

Е АНДРЕЕВ,
канд. физико-математ. наук,
НИИ ЦСУ СССР

МЕТОД КОМПОНЕНТ В АНАЛИЗЕ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ

Анализ динамики и дифференциация смертности нередко ставит перед исследователем специфические, нередко сложные задачи.

Идея разложить изменение величины средней продолжительности жизни на компоненты, соответствующие изменению смертности в отдельных возрастах, была впервые предложена Ю. А. Корчак-Чепурковским, работа которого давно стала библиографической редкостью¹. Опираясь на эту идею, нам удалось разработать метод разложения различий в средней продолжительности жизни по возрастам и причинам смерти, позволяющий оценивать влияние смертности данной возрастной группы, от данной причины и в сочетании: влияние смертности от данной причины в данном интервале возрастов на степень различия средней продолжительности жизни в населении двух регионов или в данном населении в разные периоды.

Поясним на условных примерах смысл сказанного. Надо сравнить уровни смертности двух населений. Продолжительность жизни (e_0) в первом составляет 66,4 года, а во втором—68,7 года, т. е. разность 2,3 года. Коэффициенты смертности двух населений различаются по всем возрастным группам, в частности в возрастном интервале 55—64 года, где в первом населении коэффициент смертности (m) 21,90 на 1000 человек, во втором—17,99, т. е. разность составляет 3,91. Следует также учесть, что разность коэффициентов смертности от болезней системы кровообращения в этих возрастах составляет 1,93². Нас интересует, какая часть разности e_0 связана с различиями смертности в указанном интервале и, в частности, от болезней системы кровообращения. Как будет показано далее, на данный интервал (55—64 года) приходится 0,44 года продолжительности жизни из общей разности e_0 в 2,3 года, в том числе на долю болезней системы кровообращения—0,22 года. Общий же вклад данного класса причин смерти в различия средней продолжительности жизни составляет 0,92 года.

¹ Корчак-Чепурковский Ю. А. Влияние смертности в разных возрастах на увеличение средней продолжительности жизни.—В кн.: Изучение воспроизводства населения. М., Наука, 1968, с. 134—155.

² Расчеты приведены в табл. 1—3.

Следует подчеркнуть, что переход от ряда коэффициентов смертности к величине средней продолжительности жизни новорожденного сопряжен с рядом промежуточных расчетов, и прямое влияние изменения m_x на e_0 оценить не удается. Задача решается лишь при сравнении двух независимо рассчитанных таблиц смертности. Величина средней продолжительности предстоящей жизни для новорожденного в возрасте x в таблице смертности определяется как отношение $e_x = T_x/l_x$, где T_x — число человеко-лет жизни в возрастах старше данного возраста x (для новорожденного $x=0$), l_x — число доживших до возраста x лет. Для простоты записи будем полагать в дальнейшем, что $l_0=1$, хотя обычно в таблицах смертности l_0 принимается равным 100 000. Условно разделим человеческую жизнь на две части: от рождения до некоторого возраста x и от возраста x до максимального возраста живущих. Тогда e_0 можно представить как сумму

$$e_0 = L_{0/x} + T_x,$$

где L_{x_1/x_2} — $(x_1=0, x_2=x)$ — число живущих в интервале возрастов от x_1 до x_2 , показатель таблиц смертности, равный числу человеко-лет жизни прожитых совокупностью l_0 в возрастах от x_1 до x_2 . В свою очередь $T_x = l_x \cdot e_x$. Рассмотрим две различные таблицы смертности: для одного населения, но за разные периоды или для разных населений. Показатели второй таблицы будем отличать от показателей первой штрихом (например, e_0 и e'_0 ; l_0 и l'_0 и т. д.). Тогда разность величин средней продолжительности жизни новорожденного можно записать как

$$e'_0 - e_0 = (L'_{0/x} - L_{0/x}) + (l'_x e'_x - l_x e_x). \quad (1)$$

Разность, записанная в первых скобках формулы (1), измеряет прямое влияние смертности в возрастах от 0 до x лет на различие продолжительности жизни. Это слагаемое зависит только от уровня смертности в возрастах моложе x лет. Второе слагаемое зависит как от смертности в возрастах моложе x , так и старше x . Традиционно его можно записать как

$$(l'_x e'_x - l_x e_x) = (l'_x - l_x) e_x + l_x (e'_x - e_x) + (l'_x - l_x) (e'_x - e_x). \quad (2)$$

Первое слагаемое в (2) измеряет влияние изменения смертности в возрастах моложе x на число лет жизни в возрастах старше x вследствие того, что большая или меньшая доля новорожденных достигла возраста x лет при неизменной продолжительности жизни для x -летних. Второе слагаемое, напротив, измеряет влияние только смертности в возрастах старше x лет, при неизменном дожитии до этого возраста. Наконец, третье слагаемое в (2) измеряет совокупное влияние двух указанных факторов. Отнесем на счет изменения смертности в возрастах от 0 до x лет лишь первое слагаемое (2), зависящее только от изменения смертности в этих возрастах и измеряющее косвенное ее влияние в возрастах моложе x лет на e_0 .

Обозначим через ε_{x_1, x_2} изменение средней продолжительности жизни в результате изменений смертности в возрастах от x_1 до x_2 лет. Итак,

$$\varepsilon_{0/x} = L'_{0/x} - L_{0/x} + (l'_x - l_x) e_x. \quad (3)$$

Исходим из того, что если $x_1 < x_2 < x_3$, то

$$\varepsilon_{x_1, x_3} = \varepsilon_{x_1, x_2} + \varepsilon_{x_2, x_3}, \quad (4)$$

т. е. влияние изменения смертности в суммарном возрастном интервале на величину e_0 равно сумме влияний на e_0 изменений смертности слагающих возрастных интервалов. С учетом этого выведем формулу для ε_{x_1, x_2} . Итак, $\varepsilon_{0/x_2} = \varepsilon_{0/x_1} + \varepsilon_{x_1, x_2}$ или $\varepsilon_{x_1, x_2} = \varepsilon_{0/x_2} - \varepsilon_{0/x_1}$. Записав равенство (3) для x_1 и x_2 , проведя вычитание, получим

$$\varepsilon_{x_1, x_2} = L'_{x_1/x_2} - L_{x_1/x_2} + (l'_{x_2} - l_{x_2}) e_{x_2} - (l'_{x_1} - l_{x_1}) e_{x_1}. \quad (5)$$

Равенство (5) можно упростить, если учесть, что $T_x = l_x \cdot e_x$, а $L_{x_1|x_2} = T_{x_1} - T_{x_2}$. Тогда

$$\varepsilon_{x_1|x_2} = l'_{x_1}(e'_{x_1} - e_{x_1}) - l'_{x_2}(e'_{x_2} - e_{x_2}). \quad (6)$$

Формула (6) включает лишь два типа показателей—числа доживших и среднюю продолжительность предстоящей жизни в возрасте x , что существенно облегчает расчет.

Как видно из (6), если возрастная шкала разбита на ряд непересекающихся возрастных интервалов от x_k до x_{k+1} (где k —номер интервала) и для каждого из них вычислена величина $\varepsilon_{x_k|x_{k+1}}$, то сумма $\varepsilon_{x_k|x_{k+1}}$ по всем интервалам равна разнице $e'_0 - e_0$, так как каждое из произведений (6) вида $l'_{x_k}(e'_{x_k} - e_{x_k})$, кроме первого, равного $e'_0 - e_0$, и последнего, равного 0, будет входить в эту сумму дважды—один раз со знаком минус, второй—плюс.

В работе Ю. Корчака-Чепурковского приведена несколько иная, но эквивалентная запись этой формулы³.

В табл. 1 приводится пример расчета различий в продолжительности жизни двух населений по возрастным группам.

Хотя расчет проведен с тремя знаками после запятой, его реальная точность не выше 0,1, так как именно с такой точностью рассчитан исходный ряд e_x .

Выясним, какова будет формула для $\varepsilon_{x_1|x_2}$ в случае малого (строго говоря, бесконечно малого) интервала возрастов. Пусть $x_1=x$ и $x_2=x+\Delta x$. Проведем расчет, пренебрегая величинами, содержащими сомножитель Δx^2 . При такой точности

$$l_{x+\Delta x} = l_x(1 - \mu_x \Delta x), \quad L_{x|x+\Delta x} \approx l_x \cdot \Delta x,$$

$$T_{x+\Delta x} \approx T_x - l_x \Delta x,$$

Таблица 1

Пример расчета компонент различий продолжительности жизни, соответствующих отдельным возрастным группам

| Возраст x | Средняя продолжительность жизни в x лет | | Разность величин e_x (гр. 1—гр. 2) | Числа доживающих l'_x | $l'_x (e'_x - e_x)$ (гр. 3×гр. 4) | Компонент, соответствующий данному возрастному интервалу $\varepsilon_{x_1 x_2}$ |
|----------------|---|------------------------------|--|----------------------------|--------------------------------------|---|
| | в сравниваемом населении e'_x | в базовом населении e_x | | | | |
| A | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 68,7 | 66,4 | 2,3 | 1,00000 | 2,3000 | 0,741 |
| 1 | 69,4 | 67,8 | 1,6 | 0,97446 | 1,559 | -0,087 |
| 5 | 65,8 | 64,1 | 1,7 | 0,96987 | 1,649 | -0,090 |
| 15 | 56,1 | 54,3 | 1,8 | 0,96422 | 1,736 | 0,020 |
| 25 | 46,7 | 44,9 | 1,8 | 0,95316 | 1,716 | 0,121 |
| 35 | 37,4 | 35,7 | 1,7 | 0,93802 | 1,595 | 0,320 |
| 45 | 28,3 | 26,9 | 1,4 | 0,91097 | 1,275 | 0,341 |
| 55 | 20,0 | 18,9 | 1,1 | 0,84908 | 0,934 | 0,436 |
| 65 | 12,7 | 12,0 | 0,7 | 0,71195 | 0,498 | 0,282 |
| 75 | 7,4 | 6,9 | 0,5 | 0,43570 | 0,216 | 0,216 |

¹ В этой и последующих таблицах за базу сравнения взяты данные смертности мужчин за 1975 г. по Венгерской Народной Республике, с которыми сравниваются аналогичные данные по Народной Республике Болгарии. См.: Ежегодник мировой санитарной статистики. Женева, 1980, с. 536—547, 620—631, 992—995.

³ К сожалению, при публикации формул Ю. Корчака-Чепурковского были допущены некоторые неточности.

где μ_x — сила смертности в возрасте x лет, т. е. вероятность умереть в возрастах от x до $x+\Delta x$ равна в принятой точности $\mu_x \Delta x$; $\mu_x = \frac{d l_x}{d x} : l_x$. Подставляя соответствующие значения в (6), получим.

$$\varepsilon_{x,x+\Delta x} \approx l'_x (\mu_x - \mu'_x) e_x \Delta x. \quad (7)$$

Смысл этих преобразований состоит в том, что теперь в формуле содержится показатель силы смертности, который представляет собой сумму показателей силы смертности в данном возрасте от отдельных причин (групп причин) смерти. Обозначим через $\mu_x^{(i)}$ силу смертности в возрасте x от i -й причины смерти: $i=1, \dots, m$. Тогда $\mu_x = \mu_x^{(1)} + \dots + \mu_x^{(m)}$. Изменение e_0 за счет влияния смертности от причины i в возрастном интервале $x, x+\Delta x$ приблизительно равно

$$\varepsilon_{x,x+\Delta x}^{(i)} \approx l'_x (\mu_x^{(i)} - \mu_x^{(i)'} e_x \Delta x. \quad (8)$$

В соответствии с правилами математического анализа в силу (4) получаем

$$\varepsilon_{x_1, x_2} = \int_{x_1}^{x_2} l'_x (\mu_x - \mu'_x) e_x dx, \quad (9)$$

а изменение средней продолжительности жизни e_0 вследствие изменения смертности в возрастах от x_1 до x_2 от i -й причины смерти равно

$$\varepsilon_{x_1, x_2}^{(i)} = \int_{x_1}^{x_2} l'_x (\mu_x^{(i)} - \mu_x^{(i)'} e_x) dx. \quad (10)$$

Формулы (9) и (10), к сожалению, не могут быть непосредственно использованы в расчетах, так как включают интегрирование. Для практических расчетов перейдем к приближенным арифметическим формулам. В соответствии с правилами интегрирования получаем из (10)

$$\varepsilon_{x_1, x_2}^{(i)} = (\bar{\mu}_{x_1, x_2}^{(i)} - \bar{\mu}_{x_1, x_2}^{(i)'}) \int_{x_1}^{x_2} l_x e_x dx, \quad (11)$$

где $\bar{\mu}_{x_1, x_2}^{(i)}$ и $\bar{\mu}_{x_1, x_2}^{(i)'}$ — соответственно средние значения $\mu_x^{(i)}$ и $\mu_x^{(i)'}$ в возрастах от x_1 до x_2 . Аналогично, но без индекса (i) преобразуется формула (9). Отметим, что стоящий в правой части равенства интеграл не зависит от выбора причины смерти, т. е. не содержит индекса (i). Величина ε_{x_1, x_2} в случае смертности от всех причин известна, следовательно, интеграл можно рассчитать, если узнать разность $\mu_{x_1, x_2} - \mu'_{x_1, x_2}$. Среднее значение силы смертности может быть оценено с помощью соответствующего коэффициента. Можно принять приближенно, что $\bar{\mu}_{x_1, x_2}^{(i)} = m_{x_1/x_2}^{(i)}$. Таким образом,

$$\varepsilon_{x_1, x_2}^{(i)} = \frac{m_{x_1/x_2}^{(i)} - m_{x_1/x_2}^{(i)'}}{m_{x_1/x_2} - m'_{x_1/x_2}} \varepsilon_{x_1, x_2}, \quad (12)$$

или же, подставляя соответствующее выражение,

$$\varepsilon_{x_1, x_2}^{(i)} = \frac{m_{x_1/x_2}^{(i)} - m_{x_1/x_2}^{(i)'}}{m_{x_1/x_2} - m'_{x_1/x_2}} [l_{x_1} (e'_{x_1} - e_{x_1}) - l_{x_2} (e'_{x_2} - e_{x_2})]. \quad (13)$$

Из-за того, что равенства (12) и (13) приближенные, их следует использовать лишь на минимально возможных возрастных интервалах, величиной 5, максимум 10 лет. Суммарное влияние i -й причины смерти на продолжительность жизни должно быть рассчитано как сумма по отдельным возрастным интервалам величин $\varepsilon_{x_1, x_2}^{(i)}$. Как видно из (12), если расчет проводится по полному перечню причин смерти,

то сумма по i $e_{x_1 x_2}^{(i)}$ равна соответствующей величине $\varepsilon_{x_1 x_2}$. Сумма рассчитанных таким образом компонент соответствующих отдельных причин смерти точно равна разности соответствующих величин.

Продолжим, для примера, расчет, начатый в табл. 1 для случая двух причин смерти—болезни системы кровообращения и болезни органов дыхания (см. табл. 2 и табл. 3).

Таблица 2

Возрастные коэффициенты смертности от некоторых причин смерти $m_{x_1 x_2}^{(t)}$

| Возраст x | Коэффициенты смертности от | | | | | |
|----------------|----------------------------|------------------------|---------------------------------|------------------------|--------------------------|------------------------|
| | всех причин | | болезней системы кровообращения | | болезней органов дыхания | |
| | базовое население | сравниваемое население | базовое население | сравниваемое население | базовое население | сравниваемое население |
| 0 | 0,03613 | 0,02571 | 0,00000 | 0,00016 | 0,00295 | 0,00961 |
| 1 | 0,00083 | 0,00118 | 0,00004 | 0,00002 | 0,00014 | 0,00039 |
| 5 | 0,00043 | 0,00058 | 0,00002 | 0,00001 | 0,00003 | 0,00006 |
| 15 | 0,00126 | 0,00116 | 0,00008 | 0,00008 | 0,00003 | 0,00005 |
| 25 | 0,00182 | 0,00159 | 0,00027 | 0,00023 | 0,00007 | 0,00007 |
| 35 | 0,00389 | 0,00296 | 0,00114 | 0,00078 | 0,00015 | 0,00018 |
| 45 | 0,00917 | 0,00699 | 0,00338 | 0,00246 | 0,00040 | 0,00065 |
| 55 | 0,02190 | 0,01799 | 0,00986 | 0,00793 | 0,00116 | 0,00246 |
| 65 | 0,05243 | 0,04635 | 0,02711 | 0,02506 | 0,00361 | 0,00702 |
| 75 | 0,13912 | 0,13248 | 0,08731 | 0,07820 | 0,01011 | 0,01908 |

Разности соответствующих коэффициентов смертности из табл. 2 записаны в графы (2), (3), (5) табл. 3. Туда же, в гр. 1 перенесены величины $\varepsilon_{x_1 x_2}$ из графы (6) табл. 1.

Как видно из таблицы, для населения, принятого за базу сравнения, характерна более высокая смертность от болезней системы кровообращения и более низкая смертность от болезней органов дыхания. Различия в смертности от болезней системы кровообращения приходятся на возраста старше 35 лет, а от болезней органов дыхания—на младенческие и старческие возрасты. Однако эти различия почти компенсируются: за счет указанных двух причин продолжительность жизни новорожденного в базовом и сравниваемом населении различается на $-0,286$ года. Следовательно, остальные расхождения связаны с другими причинами смерти, анализ которых выходит за рамки данной статьи. Этот пример убедительно свидетельствует о том, что предложенный способ—эффективный метод соизмерения различий смертности от отдельных причин.

Отправной точкой наших рассуждений был вопрос: за счет чего продолжительность жизни во втором населении выше или ниже, чем в первом, и мы записали разность $e'_0 - e_0$. Однако, задавшись обратным сравнением $e_0 - e'_0$, можно было бы получить иное не только по знаку, но и по абсолютной величине значение для $\varepsilon_{x_1 x_2}$. При анализе динамики более или менее ясно, что взять за базу сравнения; поскольку изучаются изменения, то базой сравнения должна служить смертность более раннего периода. Однако при территориальных сопоставлениях выбор базы затруднителен.

Очевидно, определенный недостаток изложенной методики связан

Пример расчета компонент различий продолжительности жизни,
соответствующих отдельным причинам смерти и возрастным группам $\varepsilon_{x_1, x_2}^{(l)}$

| Возраст x | Все причины смерти | | в том числе | | | |
|----------------|--------------------------|--|--|--|--|--|
| | ε_{x_1, x_2} | разность коэффициен- тов смерт- ности | болезни системы кровообращения | | болезни органов дыхания | |
| | | | разность коэффициен- тов смерт- ности | $\varepsilon_{x_1, x_2}^{(1)}$ (1) \times (3) : (2) | разность коэффициен- тов смерт- ности | $\varepsilon_{x_1, x_2}^{(2)}$ (1) \times (5) : (2) |
| A | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 0,741 | 0,01042 | -0,00016 | 0,011 | -0,00666 | -0,474 |
| 1 | -0,087 | -0,00035 | 0,00002 | 0,005 | -0,00025 | -0,062 |
| 5 | -0,090 | -0,00015 | 0,00001 | 0,006 | -0,00003 | -0,018 |
| 15 | 0,020 | 0,00010 | 0,00000 | 0,000 | -0,00002 | -0,004 |
| 25 | 0,121 | 0,00023 | 0,00004 | 0,021 | 0,00000 | 0,000 |
| 35 | 0,320 | 0,00093 | 0,00036 | 0,124 | -0,00003 | -0,010 |
| 45 | 0,341 | 0,00218 | 0,00092 | 0,144 | -0,00025 | -0,039 |
| 55 | 0,436 | 0,00391 | 0,00193 | 0,215 | -0,00130 | -0,146 |
| 65 | 0,282 | 0,00608 | 0,00205 | 0,095 | -0,00341 | -0,158 |
| 75 | 0,216 | 0,00664 | 0,00911 | 0,296 | -0,00897 | -0,292 |
| Итого | 2,300 | - | - | 0,917 | - | -1,203 |

с асимметрией формулы (6). Если бы решалась задача разложения разности $e_0 - e'_0$ по возрастам, то получилась бы следующая формула:

$$\varepsilon'_{x_1, x_2} = l_{x_1}(e_{x_1} - e'_{x_1}) - l_{x_2}(e_{x_2} - e'_{x_2})$$

и, хотя величина $(e'_0 - e_0) + (e_0 - e'_0) = 0$, сумма

$$\varepsilon_{x_1, x_2} + \varepsilon'_{x_1, x_2} = (l'_{x_1} - l_{x_1})(e'_{x_1} - e_{x_1}) - (l'_{x_2} - l_{x_2})(e'_{x_2} - e_{x_2}),$$

как правило, не равна 0. В тех случаях, когда выбор базы для сравнения не столь однозначен, целесообразно воспользоваться величиной

$$\overset{\Delta}{\varepsilon}_{x_1, x_2} = \frac{1}{2}(\varepsilon_{x_1, x_2} - \varepsilon'_{x_1, x_2}). \quad (14)$$

Величина $\overset{\Delta}{\varepsilon}_{x_1, x_2}$ антисимметрична, т. е. $\overset{\Delta}{\varepsilon}_{x_1, x_2} = -\overset{\Delta}{\varepsilon}_{x_2, x_1}$, что соответствует логике анализа. Компоненты, соответствующие отдельным причинам смерти, могут быть оценены с помощью (14), причем очевидно, что $\overset{\Delta(l)}{\varepsilon}_{x_1, x_2}$ также обладает свойством антисимметрии. В большинстве случаев численные различия между ε_{x_1, x_2} , $\overset{\Delta}{\varepsilon}_{x_1, x_2}$, $\overset{\Delta}{\varepsilon}_{x_2, x_1}$ несущественны.

Таким образом, описанный метод позволяет определять отдельные компоненты изменения продолжительности жизни, измеряющие влияние смертности в определенных возрастных группах и отдельных причин смерти на динамику продолжительности жизни. Он же может быть применен и при территориальных сопоставлениях. Что касается задачи разложения, то предложенный метод позволяет решать ее по произвольному перечню причин смерти. Этот метод также открывает возможность соизмерить отдельные изменения с точки зрения их влияния на продолжительность жизни. Точность полученных результатов не ниже, чем при расчете таблиц смертности и продолжительности жизни. Представляется также, что описание динамики и дифференциации смертности в терминах одного универсального измерителя—средней продолжительности жизни—в ряде случаев может быть эффективным способом представления результатов статистического анализа.