

УДК 519.21

А.А. Назаров, М.Г. Носова

МЕТОД ПЕРЕДВИЖКИ ВОЗРАСТНЫХ ГРУПП В ДЕМОГРАФИИ И ЕГО ПРИЛОЖЕНИЯ¹

Предлагаются методы прямой и обратной передвижки возрастных групп. Находятся основные характеристики, определяющие распределение вероятностей значений численностей групп для методов прямой и обратной передвижки. Выполняется оценка величины людских потерь Российской Федерации в годы Великой Отечественной войны с использованием данных переписей населения 1939 и 1959 гг. и методов прямой и обратной передвижки возрастных групп.

Ключевые слова: *метод передвижки, численность населения.*

В связи с возрастающей ролью демографического фактора в социально-экономическом планировании актуальными являются перспективные расчеты численности и состава населения. В решении этой задачи полезным является математическое моделирование. Разработка и использование разного рода математических моделей служат как для анализа воспроизводства населения в целом, так и для выявления закономерностей развития тех или иных демографических процессов. При моделировании принимаются определенные исходные предположения в отношении основных составляющих процесса (рождаемость, смертность, миграция и т.п.). На этой основе исчисляются другие характеристики населения и его структуры.

Особое место в математическом моделировании занимает метод передвижки возрастов (или метод компонент), разработанный П. К. Уэлптоном [1]. Исчислению половозрастной структуры населения методом возрастной передвижки занимались С.Г. Струмилин, А.Я. Боярский, П.П. Шушерин, М.С. Бедный, С. Щербов, В. Лутц, У. Сандерсон, а также Комиссия по народонаселению ООН, Государственный комитет Российской Федерации по статистике, Центр демографии и экологии человека [2, 3].

Метод передвижки возрастных групп достаточно эффективен для краткосрочных прогнозов с горизонтальным планированием на срок не более 10 – 15 лет. В данной работе методом передвижки возрастных групп определяется величина людских потерь Российской Федерации в годы Великой Отечественной войны, с использованием данных переписей населения 1939 и 1959 гг.

1. Метод передвижки по возрастам

Метод прямой передвижки применяется для определения оценок значений численности $N(x+t, t+\tau)$ группы лиц возраста $x+t$ в году $t+\tau$ при условии, что известна численность $N(x, t)$, τ – шаг прогнозирования. Численность населения рас-

¹ Работа выполнена при поддержке АБЦП «Развитие научного потенциала высшей школы (2009 – 2010 гг.)» Федерального агентства по образованию РФ по проекту «Разработка методов исследования немарковских систем массового обслуживания и их применения к сложным экономическим системам и компьютерным сетям связи».

смаатривается в совокупности, без разделения по половому признаку. Обозначим через $p(x, x+\tau)$ условную вероятность достижения возраста $x+\tau$ лицами возраста x .

Известно [4], что

$$p(x, x+\tau) = S(x+\tau) / S(x),$$

где $S(x)$ – функция дожития [4], которая есть вероятность того, что человек доживет до возраста x . При заданных $N(x, t)$ и $p(x, x+\tau)$ распределение вероятностей значений величины $N(x+\tau, t+\tau)$ определяется схемой Бернулли и является биномиальным:

$$P\{N(x+\tau, t+\tau) = m\} = C_{N(x, t)}^m p(x, x+\tau)^m (1-p(x, x+\tau))^{N(x, t)-m} \quad (1)$$

с математическим ожиданием

$$MN(x+\tau, t+\tau) = N(x, t)p(x, x+\tau) = N(x, t) \frac{S(x+\tau)}{S(x)}. \quad (2)$$

Обозначая оценку значения величины $N(x+\tau, t+\tau)$ тем же символом, равенство (2) перепишем в виде

$$N(x+\tau, t+\tau)S(x) = N(x, t)S(x+\tau) + \varepsilon_1, \quad (3)$$

где ε_1 – случайная ошибка с математическим ожиданием $M\varepsilon_1 = 0$.

Равенство (3) является основным для применения метода передвижки по возрастам. В частности, для метода прямой передвижки оно записывается в виде

$$N(x+\tau, t+\tau) = N(x, t) \frac{S(x+\tau)}{S(x)} + \varepsilon_2, \quad (4)$$

где $N(x, t)$ задано, а $N(x+\tau, t+\tau)$ является оценкой значения численности демографической группы лиц возраста $x+\tau$ в году $t+\tau$, а ε_2 – случайная ошибка с математическим ожиданием $M\varepsilon_2 = 0$.

При замене аргумента x на $x-\tau$ и t на $t-\tau$ равенство (3) перепишем в виде

$$N(x, t)S(x-\tau) = N(x-\tau, t-\tau)S(x) + \varepsilon_3,$$

где ε_3 – случайная ошибка с математическим ожиданием $M\varepsilon_3 = 0$. Откуда получим

$$N(x-\tau, t-\tau) = N(x, t) \frac{S(x-\tau)}{S(x)} + \varepsilon_4, \quad (5)$$

где $N(x, t)$ задано, а $N(x-\tau, t-\tau)$ является оценкой значения численности демографической группы лиц возраста $x-\tau$ в году $t-\tau$, ε_4 – случайная ошибка с математическим ожиданием $M\varepsilon_4 = 0$. Равенство (5) позволяет определить оценку значения численности демографической группы в прошлые моменты времени. Назовем это методом обратной передвижки. Оценка $N(x-\tau, t-\tau)$ требует дополнительного исследования, которое выполним ниже.

Из равенства (1) очевидно следует, что оценка численности $N(x+\tau, t+\tau)$, полученная прямой передвижкой, имеет дисперсию

$$DN(x+\tau, t+\tau) = N(x, t)p(x, x+\tau)(1-p(x, x+\tau)) = N(x, t) \frac{S(x+\tau)}{S(x)} \left(1 - \frac{S(x+\tau)}{S(x)}\right),$$

а коэффициент вариации V_1 этой величины составляет

$$V_1 = \frac{\sqrt{DN(x+\tau, t+\tau)}}{MN(x+\tau, t+\tau)} = \frac{1}{\sqrt{N(x, t)}} \sqrt{\frac{S(x)}{S(x+\tau)} - 1}.$$

Определим границы значений коэффициента вариации V_1 . Так как численности пятилетних возрастных групп в статистических данных РФ составляют порядка нескольких миллионов, то первый сомножитель $1/\sqrt{N(x,t)}$ имеет величину менее 10^{-3} . Используя статистические данные о зависимости функции выживания от возраста и проанализировав все возможные значения второго сомножителя для $\tau \in [1; 45]$ лет и $x \leq 70$ лет, получаем, что второй сомножитель принимает максимальное значение 12,578 при $\tau = 45$ лет. В результате получаем, что коэффициент V_1 в этом случае имеет значения менее 0,0126. Поскольку оценка (4) обладает достаточно высокой точностью, то ошибкой ε_2 можно пренебречь.

2. Метод обратной передвижки

Равенство (5), определяющее оценку численности $N(x-\tau, t-\tau)$ в методе обратной передвижки, получено применением прямой передвижки, поэтому необходимо найти характеристики этой оценки, в частности ее математическое ожидание и дисперсию. При заданном значении $N(x,t)$ найдем распределение вероятностей значений численности $N(x-\tau, t-\tau)$ группы лиц возраста $x-\tau$ в году $t-\tau$.

По формуле Байеса [5] можно записать

$$P\{N(x-\tau, t-\tau) = m / N(x,t) = n\} = \frac{P\{N(x,t) = n / N(x-\tau, t-\tau) = m\} P\{N(x-\tau, t-\tau) = m\}}{\sum_{v=n}^{\infty} P\{N(x,t) = n / N(x-\tau, t-\tau) = v\} P\{N(x-\tau, t-\tau) = v\}}. \quad (6)$$

Здесь, аналогично (1)

$$P\{N(x,t) = n / N(x-\tau, t-\tau) = m\} = C_m^n p(x-\tau, x)^n (1-p(x-\tau, x))^{m-n}, \quad (7)$$

где $p(x-\tau, x) = S(x)/S(x-\tau)$. Априорное распределение $P\{N(x-\tau, t-\tau) = m\}$ будем полагать пуассоновским с некоторым параметром a , значение которого определим ниже:

$$P\{N(x-\tau, t-\tau) = m\} = \frac{a^m}{m!} e^{-a}. \quad (8)$$

Рассмотрим сумму

$$\psi(z) = \sum_{v=n}^{\infty} z^v P\{N(x,t) = n / N(x-\tau, t-\tau) = v\} P\{N(x-\tau, t-\tau) = v\}.$$

Для краткости обозначим $p(x, x+\tau) = p$. В силу равенств (7) и (8) функцию $\psi(z)$ запишем в виде

$$\begin{aligned} \psi(z) &= \sum_{v=n}^{\infty} z^v C_v^n p^n (1-p)^{v-n} \frac{a^v}{v!} e^{-a} = p^n e^{-a} \sum_{v=n}^{\infty} z^v \frac{v!}{n!(v-n)!} (1-p)^{v-n} \frac{a^v}{v!} = \\ &= \frac{p^n e^{-a}}{n!} \sum_{v=n}^{\infty} z^v \frac{1}{(v-n)!} (1-p)^{v-n} a^v = \frac{(apz)^n}{n!} e^{-a} \sum_{v=n}^{\infty} \frac{z^{v-n}}{(v-n)!} [a(1-p)]^{v-n} = \\ &= z^n \frac{(ap)^n}{n!} e^{-a} e^{az(1-p)} = z^n \frac{(ap)^n}{n!} \exp\{a[(1-p)z-1]\}. \end{aligned}$$

Производящая функция $\varphi(z)$ распределения (6) имеет вид

$$\begin{aligned}\varphi(z) &= \sum_{m=n}^{\infty} z^m P\{N(x-\tau, t-\tau) = m / N(x, t) = n\} = \\ &= \frac{\Psi(z)}{\Psi(1)} = z^n \exp\{(z-1)a(1-p)\}.\end{aligned}\quad (9)$$

Таким образом, распределение (6) является сверткой вырожденного распределения детерминированной величины n и пуассоновского распределения с параметром

$$\lambda = a(1-p) = a\{1-p(x-\tau, x)\}.\quad (10)$$

Найдем апостериорное среднее значение величины $N(x-\tau, t-\tau)$. Очевидно, можно записать

$$MN(x-\tau, t-\tau) = n + a\{1-p(x-\tau, x)\}.$$

Полагая, что априорное и апостериорное средние значения совпадают, запишем равенство

$$a = n + a\{1-p(x-\tau, x)\},$$

из которого найдем значения параметра a в виде

$$a = n / p(x-\tau, x) = n \frac{S(x-\tau)}{S(x)}.\quad (11)$$

Таким образом, распределение (6) определяется производящей функцией (9) с параметром a вида (11). Найдем условные математическое ожидание и дисперсию величины $N(x-\tau, t-\tau)$ при условии, что выполняется равенство $n=N(x, t)$. Очевидно, что выполняется равенство

$$MN(x-\tau, t-\tau) = a = N(x, t) \frac{S(x-\tau)}{S(x)},\quad (12)$$

которое оправдывает выбор оценки в виде (5).

Найдем условную дисперсию величины оценки (5) при условии, что $n=N(x, t)$. В силу равенства (9)

$$\begin{aligned}DN(x-\tau, t-\tau) &= a\{1-p(x-\tau, x)\} = \\ &= N(x, t) \frac{S(x-\tau)}{S(x)} \left(1 - \frac{S(x)}{S(x-\tau)}\right) = N(x, t) \left(\frac{S(x-\tau)}{S(x)} - 1\right)\end{aligned}$$

а коэффициент вариации V_2 составляет величину

$$V_2 = \frac{\sqrt{DN(x-\tau, t-\tau)}}{MN(x-\tau, t-\tau)} = \frac{1}{\sqrt{N(x, t)}} \sqrt{\frac{S(x)}{S(x-\tau)} \left(1 - \frac{S(x)}{S(x-\tau)}\right)}.$$

Здесь аналогично V_1 определим диапазон значений коэффициента вариации V_2 . Первый множитель $1/\sqrt{N(x, t)}$ имеет величину менее 10^{-3} . Аналогично V_1 , проанализировав все возможные значения второго множителя для $\tau \in [1; 45]$ лет и $x \leq 70$ лет, получаем, что второй множитель максимальное значение 0,489 при-

нимает при $\tau = 45$. В результате имеем, что коэффициент вариации V_2 имеет значения менее 10^{-3} для любых τ и x . Заметим, что в смысле значений коэффициентов вариации, оценки, полученные обратной передвижкой, на порядок (в 10 раз) точнее оценок, полученных прямой передвижкой при одинаковом горизонте прогнозирования t . Поэтому случайной ошибкой ε_4 здесь также можно пренебречь.

Применим рассмотренный метод передвижки по возрастам к решению поставленной задачи определения величины людских потерь РФ в годы ВОВ.

3. Определение величины людских потерь РФ в годы ВОВ

Для решения поставленной задачи кроме статистических данных [6] распределения населения РФ по основным пятилетним возрастным группам в 1939 и 1959 гг. необходимо знать значения коэффициентов дожития. Такую информацию можно получить по результатам переписи населения РФ до 1939 г., либо после 1959 г. В данной работе воспользуемся статистическими данными 1979 г. (табл. 1).

Таблица 1

Статистические данные распределения численности населения РФ 1939 и 1959 гг. (тыс. чел.)

№ гр.: возраст	1939	1959	1979
№ 1: 0-4	13806	13353	10523
№ 2: 5-9	11735	12415	9707
№ 3: 10-14	14158	8502	9512
№ 4: 15-19	9495	8975	12385
№ 5: 20-24	8744	11552	12995
№ 6: 25-29	10454	10591	11902
№ 7: 30-34	8820	11103	8016
№ 8: 35-39	7240	6423	8399
№ 9: 40-44	5315	6177	10485
№ 10: 45-49	4268	7167	9376
№ 11: 50-54	3710	5965	9716
№ 12: 55-59	3332	4751	5595
№ 13: 60-64	2775	3589	5065
№ 14: 65-69	2079	2664	5493
№ 15: ≥ 70	2426	4303	8200

По данным 1959 и 1979 гг. найдем оценки коэффициентов дожития лиц n -й группы до возраста лиц $(n+4)$ -й группы

$$p(n, n+4) = \frac{N(n+4, 1979)}{N(n, 1959)}$$

как отношение численности $N(n+4, 1979)$ $(n+4)$ -й группы населения в 1979 г. к численности $N(n, 1959)$ n -й группы населения 1959 г., состоящих из одних и тех же лиц Российской Федерации (табл. 2).

Последнее значение 0,3855 в первой строке получено как отношение численности 8200 последней (15)-й группы 1979 г. к суммарной численности 21272

групп с 11-й по 15-ю 1959 г. Коэффициенты $p(n, n+4)$ применяются в методе прямой передвижки, а $p(n-4, n)$ соответственно в методе обратной передвижки.

Таблица 2

Оценки коэффициентов дожития

№ гр.	$p(n, n+4)$	$p(n-4, n)$
№ 1	0,9732	
№ 2	0,9587	
№ 3	0,9428	
№ 4	0,9358	
№ 5	0,9076	0,9732
№ 6	0,8853	0,9587
№ 7	0,8751	0,9428
№ 8	0,8711	0,9358
№ 9	0,8200	0,9076
№ 10	0,7664	0,8853
№ 11	0,3855	0,8751
№ 12		0,8711
№ 13		0,8200
№ 14		0,7664
№ 15		0,3855

3.1. Определение людских потерь методом прямой передвижки

Применяя значения коэффициентов дожития, приведенные в первой строке табл. 2, методом прямой передвижки, по формуле (4), найдем значения оценок численностей возрастных групп на 1959 г. по данным 1939 г. Эти значения приведены в табл. 3.

Таблица 3

Значения оценок численностей возрастных групп на 1959 г., полученных методом прямой передвижки

№ гр.	1939	$p(n, n+4)$	Передвижка на 1959	1959	Разность
№ 1	13806	0,9732			
№ 2	11735	0,9587			
№ 3	14158	0,9428			
№ 4	9495	0,9358			
№ 5	8744	0,9076	13436	11552	1884
№ 6	10454	0,8853	11250	10591	659
№ 7	8820	0,8751	13349	11103	2246
№ 8	7240	0,8711	8886	6423	2463
№ 9	5315	0,8200	7936	6177	1753
№ 10	4268	0,7664	9255	7167	2088
№ 11	14322	0,3855	7718	5965	1753
№ 12			6307	4751	1556
№ 13			4358	3589	769
№ 14			3271	2664	607
№ 15			5521	4303	1218

В последней строке этой таблицы приведены значения разностей прогнозных, полученных методом прямой передвижки, и фактических значений численностей возрастных групп 1959 г. Эти значения можно интерпретировать как величину людских потерь РФ в годы ВОВ. Их суммарное значение составляет $S_1 = 17001$ тыс.

3.2 Определение людских потерь методом обратной передвижки

Применяя значения коэффициентов дожития, приведенные во второй строке табл. 2, методом обратной передвижки по формуле (5), найдем значения оценок численностей возрастных групп на 1939 г. по данным 1959 г. Результаты приведены в табл. 4.

Таблица 4

**Значения оценок численностей возрастных групп на 1959 г.,
полученных методом обратной передвижки**

№ гр.	1959	$p(n-4, n)$	Передвижка на 1939	1939	Разность
№ 1			11870	13806	1936
№ 2			11047	11735	688
№ 3			11776	14158	2382
№ 4			6863	9495	2632
№ 5	11552	0,9732	6806	8744	1938
№ 6	10591	0,9587	8096	10454	2358
№ 7	11103	0,9428	6817	8820	2003
№ 8	6423	0,9358	5454	7240	1786
№ 9	6177	0,9076	4377	5315	938
№ 10	7167	0,8853	3476	4268	792
№ 11	5965	0,8751	11163	14322	3159
№ 12	4751	0,8711			
№ 13	3589	0,8200			
№ 14	2664	0,7664			
№ 15	4303	0,3855			

Здесь значения последней строки получено вычитанием из фактической численности возрастных групп 1939 г. значений, полученных обратной передвижкой. Эти разности можно интерпретировать как количество тех лиц данной возрастной группы, которые оказались исключенными из процесса формирования населения 1959 г. Суммарная величина S_2 этих разностей по всем возрастным группам составляет $S_2 = 20612$ тыс. Эта сумма складывается из потерь по естественным причинам, а также людских потерь, обусловленных ВОВ, значения которых по возрастным группам определяется произведением значений последней строки табл. 4 на коэффициенты $p(n, n+4)$. Эти значения приведены в табл. 5.

Значения последней строки этой таблицы совпадают со значением последней строки табл. 3, следовательно, значение людских потерь РФ в годы ВОВ составляет 17 млн человек.

Таблица 5

Значения людских потерь, обусловленных ВОВ

№ гр.		$p(n, n+4)$	
№ 1	1936	0,9732	
№ 2	688	0,9587	
№ 3	2382	0,9428	
№ 4	2632	0,9358	
№ 5	1938	0,9076	1884
№ 6	2358	0,8853	659
№ 7	2003	0,8751	2246
№ 8	1786	0,8711	2463
№ 9	938	0,8200	1759
№ 10	792	0,7664	2088
№ 11			1753
№ 12			1556
№ 13	3159	0,3855	769
№ 14			607
№ 15			1218

Заключение

Метод прямой и обратной передвижки возрастов может применяться для расчета численности возрастных групп населения в годы между датами переписи. Данный метод является простым инструментом при демографическом анализе и дает результаты, вполне адекватные реальности.

ЛИТЕРАТУРА

1. *Whelpton P.K.* Population of the United States, 1925 to 1975 // *The American Journal of Sociology*. 1928. V. 34. N 2. P. 253 – 270.
2. *Демографический энциклопедический словарь* / под ред. Д.И. Валентя. М.: Советская энциклопедия, 1985. 608 с.
3. *Медков В.М.* Демография. М.: ИНФРА-М, 2007. 683 с.
4. *Фалин Г.И., Фалин А.И.* Введение в актуарную математику. М.: Изд-во МГУ, 1994. 86 с.
5. *Назаров А.А., Терпугов А.Ф.* Теория вероятностей и случайных процессов: Учебное пособие. Томск: Изд-во НТЛ, 2006. 204 с.
6. *Российский статистический ежегодник: Статистический сборник: Официальное издание / Гос. комитет РФ по статистике (Госкомстат России) / под ред. В.Л. Соколина и др. М.: 2001. 642 с.*

Назаров Анатолий Андреевич

Носова Мария Геннадьевна

Томский государственный университет

E- mail: nazarov@fpmk.tsu.ru; nosova_mg@mail.ru

Поступила в редакцию 11 мая 2009 г.