Реферат выполнен Евгением Андреевым

Краткое содержание статьи

Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates.

by Evgeny M. Andreev, Vladimir M. Shkolnikov and Alexander Z. Begun

Demographic Research 2002, 7:14, 499-522.

Алгоритм для декомпозиции различий между агрегированными демографическими показателями и его применение к продолжительности жизни, продолжительности здоровой жизни, вероятности родить следующего ребенка и коэффициенту суммарной рождаемости.

В 1980-е годы в демографии был высокий интерес к проблеме декомпозиции различий в ожидаемой продолжительности жизни. Дискретный метод декомпозиции различий между двумя значениями продолжительностями жизни был независимо разработан в эти годы тремя исследователями из России, США, и Франции [Андреев, 1982, Arriaga, 1984, Pressat, 1985]. Формулы Андреева и Пресса полностью одинаковы. Формула Арриаги была представлена в несколько иной форме, и эта формула оказалось эквивалентной формулам Андреева и Пресса. Практически одновременно непрерывный вариант метода декомпозиции различий между двумя продолжительностями жизни по возрасту был предложен Поллардом [Pollard, 1982].

В начале 2000-х интерес к методам декомпозиции вновь повысился. На этот раз разрабатывались более общие подходы для решения общей проблемы декомпозиции – т.е. декомпозиции разницы между двумя величинами любого агрегированного демографического показателя. Решению именно этой проблемы посвящено данное исследование.

Множество агрегированных показателей может быть рассчитано на основе демографических таблиц. Каждый из них является функционалом, рассчитанным из вектора или матрицы элементарных демографических показателей (как правило, повозрастных вероятностей или коэффициентов). Анализируя изменения агрегированной демографического показателя во времени или его вариацию по странам, полезно иметь инструмент, позволяющий связать наблюдаемые изменения или различия с различиями в элементарных показателях, относящихся к конкретным возрастам и другим демографическим измерениям, таким как порядки рождения, причины смерти или группы населения.

Это исследование показало, что известные формулы для разложения различий между двумя продолжительностями жизни по возрасту представляют собой конкретную форму некоторого общего алгоритма, который включает пошаговую замену элементов одного вектора возрастных коэффициентов смертности соответствующими элементами другого вектора. В более общем случае элементы одной многомерной матрицы должны быть заменены соответствующими элементами другой матрицы. Корни такого подхода уходят в общий метод стандартизации [Kitagawa, 1964].

Продолжительность жизни

Ожидаемая продолжительность жизни рассчитывается из вектора возрастных коэффициентов смерти. Практически расчет ожидаемой продолжительности жизни осуществляется с помощью таблицы смертности, но весь этот расчет может быть записан как одна функция. Для двух населений 1 и 2 можно записать: $e_0^1 = e_0(M^1)$ и $e_0^2 = e_0(M^2)$, где вектора возрастных коэффициентов смертности $M^1 = \left\| m_x^1 \right\|$ и $M^2 = \left\| m_x^2 \right\|$, x=0, 1, 2, ... ω (ω - самый старший возраст). Переход от продолжительности жизни e_0^1 к продолжительности жизни e_0^2 соответствует замене вектора M^1 вектором M^2. Такая замена может быть выполнена пошагово: обозначим $M^{[x]}$ вектор элементарных коэффициентов смертности, в котором в возрастах $y < x$ стоят коэффициенты для второго населения m_y^2, при $y \geq x$ - для первого населения.[1] Разность $\delta_{0|x}^{2-1} = e_0(M^{[x]}) - e_0^1$ есть вклад возрастов $y < x$ в общую разность $e_0^2 - e_0^1$. Аналогично, вклад возрастов $y < x+1$ есть $\delta_{0|x+1}^{2-1} = e_0(M^{[x+1]}) - e_0^1$. Тогда вклад интервала $[x, x+1)$ может быть измерен как

$$\delta_x^{2-1} = \delta_{0|x+1}^{2-1} - \delta_{0|x}^{2-1} = e_0(M^{[x+1]}) - e_0(M^{[x]})$$

Поскольку $e_0^1 = e_0(M^{[0]})$, а $e_0^2 = e_0(M^{[\omega]})$ (,), то $\sum_{x=0}^{\omega} \delta_x^{2-1} = e_0^2 - e_0^1$. В статье доказано, что

$$\delta_x^{2-1} = l_x^2(e_x^2 - e_x^1) - l_{x+1}^2(e_{x+1}^2 - e_{x+1}^1),$$

где l_x - число доживших до возраста x (принято что $l_0 = 1$), e_x - продолжительность жизни в возрасте x, стандартные функции таблицы смертности.

Очевидно, что замены 1→2 и 2→1 – равноправны, но результат может зависеть от порядка замены т.е. δ_x^{2-1} может не быть равным $-\delta_x^{1-2}$, поэтому рекомендуется использовать симметричный компонент

$$\delta_x = \frac{1}{2} \cdot (\delta_x^{2-1} - \delta_x^{1-2}) = \frac{1}{2}\left[\left(l_x^1 + {}_x^2\right) \cdot \left(e_x^1 - e_x^2\right) - \left(l_{x+1}^1 + {}_{x+1}^2\right) \cdot \left(e_{x+1}^1 - e_{x+1}^2\right)\right].$$

Именно в таком виде была представлена формула декомпозиции в упомянутых выше работах Е.Андреева и Р.Пресса.

Представляется, что пошаговая замена от молодых к старшим возрастам, выглядит естественной, но совершенно не очевидно, почему этот путь - предпочтителен. Вообще говоря, замены могут быть организованы по-другому, например, от старших возрастов к младшим (этот подход реализован в непрерывной модели Полларда). Замена может проводится также в случайном порядке. В самом общем случае для каждого элемента потребуется сделать $2^{\omega-1}$ замен, ω - количество возрастных групп, т.е. величина порядка 100 для полных и порядка 20 для кратких таблиц смертности. Такая процедура замены - чрезвычайно трудоемка. Несколько численных экспериментов (не представленных в статье) показали, что результаты "полного" разложения различий между двумя продолжительностями жизни очень близки к оценкам, сделанным с помощью вышеприведенных формул.

[1] Для простоты все формулы в статье написаны для полных демографических таблиц (таблицы с шагом возраста 1 год). Они могут быть легко переписаны для кратких демографических таблиц.

Продолжительность здоровой жизни

Продолжительность здоровой жизни оценена с помощью формулы Д. Салливана [Sullivan, 1964]

$$h_0 = \sum_{x=0}^{\omega} {}_1L_x \cdot \pi_x ,$$

где π_x - доля человеко-лет, прожитых с "хорошим" состоянием здоровье в элементарном возрастном интервале $[x, x+1)$, а ${}_1L_x$ - число человеколет, прожитых в этом же интервале возраста, стандартный показатель таблицы смертности. Обычно веса получают из репрезентативных национальных обследований здоровья, включающих вопросы о состоянии здоровья, нетрудоспособности, инвалидности, физическом состоянии, болезнях и проч.. Продолжительность здоровой жизни есть функция двух векторов: описанного выше вектора возрастных коэффициентов смертности M и вектора возрастных весов (или долей) хорошего здоровья Π. В процессе пошаговой замены 1→2 каждом возрастном интервале можно сначала заменить m_y^1 на m_y^2, а потом π_y^1 на π_y^2. Или наоборот начать с замены весов здоровья, а потом заменить коэффициенты смертности. Компонент от замены коэффициента смертности в интервале возрастов $[x, x+1)$ равен

$$\lambda_x^{2-1} = \frac{1}{2}\{[(h_0(M^{[x+1]}, \Pi^{[x]}) - h_0(M^{[x]}, \Pi^{[x]})] + [(h_0(M^{[x+1]}, \Pi^{[x+1]}) - h_0(M^{[x]}, \Pi^{[x+1]})]\}.$$

Аналогичный компонент от замены весов здоровья равен

$$\gamma_x^{2-1} = \frac{1}{2}\{[h_0(M^{[x]}, \Pi^{[x+1]}) - h_0(M^{[x]}, \Pi^{[x]})] + [h_0(M^{[x+1]}, \Pi^{[x+1]}) - h_0(M^{[x+1]}, \Pi^{[x]})]\}.$$

Эти базовые формулы могут быть преобразованы в окончательные формулы для расчета вкладов от различий возрастных коэффициентов смертности и вкладов от различий возрастных весов здоровья выглядят:

$$\lambda_x = \frac{1}{4}(l_x^1 + l_x^2)({}_1P_x^2 - {}_1P_x^1)(\pi_x^1 + \pi_x^2) + \frac{1}{2}(h_{x+1}^1 l_x^2 + h_{x+1}^2 l_x^1)({}_1q_x^1 - {}_1q_x^2) \text{ и}$$

$$\gamma_x = \frac{1}{4}(l_x^1 + l_x^2)({}_1P_x^1 + {}_1P_x^2)(\pi_x^2 - \pi_x^1),$$

где h_x - ожидаема продолжительность здоровой жизни в возрасте x, величины ${}_1P_x$ рассчитывается как ${}_1L_x / l_x$, а ${}_1q_x$ - вероятность смерти в интервале возраста $[x, x+1)$, показатель таблицы смертности.

Вероятность родить следующего ребенка и коэффициент суммарной рождаемости

Общий алгоритм пошаговой замены может быть применен для декомпозиции различий между кумулятивными характеристиками рождаемости, рассчитанными на основе таблиц рождаемости по очередности рождений. В отличие от обычного коэффициента суммарной рождаемости (TFR), который равен сумме повозрастных коэффициентов рождаемости для всех очередностей вместе, коэффициент суммарной рождаемости с учетом очередности рождений (TFR_P) есть результат достаточно сложных вычислений. Таблица рождаемости по очередности рождений есть мультистатусная демографическая таблица, основанная на матрице коэффициентов рождаемости по возрасту матери и очередности рождения $F = \left\| f_{x,par} \right\|$, в которой каждый элементарный коэффициент рождаемости $f_{x,par}$ есть отношение числа рождений данной очередности к среднегодовому числу женщин в возрасте x имеющих $par-1$ детей.

Формулы для расчета таблицы приведены в Приложении к статье. Возраст женщин изменяется в интервале репродуктивных возрастов $[\alpha, \beta]$, а очередность меняется от 1 до максимальной очередности p. В случае, когда последняя очередность есть «пятые и следующие рождения», общее число возможных замен в каждом возрасте как минимум равно $2^4 = 16$. Вывести аналитическую формулу декомпозиции не удается, и декомпозиция, реализующая алгоритм замены, осуществляется численно. Замена показателя $f_{x,par}$ влияет на число женщин имеющих число детей par и большее число детей и на возраста $y \geq x$ и, следовательно, на все показатели таблицы рождаемости для этих порядков рождения и возрастов.

В качестве примера декомпозиции изменения показателя *TFR_P* в статье рассмотрены данные о рождаемости в России в 1989 и 1994 гг. Для этих периодов на основе данных переписи населения 1989 г. и микропереписи 1994 г. о распределении женщин по возрасту и числу рожденных детей и рутинных данных статистики рождаемости были построены две таблицы рождаемости по очередности рождений и получено два значения суммарного коэффициента рождаемости.

В статье приведены результаты разложения разницы между значениями *TFR_P* за 1994 и 1989 гг. по очередностям рождения и возрастам (рисунок 1).

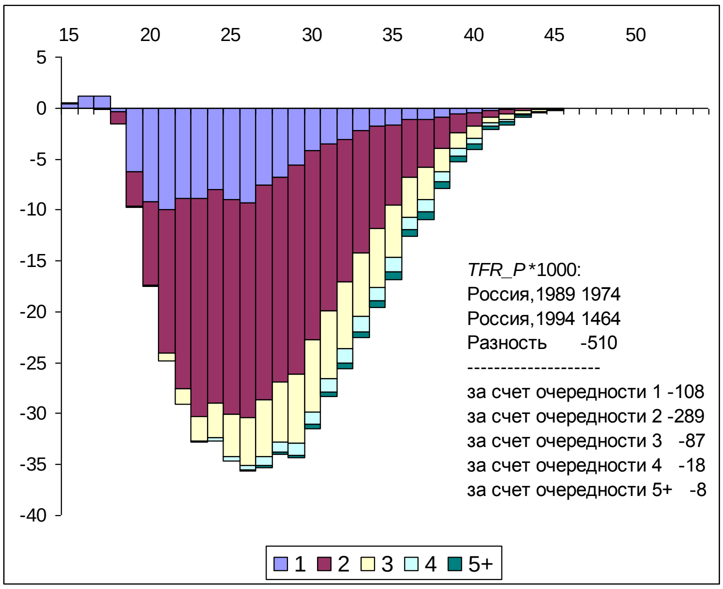

Рисунок 1. Декомпозиция снижения коэффициента суммарной рождаемости в России в 1989-1994 по возрасту и очередности рождений

Таблица рождаемости по очередности рождений содержит табличные числа рождений в каждом возрасте (в сумме равные коэффициенту суммарной рождаемости) и возникает соблазн интерпретировать разности этих коэффициентов как возрастные компоненты разложения TFR_P по возрастам. Расчет (рисунок 2) показывает, что это неверно. Дело в том, что, как уже отмечалось, – число рождений очередности выше первой в каждом возрасте зависит от интенсивности рождаемости не только данной, но и всех более низких очередностей в более молодых возрастах.

Аналогичным образом могут быть рассчитаны компоненты разностей в вероятностях рождения следующего ребенка.

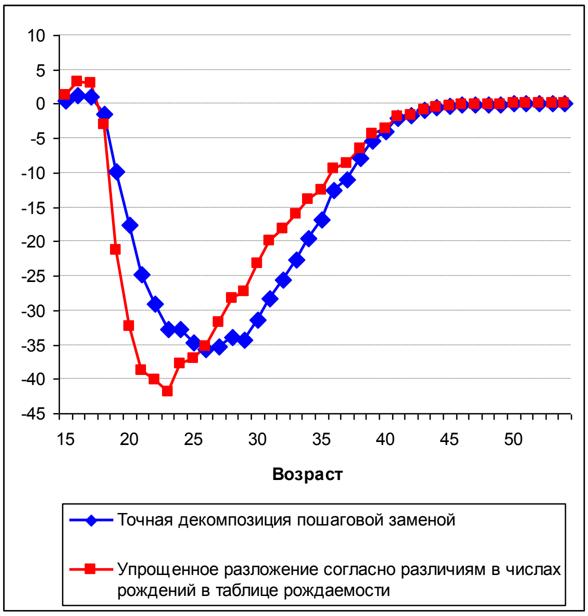

Рисунок 2. Возрастные компоненты снижения TFR_P в России между 1989 и 1994. Сравнение точного разложения на основе алгоритмом пошаговой замены с упрощенным разложением с использованием только различий в числах рождения в каждом возрасте в таблице рождаемости по очередности рождений

Заключительные соображения

Алгоритм пошаговой замены - универсальный инструмент для декомпозиции различий между обобщающими показателями, рассчитанными на основе демографических таблиц. Число примеров в данном исследовании могло быть увеличено. В частности, можно анализировать различия между коэффициентами Джини по возрастам, причинам смерти и с учетом состава населения. Особенно важна возможность учитывать состав населения, что может быть реализовано с помощью того же самого алгоритма замены, включая замену определенного для возрастной группы коэффициента и замены возрастного веса социально-демографических групп в населении.

В заключение статьи отмечаются две особенности алгоритма пошаговой замены, которые не были отмечены ранее.

Зависимость от пути. В статье повсеместно сравниваются два состояния населения обозначенные "1" и "2". Население переходит из "1" в "2" без каких-либо промежуточных состояний. В реальности может существовать последовательность промежуточных состояний, и результаты разложения могут зависеть от пути перехода из состояния "1" в состояние "2". Например, если провести серию последовательных декомпозиций изменения TFR_P в России 1989→1990, 1990→1991, 1991→1992, 1992→1993, и 1993→1994, затем суммировать компоненты соответствующие данному возрасту и очередности, то результат будет отличен от прямой декомпозиции 1989→1994.

Лучшее решение состоит в том, чтобы сделать все ежегодные переходы и затем суммировать их итоги. Но различия между прямым переходом (как 1989→1994) и последовательностью ежегодных переходов, как правило, невелики, поэтому такое решение не пользуется популярностью среди исследователей. Однако, нет никакой гарантии, что такие различия будут небольшими во всех случаях.

Возраст как особое измерение. В данной статье возраст рассматривался как особое измерение, причем особое с двух точек зрения. Во-первых, замены по возрасту всегда проводились от младших возрастов к старшим. Во-вторых, декомпозиция по всем остальным признакам проводилась в рамках данной группы возраста, т.е. возраст всегда играл роль первого измерения.

Зависимость от последовательности замен. Последовательность замен также влияет на результат. Например, при декомпозиции по возрастной смертности и по возрастным весам групп населения, при анализе различий между двумя продолжительностями жизни в зависимости от возрастной смертности и доли групп, можно сначала сделать замены возрастных коэффициентов смертности в пределах каждой группы населения, а можно заменять определенные для групп коэффициенты смертности в пределах одной возрастной группы. Вообще говоря, все схемы замены являются одинаково приемлемыми. Идеально рассчитанные компоненты должны быть основаны на усреднении *всех возможных* последовательностей замен. Этот общий подход может быть реализован только в некоторых случаях, но для сложных агрегированных показателей и больших размерностей потребует слишком много времени.

Комментарий, добавленный при подготовке данного текста. Полное число замен, которое может быть сделано для каждого из элементарных демографических показателей, ровно 2^{N-1}, где N – общее число элементарных показателей. Однако, если вклад каждого элементарного показателя в разность сравниваемых агрегированных величин оценить как среднее арифметическое вкладов при этих 2^{N-1} заменах, то может оказаться так, что такие оценки для всех N показателей в сумме не равны разнице сравниваемых величин агрегированного показателя. Для достижения равенства необходимо делать пошаговые последовательности замен 1→2, как это описано в примере для продолжительности жизни, но в произвольном порядке. В этом случае придется делать $N!$ замен, тогда как общее число обязательных замен, равно $N \cdot 2^{N-1}$, т.е. много меньше $N!$. Но можно обойтись $N \cdot 2^{N-1}$заменами: тот же количественный результат может быть достигнут с помощью расчета для каждого элементарного показателя взвешенной средней с весами $\binom{N-1}{k} \cdot 2^{-N}$, где k число пошаговых замен, выполненных перед заменой данного элементарного показателя. Здесь $\binom{a}{b}$ есть число сочетаний из a элементов по b, в России также обозначается C_a^b.

Ссылки

Андреев Е.М. (1982). Метод компонент в анализе продолжительности жизни // Вестник статистики. № 9. С. 42-48

Arriaga, E. (1984). Measuring and explaining the change in life expectancies. Demography 21(1), 83-96. Journal of Japan Statistical Society, 13: 95-98.

Kitagawa, E. (1964). Standardized comparisons in population research. Demography, 1, 296-315.

Pollard, J.H. (1982). The expectation of life and its relationship to mortality. Journal of the Institute of Actuaries, 109, Part II, No 442, 225-240.

Pollard, J.H. (1988). On the decomposition of changes in expectation of life and differentials in life expectancy. Demography, 25, 265-276.

Pressat, R. (1985). Contribution des écarts de mortalité par âge à la différance des vies moyennes. Population, 4-5, 766-770.

Sullivan, D.F. (1964). A single index of mortality and morbidity. HSMHA health report, 86, pp. 347-354.