В. П. Пересада

канд. техн. наук, ведущий научный сотрудник ЗАО «Инфоресурс» (Санкт-Петербург)

ПРОГНОЗНЫЕ ОЦЕНКИ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ ПО ВОЗРАСТНЫМ ГРУППАМ ПО ДАННЫМ ВСЕРОССИЙСКОЙ ПЕРЕПИСИ НАСЕЛЕНИЯ 2010 ГОДА

Введение

Принятие таких социально значимых решений, как реформирование системы налогообложения, принципов социальной поддержки и т. п., существенно влияют на динамику развития демографических и экономических процессов. Обоснованное прогнозирование их влияния на демографические изменения в стране возможно только при правильном выборе статистической модели реальных данных, представляемых таблицей («Распределение населения по возрастным группам»).

В этой таблице приводятся данные о численности населения каждой из возрастных групп. Число людей, доживающих от возраста t (l_t) до возраста t+1 (l_{t+1}), является случайной величиной. Поэтому данные переписи населения разных лет имеют разную численность возрастных групп.

Закон распределения вероятности этих процессов может быть описан двумя характеристиками. Одна из них — плотность распределения вероятности численности населения для возрастной группы $w(l_t)$. Вторая — закон изменения *интенсивности смертности* $h(l_t) = w(l_t)/(1 - F(l_t))$ [человек/год], где $F(l_t)$ — функция распределения вероятности числа умерших людей — $F(u) = \int_0^u w(l_t)dl_t$. Здесь $u = (l_t - l_t^*)/\sigma$, а l_t^* представляет собой математическое ожидание l_t. Разность $p_t = 1 - F(l_t)$ есть вероятность выживания в течение года t, т. е. число людей оставшихся в живых к концу года. Аналогичный показатель — *интенсивность отказов* широко используется при анализе вероятности выхода из строя сложных технических систем (Хан, Шапиро, 1969, с. 337—339).

При выборе статистической модели следует рассматривать несколько предположений о законе распределения вероятности. Предположение о нормальном распределении интенсивности смертности неприемлемо. Случайная величина l_t не может быть отрицательной. Предположение о равномерном распределении интенсивности смертности возможно. Однако это означает, что с начального момента интенсивность смертности будет непрерывно возрастать, что не соответствует действительности.

Принималось предположение, что на всем протяжении жизни выбранного поколения уровень повозрастной смертности сохраняется таким же, каким он был в начале составления этих таблиц. На основе такого предположения *расчетным* путем формировались *таблицы смертности и средней продолжительности жизни* определенного поколения населения страны (Таблицы смертности и средней продолжительности жизни населения СССР 1956—1959, 1962).

Такие таблицы считались наиболее совершенным способом изучения повозрастной смертности и такого важного показателя, как средняя продолжительность жизни населения. При анализе указанных таблиц обращает на себя внимание совершенно четкая закономерность: численность населения в этих таблицах непрерывно, без отклонений, уменьшается с увеличением возраста.

Число людей, доживающих от возраста l_{t_0}, принятого за начало отсчета, до заданного возраста K, определяется равенством

$$l_K = l_{t_0}(p_0 p_1 \dots p_t) = l_{t_0}\prod_{t=0}^{K} p_t.$$

Как видно из этого соотношения, прогнозное значение (l_K) определяется заданием только двух исходных величин. Первая из них — число людей, доживших от начального возраста $t = 0$ до первого года. Вторая — число людей, умерших в течение следующего года. В методических пояснениях (Итоги Всероссийской переписи..., с. 4), для расчета средней продолжительности предстоящей жизни, вводился вспомогательный показатель (T_x), который определялся по довольно спорному алгоритму. Именно он являлся основой для определения продолжительности предстоящей жизни для лиц, достигших возраста x лет.

Предположение, что на всем протяжении жизни поколения уровень повозрастной смертности сохраняется таким же, каким он был в начале составления таблиц, оказалось несостоятельным. Прогнозные оценки смертности и средней продолжительности жизни, даваемые на основе такой статистической модели, не соответствовали действительности. В разные возрастные периоды средняя продолжительность предстоящей жизни оказывалась разной. С 1990-х гг. в изданиях Росстата публикуются только таблицы «Распределение населения по возрастным группам». Эти данные Росстат обычно графически представляет в виде возрастно-половой пирамиды. Более подробные данные о распределении населения по возрасту за каждый год представлены на сайте Росстата.

Показатель «интенсивность смерности»

Предположение о постоянстве смертности означало, что показатель **интенсивности смертности** подчиняется экспоненциальному закону распределения. Плотность распределения вероятности экспоненциального закона имеет вид:

$$w(t;\lambda) = \begin{cases} \lambda \cdot \exp(-\lambda t),\ t \geqslant 0,\ \lambda > 0, \\ 0,\ \text{в остальных случаях.} \end{cases}$$

Для такого закона интенсивность смертности определяется соотношением $h(t) = w/(1 - F) = \lambda$. Тогда продолжительность жизни определится величиной $Tm = 1/\lambda$. Если параметр λ неизвестен, то интенсивность смертности можно оценить через его математическое ожидание. Математическое ожидание этого параметра определяется соотношением $\lambda^* = F_D/T_S$, которое получается методом максимального правдоподобия (Хан, Шапиро,1969, с. 337—339).

В качестве расчетного примера далее будем использовать данные таблицы (Итоги Всероссийской переписи...). Из них следует, что суммарное число умерших людей T_D за $T = 100$ лет, определяемое как разность между начальным и конечным значением числа умерших за это время, равно $T_D = l_1 - l_N = 1641 - 4 = 1637$.

Суммарное время T_S (число человеко-лет), в течение которого число людей оставалось постоянным, обеспечивалось заменой умерших людей людьми, вступившими в соответствующий возраст. Так как процесс ежегодного обновления протекает непрерывно, то это время определяется суммой арифметической прогрессии $S_N = (l_1 + l_N)N/2$. Суммарное время для рассматриваемого примера $T_S = S_N = 50 \cdot 1641 = 82\,250$. Тогда $\lambda^* = F_D/T_S = 1641/82\,250 = 0{,}02$, а среднее время жизни $Tm = 1/\lambda^* = 50{,}0$ лет. На рис. пунктирная прямая соответствует интенсивности смертности при $\lambda^* = F_D/T_S = 0{,}02$.

Однако такая статистическая модель не вполне соответствует реальности.

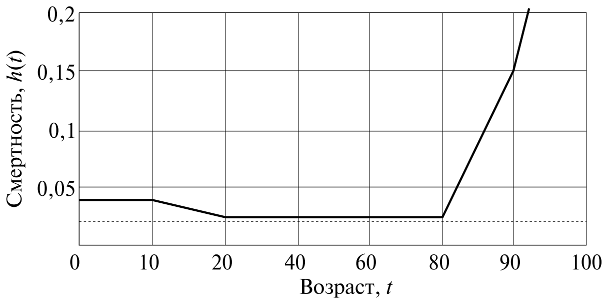

Рис. Зависимость интенсивности смертности $h(t)$ от возраста

Анализ таблиц Всероссийской переписи населения 2010 г.

Анализ указанных таблиц показывает, что характер распределения данных оказывается различным для разных возрастных категорий. Из их анализа видно, что в период *«детско-юношеский возраст»*, от $t_{00} = 0$ до $t_0 = 19$ лет, интенсивность смертности изменяется своеобразно. Ее относительно большое значение, в первые десять лет этого периода, обусловлено младенческой смертностью. Суммарная численность мужчин и женщин остается почти постоянной. Суммарное число умерших людей составляет $F_D = 1641 - 1361 = 280$, а суммарное время $T_S = 5 \cdot 1641 = 8205$. Средняя интенсивность смертности оказывается $h(t) = 280/8205 = 0{,}034$, что соответствует времени жизни $Tm = 1/h(t) = 29{,}4$ года. Далее, как это показано на рисунке, интенсивность смертности сокращается, достигая к концу первого периода $h(t) = 0{,}0234$.

Как видно из статистического анализа указанных таблиц, на протяжении второго возрастного периода, *«трудоспособный возраст»*, от $t_0 = 20$ лет до $t_1 = 80$ лет, дисперсия общего числа мужчин и женщин оказывается малой. В этот период суммарное число умерших людей составляет $F_D = 2300 - 687 = 1613$. Суммарное время $T_S = 30 \cdot 2300 = 69\,000$. Соответственно $h(t) = 1613/69\,000 = 0{,}0234$, что дает среднее время жизни в этот период $Tm = 42{,}78$ лет. Интенсивность смертности остается постоянной, и она может быть описана экспоненциальным законом.

В третьем возрастном периоде, *«постпенсионном возрасте»*, от $t_1 = 80$ лет до $T = 99$ лет, интенсивность смертности вновь возрастает вследствие физиологического износа организма. Статистическая модель третьего периода может быть описана функцией, которая изменяется от нуля до бесконечности, что естественно, так как число умерших людей не может быть отрицательным. Такой функцией может быть *логарифмически нормальное распределение*. Оно широко используется в качестве статистической модели для случайных величин, значение которых получается в результате умножения некоторого числа случайных отклонений, т. е. это мультипликативная величина $l_K = l_{t_0}\prod_{t=0}^{K} p_t$, в то время как обычное нормальное распределение является статистической моделью для случайной величины, значение которой получается в результате сложения некоторого числа случайных отклонений, т. е. это аддитивная величина

$$Y_{K+1} = (y_1 + y_2 + ... + y_K)y_{K+1} = \sum_{x=1}^{K} y_x + y_{K+1}.$$

Статистическая логнориальная модель используется в экономике при оценке плотности вероятности суммы личных доходов, суммы банковских вкладов, а также в демографических оценках и т. п. (Хан, Шапиро, 1969, с. 337—339).

Проверка справедливости гипотезы о логарифмически нормальном законе плотности распределения вероятности в третьем периоде осуществляется путем перехода к новой переменной $z_t = \ln(l_t)$. В результате такого перехода плотность вероятности новой переменной z_t станет описываться нормальным законом

$$w(z) = dF(z)/dz = \frac{1}{\sqrt{2\pi}} \exp - \left(z - z^*\right)^2 / 2\sigma_z^2.$$

Использование новой переменной позволит упростить оценку интенсивности смертности после пенсионного возраста. Затем оказывается возможным найти численность населения указанного возрастного периода, вычисляя антилогарифм $l_t = e^z$.

Следует отметить, что математическое ожидание исходной переменной l_t^* не является параметром, который характеризует центр распределения логарифмически нормального закона, а $\sigma = \exp(\sigma_z)$, его масштабом, как это имеет место для нормального распределения.

Обоснованность логарифмически нормального закона плотности распределения вероятности

Логарифмически нормальный закон плотности распределения вероятности в третьем периоде может быть подтвержден путем использования критерия $W = b^2/S^2$ (Хан, Шапиро, 1969, с. 337—339). Следуя методике, изложенной в этой работе, вычисляются два показателя. Первый из них — центральный момент второго порядка $S^2 = \sum_{k=1}^{K} z_k - \sum_{k=1}^{K} (z_k)^2/K$. Он связан с дисперсией σ^2 соотношением $S^2 = \sigma^2 K$. Здесь K — число возрастных групп.

Для вычисления второго показателя b^2 необходимо ранжировать имеющиеся данные по их возрастанию $z_{1(\min)} < z_2 < z_3 < ... z_{K(\max)}$. Из табл. видно, что в этот период численность выживших в каждый последующий год оказывается меньше предыдущего (кроме двух последних лет). Величина b вычисляется с использованием табличных значений коэффициентов a_k, т. е.

$$b = a_K(z_K - z_1) + a_{K-1}(z_{K-1} - z_2) + a_{K-2}(z_{K-2} - z_3) + ... + a_{K/2}(z_{K/2} - z_{K/2-1}).$$

В связи с симметричностью функции $w(z)$ оказывается возможным рассматривать только половину ранжированных данных, как это показано в табл.

Статистический анализ переменных $z_t = \ln(l_t)$ при $K = 20$ дал все необходимые значения характеристик рассматриваемого распределения населения по возрастным группам. Математическое ожидание $z^* = 4{,}4036$; $\mathrm{Sum}(z) = 88{,}07$; $b^2 = 47{,}128$; $\sigma_z = 1{,}626$; $S^2 = N\sigma_z^2 = 52{,}9$. Соответственно критерий $W = b^2/S^2 = 0{,}891$. Данные «Таблицы процентилей критерия W» (Хан, Шапиро, 1969, с. 337—339) показывают, что полученное значение W позволяет считать приемлемой гипотезу о LogNrm распределении данных третьего периода. Интенсивность смертности в течение этого периода меняется. В 80 лет он еще остается равной 0,0234, что соответствует времени жизни $Tm = 1/h(t) = 42{,}74$ года. Далее, как это показа-

Таблица

Распределение население по возрастным группам

№ п/п, годы	Годы жизни, (t)	Номер по ранжиру	Численность населения, тыс. чел. (l_t)	Новая переменная $Z_K = \ln(l_t)$	Весовые коэффи-циенты (a_k)	Состав показателя (b_k)
1	99	1	4,134	1,419245	0,4734	2,388
2	98	2	7,117	1,962486	0,3211	1,445
3	97	3	10,422	2,343919	0,2565	1,042
4	96	4	15,961	2,770148	0,2085	0,717
5	95	5	20,514	3,021076	0,1686	0,509
6	94	6	22,681	3,121528	0,1334	0.392
7	93	7	26,890	3,291754	0,1013	0,232
8	92	8	47,157	3,853482	0,0711	0,100
9	91	9	58,318	4,065911	0,0422	0,036
10	90	10	76,809	4,341322	0,0140	0,004
11	89	11	101,361	4,618688		
12	88	12	138,457	4,930560		
13	87	13	193,158	5,263508		
14	86	14	264,466	5,577713		
15	85	15	335,029	5,814217		
16	84	16	420,338	6,041059		
17	83	17	495,999	6,206574		
18	82	18	605,854	6,406639		
19	81	20	641,787	6,464256		
20	80	19	641,752	6,464202		
Суммарные значения				Sum = 88,07		Sum = b = 6,865

но на рисунке, интенсивность смертности нарастает, достигая к 99 году $h(t) = 0{,}4$. Это соответствует времени жизни $Tm = 1/h(t) = 2{,}5$ года.

Заключение

Таким образом, статистическая модель с тремя периодами изменения интенсивности смертности позволяет осуществлять боле достоверное прогнозирование интенсивности смертности населения страны (Елисеева, Пересада, 2003, с. 10—15). Она может служить конечным показателем эффективности и целесообразности принятия тех или иных социально значимых решений.

Анализ показателей интенсивности смертности за ряд прошлых лет позволит получить прогнозные оценки тенденций роста или сокращения народонаселения в перспективе. Более достоверные прогнозные оценки могут быть получены при использовании статистических асимптотических моделей с распределением Вейбулла.

Источники

Елисеева И. И., Пересада В. П. Межотраслевой баланс и экономическое прогнозирование: учеб. пособие. СПб., 2003.

Итоги Всероссийской переписи населения 2010 года. Таблица «Распределение населения по полу и возрасту». [Электронный ресурс]. — Режим доступа: http://www.gks.ru/free_doc/new_site/perepis2010/croc/perepis_itogi1612.htm

Таблицы смертности и средней продолжительности жизни населения СССР 1958—1959 гг. М., 1962.

Хан Г., Шапиро С. Статистические модели в инженерных задачах. М., 1969.