереть: $q_{x/x+5} = 1 - p_{x/x+5}$.

По мнению Л. Е. Дарского[1], гипотезу стационарного населения можно привлечь к изучению весьма широкого класса социальных явлений. При этом можно представить следующую схему исследования.

На определенный момент t произведена группировка населения по нескольким признакам: социальному, полу, возрасту, семейному

[1] См. Л. Е. Дарский. О применении модели стационарного населения в социологических исследованиях. В сб. «Тезисы докладов, представленных на совещание «Количественные методы в социальных исследованиях» (Сухуми, апрель 1967 г.) М., НИИ ЦСУ СССР, Отдел демографии, 1967.

положению, месту жительства и др. Исчисляются вероятности перехода населения в течение небольшого периода времени (скажем, одного года) из одного состояния (группы) в другое. Полученные вероятности p_i составляют квадратную матрицу, являющуюся характеристикой изучаемого социального процесса.

Полагая население замкнутым, а вероятности перехода действующими неопределенно длительное время, придем к выводу, что в соответствии с теорией цепей Маркова данный социальный процесс в конце концов приведет к тому, что независимо от начальной структуры населения конечная структура будет определяться только матрицей перехода. Так, можно определить семейную структуру стационарного населения, конечную структуру рабочей силы, меняющейся вследствие текучести и др.

Гипотеза стабильного населения

Гипотеза стабильного населения, введенная в науку В. Борткевичем и развитая А. Лоткой, состоит в предположении, что численность некоторого населения хотя и меняется, но при неизменных темпах, в геометрической прогрессии — по формуле сложных процентов. Что вытекает из этого?

1. Коэффициент естественного прироста, называемый для стабильного населения истинным коэффициентом естественного прироста[1] и обозначаемый r, должен быть постоянным, а значит, коэффициенты рождаемости и смертности тоже постоянны.

2. Постоянным коэффициент рождаемости может быть только при возрастающей тоже в геометрической прогрессии плотности рождений. Значит, для того, чтобы коэффициент рождаемости был постоянен при населении, возрастающем в геометрической прогрессии, нужно, чтобы плотность рождений возрастала тоже в геометрической прогрессии.

Следовательно, стабильное население характеризуется постоянным порядком вымирания и ростом плотности рождений в геометрической прогрессии. Если принять, что порядок вымирания населения постоянен и плотность рождений $[N(t)]$ меняется в геометрической прогрессии со знаменателем e^{rt}, то можно установить, что плотность рождений в момент t будет равна $\Pi_t = N_0 e^{rt}$, а в момент $t - x$ $\Pi_{t-x} = N_0 e^{r(t-x)}$.

Из людей, родившихся в момент $t - x$, в возрасте x лет в момент t живет часть, равная l_x. Значит, численность населения в возрасте x в момент t равна $N_0 e^{r(t-x)}\, l(x)dx$.

Общая же численность населения в момент t составит:

$$\int_0^{\omega} N_0 e^{r(t-x)}\, l(x)\, dx = N_0 e^{rt} \int_0^{\omega} e^{-rx}\, l(x)\, dx.$$

[1] В зарубежной демографии этот коэффициент имеет специальное название — интринсинк.

Подынтегральное выражение не зависит от t, и при постоянном r в любой момент t численность населения будет определяться множителем e^{rt}, т. е. меняться в геометрической прогрессии так же, как и плотность рождений.

Удельный вес лиц в возрасте x лет в стабильном населении составляет:

$$c_x = \frac{N_0 e^{r(t-x)} \ l(x)\,dx}{N_0 e^{rt} \int\limits_0^{\omega} e^{-rx} l(x)\,dx} = \frac{e^{-rx} \ l(x)\,dx}{\int\limits_0^{\omega} e^{-rx} l(x)\,dx}.$$

Каково же отличие стабильного населения от стационарного?

Возрастному составу стационарного населения графически соответствует кривая линия $c_x = l(x)$, а для стабильного же населения такой кривой линией будет

$$c_x = l(x)\,e^{-rx} = \frac{l(x)}{e^{rx}}.$$

Если $e^r > 1$, то население растет, а в этом случае в стабильном населении по сравнению со стационарным будет больше молодых возрастов. Если же $e^r < 1$, в стабильном населении по сравнению со стационарным будет меньше доля молодых и больше доля старых. Если $e^r = 1$, имеем $r = 0$, т. е. *при отсутствии прироста стабильное население превращается в стационарное.*

При рассмотренных величинах плотности рождения и численности населения коэффициент рождаемости стабильного населения в момент t получается равным:

$$K_{\text{р}\,(t)} = \frac{N_0 e^{rt}}{N_0 e^{rt} \int\limits_0^{\omega} e^{-rx} \ l(x)\,dx} = \frac{1}{\int\limits_0^{\omega} e^{-rx} l(x)\,dx} \approx$$

$$\approx \frac{r}{1-e^{-r}} \cdot \frac{1}{\sum e^{-rx} \ L_x}.$$

Так как истинный коэффициент естественного прироста, т. е. коэффициент естественного прироста стабильного населения, $r = (\ln e^{rx})'$, то можно получить коэффициент смертности:

$$K_{\text{с}}(t) = \frac{1}{\int\limits_0^{\omega} e^{-rx} l(x)\,dx} - r \approx \frac{r}{1-e^{-r}} \cdot \frac{1}{\sum e^{-rx} L_x} - r.$$

Исходя из стабильности населения и принимая тем самым в качестве гипотезы неизменный порядок вымирания и коэффициенты рождаемости, можно найти средние темпы роста этого населения в геометрической прогрессии, а затем сопоставить с фактическими. Таким образом, можно убедиться в том, насколько фактическое положение дел отлично от гипотетического.

Так, численность населения СССР по оценке ЦСУ на 1 января 1950 г. составляла 178,5 млн. человек, а на 1 января 1968 г.—236,7 млн. человек[1]. Определим среднегодовой темп роста стабильного населения:

$$e^r = \sqrt[18]{\frac{236{,}7}{178{,}5}} = \sqrt[18]{1{,}32605}.$$

Тогда $r = \frac{1}{18} \ln 1{,}32605 \approx 0{,}016$, или $16‰$. В действительности же коэффициент естественного прироста в 1967 г. составлял $9{,}8‰$[2].

ИНТЕГРАЛЬНОЕ УРАВНЕНИЕ И НЕТТО-КОЭФФИЦИЕНТ ВОСПРОИЗВОДСТВА

Интегральное уравнение воспроизводства населения

Продолжая развитие гипотезы стабильного населения, примем начальную численность стабильного женского населения N_0 и предположим, что через промежуток времени, равный t, его численность вырастет до $N(t)$. Получаем:

$$B(t) = N_0 e^{rt}.$$

Этот прирост стабильного населения может быть связан с нетто-коэффициентом воспроизводства R_0 (средним числом рожденных девочек одним женским поколением с учетом смертности). Если длину поколения обозначить T, то начальная численность стабильного населения N_0 через T лет составит $N_0 e^{rT}$. Тогда

$$N(t) = N_0 e^{rT} = N_0 R_0,$$

откуда

$$r = \frac{\ln R_0}{T}.$$

Для определения длины поколения используем снова модель стабильного населения. Возьмем $B(t)$ — число рожденных девочек за год, а число женщин в возрасте x лет получаем умножением чисел рожденных x лет назад на вероятность дожития до x лет, т. е. на p_x. Тогда число девочек, рожденных этими женщинами, составит:

$$B(t-x)\,p_x f(x)\,dx.$$

Интегрируя в пределах плодовитого возраста, получаем:

$$B(t) = \int\limits_{15}^{49} B(t-x)\,p_x f(x)\,dx.$$

---

[1] «Народное хозяйство СССР в 1967 г. Статистический ежегодник». М., «Статистика», 1968, стр. 7.

[2] Там же, стр. 36.

Это уравнение называется интегральным уравнением воспроизводства населения.

Решение этого уравнения определяет характер искомой величины, которая колеблется так же, как и численность женского населения. Однако эти колебания носят затухающий характер, и через некоторое время число рождений и численность населения в значительной степени приблизятся к тем, которые обусловлены соответствующим темпом постоянного прироста, показателем которого также является и коэффициент естественного прироста r. Подставляя вместо значения $B(t)$ равную ему величину N_0e^{rt} и вместо $B(t-x)$ равную ему $N_0e^{r(t-x)}$, имеем:

$$N_0 e^{rt} = \int_{15}^{49} N_0 e^{r(t-x)} p_x f(x)\, dx = N_0 e^{rt} \int_{15}^{49} e^{-rx} p_x f(x)\, dx$$

Отсюда

$$1 = \int_{15}^{49} e^{-rx} p_x f(x)\, dx.$$

У этого уравнения, которое называется характеристическим, бесконечно много комплексных корней и лишь один действительный корень, который одновременно является и коэффициентом естественного прироста.

Характеристическое уравнение можно решить разложением в строку Маклорена. Однако в данном случае результат является весьма приближенным:

$$e^{-rx} \approx 1 - rx + \frac{r^2x^2}{2!} - \ldots$$

Тогда

$$1 = \int_{15}^{49} \left(1 - rx + \frac{r^2x^2}{2!}\right) p_x f(x)\, dx =$$

$$= \int_{15}^{49} x^0 r^0 p_x f(x)\, dx - \int_{15}^{49} x r p_x f(x)\, dx + \frac{1}{2} \int_{15}^{49} x^2 r^2 p_x f(x)\, dx.$$

Обозначим $\int_{15}^{49} x^n p_x f(x)\, dx = R_n$. Делаем подстановку и получаем:

$$1 = R_0 - rR_1 + \frac{r^2}{2} R_2 \approx R_0 \left(1 - r\,\frac{R_1}{R_0} + \frac{r^2}{2} \cdot \frac{R_2}{R_0}\right).$$

Обозначим

$$\frac{R_1}{R_0} = T; \quad \left(\frac{R_1}{R_0}\right)^2 - \frac{R_2}{R_0} = \beta.$$

Используя логарифмы и их разложение в ряд, а также не учитывая степени r выше второй, можно выразить r в зависимости от R_0 и T.

Так, по формуле А. Лотки венгерские демографы Барши и Тейс получили

$$r = \frac{\ln R_0}{T}.$$

Видим, что

$$T = \frac{\int_{15}^{49} x\, p_x f(x)\, dx}{\int_{15}^{49} p_x f(x)\, dx}.$$

Значит, *длина поколения есть средняя продолжительность жизни матерей, взвешенная по стационарному населению*. Аналогичный расчет можно произвести для определения мужского поколения.

Возьмем теперь нетто-коэффициент воспроизводства, представляющий собой $R_0 = \int_{15}^{49} p_x f(x) dx$ среднее число девочек, рожденных одной женщиной за всю ее жизнь. R_0 отличается от R_b, т. е. от брутто-коэффициента воспроизводства, тем, что в R_0 учитывается смертность девочек.

Вычисление нетто-коэффициента воспроизводства суммированием

Можно ли произвести вычисление R_0, ограничиваясь суммированием?

Возьмем плодовитость в возрасте от x до $x+1$ (без учета мальчиков) и умножим на число живущих женщин L_x, тогда получим:

$$R_0 = \int_{15}^{49} p_x f(x)\, dx = \sum F_x L_x.$$

А. Я. Боярский предложил для непосредственного вычисления другой прием, предполагающий опробование различных значений r и получение двух значений r_1 и r_2. При подстановке в характеристическое уравнение величины r_1 результат меньше единицы, а при подстановке величины r_2 — больше единицы. Тогда путем интерполяции можно найти действительную величину r, лежащую между r_1 и r_2. Покажем расчет на примере женского населения Голландии[1], взятого по пятилетним интервалам (см. табл. 46), сохраняя символику автора.

Увеличиваем Φ_x, т. е. брутто-коэффициент воспроизводства, равный 286,2, в 5 раз и получаем 1,431, а нетто-коэффициент воспроизводства равен 1,298. Учитывая, что интервал пятилетний, имеем:

[1] А. Я. Боярский. Курс демографической статистики. М., Госпланиздат, 1945, стр. 88.

Таблица 46

Расчет нетто-коэффициента воспроизводства

Возраст x	Плодовитость (в ‰) $\overline{F}_x$	Брутто-коэффициент воспроизводства, деленный на 5, $\overline{\Phi}_x$	Сумма чисел живущих $\Sigma L'_x$	Нетто-коэффициент воспроизводства (в ‰) $\overline{\Phi}_x \Sigma L'_x$
15—19	12,3	6,0	4,67	28,020
20—24	91,6	44,3	4,63	205,109
25—29	165,6	80,2	4,58	367,316
30—34	153,6	74,3	4,53	336,579
35—39	111,1	53,8	4,46	239,948
40—44	51,7	25,0	4,38	109,500
45—49	5,3	2,6	4,28	11,128
Итого . . .	591,2	286,2	—	1 297,600

$$\int_{15}^{49} e^{-rx} p_x f(x)\,dx = \overline{p_x f(x)} \int_{x}^{x+5} e^{-rx}\,dx = \overline{p_x f(x)}\, e^{-rx}\, \frac{1-e^{-5r}}{r}.$$

Тогда для вычисления r произведем несколько подстановок в характеристическое уравнение и найдем $r_1 = 0{,}010$ и $r_2 = 0{,}008$. Подставляем значения r_1 и r_2 в выражение $\int_{15}^{49} e^{-rx} p_x f(x)\,dx$. Предварительно рассчитаем $\overline{\Phi}_x \overline{L}'_x$, а затем

1) $\sum \overline{\Phi}_x \overline{L}'_x e^{-0{,}010x} = 196{,}323$ ‰;

2) $\sum \overline{\Phi}_x \overline{L}'_x e^{-0{,}008x} = 207{,}321$ ‰.

Множитель $\frac{1-e^{-5r}}{r}$ при $r_1 = 0{,}010$ составит 4,88 и при $r_2 = 0{,}008$ составит 4,90. Значит, интеграл равен: для r_1 $0{,}196\,323 \times 4{,}88 = 0{,}960$; для r_2 $0{,}207\,321 \cdot 4{,}90 = 1{,}016$.

Интерполируя, находим:

$$r = 0{,}008 + (0{,}010 - 0{,}008) \cdot \frac{1{,}016 - 1}{1{,}016 - 0{,}960} = 0{,}0086, \text{ или } 8{,}6‰.$$

Асимптотическое образование возрастной структуры

Пользуясь интегральным уравнением воспроизводства населения, можно очень легко доказать утверждение Эйлера о том, что при длительном проявлении определенного уровня смертности и плодовитости, т. е. при постоянном режиме воспроизводства, установится постоянное, не зависящее от времени, возрастное распределение. Имеем:

$$B(t) = \int_{15}^{49} B(t-x)\, p_x f(x)\, dx.$$

Так как число рождений в стабильном населении изменяется в соответствии с корнем характеристического уравнения, т. е. величиной истинного коэффициента естественного прироста, то

$$B(t) = B(t-x)\, e^{rx}.$$

Пусть $c\,(x;\ t)$ есть доля лиц точного возраста x лет $(x + dx)$ во всем женском населении в момент t:

$$c(x;\ t) = \frac{B(t-x)\, p_x}{N(t)}.$$

Значит,

$$\int_{15}^{49} c(x,\ t)\, dx = 1.$$

Тогда

$$c(x,\ t) = \frac{B(t)}{N(t)} e^{-rx} p_x = b(t)\, e^{-rx} p_x,$$

где $b(t)$ — удельный вес рождений девочек во всем населении. Откуда

$$1 = b(t) \int_{15}^{49} e^{-rx} p_x\, dx; \quad b(t) = \frac{1}{\int_{15}^{49} e^{-rx} p_x dx}.$$

В правой части уже нет величины t. Значит, рождаемость в стабильном населении не зависит от времени, т. е. $b(t) = b$ и $c(x;\ t)$, равное $be^{-rx}p_x$, также не зависит от t.

Отсюда следует, что распределение по возрасту в стабильном населении не зависит от времени.

Сформулируем окончательный вывод: *возрастной состав стабильного населения образуется асимптотически, независимо от начального состояния.*

Таким образом, интегральное уравнение позволяет доказать, что любая возрастная структура населения при длительном сохранении неизменного порядка вымирания и неизменной повозрастной плодовитости стремится в пределе принять совершенно определенную возрастную структуру, остающуюся в дальнейшем без изменения.

Величина нетто-коэффициента воспроизводства указывает на характер воспроизводства: при $R_0 > 1$ мы имеем расширенное воспроизводство, а при $R_0 < 1$ (при сохранении этих соотношений) — перспективу неизбежной убыли, т. е. депопуляции. Исчисление нетто- и бруттс-коэффициентов воспроизводства вытекает из положений, разработанных Беком в 1884 г. Эти коэффициенты широко применялись и развивались Р. Кучинским и А. Лоткой и независимо от них советским статистиком Г. Н. Баткисом.

Таким образом, теория стабильного населения позволяет ввести в анализ теоретически обоснованные характеристики, для которых исходными являются современная плодовитость и порядок вымирания населения без учета миграций, при неизменной половой структуре рождающихся.

Значит, в качестве наилучших демографических показателей, с точки зрения теории стабильного населения, следует принять отношение рождений двух последовательных женских поколений, среднюю длительность одного поколения и предельные коэффициенты рождаемости и смертности, вытекающие из этой теории.

Очевидно, можно поставить вопрос об оптимальном типе воспроизводства населения, как это делает в советской демографической литературе А. Я. Кваша. «Идеальные» пропорции между изменяющейся численностью населения, с одной стороны, и уровнем экономического развития, а также научно-технического прогресса, — с другой, определяют оптимальные темпы воспроизводства населения.

При расчете «идеальных» пропорций и оптимальных темпов воспроизводства начинают использовать жизненный уровень (фактический и планируемый), сознательное материнство и отцовство, рациональное использование трудовых ресурсов общества и т. д. В ряде случаев поиски приводят к необходимости использования конкретных гипотез.

В качестве примера можно привести эмпирическую формулу венгерских демографов Р. Андорка и К. Мильтеньи, позволяющую представить показатель оптимального типа воспроизводства (R) как функцию нескольких переменных: фактической возрастной структуры населения и повозрастных душевых доходов. Данная связь была определена на базе теории стабильного населения и некоторых предложений видного французского демографа Ж. Буржуа-Пиша.

$$R=\sqrt{\frac{pv}{qj}},$$

где p и q — удельные веса иждивенцев соответственно младших и старших возрастов во всем стационарном населении, а j и v — отношение расходов на содержание одного иждивенца соответственно в младшем и старшем возрастах к расходам на содержание одного лица в трудоспособном возрасте.

Расчеты, произведенные А. Я. Квашой для населения СССР с использованием таблиц смертности 1959 г., позволили определить возрастную структуру стационарного населения и привели к оптимальной величине нетто-коэффициента воспроизводства, равной приблизительно 1,1.

СТАТИСТИЧЕСКОЕ ИСЧИСЛЕНИЕ ДЛИНЫ ПОКОЛЕНИЙ

В предыдущем изложении мы уже употребляли выражение «длина поколения», обозначаемое T. Разберем это понятие более конкретно.

Вопрос о смене поколений и о длине каждого поколения интересует демографов и социологов с точки зрения преемственности поколений, с точки зрения того, что передадут отцы детям. Как обстоит дело с воспроизводством населения? Воспроизводит ли поколение живущих себя? Такие показатели-индексы, как брутто- и нетто-коэффициенты воспроизводства, характеризующие население в отношении воспроизводства, можно расценивать в зависимости от того, как мы будем измерять поколение. Через сколько лет происходит смена одного поколения другим?

Очень важным для анализа демографических показателей является нахождение длины интервала поколения и как важнейшего результата — периода времени, в течение которого следуют друг за другом поколения. В широком смысле слова длина поколения есть промежуток времени, отделяющий одноименные события в жизни родителей и детей (браки, рождения, смерти и т. д.).

Способы измерения длины поколения

Большое внимание вопросу длины поколения уделял советский демограф Я. С. Улицкий[1].

Длина поколения может измеряться разными способами.

Прямой путь — определение разности между возрастом родителей и возрастом детей:

1) между средним возрастом родителей — $\bar{x}_{\text{род}}$ и средним возрастом детей — $\bar{x}_{\text{дет}}$:

$$D_1=\bar{x}_{\text{род}}-\bar{x}_{\text{дет}}\,^2;$$

2) между средним возрастом родителей и возрастом среднего ребенка:

$$D_2=\bar{x}_{\text{род}}-\bar{x}_{\text{ср. реб}};$$

3) между средним возрастом родителей и возрастом старшего (или младшего) ребенка:

$$D_3=\bar{x}_{\text{род}}-\bar{x}_{\text{ст. реб}};$$

$$D_4=\bar{x}_{\text{род}}-\bar{x}_{\text{мл. реб}};$$

[1] См. Я. С. Улицкий. Демографическое понятие поколения. В сб. «Проблемы демографической статистики». М., Госстатиздат, 1959.

[2] Целесообразно применять для обозначения длины поколения вообще знак D и оставить знак T только для длины поколения, исчисленной демографическим методом.

4) между отцом и сыновьями (всеми, старшими, средними или младшими) — мужская линия;

5) между матерью и дочерьми (всеми, старшей, средней или младшей) — женская линия.

В большом масштабе такие измерения не производятся. Для построения модели семьи нужны данные о средней длине поколений, продолжительности жизни отцов, возрасте сына к моменту смерти отца, суммарной продолжительности жизни отцов и жизни сыновей при живом отце, продолжительности жизни сыновей после смерти отца, продолжительности одновременной жизни дедов и внуков (возраст внука к моменту смерти деда) и др.

Кроме прямых методов расчета длины поколения возможны и косвенные, исследующие теоретически мыслимое (гипотетическое) поколение.

Эти методы расчета длины поколения аналогичны косвенному методу построения таблиц смертности (основаны на предположении, что показатели смертности в год переписи во всех возрастах относятся к одному поколению родившихся во все годы его жизни).

Принимая гипотезу стационарности поколения, можно для определенного года вычислить длину поколения и распространить ее на более или менее длительный период. Длина поколения зависит от времени вступления мужчины в брак и продолжительности периода брачной плодовитости. Здесь переплетаются биологические и социально-экономические факторы.

Многие французские и немецкие исследователи (Фурье, Бийо, Рюмелин, Ваше, Тюркан, Де-Фовиль) использовали для выражения длины поколений эмпирическую формулу:

$$D_{\text{пок}} = \bar{x}_{\text{в момент вступления в брак}} + \frac{D_{\text{бр. плод}}}{2},$$

где $D_{\text{бр. плод}}$ — период брачной плодовитости.

На наш взгляд, при этом допускались неточности. Чтобы получить приведенную выше формулу длины поколений путем теоретического рассуждения, мы ввели некоторые допущения, создающие погрешности. Первое допущение состоит в том, что длительность периода брачной плодовитости принимается равной разности между возрастом отца при рождении последнего и первого ребенка:

$$D_{\text{бр. плод}} = \bar{x}_{\text{отца (послед. реб)}} - \bar{x}_{\text{отца (перв. реб)}}.$$

Если для простоты вывода допустить, что в семье рождается только двое детей (первый и последний), то получим:

$$\bar{x}_{\text{отца (ср. реб)}} = \frac{\bar{x}_{\text{отца (послед. реб)}} + \bar{x}_{\text{отца (перв. реб)}}}{2} =$$

$$= \bar{x}_{\text{отца (перв. реб)}} + \frac{\bar{x}_{\text{отца (послед. реб)}} - \bar{x}_{\text{отца (перв. реб)}}}{2} =$$

$$= \bar{x}_{\text{отца (перв. реб)}} + \frac{D_{\text{бр. плод}}}{2}.$$

Следовательно, только в том случае, если за возраст в момент вступления мужчины в брак принимается средний возраст отца в момент рождения первого ребенка, получается данная формула, т. е. когда $\bar{x}_{\text{в момент вступления в брак}} = \bar{x}_{\text{отца (перв. реб)}}$.

Вторая неточность состоит в том, что мы для простоты вывода принимали наличие в семье только двух детей, а на самом деле все обстоит совсем не так и большое значение имеет то обстоятельство, что распределение рождений детей в периоде брачной плодовитости неравномерно. В первую половину, разумеется, рождения чаще.

Третья неточность при подсчете брачной плодовитости состоит в допущении выживания всех детей. На самом деле выживают не все.

Фурье определил длину поколения в 33,3 года. Бийо в период Великой французской революции определял длину поколения (по матери) и получил 28,3 года. Если учитывать разницу между средними возрастами мужчин и женщин, вступающих в брак, равную 5 годам, то получим 28,3 + 5 = 33,3, т. е. ту же величину.

Рюмелин получал для ряда стран этот показатель в границах от 32 до 39 лет. Метод Ваше был иным. Он определял длину поколения по супружеской паре, находя средний возраст родителей в момент рождения каждого ребенка. Определяя число рождений на каждый год супружеской жизни и число прожитых супругами лет, Ваше находил средний возраст брака. Далее, вычисляя средний возраст родителей при рождении среднего ребенка, Ваше получил длину поколения 33,1 года.

Демографический метод измерения длины поколения

Кроме весьма приближенных прямых и косвенных методов расчета длины поколения имеется способ, дающий точное значение величины поколений — демографический метод. Этот метод, как мы уже частично видели, основан на связи между коэффициентом естественного прироста K и нетто-коэффициентом воспроизводства:

$$R_0 = \int_{15}^{49} e^{-Kx} r(x)\,dx = 1.$$

R_0 — это величина, характеризующая отношение численности воспроизведенного населения к воспроизводящему в течение одного цикла. Это обстоятельство позволяет определить величину цикла, т. е. длину женского поколения, понимая под этим средний возраст матери при рождении дочери. Так как корреляция между возрастом матери и полом ребенка весьма незначительна, то рождение дочери можно заменить рождением вообще. Мы знаем, что

$$K = (\ln S)' = \frac{S'}{S}.$$

Обозначая, как и ранее, длину поколения $D_{\text{пок}}$, предположим, что население меняется в геометрической прогрессии со знаменателем e^K:

$$S_1 = S_0 e^K.$$

Значит, через $D_{\text{пок}}$ лет население увеличится в $e^{KD_{\text{пок}}}$ раз. Между тем за каждые $D_{\text{пок}}$ лет население должно увеличиться в R_0 раз, т. е. $e^{KD_{\text{пок}}} = R_0$ и $D_{\text{пок}} = \frac{\ln R_0}{K}$. При рассмотрении коэффициента естественного прироста стабильного населения мы пришли к такому же выводу. Ранее (см. стр. 167—168) мы получили $R_0 = 1{,}298$, а $K = 0{,}008$. Получаем:

$$D_{\text{пок}} = \frac{\ln 1{,}298}{0{,}008} = 30{,}5 \text{ года.}$$

Длину поколения, вычисленную таким образом, называют длиной демографического поколения и обозначают буквой T.

Можно трактовать длину поколения и иначе. Из формулы $e^{KD_{\text{пок}}} = R_0$ вытекает, что $R_0 e^{-KD_{\text{пок}}} = 1$. Тогда

$$\int_{15}^{49} e^{-Kx} r(x)\,dx = R_0 e^{-KD_{\text{пок}}} = e^{-KD_{\text{пок}}} \int_{15}^{49} r(x)\,dx.$$

Подводя под знак интеграла, получаем:

$$\int_{15}^{49} e^{-Kx} r(x)\,dx = \int_{15}^{49} e^{-KD_{\text{пок}}} r(x)\,dx.$$

Теперь можно заключить, что $D_{\text{пок}}$ есть средний возраст матери x, где весами являются $r(x)$, т. е. число дочерей, которое родят при данной повозрастной фертильности женщины возраста x лет при гипотезе стационарного населения.

Интересный расчет длины поколения предложили американские демографы Л. Дэблин и А. Лотка[1]. Они считали длиной поколения средний возраст всех матерей, родивших в данном году ребенка.

В табл. 47 приведено распределение рождений девочек, а возраст матерей дан по пятилетиям. В каждой строке приведены средние годовые данные, поэтому в итоге произведено умножение на пять.

Значит, 100 000 женщин за всю жизнь рожают 142 045 девочек, или 142 девочки на 100 женщин. Но из этих 142 045 не все выжили. Для расчета выживших мы вводим графы 3 и 4. В графе 4 после умножения 23 352 на 5 получаем 116 760. Значит, 100 000 женщин оставляют после себя 116 760 женщин следующего поколения.

[1] L. Dublin and A. Lotka. On the True Rate of Natural Increase as Exemplified by the Population of the United States, 1920. — «Journal of the American Statistical Assotiation», 1925, p. 305.

Таблица 47

Плодовитость женщин США по данным переписи 1920 г. (число рождений девочек на каждые 100 000 женщин)

Возрастные группы матерей	Среднее годовое число рождений девочек на один возрастной год m_a	Вероятность дожития до данного возраста p_a	$m_a p_a$	Накопление частоты по графе 2	Накопление частоты по графе 4	x (середина интервала)	x'	$x' m_a$
1	2	3	4	5	6	7	8	9
10—14	9	0,88567	8	9	8	12,5	—3	—27
15—19	2 202	0,87438	1 950	2 211	1 958	17,5	—2	—4 404
20—24	7 310	0,85509	6 251	9 521	8 209	22,5	—1	—7 310
25—29	7 481	0,82960	6 206	17 002	14 415	27,5	0	0
30—34	5 780	0,80181	4 634	22 782	19 049	32,5	+1	5 780
35—39	3 898	0,77417	3 018	26 680	22 067	37,5	+2	7 796
40—44	1 552	0,74664	1 159	28 232	23 226	42,5	+3	4 656
45—49	172	0,71610	123	28 404	23 349	47,5	+4	688
50—54	5	0,62937	3	28 409	23 352	52,5	+5	25
Всего	28 409·5= =142 045	—	23 352·5= =116 760	—	—	—	—	—11 741 +18 945 +7 204

Таким образом, нетто-коэффициент воспроизводства оказался равным 1,168. Найдем средний возраст матери, используя данные колонки 7, 8, 9, и получим:

$$D_{\text{пок}} = \bar{x}' K + x_0 = \frac{7\,204}{28\,409} \cdot 5 + 27{,}5 \approx 1{,}3 + 27{,}5 = 28{,}8 \text{ года.}$$

Исчисление среднего возраста всех матерей, родивших в данном году девочку, можно заменить для упрощения расчетом медианы. Но для этой цели можно использовать в качестве весов данные графы 2, а более правильно графы 4. Тогда в первом случае, введя графу 5 накопленных частот, получаем:

$$D_{\text{пок}} = 25 + 5\,\frac{14\,205 - 9\,521}{7\,481} \approx 28{,}1 \text{ года.}$$

Во втором случае по накопленным частотам графы 6 получаем:

$$D_{\text{пок}} = 25 + 5\,\frac{11\,676 - 8\,209}{6\,206} \approx 27{,}8 \text{ года.}$$

Все найденные нами величины 28,8; 28,1 и 27,8 года показывают длину поколения по женской линии.

Если, например, $R_0 = 1{,}3$, то это значит, что численность следующего поколения будет на 30% больше, чем в предыдущем.

Знавая нетто-коэффициент воспроизводства, можно, следовательно, решать и обратные задачи, т. е. определять, каким должен быть ежегодный прирост, чтобы обеспечить такое возрастание. Если принять длину поколения заданной и равной, допустим, 30 годам, то для ориентировочного ответа можно использовать среднюю арифметическую

$$\overline{K}_{\text{прироста}}=\frac{R_0-1}{D_{\text{пок}}}\cdot 1000$$

и получить $\overline{K}_{\text{прироста}}=\frac{(1,3-1)\cdot 100}{30}=\frac{0,3}{30}\cdot 1000=10^0/_{00}$.

Более точным, а значит, и более правильным следует признать вычисление по средней геометрической

$\overline{K}_{\text{роста}}=\sqrt[D_{\text{пок}}]{R_0}$ (в долях единицы), а $\overline{K}_{\text{прироста}}=\left(\sqrt[D_{\text{пок}}]{R_0}-1\right)\cdot 1000$ (в промилле).

Имеем:

$$\overline{K}_{\text{роста}}=\sqrt[30]{1,3}.$$

Логарифмируя, получим:

$$\lg\overline{K}_{\text{роста}}=\frac{1}{30}\lg 1,3=\frac{0,11394}{30}=0,00380.$$

Следовательно,

$$\overline{K}_{\text{роста}}\approx 1,009.$$

Отсюда

$$\overline{K}_{\text{прироста}}=(1,009-1)\cdot 1000=9^0/_{00}.$$

ДИВЕРГЕНЦИЯ КОЭФФИЦИЕНТОВ ЕСТЕСТВЕННОГО ПРИРОСТА СТАБИЛЬНОГО НАСЕЛЕНИЯ

Разность между коэффициентами естественного прироста мужчин и женщин называют дивергенцией коэффициентов. Если исходить из гипотезы стабильного населения, включающего оба пола, то это означает, что, кроме смертности и плодовитости мужчин и женщин, в расчет следует принять неизменное превышение рождений мальчиков над девочками.

Модель стабильного населения для обоих полов была разработана П. Кармелом[1]. Для доказательства того, что коэффициенты естественного прироста мужчин и женщин в стабильном населении согласуются друг с другом, Кармел вводит два критерия стабильного населения:

1) мужской и женский нетто-коэффициенты воспроизводства должны давать одинаковые показатели естественного прироста;

2) возрастная плодовитость мужчин и женщин должна проявляться так, чтобы число рождений было одинаковым.

Наличие указанных выше условий приводит к согласованности показателей коэффициентов естественного прироста мужчин и женщин стабильного населения.

Допустим, что между возрастом матерей и возрастом отцов имеется разница в 5 лет. Пусть число рожденных девочек B_0, тогда число рожденных на x лет раньше $B(-x)=B_0e^{-r_Fx}$. Число женщин в возрасте x лет может быть исчислено по формуле

$$P_F(x)=B_0\,e^{-r_Fx}\,p_F(x).$$

Численность мужчин, но уже в возрасте $x+5$ лет можно получить, введя показатель m, дающий среднее число мальчиков, приходящихся на одну рожденную девочку:

$$P_M(x+5)=mB_0\,e^{-r_F(x+5)}\,p_M(x+5).$$

Разделив численность мужчин на численность женщин, в результате преобразований получаем:

$$\frac{P_M(x+5)}{P_F(x)}=\frac{mp_M(x+5)}{p_F(x)}\,e^{-5r_F}.$$

Обозначим дробь

$$\frac{mp_M(x+5)}{p_F(x)}=D(x),$$

где $D(x)$ есть отношение численностей мужского и женского населения в отдельных возрастах в условиях стабильности. Сравнивая величину $D(x)$ и e^{-5r_F}, можно найти отклонение фактического населения от стабильного.

Далее Кармел сравнивает нетто-коэффициент воспроизводства мужского и женского населений при условии, что число рожденных мальчиков женщинами в возрасте x лет равно числу мужчин в возрасте $x+5$ лет.

$$mP_F(x)f_F(x)=P_M(x+5)f_M(x+5);$$

$$R_{0\,(M)}=\int p_M(x)f_M(x)\,dx;$$

$$f_M(x)=\frac{mP_F(x)f_F(x)}{P_M(x+5)}.$$

В результате подстановки и различных преобразований получим:

$$\frac{R_{0\,(M)}}{R_0(F)}=\frac{\int\frac{1}{D(x)}\,p_F(x)f_F(x)\,dx}{\int p_F(x)f_F(x)\,dx}.$$

Как мы увидим дальше, относительный компонент нетто-коэффициента воспроизводства по возрасту равен $w(x)$. Значит, то что мы получили, представляет собой величину, обратную средней гармонической из показателей $D(x)$, где весами служат указанные компоненты. Таким образом, $\frac{R_{0(M)}}{R_{0(F)}}=\frac{1}{\overline{D(x)}}$. Теперь можно установить, что

$$r_F=\frac{\ln R_{0\,(F)}}{T}\quad\text{и}\quad r_M=\frac{\ln R_{0\,(F)}-\ln\overline{D(x)}}{T+5}.$$

[1] P. H. Karmel. An Analysis of the Sources and Magnitudes of Inconsistencies between Male and Female Net Reproduction Rates in Actual Populations, — «Population Studies», 1948, 2, pp. 240—273.

Следовательно,

$$r_M - r_F = \frac{-\ln \overline{D(x)} - 5r_F}{T+5}.$$

Допуская, что r общее есть средняя арифметическая, т. е. $r = \frac{r_M + r_F}{2}$, окончательно получим:

$$r_M - r_F = \frac{-\ln \overline{D(x)} - 5r}{T+2,5}.$$

Найденное нами расхождение между коэффициентами естественного прироста стабильного населения мужчин и женщин и представляет дивергенцию этих коэффициентов. Это расхождение происходит вследствие отклонения удельного веса возрастов женщин во всем населении от стабильного состояния. Дивергенция рассматриваемых коэффициентов обуславливается рядом причин:

1) различным влиянием внешних факторов на мужское и женское население (войн, эмиграции и т. д.);
2) динамикой численности прежних рождений;
3) изменением смертности мужчин и женщин;
4) изменением мужского населения в течение прошлого периода.

Исследования, проводившиеся Кармелом по статистическим данным многих стран, показали, что 3-й и 4-й факторы очень слабо влияют на дивергенцию. Наибольшее влияние оказывают 1-й и 2-й факторы. Влияние войн и использование для этой цели возрастного распределения мужчин и женщин является предметом самостоятельного рассмотрения. Остановимся на действии второго фактора на дивергенцию коэффициентов естественного прироста стабильного населения.

Рассмотрим величину $D(x)$, исходя из предположения, соответствующего действительности, о сохранении показателя смертности и превышении рождаемости мальчиков. Как нетрудно заметить, $D(x)$ в случае стабильного населения будет зависеть главным образом от расхождений, имевших место в численности прежних рождений:

$$D(x) = \frac{B(-x-5)}{B(-x)}.$$

Возьмем обратную величину

$$v(x) = \frac{B(-x)}{B(-x-5)}.$$

$v(x)$ — отношение числа рождений в отдельной возрастной группе к числу рождений за прошедшие 5 лет. Величина $v(x) < 1$ показывает, что число рождений по сравнению с прошлым периодом сократилось.

Тогда отношение двух нетто-коэффициентов воспроизводства даст:

$$\frac{R_{0(M)}}{R_{0(F)}} = v = \int v(x)\, w(x)\, dx.$$

Разность же коэффициентов естественного прироста

$$r_M - r_F = \frac{\ln v(x) - 5r}{T+2,5}.$$

Произведенные расчеты дают основание утверждать, что основной причиной дивергенции коэффициентов естественного прироста является первый фактор, т. е. войны. Влияние же второго фактора значительно слабее, но с ним нужно считаться при анализе воспроизводства населения.

Дальнейшее рассмотрение истинного коэффициента естественного прироста населения показывает, что его колебания для двух полов определяется еще и тем, какой пол имеет абсолютное превосходство при вступлении в брак.

В соответствии с моделью Интема, коэффициент фактической брачности устанавливается следующим образом: исчисляется количество вступивших в брак женщин и мужчин в возрасте x и y лет $M\ (x,\ y)$ в период времени t; на основании величин $N_M\ (x,\ t)$ и соответственно $N_F\ (x,\ t)$, используя среднюю арифметическую, устанавливают коэффициент фактической брачности:

$$M(x,\ y,\ t) = 2v_a(x,\ y)\,[aN_M(y,\ t) - (1-a)\,N_F(x,\ t)],$$

где a — показатель брачного превосходства мужчин по сравнению с женщинами, получаемый при помощи формулы $\frac{a}{1-a} = \frac{v_M(y)}{v(x)}$; $v_M(y)$ и $v(x)$ — коэффициенты брачности мужчин и женщин в возрасте y и x лет; $v_a\ (x,\ y)$ — коэффициент брачности возрастной комбинации x, y, уточненной при помощи веса a.

В результате использования этой формулы показатель воспроизводства всего населения получается в виде одной величины[1].

КОЭФФИЦИЕНТЫ ЗАМЕЩЕНИЯ

Разработанные Беком, Кучинским и Лоткой построения брутто- и нетто-коэффициентов воспроизводства предназначены, по сути дела, для ответа на вопросы, в какой мере фактическая плодовитость обеспечивает дальнейший рост численности населения. Гарантирует ли современная плодовитость рост численности населения или ее уровня только достаточно для стабилизации численности? И наконец, может быть этой плодовитости недостаточно для поддержания численности населения на определенном уровне? Размер же прироста или уменьшения населения, явившихся результатом недостаточной рождаемости, брутто- и нетто-коэффициенты воспроизводства не дают. Между тем определение коэффициентов действительного прироста является конечной целью изучения воспроизводства. Для решения этого вопроса и привлекается теория стабильного населения. Для ответа на указан-

[1] Модель Интема носит несколько искусственный характер. Дело в том, что параметр a не имеет реального демографического смысла.

ные выше вопросы, а также для получения действительной картины динамики населения может быть рекомендован целый ряд методов, основной задачей которых является получение истинных коэффициентов рождаемости, смертности и естественного прироста, характеризующих фактический, а не зависящий от возрастной структуры уровень этих показателей.

Обратимся к коэффициенту замещения старого поколения новым (replacement rate). Воспользуемся для этого гипотезой стационарного населения и плодовитостью в ее самой простой форме.

Потребуем от коэффициента замещения ответа на коренной вопрос, аналогичный тем, которые ставили Кучинский и Лотка перед брутто- и нетто-коэффициентами воспроизводства. А именно, достаточен ли уровень рождаемости для пополнения состава имеющегося населения?

Коэффициент замещения по соотношению удельных весов детей

Одним из вариантов коэффициента замещения ($K_з$), частично дающим ответ на поставленный нами вопрос, является такой коэффициент замещения, который может быть рассчитан дифференцированно по полу и представляет собой отношение удельного веса детей моложе определенного возраста в плодовитом контингенте (мужчин и женщин) по данным переписи — D_1 к удельному весу детей того же возраста по данным стационарного населения (из таблиц смертности) — D_2:

$$K_{з\,(1)} = \frac{D_1}{D_2}.$$

Можно под «определенным возрастом» понимать 5 лет или 7 лет, т. е. дошкольный возраст. Этот показатель носит приближенный характер, но может применяться в тех случаях, когда отсутствуют необходимые данные о плодовитости.

Недостатком данного коэффициента замещения является то, что в нем рассматривается воспроизводство населения за несколько лет до переписи. При этом картина получается искаженной в силу неполноты учета при переписи самых молодых возрастов.

Коэффициент замещения по содержанию показателей рождаемости

Коэффициент замещения может основываться также на использовании «грубого» коэффициента рождаемости календарного года и коэффициента смертности стационарного населения, взятого из таблиц смертности. Если учесть, что в стационарном населении коэффициент рождаемости равен коэффициенту смертности и обратному значению средней продолжительности жизни новорожденных, то можно сопоставить показатели рождаемости:

$$K_{з\,(2)} = \frac{n_{\text{фактический}}}{n_{\text{стационарн}}} = \frac{n_{\text{фактический}}}{\frac{1}{\mathring{e}_0}} = n_{\text{фактический}} \cdot \mathring{e}_0.$$

Недостаток этого коэффициента замещения заключается в сравнении неоднородных показателей.

Коэффициент замещения Бургдёрфера

Третий вариант коэффициента замещения под влиянием идей немецкого демографа Бека был разработан Бургдёрфером. Он заключается в получении «очищенного» баланса населения. «Очищенный» коэффициент рождаемости получается путем стандартизации плодовитости женщин по стационарному населению. Из него вычитается «истинный» коэффициент смертности. Разность этих показателей дает «очищенный» естественный прирост, который правильно рассматривать лишь в качестве показателя замещения.

Рассмотрим сущность этого метода, его практическое использование в работе Б. Ц. Урланиса[1] при выяснении вопроса о величине естественного прироста населения в Западной Европе и его дальнейшее развитие.

Для использования этого метода необходимо:

1) предварительное исчисление таблиц смертности и получение из них средней продолжительности жизни (e_0^0);

2) знание удельных весов: женщин плодовитого возраста в стационарном населении c_{15-49}^S; женщин плодовитого возраста во всем населении c_{15-49}; замужних женщин в контингенте плодовитого возраста двух периодов (например, до войны и после нее или до первой мировой войны и накануне второй);

3) численность населения на 1 января и число родившихся за год.

Из таблиц смертности Б. Ц. Урланис берет $\overline{e_0^0}$, т. е. средневзвешенную среднюю продолжительность жизни в Западной Европе (весами являются численности населения). В 30-х годах для мужчин $e_0^0 = 53{,}8$ года, а для женщин $e_0^0 = 57{,}0$ лет, в среднем $\overline{e_0^0} = 55{,}5$.

Для получения удельного веса женщин плодовитого возраста в стационарном населении нужно разделить численность женщин плодовитого возраста на численность стационарного населения. Числитель находим из таблиц смертности суммированием L_x в пределах 15—49 лет. По итальянской таблице смертности 1930—1932 гг. получаем, что $\sum\limits_{x=15}^{x=49} L_x = 2\,681\,044$. Общая же численность стационарного населения получена умножением числа родившихся мальчиков и девочек на среднюю продолжительность жизни и последующим суммированием. При гипотезе стационарного населения рождается 100 000 мальчиков и 100 000 девочек, а пропорция среди фактически родившихся мальчиков и девочек 106 и 100, получаем:

$$S_{\text{стационарн}} = 100\,000 \cdot 56{,}0 + 106\,000 \cdot 53{,}8 \approx 11\,303\,000.$$

[1] См. Б. Ц. Урланис. Рост населения в Европе. М., Огиз-Госполитиздат, 1941, стр. 306—312.

Теперь определяем долю женщин плодовитого возраста в стационарном населении Италии: $c^s_{15-49} = \frac{2\,681\,044}{11\,303\,000} = 0{,}237$.

Ввиду отсутствия единой таблицы смертности всей Западной Европы Б. Ц. Урланис использует таблицы смертности нескольких стран и получает для Западной Европы эту величину, равную 0,239.

Удельный вес женщин плодовитого возраста во всем населении находят по переписным данным. В целом по Западной Европе Б. Ц. Урланис получил $C_{15-49} = 0{,}267$. Удельный вес замужних женщин в этом контингенте необходим, по мысли автора этого метода, для обоснования влияния «временных причин», т. е. снижения брачности из-за войны. Однако, как справедливо указывает Б. Ц. Урланис, снижение удельного веса замужних женщин может быть результатом не только войны, но и действия социально-экономических причин. Этот и ряд других мотивов позволяют обойтись без этой поправки.

Определив численность населения Западной Европы на 1 января 1939 г. в 398,6 млн. человек, а количество рождений за год 8011 тыс. человек, получаем фактический коэффициент рождаемости $20{,}2^0/_{00}$. Теперь можно исчислить «истинные» коэффициенты рождаемости и смертности, т. е. перейти от $20{,}2^0/_{00}$ — «грубого» коэффициента фактической рождаемости — к «истинному» n'. Для этого нужно найти поправочный коэффициент, равный отношению доли женщин плодовитого возраста в стационарном населении (0,239) к фактической доле тех же женщин (0,267): $Q = \frac{0{,}239}{0{,}267} = 0{,}895$.

Определим исправленный коэффициент рождаемости:

$$n' = nQ = 20{,}2 \cdot 0{,}895 = 18{,}1^0/_{00}.$$

Найдем исправленный коэффициент смертности. Фактический коэффициент смертности $m = 13{,}6^0/_{00}$, а для стационарного населения $m' = \frac{1}{e^0_0} = \frac{1}{55{,}5} = 0{,}018$, или $18^0/_{00}$.

Таким образом, разница между «грубыми» коэффициентами рождаемости ($20{,}2^0/_{00}$) и смертности ($13{,}6^0/_{00}$), равная $6{,}6^0/_{00}$, показывала меньшую остроту демографического кризиса ($18{,}1-18{,}0=0{,}1$).

ВЗВЕШИВАНИЕ КОЭФФИЦИЕНТОВ ЗАМЕЩЕНИЯ

Можно использовать коэффициент замещения в узком смысле слова как отношение фактического числа рождений к предельному числу. Под предельным числом рождений будем понимать такое число, которое необходимо для поддержания неизменного состава населения плодовитых возрастов. Если $\varphi(x)$ — среднегодовое число рожденных девочек женщинами в возрасте x лет, то, как известно, брутто-коэффициент воспроизводства будет равен:

$$R_b = \int_{15}^{49} \varphi(x)\, dx.$$

Если $P(x)$ — фактическое среднегодовое число женщин возраста x лет, а p_x — вероятность дожития, то невзвешенный коэффициент замещения будет равен:

$$K_{\text{з. общ}} = \frac{\int_{15}^{49} P(x)\,\varphi(x)\,dx}{\dfrac{\int_{15}^{49} P(x)\,dx}{\int_{15}^{49} p(x)\,dx}}.$$

В числителе годовое число рожденных девочек, в знаменателе предельное число рождений, рассчитанное на соответствующий год.

Этот коэффициент можно и взвешивать. Тогда после преобразований получим:

$$K_{\text{з. общ}} = \frac{\int_{15}^{49} P(x)\,\varphi(x)\,dx}{\frac{1}{R_b}\int_{15}^{49} \frac{P(x)\,\varphi(x)\,dx}{p(x)}} = R_b\,\overline{p(x)}.$$

Итак, взвешенный коэффициент замещения получается умножением брутто-коэффициента воспроизводства на среднюю гармоническую вероятность дожития, взвешенную по числу рождений, приходящихся на отдельные возрасты. Некоторые примеры применения этих показателей для анализа демографических процессов можно найти в работах венгерских демографов[1].

СОВЕРШЕНСТВОВАНИЕ МОДЕЛИ СТАБИЛЬНОГО НАСЕЛЕНИЯ НА ОСНОВЕ АНАЛИЗА БРАЧНОСТИ

Метод «женихов и невест»

Характеризуя динамику брачности с помощью средних возрастов вступающих в брак, используют среднюю разность возрастов вступающих в брак или лучше разность средних возрастов вступающих в брак.

Возьмем комбинационную таблицу повозрастных сочетаний вступающих в брак, обозначим возраст жениха x, возраст невесты y.

Очевидно, что средний возраст жениха $(\overline{x_i})$ каждой группы есть середина (центр) возрастных интервалов женихов, а средний возраст

[1] См. Д. Барши и Э. Тейс. Расчеты воспроизводства населения на основе показателей замещения поколений и модели стабильного населения. В сб. «Методы демографических исследований». М., «Статистика», 1969.

невест $(\bar{y}_j)$ соответствующей i-й группы женихов можно получить по формуле

$$\bar{y}_i = \frac{\sum y_j m_i}{\sum m_i},$$

где y_j — середина j-го возрастного интервала невест,

m_i — число браков в i-м интервале.

Тогда, например, для центрального возраста первого возрастного интервала женихов, т. е. $\bar{x}_1$ лет, можно вычислить соответствующий ему средний возраст невест

$$\bar{y}_2 = \frac{\sum y_j m_1}{\sum m_1}.$$

Для центрального возраста второго интервала женихов, т. е. $\bar{x}_2$ лет, средний возраст невест будет найден по формуле

$$\bar{y}_2 = \frac{\sum y_j m_2}{\sum m_2} \text{ и т. д.}$$

В качестве примера рассмотрим табл. 48[1].

Надо полагать, что существует весьма тесная связь между возрастом жениха и невесты и разностью их средних возрастов и что эта разность средних возрастов может быть представлена как функция среднего возраста жениха или невесты. В качестве функции, апроксимирующей искомую, можно принять корреляционное уравнение связи по прямой линии или параболе второго порядка.

Многие демографы считают разность средних возрастов весьма устойчивой во времени и заключенной в узких пределах для разных народностей. Тем самым это единообразие разности средних возрастов признается величиной, не зависимой от социально-экономических и территориальных факторов. На наш взгляд, такое мнение ошибочно. Однако фактическими данными для одной и той же территории, но в разные периоды во времени мы не располагаем. Можно лишь для сравнения привести результат наших расчетов разности между средним возрастом женихов и невест, распределенных по среднему возрасту жениха, по данным, относящимся к населению Советского Союза в 1966 г., и результат аналогичных расчетов, произведенных В. В. Паевским в 1922 г. для Ленинграда (другие данные у нас отсутствуют).

Покажем расчет среднего возраста невест первой группы женихов, т. е. тех, возраст которых моложе 20 лет. Условно примем за значение этого интервала средний возраст этой группы для женихов 19 лет, а для невест 18 лет, а для последующих групп 22,5; 27,5, 32,5 и т. д.

---
[1] «Вестник статистики» № 1 за 1967 г., стр. 94.

Таблица 48

Распределение браков в СССР по возрасту жениха и невесты за 1966 г.

Возраст жениха (x_i) / Возраст невесты (y_j)	Всего	В том числе в возрасте (лет): моложе 20 $\bar{x}_1=19,0$	20—24 $\bar{x}_2=22,5$	25—29 $\bar{x}_3=27,5$	30—34 $\bar{x}_4=32,5$	35—39 $\bar{x}_5=37,5$	40—44 $\bar{x}_6=42,5$	45—49 $\bar{x}_7=47,5$	50—54 $\bar{x}_8=52,5$	55—60 $\bar{x}_9=57,5$	60 лет и старше $\bar{x}_{10}=62,5$	неизвестно
Всего	2 087 599	105 491	557 691	783 678	202 433	123 862	60 196	35 046	44 572	57 094	115 358	2 178
в том числе: моложе 20 лет ($\bar{y}_1=18,0$)	460 813	70 583	203 261	172 850	12 524	1 302	225	51	—	—	—	17
20—24 $\bar{y}_2=22,5$	616 245	30 302	257 983	286 476	33 977	6 022	1 177	285	—	—	—	23
25—29 $\bar{y}_3=27,5$	444 399	3 896	83 985	250 189	74 350	25 220	5 043	1 249	451	—	—	16
30—34 $\bar{y}_4=32,5$	165 123	557	10 070	56 178	50 226	33 109	9 903	2 752	1 166	612	531	19
35—39 $\bar{y}_5=37,5$	117 400	127	2 067	15 415	25 300	39 573	19 570	7 697	3 959	2 087	1 595	10
40—44 $\bar{y}_6=42,5$	72 676	11	264	2 472	6 028	15 525	17 870	11 583	9 508	5 191	4 221	3
45—49 $\bar{y}_7=47,5$	41 748	—	—	—	—	2 584	4 983	7 808	10 944	8 035	7 391	3
50 и старше $\bar{y}_8=52,5$	166 833	—	—	—	—	495	1 417	3 613	18 539	41 156	101 573	40
Неизвестно	2 362	15	61	98	28	32	8	8	5	13	47	2 047

до 52,5 года у невест и до 62,5 года у женихов. Тогда, исключая невест, возраст которых неизвестен, получаем[1]:

$$\bar{y}_1 = \frac{18{,}0\cdot 70583 + 22{,}5\cdot 30302 + 27{,}5\cdot 3896 + 32{,}5\cdot 557 + 37{,}5\cdot 127 + 42{,}5\cdot 11}{70583 + 30302 + 3896 + 557 + 127 + 11} \approx$$

$$\approx 19{,}8 \text{ года.}$$

Значит, разность между средними возрастами женихов и невест для первой группы женихов, т. е. $\bar{x}_1 - \bar{y}_1$, составляет 19,0—19,8 = = — 0,8 лет.

Произведя такой расчет по всем группам женихов, получим данные для табл. 49.

Таблица 49

Разности между средними возрастами женихов и невест, распределенные по среднему возрасту жениха

Средний возраст жениха ($\bar{x}_i$)	19,0	22,5	27,5	32,5	37,5	42,5	47,5	52,5	57,5	62,5
Разность между средними возрастами женихов и невест ($\bar{x}_i - \bar{y}_i$)	−0,8	+0,6	+3,9	+3,5	+3,6	+4,4	+5,4	+5,5	+5,6	+11,0

Примечание. В табл. 48, на основании которой был произведен данный расчет, — первый и последний возрастные интервалы у женихов и невест даны открытыми. Поэтому результаты —0,8 и +11,0 в некоторой степени отражают приближенный и условный характер расчета.

Результаты расчета В. В. Паевского по данным Ленинграда[2] приведены в табл. 50.

Представленные в табл. 49 и 50 числа, полученные статистической обработкой данных, отражают корреляционную связь средних возрастов женихов и невест. Хотя эти данные относятся к разным территориям (Ленинград и весь Советский Союз) и в разное время (1922 и 1966 гг.), они позволяют, однако, сделать определенные выводы о том, что соответствие возрастов женихов и невест весьма неустойчиво.

[1] Можно условно предположить, что невесты, возраст которых неизвестен, распределены по всем группам пропорционально их численности.

[2] В. В. Паевский. Средняя разность возрастов женихов и невест. «Бюллетень Ленинградского Губернского отдела статистики» № 8 за 1924 г., стр. 41.

Таблица 50

Разности между возрастом жениха и невесты, распределенные по возрасту жениха

Возраст жениха	19	20	23	27	32	37	42	47	52	57	65
Средняя разность возрастов жениха и невесты	−1,5	−1,0	0,8	3,2	6,0	8,2	10,2	12,8	14,8	15,5	20,9

Для более конкретных выводов нужно привлечь статистические данные по одной и той же территории. В какой мере совершаются браки в пределах определенных возрастных групп, мы можем определить с помощью средних возрастов невест, вычисленных для определенного возраста женихов. А как же ответить на вопрос, совершаются ли браки в пределах определенной общественной группы, семейного состояния, национальности и др.?

Демографическая статистика разработала метод «женихов и невест», который может быть применен для решения этих и аналогичных задач.

Совокупность лиц, вступающих (или вступивших) в брак в течение определенного периода, группируют по определенным признакам (национальность, прежнее семейное состояние, грамотность и т. д.) и составляют таблицы, указывающие на связь этих признаков у женихов и невест. Например, в табл. 51 даны возможные комбинации по прежнему семейному состоянию.

Таблица 51

Схема комбинации супружеских пар

Женихи	Невесты		
	1 — девицы	2 — вдовы	3 — разведенные
I — холостые	I—1	I—2	I—3
II — вдовцы	II—1	II—2	II—3
III — разведенные	III—1	III—2	III—3

Значит, можно сопоставить для каждого периода следующие комбинации брачных пар по рассматриваемому признаку:

I — 1 — холостые и девицы;
I — 2 — холостые и вдовы;
I — 3 — холостые и разведенные;
II — 1 — вдовцы и девицы;
II — 2 — вдовцы и вдовы;
II — 3 — вдовцы и разведенные;
III — 1 — разведенные и девицы;

III — 2 — разведенные и вдовы;

III — 3 — разведенные и разведенные.

Аналогично выбираются возможные комбинации и для других признаков. Изучая взаимные сочетания признаков женихов и невест, можно установить, какие из перечисленных комбинаций встречаются часто, какие реже, а какие совсем редко.

Для изучения «притяжения» признаков женихов и невест друг к другу были разработаны индексы брачности. Рассмотрим применительно к данному примеру сущность индекса Бенини, модифицированного М. В. Птухой.

Возьмем величины, характеризующие совокупности, дающие данную комбинацию, например I — 1 (холостые и девицы), и обозначим общее число вступивших в брак холостых мужчин — B, а общее число девиц — C. Их произведение составит BC. Разделим это произведение на общее число заключенных браков D и получим $\frac{BC}{D}$ — число холостых и девиц при безразличии к наличию признака.

Возьмем все браки, заключенные женихами B. При безразличии к признаку они должны разбиться на две группы: однородные (I — 1, II — 2, III — 3) и неоднородные (все остальные). Однородных будет $\frac{BC}{D}$, а неоднородных $B - \frac{BC}{D}$. При наличии «притяжения» однородных совокупностей первая группа будет увеличиваться за счет второй. При наличии «отталкивания» — наоборот. Обозначим число браков в комбинациях через A, найдем разность $A - \frac{BC}{D}$, показывающую размер увеличения первой группы за счет второй. Если теперь эту разность отнести к величине той совокупности, из которой она взята, то получим индекс «притяжения» $I_{\text{пр}}$:

для женихов

$$I_{\text{пр. (ж)}} = \frac{A - \frac{BC}{D}}{B - \frac{BC}{D}} = \frac{AD - BC}{B(D - C)};$$

для невест

$$I_{\text{пр. (н)}} = \frac{A - \frac{BC}{D}}{C - \frac{BC}{D}} = \frac{AD - BC}{C(D - B)}.$$

Для обоих полов индекс «притяжения» можно получать как среднюю геометрическую между $I_{\text{пр. (ж)}}$ и $I_{\text{пр. (н)}}$

$$\bar{I}_{\text{пр}} = \sqrt{I_{\text{пр. (ж)}} I_{\text{пр. (н)}}} = \frac{AD - BC}{\sqrt{CB(D - C)(D - B)}}.$$

Аналогично получается индекс «отталкивания», имеющий тот же логический смысл и одинаковый для обоих полов, но с отрицательным знаком:

$$I_{\text{отт}} = \frac{A - \frac{BC}{D}}{\frac{BC}{D}} = \frac{AD}{BC} - 1.$$

Значение получаемых индексов колеблется в пределах от — 1 (полное «отталкивание») через 0 (безразличие) к + 1 (полное «притяжение»). Распространяя рассматриваемые нами индексы не только на комбинацию семейного положения, а на национальные, возрастные и другие признаки, можно изучать изменение брачности во времени. Сравнивая индексы брачности в СССР с индексами в капиталистических странах, можно установить различия, обусловленные социально-экономическими факторами.

Можно предположить, что индекс взаимного притяжения между группами однородных молодых возрастов вступающих в брак в СССР будет выше, чем в капиталистических странах. Тем самым будет подтверждено уменьшение влияния косвенных мотивов при вступлении в брак (социально-экономических и имущественных).

Интересный вывод, на наш взгляд, может получиться при изучении индекса брачности по национальному признаку. Следует полагать, что индекс «протяжения» женихов всех национальностей по отношению к невестам той же национальности уменьшится. Это будет свидетельствовать о том, что национальные различия все меньше и меньше перестают служить препятствием для вступления в брак.

Метод «женихов и невест» служит ограниченно тем же целям, что и показатель тесноты корреляционной связи, и может быть успешно применен для ответа на частный вопрос корреляционной теории о тесноте «притяжения» или «отталкивания» каких-нибудь двух конкретных значений взаимосвязанных признаков у двух объектов, относительно которых известно фактическое число имевших место случаев комбинации. Продолжим наш пример о соответствии прежнего семейного состояния женихов и невест (пример условный).

Таблица 52

Комбинации супружеских пар

Женихи	Невесты			
	девицы	вдовы	разведенные	всего
Холостые	30-A	5	5	40-B
Вдовцы	20	10	5	35
Разведенные	10	5	10	25
Всего	60-C	20	20	100-D

Покажем вычисление индексов «притяжения» для одной комбинации семейного состояния вступающих в брак, например, когда женихи — холостые, а невесты — девицы.

Из табл. 52 имеем $A = 30$, $B = 40$, $C = 60$, $D = 100$.

Найдем индекс «притяжения» для женихов и невест:

$$I_{\text{пр. (ж)}} = \frac{30 \cdot 100 - 40 \cdot 60}{40 \cdot (100 - 60)} = \frac{3}{8},$$

$$I_{\text{пр. (н)}} = \frac{30 \cdot 100 - 40 \cdot 60}{60 \cdot (100 - 40)} = \frac{1}{6}.$$

Теперь можно определить индекс «притяжения» для обоих полов:

$$\bar{I}_{\text{пр}} = \sqrt{\frac{3}{8} \cdot \frac{1}{6}} = \frac{1}{4}.$$

При наличии показателей в динамике можно выяснить, каким изменениям подвергаются фиксированные для каждого периода соответствия семейного, возрастного, национального и других признаков женихов и невест. Однако, пользуясь этим методом анализа, следует помнить, что в основе системы индексов «притяжения» лежит гипотеза, соответствующая урновым схемам, но не подтвержденная действительностью о равновозможности «встреч» потенциальных женихов и невест. Так, вероятность «встречи» женихом-узбеком русской невесты меньше, чем невесты узбечки, хотя в СССР русских женщин в несколько десятков раз больше, чем узбечек.

До середины 30-х годов разработанная Кучинским и Лоткой модель воспроизводства населения довольно успешно использовалась для изучения воспроизводства женского населения.

Между тем показатели воспроизводства имеют различные значения для мужчин и женщин, и общее представление о движении населения в целом только по динамике женского населения является неполным. Следовательно, недостаток этой модели в игнорировании воспроизводства мужчин.

Модификация показателей воспроизводства населения

В 40-х годах началось совершенствование моделей стабильного населения на основе анализа брачности. Дело в том, что, основываясь на модели воспроизводства населения Кучинского и Лотки, можно допустить большие неточности, так как эта модель не учитывает влияния изменения брачности в стабильном населении на такие показатели воспроизводства, как брутто- и нетто-коэффициенты.

Учет поправок к этим показателям в рамках первоначальной модели приводит к получению модифицированных показателей воспроизводства.

Первое уточнение состояло в расчленении функций плодовитости на две: брачной и внебрачной. Обозначая их $\varphi(x)$ и $\varphi'(x)$, а процент замужних женщин в возрасте x лет $z(x)$, получим нетто-коэффициент воспроизводства для женщин:

$$R_{0\,(F)} = \sum_{15}^{49} z(x)\, p(x)\, \varphi(x) + \sum_{15}^{49} [1 - z(x)]\, p(x)\, \varphi'(x).$$

Аналогично можно получить этот показатель и для мужчин. При этом за основу принимается таблица брачности какого-нибудь года, по ней находят удельные веса состоящих в браке по возрастным группам.

Дальнейшее совершенствование этой модели состоит в более детальном учете брачной плодовитости. Некоторые исследователи пытались использовать связь брачной плодовитости не только с возрастом матери, но и с продолжительностью брака. Обозначим плодовитость женщин в возрасте x лет, состоящих в браке в течение τ времени, $\Phi(x, \tau)$, а удельный вес таких женщин $n(x, \tau)$ и найдем плодовитость по возрасту $\Phi(x)$

$$\Phi(x) = \sum_{\tau=0}^{x=15} n(x, \tau)\, \Phi(x, \tau).$$

Ряд моментов позволяет утверждать, что общая плодовитость женщин может вырасти в случае увеличения удельного веса браков малой продолжительности, т. е. $n(x, \tau)$ при малых τ увеличивается при неизменной функции плодовитости $\Phi(x, \tau)$. Это произойдет потому, что плодовитость, относящаяся к малым величинам τ, выше плодовитости более длительных браков. Именно поэтому нетто-коэффициент воспроизводства, исчисляемый по первоначальной модели Лотки, завышен.

Но указанное обстоятельство имеет преходящее значение. Исходя из этого, требуется учитывать и продолжительность браков.

Известны попытки демографов скандинавских стран и американского демографа Хайнала заменить показатели плодовитости по возрасту и продолжительности брака показателем, являющимся функцией возраста женщины при вступлении в брак. Рассмотрим математическую последовательность таких рассуждений. Если $\Phi(x, y, \tau)$ — плодовитость браков, заключенных между женщинами в возрасте x лет и мужчинами в возрасте y лет и длившихся τ лет, то брутто-продуктивность таких браков может быть рассчитана по формуле

$$F_b(x, y) = \int \Phi(x, y, \tau)\, d\tau.$$

Вероятность сохранения брака между женщинами x лет и мужчинами y лет в течение τ лет обозначим $q(x, y, \tau)$. Тогда формула нетто-продуктивности этих браков примет вид:

$$F_{n\,(x, y)} = \int \Phi(x, y, \tau)\, q(x, y, \tau)\, d\tau.$$

Обозначая далее удельный вес браков, заключенных между женщинами возраста x лет и мужчинами возраста y лет в поколении мужчин $z\,(x,\ y)$ и этот же удельный вес в поколении женщин $z'\,(x,\ y)$, получим нетто-коэффициент, соответствующий мужской брачности, в виде двойного интеграла:

$$R_{0\,(M)}=\iint z\,(x,\ y)\,F_0(x,\ y)\,dx\,dy$$

и для женской брачности:

$$R_{0\,(F)}=\iint z'\,(x,\ y)\,F_0(x,\ y)\,dx\,dy.$$

Можно представить и число браков для поколений мужчин:

$$\gamma_M=\iint z\,(x,\ y)\,dx\,dy$$

и для женщин:

$$\gamma_F=\iint z'\,(x,\ y)\,dx\,dy.$$

На основании этих показателей рассчитаем среднюю нетто-продуктивность браков в течение одного поколения:

1) по мужской брачности:

$$F_{0\,(M)}=\frac{\iint z\,(x,\ y)\,F_0\,(x,\ y)\,dx\,dy}{\iint z\,(x,\ y)\,dx\,dy},$$

2) по женской брачности:

$$F_{0\,(F)}=\frac{\iint z'\,(x,\ y)\,F_0\,(x,\ y)\,dx\,dy}{\iint z'\,(x,\ y)\,dx\,dy}.$$

Значит, формулы нетто-коэффициента воспроизводства примут вид:

1) $R_{0\,(M)}=\gamma_M F_{0\,(M)}$,

2) $R_{0\,(F)}=\gamma_F F_{(0)\,F}$.

Ввиду того, что конкретные расчеты, произведенные Хайналом, показали отклонения в этих показателях, достигающие 1—3%, было предложено использовать невзвешенную среднюю арифметическую.

Изучение воспроизводства населения методом когорт

Весьма перспективным при изучении воспроизводства населения является использование метода когорт, при котором обобщенные демографические показатели, относящиеся к населению, включающему в себя различные группы, сформированные в разные периоды, заменяют единым демографическим показателем, суммирующим фазы развития определенной группы населения, сформированной в один период (когорты).

Плодовитость женщин есть величина переменная. С одной стороны, ее можно рассмотреть как функцию возраста, с другой — для женщин одного и того же возраста она меняется во времени. При этом внутри когорты действует своя закономерность. На когорту женщин одного года рождения очень сильное влияние оказывает демографическое прошлое. Это означает, что уровень плодовитости молодых женщин в настоящее время может резко отличаться от плодовитости пожилых женщин в их молодые годы, через некоторое время она не будет равна плодовитости женщин, относящихся к более старшим возрастным группам в настоящее время.

Метод когорт дает возможность изучать плодовитость путем расчета конечной величины семьи[1].

Можно, например, допустить, что плодовитость в настоящее время так же влияет на полную величину семьи, как и в последней когорте с законченной плодовитостью. При этом требуется экстраполяция. Пусть продолжительность брака — τ, а время формирования когорты — t, тогда показатель плодовитости когорты в настоящее время — $\varphi\,(t,\ \tau)$. Число рожденных от этих браков в течение τ лет составит:

$$\Psi\,(t,\ \tau)=\int_0^{\tau}\varphi\,(t,\ \tau)\,d\tau.$$

Если принять изменение этой функции в указанном интервале по параболе высокого порядка, то получим, например:

$$\Psi\,(t,\ \tau)=A_t+B_t\tau+C_t\tau^2+D_t\tau^3+\ldots$$

Останавливаясь на параболе второго порядка, имеем: $A_t=A_0+A_1t+A_2t^2$. Используя метод наименьших квадратов, можно определить A_t. Аналогично находят B_t, C_t и т. д. При этом точность повышается, если плодовитость устанавливается для отдельных групп по возрасту вступления в брак, т. е. учитывается возраст x. Тогда за основу принимается функция $\psi\,(t,\ \tau,\ x)$.

Устанавливая эту функцию для каждого возраста и находя соответствующие числа рождений, можно получить конечную величину семьи, характеризующую когорту. При экстраполяции конечной величины семьи исследователь встречается с большими трудностями. Как показали практические расчеты, сравнительно неплохие результаты в этом случае дает применение сложных модифицированных логистических кривых.

[1] Затруднения при этом состоят в том, что когорты женщин, для которых период плодовитости уже закончился и величина семьи известна, сформированы в результате браков двадцатилетней и большей давности и отражают уровень предшествующих периодов, в то время как для женщин, рожающих сейчас, конечную величину семьи еще нельзя установить, но можно оценить путем некоторых предположений.

ВЛИЯНИЕ ВОЗРАСТНОЙ СТРУКТУРЫ НА НЕТТО-КОЭФФИЦИЕНТ ВОСПРОИЗВОДСТВА

Попробуем найти влияние каждого возраста на величину нетто-коэффициента воспроизводства R_0. Возьмем отношение $\frac{p(x)\varphi(x)dx}{\int_{15}^{49} p(x)\varphi(x)dx} = w(x)$ и назовем его относительным компонентом нетто-коэффициента воспроизводства в разбивке по возрасту. Аналогичный компонент для брутто-коэффициента воспроизводства составит:

$$\frac{\varphi(x)\,dx}{\int_{15}^{49} \varphi(x)\,dx} = w'(x).$$

Зная, что

$$R_0 = \int_{15}^{49} p(x)\,\varphi(x)\,dx = R_b\,\frac{\int_{15}^{49} p(x)\,\varphi(x)\,dx}{\int_{15}^{45} \varphi(x)\,dx} = R_b\,\overline{p(x)},$$

можно сделать вывод, что R_0 есть R_b, умноженный на взвешенную среднюю арифметическую из вероятностей дожития, где весами являются относительные компоненты брутто-коэффициента воспроизводства в разбивке по возрасту.

МАТЕМАТИЧЕСКАЯ ТЕОРИЯ ВОСПРОИЗВОДСТВА

Представляет интерес математическая теория воспроизводства населения, разработанная сотрудниками института прикладной математики берлинского университета Бонцом и Хилбунгом, в которой авторы доказывали, что при наличии 4 условий: 1) закрытого населения; 2) неизменного порядка вымирания; 3) постоянного общего коэффициента плодовитости женщин 15—49 лет; 4) постоянной половой пропорции среди родившихся — при зафиксированном начале отсчета, принятом за нуль, отношение числа родившихся в n-м году к n-й степени некоторого числа λ имеет пределом постоянную величину C, т. е.

$$\lim \frac{L_n}{\lambda^n} = C.$$

При этом λ может быть определена из уравнения:

$$\lambda^{45} = K_w\,(l_{15}\,\lambda^{30} + l_{16}\,\lambda^{29} + \ldots + l_{44}\,\lambda + l_{45}),$$

где K_w — отношение числа родившихся девочек к числу женщин в возрасте 15—49 лет; l_{15}, l_{16}, ..., l_{45} — числа доживающих женщин из таблицы смертности при $l_0 = 1$. Анализ показывает наличие трех возможностей: $\lambda < 0$; $\lambda = 0$; $\lambda > 0$. При $\lambda < 0$ в условиях длительного не-

изменного порядка вымирания и коэффициента плодовитости число родившихся, а следовательно, и численность населения убывает в геометрической прогрессии. При $\lambda = 0$ число родившихся стремится к стационарности, а при $\lambda > 0$ — возрастает.

В 1932 г. В. В. Паевский и С. А. Новосельский дали приближенное решение этого уравнения и получили для СССР $\lambda = 1{,}0168$, что соответствовало предположению о предельном годовом приросте населения СССР, равном 16,8‰; фактически же он составлял около 20‰.

Р. Мизес пытался усовершенствовать построения Бонца и Хилбунга, заменив их уравнение приближенным:

$$\lambda = 1 + 0{,}003 \cdot K_w \sum_{15}^{49} l_x$$

Как мы видим, фактический коэффициент естественного прироста заметно отличается от предсказанного предельного асимптотического. Разумеется, что предельно асимптотические построения имеют лишь теоретический интерес при изучении характеристик плодовитости и смертности. В настоящее время для решения задач, связанных с построением моделей воспроизводства населения, с успехом применяется матричное исчисление[1].

ВЗАИМОСВЯЗЬ ПОКАЗАТЕЛЕЙ

Как удалось доказать В. Борткевичу, между средней продолжительностью жизни (e_0^0), коэффициентом рождаемости (n), коэффициентом смертности (m) и средним возрастом умерших ($\bar{x}$) существуют взаимосвязи, определяемые различными допущениями о порядке вымирания населения во времени и изменении плотности рождений.

При этом под неизменным во времени порядком вымирания В. Борткевич понимал гипотезу о том, что числа доживающих до того же возраста из разных совокупностей родившихся пропорциональны этим совокупностям. Обозначая $v\,(x,\,t)$ число лиц, которые достигли или достигнут возраста x из родившихся в промежуток времени от 0 до t; $v\,(0,\,t)$ — число родившихся в этот же промежуток времени, а $f(x)$ — вероятность для новорожденного дожить до возраста x. В. Борткевич получил функцию двух переменных, выраженную произведением двух функций, из которых первая есть функция переменного t, а вторая — x:

$$v\,(x,\,t) = v\,(0,\,t)\,f\,(x).$$

Далее, В. Борткевич вводит плотность переживания возраста x для времени рождения t, понимая под этим термином частное от де-

---

[1] См. А. Я. Боярский и др. Курс демографии.

ления чисел переживших определенный возраст в течение бесконечно малого промежутка времени на этот промежуток:

$$\lim_{\Delta t \to 0} \frac{\Delta v(x, t)}{\Delta t} = \frac{\partial v(x, t)}{\partial t}.$$

Принимая x, равным нулю, можно получить плотность рождений для времени t, т. е. $F(t)$:

$$\lim_{\Delta t \to 0} \frac{\Delta v(0, t)}{\Delta t} = \frac{dv(0, t)}{dt} = F'(t) \text{— плотность рождений.}$$

Пользуясь введенными в анализ величинами, В. Борткевич выявил гипотетические взаимосвязи показателей воспроизводства населения[1].

Таблица 53

Взаимосвязь показателей $\mathring{e}_0$, n, m и $\bar{x}$ при неизменном порядке вымирания

Плотность рождений	Формулы, связывающие $\mathring{e}_0$, n, m и $\bar{x}$
Постоянна	$\mathring{e}_0 = \frac{1}{n} = \frac{1}{m} = \bar{x}$ (стационарное население)
Возрастает	$\mathring{e}_0 > \frac{1}{n}$; $\frac{1}{m} > \bar{x}$; $\frac{1}{m} > \frac{1}{n}$; $\mathring{e}_0 > \bar{x}$
Убывает	$\mathring{e}_0 < \frac{1}{n}$; $\frac{1}{m} < \bar{x}$; $\frac{1}{m} < \frac{1}{n}$; $\mathring{e}_0 < \bar{x}$

Примечание. Знак неравенства между $\bar{x}$ и $\frac{1}{m}$ определяется не только возрастанием или убыванием плотности рождений, но зависит еще от порядка вымирания. Когда же речь идет о неравенстве между $\bar{x}$ и $\frac{1}{n}$, это условие еще должно быть дополнено указанием на закон, по которому происходит возрастание или убывание плотности рождений

В результате анализа В. Борткевич делает вывод о самом малом значении n, m и $\bar{x}$ для характеристики «истинной меры жизни» e_0^0, так как все эти показатели зависят не только от смертности, но и обусловлены плотностью рождений. При различных допущениях $\frac{1}{n}$ и $\bar{x}$ всегда лежат по одну сторону e_0^0 и не определяют границ ее изменения.

[1] См. В. Борткевич. Смертность и долговечность мужского православного населения Европейской России. Записки Императорской Академии Наук, т. 63, кн. II. СПб., 1890.

Глава 4 ВЫРАВНИВАНИЕ (СГЛАЖИВАНИЕ) ДЕМОГРАФИЧЕСКИХ ПОКАЗАТЕЛЕЙ. ИНТЕРПОЛЯЦИЯ И ЭКСТРАПОЛЯЦИЯ

ВОЗРАСТНАЯ АККУМУЛЯЦИЯ И МЕТОДЫ ЕЕ ИЗМЕРЕНИЯ

При регистрации возраста в момент переписи населения или в загсах в результате склонности населения к округлению своего возраста вместо указания точного численность в возрастах, кратных пяти, т. е. оканчивающихся на 0 и 5, часто оказывается сильно завышенной. Такое скопление численности населения на круглых возрастах и называется аккумуляцией возрастов.

Измерение аккумуляции, т. е. преобладания круглых возрастов, может производиться различными методами.

С. А. Новосельский предложил следующую формулу для определения коэффициента аккумуляции:

$$\overline{K}_{\text{аккум}} = \frac{\sum_{K=5}^{12} S_{5K}}{\frac{1}{5} \sum_{K=23}^{62} S_K}.$$

В числителе суммируется численность населения возрастов, кратных 5, в интервале 23—62 года. В знаменателе же берется 20% общей численности населения в тех же возрастах. Данный метод исходит из предположения, что численность населения в пределах 5-летнего интервала возраста изменяется по прямой $y = a_0 + a_1 x$. Отсюда равенство:

$$5S_{25} = S_{23} + S_{24} + S_{25} + S_{26} + S_{27}$$

и т. д. для каждой пятилетней группы в возрасте от 23 до 62 лет.

Можно предложить другой метод, более трудоемкий, но устанавливающий аккумуляцию точнее:

1) выравнивают численность населения (S) по одногодичным возрастам каким-нибудь методом (допустим, скользящей средней) и получают $\overline{S}_x$;

2) находят отношение фактических численностей аккумулирующих возрастов к выровненным $\frac{S}{\overline{S}_x}$ и получают соответствующие коэффициенты:

$$K_0=\frac{S_0}{\bar{S}_{x=0}};\quad K_1=\frac{S_5}{\bar{S}_{x=5}};\quad K_2=\frac{S_{10}}{\bar{S}_{x=10}}\ \text{и т. д.};$$

3) из этих коэффициентов находят среднюю гармоническую, где весами является фактическая численность населения:

$$\bar{K}_{\text{аккум}}=\frac{\Sigma S}{\sum \frac{S}{K}}=\frac{S_0+S_5+S_{10}+\ldots}{\frac{S_0}{\frac{S_0}{\bar{S}_{x=0}}}+\frac{S_5}{\frac{S_5}{\bar{S}_{x=5}}}+\frac{S_{10}}{\frac{S_{10}}{\bar{S}_{x=10}}}+\ldots}=$$

$$=\frac{S_0+S_5+S_{10}+\ldots}{\bar{S}_{x=0}+\bar{S}_{x=5}+\bar{S}_{x=10}+\ldots},$$

или среднюю арифметическую, где в качестве весов принимается выровненная численность:

$$\bar{K}_{\text{аккум}}=\frac{\Sigma K\bar{S}_x}{\Sigma \bar{S}_x}.$$

Найденные средние и будут представлять числовую меру того, насколько в среднем аккумулирующие возраста дают численность большую, чем следовало бы ожидать, если бы аккумуляции не было. Эта числовая мера и может служить коэффициентом аккумуляции.

Английские демографы использовали интересный прием оценки неточности возрастных распределений. Население в возрасте 23—72 года расчленяется на 10 пятилетних групп, составленных из возрастов, оканчивающихся одной цифрой. Например, первая группа состоит из возрастов 23, 33, 43, 53, 63; вторая группа включает возраста 24, 34, 44, 54, 64 и т. д.; 10-я группа включает возраста 32, 42, 52, 62, 72 года. Определяется удельный вес каждой из 10 групп в одной десятой численности населения в возрасте 23—72 года. При очень благоприятных обстоятельствах, т. е. когда аккумуляция отсутствует, этот показатель, называемый «возрастной скученностью», дает во всех группах величину, близкую к единице.

Вычисления методом С. А. Новосельского по переписи населения 1897 г. в России дали по европейской части $K = 1{,}75$; в СССР по переписи населения 1926 г. общий $K = 1{,}59$, а для Сибири, например, 1,62. По переписи населения 1959 г. коэффициент аккумуляции был равен по Советскому Союзу 1,09, а по республикам колебался в сравнительно небольших пределах.

Многие исследователи полагают, что искажение истинного возрастного состава округлением чисел, оканчивающихся на 0 и 5 лет, обратно пропорционально уровню грамотности и культуры.

С этим мнением можно частично согласиться. Но нужно еще учитывать, что коэффициент аккумуляции является числовой мерой, характеризующей также совершенство методов статистического наблюдения и опроса населения, и может быть применен для измерения степени неточности регистрации возраста. Вычисления показали, что

по данным переписи населения 1926 г. и по уточненным результатам выборочного обследования лиц старше 80 лет по тем республикам, где такие обследования проводились в 1959 г., при ответе на вопросы о возрасте чаще ошибаются женщины, чем мужчины, сельское население, чем городское, население старших возрастов, чем молодое, и т. д.

Из результатов выборочного обследования, произведенного Институтом геронтологии АМН СССР по уточнению возрастного распределения по переписи населения 1959 г., выявилось, что старики 80 лет и старше чаще всего (в 3—4 раза) преувеличивают свой возраст, чем преуменьшают. Установлено, что неграмотные допускают при определении возраста значительно больше ошибок, чем грамотные. Следует отметить, что для стариков, помимо обычной склонности к округлению, характерно еще так называемое старческое кокетство — склонность преувеличивать свой возраст; женщины же в возрастах примерно 30—65 лет склонны его уменьшать.

Произведенные нами расчеты показали, что аккумуляция возраста умерших значительно ниже, чем живущих. С точки зрения исследователя, особого внимания заслуживают вероятности умереть или дожить, относящиеся к возрастам, кратным пяти. С учетом аккумуляции на этих возрастах концентрируются числа умирающих или доживающих больше, чем на соседних возрастах. Вероятностям, исчисленным на основе большей численности, присуща большая надежность, чем вероятностям, полученным из малого числа умерших и живых. Так, с вероятностью, близкой к достоверности (к единице), можно утверждать на основании теорем П. Л. Чебышева и А. М. Ляпунова, что разность между искомой вероятностью, например, дожития p и фактической, полученной из наблюдения, p' не больше $3\frac{\sqrt{pq}}{S}$ или вероятность неравенства $P[(p-p')<3\frac{\sqrt{pq}}{S}]$ равна $\frac{2}{\sqrt{\pi}}\int_0^3 e^{\frac{-t^2}{2}}dt = 0{,}997$, где S — численность населения. Чем больше S, тем ближе p и p'.

Все изложенное можно было бы считать правильным в том случае, если численность населения возрастов, кратных 5, была бы действительно больше соседних возрастов. А так как этого нет, исследователь прибегает к математической обработке данных, обеспечивающей дальнейшее использование фактических численностей.

Можно рекомендовать рассматривать вместо вероятности, относящейся к возрасту, кратному 5, некоторую среднюю вероятность для возрастов $5t+i$.

Это означает, что для возраста $5t$ указывается не та вероятность, которая получается с использованием численности этого возраста, а средняя вероятность из 7 или 13 соседних одногодичных интервалов.

Если берется 7 соседних возрастов, то получается средняя из показателей возрастов: $5t-3$; $5t-2$; $5t-1$; $5t$; $5t+1$; $5t+2$; $5t+3$. В этом случае допускается, что опрашиваемый мог допустить ошибку при ответе на вопрос о возрасте в пределах трех лет. Конечно, могут быть ошибки и в 5 и даже в 10 лет. Но, как показывает практика, вероятность отнесения возраста к числам $5t$ больше, чем вероятность отнесения его к виду $10t$.

РАСЩЕПЛЕНИЕ ВОЗРАСТНЫХ ИНТЕРВАЛОВ

Оценка группировок

Оценим достоинства и недостатки применяющихся группировок.

Общепринятая группировка по пятилетним интервалам (20—24; 25—29; 30—34 и т. д.) имеет тот недостаток, что аккумулирующий возраст стоит не в центре группы, а на краю, и поэтому некоторая часть единиц из-за аккумуляции попадает из одной группы в другую.

Так, рассматривая группу 25—29 лет, заключаем, что действительная численность населения равна:

$$\sum_{x=25}^{x=29} S_x = S_{25} + S_{26} + S_{27} + S_{28} + S_{29}.$$

Перепись же учтет другую величину:

$$S_{25} + (a_{24} + a_{23}) + S_{26} + S_{27} + S_{28} - a_{28} + S_{29} - a_{29}.$$

Значит, разница между численностью пятилетнего интервала по переписи и той, которая должна была бы быть при отсутствии аккумуляции, составляет $(a_{24} + a_{23}) - (a_{28} + a_{29})$.

Эта величина означает разность между оторванными численностями возрастов от предыдущего интервала $(a_{24} + a_{23})$ аккумулирующей силой возраста 25 лет и оторванными численностями от рассматриваемой группы аккумулирующей силой возраста 30 лет. Но аккумуляция, как известно, на возрасте, кратном нулю, сильнее, чем на возрастах, кратных пяти. Поэтому аккумулирующие числа, притянутые из предыдущей группы, будут меньшими, чем оторванные из рассматриваемой группы и учтенные в последующей.

Окончательный вывод, следовательно, такой: численности групп, заключающих в себе возраст, оканчивающийся на 5, будут занижены за счет некоторого завышения численности интервалов, содержащих в себе возраст, оканчивающийся на нуль.

Возникает заманчивая мысль, помещать аккумулирующий возраст в центр пятилетних групп: 23—27; 28—32; 33—37 и т. д. Однако и в этом случае скажется аккумулирующая сила возрастов 0 и 5 лет. С увеличением возраста притяжение к возрасту, кончающемуся нулем, возрастает по сравнению с притяжением возрастов, оканчивающихся пятеркой. Поэтому материалы, полученные в результате такой группировки, очень ненадежны, так как неточность суммарных численностей увеличивается с ростом возрастов (особенно сильно в старческих).

Выход состоит в использовании интервалов, включающих в себя аккумулирующие возраста, кратные и нулю и пяти, т. е. десятилетних интервалов, либо в выравнивании однолетних интервалов.

Иногда поступают так: до 25-летнего возраста группировку ведут по пятилетним интервалам, а для возрастов старше 25 лет — по десятилетним. При этом лучше всего создавать десятилетние интервалы так, чтобы наибольшая аккумулирующая цифра, т. е. нуль, оказалась в центре возрастного интервала. Если такое построение производить без дробления возрастов, то наиболее приемлемой оказывается группировка по интервалам: 25—34; 35—44 и т. д. или 26—35; 36—45 и т. д.

Во всяком случае следует иметь в виду, что решить задачу подбора возрастных интервалов, в какой-то степени уменьшающих влияние аккумуляции, можно в каждом конкретном случае лишь путем неоднократного экспериментирования.

Дробление интервалов

Нередко при обработке демографических данных возникает необходимость в дроблении, т. е. расщеплении уже имеющихся возрастных интервалов. Рассмотрим расщепление, основанное на предположении, что плотность вариационного ряда изменяется по параболе второго порядка. При этом будем вести отсчет в единицах, равных величине расщепляемого интервала. Если обозначить эту величину x, то при этом может быть два случая:

1) при равенстве интервалов расщепляемому интервалу будет соответствовать изменение x от 0 до 1. Интервалу, предшествующему расщепляемому, — от -1 до 0, а интервалу, последующему за расщепляемым, — от 1 до 2;

2) при неравных интервалах пределы изменения и величина x устанавливаются в зависимости от величины интервальных разностей.

Для использования частостей предположим, что они представляют собой интегралы плотностей распределения в соответствующих пределах, отвечающих величине x. При расщеплении интервала этот предел разбивается на 2 части: первую, составляющую долю θ в величине расщепляемого интервала, и вторую $1 - \theta$. Соответственно частость расщепляемого интервала F распадается на F_0^θ и $F - F_0^\theta$.

Обозначим m_{-1} — частость интервала, предшествующего расщепляемому; m_0 — частость расщепляемого интервала; m_{+1} — частость интервала, последующего за расщепляемым.

Возьмем плотность вариационного ряда $f(x) = a_0 + a_1 x + a_2 x^2$. Для определения параметров a_0, a_1 и a_2 используем соответствующие частости m_{-1}, m_0, m_{+1}. Получаем:

$$\int_{-1}^{0} f(x)\,dx = \int_{-1}^{0} (a_0 + a_1 x + a_2 x^2)\,dx = a_0 - \frac{a_1}{2} + \frac{a_2}{3} = m_{-1},$$

$$\int_{0}^{1} f(x)\,dx = \int_{0}^{1} (a_0 + a_1 x + a_2 x^2)\,dx = a_0 + \frac{a_1}{2} + \frac{a_2}{3} = m_0,$$

$$\int_{1}^{2} f(x)\,dx = \int_{1}^{2} (a_0 + a_1 x + a_2 x^2)\,dx = a_0 + \frac{3}{2} a_1 + \frac{7}{3} a_2 = m_{+1}.$$

Произведя алгебраические преобразования и решив систему трех уравнений, получим значения параметров.

Введем еще три символа:

ΔF_{-1} — приращение частостей, равное $m_0 - m_{-1}$;

$\Delta F_0 = m_{+1} - m_0$;

$\Delta^2 F$ — второе приращение частостей, равное $(m_{+1} - m_0) - (m_0 - m_{-1})$. Тогда

$$a_0 = \frac{m_{-1} + m_0}{2} - \frac{\Delta^2 F}{6},$$

$$a_1 = \Delta F_{-1} = m_0 - m_{+1},$$

$$a_2 = \frac{\Delta^2 F}{2} = \frac{(m_{+1} - m_0) - (m_0 - m_{-1})}{3} = \frac{m_{+1} - 2m_0 + m_{-1}}{2}.$$

Таким образом, $f(x) = \frac{m_{-1} + m_0}{2} - \frac{\Delta^2 F}{6} + \Delta F_{-1} x + \frac{\Delta^2 F}{2} x^2.$

Если теперь взять часть расщепляемого интервала, то его частость будет исчисляться по формуле

$$F_0^\theta = \int_0^\theta f(x)\,dx = a_0 \theta + \frac{a_1}{2} \theta^2 + \frac{a_2}{3} \theta^3 = \theta \left(a_0 + \frac{a_1}{2} \theta + \frac{a_2}{3} \theta^2 \right).$$

Подставляя найденные значения параметров, получим:

$$F_0^\theta = \theta \left[\frac{m_{-1} + m_0 + \Delta F_{-1} \theta}{2} - \frac{\Delta^2 F (1 - \theta^2)}{6} \right].$$

В качестве примера возьмем следующие данные о распределении мигрирующего населения в РСФСР в 1953 г. по возрасту[1].

Таблица 54

Прибытие мигрантов

Возраст (годы)	Частость %
0— 5	4,6
5—10	2,4
10—15	2,5

Произведем расщепление интервала 5—10 лет на две части, выделим прибывших в возрасте от 5 до 8 лет и определим их удельный вес:

$$\theta = \frac{8-5}{10-5} = \frac{3}{5} = 0{,}6;$$

$$m_{-1} = 4{,}6; \quad m_0 = 2{,}4; \quad m_{+1} = 2{,}5;$$

$$\Delta F_{-1} = 2{,}4 - 4{,}6 = -2{,}2; \quad \Delta^2 F = 2{,}5 - 2 \cdot 2{,}4 + 4{,}6 = 2{,}3.$$

Получаем частость по соответствующей формуле

$$F_0^{0,6} = 0{,}6 \left[\frac{4{,}6 + 2{,}4 + (-2{,}2) \cdot 0{,}6}{2} - \frac{2{,}3\,(1 - 0{,}6^2)}{6} \right] \approx 1{,}6\%.$$

При неравных интервалах расчет усложняется. Пределы интегрирования берутся в зависимости от величины интервалов, предшествующего расщепляемому и последующего за ним.

Для примера возьмем распределение городов и поселков городского типа СССР по числу жителей в 1959 г.[1]

Таблица 55

Число жителей в городах и поселках городского типа (в тыс.)	Число городов и поселков
От 3 до 5	904
» 5—10	1 296
» 10—20	798

Расщепим интервал от 5 до 10 тыс. человек и выделим группу городов с населением от 5 до 6 тыс. человек[2].

Исходя из общей численности городов и поселков городского типа 4619, найдем частости указанных трех групп городов и поселков. Получаем для первой группы частость 19,57%, для второй — 28,06% и для третьей — 17,28%.

Тогда $\theta = 0{,}2$ и $1 - \theta = 0{,}8$. Требуется найти частость $F_0^{0,2}$.

Полагая, что, как и при равных интервалах, счет ведется в единицах расщепленного интервала 10—5 = 5 тыс. человек, для него пределы интегрирования будут от 0 до 1, для предшествующего интервала, где интервальная разность равна 5—3 = 2 тыс. человек, пределы интегрирования будут от — 0,4 до 0, а для последующего интервала при интервальной разности 20—10 = 10 тыс. человек пределы будут соответственно равны 1 и 3. Сохраняя принятые обозначения, получим:

$$\int_{-0,4}^{0} f(x)\,dx = \left| a_0 + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3 \right|_{-0,4}^{0} = 0{,}4a_0 - 0{,}08a_1 + 0{,}02133a_2 = m_{-1};$$

$$\int_0^1 f(x)\,dx = a_0 + 0{,}5a_1 + 0{,}33333\,a_2 = m_0;$$

$$\int_1^3 f(x)\,dx = \left| a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3 \right|_1^3 = 2a_0 + 4a_1 + 8{,}66667a_2 = m_{+1}.$$

[1] См. А. И. Ежов. Выравнивание и вычисление рядов распределения. М., Госстатиздат, 1961, стр. 33.

[1] «Итоги Всесоюзной переписи населения 1959 года. СССР», стр. 35.

[2] Разумеется, что при наличии первичных данных можно требуемый интервал получить без расщепления.

Система уравнений для нахождения параметров примет вид:

$$\begin{cases} 0{,}4a_0-0{,}08a_1+0{,}02133a_2=19{,}57, \\ a_0+0{,}5a_1+0{,}33333a_2=28{,}06, \\ 2a_0+4a_1+8{,}66667a_2=17{,}28. \end{cases}$$

Откуда $a_0=41{,}972$; $a_1=-32{,}783$; $a_2=7{,}438$.

Теперь можно найти частоту F_0^0 группы, выделяемой при расщеплении, т. е. от 5 до 6 тыс. человек:

$$F_0^{0,2}=\int\limits_0^{0,2} f(x)\,dx=0{,}2\left(a_0+\frac{a_1}{2}0{,}2+\frac{a_2}{3}0{,}2^2\right)=$$

$$=0{,}2\left(41{,}972-\frac{32{,}783}{2}\cdot 0{,}2+\frac{7{,}438}{3}\cdot 0{,}04\right)\approx 7{,}76\%.$$

Следовательно, при расщеплении интервала 5—10 тыс. человек на две части: 5—6 тыс. человек и 6—10 тыс. человек — частость расщепленного интервала, равная 28,06%, распадается соответственно на 7,76% и 28,06—7,76 = 20,3%.

Для приведения данных, сгруппированных в 10-летние интервалы, к 5-летним интервалам можно использовать формулу Ньютона

$$f_{na}=\frac{1}{2}\left[f_n+\frac{1}{8}(f_{n-1}-f_{n+1})\right],$$

где f_{na} — численность в первой половине данной десятилетней группы, которая должна быть вычислена;
f_n — численность всей десятилетней группы;
f_{n-1} и f_{n+1} — численности в предшествующей и последующей десятилетних группах.

Среднее значение интервала

При изучении демографических процессов исследователю чаще всего приходится иметь дело с интервальными вариационными рядами. При этом за значение интервала принимается его середина, т. е. центр. Такое допущение может быть уточнено для всех интервалов, кроме двух первых и двух последних.

Обозначим возрастной признак x, а численность единиц, обладающих данным признаком, $f(x)$. Если рассматривать $f(x)$ как линейную функцию, то значение интервала будет средней арифметической его границ (верхней и нижней). Для уточнения численности будем рассматривать в качестве функции возраста не прямую линию, а параболу второго порядка $f(x)=a_0+a_1x+a_2x^2$. Тогда нижняя граница интервала будет x, верхняя $x+dx$, а число единиц в интервале $f(x)dx$. В конечных пределах от x_1 до x_2 получим:

$$\Sigma_0=\int\limits_{x_1}^{x_2} f(x)\,dx.$$

Требуется найти среднюю величину $(\bar{x})$ признака для данного интервала:

$$\bar{x}=\frac{\int\limits_{x_1}^{x_2} xf(x)\,dx}{\int\limits_{x_1}^{x_2} f(x)\,dx}=\frac{\int\limits_{x_1}^{x_2} xf(x)\,dx}{\Sigma_0}.$$

Допустим, что нам известны численные значения соседних интервалов: предыдущего Σ_{-1} и последующего Σ_{+1}. Предыдущий интервал ограничен пределами от $x_1-(x_2-x_1)$ до x_1, т. е. от $(2x_1-x_2)$ до x_1. Последующий интервал ограничен пределами от x_2 до $(2x_2-x_1)$.

Заменяем $f(x)$ параболой 2-го порядка и получаем примерно то же, что и при расщеплении интервалов:

$$\Sigma_{-1}=\int\limits_{2x_1-x_2}^{x_1} f(x)\,dx=\int\limits_{2x_1-x_2}^{x_1}(a_0+a_1x+a_2x^2)\,dx;$$

$$\Sigma_0=\int\limits_{x_1}^{x_2} f(x)\,dx=\int\limits_{x_1}^{x_2}(a_0+a_1x+a_2x^2)\,dx;$$

$$\Sigma_{+1}=\int\limits_{x_2}^{2x_2-x_1} f(x)\,dx=\int\limits_{x_2}^{2x_2-x_1}(a_0+a_1x+a_2x^2)\,dx.$$

Найдем параметры a_0, a_1 и a_2 и подставим в значение средней. Тогда получим:

$$\bar{x}=\frac{x_1+x_2}{2}+\frac{1}{24}(x_2-x_1)\cdot\left(\frac{\Sigma_{+1}-\Sigma_{-1}}{\Sigma_0}\right).$$

Если теперь допустим, что предыдущая и последующая численности равны друг другу, а разность x_2-x_1 есть величина интервала, равная k, то получим:

$$\bar{x}=\frac{x_1+x_2}{2}+\frac{1}{24}k.$$

Примерно такое же решение дал В. И. Борткевич.

Известно и другое решение аналогичной задачи, данное статистиком К. М. Беккером (при вычислении таблиц смертности для Германии по данным об умерших). Его формула отличается от ранее приведенной только коэффициентом при втором члене:

$$\bar{x}=\frac{x_1+x_2}{2}+\frac{1}{16}k.$$

Есть основание полагать, что К. М. Беккер, выводя свою формулу, без должной строгости отнесся к отбору исходных посылок. К.М. Беккер предположил, что 6 площадей элементарных совокупностей умер-

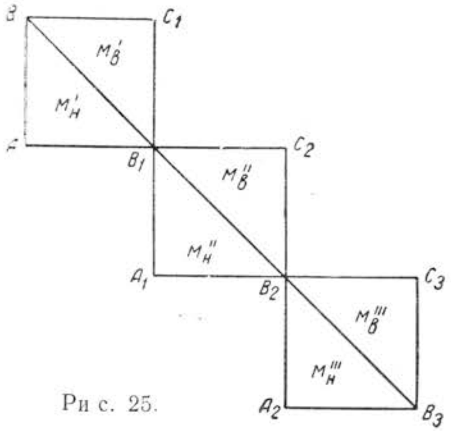

Рис. 25.

ших за примыкающие три года (см. рис. 25) $M'_{н}$, $M'_{в}$, $M''_{н}$, $M''_{в}$, $M'''_{н}$, $M'''_{в}$ расположены по параболе 2-го порядка. Однако, как справедливо указывает В. И. Борткевич, между этими неоднородными совокупностями нет той простой зависимости, которую допустил К. М. Беккер.

УСТРАНЕНИЕ ВОЗРАСТНОЙ АККУМУЛЯЦИИ

Колеблемость показателей (чисел рождений, смертей, браков, увеличение или уменьшение численности населения, изменения в направлении миграции и т. д.) является непременной особенностью демографических явлений. Для всех показателей последовательность во времени является решающей. Наблюдая динамику демографических процессов, исследователю следует иметь в виду наличие в них определенных тенденций, подчас скрытых и завуалированных причудливыми колебаниями уровней. Изучение динамических рядов предполагает непременное установление характера и особенностей их изменений. Иногда такое изучение рядов осложняется случайными кратковременными факторами, затрудняющими выявление основной тенденции движения ряда. Демография использует различные приемы для сглаживания динамических рядов, т. е. устранения случайных колебаний и обнаружения основной тенденции. В демографии могут найти применение такие методы, как исследование средней скорости ряда или среднего коэффициента роста и т. д. Однако подобные методы не являются специфичными для демографии, поэтому мы на них не останавливаемся и отсылаем интересующихся к специальной литературе[1]. Мы рассмотрим те приемы, которые учитывают особенности демографических показателей, связанных главным образом с возрастом.

[1] См. И. Г. Венецкий и Г. С. Кильдишев. Основы математической статистики, стр. 207—237.

Приступая к выравниванию данных об абсолютной численности населения по состоянию на критический момент переписи или чисел умерших за какой-нибудь период, распределенных по возрасту, нужно отчетливо понимать, что задачей этого выравнивания является сглаживание округленных при регистрации возрастов, но не самого характера изменений этих численностей при переходе из одного возраста к другому. Для устранения следов аккумуляции и элементов случайности в числах живущих или умерших их подвергают выравниванию.

Иногда выравниваются не абсолютные численности, а вероятности дожития или смерти, взятые из таблиц смертности. Выравнивание демографических данных имеет недостаток, свойственный вообще многим математическим методам. Допустим, что в некоторых данных, например числах живущих или умерших, имеют место резкие скачки, вызванные либо существенными, историческими причинами (например, резким понижением рождаемости, военными потерями мужского населения и аналогичными им), либо случайными (неправильной регистрацией возраста и др.). Наша задача состоит в устранении влияния случайных причин, искажающих изменение указанных численностей. Применив выравнивание, мы получаем гладкую математически правильную кривую. При этом у нас нет полной уверенности в том, что мы не потеряли каких-либо существенных черт исследуемого явления. Больше того, получив определенный порядок изменения чисел живущих или умерших в том или ином возрасте, мы не сможем сформулировать вывод о заведомой пригодности или непригодности принятого в данном характерном случае метода выравнивания. Дело в том, что нет однозначного ответа на вопрос о том, какие колебания в возрастной кривой следует отнести за счет случайных искажений, а какие выражают объективную действительность, т. е. являются существенными.

Частичный выход состоит в применении методов выравнивания не только к абсолютным числам живущих и умерших, а к относительным величинам интенсивности, представляющим результат сравнения численностей, т. е. показателям смертности, ибо вероятнее всего, что кривая смертности представляет собой более или менее правильную плавную кривую. Для проверки результатов можно рекомендовать выравнивание и абсолютных чисел живущих и умерших и относительных показателей смертности с последующим сопоставлением полученных результатов.

Если к методам выравнивания не предъявлять чрезмерных требований, противоречащих самой их идее, то можно принять в качестве аксиомы, что выравнивание является научным приемом и может быть широко использовано в познавательных целях. Основная задача состоит в том, чтобы выбрать такой метод, который, учитывая специфичность имеющихся данных, давал бы все-таки некоторые гарантии того, что выровненные данные полностью или частично свободны от неправильностей фактических данных.

Некоторые исследователи идут по пути усложнения используемых методов, наивно полагая при этом, что чем сложнее применяемый метод, тем больше гарантия правильности приближения выровненных данных к действительности. Между тем выдающийся советский статистик и актуарий Б. С. Ястремский писал, что «...нельзя измерять длину резинового жгута прибором, рассчитанным на точность до одного микрона, или подносить ко рту стакан чаю при помощи домкрата»[1].

Многие исследователи на основании изучения обширного материала сделали заключение, что усложненные формулы могли бы иногда точнее описать зависимость демографических явлений от возраста, но это уточнение не в состоянии компенсировать неудобства, связанные с усложнением. Такая неопределенность при выборе того или иного метода выравнивания отражает, по сути дела, неопределенность самой задачи выравнивания. Можно привести несколько кривых линий, каждая из которых будет с конкретной точки зрения самой близкой к фактическим данным. Соответствие найденных теоретических кривых эмпирическим вовсе не доказывает правильности исходных теоретических позиций, так как такое же соответствие на данном конкретном материале может получиться и при других гипотезах. Этот очень важный для выравнивания демографических показателей вывод часто игнорируется многими исследователями.

Если поставить задачу подобрать кривую, наименее отклоняющуюся (соблюдая принцип наименьших квадратов) от эмпирической, то нетрудно доказать, что эта кривая может быть элементарной алгебраической функцией не выше 4-го (или максимум 5-го) порядка.

При этом процесс выравнивания, т. е. нахождения выровненных значений вместо фактических, будет одновременно также и процессом интерполяции уровней.

Вопрос о выравнивании или интерполировании функции вообще поднимался давно. Что касается методов, применяемых в демографии, то, например, для выравнивания кривой смертности многие математики искали такие алгебраические формулы, которые давали бы плавную кривую линию, близкую к линии смертности, получаемой из таблицы смертности, принятой за эталон.

Первая гипотеза была построена Моавром на основании таблицы смертности, составленной Галлеем. Используя неполные данные Бреславля (1886—1891 гг.), Моавр делает вывод о том, что число доживающих снижается по прямой линии, и, следовательно, в качестве плавной линии, выражающей результат выравнивания, может быть принята прямая. Однако привлечение статистических данных показало в дальнейшем, что закон смертности более сложен, чем это представляется по отрывочным и несовершенным данным о смертности.

В основе всех методов выравнивания лежит идея «плавности» выровненной линии. Такая идея могла оказаться и оказалась весьма полезной в тех случаях, когда выравниванию подвергаются участки, относительно которых исследователь уверен, что в них отсутствуют резкие изломы, или такие показатели, которые по своей природе могут изменяться только плавно. Между тем сплошь и рядом выравнивание по сгруппированным данным применяется без учета вышесказанного. Некоторым оправданием тому может служить тот факт, что порой трудно, а иногда и невозможно отличить плавное изменение показателей от неплавного или найти критерий возможности использования методов выравнивания.

[1] Б. С. Ястремский. Труды по статистике. М., Редакционно-издательское управление ЦУНХУ Госплана СССР и В/О «Союзоргучет», 1937, стр. 297.

Возникает два вопроса. Насколько правильно, т. е. научно аргументированно, произведена возрастная группировка? Какой из известных методов выравнивания следует для данного случая считать наиболее приемлемым?

При ответе на первый вопрос следует учитывать главные, существенные черты изучаемой совокупности, а также национальные особенности. Так, для населения СССР характерна аккумуляция на возрастах, кратных пяти. Поэтому наилучшей группировкой следует признать такую, при которой на каждый возрастной интервал приходится только один аккумуляционный возраст. Очевидно, наиболее удачным будет то решение, при котором этот возраст не приходится на начало или конец интервала. Следовательно, наилучшая группировка населения с интервалом в пять лет с помещением аккумуляционного возраста в середину интервала.

Иногда при затруднениях выбора границ интервалов можно использовать сразу 5 способов группировки населения по возрасту:

1—5, 6—10, 11—15 лет и т. д.
2—6, 7—11, 12—16 лет и т. д.
3—7, 8—12, 13—17 лет и т. д.
4—8, 9—13, 14—18 лет и т. д.
5—9, 10—14, 15—19 лет и т. д.

По каждой группировке проводится выравнивание показателей т. е. интерполяция всех промежуточных возрастов. Получается пять выровненных рядов показателей, относящихся к одногодичным интервалам.

Из каждых пяти показателей, относящихся к одному году, вычисляются средние, которые и образуют окончательный ряд выровненных показателей. Этот метод был использован при обработке материалов переписи населения 1926 г. по сельскому населению Ленинградско-Карельского района. Выравниванию подверглись вероятности умереть и находились выровненные (интерполированные) значения для каждого возраста, т. е. $\overline{q}_x$.

Например, для мужского населения в возрасте 20 лет при первой системе группировки, т. е. 1—5, 6—10, 11—15 и т. д. лет, выровненное значение вероятности умереть составило $q_{20} = 0{,}00574$, при второй системе группировки — 0,00615, при третьей — 0,00638, при четвертой — 0,00625, при пятой — 0,00576. Отсюда $\overline{q}_{20} = 0{,}00606$.

Чтобы ответить на второй вопрос, необходимо решить проблему оценочных критериев. Можно предложить два критерия. Один должен оценивать степень плавности ряда, полученного при выравнивании, а второй — степень соответствия выровненных данных фактическим. Эти два критерия несколько противоречат друг другу.

Дело в том, что фактические данные демографических явлений и процессов, связанные с возрастными группировками и подвергающиеся выравниванию на сравнительно небольших участках, отличаются значительной плавностью и совпадение ряда с фактическим неплавным обычно не имеет места.

При стремлении как можно полнее удовлетворить первое требование, мы ухудшаем соответствие между выровненными и фактическими данными. Стремление же максимально приблизить выровненный ряд к фактическому сводит на нет стремление к «плавности».

Выявить, насколько выровненный ряд является плавным, несложно. Для этого составляют следующую таблицу:

Таблица 56

Разности различных порядков

Уровни y	Разности			
	Первая $\Delta^1 y$	Вторая $\Delta^2 y$	Третья $\Delta^3 y$	Четвертая $\Delta^4 y$
y_1 y_2 y_3 y_4 y_5	$\Delta^1 y_1$ $\Delta^1 y_2$ $\Delta^1 y_3$ $\Delta^1 y_4$	$\Delta^2 y_1$ $\Delta^2 y_2$ $\Delta^2 y_3$	$\Delta^3 y_1$ $\Delta^3 y_2$	$\Delta^4 y_1$

Исследуется уменьшение величины разностей. Чем ниже порядок исчезающих разностей, тем лучше выравнивание. В соответствии с конечными разностями при выравнивании по прямой линии вторые разности равны нулю, при выравнивании по параболе 2-го порядка третьи разности равны нулю и т. д. Значит, каждое повышение порядка используемой для выравнивания параболы дает еще одно значение $\Delta^n y$.

Опыт показывает, что если абсолютные величины разностей 3-го порядка малы и знаки их часто меняются, это является признаком «плавности».

Для наилучшего «примыкания» выровненных показателей друг к другу или к фактическим в местах стыковки используются соприкасательные параболы, так как между стыками выровненные данные дают разности первых трех или четырех порядков, а в местах стыков их изменение происходит скачками.

Произведенные расчеты показывают, что обычно использование соприкасательных парабол приводит к тому, что скачки в первых разностях исчезают совершенно, а во вторых и третьих становятся при построении таблиц смертности малозаметными.

Вопрос о замене фактических показателей смертности выровненными специально рассматривался Б. С. Ястремским. В результате было сформулировано два положения:

«То предположение, что *значения вероятности* **y** *должны уменьшаться на некоторой плавной кривой, является не только естественным, но и логически необходимым*»[1];

«*Замена скачущего ряда частостей* w *плавным рядом* **y** *представляет собой логически необходимое завершение процесса обработки данных о смертности. Мы нашли если не наивероятнейший ход изменения смертности во времени, то весьма близкое к истине приближение*»[2].

Что же касается соответствия выровненных данных фактическим, то в качестве меры такого соответствия при выравнивании, например, повозрастных коэффициентов смертности может служить близость вычисленных абсолютных чисел, т. е. ожидаемых смертных случаев, к действительным числам в каком-нибудь возрастном интервале.

Чтобы исчислить ожидаемое число умерших в каждом возрасте, надо непосредственные числа живущих (по данным переписи) умножить на соответствующие коэффициенты смертности.

Выбирается достаточно широкий возрастной интервал и найденные числа умерших суммируются в пределах выбранного интервала. В пределах этого же интервала суммируются фактические числа умерших. Найденные суммы ожидаемых и наблюдаемых чисел умерших сопоставляют.

Произведенные расчеты показывают, что до возрастов 70—75 лет числа смертных случаев, вычисленных по выровненным данным, близки к фактическим. В возрастах же старше 70—75 лет расчетные данные значительно отличаются от фактических. Это дало основание В. В. Паевскому и С. А. Новосельскому, использовавшим при обработке материалов переписи 1926 г. этот метод, утверждать, что в старческих возрастах, помимо дефектов самого метода выравнивания, большое значение имеет качество исходного материала, привлекаемого для выравнивания.

Для примера в табл. 57 сопоставим фактические и ожидаемые числа умерших мужчин европейской части СССР по таблицам смертности 1926—1927 гг.[3] (интервал семилетний).

Общее расхождение +0,148% весьма невелико. Что же касается возрастов от 5 до 75 лет, то в них совпадение очень близкое и расхождение составляет всего 0,07%. Для старческих же возрастов расхождение достигает 1,6%.

Иногда приходится мириться и с бо́льшим расхождением. В связи с противоречиями между двумя критериями возникает необходимость поисков оптимального (в данном случае компромиссного) решения,

[1] Б. С. Ястремский. Дисперсионный метод изучения эволюции смертности. СПб., 1913, стр. 22.

[2] Там же, стр. 24.

[3] «Смертность и продолжительность жизни населения СССР. 1926—1927. Таблицы смертности», стр. XXXIV.

Таблица 57

Сопоставление ожидаемого и действительного числа умерших

Возрастные группы (в годах)	Ожидаемое число умерших	Действительное число умерших	Разность: абсолютная	Разность: в процентах к действительному числу умерших
5—11	93 945	93 764	+181	+0,19
12—18	61 316	61 159	+157	+0,26
19—25	81 236	81 655	—419	—0,51
26—32	64 791	64 724	+ 67	+0,10
33—39	61 192	61 083	+109	+0,18
40—46	77 080	77 583	—509	—0,66
47—53	85 056	85 336	—280	—0,33
54—60	106 985	106 586	+399	+0,37
61—67	108 953	109 302	—349	—0,32
68—74	92 423	93 373	+ 50	+0,05
75—81	73 686	72 225	+1 461	+ 2,02
82—88	33 569	35 405	—1 836	— 5,18
89—95	15 218	13 284	+1 934	+14,55
96—99	3 529	3 074	+ 455	+14,80
Всего: до 75	832 977	833 571	—594	—0,07
Всего: от 75 и старше	126 002	123 988	+2 014	+1,60
Всего: 5—99	958 979	957 559	+1 420	+0,148

при котором при наименьшем удалении выровненных данных от фактических наблюдалась максимальная (в данном случае достаточная) плавность.

Следует отметить, что при выравнивании данных, полученных при переписи населения 1926 г., предпочтение чаще отдавалось соответствию выровненных данных фактическим. Критерием пригодности используемого метода выравнивания для каждого конкретного случая может служить соблюдение требования о том, чтобы сумма выровненных данных в пределах какого-нибудь возрастного интервала, например 5 или 10 лет, совпадала с суммой фактических чисел умерших.

При подборе того или иного метода выравнивания, т. е. при выборе кривой, по которой производится выравнивание, во-первых, следует следить за тем, чтобы привлекаемая кривая была пригодна для сравнительно небольшого возрастного интервала, т. е. участка конкретных возрастов. Это необходимо для ограничения влияния далеко отстоящих возрастов на характер изменения показателей выравниваемого участка. Во-вторых, выбор метода определяется исходным допущением о характере изменения явления в 5- или 10-летнем возрастном интервале, подвергаемом выравниванию в целях интерполяции. Так, в случае самого простого допущения об относительной равномерности изменения фактических показателей в пределах группы следует применять прямолинейное выравнивание. Когда же делается предположение об отсутствии прямой пропорциональности между изменением величины выравниваемого демографического признака и возрастом, используют выравнивание не абсолютных уровней, а логарифмов этих уровней и т. д.

В тех случаях, когда применение математических формул при выравнивании почему-либо нежелательно, можно рекомендовать метод исправления повозрастных численностей живущих, пригодный для случаев, когда в распоряжении исследователя имеются данные переписи о возрастном составе населения и известно распределение умерших по одногодичным возрастным группам за два года (до переписи и после нее) либо за 4 года (по два с каждой стороны). Данные о распределении численностей живущего населения и распределении умерших по возрастам искажены аккумуляцией, однако в неравной степени. Возраст умерших несет меньшие следы аккумуляции, чем возраст живущих. Допуская, что повозрастная смертность не изменяется, т. е. остается такой, какой она была в год переписи, можно найти среднюю возрастную численность умерших путем суммирования повозрастных численностей умерших за ряд последовательных лет и деления на число лет наблюдения. Таким образом, получается величина d_x. Разделив d_x на коэффициенты смертности m_x, получим число живущих в возрасте x лет. Этот метод довольно успешно был использован В. В. Паевским и С. А. Новосельским.

МЕХАНИЧЕСКОЕ ВЫРАВНИВАНИЕ

Конечной целью выравнивания является получение плавных уровней и последующее сопоставление их с фактическими. При этом предполагается, что наблюдаемые значения в пределах группы следуют какому-то закону, который определяется с помощью средней арифметической некоторых фактических уровней. Следовательно, если взять какой-нибудь фактический уровень и одинаковое число соседних с ним с той и другой стороны, можно сконструировать множество различных формул выравнивания. Общим в этих формулах является то, что полученные с их помощью результаты являются симметричными по отношению к центральному уровню.

Такими являются формулы, полученные методом невзвешенной (простой) и взвешенной скользящей средней; методом, основанным на привлечении численностей 15 (Воольхауза) и 20 (Карупа) одногодичных интервалов; методом Хайэма.

Метод невзвешенных и взвешенных скользящих средних

Пусть a_x фактическое число живущих или умерших в возрасте x лет, подлежащее выравниванию. В качестве периода выравнивания может быть взято произвольное число членов[1]. Если взять 5-членный период, то за первое выровненное значение принимается среднее из 5 смежных a_x, из которых два предшествуют выравниваемому и два следуют за ним. Тогда

$$a_x' = \frac{a_{x-2} + a_{x-1} + a_x + a_{x+1} + a_{x+2}}{5},$$

$$a_{x-1}' = \frac{a_{x-3} + a_{x-2} + a_{x-1} + a_x + a_{x+1}}{5}.$$

$$a_{x-2}' = \frac{a_{x-4} + a_{x-3} + a_{x-2} + a_{x-1} + a_x}{5}$$

и т. д. Найденные средние и будут представлять собой невзвешенные скользящие средние.

Применяя этот же прием вторично, находим путем скольжения среднюю из 5 найденных уже скользящих средних, т. е. второе выровненное значение a_x''.

Подставляя значения a_x', получаем скользящие средние уже для девятичленных периодов:

$$a_x'' = \frac{a_{x-4} + 2a_{x-3} + 3a_{x-2} + 4a_{x-1} + 5a_x + 4a_{x+1} + 3a_{x+2} + 2a_{x+3} + a_{x+4}}{25}.$$

Преобразуя, получаем:

$$a_x'' = \frac{(a_{x-4} + a_{x+4}) + 2(a_{x-3} + a_{x+3}) + 3(a_{x-2} + a_{x+2}) + 4(a_{x-1} + a_{x+1}) + 5a_x}{25}.$$

Применяя еще раз скольжение, аналогичным путем получим среднюю для 13-членного периода:

$$a_x''' = \frac{(a_{x-6} + a_{x+6}) + 3(a_{x-5} + a_{x+5}) + 6(a_{x-4} + a_{x+4}) + 10(a_{x-3} + a_{x+3})}{125} + \frac{15(a_{x-2} + a_{x+2}) + 18(a_{x-1} + a_{x+1}) + 19a_x}{125}$$

или

$$a_x''' = 0{,}152a_x + 0{,}008(a_{x-6} + a_{x+6}) + 0{,}024(a_{x-5} + a_{x+5}) + 0{,}048(a_{x-4} + a_{x+4}) + 0{,}080(a_{x-3} + a_{x+3}) + 0{,}120(a_{x-2} + a_{x+2}) + 0{,}144(a_{x-1} + a_{x+1}).$$

[1] См. И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики, стр. 214 (в качестве такого периода берутся два члена).

Использование формулы взвешенной скользящей средней с 13-членным периодом может быть облегчено применением трафарета со следующими весами:

a_{x-6}	1
a_{x-5}	3
a_{x-4}	6
a_{x-3}	10
a_{x-2}	15
a_{x-1}	18
a_x	19
a_{x+1}	18
a_{x+2}	15
a_{x+3}	10
a_{x+4}	6
a_{x+5}	3
a_{x+6}	1
Итого	125

Этот трафарет прикладывается (слева) к фактическим данным, подлежащим выравниванию. Умножая фактические данные на веса (коэффициенты трафарета), находим сумму произведений. Первую сумму записываем против коэффициента 19, затем передвигаем трафарет на строку вниз и умножением фактических данных на коэффициенты трафарета находим вторую сумму и т. д. Все найденные суммы делятся на 125 (сумму всех весов). Полученные скользящие средние покажут плавный выровненный уровень. Этот способ выравнивания называют методом Финлезона.

Если ограничиться взвешенной скользящей средней с 9-членным периодом, то трафарет будет с другими весами: 1, 2, 3, 4, 5, 4, 3, 2, 1. В этом случае все найденные суммы делятся на 25.

Эти формулы применялись С. А. Новосельским при построении таблиц смертности для разных возрастных контингентов. Например, число живущих мужчин в одногодичных возрастах 20—114 лет выравнивалось по взвешенной 13-членной скользящей средней. Для периода 9—20 лет использовалась взвешенная скользящая средняя с 9-членным периодом, а числа живущих в возрасте 0—9 лет выравнивались по невзвешенной скользящей средней с 5-членным периодом.

Используя в качестве периода два члена и производя дальнейшие скольжения, можно получить следующие веса для взвешивания, образующие биномиальные коэффициенты (см. табл. 58).

Производя центрирование, получаем уже одиннадцать весов:

+		1;	9;	36;	84;	126;	126;	84;	36;	9;	1
	1;	9;	36;	84;	126;	126;	84;	36;	9;	1;	
	1;	10;	45	120;	210;	252;	210	120;	45;	10;	1

Таблица 58

Биномиальные коэффициенты для взвешивания

Число членов периода	Коэффициенты для взвешивания
3	1 2 1
4	1 3 3 1
5	1 4 6 4 1
6	1 5 10 10 5 1
7	1 6 15 20 15 6 1
8	1 7 21 35 35 21 7 1
9	1 8 28 56 70 56 28 8 1
10	1 9 36 84 126 126 84 36 9 1

Такой способ десятилетней центрированной скользящей средней взвешенной по биномиальным коэффициентам использовался для выравнивания коэффициентов чисел живущих и умерших в возрасте 20—44 года при построении таблиц смертности и средней продолжительности жизни населения СССР 1958—1959 гг.

Метод, основанный на привлечении показателей 15 одногодичных интервалов (метод Воольхауза)

Взяв численность 15 одногодичных интервалов, можно вывести формулу для выравнивания среднего показателя:

$$\bar{y}_x = a_0 w_x + a_1 (w_{x+1} + w_{x-1}) + a_2 (w_{x+2} + w_{x-2}) + \ldots + a_n (w_{x+n} + w_{x-n}).$$

При этом $a_0 + 2(a_1 + a_2 + \ldots + a_n) = 1$.

Рассмотрим соображения, приводящие к этой формуле.

Для выравнивания показателя, относящегося к срединному возрасту, привлекаем до него и после него по 7 значений соседних возрастов и получаем 15 равноотстоящих значений:

$$w_{-7},\ w_{-6},\ w_{-5},\ \ldots\ w_0;\ w_1;\ \ldots;\ w_5;\ w_6;\ w_7.$$

Через каждые три значения, отделенные 5-летним интервалом, проводим параболу 2-го порядка и получаем 5 парабол. Эти параболы объединят следующие точки:

$$\left.\begin{matrix} w_{-7} \\ w_{-2} \\ w_3 \end{matrix}\right\} \quad \left.\begin{matrix} w_{-6} \\ w_{-1} \\ w_4 \end{matrix}\right\} \quad \left.\begin{matrix} w_{-5} \\ w_0 \\ w_5 \end{matrix}\right\} \quad \left.\begin{matrix} w_{-4} \\ w_1 \\ w_6 \end{matrix}\right\} \quad \left.\begin{matrix} w_{-3} \\ w_2 \\ w_7 \end{matrix}\right\}.$$

Для нахождения y_0, т. е. выровненного значения срединного пятилетия, находим среднее арифметическое из 5 значений.

Полагая уравнение параболы с вершиной n и обозначая значения, расположенные по обе стороны от выравниваемого, приравненного нулевому, ± 1 и ± 2, получим систему трех уравнений:

$$\begin{cases} w_{n-5} = a_n + b_n (n-5) + c_n (n-5)^2; \\ w_n = a_n + b_n n + c_n n^2; \\ w_{n+5} = a_n + b_n (n+5) + c_n (n+5)^2. \end{cases}$$

Решая ее, находим:

$$a_n = w_n \frac{(5-n)(5+n)}{25} + w_{n-5} \frac{n(n+5)}{50} + w_{n+5} \frac{n(n-5)}{50};$$

$$b_n = \frac{1}{5a} \Delta w_{n-5} - \frac{1}{50a^2} \Delta^2 w_{n-5} (2n-5);$$

$$c_n = \frac{1}{50a^2} \Delta^2 w_{n-5}.$$

Каждая парабола даст для любого значения x выровненное значение, а 5 парабол для этого же значения x дадут 5 точек. При $x = 0$ значения функции будут a_{-2}, a_{-1}, a_0, a_1, a_2. Находя среднюю арифметическую, получаем:

$$\bar{y}_0 = \frac{1}{5} (a_{-2} + a_{-1} + a_0 + a_1 + a_2).$$

В конечном итоге после подстановки w формула примет вид:

$$\bar{y}_x = \frac{1}{125} [25 w_x + 24 (w_{x+1} + w_{x-1}) + 21 (w_{x+2} + w_{x-2}) + 7 (w_{x+3} + w_{x-3}) + 3 (w_{x+4} + w_{x-4}) - 2 (w_{x+6} + w_{x-6}) - 3 (w_{x+7} + w_{x-7})].$$

Эта формула может быть преобразована разными методами. Разберем один из них.

Возьмем сумму 5 соседних членов с w_{x+k} в середине:

$${}_k\Sigma_5 = w_{x+k} + (w_{x+k+1} + w_{x+k-1}) + (w_{x+k+2} + w_{x+k-2}) + \ldots;$$
$${}_0\Sigma_5 = w_{x-2} + w_{x-1} + w_x + w_{x+1} + w_{x+2};$$

для повторной функции:

$${}_0\Sigma_{5;\,5} = {}_{-2}\Sigma_5 + {}_{-1}\Sigma_5 + {}_0\Sigma_5 + {}_1\Sigma_5 + {}_2\Sigma_5 = 5 w_x + 4 (w_{x+1} + w_{x-1}) + 3 (w_{x+2} + w_{x-2}) + 2 (w_{x+3} + w_{x-3}) + (w_{x+4} + w_{x-4});$$

наконец, для двукратно-повторной функции:

$${}_0\Sigma_{5;\,5;\,5} = {}_{-2}\Sigma_{5;\,5} + {}_{-1}\Sigma_{5;\,5} + {}_0\Sigma_{5;\,5} + {}_1\Sigma_{5;\,5} + {}_2\Sigma_{5;\,5} =$$

$$= 19w_x + 18(w_{x+1} + w_{x-1}) + 15(w_{x+2} + w_{x-2}) + 10(w_{x+3} + w_{x-3}) +$$
$$+ 6(w_{x+4} + w_{x-4}) + 3(w_{x+5} + w_{x-5}) + (w_{x+6} + w_{x-6}).$$

Заменяя переменные, можно получить формулу

$$\bar{y}_x = \frac{1}{125}\left[10\,{}_0\Sigma_{5;\,5;\,5} - 3\left({}_{-1}\Sigma_{5;\,5;\,5} + {}_0\Sigma_{5;\,5;\,5}\right)\right].$$

Одним из вариантов этой формулы пользовался В. И. Гребенщиков. Выравнивая числа доживающих таблицы смертности для русских врачей от возраста 30 лет и привлекая для этого данные с 24 лет, В. И. Гребенщиков использовал следующую модифицированную формулу Воольхауза:

$$l_x = 0{,}2\,(\lambda_x + v_1 + v_2) - 0{,}008\,(f - 4q - 2h + 3k),$$

где l_x — исправленные, т. е. выровненные числа доживающих;
λ_x — фактическое число доживающих до возраста x лет.

$$v_1 = \lambda_{x-1} + \lambda_{x+1}; \quad v_2 = \lambda_{x-2} + \lambda_{x+2};$$
$$v_3 = \lambda_{x-3} + \lambda_{x+3}; \quad f = v_1 - v_3;$$
$$v_4 = \lambda_{x-4} + \lambda_{x+4}; \quad q = v_2 - v_4;$$
$$v_6 = \lambda_{x-6} + \lambda_{x+6}; \quad h = v_6 - v_4;$$
$$v_7 = \lambda_{x-7} + \lambda_{x+7}; \quad k = v_7 - v_3.$$

Выравнивание же чисел доживающих в возрасте 24—30 лет проводилось В. И. Гребенщиковым по другой формуле того же автора:

$$l_{7-n} = l_7 + \frac{n(n+3)}{2}\,d_7 - \frac{n(n-1)}{2}\,d_8 - \frac{n(n-1)(n+2)}{1\cdot 2\cdot 3}\,\Delta,$$

где l_7 — первое выровненное число умерших при переходе от возраста 30 лет к 31 году, а d_8 — первое выровненное число умерших при переходе от возраста 31 год к 32. При этом

$$\Delta = \frac{1}{12}\left(-\frac{\lambda_0 - \lambda_7}{7} + 5d_7 - 4d_8\right), \quad \text{где } \lambda_0 = 10\,000.$$

Придавая n в рассматриваемой формуле все значения от 1 до 7, получаем l_x для x от 25 до 30 лет.

Используя данные о числе доживающих по «неисправленной» таблице смертности для врачей России, В. И. Гребенщиков произвел расчеты по схеме, приведенной в табл. 59, и получил выровненные числа доживающих начиная с возраста 30 лет и старше.

Привлекая некоторые показатели l_x и d_x, полученные при выравнивании возрастов от 30 лет и старше, В. И. Гребенщиков произвел выравнивание возрастов 24—30 лет.

Таблица 59

Макет таблицы для расчета «исправленного» числа доживающих и умирающих

Возраст от 24 лет и старше, x	По неисправленной таблице доживает λ_x	$v_1 = \lambda_{x-1} + \lambda_{x+1}$	$v_2 = \lambda_{x-2} + \lambda_{x+2}$	$v_3 = \lambda_{x-3} + \lambda_{x+3}$	$v_4 = \lambda_{x-4} + \lambda_{x+4}$	$v_6 = \lambda_{x-6} + \lambda_{x+6}$	$v_7 = \lambda_{x-7} + \lambda_{x+7}$	$f = v_1 - v_3$	$q = v_2 - v_4$	$h = v_6 - v_4$	$k = v_7 - v_3$	$4q$	$2h$	$3k$	$0{,}2\,(\lambda_x + v_1 + v_2)$	«Исправленное» число	
																доживающих l_x	умерших d_x

Метод, основанный на привлечении показателей 20 одногодичных интервалов (метод Карупа)

Выравнивание этим методом производится по формуле, вывод которой аналогичен выводу предыдущей формулы (Воольхауза). Данный метод предназначен для выравнивания показателя десятого возраста вместо восьмого и привлечения для этого показателей 20 возрастов. Все привлеченные возраста разбиваются на 5 четверок, выравнивание производится по параболе 3-го порядка: $\bar{y}_x = a_0 + a_1x + a_2x^2 + a_3x^3$. Каждая из парабол проходит через 4 точки:

$$\begin{vmatrix} w_{-9} \\ w_{-4} \\ w_1 \\ w_6 \end{vmatrix} \quad \begin{vmatrix} w_{-8} \\ w_{-3} \\ w_2 \\ w_7 \end{vmatrix} \quad \begin{vmatrix} w_{-7} \\ w_{-2} \\ w_3 \\ w_8 \end{vmatrix} \quad \begin{vmatrix} w_{-6} \\ w_{-1} \\ w_4 \\ w_9 \end{vmatrix} \quad \begin{vmatrix} w_{-5} \\ w_0 \\ w_5 \\ w_{10}. \end{vmatrix}$$

Параболы 3-го порядка дадут для y_0 некоторые значения. Средняя арифметическая из пяти вычисленных значений и будет выровненной величиной.

Перенесем 10-е значение возраста в абсциссу, равную нулю, и решим для каждой параболы, проходящей через n и $n + 5$, систему уравнений:

$$w_n = a_n + b_n\,n + c_n\,n^2 + d_n\,n^3;$$
$$w_{n+5} = a_n + b_n(n+5) + c_n(n+5)^2 + d_n(n+5)^3.$$

В свою очередь

$$w_{n+5} = w_n + 5w'_n + \frac{25}{2}w''_n \quad \text{и}$$

$$w_{n-5}=w_n - 5w_n'+\frac{25}{2}w_n'',$$

где w_n' и w_n'' — производные.

Тогда получим

$$w_n'=\frac{w_{n+5}-w_{n-5}}{10}=b_n+2c_n n+3d_n n^2;$$

$$w_{n+5}'=\frac{w_{n+10}-w_n}{10}=b_n+2c_n(n+5)+3d_n(n+5)^2.$$

Обозначим разности разных порядков:

$$\Delta w_{n-5}=w_n-w_{n-5};$$

$$\Delta^2 w_{n-5}=w_{n+5}-2w_n+w_{n-5};$$

$$\Delta^3 w_{n-5}=w_{n+10}-3w_{n+5}+3w_n-w_{n-5}.$$

Решая систему, получим следующие значения параметров:

$$a_n=w_n-\frac{n}{5}\Delta w_n+\frac{n(n+5)}{50}\Delta^2 w_{n-5}-\frac{n^2(n+5)}{250}\Delta^3 w_{n-5};$$

$$b_n=\frac{1}{10}\left[2\Delta w_n-\frac{1}{5}\Delta^2 w_{n-5}(2n+5)+\frac{n}{25}\Delta^3 w_{n-5}(3n+10)\right];$$

$$c_n=\frac{1}{50}\left[\Delta^2 w_{n-5}-\frac{1}{5}\Delta^3 w_{n-5}(3n+5)\right];$$

$$d_n=\frac{\Delta^3 w_{n-5}}{250}.$$

В каждой из парабол значение a_n дает выровненное значение 10-го наблюдения. Ограничиваясь a_{-4}, a_{-3}, a_{-2}, a_{-1}, a_0, найдем из них среднюю арифметическую:

$$\bar{y}_x=\frac{1}{5}(a_{-4}+a_{-3}+a_{-2}+a_{-1}+a_0).$$

Подставляя вместо a соответствующие значения w, в конечном итоге получим формулу

$$\bar{y}_x=\frac{1}{1\,250}[250w_x+228(w_{x+1}+w_{x-1})+174(w_{x+2}+w_{x-2})+$$
$$+106(w_{x+3}+w_{x-3})+42(w_{x+4}+w_{x-4})-16(w_{x+6}+w_{x-6})-$$
$$-18(w_{x+7}+w_{x-7})-12(w_{x+8}+w_{x-8})-4(w_{x+9}+w_{x-9})].$$

Этой формуле идентична следующая:

$$\bar{y}_x=\frac{1}{1250}\Big[6({}_{-1}\Sigma_{5;\,5;\,5}+{}_0\Sigma_{5;\,5;\,5}+{}_1\Sigma_{5;\,5;\,5})-$$
$$-4({}_{-3}\Sigma_{5;\,5;\,5}+{}_3\Sigma_{5;\,5;\,5})\Big].$$

Метод Хайэма

Хайэм вывел формулу, обозначив показатели выравниваемых возрастов u_0, u_1, u_2, ... и получая p-й выровненный уровень по формуле $(1+\Delta)^{p-1}u_0$. Сумму уровней $u_0+u_1+u_2+\ldots+u_{p-1}$ можно записать в виде:

$$\frac{(1+\Delta)^{p-1}}{\Delta}u_0.$$

Второе сложение даст следующие члены:

$$\frac{(1+\Delta)^{p-1}}{\Delta}u_0;\quad \frac{(1+\Delta)^{p-1}}{\Delta}u_1;\ \ldots;\quad \frac{(1+\Delta)^{p-1}}{\Delta}u_{p-1}.$$

Как результат q членов имеем:

$$\left[\frac{(1+\Delta)^{p+q}-(1-\Delta)^p}{\Delta^2}-\frac{(1+\Delta)^{q-1}}{\Delta^2}\right]u_0=$$
$$=\left[\frac{(1+\Delta)^{p-1}}{\Delta}\cdot\frac{(1+\Delta)^{q-1}}{\Delta}\right]u_0.$$

Сложив в полученных таким же образом членах каждые r соседних членов, в итоге получаем:

$$\left[\frac{(1+\Delta)^{p-1}}{\Delta}\cdot\frac{(1+\Delta)^{q-1}}{\Delta}\cdot\frac{(1+\Delta)^{r-1}}{\Delta}\right]u_0.$$

Сравнивая каждый из заключительных результатов с

$$pu_{\frac{p-1}{2}}=p(1+\Delta)^{\frac{p-1}{2}}u_0;$$

$$pqu_{\frac{p+q-2}{2}}=pq(1+\Delta)^{\frac{p+q-2}{2}}u_0;$$

$$pqru_{\frac{p+q+r-3}{2}}=pqr(1+\Delta)^{\frac{p+q+r-3}{2}}u_0,$$

получим:

$$\frac{(1+\Delta)^{p_1-1}}{\Delta}u_i=$$
$$=p_1\left[(1+\Delta)^{\frac{p_1-1}{2}}+\frac{p_1^2-1}{1\cdot2\cdot3\cdot4}\Delta^2+\frac{(p_1^2-1)(p_1-3)}{1\cdot2\cdot4\cdot6}\Delta^3+\ldots\right]u_0=$$
$$=\frac{(1+\Delta)^{p_1-1}}{\Delta}\cdot\frac{(1+\Delta)^{p_2-1}}{\Delta}u_i=p_1p_2\times$$
$$\times\left[(1+\Delta)^{\frac{p_1+p_2-2}{2}}+\frac{p_1^2+p_2^2-2}{1\cdot2\cdot3\cdot4}\Delta^2+\frac{(p_1^2+p_2^2-2)(p_1+p_2-4)}{2\cdot4\cdot6}\Delta^3+\ldots\right]u_i.$$

Обозначим:

$$p_1 p_2 p_3 \dots p_n = M;$$
$$p_1 + p_2 + p_3 + \dots + p_n = A;$$
$$p_1^2 + p_2^2 + p_3^2 + \dots + p_n^2 = B.$$

Преобразуем:

$$\frac{(1+\Delta)^{p_1-1}}{\Delta} \cdot \frac{(1+\Delta)^{p_2-1}}{\Delta} \dots \frac{(1+\Delta)^{p_n-1}}{\Delta} u_i =$$
$$= M\left[(1+\Delta)^{\frac{A-n}{2}} + \frac{B-n}{1\cdot2\cdot3\cdot4}\left(\Delta^2 + \frac{A-n+2}{2}\Delta^3\right) + \dots\right] u_0.$$

Произведем повторное сложение B членов и в качестве формулы получим:

$$u_{\frac{A-n}{2}} = \frac{\sum p_1 p_2 \dots p_n}{M} - \frac{B-n}{1\cdot2\cdot3\cdot4}\left(\Delta^2 + \frac{A-n+2}{2}\Delta^3\right) u_3,$$

где $u_{\frac{A-n}{2}} = (1+\Delta)^{\frac{A-n}{2}} u_0$.

Взяв при повторном сложении тот же период, что и при первом, в результате преобразований получаем, например, следующую формулу выравнивания:

$$u_8 = 0{,}2106\, u_0 + 0{,}192\,(u_1 + u_{-1}) + 0{,}14186\,(u_2 + u_{-2}) +$$
$$+ 0{,}07946\,(u_3 + u_{-3}) + 0{,}024\,(u_4 + u_{-4}) - 0{,}0053\,(u_5 + u_{-5}) -$$
$$- 0{,}016\,(u_6 + u_{-6}) - 0{,}0144\,(u_7 + u_{-7}) - 0{,}00693\,(u_8 + u_{-8}).$$

АНАЛИТИЧЕСКОЕ ВЫРАВНИВАНИЕ. ИНТЕРПОЛЯЦИЯ И ЭКСТРАПОЛЯЦИЯ

Метод наименьших квадратов. Выравнивание по прямой линии и показательной функции

Если предположить, что наблюдаемые значения таких демографических показателей, как фактическое число живущих или умерших в определенном возрасте, или вероятности дожития или смерти, следуют какому-то закону, выраженному в виде определенной аналитической формулы, то перед исследователем возникает задача найти данную формулу.

Нахождение этой формулы называется *аналитическим выравниванием*, представляющим весьма удобный способ описания эмпирических данных. При этом обычно применяют ограниченное количество линий, по которым производится аналитическое выравнивание: прямую линию, параболы 2, 3, 4 или 5-го порядка и показательную (экспонентную) функцию. При выборе типа линии исходят из следующих положений:

если абсолютные приросты выравниваемых уровней, т. е. первые разности, колеблются около постоянной величины, то математической функцией, уравнение которой можно принять за основу аналитического выравнивания, следует считать прямую линию

$$\overline{y}_x = a_0 + a_1 x;$$

если приросты приростов уровней, т. е. ускорения или вторые разности, колеблются около постоянной величины, то за основу аналитического выравнивания следует признать параболу 2-го порядка:

$$\overline{y}_x = a_0 + a_1 x + a_2 x^2;$$

если же близки к постоянной величине третьи, четвертые или пятые разности, то за основу принимаются параболы более высоких порядков:

$\overline{y}_x = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ (парабола третьего порядка);
$\overline{y}_x = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$ (парабола четвертого порядка);
$\overline{y}_x = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$ (парабола пятого порядка).

Если уровни изменяются с приблизительным постоянным относительным приростом, то выравнивание производится по показательной (экспонентной) функции, например, по формуле

$$\overline{y}_x = a_0 a_1^x.$$

Для нахождения аналитического уравнения, по которому производится выравнивание, можно использовать несколько методов.

Первый из них (при выравнивании по прямой линии) состоит в расчленении всех выравниваемых показателей на две примерно равные части (например, от возраста x до $x + n$ и от $x + n$ до $x + 2n$), при этом сумма выровненных уровней в каждой части, а следовательно, и в общем должна совпадать с суммой фактических значений, или $\Sigma\ (y - \overline{y}_x) = 0$.

Предъявляя это требование к каждой из двух частей, получим систему двух уравнений с неизвестными параметрами a_0 и a_1.

При выравнивании по параболе 2-го порядка (с тремя параметрами) выравниваемые уровни расчленяются на три части и решается система трех уравнений и т. д.[1]

Второй метод — метод наименьших квадратов — основан на требовании о том, чтобы сумма квадратов отклонений фактических уровней от выровненных была наименьшей:

$$\sum (y - \overline{y}_x)^2 = \min.$$

[1] См. И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики, стр. 216—218.

Для нахождения параметров a_0, a_1, a_2 и т. д. следует решать систему нормальных уравнений. Так, для выравнивания по прямой линии система нормальных уравнений примет вид:

$$\begin{cases} a_0 n + a_1 \sum x = \sum y, \\ a_0 \sum x + a_1 \sum x^2 = \sum xy. \end{cases}$$

Применяя особый способ отсчета возраста, чтобы $\Sigma x = 0$, можно добиться значительных упрощений. Если число возрастов нечетное, то отсчет x может производиться от срединного возраста. При этом отсчете значение срединного возраста принимают равным нулю. Младшие возраста имеют отрицательное значение (-1, -2, -3 и т. д.), а старшие — положительные ($+1$, $+2$, $+3$ и т. д.). Если же число возрастов четное, то отсчет производится от срединной пары возрастов и значения x для этой пары принимают -1 и $+1$, а далее для молодых идут значения -3, -5, -7 и т. д., а для старших $+3$, $+5$, $+7$ и т. д.

Интерполяцией в демографии называется выравнивание известных данных, относящихся к крайним возрастам, и нахождение уровней, относящихся к промежуточным. При этом в качестве основы выравнивания принимается какое-нибудь правдоподобное предположение (гипотеза) о порядке изменения уровней, относящихся к промежуточным возрастам.

Выровненные, т. е. интерполированные, уровни могут либо исправлять фактические уровни, уже известные, но искаженные в силу каких-то причин, либо давать значение уровней возрастов, для которых они совсем не были известны.

Экстраполяцией же называют выравнивание уровней, относящихся к определенным возрастам, для нахождения уровней, относящихся к последующим возрастам — более старшим (или предшествующим — более младшим), чем те, на основе которых производится выравнивание.

При этом в качестве основного предположения принимается гипотеза о том, что за пределами возрастов, принятых за опорные, действуют те же законы, что и для опорных.

При выравнивании числа живущих или умерших мы уже предполагали, что $\Sigma\ (y - \bar{y}_x) = 0$, где $\bar{y}_x$ — выровненный, т. е. плавный уровень. Если учесть, что у каждого возрастного интервала могут быть свои плавные уровни, то, следовательно, смежные периоды отличаются угловыми коэффициентами (наклоном к оси x).

Если предположить, что плавные уровни каждого периода не очень сильно отличаются от прямой линии, то в качестве функции, выражающей плавный уровень, можно было бы принять параболу, имеющую столько изгибов, сколько периодов в изучаемом интервале.

Признавая, что парабола — очень удобная функция для интерполирования (это подтверждается опытом и практикой интерполирования), следует отметить, что она лишь в редких случаях пригодна для экстраполяции.

Какими же свойствами должна обладать кривая, которая, на наш взгляд, годилась бы и для интерполирования, и для экстраполирования? Во-первых, сумма плавных периодов нескольких участков этой кривой должна равняться общей сумме периода, охватывающего все участки.

Это значит, что если имеются плавные уровни участков, то их сумма должна дать и дает:

$$\bar{y}_{x\,(1)} = a_{0\,(1)} + a_{1\,(1)}\,x + a_{2\,(1)}\,x^2 + \ \dots\ + a_{n\,(1)}\,x^n,$$

$$\bar{y}_{x\,(2)} = a_{0\,(2)} + a_{1\,(2)}\,x + a_{2\,(2)}\,x^2 + \ \dots\ + a_{n\,(2)}\,x^n,$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\sum \bar{y}_x = na_0 + x \sum a_1 + x^2 \sum a_2 + \ \dots\ + x^n \sum a_n.$$

При этом показатель степени общего результата такой же, как и для отдельных участков.

А вот чтобы получить произведение, парабола не годится. Для этого можно воспользоваться функцией

$$\bar{y}_x = 10^{a_0 + a_1 x + a_2 x^2 + \ \dots\ + a_n x^n}.$$

Обозначив $a_0 + a_1 x + a_2 x^2 + \ \dots\ + a_n x^n$ через z, имеем:

$$\bar{y}_{x\,(1)} = 10^{z_1}$$

$$\bar{y}_{x\,(2)} = 10^{z_2}$$

$$\dots\dots$$

$$\Pi\,(\bar{y}_x) = 10^{\Sigma z},$$ где Π — знак произведения.

Примем в качестве экстраполируемой кривой функцию $\bar{y}_x = 10^z$ на том основании, что ее можно легко преобразовать в маклореновский ряд.

Имеем: $\bar{y}_x = 10^z$. Логарифмируя, получаем:

$$\lg \bar{y}_x = z = a_0 + a_1 x + a_2 x^2 + \ \dots\ + a_n x^n.$$

Этот результат означает, что выравниванию следует подвергать не сами эмпирические уровни, а их логарифмы.

Используя метод наименьших квадратов и ограничиваясь параболой 3-го порядка, получим следующую систему нормальных уравнений, состоящую из 4 уравнений с 4 неизвестными параметрами a_0, a_1, a_2 и a_3:

$$\begin{cases} na_0 + a_1 \sum x + a_2 \sum x^2 + a_3 \sum x^3 = \sum \lg y, \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 + a_3 \sum x^4 = \sum x \lg y, \\ a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 + a_3 \sum x^5 = \sum x^2 \lg y, \\ a_0 \sum x^3 + a_1 \sum x^4 + a_2 \sum x^5 + a_3 \sum x^6 = \sum x^3 \lg y. \end{cases}$$

Можно упростить эту систему, принимая в качестве x одногодичный возраст и ведя отсчет от возраста x, принимаемого за условный нуль. Тогда $\Sigma x = 0$.

$$\begin{cases} na_0 + a_2 \sum x^2 = \sum \lg y, \\ a_1 \sum x^2 + a_3 \sum x^4 = \sum x \lg y, \\ a_0 \sum x^2 + a_2 \sum x^4 = \sum x^2 \lg y, \\ a_1 \sum x^4 + a_3 \sum x^6 = \sum x^3 \lg y. \end{cases}$$

Определив параметры, подставив их и найдя $\lg \overline{y}_x = z$, по логарифмам определим плавный ряд, т. е. $\overline{y}_x = 10^z$.

Экстраполяция и интерполяция путем накладывания ограничений на параметры

Для экстраполяции численности возрастов пятилетней группы методом наименьших квадратов с привлечением численностей двух пятилетних возрастов, предшествующих экстраполируемому, можно рекомендовать метод парабол с наложенными ограничениями.

Выравниваем численности каждой из двух пятилетних групп по параболе 2-го порядка и получаем:

$$\overline{y}_{x.\,1} = a_{0\,(1)} + a_{1\,(1)} x + a_{2\,(1)} x^2,$$
$$\overline{y}_{x.\,2} = a_{0\,(2)} + a_{1\,(2)} x + a_{2\,(2)} x^2.$$

Определим темпы роста параметров a_0, a_1 и a_2, т. е. коэффициенты изменяемости этих параметров:

$$K_{\text{и}\,(0)} = \frac{a_{0\,(2)}}{a_{0\,(1)}}, \quad K_{\text{и}\,(1)} = \frac{a_{1\,(2)}}{a_{1\,(1)}}, \quad K_{\text{и}\,(2)} = \frac{a_{2\,(2)}}{a_{2\,(1)}}.$$

Накладываем на вторую параболу ограничения в виде найденных коэффициентов изменяемости и получаем новое уравнение для экстраполируемых возрастов:

$$\overline{y}_{x.\,2} = a_{0\,(2)} K_{\text{и}\,(0)} + a_{1\,(2)} K_{\text{и}\,(1)} x + a_{2\,(2)} K_{\text{и}\,(2)} x^2.$$

Если же до экстраполируемого периода привлекаются три или даже четыре пятилетних возрастных интервала, то коэффициенты изменяемости определяются как средняя геометрическая. Например, для параметра a_0 в случае привлечения трех пятилетних интервалов имеем:

$$K_{\text{и}\,(0)} = \sqrt{\frac{a_{0\,(3)}}{a_{0\,(1)}}};$$

в случае привлечения четырех пятилетних интервалов

$$K_{\text{и}\,(0)} = \sqrt[3]{\frac{a_{0\,(4)}}{a_{0\,(1)}}}.$$

Аналогично находят коэффициенты изменяемости параметров a_1 и a_2.

В качестве примера используем данные о численности мужчин в 50 губерниях Европейской России по переписи 1897 г. До возраста, подлежащего экстраполяции, т. е. 50—54 лет, привлекаем численности возрастов за два предшествующих пятилетних интервала, т. е. 40—44 и 45—49 лет.

Возраст x	40	41	42	43	44	45	46	47	48	49	Итого
Численность мужчин (в тыс.) y	843,0	284,3	468,6	394,6	316,2	668,2	356,0	332,7	409,1	224,5	4297,2

Выравниваем численности возрастов 40—44 лет по параболе 2-го порядка (см. табл. 60).

Таблица 60

Возраст (в годах)	Условный отсчет возраста x	y	x^2	xy	x^4	x^2y	$\overline{y}_x$
40	—2	843,0	4		16		750,32
41	—1	284,3	1		1		505,51
42	0	468,6	0		0		361,02
43	+1	394,6	1		1		316,85
44	+2	316,2	4		16		373,00
Итого	0	2306,7	10	—943,3	34	5315,7	2306,7

Имеем:

$$\begin{cases} a_{0\,(1)} n + a_{2\,(1)} \sum x^2 = \sum y, \\ a_{0\,(1)} \sum x^2 + a_{2\,(1)} \sum x^4 = \sum x^2 y, \\ a_{1\,(1)} \sum x^2 = \sum xy. \end{cases}$$

Решаем систему нормальных уравнений и находим параметры:

$$a_{0\,(1)} = 361{,}02; \quad a_{1\,(1)} = -94{,}33; \quad a_{2\,(1)} = 50{,}16.$$

Уравнение параболы для возрастов 40—44 лет имеет вид:

$$\overline{y}_x = 361{,}02 - 94{,}33\,x + 50{,}16\,x^2.$$

Используем также параболу 2-го порядка и для выравнивания численности мужчин возраста 45—49 лет (см. табл. 61).

Таблица 61

Возраст (в годах)	Условный отсчет возраста x	y	x^2	xy	x^4	x^2y	$\bar{y}_x$
45	—2	668,2	4		16		615,66
46	—1	356,0	1		1		456,18
47	0	332,7	0		0		347,40
48	+1	409,1	1		1		289,32
49	+2	224,5	4		16		281,94
Итого	0	1 990,5	10	—834,3	34	4 335,9	1 990,50

Находим параметры: $a_{0\,(2)} = 347{,}40$; $a_{1\,(2)} = -83{,}43$; $a_{2\,(2)} = 25{,}35$.

Уравнение параболы для возрастов 45—49 лет примет вид:

$$\bar{y}_x = 347{,}40 - 83{,}43\,x + 25{,}35\,x^2.$$

Находим коэффициенты изменяемости:

$$K_{\text{и}\,(0)} = \frac{a_{0\,(2)}}{a_{0\,(1)}} = 0{,}96227,$$

$$K_{\text{и}\,(1)} = \frac{a_{1\,(2)}}{a_{1\,(1)}} = 0{,}88445,$$

$$K_{\text{и}\,(2)} = \frac{a_{2\,(2)}}{a_{2\,(1)}} = 0{,}50538.$$

Предположим далее, что параметры для уравнения параболы экстраполируемой пятилетней группы будут отличаться от параметров предыдущей группы (45—49 лет) в тех же размерах, которые были нами учтены в коэффициентах изменяемости.

Тогда численности экстраполируемых возрастов, т. е. возрастов 50—54 лет, можно найти по уравнению параболы возрастной группы 45—49 лет с наложенными коэффициентами изменяемости:

$$\bar{y}_x = a_{0\,(2)} K_{\text{и}\,(0)} + a_{1\,(2)} K_{\text{и}\,(1)}\,x + a_{2\,(2)} K_{\text{и}\,(2)}\,x^2 =$$
$$= 347{,}40 \cdot 0{,}96227 - 83{,}43 \cdot 0{,}88445\,x + 25{,}35 \cdot 0{,}50538\,x^2 =$$
$$= 334{,}30 - 73{,}78\,x + 12{,}81\,x^2.$$

При такой экстраполяции следует учитывать, что численности мужчин с увеличением возраста вследствие вымирания обнаруживают тенденцию к снижению. Поэтому экстраполированные по найденной формуле с наложенными ограничениями на параметры численности мужчин следует уменьшить.

Величину уменьшения находят путем сопоставления фактической численности мужчин экстраполируемого пятилетнего интервала с найденной общей экстраполируемой численностью.

Можно (более приближенно) для внесения поправок в найденные экстраполированные численности использовать отношения фактических численностей двух пятилетних интервалов, привлекаемых для экстраполяции.

Возьмем поправки:

$$z_1 = \frac{\text{Фактическая численность экстраполируемого интервала (50—54)}}{\text{Суммарная численность, полученная экстраполяцией}} =$$
$$= \frac{1653{,}1}{1797{,}6} \approx 0{,}92;$$

$$z_2 = \frac{\text{Фактическая численность интервала 45—49 лет}}{\text{Фактическая численность интервала 40—44 года}} =$$
$$= \frac{1990{,}5}{2306{,}7} \approx 0{,}86.$$

В результате использования найденной формулы и величины поправок получаем экстраполированную численность мужчин:

Таблица 62

(в тыс.)

Возраст (в годах)	Условный отсчет возраста x	Экстраполированная численность мужчин $\bar{y}_x = 334{,}30 - 73{,}78x + 12{,}81x^2$	Экстраполированная численность мужчин с поправкой	
			$\bar{y}_x z_1$	$\bar{y}_x z_2$
50	—2	531,10	488,6	456,7
51	—1	420,89	387,2	362,0
52	0	334,30	307,5	287,5
53	+1	273,33	251,5	235,1
54	+2	237,98	218,3	204,7
Итого . .	0	1 797,60	1 653,1	1 546,0

Как и следовало ожидать, экстраполяция численностей в возрасте 50—54 лет довольно хорошо воспроизводит общие тенденции фактических численностей в двух предыдущих пятилетних интервалах (см. рис. 26).

Следует заметить, что применение этого метода к данным 1897 г. дает хорошие результаты потому, что, если не считать искажений в результате возрастной аккумуляции, возрастная структура населения тех лет была очень ровной и не имела существенных изъянов; для населений с серьезными нарушениями возрастной структуры использовать экстраполяцию надо очень осторожно.

Еще лучшие результаты можно получить при использовании этого метода для интерполяции.

Возьмем по одному пятилетнему возрастному интервалу с фактической повозрастной численностью населения до интерполируемого пятилетнего периода и после него. Произведем выравнивание численностей в каждом привлеченном пятилетнем периоде по параболе 2-го порядка:

$$\bar{y}_x = a_{0\,(-1)} + a_{1\,(-1)}\,x + a_{2\,(-1)}\,x^2,$$

$$\bar{y}_x = a_{0\,(+1)} + a_{1\,(+1)}\,x + a_{2\,(+1)}\,x^2.$$

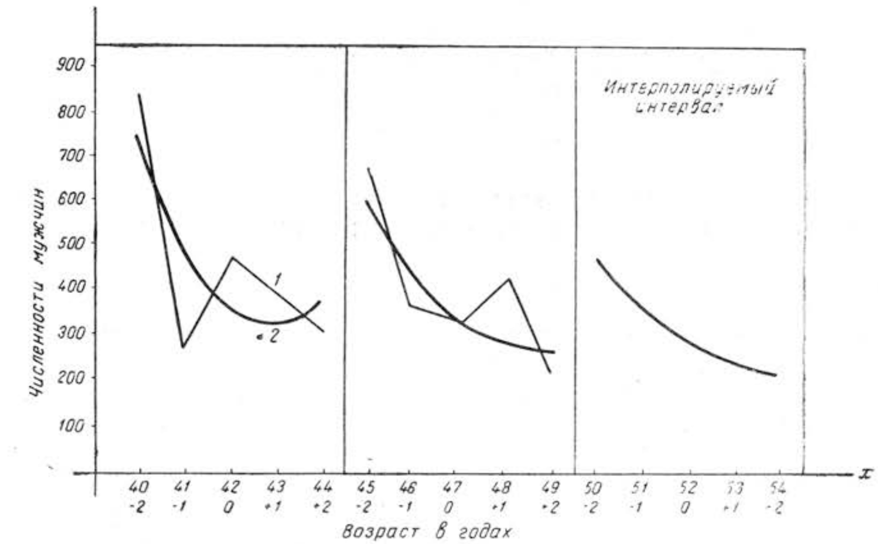

Рис. 26. Фактические (*1*) и выровненные (*2*) численности мужчин в трех пятилетних интервалах.

Параметры для интерполируемого возрастного пятилетнего периода определяем путем исчисления средней арифметической взвешенной из найденных параметров. В качестве весов используем численности населения привлекаемых пятилетних групп:

$$a_{0\,(0)} = \frac{\sum a_0 S}{\sum S} = \frac{a_{0\,(-1)}\,S_{-1} + a_{0\,(+1)}\,S_{+1}}{S_{-1} + S_{+1}},$$

$$a_{1\,(0)} = \frac{\sum a_1 S}{\sum S} = \frac{a_{1\,(-1)}\,S_{-1} + a_{1\,(+1)}\,S_{+1}}{S_{-1} + S_{+1}},$$

$$a_{2\,(0)} = \frac{\sum a_2 S}{\sum S} = \frac{a_{2\,(-1)}\,S_{-1} + a_{2\,(+1)}\,S_{+1}}{S_{-1} + S_{+1}}.$$

Далее получаем интерполяционную формулу

$$\bar{y}_x = a_{0\,(0)} + a_{1\,(0)}\,x + a_{2\,(0)}\,x^2,$$

где x принимает значения -2, -1, 0, 1, 2.

Продолжая наш пример, привлекаем для интерполирования численностей возрастов 50—54 лет численности двух пятилетних периодов: предшествующего интерполируемому, т. е. 45—49 лет, и последующего за ним, т. е. 55—59 лет.

Для возрастов 45—49 лет параметры a_0, a_1 и a_2 уже найдены:

$$a_{0\,(-1)} = +347{,}40;\quad a_{1\,(-1)} = -83{,}43;\quad a_{2\,(-1)} = +25{,}35.$$

Повозрастная численность мужчин 55—59 лет приведена в табл. 63.

Таблица 63

Возраст	x	y	x^2	xy	x^4	x^2y
55	−2	507,6	4		16	
56	−1	266,7	1		1	
57	0	213,5	0		0	
58	+1	233,5	1		1	
59	+2	128,7	4		16	
Итого . . .	0	1 350,0	10	−791,0	34	3 045,4

Решаем систему нормальных уравнений и находим параметры:

$$a_{0\,(+1)} = 220{,}66;\quad a_{1\,(+1)} = -79{,}10;\quad a_{2\,(+1)} = 24{,}67.$$

Находим средние взвешенные из полученных параметров:

$$a_{0\,(0)} = \frac{347{,}40\cdot 1990{,}5 + 220{,}66\cdot 1350{,}0}{1990{,}5 + 1350{,}0} = \frac{989390{,}700}{3340{,}5} \approx 296{,}2;$$

$$a_{1\,(0)} = \frac{(-83{,}43)\cdot 1990{,}5 + (-79{,}10)\cdot 1350{,}0}{1990{,}5 + 1350{,}0} = \frac{-272852{,}415}{3340{,}5} \approx -81{,}7;$$

$$a_{2\,(0)} = \frac{25{,}35\cdot 1990{,}5 + 24{,}67\cdot 1350{,}0}{1990{,}5 + 1350{,}0} = \frac{83763{,}675}{3340{,}5} \approx 25{,}1.$$

Получаем интерполяционную формулу

$$\bar{y}_x = 296{,}2 - 81{,}7\,x + 25{,}1\,x^2.$$

Подставляя вместо x соответствующие его значения, находим интерполированные численности мужчин в возрасте 50—54 лет (см. табл. 64).

Сумма всех численностей составит 1732. Вычислим поправку:

$$z = \frac{1653,1}{1732,0} = 0,955.$$

Затем получим интерполированные численности мужчин с учетом поправки.

Сводим фактическую численность мужчин в возрасте 50—54 лет и результаты экстраполяции и интерполяции (без поправки и с поправкой) в табл. 64.

Таблица 64

Возраст (в годах)	Условный отсчет возраста x	Численность мужчин (в тыс.)				
		по переписи (фактическая)	экстраполированная		интерполированная	
			без поправки	с поправкой	без поправки	с поправкой
50	−2	761,3	531,1	488,6	560,0	534,8
51	−1	167,8	420,9	387,2	403,0	384,9
52	0	290,4	334,3	307,5	296,2	282,9
53	+1	240,0	273,3	251,5	239,6	228,8
54	+2	193,6	238,0	218,3	233,2	221,7
Итого . . .	0	1 653,1	1 797,6	1 653,1	1 732,0	1 653,1

Как видно из таблицы, результаты экстраполяции и интерполяции с учетом поправок не очень отличаются друг от друга (численности мужчин в возрастах 51 и 54 лет почти совпадают). В рассмотренном примере для интерполяции возрастов 50—54 лет мы привлекли фактические данные о численности мужчин двух пятилетних групп (одну до интерполируемого интервала, другую — после).

Для интерполяции численности пятилетней возрастной группы можно привлекать не по одной, а по две пятилетние группы с каждой стороны и более. Тогда параметры для интерполируемой пятилетней группы будут среднеарифметическими взвешенными по численностям населения привлекаемых пятилетних групп, а общая численность населения, полученная при интерполяции или экстраполяции, будет приближаться к общей фактической численности.

Выравнивание по методу наименьших квадратов обладает очень серьезным недостатком. Этот метод оставляет кривой очень большую свободу. Практически при требовании $\Sigma\,(y - \bar{y}_x)^2 = \min$ выровненная кривая может не проходить ни через одну из фактических точек.

Между тем известны случаи, когда исследователь стремится к большей близости кривой не ко всем фактическим данным, а к избранным, внушающим ему большое доверие. В этом случае предъявляется дополнительное требование, чтобы кривая проходила через конкретные точки, в частности когда речь идет о числах доживающих или вероятности умереть, она может проходить через точки, соответствующие возрастам 15, 20, 25 лет и т. д.

Иногда интерполируют данные о продолжительности жизни в какой-либо стране по данным о продолжительности жизни в соседних странах, сходных с исследуемой с экономической точки зрения. Такую интерполяцию производил Б. Ц. Урланис[1].

Использование для выравнивания парабол различных порядков Соприкасательные параболы

Некоторые исследователи считают наилучшим методом выравнивания привлечение парабол 3-го и 4-го порядков.

При этом для возрастов 3—13 лет в некоторых случаях используется парабола 4-го порядка, а для возрастов старше 13 лет — парабола 3-го порядка. Рассмотрим последовательность выравнивания по параболе 4-го порядка. Уравнение кривой имеет вид:

$$\bar{y}_x = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4.$$

Начиная счет от 8 лет, т. е. принимая возраст 8 лет равным нулю, имеем:

Возраст	$\mathbf{x_3}$	x_4	x_5	$\mathbf{x_6}$	x_7	$\mathbf{x_8}$	x_9	$\mathbf{x_{10}}$	x_{11}	x_{12}	$\mathbf{x_{13}}$
Условный отсчет возраста	x_{-5}	x_{-4}	x_{-3}	x_{-2}	x_{-1}	x_0	x_1	x_2	x_3	x_4	x_5

Условно за опорные точки примем возраста 3, 6, 8, 10 и 13 лет (выделенные полужирным). Считая возраст 8 лет центральным (нулевым), располагаем остальные точки симметрично по обе стороны от него.

Тогда получаем:

$$\bar{y}_{x=3} = a_0 - 5a_1 + 25a_2 - 125a_3 + 625a_4;$$
$$\bar{y}_{x=6} = a_0 - 2a_1 + 4a_2 - 8a_3 + 16a_4;$$
$$\bar{y}_{x=8} = a_0;$$
$$\bar{y}_{x=10} = a_0 + 2a_1 + 4a_2 + 8a_3 + 16a_4;$$
$$\bar{y}_{x=13} = a_0 + 5a_1 + 25a_2 + 125a_3 + 625a_4.$$

Откуда

$$\bar{y}_{x=3} + \bar{y}_{x=13} = 2a_0 + 50a_2 + 1250a_4;$$
$$\bar{y}_{x=6} + \bar{y}_{x=10} = 2a_0 + 8a_2 + 32a_4;$$
$$\bar{y}_{x=13} - \bar{y}_{x=3} = 10a_1 + 250a_3;$$
$$\bar{y}_{x=10} - \bar{y}_{x=6} = 4a_1 + 16a_3.$$

[1] См. Б. Ц. Урланис. Рост населения в Европе, стр. 306.

Решая систему уравнений, получаем (запись ведем упрощенно):

$$a_0 = y_8;$$

$$a_1 = \frac{1}{420}(125y_{10} - 125y_6 - 8y_{13} + 8y_3);$$

$$a_2 = \frac{1}{4\,200}(625y_6 + 625y_{10} - 16y_3 - 16y_{13} - 1218y_8);$$

$$a_3 = \frac{1}{420}(2y_{13} - 2y_3 + 5y_6 - 5y_{10});$$

$$a_4 = \frac{1}{4\,200}(4y_{13} + 4y_3 - 25y_{10} - 25y_6 + 42y_8).$$

Можно упростить расчет, используя метод конечных разностей.

Таблица 65

Разности различных порядков

x	y	Значение y, выраженное параметрами	Первая разность Δy	Вторая разность $\Delta^2 y$	Третья разность $\Delta^3 y$	Четвертая разность $\Delta^4 y$
0	y_0	a_0				
			$a_1+a_2+a_3+a_4$			
1	y_1	$a_0+a_1+a_2+a_3+a_4$		$2a_2+6a_3+14a_4$		
			$a_1+3a_2+7a_3+15a_4$		$6a_3+36a_4$	
2	y_2	$a_0+2a_1+4a_2+$ $+8a_3+16a_4$		$2a_2+12a_3+50a_4$		$24a_4$
			$a_1+5a_2+19a_3+65a_4$		$6a_3+60a_4$	
3	y_3	$a_0+3a_1+9a_2+$ $+27a_3+81a_4$		$2a_2+18a_3+$ $110a_4$		
			$a_1+7a_2+37a_3+$ $+175a_4$			
4	y_4	$a_0+4a_1+16a_2+$ $+64a_3+256a_4$				

Таким образом, из уравнения $\overline{y}_x = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$ при $x = 0$ получаем следующие разности:

$$\Delta y_0 = a_1 + a_2 + a_3 + a_4;$$

$$\Delta^2 y_0 = 2a_2 + 6a_3 + 14a_4;$$

$$\Delta^3 y_0 = 6a_3 + 36a_4;$$

$$\Delta^4 y_0 = 24a_4.$$

Подставляя в полученные уравнения разностей найденные значения a_0, a_1, a_2, a_3, a_4, производя необходимые преобразования и используя коэффициенты табл. 66, можно произвести интерполяцию.

Таблица 66

Разности различных порядков	y_3	y_6	y_8	y_{10}	y_{13}
Δy_0	0,01143	—0,14286	—0,28	0,42857	—0,01714
$\Delta^2 y_0$	—0,02286	0,28571	—0,44	0,14286	0,03429
$\Delta^3 y_0$	0,00571	—0,14286	0,36	—0,28571	0,06286
$\Delta^4 y_0$	0,02286	—0,14286	0,24	—0,14286	0,02286

Принимая во внимание следующие связи между разностями: $y_{x+1} = y_x + \Delta y_x$; $\Delta y_{x+1} = \Delta y_x + \Delta^2 y_x$; $\Delta^2 y_{x+1} = \Delta^2 y_x + \Delta^3 y_x$; $\Delta^3 y_{x+1} = \Delta^3 y_x + \Delta^4 y_x$ и укладывая дальнейшие вычисления в схему, изложенную в табл. 67, можно весь процесс выравнивания значительно облегчить.

Таблица 67

Возраст	Показатели, подлежащие выравниванию	Δy_x	$\Delta^2 y_x$	$\Delta^3 y_x$	$\Delta^4 y_x$
x	y_x	Δy_x	$\Delta^2 y_x$	$\Delta^3 y_x$	$\Delta^4 y_x$
$x+1$	y_{x+1}	Δy_{x+1}	$\Delta^2 y_{x+1}$	$\Delta^3 y_{x+1}$	$\Delta^4 y_{x+1}$
$x+2$	y_{x+2}	Δy_{x+2}	$\Delta^2 y_{x+2}$	$\Delta^3 y_{x+2}$	$\Delta^4 y_{x+2}$

В качестве опорных можно выбрать и другие возраста, тогда в каждом конкретном случае выводится своя система формул.

Взяв в качестве опорных возраста 3, 5, 8, 11 и 13 лет, получим:

Возраст	x_3	x_5	x_8	x_{11}	x_{13}
Условный отсчет возраста	x_{-5}	x_{-3}	x_0	x_3	x_5

Привлекая параболу 4-го порядка, составим уравнения:

$$y_3 = a_0 - 5a_1 + 25a_2 - 125a_3 + 625a_4;$$
$$y_5 = a_0 - 3a_1 + 9a_2 - 27a_3 + 81a_4;$$
$$y_8 = a_0;$$
$$y_{11} = a_0 + 3a_1 + 9a_2 + 27a_3 + 81a_4;$$
$$y_{13} = a_0 + 5a_1 + 25a_2 + 125a_3 + 625a_4.$$

Откуда

$$y_3 + y_{13} = 2a_0 + 50a_2 + 1250a_4;$$
$$y_{13} - y_3 = 10a_1 + 250a_3;$$
$$y_5 + y_{11} = 2a_0 + 18a_2 + 162a_4;$$
$$y_{11} - y_5 = 6a_1 + 54a_3.$$

Решая систему уравнений, находим параметры:

$$a_0 = y_8;$$
$$a_1 = \frac{1}{480}(125y_{11} - 125y_5 - 27y_{13} + 27y_3);$$
$$a_2 = \frac{1}{7\,200}(625y_5 + 625y_{11} - 81y_3 - 81y_{13} - 1\,088y_8);$$
$$a_3 = \frac{1}{480}(3y_{13} - 3y_3 - 5y_{11} + 5y_5);$$
$$a_4 = \frac{1}{7\,200}(-25y_5 - 25y_{11} + 9y_3 + 9y_{13} + 32y_8).$$

Используем в качестве примера данные из таблицы смертности городского и сельского населения СССР (1958—1959 гг.)[1]. Возьмем числа доживающих до возраста x лет, т. е. l_x. В качестве опорных выберем возраста 3, 6, 8, 10 и 13 лет.

Таблица 68

Возраст (в годах)	Числа доживающих	Условный отсчет возраста x
3	$l_3 = 94\,780$	—5
4		—4
5		—3
6	$l_6 = 94\,292$	—2
7		—1
8	$l_8 = 94\,069$	0
9		1
10	$l_{10} = 93\,885$	2
11		3
12		4
13	$l_{13} = 93\,654$	5
Итого...		0

Находим параметры:

$$a_0 = l_8 = 94\,069;$$
$$a_1 = \frac{1}{420}(125 \cdot 93\,885 - 125 \cdot 94\,292 - 8 \cdot 93\,654 + 8 \cdot 94\,780) =$$
$$= \frac{-41\,867}{420} = -99{,}68333;$$
$$a_2 = \frac{1}{4\,200}(625 \cdot 94\,292 + 625 \cdot 93\,885 - 16 \cdot 94\,780 - 16 \cdot 93\,654 -$$

[1] «Итоги Всесоюзной переписи населения 1959 года. СССР», стр. 262.

$$-1\,218 \cdot 94\,069) = \frac{19\,639}{4\,200} = 4{,}67595;$$
$$a_3 = \frac{1}{420}(2 \cdot 93\,654 - 2 \cdot 94\,780 + 5 \cdot 94\,292 - 5 \cdot 93\,885) =$$
$$= \frac{-217}{420} = -0{,}51667;$$
$$a_4 = \frac{1}{4\,200}(4 \cdot 93\,654 + 4 \cdot 94\,780 - 25 \cdot 93\,885 - 25 \cdot 94\,292 +$$
$$+ 42 \cdot 94\,069) = \frac{209}{4\,200} = 0{,}04976.$$

Тогда

$$\bar{y}_x = 94\,069 - 99{,}68333\,x + 4{,}67595\,x^2 - 0{,}51667\,x^3 + 0{,}04976\,x^4.$$

Подставляя различные значения x, получим интерполированные значения чисел доживающих для любых возрастов в интервале от 3 до 13 лет. Так, для возраста четыре года (т. е. при $x = -4$) получаем:

$$\bar{l}_4 = 94\,069 - 99{,}68333 \cdot (-4) + 4{,}67595 \cdot 16 - 0{,}51667 \cdot (-64) + $$
$$+ 0{,}04976 \cdot 256 \approx 94\,588;$$

Интерполированная численность доживающих до 5 лет равна:

$$\bar{l}_5 = 94069 - 99{,}68333 \cdot (-3) + 4{,}67595 \cdot 9 - 0{,}51667 \cdot (-27) +$$
$$+ 0{,}04976 \cdot 81 \approx 94\,428 \text{ и т. д.}$$

Интересно, что числа доживающих, приведенные в таблице смертности городского и сельского населения СССР, очень мало отличаются от вычисленных нами. Так, по таблице смертности $l_4 = 94\,568$, а $l_5 = 94\,416$.

Для возрастов старше 13 лет для выравнивания может быть использована парабола 3-го порядка:

$$\bar{y}_x = a_0 + a_1 x + a_2 x^2 + a_3 x^3.$$

За начало координат в соответствии с предыдущим изложением примем пять возрастов, кратных пяти, с интервалом в 15 лет начиная с 15 лет, т. е. 15, 30, 45, 60 и 75 лет. Значит, на этих возрастах $x = 0$.

Составим систему уравнений:

$$y_{x=0} = a_0;$$
$$y_{x=5} = a_0 + 5a_1 + 25a_2 + 125a_3;$$
$$y_{x=10} = a_0 + 10a_1 + 100a_2 + 1000a_3;$$
$$y_{x=15} = a_0 + 15a_1 + 225a_2 + 3375a_3.$$

Решая систему уравнений, получаем:

$$a_0 = y_{x=0};$$
$$a_1 = \frac{1}{60}(-22y_{x=0} + 36y_{x=5} - 18y_{x=10} + 4y_{x+15});$$

$$a_2 = \frac{1}{100}(4y_{x=0} - 10y_{x=5} + 8y_{x=10} - 2y_{x=15});$$

$$a_3 = \frac{1}{1\,500}(-2y_{x=0} + 6y_{x=5} - 6y_{x=10} + 2y_{x=15});$$

Используя конечные разности, получаем при $x = 0$:

$$\Delta y_0 = a_1 + a_2 + a_3;$$
$$\Delta^2 y_0 = 2a_2 + 6a_3;$$
$$\Delta^3 y_0 = 6a_3.$$

Делаем подстановку:

$$\Delta y_0 = \frac{-41y_{x=0} + 63y_{x=5} - 28y_{x=10} + 6y_{x=15}}{125};$$

$$\Delta^2 y_0 = \frac{9y_{x=0} - 22y_{x=5} + 17y_{x=10} - 4y_{x=15}}{125};$$

$$\Delta^3 y_0 = \frac{-y_{x=0} + 3y_{x=5} + 3y_{x=10} + y_{x=15}}{125}.$$

Следует иметь в виду, что в местах стыка выравнивающих парабол 4-го и 3-го порядков могут быть некоторые расхождения между выровненными значениями уровней. Для ликвидации этих расхождений можно использовать метод соприкасательных парабол.

Возьмем уравнение 1-й выравнивающей параболы $y_1 = f_1(x)$,
уравнение 2-й выравнивающей параболы $y_2 = f_2(x)$,
уравнение соприкасательной параболы $y = f(x)$.

Для нахождения параболы, соприкасательной к 1-й выравнивающей параболе в точке x_1, а ко 2-й выравнивающей параболе в точке x_2, введем некоторые дополнительные условия, а именно равенство производных выравнивающих и соприкасательной парабол:

$$f_1(x_1) = f(x_1) \quad \text{и} \quad f_2(x_2) = f(x_2);$$
$$f_1'(x_1) = f'(x_1) \quad \text{и} \quad f_2'(x_2) = f'(x_2);$$
$$f_1''(x_1) = f''(x_1) \quad \text{и} \quad f_2''(x_2) = f''(x_2).$$

Запишем формулу в общем виде:

$$f_1^{(n)}(x_1) = f^{(n)}(x_1) \quad \text{и} \quad f_2^{(n)}(x_2) = f^{(n)}(x_2).$$

Если устраивать соприкосновение в точках на расстоянии 5 лет до стыка и 5 лет после него, то графически это будет выглядеть как на рис. 27.

Из теории конечных разностей известно, что равенству некоторого числа производных соответствует равенство конечных разностей. Чтобы найти уровень соприкасательной параболы, используют разложение функции в ряд Тейлора:

$$\Delta f(x+h) = hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f'''(x) + \ldots.$$

Пусть в точке, соответствующей возрасту, от которого до стыка выравнивающих парабол остается 5 лет (на рис. 27 это возраст $x_1 = 30$ годам), соответствующие значения выравнивающих парабол будут $\bar{y}_{-5}$, $\Delta\bar{y}_{-5}$, $\Delta^2\bar{y}_{-5}$, $\Delta^3\bar{y}_{-5}$. В точке, соответствующей тому возрасту, до которого от стыка выравнивающих парабол 5 лет, соответствующие значения второй выравнивающей параболы будут $\bar{\bar{y}}_{+5}$, $\Delta\bar{\bar{y}}_{+5}$, $\Delta^2\bar{\bar{y}}_{+5}$, $\Delta^3\bar{\bar{y}}_{+5}$.

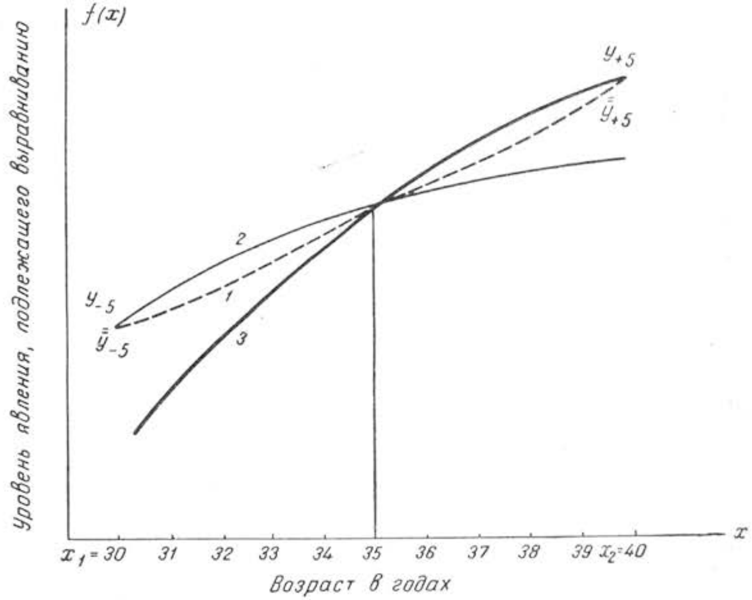

Рис. 27. Соприкасательная парабола к двум выравнивающим параболам:

1 — соприкасательная парабола; *2* — 1-я выравнивающая парабола; *3* — 2-я выравнивающая парабола.

Для нахождения тех же значений соприкасательной параболы предположим:

$$y_{-5} = \bar{y}_{-5}; \qquad y_{+5} = \bar{\bar{y}}_{+5};$$
$$\Delta y_{-5} = \Delta\bar{y}_{-5}; \qquad \Delta y_{+5} = \Delta\bar{\bar{y}}_{+5}.$$

Рассмотрим некоторые свойства разностей:

$$y_{x+k} = y_x + k\,\Delta_1\, y_x = y_x + k\,\Delta_2\, y_x + \frac{k(k-1)}{2!}\Delta_2^2\, y_x;$$

$$\Delta y_{x+k} = \Delta y_x + k\,\Delta_1^2\, y_x = \Delta y_x + k\,\Delta_2^2\, y_x + \frac{k(k+1)}{2!}\Delta_2^3\, y_x.$$

Возьмем конкретный пример:

$$y_{x+k} = 10; \quad y_x = 0; \quad \Delta_1\, y_x = 1; \quad \Delta_2\, y_x = 0{,}55; \quad \Delta_2^2\, y_x = 0{,}1.$$

Получим: $10=0+10\cdot 1=0+10\cdot 0,55+\frac{10\cdot 9}{2!}\cdot 0,1.$

Отметив, что значки при Δy_x разные в зависимости от того, сколько слагаемых выражает конечную сумму, можно найти $\Delta^2 y$ и $\Delta^3 y$, решая следующую систему уравнений:

$$\begin{cases} y_{+5}=\overline{y}_{+5}=\overline{y}_{-5}+10\Delta\overline{y}_{-5}+45\Delta^2 y+120\Delta^3 y, \\ \Delta y_{+5}=\Delta\overline{\overline{y}}_{+5}=\Delta\overline{\overline{y}}_{-5}+10\Delta^2 y+45\Delta^3 y. \end{cases}$$

После определения $\Delta^2 y$ и $\Delta^3 y$ нетрудно найти Δy и y, т. е. все, что необходимо для дальнейших расчетов. В местах стыка выровненные уровни рассчитываются по уравнению соприкасательной параболы.

Скользящие параболы

Иногда данные, подлежащие выравниванию, надежны лишь в пределах укрупненных интервалов (пяти- или десятилетних). В этих случаях выравнивание возрастного распределения (т. е. интерполяция) состоит в нахождении численности населения по каждому одногодичному интервалу внутри укрупненных. При этом используются скользящие параболы различных порядков.

Метод С. А. Новосельского

Метод С. А. Новосельского основан на применении для каждого пятилетнего интервала своей параболы 2-го порядка:

$$\overline{y}_x=a_0+a_1x+a_2x^2,$$

где x — возраст, $\overline{y}_x$ — выровненное значение численности населения в каждом возрасте, a_0, a_1 и a_2 — параметры, определяемые суммарной численностью населения трех пятилетних интервалов (сглаживаемого и двух примыкающих к нему). Рассмотрим табл. 69.

На основании данных таблицы составим систему трех уравнений:

$$\begin{cases} \Sigma_{-1}=5a_0-25a_1+135a_2, \\ \Sigma_0=5a_0+10a_2, \\ \Sigma_1=5a_0+25a_1+135a_2. \end{cases}$$

Решая ее относительно параметров a_0, a_1 и a_2, получаем:

$$a_0=0,2\Sigma_0-0,008(\Sigma_1-2\Sigma_0+\Sigma_{-1}),$$

$$a_1=0,02(\Sigma_1-\Sigma_{-1}),$$

$$a_2=0,004(\Sigma_1-2\Sigma_0+\Sigma_{-1}).$$

Теперь можно получить выровненные значения численности каждого одногодичного интервала:

$$\breve{y}_{-2}=\quad 0,048\Sigma_{-1}+0,184\Sigma_0-0,032\Sigma_1,$$

$$\breve{y}_{-1}=\quad 0,016\Sigma_{-1}+0,208\Sigma_0-0,024\Sigma_1,$$

$$\breve{y}_0=-0,008\Sigma_{-1}+0,216\Sigma_0-0,008\Sigma_1,$$

$$\breve{y}_1=-0,024\Sigma_{-1}+0,208\Sigma_0+0,016\Sigma_1,$$

$$\breve{y}_2=-0,032\Sigma_{-1}+0,184\Sigma_0+0,048\Sigma_1.$$

Таблица 69

Возрастные пятилетние интервалы	Условный отсчет возраста x	Численность населения y	Выровненная численность населения
Предшествующий сглаживаемому	−7	y_{-7}	$\breve{y}_{x=-7}=a_0-7a_1+49a_2$
	−6	y_{-6}	$\breve{y}_{x=-6}=a_0-6a_1+36a_2$
	−5	y_{-5}	$\breve{y}_{x=-5}=a_0-5a_1+25a_2$
	−4	y_{-4}	$\breve{y}_{x=-4}=a_0-4a_1+16a_2$
	−3	y_{-3}	$\breve{y}_{x=-3}=a_0-3a_1+9a_2$
	Итого...	Σ_{-1}	$5a_0-25a_1+135a_2$
Сглаживаемый	−2	y_{-2}	$\breve{y}_{x=-2}=a_0-2a_1+4a_2$
	−1	y_{-1}	$\breve{y}_{x=-1}=a_0-a_1+a_2$
	0	y_0	$\breve{y}_{x=0}=a_0$
	1	y_1	$\breve{y}_{x=1}=a_0+a_1+a_2$
	2	y_2	$\breve{y}_{x=2}=a_0+2a_1+4a_2$
	Итого...	Σ_0	$5a_0\quad+10a_2$
Последующий за сглаживаемым	3	y_3	$\breve{y}_{x=3}=a_0+3a_1+9a_2$
	4	y_4	$\breve{y}_{x=4}=a_0+4a_1+16a_2$
	5	y_5	$\breve{y}_{x=5}=a_0+5a_1+25a_2$
	6	y_6	$\breve{y}_{x=6}=a_0+6a_1+36a_2$
	7	y_7	$\breve{y}_{x=7}=a_0+7a_1+49a_2$
	Итого...	Σ_1	$5a_0+25a_1+135a_2$

Если нужно произвести интерполяцию и получить выровненные численности населения каждого одногодичного интервала, обозначения соответственно меняются.

Этот метод впервые был применен для устранения аккумуляции возрастного распределения населения по переписи 1897 г. при построении таблиц смертности 1896—1897 гг. Недостатком данного метода является плохое смыкание парабол соседних пятилетних интервалов.

Используя данные о численности мужчин Европейской России по переписи населения 1897 г., произведем интерполяцию возрастного интервала 45—49 лет и найдем сглаженные численности возрастов 45, 46, 47, 48, 49 лет:

Возрастные интервалы	Численность мужчин (в тыс.)
40—44	2 306,7 (Σ_{-1})
45—49	1 990,5 (Σ_0)
50—54	1 653,1 (Σ_1)

Используя вышеприведенные формулы, получаем:

$$\breve{y}_{-2} = \breve{y}_{45} = 0{,}048\,\Sigma_{-1} + 0{,}184\,\Sigma_0 - 0{,}032\,\Sigma_1 = \\ = 0{,}048\cdot 2306{,}7 + 0{,}184\cdot 1990{,}5 - 0{,}032\cdot 1653{,}1 = \\ = 110{,}7216 + 366{,}2520 - 52{,}8992 \approx 424{,}1;$$

$$\breve{y}_{-1} = \breve{y}_{46} = 0{,}016\cdot 2306{,}7 + 0{,}208\cdot 1990{,}5 - 0{,}024\cdot 1653{,}1 \approx \\ \approx 411{,}3;$$

$$\breve{y}_0 = \breve{y}_{47} = -0{,}008\cdot 2306{,}7 + 0{,}216\cdot 1990{,}5 - 0{,}008\cdot 1653{,}1 \approx 398{,}3;$$

$$\breve{y}_1 = \breve{y}_{48} = -0{,}024\cdot 2306{,}7 + 0{,}208\cdot 1990{,}5 + \\ + 0{,}016\cdot 1653{,}1 \approx 385{,}1;$$

$$\breve{y}_2 = \breve{y}_{49} = -0{,}032\cdot 2306{,}7 + 0{,}184\cdot 1990{,}5 + 0{,}048\times 1653{,}1 \approx 371{,}8.$$

Сопоставив непосредственные данные, взятые из результатов переписи 1897 г. и искаженные аккумуляцией, с выровненными, увидим, что суммы фактических и выровненных данных по пятилетнему интервалу совпадают.

Метод Б. С. Ястремского

Метод Б. С. Ястремского, основанный на применении парабол 3-го порядка, позволяет лучше смыкать соседние интервалы. Уравнение параболы 3-го порядка $\breve{y}_x = a_0 + a_1x + a_2x^2 + a_3x^3$. Для нахождения четырех параметров a_0, a_1, a_2 и a_3 привлекают 4 десятилетних интервала, каждая смежная пара которых имеет по одному общему пятилетнему интервалу. Решая задачу в общем виде и привлекая численности населения еще двух пятилетних интервалов, примыкающих к тем, которые были использованы при способе С. А. Новосельского, т. е. Σ_{-2} и Σ_2, Б. С. Ястремский получил следующую систему уравнений (сохраняем принятую символику):

$$\breve{y}_{-2} = -0{,}0177\,\Sigma_{-2} + 0{,}1044\,\Sigma_{-1} + 0{,}1210\,\Sigma_0 - 0{,}0044\,\Sigma_1 - 0{,}0033\,\Sigma_2;$$
$$\breve{y}_{-1} = -0{,}0175\,\Sigma_{-2} + 0{,}0780\,\Sigma_{-1} + 0{,}1270\,\Sigma_0 + 0{,}0220\,\Sigma_1 - 0{,}0095\,\Sigma_2;$$
$$\breve{y}_0 = -0{,}0145\,\Sigma_{-2} + 0{,}0500\,\Sigma_{-1} + 0{,}1290\,\Sigma_0 + 0{,}0500\,\Sigma_1 - 0{,}0145\,\Sigma_2;$$
$$\breve{y}_1 = -0{,}0095\,\Sigma_{-2} + 0{,}0220\,\Sigma_{-1} + 0{,}1270\,\Sigma_0 + 0{,}0780\,\Sigma_1 - 0{,}0175\,\Sigma_2;$$
$$\breve{y}_2 = -0{,}0033\,\Sigma_{-2} - 0{,}0044\,\Sigma_{-1} + 0{,}1210\,\Sigma_0 + 0{,}1044\,\Sigma_1 - 0{,}0177\,\Sigma_2.$$

Используя данные предыдущего примера о численности мужского населения по переписи 1897 г. для трех пятилетних интервалов и привлекая дополнительно данные о численности населения в пятилетних интервалах 35—39 и 55—59 лет, произведем интерполяцию пятилетнего интервала 45—49 лет:

Возрастные интервалы	Численность мужчин (в тыс.)
35—39	2 789,9 (Σ_{-2})
40—44	2 306,7 (Σ_{-1})
45—49	1 990,5 (Σ_0)
50—54	1 653,1 (Σ_1)
55—59	1 350,0 (Σ_2)

Подставляем и получаем:

$$\breve{y}_{-2} = \breve{y}_{45} = -0{,}0177\cdot 2789{,}9 + 0{,}1044\cdot 2306{,}7 + \\ + 0{,}1210\cdot 1990{,}5 - 0{,}0044\cdot 1653{,}1 - 0{,}0033\cdot 1350{,}0 = \\ = -49{,}38 + 240{,}82 + 240{,}86 - 7{,}27 - 4{,}45 \approx 420{,}6.$$

Аналогично делаем расчет для всех последующих возрастов:

Возраст	Выровненная численность мужчин (в тыс.)
45	420,6 ($\breve{y}_{-2} = \breve{y}_{45}$)
46	407,5 ($\breve{y}_{-1} = \breve{y}_{46}$)
47	394,8 ($\breve{y}_0 = \breve{y}_{47}$)
48	382,4 ($\breve{y}_1 = \breve{y}_{48}$)
49	370,2 ($\breve{y}_2 = \breve{y}_{49}$)
Итого ...	1 975,5

Метод Б. С. Ястремского применялся при разработке материалов переписи 1920 г.

Суммируя уровни для пятилетнего интервала 45—49, т. е. $\breve{y}_{-2} + \breve{y}_{-1} + \breve{y}_0 + \breve{y}_1 + \breve{y}_2 = \sum_{-2}^{+2} y$, получаем 1975,5 тыс. человек. Эту же

9*

величину можно получить по дополнительно выведенной Б. С. Ястремским формуле

$$\sum_{-2}^{+2} y=\frac{1}{16}(-\Sigma_{-2}+4\Sigma_{-1}+10\Sigma_0+4\Sigma_1-\Sigma_2).$$

Используя эту формулу, действительно получаем:

$$\sum_{-2}^{+2} y=\frac{1}{16}(-2789{,}9+4\cdot 2306{,}7+10\cdot 1990{,}5+$$
$$+4\cdot 1653{,}1-1350{,}0)\approx 1975{,}5 \text{ тыс. человек.}$$

Заметим, однако, что полученная сумма не совпадает с фактической, равной 1990,5 тыс. человек.

Метод В. В. Паевского

Метод В. В. Паевского учитывает, что аккумуляция в возрастах, кратных 0 и 5, неодинакова и поэтому привлечение пятилетних численностей является приближенным приемом. Более правильным является использование десятилетних интервалов. Но по способу Б. С. Ястремского десятилетняя сумма выровненных значений не совпадает с численностью десятилетнего фактического интервала. Так, десятилетняя сумма выровненных данных равна сумме двух выровненных пятилетних интервалов:

$$\frac{1}{16}(-\Sigma_{-2}+4\Sigma_{-1}+10\Sigma_0+4\Sigma_1-\Sigma_2)+$$
$$+\frac{1}{16}(-\Sigma_{-1}+4\Sigma_0+10\Sigma_1+4\Sigma_2-\Sigma_3)=$$
$$=\frac{1}{16}(-\Sigma_{-2}+3\Sigma_{-1}+14\Sigma_0+14\Sigma_1+3\Sigma_2-\Sigma_3).$$

Найденная сумма двух выровненных пятилетних интервалов отличается от фактического десятилетнего интервала $\Sigma_0+\Sigma_1$ на величину разности:

$$K=\frac{1}{16}(\Sigma_{-2}-3\Sigma_{-1}+2\Sigma_0+2\Sigma_1-3\Sigma_2+\Sigma_3).$$

Эта разность может быть довольно велика. Практически можно эту разность распределить пропорционально выровненным численностям.

По рекомендации В. В. Паевского для интерполяции можно использовать параболу 4-го порядка: $\breve{y}_x=a_0+a_1x+a_2x^2+a_3x^3+a_4x^4$.

Параметр a_4 может быть определен на основании требования о равенстве численностей фактических и выровненных пятилетних интервалов, так как при любом способе группировки аккумулирующие возраста 5 и 0, а также все промежуточные войдут в один интервал и

вместят в себя всю «длину волны». Получается система формул, при которой имеет место следующее соотношение фактических и выровненных численностей по пятилетним и десятилетним интервалам (сохраняем обозначения автора): $x_1+x_2=V_1+V_2$; $x_2+x_3=V_2+V_3$; $x_3+x_4=V_3+V_4$; $x_4+x_5=V_4+V_5$, где V — фактические численности пятилетних групп, x — суммы численностей пятилетних групп, полученных выравниванием.

$$y_{10}=-0{,}0128V_1+0{,}0899V_2+0{,}1306V_3+0{,}0052V_4-0{,}0175V_5+0{,}0048V_6$$
$$y_{11}=-0{,}0105V_1+0{,}0571V_2+0{,}1409V_3+0{,}0359V_4-0{,}0304V_5+0{,}0069V_6$$
$$y_{12}=-0{,}0068V_1+0{,}0269V_2+0{,}1444V_3+0{,}0654V_4-0{,}0376V_5+0{,}0077V_6$$
$$y_{13}=-0{,}0025V_1+0{,}0011V_2+0{,}1409V_3+0{,}0919V_4-0{,}0383V_5+0{,}0069V_6$$
$$y_{14}=+0{,}0015V_1-0{,}0188V_2+0{,}1306V_3+0{,}1140V_4-0{,}0321V_5+0{,}0048V_6$$

$$\Sigma_1=-0{,}0312V_1+0{,}1562V_2+0{,}6875V_3+0{,}3125V_4-0{,}1562V_5+0{,}0312V_6$$

$$y_{15}=0{,}0048V_1-0{,}0321V_2+0{,}1140V_3+0{,}1306V_4-0{,}0188V_5+0{,}0015V_6$$
$$y_{16}=0{,}0069V_1-0{,}0383V_2+0{,}0919V_3+0{,}1409V_4+0{,}0011V_5-0{,}0025V_6$$
$$y_{17}=0{,}0077V_1-0{,}0375V_2+0{,}0654V_3+0{,}1444V_4+0{,}0269V_5-0{,}0068V_6$$
$$y_{18}=0{,}0069V_1-0{,}0304V_2+0{,}0352V_3+0{,}1409V_4+0{,}0571V_5-0{,}0105V_6$$
$$y_{19}=0{,}0048V_1-0{,}0175V_2+0{,}0052V_3+0{,}1306V_4+0{,}0899V_5-0{,}0129V_6$$

$$\Sigma_2=0{,}0312V_1-0{,}1562V_2+0{,}3125V_3+0{,}6975V_4+0{,}1562V_5-0{,}0312V_6$$

Откуда $\Sigma_1+\Sigma_2$ равны фактической численности, полученной по переписи, т. е. V_3+V_4.

В качестве примера приведем результаты, полученные выравниванием численности мужчин в 50 губерниях Европейской России по данным переписи 1897 г. методами Б. С. Ястремского и В. В. Паевского (см. табл. 70).

Мы видим при использовании метода В. В. Паевского математически точное равенство между выровненными и фактическими суммами. Плавность при обоих методах получается вполне удовлетворительной (проверка производилась нами методом третьих разностей).

По сравнению с методом Б. С. Ястремского метод В. В. Паевского дает худшие результаты в местах стыков парабол. Методом В. В. Паевского можно пользоваться и при интерполяции старческих возрастов, где аккумуляция в возрастах, оканчивающихся на 0 и 5 лет, проявляется особенно резко.

Вот, например, расчет, произведенный методом В. В. Паевского для возрастов от 90 лет по данным переписи населения 1897 г. о числе мужчин Европейской России (см. табл. 71).

Результаты выравнивания плавные, достаточно близкие к фактическим, и общие численности в интервале 90—109 лет равны друг другу.

9В Зак. 1603

Таблица 70

Сопоставление выровненных численностей мужчин (в тыс.)

Возраст	Численность по переписи*	Численность мужчин, выровненная методом	
		Б. С. Ястремского	В. В. Паевского
30	989,9	612,8	612,9
31	374,0	601,0	601,0
32	563,5	590,0	590,1
33	535,9	579,7	579,8
34	402,0	569,9	569,9
35	759,5	560,2	560,3
36	530,2	550,6	550,6
37	541,7	540,7	540,7
38	588,4	530,4	530,4
39	370,1	519,4	519,6
Итого 30—39	5 655,3	5 654,7	5 655,3
40	843,0	505,5	503,0
41	284,3	488,5	484,9
42	468,6	471,1	467,1
43	394,6	453,5	449,8
44	316,2	435,9	433,4
45	668,3	420,6	418,1
46	356,0	407,5	403,8
47	332,7	394,8	390,8
48	409,1	382,4	378,8
49	224,5	370,2	367,7
Итого 40—49	4 297,4	4 330,0	4 297,4
50	761,3	357,1	354,6
51	167,8	343,1	339,5
52	290,4	329,3	325,3
53	240,0	316,0	312,4
54	193,6	303,2	300,7
55	507,6	292,9	290,4
56	266,7	284,9	281,3
57	213,5	277,3	273,3
58	233,5	269,8	266,1
59	128,7	262,1	259,6
Итого 50—59	3 003,2	3 035,7	3 003,2

* С. А. Новосельский. Смертность и продолжительность жизни в России. Пг., 1916, стр. 101—102.

Продолжение

Возраст	Численность по переписи*	Численность мужчин, выровненная методом	
		Б. С. Ястремского	В. В. Паевского
60	642,4	252,6	253,1
61	107,4	241,1	241,9
62	181,9	229,2	230,1
63	163,2	216,9	217,7
64	112,8	204,3	201,8
65	335,9	191,0	191,6
66	128,8	177,3	178,1
67	136,4	163,7	167,6
68	105,8	150,2	151,1
69	56,0	137,2	137,8
Итого 60—69	1 970,7	1 963,4	1 970,7
Всего 30—69	14 926,6	14 983,5	14 926,6

Таблица 71

Возраст в годах	Численность мужчин по переписи*	Численность мужчин, выровненная методом В. В. Паевского
90	22,00	6,52
91	1,30	5,58
92	1,67	4,87
93	1,42	4,30
94	1,19	3,83
95	4,72	3,40
96	1,84	2,97
97	1,43	2,52
98	1,40	2,05
99	0,61	1,54
100	2,04	0,92
101	0,23	0,70
102	0,28	0,53
103	0,20	0,39
104	0,12	0,28
105	0,25	0,21
106	0,08	0,14
107	0,07	0,10
108	0,05	0,07
109	0,07	0,05
Итого	40,97	40,97

* С. А. Новосельский. Смертность и продолжительность жизни в России, стр. 103.

Методы, использованные В. И. Борткевичем и М. В. Птухой

Имея численности пятилетних возрастных групп, В. И. Борткевич и М. В. Птуха осуществляли переход к одногодичным группам по формулам, взятым ими из «*Annali di Statistica*»[1].

Обозначая численность соответствующих групп живущих или умерших, как и ранее, Σ_{-1}, Σ_0, Σ_1, а искомые величины через $\breve{y}_1$, $\breve{y}_2$, $\breve{y}_3$, $\breve{y}_4$, $\breve{y}_5$, значение центрального интерполируемого уровня определяем по формуле

$$\breve{y}_3 = \frac{\Sigma_0}{5} - \frac{\Sigma_{-1} + \Sigma_1 - 2\Sigma_0}{125}.$$

Все остальные интерполируемые уровни выражены через $\breve{y}_3$:

$$\breve{y}_1 = \breve{y}_3 - \frac{\Sigma_1 - \Sigma_{-1}}{25} + \frac{\Sigma_{-1} + \Sigma_1 - 2\Sigma_0}{62,5},$$

$$\breve{y}_2 = \breve{y}_3 - \frac{\Sigma_1 - \Sigma_{-1}}{50} + \frac{\Sigma_{-1} + \Sigma_1 - 2\Sigma_0}{250},$$

$$\breve{y}_4 = \breve{y}_3 + \frac{\Sigma_1 - \Sigma_{-1}}{50} + \frac{\Sigma_{-1} + \Sigma_1 - 2\Sigma_0}{250},$$

$$\breve{y}_5 = \breve{y}_3 + \frac{\Sigma_1 - \Sigma_{-1}}{25} + \frac{\Sigma_{-1} + \Sigma_1 - 2\Sigma_0}{62,5}.$$

Привлечем данные о численности мужчин Европейской России по переписи 1897 г., уже использованные нами для интерполяции методом С. А. Новосельского, произведем для сравнения интерполяцию данным методом.

Возрастные интервалы	Численность мужчин (в тыс.)
40—44	2 306,7 (Σ_{-1})
45—49	1 990,5 (Σ_0)
50—54	1 653,1 (Σ_1)

Найдем численность мужчин центрального возраста интерполируемого периода, т. е. 47 лет:

$$\breve{y}_3 = \breve{y}_{47} = \frac{1990,5}{5} - \frac{2306,7 + 1653,1 - 2 \cdot 1990,5}{125} =$$

$$= 398,1 + \frac{21,2}{125} = 398,3.$$

Откуда

$$\breve{y}_1 = \breve{y}_{45} = 398,3 - \frac{-653,6}{25} + \frac{-21,2}{62,5} = 398,3 + 26,1 - 0,3 \approx 424,1,$$

[1] См. М. В. Птуха. Очерки по статистике населения. М., Госстатиздат, 1960, стр. 188.

$$\breve{y}_2 = \breve{y}_{46} = 398,3 - \frac{-653,6}{50} + \frac{-21,2}{250} \approx 411,3,$$

$$\breve{y}_4 = \breve{y}_{48} = 398,3 + \frac{-653,6}{50} + \frac{-21,2}{250} \approx 385,1,$$

$$\breve{y}_5 = \breve{y}_{49} = 398,3 + \frac{-653,6}{25} + \frac{-21,2}{62,5} \approx 371,8.$$

Таким образом, полученные этим методом интерполированные годовые численности ничем не отличаются от полученных методом С. А. Новосельского, и сумма выровненных чисел мужчин в пятилетнем интервале совпадает с фактической (расхождение в 0,1 объясняется приближенностью расчета).

При несовпадении выровненных и фактических сумм живущих и умерших за соответствующие действительности принимают фактические, устанавливается коэффициент пропорциональности (отношение общей фактической суммы к сумме выровненных показателей), и численность каждого одногодичного возраста умножается на этот коэффициент. Вычислив данным методом интерполируемые уровни, М. В. Птуха прибегал к их выравниванию методом взвешенных скользящих средних.

При интерполяции численности одногодичных возрастов в пятилетнем интервале путем привлечения численностей двух соседних пятилетних интервалов можно применять иной метод.

Возьмем все те же данные о численности мужского населения Европейской России по переписи 1897 г. Следовательно, мы имеем:

$$S_{40-44} = 2306,7, \quad S_{45-49} = 1990,5 \text{ и } S_{50-54} = 1653,1.$$

Обозначая численности населения в каждом одногодичном интервале y и присваивая им соответствующие индексы, получаем:

$$S_{40-44} = 2306,7 = y_0 + y_1 + y_2 + y_3 + y_4,$$
$$S_{45-49} = 1990,5 = y_5 + y_6 + y_7 + y_8 + y_9,$$
$$S_{50-54} = 1653,1 = y_{10} + y_{11} + y_{12} + y_{13} + y_{14}.$$

Предполагая, что изменение численностей внутри каждого пятилетнего интервала происходит по параболе 2-го порядка ($y = a_0 + a_1x + a_2x^2$), получаем следующую систему уравнений с тремя неизвестными:

$$\begin{cases} S_{40-44} = \int\limits_0^5 (a_0 + a_1x + a_2x^2)\,dx = 5a_0 + 12,5a_1 + 41,66667a_2, \\ S_{45-49} = \int\limits_5^{10} (a_0 + a_1x + a_2x^2)\,dx = 5a_0 + 37,5a_1 + 291,66667a_2, \\ S_{50-54} = \int\limits_{10}^{15} (a_0 + a_1x + a_2x^2)\,dx = 5a_0 + 62,5a_1 + 791,66667a_2. \end{cases}$$

Решая систему, находим параметры:

$$a_0 = 0{,}36667\,S_{40-44} - 0{,}23333\,S_{45-49} + 0{,}06667 S_{50-54} = 491{,}5,$$
$$a_1 = -0{,}08\,S_{40-44} + 0{,}12\,S_{45-49} - 0{,}04\,S_{50-54} = -11{,}792,$$
$$a_2 = 0{,}004\,S_{40-44} - 0{,}008\,S_{45-49} + 0{,}004\,S_{50-54} = -0{,}0852.$$

Вычисляем численность одногодичного возраста, например 45 лет, т. е. y_5:

$$y_5 = \int_5^6 (a_0 + a_1x + a_2x^2)\,dx = \left| a_0x + a_1\frac{x^2}{2} + a_2\frac{x^3}{3}\right|_5^6 =$$
$$= 6a_0 + 18a_1 + 72a_2 - (5a_0 + 12{,}5a_1 + 41{,}67a_2) =$$
$$= a_0 + 5{,}5a_1 + 30{,}33a_2.$$

Подставляя значения параметров, получаем:

$$S_{45} = y_5 = 491{,}5 + 5{,}5(-11{,}792) + 30{,}33(-0{,}0852) = 424{,}1.$$

Аналогично найдем численность одногодичного возраста 46 лет:

$$y_6 = \int_6^7 (a_0 + a_1x + a_2x^2)\,dx = \left| a_0x + a_1\frac{x^2}{2} + a_2\frac{x^3}{3}\right|_6^7 =$$
$$= 7a_0 + 24{,}5a_1 + 114{,}33a_2 - (6a_0 + 18a_1 + 72a_2) = a_0 + 6{,}5a_1 + 42{,}33a_2,$$

откуда

$$S_{46} = y_6 = 491{,}5 + 6{,}5(-11{,}792) + 42{,}33(-0{,}0852) = 411{,}2 \text{ и т. д.}$$

Можно получить сумму численностей двух интерполируемых возрастов сразу, т. е. $y_5 + y_6$:

$$S_{45-46} = y_5 + y_6 = \int_5^7 (a_0 + a_1x + a_2x^2)\,dx = \left| a_0x + a_1\frac{x^2}{2} + a_2\frac{x^3}{3}\right|_5^7 =$$
$$= 7a_0 + 24{,}5a_1 + 114{,}33a_2 - (5a_0 + 12{,}5a_1 + 41{,}67a_2) =$$
$$= 2a_0 + 12a_1 + 72{,}67a_2.$$

Подставляя значения параметров, получаем:

$$S_{45-46} = y_5 + y_6 = 2\cdot 491{,}5 + 12(-11{,}792) + 72{,}67(-0{,}0852) = 835{,}3.$$

КОМБИНИРОВАНИЕ ПАРАБОЛ

Получив методом скользящих парабол определенную систему формул для уровней определенного периода, подлежащего выравниванию, можно ограничиваться расчетом лишь некоторых значений выровненных уровней, которые назовем опорными; а далее, используя новую параболу, произвести выравнивание, основываясь на опорных возрастах.

Возьмем систему формул, полученную С. А. Новосельским:

$\breve{y}_0 = -0{,}008\,\Sigma_{-1} + 0{,}216\,\Sigma_0 - 0{,}008\,\Sigma_1$ и т. д. (см. стр. 241).

Интерполируем числа живущих L_x и умерших d_x и вычисляем q_x и p_x по формулам:

$$q_x = \frac{d_x}{L_x + \frac{1}{2}d_x}, \quad p_x = 1 - q_x.$$

Находим q_x для опорных — центральных возрастов пятилетних групп: 12, 17, 22 и т. д. до 82 лет.

Для исчисления q_x всех возрастов между опорными используют параболу 3-го порядка $u_{x+1} = a_0 + ax + bx^2 + cx^3$, в окончательном виде представленную следующим уравнением:

$$u_{x+1} = u_1 + x\Delta u_0 + \frac{x + x^2}{2}\Delta^2 u_0 + \frac{x^2 + x^3}{2}\Delta^3 u_0.$$

Подставляя вместо x последовательно его значения 0, $\frac{1}{5}$, $\frac{2}{5}$, $\frac{3}{5}$, $\frac{4}{5}$, 1 и производя вычитание, получаем следующие разности одногодичных групп:

$$\delta u_1 = 0{,}2\Delta u_0 + 0{,}12\Delta^2 u_0 - 0{,}016\Delta^3 u_0,$$
$$\delta^2 u_1 = 0{,}04\Delta^2 u_0 - 0{,}016\Delta^3 u_0,$$
$$\delta^3 u_1 = 0{,}024\Delta^3 u_0.$$

Далее интерполированием находим одногодичные значения функции, причем для интерполирования берем не значения q_x, а их логарифмы. Такая искусственная форма функции дает наибольшую плавность. Для избежания большого количества нулей перед значащими цифрами берем не $\lg q_x$, а $\lg(q_x + 0{,}1)$.

Итак, сначала получают разности 1, 2 и 3-го порядка между значениями пятилетних групп путем вычитания (Δ); затем находят разности между одногодичными значениями (δ), а по ним путем суммирования и сами значения q. Таким методом выравнивались q_x при построении таблицы смертности 1926—1927 гг. для городов Восточной Сибири[1]. Произведем интерполяцию q_x для возрастов с 17 до 27 лет. Для этого привлекаем данные для центральных возрастов:

Возраст (x)	12	17	22	27	32	37
Вероятность умереть (q_x)	0,003092	0,005590	0,005404	0,009570	0,011581	0,014471

Расчет делается по схеме, приведенной в табл. 72.

При комбинировании парабол предварительно найдем опорные уровни путем объединения двух пятилетних интервалов в один десятилетний, как это делал Б. С. Ястремский, с применением скольжения. Обозначая суммы фактических численностей пятилетних возрастов через S, а суммы выровненных численностей через V и исходя из требования равенства сумм десятилетних интервалов, получаем систему трех уравнений:

[1] К. И. Романов. Смертность и продолжительность жизни населения городов Восточной Сибири за 1926—1927 гг. Иркутск, издание Восточносибирского Мединститута, 1935.

Таблица 72

Интерполяция вероятностей умереть методом комбинирования парабол

Возраст x	q_x	$q_x+0{,}1$	$\lg(q_x+0{,}1)$	Δu_0	$\Delta^2 u_0$	$\Delta^3 u_0$	$0{,}2\,\Delta u_0$	$0{,}12\,\Delta^2 u_0$	$0{,}016\,\Delta^3 u_0$	$0{,}04\,\Delta^2 u_0$
1	2	3	4	5	6	7	8	9	10	11
12	0,003092	0,103092	$\bar{1}$,013229							
				+10 396						
17	0,005590	0,105590	$\bar{1}$,023625		−11 165		+2079,200	−1339,800	+460,320	−446,600
				−769		+28 770				
22	0,005404	0,105404	$\bar{1}$,022856		+17 605		−153,800	+2112,600	−424,512	+704,200
				+16 836		−26 532				
27	0,009570	0,109570	$\bar{1}$,039692							
					−8 927		+3367,200	−1071,240	+193,872	−357,080
32	0,011581	0,111581	$\bar{1}$,047601	+7 909		+12 117				
					+3 190					
37	0,014471	0,114471	$\bar{1}$,058700	+11 099						

$0{,}024\,\Delta^3 u_0 = \delta^3 u_1$	$\delta^2 u_1 = 0{,}04\,\Delta^2 u_0 - 0{,}016\,\Delta^3 u_0$	$\delta u_1 = 0{,}2\Delta u_0 + 0{,}12\Delta^2 u_0 - 0{,}016\Delta^3 u_0$	$\delta^3 u_1$	$\delta^2 u_1$	δu_1	$\lg(q_x+0{,}1)$	$q_x+0{,}1$	q_x	p_x	Возраст x
12	13	14	15	16	17	18	19	20	21	22
+690,480	−906,920	279,080	+690,480	−906,920	279,080	$\bar{1}$,023625	0,105590	0,005590	0,994410	17
				−216,440	−627,840	$\bar{1}$,023904	0,105657	0,005657	0,994343	18
				+474,040	−844,280	$\bar{1}$,023276	0,105505	0,005505	0,994495	19
				+1166,520	−370,240	$\bar{1}$,022432	0,105201	0,005201	0,994799	20
					794,280	$\bar{1}$,022062	0,105212	0,005212	0,994788	21
−636,768	1128,712	2383,312	−636,768	+1128,712	2383,312	$\bar{1}$,022856	0,105404	0,005494	0,994596	22
				+491,944	3512,024	$\bar{1}$,025239	0,105985	0,005985	0,994015	23
				−144,824	4003,968	$\bar{1}$,028751	0,106842	0,006842	0,993158	24
				−781,592	3859,144	$\bar{1}$,032755	0,107833	0,007833	0,992167	25
					3077,552	$\bar{1}$,036614	0,108796	0,008796	0,991204	26
+290,808	−550,952	2102,088	+290,808	−550,952	2102,088	$\bar{1}$,039692	0,109570	0,009570	0,990430	27

$$\begin{cases} V_1+V_2=S_1+S_2, \\ V_2+V_3=S_2+S_3, \\ V_3+V_4=S_3+S_4. \end{cases}$$

Имея 4 пятилетия, т. е. 20 лет, охватываем их параболой 2-го порядка $\breve{y}_x = a_0 + a_1x + a_2x^2$.

Возьмем возраста начиная с 10 лет и составим систему обозначений так, чтобы $\Sigma x = 0$ (см. табл. 73). Составим три уравнения и, решая их, находим параметры a_0, a_1, a_2.

$$a_0 = -0{,}0165S_1 + 0{,}1165S_2 + 0{,}1165S_3 - 0{,}0165S_4,$$
$$a_1 = 0{,}0050S_1 + 0{,}0050S_2 - 0{,}0050S_3 - 0{,}0050S_4,$$
$$a_2 = 0{,}0005\,S_1 - 0{,}0005S_2 - 0{,}0005S_3 + 0{,}0005S_4.$$

Исходя из того, что q_x — плавная функция возраста, изменяющаяся без скачков, находим q_x для опорных возрастов, в качестве которых используют центральные возраста пятилетних интервалов 17, 22, 27, 32, 37 и т. д.

В соответствии с нашими обозначениями возраст 17 лет соответствует условному обозначению возраста $x = -5$, уровень которого y_8, а возраст 22 соответствует $x = +5$ с уровнем y_{13}. Подставляя в уравнение параболы $\breve{y}_x = a_0 + a_1x + a_2x^2$ значения параметров a_0, a_1, a_2 и значения x, равные -5 и $+5$, получим после преобразований:

$$y_8 = 0{,}021S_1 + 0{,}129S_2 + 0{,}079S_3 - 0{,}029S_4,$$
$$y_{13} = -0{,}029S_1 + 0{,}079S_2 + 0{,}129S_3 + 0{,}021S_4.$$

Далее произведем передвижку, приравнивая, например, y_8 к 22 годам, а y_{13} к 27; тогда S_1 выразит сумму численностей возрастов в пятилетнем интервале от 15 до 19 лет; S_2 — 20—24; S_3 — 25—29; S_4 — 30—34 лет и т. д.

Находим выравниваемые численности живущих и умерших, относящиеся к опорным возрастам. Зная эти численности, определяем q_x. Допустим, например, что выровненные численности живущих в опорных возрастах составляют: $y_{22} = 120\,731$, $y_{27} = 93\,147$.

Тогда на основании уравнения $q_x = \dfrac{d_x}{L_x + \frac{1}{2}d_x}$ получаем:

$$q_{22} = \frac{813}{120\,731 + 406} \approx 0{,}00671,$$
$$q_{27} = \frac{757}{93\,147 + 379} \approx 0{,}00810.$$

Для интерполирования q_x между опорными возрастами можно использовать параболу 4-го порядка

$$\breve{y}_x = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4.$$

Введем систему обозначений, аналогичную табл. 75, для возрастов с 12 лет до 32 (см. табл. 74).

Таким образом, при соответствующем условном обозначении $x = 10$ мы имеем y_{12}, при $x = 5$ y_{17}, при $x = 0$ y_{22}, при $x = -5$ y_{27} и при $x = -10$ y_{32}. Обозначим q_x опорных возрастов: $y_{12} = A_1$, $y_{17} = A_2$, $y_{22} = A_0$, $y_{27} = A_3$, $y_{32} = A_4$.

Таблица 73

Возраст в годах	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Символическое обозначение возраста x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
Условное обозначение возраста x	—19	—17	—15	—13	—11	—9	—7	—5	—3	—1	+1	+3	+5	+7	+9	+11	+13	+15	+17	+19
Выравниваемый уровень y	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}	y_{18}	y_{19}	y_{20}

Таблица 74

Возраст в годах	**12**	*13*	*14*	*15*	*16*	**17**	18	19	20	21	22	23	24	25	26	**27**	*28*	*29*	*30*	*31*	**32**
Символические обозначения возраста	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}	x_{31}	x_{32}
Условное обозначение возраста x	+10	+9	+8	+7	+6	+5	+4	+3	+2	+1	0	—1	—2	—3	—4	—5	—6	—7	—8	—9	—10
Вероятность умереть y	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}	y_{18}	y_{19}	y_{20}	y_{21}	y_{22}	y_{23}	y_{24}	y_{25}	y_{26}	y_{27}	y_{28}	y_{29}	y_{30}	y_{31}	y_{32}

Примечание. Опорные возраста в таблице выделены полужирным шрифтом; возраста, между которыми производится выравнивание, набраны курсивом.

Подставляя значения x, равные 10, 5, 0, — 5, — 10, в уравнение параболы 4-го порядка, составляем 5 уравнений с пятью неизвестными:

$$\begin{cases} A_1 = a_0 + 10a_1 + 100a_2 + 1\,000a_3 + 10\,000a_4, \\ A_2 = a_0 + 5a_1 + 25a_2 + 125a_3 + 625a_4, \\ A_0 = a_0, \\ A_3 = a_0 - 5a_1 + 25a_2 - 125a_3 + 625a_4, \\ a_4 = a_0 - 10a_1 + 100a_2 - 1\,000a_3 + 10\,000a_4, \end{cases}$$

Решая систему в общем виде, получаем:

$$\begin{aligned} a_0 &= A_0, \\ a_1 &= -0{,}00150A_1 + 0{,}13325A_2 - 0{,}13325A_3 + 0{,}01650A_4, \\ a_2 &= -0{,}05000A_0 - 0{,}00150A_1 + 0{,}02650A_2 + 0{,}02650A_3 - 0{,}00150A_4, \\ a_3 &= 0{,}00066A_1 - 0{,}00133A_2 + 0{,}00133A_3 - 0{,}00066A_4, \\ a_4 &= 0{,}00040A_0 + 0{,}00006A_1 - 0{,}00026A_2 - 0{,}00026A_3 + 0{,}00006A_4. \end{aligned}$$

Подставляя найденные значения параметров в уравнение параболы, получаем для выровненных значений $y_x = q_x$ одиннадцать уравнений:

$$\begin{aligned} y_{17} &= A_2, \\ y_{18} &= 0{,}30240A_0 - 0{,}03240A_1 + 0{,}80532A_2 - 0{,}09044A_3 + 0{,}01512A_4, \\ y_{19} &= 0{,}58240A_0 - 0{,}04032A_1 + 0{,}58128A_2 - 0{,}14640A_3 + 0{,}02304A_4, \\ y_{20} &= 0{,}80640A_0 - 0{,}03276A_1 + 0{,}35770A_2 - 0{,}15402A_3 + 0{,}02268A_4, \\ y_{21} &= 0{,}95040A_0 - 0{,}01728A_1 + 0{,}15160A_2 - 0{,}10568A_3 + 0{,}01440A_4, \\ y_{22} &= A_0, \\ y_{23} &= 0{,}95040A_0 + 0{,}01440A_1 - 0{,}15160A_2 + 0{,}15816A_3 - 0{,}01728A_4, \\ y_{24} &= 0{,}80640A_0 + 0{,}02268A_1 - 0{,}35770A_2 + 0{,}35770A_3 - 0{,}03276A_4, \\ y_{25} &= 0{,}58240A_0 + 0{,}02304A_1 - 0{,}58128A_2 + 0{,}58128A_3 - 0{,}04032A_4, \\ y_{26} &= 0{,}30240A_0 + 0{,}01512A_1 - 0{,}80532A_2 + 0{,}80532A_3 - 0{,}03240A_4, \\ y_{27} &= A_3. \end{aligned}$$

Аналогичный метод комбинации двух парабол был использован при обработке данных переписи населения 1926 г. по Уральской области для составления таблиц смертности. При выравнивании q_x возможна комбинация двух парабол с дополнительным использованием метода конечных разностей.

ПАРАБОЛЫ 4-ГО И 5-ГО ПОРЯДКА

Методы Спрейга

Рассмотрим особенности метода выравнивания, разработанного Спрейгом, и укажем возможности использования этого метода применительно к демографии.

Допустим, что требуется найти выровненный ряд из пяти одногодичных вероятностей смерти (или дожития). При этом должно удовлетворяться основное требование, состоящее в том, что фактическая и выровненная пятилетние вероятности смерти должны быть равны.

Имеем шесть ординат[1]: y_{-2}, y_{-1}, y_0, y_1, y_2, y_3, принятых за опорные. Следовательно, x равен соответственно -2, -1, 0, $+1$, $+2$, $+3$. Допустим, что между двумя опорными ординатами, например y_0 и y_1, требуется определить несколько промежуточных членов (см. рис. 28).

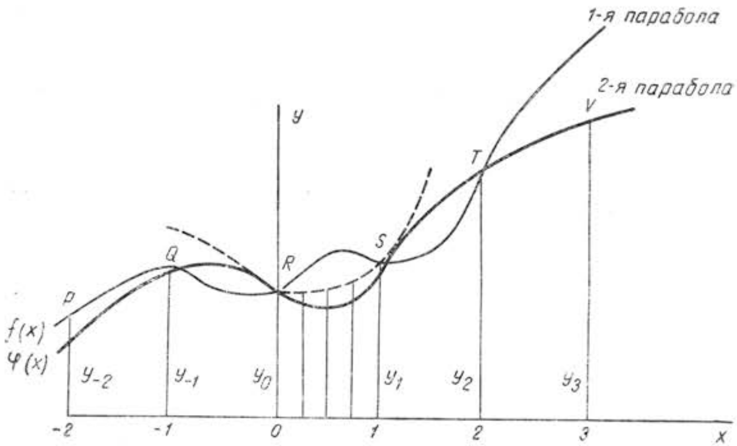

Рис. 28. Выравнивание методом Спрейга.

Рассматриваем две группы ординат по пяти ординат в каждой: первая — y_{-2}, y_{-1}, y_0, y_1, y_2; вторая — y_{-1}, y_0, y_1, y_2, y_3 и проводим через каждую группу ординат параболу 4-го порядка.

Первая парабола проходит через точки P, Q, R, S, T, а вторая — через точки Q, R, S, T, V. Составим уравнение функций для первой группы ординат:

$$f(x) = y_{-2} + (x+2)\Delta y_{-2} + \frac{(x+2)(x+1)}{1\cdot 2}\Delta^2 y_{-2} + \frac{(x+2)(x+1)x}{1\cdot 2\cdot 3}\Delta^3 y_{-2} + \frac{(x+2)(x+1)x(x-1)}{1\cdot 2\cdot 3\cdot 4}\Delta^4 y_{-2},$$

для второй группы ординат:

$$\varphi(x) = y_{-1} + (x+1)\Delta y_{-1} + \frac{(x+1)x}{1\cdot 2}\Delta^2 y_{-1} + \frac{(x+1)x(x-1)}{1\cdot 2\cdot 3}\Delta^3 y_{-1} + \frac{(x+1)x(x-1)(x-2)}{1\cdot 2\cdot 3\cdot 4}\Delta^4 y_{-1}.$$

[1] Последовательность действий дана в соответствии с рекомендациями С. А. Новосельского и В. В. Паевского, заимствовавших вывод у Гловера из работы «United States Life — Tables» (Washington, 1921), и во многом аналогична действиям при использовании метода соприкасательных парабол.

Так как эти две параболы в точках, где $x = -1$, $x = 0$, $x = +1$, $x = +2$, пересекаются, то $f(-1) = \varphi(-1)$, $f(0) = \varphi(0)$, $f(1) = \varphi(1)$ и $f(2) = \varphi(2)$.

Нужно найти уравнение параболы $y(x) = f(x) + u(x)$, которая имела бы, во-первых, общие ординаты с $f(x)$ и $\varphi(x)$ в точках $x = 0$ и $x = 1$ и, во-вторых, общую касательную и радиус кривизны с первой параболой, т. е. $f(x)$, в точке, где $x = 0$, и со второй параболой, т. е. $\varphi(x)$, в точке, где $x = 1$. Следовательно,

$$y(0) = f(0), \quad y'(0) = f'(0), \quad y''(0) = f''(0),$$

$$y(1) = \varphi(1), \quad y'(1) = \varphi'(1), \quad y''(1) = \varphi''(1).$$

Это 6 ограничительных условий, позволяющих вывести уравнение параболы 5-го порядка. Из уравнения искомой параболы находим:

$$u(x) = y(x) - f(x),$$

$$u'(x) = y'(x) - f'(x),$$

$$u''(x) = y''(x) - f''(x).$$

Подставим вместо x соответствующие значения 0 и 1 и получаем:

$$u(0) = 0, \quad u'(0) = 0, \quad u''(0) = 0, \quad u(1) = y(1) - f(1).$$

Но $f(1) = \varphi(1)$, так как кривые $f(x)$, $\varphi(x)$ пересекаются в точке, где $x = 1$. Отсюда $u(1) = 0$. Таким образом, функция $u(x)$ содержит в себе множители $x^3(x-1)$. Можно определить еще один множитель, входящий в состав $u(x)$, зная, что $\Delta^5 y_{-1} = \Delta^5 y_{-2} + \Delta^{5+1} y_{-2}$, и ограничиваясь пятой разностью, которая для параболы 5-го порядка является постоянной величиной. Запишем тождество:

$$\varphi(x) = f(x) + \frac{(x+1)x(x-1)(x-2)}{1\cdot 2\cdot 3\cdot 4}\Delta^5 y_{-2}.$$

Рассмотрим разность:

$$y(x) - \varphi(x) = f(x) + u(x) - \varphi(x).$$

Заменяя значение $\varphi(x)$ в правой части предыдущего равенства, получаем:

$$y(x) - \varphi(x) = u(x) - \frac{(x+1)x(x-1)(x-2)}{1\cdot 2\cdot 3\cdot 4}\Delta^5 y_{-2}.$$

Из шести ограничительных условий, изложенных выше, возьмем одно: $y(1) - \varphi(1) = 0$. Дифференцируя дважды, получаем $y'(1) - \varphi'(1) = 0$, $y''(1) - \varphi''(1) = 0$, отсюда

$$u'(1) = u''(1) = -\frac{2\Delta^5 y_{-2}}{1\cdot 2\cdot 3\cdot 4}.$$

Теперь можно задать функцию $u(x)$, считая ее параболой 5-го порядка: $u(x) = x^3 (x-1)(a_0 + a_1 x) \frac{\Delta^5 y_{-2}}{1\cdot 2\cdot 3\cdot 4}$. Снова дифференцируя дважды, подставляя $x = 1$ и решая полученную систему двух уравнений, находим параметры a_0 и a_1: $a_0 = -7, a_1 = 5$. Тогда $u(x) = x^3(x-1)(5x-7)\frac{\Delta^5 y_{-2}}{1\cdot 2\cdot 3\cdot 4}$. А уравнение искомой параболы $y(x)$ примет вид:

$$y(x) = y_{-2} + \frac{x+2}{1}\Delta y_{-2} + \frac{(x+2)(x+1)}{1\cdot 2}\Delta^2 y_{-2} + \frac{(x+2)(x+1)x}{1\cdot 2\cdot 3}\Delta^3 y_{-2} + \frac{(x+2)(x+1)x(x-1)}{1\cdot 2\cdot 3\cdot 4}\Delta^4 y_{-2} + \frac{x^3(x-1)(5x-7)}{1\cdot 2\cdot 3\cdot 4}\Delta^5 y_{-2}.$$

Полученная выше формула может быть использована для интерполяции между двумя опорными точками. Обозначим невыровненные одногодичные коэффициенты смертности m'_x, а выровненые — m_x. Если допустить, что $p_x\, p_{x+1}\, p_{x+2}\, p_{x+3}\, p_{x+4} = p'_x\, p'_{x+1}\, p'_{x+2}\, p'_{x+3}\, p'_{x+4}$, то при постоянстве силы смертности на протяжении одногодичного возрастного интервала имеем $p'_x = e^{-m'_x}$ и $p = e^{-m_x}$; тогда $e^{-(m_x+m_{x+1}+m_{x+2}+m_{x+3}+m_{x+4})} = e^{-(m'_x+m'_{x+1}+m'_{x+2}+m'_{x+3}+m'_{x+4})}$, т. е. $\sum_x^{x+4} m_x = \sum_x^{x+4} m'_x$.

Введем в рассмотрение функцию $y(x) = \sum_0^{x-1} m'_x$. Обозначим разности этой функции при приращении аргумента h, равного 5, Δy, а разности этой функции при приращении аргумента h, равного 1, δ'_y. Очевидно, что $\Delta y_z = y_{z+5} - y_z = m'_z + m'_{z+1} + m'_{z+2} + m'_{z+3} + m'_{z+4}$, т. е. Δy_z — это пятилетняя сумма невыровненных коэффициентов, а $\delta' y_z$ — это одногодичные коэффициенты смертности. Используя полученную формулу интерполирующей параболы $y(x)$, найдем зависимость между Δy_{-2}, Δy_{-1}, Δy_0, Δy_1, Δy_2 и $\delta y_{0,0}$, $\delta y_{0,2}$, $\delta y_{0,4}$, $\delta y_{0,6}$, $\delta y_{0,8}$.

При обработке материалов переписи населения 1926 г. С. А. Новосельский и В. В. Паевский поступали следующим образом.

В формуле интерполирующей параболы заменяли разности высоких порядков относительно y_{-2} разностями 1-го порядка при помощи следующих переходных формул:

$$\Delta^2 y_{-2} = \Delta y_{-1} - \Delta y_{-2},$$
$$\Delta^3 y_{-2} = \Delta y_0 - 2\Delta y_{-1} + \Delta y_{-2},$$
$$\Delta^4 y_{-2} = \Delta y_1 - 3\Delta y_0 + 3\Delta y_{-1} - \Delta y_{-2},$$
$$\Delta^5 y_{-2} = \Delta y_2 - 4\Delta y_1 + 6\Delta y_0 - 4\Delta y_{-1} + \Delta y_{-2}.$$

Придавая x значения 0,0; 0,2; 0,4; 0,6; 0,8; 1,0 и производя последовательное вычитание, находим:

$$\delta y_{0,0} = 0{,}0008(-16\Delta y_{-2} + 106\Delta y_{-1} + 188\Delta y_0 - 30\Delta y_1 + 2\Delta y_2),$$
$$\delta y_{0,2} = 0{,}0008(-2\Delta y_{-2} + 18\Delta y_{-1} + 278\Delta y_0 - 52\Delta y_1 + 8\Delta y_2),$$
$$\delta y_{0,4} = 0{,}0008(8\Delta y_{-2} - 42\Delta y_{-1} + 318\Delta y_0 - 42\Delta y_1 + 8\Delta y_2),$$
$$\delta y_{0,6} = 0{,}0008(8\Delta y_{-2} - 52\Delta y_{-1} + 278\Delta y_0 + 18\Delta y_1 - 2\Delta y_2),$$
$$\delta y_{0,8} = 0{,}0008(2\Delta y_{-2} - 30\Delta y_{-1} + 188\Delta y_0 + 106\Delta y_1 - 16\Delta y_2).$$

Эти формулы позволили, привлекая пятилетние интервалы, вычислить выровненные одногодичные коэффициенты смертности в пятилетнем интервале, находящемся в середине. При этом для каждого из пяти входящих в формулы пятилетий должны быть известны суммы одногодичных невыровненных коэффициентов. Постепенное передвижение позволяет определить плавный ряд одногодичных коэффициентов. Однако при этом 4 пятилетия остаются незаполненными (первые два и последние два).

Для возрастов 0—5 и 5—10 лет используется кривая $f(x)$ и в ней x задается равным $-2{,}0$; $-1{,}8$; $-1{,}6$; $-1{,}4$; $-1{,}2$; $-1{,}0$; $-0{,}8$; $-0{,}6$; $-0{,}4$; $-0{,}2$; 0. Затем путем последовательного вычитания находят выражения для 10 одногодичных разностей через пятилетние. Например, $\delta y_{-2,0} = 0{,}0008\ (452\,\Delta y_{-2} - 346\,\Delta y_{-1} + 186\,\Delta y_0 - 42\,\Delta y_1)$.

Аналогично поступаем с крайней группой старческих возрастов и находим 10 показателей от $\delta y_{1,0}$ до $\delta y_{1,8}$. Например, $\delta y_{1,0} = 0{,}0008(-18\Delta y_{-1} + 114\Delta y_0 + 176\Delta y_1 - 22\Delta y_2)$.

Все полученные нами формулы для $\delta\, y_x$ можно упростить, если использовать интерполяционный метод Гловера, рекомендованный С. А. Новосельским и В. В. Паевским и состоящий в использовании средней арифметической из пяти невыровненных одногодичных коэффициентов смертности:

$$\overline{m}_z = \frac{m'_z + m'_{z+1} + m'_{z+2} + m'_{z+3} + m'_{z+4}}{5}.$$

Тогда $\Sigma y_z = 5\overline{m}_z$.

Вместо $\overline{m}_z$ можно брать частные от деления пятилетней суммы умерших на соответствующую пятилетнюю сумму живущих. Если обозначить пять последовательных средних пятилетних коэффициентов через M_{-2}, M_{-1}, M_0, M_1, M_2, а выровненные 25 коэффициентов через m_{-10}, m_{-9}, ..., m_0, m_1, m_2, ..., m_{14}, то значения этих коэффициентов можно свести в табл. 75.

Практически вычисление одногодичных выровненных коэффициентов смертности рекомендуется производить с помощью трафарета[1].

[1] См. С. А. Новосельский и В. В. Паевский. Таблицы смертности населения СССР. В кн.: «Смертность и продолжительность жизни населения СССР. 1926—1927. Таблицы смертности», стр. XXII.

Таблица 75

Коэффициенты для вычисления одногодичных выровненных показателей смертности

	M_{-2}	M_{-1}	M_0	M_1	M_2	
$m_{-10}=8\cdot 10^{-3}$	226	—173	93	—21	0	Крайние молодые группы
$m_{-9}=8\cdot 10^{-3}$	165	—60	25	—5	0	
$m_{-8}=8\cdot 10^{-3}$	115	25	—20	5	0	
$m_{-7}=8\cdot 10^{-3}$	75	85	—45	10	0	
$m_{-6}=8\cdot 10^{-3}$	44	123	—53	11	0	
$m_{-5}=8\cdot 10^{-3}$	21	142	—47	9	0	
$m_{-4}=8\cdot 10^{-3}$	5	145	—30	5	0	
$m_{-3}=8\cdot 10^{-3}$	—5	135	—5	0	0	
$m_{-2}=8\cdot 10^{-3}$	—10	115	25	—5	0	
$m_{-1}=8\cdot 10^{-3}$	—11	88	57	—9	0	
$m_0=8\cdot 10^{-3}$	—8	53	94	—15	1	Средние группы
$m_1=8\cdot 10^{-3}$	—1	9	139	—26	4	
$m_2=8\cdot 10^{-3}$	4	—21	159	—21	4	
$m_3=8\cdot 10^{-3}$	4	—26	139	9	—1	
$m_4=8\cdot 10^{-3}$	1	—15	94	53	—8	
$m_5=8\cdot 10^{-3}$	0	—9	57	88	—11	Крайние старческие группы
$m_6=8\cdot 10^{-3}$	0	—5	25	115	—10	
$m_7=8\cdot 10^{-3}$	0	0	—5	135	—5	
$m_8=8\cdot 10^{-3}$	0	5	—30	145	5	
$m_9=8\cdot 10^{-3}$	0	9	—47	142	21	
$m_{10}=8\cdot 10^{-3}$	0	11	—53	123	44	
$m_{11}=8\cdot 10^{-3}$	0	10	—45	85	75	
$m_{12}=8\cdot 10^{-3}$	0	5	—20	25	115	
$m_{13}=8\cdot 10^{-3}$	0	—5	25	—60	165	
$m_{14}=8\cdot 10^{-3}$	0	—21	93	—173	226	

Спрейгом найдена также другая формула интерполяции, которую можно использовать для нахождения выравниваемого возраста путем привлечения 30 соседних возрастов. Через каждые 6 возрастов, отстоящих друг от друга на 5 одногодичных интервалов, проводят 5 парабол 5-го порядка, причем таким образом, что искомая парабола пройдет через каждое из двух разделенных 5 точками значений и будет иметь в них производные 1-го и 2-го порядка, совпадающие с аналогичными производными той параболы 4-го порядка, которая, в свою очередь, будет проложена через обе точки и пройдет через две последующие и две предшествующие точки. В дальнейшем эти 5 парабол служат для исчисления среднего значения для 15-го фактического уровня аналогично тому, как это делается в предыдущих методах (Воольхауза и Карупа).

Обозначим w_0 и w_1 два значения, через которые пройдет одна из пяти парабол. Отнесем точку w_0 в начало координат. Тогда уравнение этой параболы примет вид:

$$\overline{y_x}=w_0+ax+bx^2+cx^3+dx^4+ex^5.$$

Прохождение кривой через значение w_1 приводит к ряду условий. 1) $w_1=w_0+a+b+c+d+e$, так как парабола 4-го порядка для точек $w_{-2}, w_{-1}, w_0, w_1, w_2$ дает уравнение в конечных разностях:

$$\overline{y_x}=w_{-2}+\frac{x+2}{1}\Delta w_{-2}+\frac{(x+2)(x+1)}{1\cdot 2}\Delta^2 w_{-2}+$$
$$+\frac{(x+2)(x+1)x}{1\cdot 2\cdot 3}\Delta^3 w_{-2}+\frac{(x+2)(x+1)x(x-1)}{1\cdot 2\cdot 3\cdot 4}\Delta^4 w_{-2},$$

где $\Delta w_{-2}=w_{-1}-w_{-2}$.

Так как согласно условию две первые производные этой параболы совпадают с соответствующими производными искомой параболы, получаем:

$$2)\quad \left(\frac{dy}{dx}\right)_0=a=\Delta w_{-2}+\frac{3}{2}\Delta^2 w_{-2}+\frac{1}{3}\Delta^3 w_{-2}-\frac{1}{12}\Delta^4 w_{-2},$$

$$3)\quad \left(\frac{d^2 y}{dx^2}\right)_0=2b=\Delta^2 w_{-2}+\Delta^3 w_{-2}-\frac{1}{12}\Delta^4 w_{-2}.$$

Однако в связи с требованием совпадения производных параболы 5-го порядка в точке 1 с параболой 4-го порядка

$$\breve{y}_x=w_{-1}+\frac{x+1}{1}\Delta w_{-1}+\frac{(x+1)x}{1\cdot 2}\Delta^2 w_{-1}+$$
$$+\frac{(x+1)x(x-1)}{1\cdot 2\cdot 3}\Delta^3 w_{-1}+\frac{(x+1)x(x-1)(x-2)}{1\cdot 2\cdot 3\cdot 4}\Delta^4 w_{-1}$$

либо с тождественной ей параболой

$$w_{-2}+\frac{x+2}{1}\Delta w_{-2}+\frac{(x+2)(x+1)}{1\cdot 2}\Delta^2 w_{-2}+\frac{(x+2)(x+1)x}{1\cdot 2\cdot 3}\Delta^3 w_{-2}+$$
$$+\frac{(x+2)(x+1)x(x-1)}{1\cdot 2\cdot 3\cdot 4}\Delta^4 w_{-2}+\frac{(x+1)x(x-1)(x-2)}{1\cdot 2\cdot 3\cdot 4}\Delta^5 w_{-2}.$$

Отсюда необходимо выполнение еще 2 условий:

$$4)\quad a+2b+3c+4d+5e=\Delta w_{-2}+\frac{5}{2}\Delta^2 w_{-2}+\frac{11}{16}\Delta^3 w_{-2}+$$
$$+\frac{1}{4}\Delta^4 w_{-2}-\frac{1}{12}\Delta^5 w_{-2},$$

$$5)\quad 2b+6c+12d+20e=\Delta^2 w_{-2}+2\Delta^3 w_{-2}+\frac{11}{12}\Delta^4 w_{-2}-\frac{1}{12}\Delta^5 w_{-2}.$$

Решив указанные 5 уравнений, получим параметры (при записи $\Delta^k w_{-2}$ обозначаем через Δ^k);

$$
\begin{aligned}
a &= \Delta + \frac{3}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \frac{1}{12}\Delta^4,\\
b &= \frac{1}{2}\Delta^2 + \frac{1}{2}\Delta^3 - \frac{1}{24}\Delta^4;\\
c &= \frac{1}{6}\Delta^3 + \frac{1}{12}\Delta^4 + \frac{7}{24}\Delta^5,\\
d &= \frac{1}{24}\Delta^4 - \frac{1}{2}\Delta^5,\\
e &= \frac{5}{24}\Delta^5.
\end{aligned}
$$

Искомое уравнение будет выглядеть так:

$$
\overline{y_x} = w_0 + \left(\Delta + \frac{3}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \frac{1}{12}\Delta^4\right)x + \left(\frac{1}{2}\Delta^2 + \frac{1}{2}\Delta^3 - \frac{1}{24}\Delta^4\right)x^2 + \left(\frac{1}{6}\Delta^3 + \frac{1}{12}\Delta^4 + \frac{7}{24}\Delta^5\right)x^3 + \left(\frac{1}{24}\Delta^4 - \frac{1}{2}\Delta^5\right)x^4 + \frac{5}{24}\Delta^5 x^5.
$$

Данная формула позволяет производить выравнивание и находить промежуточные члены между опорными с помощью разностей и начального члена в рассматриваемом интервале.

Из этого основного уравнения выводится уравнение для всех 5 парабол с вершиной в начальной точке координат, если к значениям w_0 и w_1 приплюсовать по порядку значения в точках 10 и 15, 11 и 16, 12 и 17, 13 и 18, 14 и 19.

Но от каждой из этих парабол всегда требуется только одно значение: $x = \frac{2}{5}$ той параболы, которая проходит через точки 10 и 15; значения для $x = \frac{1}{5}$ той параболы, которая проходит через точки 11 и 16; значение для $x = 0$ той параболы, которая проходит через 12 и 17 и т. д. Когда складываются исчисленные таким образом значения и сумма делится на 5, в конечном итоге получается следующая формула

$$
\begin{aligned}
\overline{y_x} = 0{,}2\,w_x &+ 0{,}18688\,(w_{x+1} + w_{x-1}) + 0{,}14528\,(w_{x+2} + w_{x-2}) +\\
&+ 0{,}08768\,(w_{x+3} + w_{x-3}) + 0{,}03488\,(w_{x+4} + w_{x-4}) -\\
&- 0{,}01952\,(w_{x+6} + w_{x-6}) - 0{,}02272\,(w_{x+7} + w_{x-7}) -\\
&- 0{,}01472\,(w_{x+8} + w_{x-8}) - 0{,}00512\,(w_{x+9} + w_{x-9}) +\\
&+ 0{,}00256\,(w_{x+11} + w_{x-11}) + 0{,}00288\,(w_{x+12} + w_{x-12}) +\\
&+ 0{,}00160\,(w_{+x13} + w_{x-13}) + 0{,}00032\,(w_{x+14} + w_{x-14}).
\end{aligned}
$$

Вернемся к формуле

$$
\overline{y_x} = w_0 + x\left(\Delta + \frac{3}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \frac{1}{12}\Delta^4\right) + x^2\left(\frac{1}{2}\Delta^2 + \frac{1}{2}\Delta^3 - \frac{1}{24}\Delta^4\right) + x^3\left(\frac{1}{6}\Delta^3 + \frac{1}{12}\Delta^4 + \frac{7}{24}\Delta^5\right) +
$$

$$
+ x^4\left(\frac{1}{24}\Delta^4 - \frac{1}{2}\Delta^5\right) + \frac{5}{24}\Delta^5 x^5
$$

и, проводя интерполяцию между опорными возрастами y_0 и y_1, вставим 4 новых равноотстоящих члена, приписывая x значения 0, $\frac{1}{5}$, $\frac{2}{5}$, $\frac{3}{5}$, $\frac{4}{5}$, 1. Вычитая полученные значения, найдем выражения для разностей начального члена y_{-2}:

$$
\begin{aligned}
\delta y_0 &= \frac{\Delta y_{-2}}{5} + 8\,\frac{\Delta^2 y_{-2}}{5^2} + 11\,\frac{\Delta^3 y_{-2}}{5^3} - 11\,\frac{\Delta^4 y_{-2}}{5^4} + \frac{\Delta^5 y_{-2}}{5^4},\\
\delta^2 y_0 &= \frac{\Delta^2 y_{-2}}{5^2} + 6\,\frac{\Delta^3 y_{-2}}{5^3} + \frac{\Delta^4 y_{-2}}{5^4} + 3\,\frac{\Delta^5 y_{-2}}{5^4},\\
\delta^3 y_0 &= \frac{\Delta^3 y_{-2}}{5^3} + 4\,\frac{\Delta^4 y_{-2}}{5^4} - 3\,\frac{\Delta^5 y_{-2}}{5^4},\\
\delta^4 y_0 &= \frac{\Delta^4 y_{-2}}{5^4} - 2\,\frac{\Delta^5 y_{-2}}{5^4},\\
\delta^5 y_0 &= 5\,\frac{\Delta^5 y_{-2}}{5^4}.
\end{aligned}
$$

Рассмотрим применение этой формулы для интерполяции вероятности смерти на 4-м году по данным таблицы смертности, составленной применительно к застраховавшимся в возрасте x лет.

x	20	25	30	35	40	45
$q_{x/3}$	0,010709	0,009585	0,009225	0,010177	0,011313	0,013955

Составим таблицу разностей.

Таблица 76

x	$q_{x/3}$	Δ	Δ^2	Δ^3	Δ^4	Δ^5
20	0,010709					
		−1 124				
25	0,009585		+764			
		−360		+548		
30	0,009225		+1 312		−1 676	
		+952		−1 128		+4 126
35	0,010177		+184		+2 450	
		+1 136		+1 322		
40	0,011313		+1 506			
		+2 642				
45	0,013955					

Подготовительные расчеты приведены в табл. 77.

Таблица 77

Расчеты для интерполяции

x	$\frac{\Delta}{5}$	$\frac{\Delta^2}{5^2}$	$\frac{\Delta^3}{5^3}$	$\frac{\Delta^4}{5^4}$	$\frac{\Delta^5}{5^4}$
20	−224,80	+30,56	+4,38	−2,68	+6,60
25	−72,00	+52,48	−9,02	+3,92	−6,13
30	+190,40	+7,36	+10,58	−2,21	
35	+227,20	+60,24	−0,45		
40	+528,40	+58,00			
45	+818,40				

Интерполируем интервал 30—35 лет. Для этого используем шесть показателей; $q_{20/3}$, $q_{25/3}$, $q_{30/3}$, $q_{35/3}$, $q_{40/3}$, $q_{45/3}$. Возьмем из предыдущей таблицы первый ряд разностей: —224,80; +30,56; +4,38; —2,68; +6,60 и вычисляем одногодичные разности:

$$\delta y_{30/3} = -224{,}80 + 8 \cdot 30{,}56 + 11 \cdot 4{,}38 - 11 \cdot (-2{,}68) + 6{,}60 = -224{,}80 + 244{,}48 + 48{,}18 + 29{,}48 + 6{,}60 = +103{,}94;$$

$$\delta^2 y_{30/3} = 30{,}56 + 26{,}28 - 2{,}68 + 19{,}80 = 73{,}96;$$

$$\delta^3 y_{30/3} = 4{,}38 - 10{,}72 - 19{,}80 = -26{,}14;$$

$$\delta^4 y_{30/3} = -2{,}68 - 13{,}20 = -15{,}88;$$

$$\delta^5 y_{30/3} = +33{,}00.$$

Составляем табл. 78.

Таблица 78

Интерполяция между опорными возрастами

$\frac{\Delta^5}{5^4}$	$\frac{\Delta^4}{5^4}$	$\frac{\Delta^3}{5^3}$	$\frac{\Delta^2}{5^2}$	$\frac{\Delta}{5}$	Накопленные разности	$q_{x/3}$	x
+6,60	−2,68	+4,38	30,56	−224,80	—	0,00922500	30
+6,60	+29,48	+48,18	+244,48	−224,80	**+103,94**	0,00932894	31
+19,80	−2,68	+26,28	+30,56	**+73,96**	+177,90	0,00950684	32
−19,80	−10,72	+4,38	**−26,14**	+47,82	+225,72	0,00973256	33
−13,20	−2,68	**−15,88**	−42,02	+5,80	+231,52	0,00996408	34
+33,00	**+33,00**	+17,12	−24,90	−19,10	+212,42	0,01017650	35

Эта таблица позволяет путем суммирования получать разности вероятностей умереть δ, а затем и сами вероятности. Так, величина опорного возраста 30 лет равна $q_{30/3} = 0{,}00922500$. Откуда

$$q_{31/3} = 0{,}00922500 + 0{,}00010394 = 0{,}00932894,$$

$$q_{32/3} = 0{,}00932894 + 0{,}00017790 = 0{,}00950684 \text{ и т. д.}$$

Фактическая вероятность умереть в следующем опорном возрасте, т. е. 35 лет, и выровненная вероятность этого же возраста должны совпадать. В нашем примере $q_{35/3}$ фактическое составляет 0,01017700, а $q_{35/3}$ выровненное — 0,01017650, т. е. мало отличаются друг от друга. Аналогично можно продолжить процесс выравнивания и дальше.

Выравнивание численности населения статистиками ООН

При выравнивании численности населения в возрасте от 10 до 74 лет по пятилетним интервалам отдел по народонаселению социального бюро ООН использует метод, также имеющий в основе параболу 5-го порядка. Для внесения поправки в численность какой-нибудь пятилетней возрастной группы в расчет включаются данные за две предшествующие и за две последующие пятилетние группы. Обозначаем:

S — исправленное число лиц в интерполируемом пятилетнем интервале;

Σ — фактически зарегистрированное число лиц в пятилетних интервалах; при этом Σ_{-2} и Σ_{-1} численности в предшествующих интерполируемому пятилетних возрастах; Σ_0 — в интерполируемом интервале, а Σ_{+1} и Σ_2 — в двух последующих за интерполируемым пятилетних возрастах.

Формула имеет вид:

$$S = \frac{1}{16}(-\Sigma_{-2} + 4\Sigma_{-1} + 10\Sigma_0 + 4\Sigma_{+1} - \Sigma_{+2}).$$

В качестве примера возьмем уже использованные ранее данные (стр. 243) и найдем выровненную численность населения в возрасте 45—49 лет.

Имеем:

$$\Sigma_{-2} = 2789{,}9; \quad \Sigma_{-1} = 2306{,}7; \quad \Sigma_0 = 1990{,}5;$$
$$\Sigma_1 = 1653{,}1; \quad \Sigma_2 = 1350{,}0;$$

$$S = \frac{1}{16}(-2789{,}9 + 4 \cdot 2306{,}7 + 10 \cdot 1990{,}5 + 4 \cdot 1653{,}1 - 1350{,}0) = 1975{,}3.$$

Для интерполяции же численности населения по одногодичным возрастам в пределах пятилетних групп отдел по народонаселению социального бюро ООН предлагает метод Спрейга — Гловера (множители Спрейга), основанный также на привлечении данных за две предшествующие и две последующие пятилетние группы.

Обозначим дополнительно численность интерполируемых одногодичных возрастов центральной пятилетней группы через y_{-2}, y_{-1}, y_0, y_{+1} и y_{+2}. Тогда для определения интерполированной численности каждой одногодичной возрастной группы показатели нескольких пятилетних групп умножаются на коэффициенты, приведенные в соответствующей строке табл. 79, и результаты складываются.

Таблица 79

y	Σ_{-2}	Σ_{-1}	Σ_0	Σ_{+1}	Σ_{+2}
y_{-2}	−0,0128	+0,0848	+0,1504	−0,0240	+0,0016
y_{-1}	−0,0016	+0,0144	+0,2224	−0,0416	+0,0064
y_0	+0,0064	−0,0336	+0,2544	−0,0336	+0,0064
y_{+1}	+0,0064	−0,0416	+0,2224	+0,0144	−0,0016
y_{+2}	+0,0016	−0,0240	+0,1504	+0,0848	−0,0128

Например, $y_{-2} = -0{,}0128\,\Sigma_{-2} + 0{,}0848\,\Sigma_{-1} + 0{,}1504\,\Sigma_0 - 0{,}0240\,\Sigma_1 + 0{,}0016\Sigma_2$ и т. д. При этом методе первые две пятилетние возрастные группы, т. е. население до 10 лет, и последние две группы не могут быть интерполированы по этим множителям; для них множители приведены в табл. 80.

Таблица 80

Множители Спрейга

n	N_1	N_2	N_3	N_4	N_5
			Первая конечная группа		
n_1	+0,3616	−0,2768	+0,1488	−0,0336	—
n_2	+0,2640	−0,0960	+0,0400	−0,0080	—
n_3	+0,1840	+0,0400	−0,0320	+0,0080	—
n_4	+0,1200	+0,1360	−0,0720	+0,0160	—
n_5	+0,0704	+0,1968	−0,0848	+0,0176	—
			Первая ближайшая к конечной группа		
n_1	+0,0336	+0,2278	−0,0752	+0,0144	—
n_2	+0,0080	+0,2320	−0,0480	+0,0080	—
n_3	−0,0080	+0,2160	−0,0080	+0,0000	—
n_4	−0,0160	+0,1840	+0,0400	−0,0080	—
n_5	−0,0176	+0,1408	+0,0912	−0,0144	—
			Средняя группа		
n_1	−0,0128	+0,0848	+0,1504	−0,0240	+0,0016
n_2	−0,0016	+0,0144	+0,2224	−0,0416	+0,0064
n_3	+0,0064	−0,0336	+0,2544	−0,0336	+0,0064
n_4	+0,0064	−0,0416	+0,2224	+0,0144	−0,0016
n_5	+0,0016	−0,0240	+0,1504	+0,0848	−0,0128
			Последняя ближайшая к конечной группа		
n_1	−0,0144	+0,0912	+0,1408	−0,0176	—
n_2	−0,0080	+0,0400	+0,1840	−0,0160	—
n_3	+0,0000	−0,0080	+0,2160	−0,0080	—
n_4	+0,0080	−0,0480	+0,2320	+0,0080	—
n_5	+0,0144	−0,0752	+0,2272	+0,0336	—
			Последняя конечная группа		
n_1	+0,0176	−0,0848	+0,1968	+0,0704	—
n_2	+0,0160	−0,0720	+0,1360	+0,1200	—
n_3	+0,0080	−0,0320	+0,0400	+0,1840	—
n_4	−0,0080	+0,0400	−0,0960	+0,2640	—
n_5	−0,0336	+0,1488	−0,2768	+0,3616	—

Произведем интерполирование по этой формуле данных, уже использованных нами при рассмотрении метода Б. С. Ястремского (см. стр. 243).

Получаем для возраста 45 лет:

$$\bar{y}_{-2} = -0{,}0128\cdot 2789{,}9 + 0{,}0848\cdot 2306{,}7 + 0{,}1504\cdot 1990{,}5 - 0{,}0240\cdot 1653{,}1 + 0{,}0016\cdot 1350{,}0 \approx 424{,}9 \text{ и т. д.}$$

Метод Спрейга — Гловера дает определенную систему множителей. С. А. Новосельский и В. В. Паевский, приспособив данный метод для интерполяции коэффициентов смертности, получили другие множители (см. табл. 75). Для получения спрейговских множителей нужно множители, рекомендованные С. А. Новосельским и В. В. Паевским, умножить на 0,2.

ВЫРАВНИВАНИЕ ПОКАЗАТЕЛЕЙ, ОТНОСЯЩИХСЯ К СТАРЧЕСКИМ ВОЗРАСТАМ

Метод Гомперца — Макегама

Принято считать, что первой более или менее удачной формулой, предложенной для выравнивания (интерполирования) смертности, является формула Гомперца[1], усовершенствованная актуарием Макегамом. Эту кривую иногда называют кривой роста, используют в страховом деле и при построении таблиц смертности, т. е. в тех случаях, когда требуется кумулятивное возрастание к некоторому максимальному уровню. Это имеет место, когда приращения выравниваемых уровней все время уменьшаются. Поэтому особенно часто этот метод применяется для выравнивания уровней, относящихся к старческим возрастам.

С возраста 70 лет и более коэффициенты и вероятности дожития (или смерти) дают значительные колебания и скачки, которые могут быть объяснены неточностью исходных возрастных данных, а помимо этого иногда и малыми абсолютными величинами.

Гомперц предложил свою формулу, полагая, что смерть вызывается совокупностью причин, действующих в связи с возрастом x, усиливающих свое влияние в геометрической прогрессии по отношению к возрасту:

$$\mu_x = bc^x,$$

где μ_x — сила смертности,
x — возраст.

[1] Compertc. On the nature of the function exspressive of the law human mortality. London. Transactions, 1825.

Макегам полагал, что, помимо факторов первого рода, т. е. возрастных, действующих на смертность, имеются еще факторы, действующие на смертность независимо от возраста.

$$\mu_x = a + bc^x.$$

Преобразуем формулу силы смертности, учитывая, что

$$\mu_x = -\frac{l'(x)}{l(x)} = -[\ln l(x)]' \quad \text{и} \quad a + bc^x = \left[ax + \frac{b}{\ln c} c^x\right]'.$$

Тогда

$$[\ln l(x)]' = -\left[ax + \frac{b}{\ln c} c^x\right]'.$$

Так как производные двух выражений равны, то эти выражения отличаются на одинаковую величину, которую обозначим $\ln k$. Значит,

$$\ln l_x = \ln k - ax - \frac{b}{\ln c} c^x.$$

Предположим, что

$$-a = \ln S \quad \text{и} \quad -\frac{b}{\ln c} = \ln z.$$

Имеем:

$$\ln l_x = \ln k + x \ln S + c^x \ln z;$$

переходя от логарифмов, получаем:

$$l_x = kS^x z^{c^x}.$$

Практическое использование метода Гомперца—Макегама по Кинг—Харди

Использование формулы Гомперца — Макегама в практических целях возможно при методе Кинг — Харди. Идея этого метода состоит в том, что весь подлежащий выравниванию ряд делится на четыре равных возрастных группы. В каждой группе вычисляют сумму логарифмов фактических значений, конечные разности 1-го и 2-го порядка, а затем находят все параметры. Возьмем $x = 17$, а $t = 18$ и найдем четыре соответствующие суммы доживающих в пределах возрастов:

от x до $x + t - 1$, т. е. от 17 до 34 лет,
от $x + t$ до $x + 2t - 1$, т. е. от 35 до 52 лет,
от $x + 2t$ до $x + 3t - 1$, т. е. от 53 до 70 лет,
от $x + 3t$ до $x + 4t - 1$, т. е. от 71 до 88 лет.

Сумма $I = \sum\limits_{x}^{x+t-1} \ln l_x = \sum\limits_{x}^{x+t-1} (\ln k + x \ln S + \ln zc^x) =$

$$= t \ln k + [x + (x+1) + (x+2) + \ldots + (x+t-1)] \ln S +$$
$$+ c^x (1 + c + c^2 + \ldots + c^{t-1}) \ln z = t \ln k + \frac{t}{2}(2x + t - 1) \ln S +$$
$$+ c^x \left(\frac{c^t - 1}{c - 1}\right) \ln z.$$

Аналогично сделаем другие вычисления и после преобразований получим:

$$\text{Сумма II} = \sum_{x+t}^{x+2t-1} \ln l_x = t \ln k + \frac{t}{2}(2x + 3t - 1) \ln S + c^{x+t} \frac{c^t - 1}{c - 1} \ln z;$$

$$\text{Сумма III} = \sum_{x+2t}^{x+3t-1} \ln l_x = t \ln k + \frac{t}{2}(2x + 5t - 1) \ln S + c^{x+2t} \frac{c^t - 1}{c - 1} \ln z;$$

$$\text{Сумма IV} = \sum_{x+3t}^{x+4t-1} \ln l_x = t \ln k + \frac{t}{2}(2x + 7t - 1) \ln S + c^{x+3t} \frac{c^t - 1}{c - 1} \ln z.$$

Находим конечные разности сумм 1-го порядка:

$$\text{I} - \text{II} = \Delta\text{II} = -t^2 \ln S - c^x \frac{(c^t - 1)^2}{c - 1} \ln z;$$
$$\text{II} - \text{III} = \Delta\text{III} = -t^2 \ln S - c^{x+t} \frac{(c^t - 1)^2}{c - 1} \ln z;$$
$$\text{III} - \text{IV} = \Delta\,\text{IV} = -t^2 \ln S - c^{x+2t} \frac{(c^t - 1)^2}{c - 1} \ln z;$$

затем конечные разности сумм 2-го порядка:

$$\Delta\,\text{II} - \Delta\,\text{III} = \Delta^2\,\text{III} = c^x \frac{(c^t - 1)^3}{c - 1} \ln z;$$
$$\Delta\,\text{III} - \Delta\,\text{IV} = \Delta^2\,\text{IV} = c^{x+t} \frac{(c^t - 1)^3}{c - 1} \ln z.$$

Разделим вторые разности и получим:

$$c^t = \frac{\Delta^2\,\text{IV}}{\Delta^2\,\text{III}}.$$

Теперь находим:

$$\ln S = \frac{1}{t^2} [\Delta\text{II} + c^{x \ln c} + 2 \ln (c^t - 1) + \ln \ln (-z) - \ln (c - 1)];$$

$$\ln k = \frac{1}{t}\left[\text{I} - \frac{t}{2}(2x - t - 1) \ln S - c^x \left(\frac{c^t - 1}{c - 1}\right) \ln z\right],$$

Упрощенный метод Гомперца—Макегама

Формула Гомперца — Макегама может быть использована в более простой форме: $y = a \cdot b^{c^x}$ или $\lg y = \lg a + (\lg b)c^x$. Первое слагаемое ($\lg a$) максимальный уровень, а второе — разность между выровненным уровнем и максимумом в каждом последующем возрасте.

Практически поступают так: делят весь возрастной интервал на 3 части, суммы логарифмов каждой части обозначают S_1, S_2, S_3, $S_2 - S_1 = d_1$, $S_3 - S_2 = d_2$. Тогда $\frac{d_2}{d_1} = c^n$.

Имеем:

$$S_1 = \Sigma \lg y_1 = n \lg a + (\lg b)\, c\, \Sigma x_1;$$
$$S_2 = \Sigma \lg y_2 = n \lg a + (\lg b)\, c\, \Sigma x_2;$$
$$S_3 = \Sigma \lg y_3 = n \lg a + (\lg b)\, c\, \Sigma x_3;$$
$$d_1 = S_2 - S_1 = (\lg b)\, c\, (\Sigma x_2 - \Sigma x_1);$$
$$d_2 = S_3 - S_2 = (\lg b)\, c\, (\Sigma x_3 - \Sigma x_2).$$

Для определения параметров a, b и c используют следующие соотношения:

$$c^n = \frac{d_2}{d_1} = \frac{\Sigma x_3 - \Sigma x_2}{\Sigma x_2 - \Sigma x_1}; \qquad c = \sqrt[n]{\frac{\Sigma x_3 - \Sigma x_2}{\Sigma x_2 - \Sigma x_1}};$$
$$\lg b = \frac{d_1 (c-1)}{(c^n - 1)^2};$$
$$\lg a = \frac{1}{n}\left(S_1 - \frac{d_1}{c^n - 1}\right).$$

Схема расчетов по формуле Гомперца — Макегама дана в табл. 81.

Таблица 81

Возраст	Числа, подлежащие выравниванию, y	$\lg y$	Частные итоги	Первые разности
			S_1 S_2 S_3	$S_2 - S_1 = d_1$ $S_3 - S_2 = d_2$

Затем находим c, а далее расчет ведется по табл. 82.

Таблица 82

Возраст x	c^x	$(\lg b)\ c^x$	$\lg y = \lg a + (\lg b)\ c^x$	Антилогарифмы, т. е. выровненные значения
0	1,00000			
1				
2				
3				
и т. д.				

Формула Тиле

Одна из модификаций формулы Гомперца — Макегама — формула Тиле, которая является выражением показательной кривой и используется при выравнивании вероятностей умереть для старческих возрастов:

$$\mu_x = ae^{bx},$$

где $\mu_x = \frac{2q_x}{1 + p_x}$; x — возраст.

Зная q_x и p_x для двух возрастов, можно определить a и b, затем, решая уравнение относительно μ_x для произвольного x, находить q_x для любого возраста.

Используем формулу Тиле для выравнивания данных о населении Уральской области старше 87 лет по переписи 1926 г. Примем в качестве известных величин:

$$q_{82} = 0{,}10894 \quad \text{и} \quad q_{87} = 0{,}14473;$$
$$p_{82} = 0{,}89106 \quad \text{и} \quad p_{87} = 0{,}85527.$$

Получаем:

$$\mu_{82} = \frac{2 \cdot 0{,}10894}{1 + 0{,}89106} = 0{,}11522,$$
$$\mu_{87} = \frac{2 \cdot 0{,}14473}{1 + 0{,}85527} = 0{,}15602.$$

Решаем систему уравнений относительно a и b.

$$\begin{cases} ae^{82b} = 0{,}11522, \\ ae^{87b} = 0{,}15602 \end{cases}$$

и получаем:

$$a = 0{,}0007988; \quad b = 0{,}060628.$$

Следовательно,

$$\mu_x = 0{,}0007988\, e^{0{,}060628\, x}.$$

Подставляя значения x, равные 88, 89 и т. д., получаем μ_{88}, μ_{89} и т. д., а затем $q_{88} = 0{,}15308$, $q_{89} = 0{,}16187$ и т. д. Интервал 82—87 лет может быть заполнен по параболе какого-нибудь порядка. Для населения городов Восточной Сибири при обработке данных переписи 1926 г. заполнение возрастного интервала 82—86 лет производилось по параболе 4-го порядка.

Формула Гомперца — Макегама со всеми ее модификациями является очень гибкой, дает плавный выровненный ряд. Однако использование ее имеет ограничения. Для детских возрастов она вообще неприменима, а для возрастов до 30 лет, как показали произведенные нами расчеты, она завышает коэффициенты. Большим недостатком этой формулы является то, что при вычислении параметров k, S, c и q по способу Кинг — Харди выровненные значения зависят от того, какие возраста мы принимаем за опорные. Как выход при расчетах параметров можно предложить использование метода наименьших квадратов.

Еще одна модификация формулы Гомперца — Макегама

При выравнивании вероятности дожития применяют еще одну модификацию формулы Гомперца — Макегама:

$$p_x = 10^{a+bc^x}$$

или

$$\lg p_x = a + bc^x.$$

Эта формула часто используется для старческих возрастов, реже для средних возрастов. Допустим, что для каких-нибудь трех возрастов, образующих арифметическую прогрессию (например, 50, 60, 70; 60, 70, 80; 70, 80, 90 или любых других, некратных десяти) известны вероятности дожития p_x.

Возьмем:

$$A = 50,\ k = 10,\ A + k = 60,\ A + 2k = 70.$$

Тогда

$$p_A = 10^{a+bc^A};\ p_{A+k} = 10^{a+bc^{A+k}};\ p_{A+2k} = 10^{a+bc^{A+2k}}.$$

После логарифмирования получим:

$$\lg p_A = a + bc^A;$$
$$\lg p_{A+k} = a + bc^{A+k};$$
$$\lg p_{A+2k} = a + bc^{A+2k}.$$

Находим отношение двух первых разностей и, произведя преобразования, получим:

$$\frac{\lg p_{A+2k} - \lg p_{A+k}}{\lg p_{A+k} - \lg p_A} = c^k.$$

Откуда легко определяется c. Первую разность

$$\lg p_{A+k} - \lg p_A = bc^A (c^k - 1)$$

разделим на $c^k - 1$ и получим:

$$bc^A = \frac{\lg p_{A+k} - \lg p_A}{c^k - 1},$$

откуда определяется b.

Параметр a определяется из разности

$$\lg p_A - bc^A = a + bc^A - bc^A = a.$$

Рассмотрим применение этого метода на примере возрастов 70, 80 и 90 лет таблицы смертности городского и сельского населения 1958—1959 гг. Имеем: $p_{70} = 0{,}96650$; $p_{80} = 0{,}91949$; $p_{90} = 0{,}84674$; $A = 70$; $k = 10$.

Находим:

$$\lg p_{70} = -0{,}0147981;\ \lg p_{80} = -0{,}0364530;\ \lg p_{90} = -0{,}0722499;$$

$$\lg p_{80} - \lg p_{70} = -0{,}0216549;\quad \lg p_{90} - \lg p_{80} = -0{,}0357969;$$

$$c = \sqrt[10]{\frac{-0{,}0357969}{-0{,}0216549}} = \sqrt[10]{1{,}6530623} = 1{,}0515475.$$

Находим $c^{70} = (1{,}0515475)^{70} = 33{,}7306822$ и получаем:

$$bc^{70} = b\,33{,}7306822 = \frac{\lg p_{80} - \lg p_{70}}{c^k - 1} = \frac{-0{,}0216549}{0{,}6530623}.$$

Откуда $b = -0{,}00098305$.

Предварительно определив $c^{95} = (1{,}0515475)^{95} = 118{,}5081967$ и параметр $a = \lg p_{70} - bc^{70} = -0{,}0147981 - (-0{,}00098305) \cdot 33{,}7306822 = -0{,}0147981 + 0{,}0331606 = 0{,}0183625$, можно найти p_{95}:

$\lg p_{95} = a + bc^{95} = 0{,}0183625 - 0{,}00098305 \cdot 118{,}5081967 = -0{,}0981387$, откуда $p_{95} = 0{,}79774$.

Интерполяционная формула Ньютона и ее использование при выравнивании по параболической кривой

Для выравнивания вероятностей дожития или смерти старческих возрастов может быть использована интерполяционная формула Ньютона.

Предположим, что вероятность дожития p_x начиная с какого-нибудь возраста укладывается в логарифмическую кривую $z = \lg p_x$. Выбираем сначала в качестве опорных четыре следующие друг за другом вероятности дожития p_x, p_{x+1}, p_{x+2}, p_{x+3}, а затем еще одну, внушающую уверенность в ее правильности и отстоящую от p_{x+3} на несколько лет — p_{x+y}. Эта вероятность может быть фактической или полученной ранее каким-либо методом выравнивания.

Берем $\lg p_x$, $\lg p_{x+1}$, $\lg p_{x+2}$, $\lg p_{x+3}$, $\lg p_{x+y}$ и определяем разности 1, 2, и 3-го порядков, т. е. Δ, Δ^2 и Δ^3.

Используем интерполяционную формулу Ньютона для определения разности 4-го порядка. При разности аргумента $h = 1$ имеем:

$$u_{k+y} = u_k + y\,\Delta u_k + \frac{y(y-1)}{2!}\Delta^2 u_k + \frac{y(y-1)(y-2)}{3!}\Delta^3 u_k +$$
$$+ \frac{y(y-1)(y-2)(y-3)}{4!}\Delta^4 u_k + \frac{y(y-1)(y-2)(y-3)(y-4)}{5}\Delta^5 u_k + \ldots$$

Примем разность 4-го порядка постоянной, а значит, разность 5-го порядка равной нулю, и для возраста старше 70 лет ($x > 70$) получаем:

$$\lg p_{x+y} = \lg p_x + y\,\Delta \lg p_x + \frac{y(y-1)}{2!}\Delta^2 \lg p_x +$$
$$+ \frac{y(y-1)(y-2)}{3!}\Delta^3 \lg p_x + \frac{y(y-1)(y-2)(y-3)}{4!}\Delta^4 \lg p_x.$$

Зная $\lg p_{x+y}$ и разности 1, 2 и 3-го порядков, можно определить четвертую разность, т. е. $\Delta^4 \lg p_x$. Производя дальнейшее суммирование, получаем все необходимые вероятности дожития.

Допустим, что мы имеем:

$$p_{74}=0{,}89921; \qquad \lg p_{74}=\bar{1}{,}950465;$$
$$p_{75}=0{,}88641; \qquad \lg p_{75}=\bar{1}{,}947635;$$
$$p_{76}=0{,}88071; \qquad \lg p_{76}=\bar{1}{,}944835;$$
$$p_{77}=0{,}87516; \qquad \lg p_{77}=\bar{1}{,}942090;$$
$$p_{82}=p_{74+8}=0{,}85014; \qquad \lg p_{78}=\bar{1}{,}929483.$$

Находим разности:

$$\Delta \lg p_{74}=\bar{1}{,}947635-\bar{1}{,}950465=-2831^1;$$
$$\Delta^2 \lg p_{74}=82; \quad \Delta^3 \lg p_{74}=-125 \quad \text{и} \quad y=82-74=8.$$

Подставляем все данные в формулу и решаем уравнение относительно $\Delta^4 \lg p_{74}$:

$$929\,483=950\,465+8(-2\,831)+\frac{8\cdot 7}{1\cdot 2}\cdot 82+$$
$$+\frac{8\cdot 7\cdot 6}{1\cdot 2\cdot 3}(-125)+\frac{8\cdot 7\cdot 6\cdot 5}{1\cdot 2\cdot 3\cdot 4}\Delta^4 \lg p_{74}.$$

Откуда $\Delta^4 \lg p_{74}=91$.

Теперь получаем суммированием, например, $\lg p_{78}$:

$$\lg p_{78}=91+4\lg p_{77}-6\lg p_{76}+4\lg p_{75}-\lg p_{74}=\bar{1}{,}93922.$$

Откуда $p_{78}=0{,}86940$ и т. д.

Этот способ выравнивания применялся при обработке данных переписи населения СССР 1926 г. для населения городов Восточной Сибири. Использовались возраста 74, 75, 76, 77 и 82 года.

Интерполяционную формулу Ньютона можно использовать при выравнивании по параболической кривой, например при выравнивании коэффициентов смертности или вероятности умереть в старческих возрастах. По-прежнему главным условием выравнивания является сохранение основных черт тенденций предшествующих возрастов и совпадение сумм фактических коэффициентов (или вероятностей) с суммой выровненных на интерполируемом участке. Кроме того, вводится дополнительное условие для очень старческих возрастов: вероятность умереть должна получать значения, весьма близкие к достоверным.

Изложим идею этого метода, используя целую рациональную функцию $z=f(x)$, степень которой не выше n.

[1] Для упрощения в дальнейшем записывается только мантисса.

Пусть имеется n пар значений x и z, где x — возраст, а z — коэффициенты смертности или вероятности умереть.

$$\begin{matrix} x & x_0 & x_1 & x_2\ldots x_n \\ z & z_0 & z_1 & z_2\ldots z_n \end{matrix}$$

Имеем при неравных значениях разности аргумента, т. е. возраста, $z=z_0+(x-x_0)f_1(x_1)+(x-x_0)(x-x_1)f_2(x_2)+\ldots+(x-x_0)(x-x_1)\ldots$ $\ldots(x-x_{n-1})f_n(x_n)$, где вспомогательные функции $f_i(x_i)$ имеют следующие значения:

$$f_1(x)=\frac{f_1(x)-f(x_0)}{x-x_0}; \quad f_2(x)=\frac{f_2(x)-f_1(x)}{x-x_1}$$

и т. д.

Если же разности возрастов равны друг другу и равны одному году, то интерполяционная формула примет вид:

$$z=f(x)=z_0+\Delta z_0(x-x_0)+\Delta^2 z_0\frac{(x-x_0)(x-x_1)}{2!}+$$
$$+\ldots+\Delta^n z_0\frac{(x-x_0)(x-x_1)\ldots(x-x_{n-1})}{n!},$$

где Δ, Δ^2, ..., Δ^n — разности 1, 2, ..., n-го порядка.

Применительно к выравниванию коэффициентов смертности старческих возрастов получим:

$$m_x=m_{79}+(x-79)\Delta m_{79}+\frac{(x-79)(x-80)}{2!}\Delta^2 m_{79}+$$
$$+\frac{(x-79)(x-80)(x-81)}{3!}\Delta^3 m_{79} \quad \text{и т. д.}$$

Эта формула была частично использована при обработке материалов переписи населения 1926 г., в частности для расчетов численности мужчин европейской части СССР без 36 городов[1].

Возьмем $m_{79}=0{,}10479$ и $m_{80}=0{,}11515$. Введение дополнительного условия требует, чтобы коэффициент смертности, скажем, в предельном возрасте был очень близок к единице. Возьмем, например, 125 лет и, учитывая, что число умерших бралось за два года, примыкающие к переписи, получим $m_{125}=2{,}00000$. Кроме этого, требуется, чтобы исчисленное число смертных случаев в возрасте 82—99 лет по выровненным коэффициентам равнялось 36 613 человек, что соответствовало бы фактическому числу умерших за 1926—1927 гг. Это можно выразить следующим образом: $\sum\limits_{82}^{99} m_x S_x=36\,613$ (где S_x — удвоенная численность населения по переписи).

Привлекая необходимые данные, можно, умножая все коэффициенты смертности на 10^5, записать:

[1] «Смертность и продолжительность жизни населения СССР. 1926—1927. Таблицы смертности», стр. XXIX.

$$\sum_{82}^{99}\Big[10\,479+(11\,515-10\,479)(x-79)\,S_x+\frac{\Delta^2 m_{79}}{1\cdot 2}\sum(x-79)\times$$
$$\times(x-80)\,S_x+\frac{\Delta^3 m_{79}}{1\cdot 2\cdot 3}\sum(x-79)(x-80)(x-81)\,S_x\Big]=36\,613\cdot 10^5.$$

Вводя соответствующие данные, получим:

$$200\,000=10\,479+1\,036\,(125-79)+(125-79)(125-80)\,\frac{\Delta^2 m_{79}}{1\cdot 2}+$$
$$+(125-79)(125-80)(125-81)\,\frac{\Delta^3 m_{79}}{1\cdot 2\cdot 3}.$$

Следовательно,

$$58\,135+2\,070\,\frac{\Delta^2 m_{79}}{1\cdot 2}+91\,080\,\frac{\Delta^3 m_{79}}{1\cdot 2\cdot 3}=200\,000.$$

Решая систему двух уравнений, находим:

$$\frac{\Delta^2 m_{79}}{1\cdot 2}=1{,}912 \quad \text{и} \quad \frac{\Delta^3 m_{79}}{1\cdot 2\cdot 3}=-15{,}594.$$

Теперь уже окончательно получаем:

$$m_x=10\,479+1\,036\,(x-79)+1{,}912\,(x-79)\,(x-80)-$$
$$-15{,}594\,(x-79)\,(x-80)\,(x-81).$$

Подставляя соответствующие значения, находим коэффициенты смертности: $m_{81}=0{,}12520$; $m_{82}=0{,}13504$; $m_{83}=0{,}14482$ и т. д.

ИНТЕРПОЛИРОВАНИЕ МЕТОДОМ НЬЮТОНА С РАЗДЕЛЕННЫМИ РАЗНОСТЯМИ

Интерполяция этим методом может быть использована для старческих возрастов (свыше 70 лет)

Обозначим:

x — возраст;

$\overline{y}_x$ — искомая величина для данного возраста;

x_1, x_2 и т. д. — возраста, для которых определяемые величины уже известны;

δ', δ'', δ''' — разделенные разности:

$$\delta_1'=\frac{y_2-y_1}{x_2-x_1};\quad \delta_2'=\frac{y_3-y_2}{x_3-x_2};$$
$$\delta_3'=\frac{y_4-y_3}{x_4-x_3};\quad \delta_1''=\frac{\delta_2'-\delta_1'}{x_3-x_1};$$
$$\delta_2''=\frac{\delta_3'-\delta_2'}{x_4-x_2};\quad \delta_1'''=\frac{\delta_2''-\delta_1''}{x_4-x_1}.$$

Интерполяционная формула Ньютона с разделенными разностями для случая, когда известны демографические показатели опорных возрастов, примет вид:

$$\overline{y}_x=y_{x_1}+(x-x_1)\,\{\delta_1'+(x-x_2)\,[\delta_1''+(x-x_3)\,\delta_1''']\}=$$
$$=y_{x_1}+(x-x_1)\,\delta_1'+(x-x_1)(x-x_2)\,\delta_1''+(x-x_1)(x-x_2)(x-x_3)\,\delta_1'''.$$

Если известны q_{85}; q_{90}, то для исчисления q_x промежуточных возрастов, т. е. q_{86}, q_{87}, q_{88}, q_{89}, привлекаем еще значения q_x, например q_{70} и q_{100}. В качестве примера используем данные о смертности мужчин по таблицам смертности СССР 1958—1959 гг.[1] (см. табл. 83).

Для нахождения значений $\overline{y}_x$ в табл. 83 с использованием разностей удобнее всего расчеты проводить по схеме, изложенной в табл. 84.

ИНТЕРПОЛИРОВАНИЕ МЕТОДОМ ЛАГРАНЖА

Для некоторых возрастных интервалов при выравнивании (интерполяции) можно привлечь формулу Лагранжа, применяемую при определении функции с неравными интервалами.

Пусть фактические уровни в пяти заданных возрастах (это наиболее типичный случай в демографии), отделенных друг от друга разным числом лет, имеют следующие значения: w_a, w_b, w_c, w_d, w_e, а выровненные значения в возрастном интервале от a до e обозначим $\overline{y}_x$ (где $a<x<e$). Тогда по формуле Лагранжа получаем:

$$\overline{y}_x=w_a\,\frac{(x-b)(x-c)(x-d)(x-e)}{(a-b)(a-c)(a-d)(a-e)}+w_b\,\frac{(x-a)(x-c)(x-d)(x-e)}{(b-a)(b-c)(b-d)(b-e)}+$$
$$+w_c\,\frac{(x-a)(x-b)(x-d)(x-e)}{(c-a)(c-b)(c-d)(c-e)}+w_d\,\frac{(x-a)(x-b)(x-c)(x-e)}{(d-a)(d-b)(d-c)(d-e)}+$$
$$+w_e\,\frac{(x-a)(x-b)(x-c)(x-d)}{(e-a)(e-b)(e-c)(e-d)}.$$

Для контроля хода расчетов рекомендуется следить за соблюдением требования о том, чтобы сумма всех коэффициентов при фактических уровнях при нахождении $\overline{y}_x$ для каждого значения x равнялась единице.

Рассмотрим применение этого метода на конкретных данных. Возьмем вероятности умереть по таблице смертности для населения городов Восточной Сибири по данным переписи населения 1926 г. и текущего учета смертности[2] в промежутке возрастов от 4 до 17 лет.

В качестве опорных точек возьмем возрасты 3, 4, 12, 17, 18 лет. Имеем:

$$q_{a=3}=0{,}01876;\quad q_{b=4}=0{,}01438;\quad q_{c=12}=0{,}00309;$$
$$q_{d=17}=0{,}00559;\quad q_{e=18}=0{,}00565.$$

[1] «Итоги Всесоюзной переписи населения 1959 года. СССР», стр. 265.

[2] К. И. Романов. Смертность и продолжительность жизни населения городов Восточной Сибири за 1926—1927 гг. Иркутск, издание Восточносибирского Мединститута, 1935.

Таблица 83

x	Вероятность умереть 100000 ($q_x = y$)	δ'	δ''	δ'''	Вероятность умереть, умноженная на 100000	
					по переписи	интерполированные
$x_1 = 70$	44,04	$\delta_1' = \frac{132,88-44,04}{85-70} = \frac{88,84}{15} = 5,92$				
$x_2 = 85$	132,88	$\delta_2' = \frac{176,68-132,88}{90-85} = \frac{43,80}{5} = 8,76$	$\delta_1'' = \frac{8,76-5,92}{90-70} = \frac{2,84}{20} = 0,142$	$\delta_1''' = \frac{0,111-0,142}{100-70} = \frac{-0,031}{30} = -0,001$	141,46 149,50 156,56 167,42	141,096 149,610 158,376 167,388
$x_3 = 90$	176,68	$\delta_3' = \frac{281,00-176,68}{100-90} = \frac{104,32}{10} = 10,432$	$\delta_2'' = \frac{10,432-8,76}{100-85} = \frac{1,672}{15} = 0,111$			
$x_4 = 100$	281,00					

Таблица 84

x	$x - x_1 = t_1$	$x - x_2 = t_2$	$x - x_3 = t_3$	$(x-x_1)(x-x_2) = t_1 \cdot t_2$	$(x-x_1)(x-x_2) \times (x-x_3) = t_1 \cdot t_2 \cdot t_3$	$t_1 \delta_1'$	$t_1 t_2 \delta_1''$	$t_1 t_2 t_3 \delta_1'''$	$\bar{y}_x$
86	86—70=16	86—85=1	86—90=—4	16·1=16	16·1·(—4)=—64	94,72	2,272	+0,664	141,096
87	87—70=17	87—85=2	87—90=—3	17·2=34	17·2·(—3)=—102	100,64	4,828	+0,102	149,610
88	88—70=18	88—85=3	88—90=—2	18·3=54	18·3·(—2)=—108	106,56	7,668	+0,108	158,376
89	89—70=19	89—85=4	89—90=—1	19·4=76	19·4·(—1)=—76	112,48	10,792	+0,076	167,388

Тогда для $\bar{q}_{x=6}$ получим:

$$\bar{q}_{x=6} = 0{,}01876\ \frac{(6-4)(6-12)(6-17)(6-18)}{(3-4)(3-12)(3-17)(3-18)} +$$
$$+ 0{,}01438\ \frac{(6-3)(6-12)(6-17)(6-18)}{(4-3)(4-12)(4-17)(4-18)} +$$
$$+ 0{,}00309\ \frac{(6-3)(6-4)(6-17)(6-18)}{(12-3)(12-4)(12-17)(12-18)} +$$
$$+ 0{,}00559\ \frac{(6-3)(6-4)(6-12)(6-18)}{(17-3)(17-4)(17-12)(17-18)} +$$
$$+ 0{,}00565\ \frac{(6-3)(6-4)(6-12)(6-17)}{(18-3)(18-4)(18-12)(18-17)} =$$
$$= -0{,}83809 \cdot 0{,}01876 + 1{,}63186 \cdot 0{,}01438 +$$
$$+ 0{,}36667 \cdot 0{,}00309 - 0{,}47472 \cdot 0{,}00559 + 0{,}31429 \cdot 0{,}00565.$$

Произведем проверку и получим сумму коэффициентов: $-0{,}83809 + 1{,}63186 + 0{,}36667 - 0{,}47472 + 0{,}31429 = 1$. Умножаем известные вероятности на найденные коэффициенты и получаем:

$$-0{,}83809 \cdot 0{,}01876 = -0{,}01572$$
$$1{,}63186 \cdot 0{,}01438 = 0{,}02347$$
$$0{,}36667 \cdot 0{,}00309 = 0{,}00113$$
$$-0{,}47472 \cdot 0{,}00559 = -0{,}00265$$
$$0{,}31429 \cdot 0{,}00565 = 0{,}00178$$

Итого 0,00801

Значит, $\bar{q}_{x=6} = 0{,}00801$.

Аналогично находят $\bar{q}_{x=7}$, $\bar{q}_{x=8}$, …, $\bar{q}_{x=17}$.

ИНТЕРПОЛИРОВАНИЕ МЕТОДОМ В. Я. БУНЯКОВСКОГО

В. Я. Буняковский применил особый способ выравнивания чисел умерших. При этом принимались в качестве основы два положения: после выравнивания последнее число умерших в каждом пятилетнем интервале должно быть равно средней арифметической возрастов, смежных с ним (т. е. предшествующего и последующего); разность (положительная) невыровненных чисел умерших в последний год пятилетнего интервала по сравнению с выровненным делится на две части пропорционально численности умерших двух смежных возрастов, которые и прибавляются к этим численностям.

Обозначая через a, b и c фактические числа умерших трех смежных возрастов, а через a', b' и c' — выровненные и полагая, что b и b' соответствуют последнему году пятилетнего интервала, имеем:

$$b' = \frac{a+b+c}{3}, \quad a' = \frac{2}{3}\, a\, \frac{a+b+c}{a+c}, \quad c' = \frac{2}{3}\, c\, \frac{a+b+c}{a+c}.$$

Этот способ применяется для устранения возрастной аккумуляции только в случае, когда она не выходит за пределы трех возрастов[1].

В качестве упрощенного примера произведем следующий расчет. Пусть для возрастов 49, 50 и 51 год численность умерших составляла 10, 20 и 6 тыс. человек. Имеем: $a = 10$, $b = 20$, $c = 6$, $a + b + c = 36$. Тогда

$$b' = \frac{a+b+c}{3} = \frac{36}{3} = 12,$$

$$a' = \frac{2}{3} \cdot 10 \cdot \frac{36}{16} = 15,$$

$$c' = \frac{2}{3} \cdot 6 \cdot \frac{36}{16} = 9.$$

Суммируя, получаем: $a' + b' + c' = 15 + 12 + 9 = 36$. Как видим, сумма выровненных чисел умерших равна сумме невыровненных.

Исходя из зависимости между возрастом и средней продолжительностью предстоящей жизни, а также из допущения о том, что убывание средней продолжительности жизни в интервале 10—68 лет с увеличением возраста происходит прямолинейно, с постоянной первой разностью, В. Я. Буняковский вывел формулу «эмпирического закона смертности»[2].

Если обозначить x возраст, $\overset{\circ}{e}_x$ — среднюю продолжительность предстоящей жизни в возрасте x лет, n — число возрастов, привлекаемых к анализу, то

$$\overline{\overset{\circ}{e}}_x = a_0 - a_1 x.$$

Параметры a_0 и a_1 В. Я. Буняковский определял методом наименьших квадратов по формулам:

$$a_0 = \frac{\Sigma (x^2) \Sigma \overset{\circ}{e}_x - \Sigma (x) \Sigma (x\overset{\circ}{e}_x)}{n \Sigma (x^2) - [\Sigma (x)]^2},$$

$$a_1 = \frac{\Sigma (x) \Sigma \overset{\circ}{e}_x - n \Sigma (x\overset{\circ}{e}_x)}{n \Sigma (x^2) - [\Sigma (x)]^2}.$$

Принимая во внимание связь между средней продолжительностью жизни и числом доживающих l_x, В. Я. Буняковский получил формулу, выражающую закон смертности:

$$l_x = \frac{(\lambda - x)(\lambda - x + 1)(\lambda - x + 2) \ldots (\lambda - x_0 - 1)}{(\mu - x)(\mu - x + 1)(\mu - x + 2) \ldots (\mu - x_0 - 1)} \, l_{x_0},$$

[1] В старческих возрастах наблюдается аккумуляция, выходящая за пределы более 10 лет.

[2] В. Я. Буняковский. Об одном эмпирическом выражении закона смертности. Записки Императорской Академии Наук, т. 15, кн. II. СПб., 1869.

где

$$\lambda = \frac{a_0 + a_1 - \frac{1}{2}}{a_1}, \quad \mu = \frac{a_0 + \frac{1}{2}}{a_1},$$

$x_0 = 10$, l_{x_0} — число доживающих до 10 лет.

Извлекая из таблиц смертности Депарсье семь значений средней продолжительности предстоящей жизни для возрастов 10, 20, 30, 40, 50, 60, 70 соответственно 46 лет 10 месяцев, 40 лет 3 месяца, 34 года 1 месяц, 27 лет 6 месяцев, 20 лет 5 месяцев, 14 лет 3 месяца, 8 лет 8 месяцев и привлекая приведенные выше формулы, В. Я. Буняковский получил: $n = 7$; $\Sigma (x) = 280$; $\Sigma (x^2) = 14\,000$; $[\Sigma (x)]^2 = 78\,400$; $\Sigma (e_x^0) = 192$ года; $\Sigma (xe_x^0) = 5878$ лет 4 месяца; $a_0 = 53$ года 2 месяца; $a_1 = 7$ месяцев 22 дня. Тогда
$\overline{e}_x^0 = 53$ года 2 месяца — 7 месяцев 22 дня $\cdot x$.

Далее были найдены следующие значения: $\lambda = 82{,}91$ года; $\mu = 83{,}46$ года; $l_{x_0} = l_{10} = 880$.

Окончательно

$$l_x = \frac{(82{,}91 - x)(82{,}91 - x + 1) \ldots (82{,}91 - 11)}{(83{,}46 - x)(83{,}46 - x + 1) \ldots (83{,}46 - 11)} \cdot 880.$$

Все полученные выровненные числа остающихся в живых и взятые из таблицы Депарсье в различных возрастах В. Я. Буняковский уложил в таблицу.

Полученные результаты В. Я. Буняковский считал «очень удовлетворительными». Произведенные нами аналогичные расчеты по таблице смертности в СССР 1958—1959 гг. для обоих полов по возрастам, указанным В. Я. Буняковским, дали также удовлетворительные результаты.

Таблица 85

Возраст в годах	Число остающихся в живых по Депарсье	Числа остающихся в живых по формуле В. Я. Буняковского	Отклонения в процентах
10	880	880	—
15	848	846	0,24
20	814	811	0,36
25	774	771	0,39
30	734	737	0,41
35	694	693	0,14
40	657	657	—
45	622	607	2,41
50	581	567	2,41
55	526	508	3,42
60	463	465	0,43
65	395	386	2,28
70	310	337	8,71

ИНТЕРПОЛИРОВАНИЕ МЕТОДОМ ИСПОЛЬЗОВАНИЯ КОЭФФИЦИЕНТОВ РОСТА

Можно предложить очень простой метод выравнивания, основанный на использовании коэффициентов роста.

Выравнивание при равных интервалах

Сохраняя прежнюю символику, имеем:

фактическую численность населения в пятилетнем интервале, предшествующем выравниваемому, с возрастами x_{-7}, x_{-6}, x_{-5}, x_{-4}, x_{-3}, равную Σ_{-1};

фактическую численность населения в пятилетнем интервале, подлежащем выравниванию, с возрастами x_{-2}, x_{-1}, x_0, x_1, x_2, равную Σ_0;

фактическую численность населения в пятилетнем интервале, последующем за выравниваемым, с возрастами x_3, x_4, x_5, x_6, x_7, равную Σ_{+1}.

Таким образом, для интерполяции численностей внутри какого-нибудь пятилетнего интервала привлекаются численности двух соседних пятилетних интервалов.

Допускаем, что численности населения в каждом одногодичном возрасте во всех трех пятилетних интервалах представляют 1/5 всей фактической численности. Таким образом, численность центровых возрастов в каждом интервале равна:

для возраста x_{-5} лет $\frac{\Sigma_{-1}}{5}$,

для возраста x_0 лет $\frac{\Sigma_0}{5}$,

для возраста x_5 лет $\frac{\Sigma_{+1}}{5}$.

Допустим далее, что численность населения убывает (или возрастает) равномерно от $\frac{\Sigma_{-1}}{5}$ до $\frac{\Sigma_0}{5}$ с одинаковым коэффициентом роста, равным K_1, а от $\frac{\Sigma_0}{5}$ до $\frac{\Sigma_{+1}}{5}$ с одинаковым коэффициентом роста, равным K_2 (см. табл. 86).

Тогда получим сумму одногодичных численностей населения в промежутке от 1-го до 2-го центра:

$$\frac{\Sigma_{-1}}{5}+\frac{\Sigma_{-1}}{5}K_1+\frac{\Sigma_{-1}}{5}K_1^2+\frac{\Sigma_{-1}}{5}K_1^3+\frac{\Sigma_{-1}}{5}K_1^4;$$

в промежутке от 2-го и 3-го центра:

$$\frac{\Sigma_0}{5}+\frac{\Sigma_0}{5}K_2+\frac{\Sigma_0}{5}K_2^2+\frac{\Sigma_0}{5}K_2^3+\frac{\Sigma_0}{5}K_2^4,$$

$$\frac{\Sigma_{-1}}{5}K_1^5=\frac{\Sigma_0}{5} \text{ и, следовательно, } K_1=\sqrt[5]{\frac{\Sigma_0}{\Sigma_{-1}}},$$

$$\frac{\Sigma_0}{5}K_2^5=\frac{\Sigma_{+1}}{5} \text{ и, следовательно, } K_2=\sqrt[5]{\frac{\Sigma_{+1}}{\Sigma_0}}.$$

Таблица 86

Условное обозначение возраста	Центральный возраст в пятилетнем интервале	Численность населения в центральном возрасте	Численность населения во всех возрастах в интервале от первого до третьего центра
x_{-7}			$\frac{\Sigma_{-1}}{5}$
x_{-6}			
x_{-5}	x_{-5}	$\frac{\Sigma_{-1}}{5}$	$\frac{\Sigma_{-1}}{5}K_1$
x_{-4}			
x_{-3}			$\frac{\Sigma_{-1}}{5}K_1^2$
Итого	—	Σ_{-1}	
x_{-2}			$\frac{\Sigma_{-1}}{5}K_1^3$
x_{-1}			$\frac{\Sigma_{-1}}{5}K_1^4$
x_0	x_0	$\frac{\Sigma_0}{5}$	$\frac{\Sigma_0}{5}=\frac{\Sigma_{-1}}{5}K_1^5$
x_1			$\frac{\Sigma_0}{5}K_2$
x_2			$\frac{\Sigma_0}{5}K_2^2$
Итого	—	Σ_0	
x_3			
x_4			$\frac{\Sigma_0}{5}K_2^3$
x_5	x_5	$\frac{\Sigma_{+1}}{5}$	$\frac{\Sigma_0}{5}K_2^4$
x_6			
x_7			$\frac{\Sigma_{+1}}{5}=\frac{\Sigma_0}{5}K_2^5$
Итого	—	Σ_{+1}	
Всего	—	$\Sigma_{-1}+\Sigma_0+\Sigma_{+1}$	

Теперь выразим общую численность населения по одногодичным возрастам в интерполируемом пятилетнем интервале следующим образом:

Возраст, x	x_{-2}	x_{-1}	x_0	x_{+1}	x_{+2}	Сумма интерполируемых численностей
Численность населения по возрастам, y . . .	$\frac{\Sigma_{-1}}{5}K_1^3$	$\frac{\Sigma_{-1}}{5}K_1^4$	$\frac{\Sigma_0}{5}$	$\frac{\Sigma_0}{5}K_2$	$\frac{\Sigma_0}{5}K_2^2$	Σy

Найденная сумма выровненных численностей может не совпасть с фактической численностью интерполируемого пятилетия, т. е. с Σ_0. Если фактические численности населения по пятилетиям считать соответствующими действительности и в качестве основного требования выдвинуть необходимость равенства выровненных численностей сумме фактических, то для совпадения нужно умножить все полученные выше численности на коэффициент поправки

$$z = \frac{\Sigma y}{\Sigma_0}.$$

Тогда получим окончательно:

$$\Sigma_0 = z\left(\frac{\Sigma_{-1}}{5}K_1^3 + \frac{\Sigma_{-1}}{5}K_1^4 + \frac{\Sigma_0}{5} + \frac{\Sigma_0}{5}K_2 + \frac{\Sigma_0}{5}K_2^2\right).$$

Значит, интерполируемые численности населения для каждого года срединного пятилетнего интервала будут иметь следующие значения:

Годы	Интерполируемые численности населения
1	$\frac{\Sigma_{-1}}{5}K_1^3 z$
2	$\frac{\Sigma_{-1}}{5}K_1^4 z$
3	$\frac{\Sigma_0}{5} z$
4	$\frac{\Sigma_0}{5}K_2 z$
5	$\frac{\Sigma_0}{5}K_2^2 z$
Итого . .	Σ_0

В качестве примера используем данные, рассмотренные нами при исследовании метода С. А. Новосельского.

Таблица 87

Исходные данные для интерполяции

Пятилетние возрастные интервалы	40—44	45—49	50—54
Численности мужчин (в тыс.) . . . . . Σ_i	2306,7 Σ_{-1}	1990,5 Σ_0	1653,1 Σ_{+1}

Следовательно,

$$K_1 = \sqrt[5]{\frac{1990,5}{2306,7}} \approx 0,9709, \quad K_2 = \sqrt[5]{\frac{1653,1}{1990,5}} \approx 0,9637.$$

Найдем Σy.

Возраст (в годах)	Интерполированные численности мужчин (в тыс.) без учета поправочного коэффициента
45	$\frac{2306,7}{5}\cdot 0,9709^3 \approx 422,20$
46	$\frac{2306,7}{5}\cdot 0,9709^4 \approx 409,91$
47	$\frac{1990,5}{5} \approx 398,12$
48	$\frac{1990,5}{5}\cdot 0,9637 \approx 383,67$
49	$\frac{1990,5}{5}\cdot 0,9637^2 \approx 369,74$
Итого . .	$\Sigma y = 1983,64$

Определяем поправочный коэффициент:

$$z = \frac{1990,50}{1983,64} \approx 1,00351.$$

Теперь найденные интерполированные численности умножаем на z и получаем окончательно:

Возраст (в годах)	Численности мужчин (в тыс.), найденные методом интерполяции
45	$422,20\cdot 1,00351 = 423,7$
46	$409,91\cdot 1,00351 = 411,3$
47	$398,12\cdot 1,00351 = 399,5$
48	$383,67\cdot 1,00351 = 385,0$
49	$369,74\cdot 1,00351 = 371,0$
Итого . .	1990,5

Как видно, полученные довольно элементарными приемами результаты мало отличаются от результатов, полученных по методу С. А. Новосельского. Так, численность населения в возрасте 46 лет, равная 411,3 совпала полностью, а для возрастов 45, 47, 48 и 49 — численности довольно близки.

Выравнивание при неравных интервалах

Использование данного метода может производиться и при неравных интервалах. Рассмотрим два случая.

Интерполируемый интервал больше привлекаемых с двух сторон

Пусть интерполируемый интервал равен 9 годам, а привлекаемые — по 5 лет. Обозначим возраста верхнего привлекаемого 5-летнего интервала $x_{-9}, x_{-8}, x_{-7}, x_{-6}$ и x_{-5}, возраста интерполируемого 9-летнего интервала $x_{-4}, x_{-3}, x_{-2}, x_{-1}, x_0, x_1, x_2, x_3, x_4$ и возраста нижнего привлекаемого 5-летнего интервала x_5, x_6, x_7, x_8, x_9. Значит, численности центровых возрастов в каждом из трех интервалов равны:

$$\text{для возраста } x_{-7} \text{ лет} - \frac{\Sigma_{-1}}{5}, \text{ для возраста } x_0 \text{ лет} - \frac{\Sigma_0}{9},$$

$$\text{для возраста } x_7 \text{ лет} - \frac{\Sigma_{+1}}{5}.$$

Общую картину изменений численностей населения можно проследить по табл. 88.

Общая численность в интерполируемом 9-летнем интервале выразится величинами, заключенными между $\frac{\Sigma_{-1}}{5} K_1^3$ (для возраста x_{-4}) и $\frac{\Sigma_0}{9} K_2^4$ (для возраста x_4).

Если найденная сумма выровненных численностей Σy не совпадает с фактической численностью интерполируемого интервала, т. е. Σ_0, то так же, как и в предыдущем случае, вводится поправка $z = \frac{\Sigma y}{\Sigma_0}$. Окончательно интерполированные уровни примут следующие значения:

$$\frac{\Sigma_{-1}}{5} K_1^3 z; \quad \frac{\Sigma_{-1}}{5} K_1^4 z; \quad \frac{\Sigma_{-1}}{5} K_1^5 z; \quad \frac{\Sigma_{-1}}{5} K_1^6 z; \quad \frac{\Sigma_0}{9} z; \quad \frac{\Sigma_0}{9} K_2 z;$$

$$\frac{\Sigma_0}{9} K_2^2 z; \quad \frac{\Sigma_0}{9} K_2^3 z; \quad \frac{\Sigma_0}{9} K_2^4 z.$$

Интерполируемый интервал меньше привлекаемых двух, не равных друг другу

Пусть интерполируемый интервал равен 5 годам, интервал, предшествующий интерполируемому, равен 7, а последующий за ним 3 годам. В результате произведенных действий получим:

$$K_1 = \sqrt[6]{\frac{7}{5} \frac{\Sigma_0}{\Sigma_{-1}}}, \quad K_2 = \sqrt[4]{\frac{5}{3} \frac{\Sigma_{+1}}{\Sigma_0}}.$$

В интерполируемом интервале 5 уровней вычисляются по формулам:

$$\frac{\Sigma_{-1}}{7} K_1^4 z; \quad \frac{\Sigma_{-1}}{7} K_1^5 z; \quad \frac{\Sigma_0}{5} z; \quad \frac{\Sigma_0}{5} K_2 z; \quad \frac{\Sigma_0}{5} K_2^2 z.$$

Теперь можно вывести систему формул в общем виде. Обозначим число уровней в интервале, предшествующем интерполируемому, a, число уровней в интерполируемом интервале b, число уровней в интервале, последующем за интерполируемым, c. Формулы коэффициентов роста, используемых для интерполяции, примут вид

$$K_1 = \sqrt[\frac{a+b}{2}]{\frac{a}{b} \frac{\Sigma_0}{\Sigma_{-1}}}, \quad K_2 = \sqrt[\frac{b+c}{2}]{\frac{b}{c} \frac{\Sigma_{+1}}{\Sigma_0}}.$$

Таблица 88

Условное обозначение возраста	Центральный возраст в интервале	Численность населения в центральном возрасте	Численность населения во всех возрастах в интервале от первого до третьего центра
x_{-9}			
x_{-8}			
x_{-7}	x_{-7}	$\frac{\Sigma_{-1}}{5}$	$\frac{\Sigma_{-1}}{5}$
x_{-6}			$\frac{\Sigma_{-1}}{5} K_1$
x_{-5}			$\frac{\Sigma_{-1}}{5} K_1^2$
Итого	—	Σ_{-1}	
x_{-4}			$\frac{\Sigma_{-1}}{5} K_1^3$
x_{-3}			$\frac{\Sigma_{-1}}{5} K_1^4$
x_{-2}			$\frac{\Sigma_{-1}}{5} K_1^5$
x_{-1}			$\frac{\Sigma_{-1}}{5} K_1^6$
x_0	x_0	$\frac{\Sigma_0}{9}$	$\frac{\Sigma_0}{9} = \frac{\Sigma_{-1}}{5} K_1^7$
x_1			$\frac{\Sigma_0}{9} K_2$
x_2			$\frac{\Sigma_0}{9} K_2^2$
x_3			$\frac{\Sigma_0}{9} K_2^3$
x_4			$\frac{\Sigma_0}{9} K_2^4$
Итого	—	Σ_0	
x_5			$\frac{\Sigma_0}{9} K_2^5$
x_6			$\frac{\Sigma_0}{9} K_2^6$
x_7	x_7	$\frac{\Sigma_{+1}}{5}$	$\frac{\Sigma_{+1}}{5} = \frac{\Sigma_0}{9} K_2^7$
x_8			
x_9			
Итого	—	Σ_{+1}	

Сами интерполируемые уровни до привлечения поправки z будут иметь вид: $\frac{\Sigma_{-1}}{a} K_1^{\frac{a+1}{2}}$, $\frac{\Sigma_{-1}}{a} K_1^{\frac{a+1}{2}+1}$, ..., $\frac{\Sigma_{-1}}{a} K_1^{\frac{a+b-2}{2}}$; $\frac{\Sigma_0}{b}$; $\frac{\Sigma_0}{b} K_2$,

$$\frac{\Sigma_0}{b} K_2^2, \ldots, \frac{\Sigma_0}{b} K_2^{\frac{b-1}{2}}.$$

С учетом поправки все найденные уровни должны быть умножены на z. Предлагаемый метод интерполяции данных о численности населения легко распространяется на числа умерших, вероятности умереть и т. д.

ДИСПЕРСИОННЫЙ МЕТОД ВЫРАВНИВАНИЯ

Дисперсионный метод выравнивания данных о смертности был разработан Б. С. Ястремским.

В 1910 г. на III съезде представителей обществ страхования жизни при бюро съездов в России была создана Комиссия актуариев, секретарем которой был избран Б. С. Ястремский. С 1912 г. комиссия приступила к обработке материалов для построения таблиц смертности для застрахованных. В 1913 г. Б. С. Ястремский в ряде работ изложил данный метод выравнивания таблиц смертности[1]. Рассмотрим теоретические основы этого метода и целесообразность его практического использования.

Опыт показывает, что если производить возрастную группировку, используя не одногодичные, а другие — бо́льшие или меньшие интервалы, то при любой системе группировок полученные коэффициенты смертности (вероятности умереть) сохраняют ряд закономерностей.

1. Чем меньше возрастной интервал, тем больше колеблются коэффициенты смертности и очертания описывающей их кривой носят малоопределенный характер. С увеличением возрастного интервала колебания коэффициентов смертности уменьшаются и приближаются к некоторой плавной линии.

2. Сравнивая коэффициенты смертности, полученные при различных системах группировок, с коэффициентами смертности, расположенными по плавной линии, увидим, что полученные ранжированные разности $K_{\text{см (факт)}} - K_{\text{см (плавн)}}$ (или $w_x - \bar{y}_x$) уложатся в закон ошибок Гаусса — Лапласа, т. е. в закон нормального распределения[2]. В этом случае w_x — это фактические вероятности умереть, а $\bar{y}_x$ — выровненные. Положительные и отрицательные отклонения расположатся симметрично, и их суммы будут приблизительно равны друг другу по абсолютной величине; общая их сумма не будет значительно отличаться от нуля. Большие отклонения будут встречаться реже, чем маленькие.

[1] См., например, Б. Ястремский. Дисперсионный метод изучения эволюции смертности.

[2] И. Г. Венецкий., Г. С. Кильдишев. Основы математической статистики, стр. 164—177.

3. Распределение указанных разностей будет укладываться в закон нормального распределения тем лучше, чем однороднее и богаче статистический материал.

Указанные обстоятельства позволяют производить выравнивание дисперсионным методом. При этом плавная линия, воспроизводящая фактические уровни коэффициентов смертности, должна как можно лучше улавливать нормальное распределение разностей $w_{x(\text{факт})} - \bar{y}_{x\,(\text{плавн})}$.

Теоретически $\bar{y}_{x\,(\text{плавн})} = f(x, a_0, a_1, a_2, \ldots, a_n)$. Число параметров a_0, a_1, a_2, ..., a_n зависит от того, какую кривую мы примем в качестве функции, достаточно точно улавливающей изменение смертности. Если в общем виде принята функция $\bar{y}_{x(\text{плавн})} = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$, а исчисление параметров будет производиться методом наименьших квадратов, то по мере увеличения параметров число уравнений в системе нормальных уравнений будет увеличиваться, а кривая $\bar{y}_{x(\text{плавн})}$ будет усложняться. При зависимости по прямой линии $\bar{y}_{x(\text{плавн})} = a_0 + a_1 x$ система нормальных уравнений примет вид:

$$\begin{cases} \Sigma (a_0 + a_1 x) = a_0 n + a_1 \Sigma x = \Sigma \bar{y}_{x\,(\text{плавн})} = \Sigma w_{x(\text{факт})}, \\ \Sigma (a_0 + a_1 x) x = a_0 \Sigma x + a_1 \Sigma x^2 = \Sigma \bar{y}_{x(\text{плавн})}\, x = \Sigma w_{x(\text{факт})}\, x. \end{cases}$$

При параболе 2-го порядка $\bar{y}_{x(\text{плавн})} = a_0 + a_1 x + a_2 x^2$ система нормальных уравнений примет вид:

$$\begin{cases} \Sigma (a_0 + a_1 x + a_2 x^2) = a_0 n + a_1 \Sigma x + a_2 \Sigma x^2 = \\ \qquad = \Sigma \bar{y}_{x(\text{плавн})} = \Sigma w_{x(\text{факт})}, \\ \Sigma (a_0 + a_1 x + a_2 x^2)\, x = a_0 \Sigma x + a_1 \Sigma x^2 + a_2 \Sigma x^3 = \\ \qquad = \Sigma \bar{y}_{x(\text{плавн})}\, x = \Sigma w_{x(\text{факт})}\, x, \\ \Sigma (a_0 + a_1 x + a_2 x^2)\, x^2 = a_0 \Sigma x^2 + a_1 \Sigma x^3 + a_2 \Sigma x^4 = \\ \qquad = \Sigma \bar{y}_{x(\text{плавн})}\, x^2 = \Sigma w_{x(\text{факт})}\, x^2 \end{cases}$$

и т. д.

Рассуждая теоретически, будем увеличивать число параметров до тех пор, пока не найдем такие разности $w_{x\text{факт}} - \bar{y}_{x(\text{плавн})}$, которые, на наш взгляд, достаточно близко укладываются в закон нормального распределения.

Практическое использование дисперсионного метода намного сложнее, чем указывает Б. С. Ястремский, ввиду того, что при различных возрастных группировках получаются группы с различными статистическими весами q, которыми при расчетах пренебрегать нельзя. Если обозначить S численность возрастной группы, то $q = \frac{S}{y(y-1)}$. Но ведь это и есть показатель вероятности, который является целью наших поисков. Поэтому q определяется приблизительно, методом Карупа. С учетом q система нормальных уравнений примет измененный вид, например для параболы 2-го порядка:

$$\begin{cases} \Sigma q(a_0 + a_1 x + a_2 x^2) = \Sigma \overline{y}_{x(\text{плавн})}\, q = \Sigma w_{x(\text{факт})}\, q, \\ \Sigma q(a_0 x + a_1 x^2 + a_2 x^3) = \Sigma \overline{y}_{x(\text{плавн})}\, qx = \Sigma w_{x(\text{факт})}\, qx, \\ \Sigma q(a_0 x^2 + a_1 x^3 + a_2 x^4) = \Sigma \overline{y}_{x(\text{плавн})}\, qx^2 = \Sigma w_{x(\text{факт})}\, qx^2. \end{cases}$$

Обозначая $\Sigma qx^m = A_m$ и $\Sigma q w_{x(\text{факт})}\, x^m = B_m$, получим:

$$\begin{cases} a_0 A_0 + a_1 A_1 + a_2 A_2 = B_0, \\ a_0 A_1 + a_1 A_2 + a_2 A_3 = B_1, \\ a_0 A_2 + a_1 A_3 + a_2 A_4 = B_2. \end{cases}$$

Параболы более высоких степеней привлекаются до тех пор; пока мы не достигнем достаточного согласия между фактическими и теоретическими отклонениями.

Оценка близости этих отклонений может производиться путем приведения их к весу, равному единице, и вычисления λ:

$$\lambda = \Sigma q \left[\overline{y}_{x(\text{плавн})} - w_{x(\text{факт})}\right].$$

При этом можно использовать один из трех способов оценки.

1. Подсчитывают $\lambda_{(\text{факт})}$, заключенные в пределах λ_1 и λ_2, и сравнивают с теоретическим значением, подсчитанным по формуле

$$P(\lambda_1, \lambda_2) = \frac{z}{\sqrt{2\pi}} \int\limits_{\lambda_1}^{\lambda_2} e^{-\frac{\lambda^2}{2}}\, d\lambda,$$

где z — число наблюдаемых коэффициентов смертности (или вероятностей).

2. Располагают λ в ранжированном порядке и сравнивают с огивой Гальтона, т. е. графическим изображением зависимости, обратной той, которая выражается уравнением

$$P(\lambda) = \frac{Z}{\sqrt{2\pi}} \int\limits_0^{\lambda} e^{-\frac{\lambda^2}{2}}\, d\lambda,$$

где $P(\lambda)$ — количество отклонений, не выходящих за пределы $0 - \lambda$.

3. Находят отношение фактических и теоретических средних квадратов отклонений и получают коэффициенты дисперсии Q^2.

Подсчет фактического среднего квадрата отклонений производится по формуле $\frac{\Sigma\lambda^2}{z}$. Подсчет теоретического значения среднего квадрата случайных отклонений, приведенного к весу, равному единице, производится по формуле

$$\int\limits_{-\infty}^{+\infty} \frac{z}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} \lambda^2\, d\lambda = 1.$$

Следовательно, $Q^2 = \frac{\Sigma\lambda^2}{z}$. Значение $\Sigma\lambda^2$ может быть найдено и без вычисления всех последовательных приближений следующим образом: если

$$\overline{y}_{x(\text{плавн})} = a_0' + a_1' x;$$
$$\overline{y}_{x(\text{плавн})} = a_0'' + a_1'' x + a_2'' x^2,$$

то

$$\Sigma\lambda^2 = \Sigma q \left[w_{x(\text{факт})} - \overline{y}_{x(\text{плавн})}\right]^2 =$$
$$= \Sigma q w_{x(\text{факт})} - (a_0 B_0 + a_1 B_1 + a_2 B_2 + \ldots).$$

При возрастании порядка используемой для выравнивания параболы значение коэффициента дисперсии уменьшается.

При $\overline{y}_{x(\text{плавн})} = a_0$ коэффициент дисперсии Q_0^2;
при $\overline{y}_{x(\text{плавн})} = a_0 + a_1 x$ коэффициент дисперсии Q_1^2;
при $\overline{y}_{x(\text{плавн})} = a_0 + a_1 x + a_2 x^2$ коэффициент дисперсии Q_2^2 и т. д.

Как было доказано Б. С. Ястремским, значения Q_i^2 располагаются следующим образом: $Q_0^2 > Q_1^2 > Q_2^2 > Q_3^2 > \ldots, > Q_n^2$ и справедливы следующие приближенные равенства:

$$Q_0^2 - Q_1^2 = \frac{1}{12} q (\Delta y)^2; \quad Q_1^2 - Q_2^2 = \frac{1}{45} q (\Delta^2 y)^2, \quad Q_2^2 - Q_3^2 = \frac{1}{138} q (\Delta^3 y)^2.$$

Если взвешенные отклонения λ укладываются в нормальное распределение, то $Q^2 \to 1$.

Таким образом, в качестве достаточно сложной параболы — плавной и хорошо воспроизводящей фактические данные о повозрастной смертности — можно считать такую параболу, при которой коэффициент дисперсии Q^2 достаточно близок к единице. Степень близости Q^2 к единице может быть измерена величиной $\sqrt{\frac{2}{z}}$.

Если $Q^2 - 1 > \sqrt{\frac{2}{z}}$, то следует признать, что выравнивание не дает еще нужной плавности. Если же использована очень сложная кривая, то $Q^2 - 1 < \sqrt{\frac{2}{z}}$.

Дисперсионный метод выравнивания был использован при обработке данных 9 страховых обществ в России. Выравнивались показатели смертности застрахованных. Для оценки качества выравнивания мы используем общепринятый критерий — соблюдение равенства сумм фактических и выровненных данных.

Предположим, что кривая $\overline{y}_{x(\text{плавн})}$ удовлетворяет условию основного равенства сумм, когда

$$\Sigma q w_{x(\text{факт})}\, \overline{y}_{x(\text{плавн})} = \Sigma q \overline{y}_{x\ (\text{плавн})}^2.$$

Так как $q = \frac{S}{\bar{y}_{x(\text{плавн})}\,[1-\bar{y}_{x(\text{плавн})}]}$, делаем подстановку и после соответствующих сокращений приходим к равенству $\Sigma S w_{x(\text{факт})} = \Sigma S \bar{y}_{x(\text{плавн})}$, а это означает равенство действительного числа смертных случаев выровненным данным.

Таким образом, мы видим, что дисперсионный метод удовлетворяет требованию необходимой плавности и, как нам удалось показать, выровненные данные довольно хорошо воспроизводят порядок изменения фактических.

ВЫРАВНИВАНИЕ МЕТОДОМ А. Я. БОЯРСКОГО[1]

Условия аккумуляции различных возрастов могут быть определены при сопоставлении результатов двух переписей. Имея, например, материалы переписи населения 1926 и 1939 гг.[2], можно сравнить число людей, которым по переписи населения 1926 г. было 50 лет, с числом людей, которым по переписи населения 1939 г. было 50 + + (1939—1926), т. е. от 62 до 63 лет. А. Я. Боярский исходит из предположения, что дожитие поколений за промежуток между переписями меняется в зависимости от возраста плавно. Единственным источником нарушений этой плавности может быть только аккумуляция. Свой метод он иллюстрировал на примерах переписей 1920 и 1926 гг.

Пусть по переписи населения 1920 и 1926 гг. фактические численности составляли y'_x и y_x, а выровненные: S'_x и S_x. Коэффициент аккумуляции возраста x лет составлял для переписи населения 1920 г. $1+\gamma'_x$, а для 1926 г. $1+\gamma_x$[3].

Откуда $\qquad S_x = y_x(1+\gamma_x); \quad S'_x = y'_x(1+\gamma'_x).$

Надо полагать, что отношение показателей аккумуляции есть величина постоянная для всех возрастов:

$$\frac{\gamma'_x}{\gamma_x} = c.$$

Коэффициент дожития K_x можно получить по формуле

$$K_x = \frac{y_x}{y'_{x-6}} = \frac{S_x}{S'_{x-6}} \cdot \frac{1+\gamma'_{x-6}}{1+\gamma_x}.$$

Получим коэффициент дожития всех возрастов выравниванием по 10-летней скользящей средней геометрической. Найденные выровненные значения обозначим $\bar{K}_x$. При этом практически находим $\log \bar{K}_x$ как скользящую среднюю арифметическую $\log \bar{K}_x$.

[1] Этот метод изложен А. Я. Боярским в докторской диссертации на тему «Проект методологии таблиц смертности и таблиц плодовитости населения СССР (1938—1939 гг.)» (диссертация не опубликована).

[2] В качестве примера взяты эти две переписи населения, а не переписи 1939 и 1959 гг., так как в период между ними не было такого фактора, как война.

[3] Для неаккумулирующих возрастов $1+\gamma_x<1$.

Примем $\bar{K}_x = \frac{S_x}{S'_{x-6}}$ и следовательно,

$$\frac{\bar{K}_x}{K_x} = Q_x = \frac{S_x}{S'_{x-6}} : \frac{y_x}{y'_{x-6}} = \frac{1+\gamma_x}{1+\gamma'_x}.$$

Показатель Q_x — рафинированный продукт возрастной аккумуляции. Произведем алгебраические преобразования и получим рекурентную формулу

$$\gamma_x = Q(1+c\gamma_{x-6}) - 1.$$

Для вычисления c, т. е. коэффициента изменения силы возрастной аккумуляции, введем дополнительное условие. Обозначим через Σ сумму численностей по всем возрастам интервала 13—82 года, а $\underset{\bullet}{\Sigma}$ сумму по входящим в него возрастам, кратным пяти. Получаем:

$$5\underset{\bullet}{\Sigma} S_x = \Sigma S_x, \text{ или } 5\underset{\bullet}{\Sigma} y_x(1+\gamma_x) = \Sigma y_x(1+\gamma_x),$$

$$\text{или } 5\underset{\bullet}{\Sigma} y_x Q_x(1+c\gamma_{x-6}) = \Sigma y_x Q_x(1+c\gamma_{x-6}).$$

Из последнего уравнения находим c:

$$c = \frac{\Sigma y_x Q_x - 5\underset{\bullet}{\Sigma} y_x Q_x}{5\underset{\bullet}{\Sigma} y_x Q_x \gamma_{x-6} - \Sigma y_x Q_x \gamma_{x-6}}.$$

Кроме того, нужно соблюдать дополнительное условие, т. е.

$$\dot{\gamma}_x = Q(1+c\gamma_{x-6}) - 1.$$

Далее сначала c принимаем равным единице, определяем все γ_x и подставляем их в написанные выше суммы, находим из дополнительного условия исправленное c.

Производим расчет снова, находим все γ_x и вновь исправляем c. И так до тех пор, пока не будет найдено такое c, которое удовлетворяет дополнительному условию и условию равенства единице коэффициента аккумуляции для исправленного с их помощью возрастного распределения.

Расчеты по определению c лучше всего начинать с подстановки такого значения c, которое вытекает из фактических общих коэффициентов аккумуляции. Пользуясь этой формулой, А. Я. Боярский определил c. Однако предположение о постоянстве c не совсем правильно. Расчеты показали, что, например, для интервала 13—82 года c равно 1,3014, а для интервала 13—62 c равно 1,400. Надо полагать, что этот прием выравнивания для определенных возрастных контингентов может быть применен в комбинациях с другими.

КОМБИНИРОВАНИЕ РАЗЛИЧНЫХ МЕТОДОВ ВЫРАВНИВАНИЯ

Комбинированный метод выравнивания состоит из последовательного применения трех операций. Первая состоит в выравнивании чисел живущих и умерших методом взвешенной или невзвешенной скользящей средней.

Вторая операция — выравнивание по одному из методов Гомперца — Макегама.

Третья операция — сопоставление результатов второго выравнивания с первым и объединение отклонений в пределах одного знака на протяжении отдельных промежутков возраста.

Совокупности этих отклонений в пределах одного знака подвергаются последнему выравниванию. Выровненные отклонения алгебраически суммируются с результатом второй операции.

При этом методе затрата вычисленных работ весьма велика.

Расчеты, произведенные С. А. Новосельским, по данным переписи населения 1926 г. по Украинской ССР привели к результатам, близким к результатам, полученным общим способом, предложенным В. В. Паевским.

Некоторые исследователи полагают, что в затруднительных положениях, требующих конкретных решений демографических задач, можно ориентироваться на здравый смысл и интуицию. Это заблуждение. Одного здравого смысла исследователя и его критического чутья для решения различных задач демографии недостаточно. Они способны заострить внимание исследователя на демографической проблеме, заставить его усомниться в правильности ее разрешения, поставить ряд вопросов об отличительных чертах той или иной демографической ситуации. Решение же частных и общих вопросов демографии возможно лишь на основе математической обработки массовых статистических данных.

В настоящее время вопрос об использовании математических методов в области социально-экономических наук и, следовательно, в области демографии не является дискуссионным. Однако в мнениях о том, какие следует при этом привлекать методы, нет полного единодушия. Особенно это касается соотношения математических методов с другими.

Конечно, не стоит привлекать математические и математико-статистические методы только ради их самих, без надлежащей теоретической базы. Наоборот, правильное использование этих методов не отвергает, а предполагает их ориентировку на теоретические положения, введенные марксистско-ленинской материалистической философией и социально-экономическими науками.

ОГЛАВЛЕНИЕ

ГЛАВА 3. ПОКАЗАТЕЛИ ВОСПРОИЗВОДСТВА НАСЕЛЕНИЯ

ГЛАВА 4. ВЫРАВНИВАНИЕ (СГЛАЖИВАНИЕ) ДЕМОГРАФИЧЕСКИХ ПОКАЗАТЕЛЕЙ. ИНТЕРПОЛЯЦИЯ И ЭКСТРАПОЛЯЦИЯ

2р.16к.