



**Никола Иванов Чолаков** е роден през 1944 г. в София. През 1971 г. завършва СУ „Св. Климент Охридски“, специалност Математика. В УНСС е на работа от 1977 г., доктор по икономика от 1996 г.

Специализирал е в Московския държавен университет „Ломоносов“ (1978 г.) и в Международната организация по труда – Женева (1980 г.).

Бил е консултант на ООН по демография (1984 и 1988 г.), пълномощен министър в постоянното представителство на Република България в ООН в Ню Йорк (1997 г.) и пълномощен министър в посолството на Република България в САЩ (1998 – 2001 г.).

Основни публикации: „Демографските процеси и населението на България“ (1972 г.), „Прогноза за населението на НРБ по окръзи (градове и села) през периода 1970 – 2000 година“ (1975 г.), „Демографическа ситуация, принципи, проблеми и меры демографической политики НР Болгарии“ (1985 г.), „Population ageing and its social and economic consequences“, Economic Studies No. 3, United Nations, Geneva, 1992, „Методологически и методически проблеми при статистическото изучаване на смъртността“ (1992 г.), „Математически методи и модели в животозастраховането и пенсионното осигуряване (2003 г.), „Статистическата информация за естествено-то движение на населението и миграцията в България и някои насоки за нейното усъвършенстване“, сб. „Научни трудове 2003“ № 2, VII „Стопанство“, УНСС и гр.

## **СЛУЧАЙНИ ОТКЛОНЕНИЯ И СИСТЕМАТИЧНИ ГРЕШКИ И ВЪЗДЕЙСТВИЕТО ИМ ВЪРХУ ПОКАЗАТЕЛИТЕ НА ТАБЛИЦИТЕ ЗА СМЪРТНОСТ**

гл.ас. д-р Никола Чолаков

С развитието на статистическите методи след Втората световна война в посока на емпиричните социологически проучвания се забелязва определен интерес и към таблиците за смъртност като репрезентативна картина на възрастовия профил на риска за умирање на дадено население към определен момент. Редица автори (Chiang 1960, 1961, 1968, Keyfitz 1968, Ноем 1969, 1976, Feichtinger 1972) разработват и развиват модела на стареенето на условното поколение от таблиците на смъртност като верига от събития в динамични съвкупности и извличат основните показатели на този стохастичен процес. В студията се изхожда от тези резултати и се търсят решения в по-широк план за въздействието върху показателите на таблиците за смъртност, което се определя не само от стоха-

стиката в стареенето на условното поколение и въобще на демографските процеси, но и от разнообразни външни за демографията фактори.

Във връзка с това уместно би било да се привлече вниманието върху факта, че движението на демографската статистическа информация у нас изглежда излишно удължено и усложнено – от съобщенията за смърт, през различните общински, регионални и централни административни звена (регионалните центрове по здравеопазване, териториалните статистически бюра, териториалните звена на ГРАО), гражданския регистър за административно обслужване до националния статистически орган. Подобно прехвърляне на информацията на различни носители и през различни ведомства би могло да се окаже предпоставка за пропуски и допускане на грешки и неточности от разнообразно естество, което в крайна сметка би могло да се отрази по някакъв начин и на най-аналитичните показатели като тези на таблиците за смъртност. Симптоматичен факт е, че вече три години след преброяването на населението у нас (март 2001 г.) държавната статистика се въздържа от публикуване на окончателни данни за основните възрастово-полови структури към критичния момент на наблюдението. Не бива да се подценява и рискът от невнимателно, неточно или неправилно прилагане на съответните методики за обработка на информацията и за получаване на различните показатели.

В хода на изследването сме показали как някои междинни резултати биха могли да се използват при сравнителния анализ на смъртността по данни от таблици за смъртност с отчитане на вариацията във величината на показателите, породена от различните фактори.

В известна степен тази разработка е продължение на работата ни по усъвършенстване на статистическата информация за населението в България (Чолаков, Малешков, 2003 г.).

## 1. ТАБЛИЦИ ЗА СМЪРТНОСТ

Таблиците за смъртност (*life table*, среща се и *mortality table*) представят условни, хипотетични поколения, водещи начало от примерно 100 000 живородени, и разкриват с помощта на редица специфични показатели как се формира продължителността на живота в хода на напредването във възрастта и как се стопява това поколение в условията на определен възрастов профил на смъртността. Профилът на смъртността обикновено съответства на фактическите данни за смъртността на определено население през определен период от една-две до три години. По такъв начин с условното поколение се имитира фактическата смъртност във възрастовия профил, но в чист вид, независимо от възрастовата структура на фактическото население. Населенията в Скандинавия например бележат сравнително висок годишен брой умрели не защото смъртността там е висока, напротив, там смъртността е една от най-ниските в света, а защото тези населения са застарели, делът на възрастните е сравнително висок и най-рисковите съвкупности са сравнително многобройни. Повечето населения в Африка представят обратния пример. Там делът на младите доминира, в резултат на което годишният брой на умрелите е сравнително нисък, въпреки че смъртността е висока и средната продължителност на живота е доста по-ниска в сравнение с Европа.

Предвид това таблиците за смъртност обикновено се получават по данни за смъртността през отделна година (или две-три години поред), като смъртността при условното поколение фактически се сглобява от възрастовата смъртност на всичките поколения през този период. Такива таблици обикновено се наричат *периодни* или *трансверзални*, тъй като представят възрастовия профил на смъртността за определен период. Срещат се макар и рядко и *кохортни* таблици за смъртност, които представят едно действително поколение с еднородна и минимална условност: стареенето или напредването във възрастта е представено на скала с начало 100 000 живородени. Трудността при изработването на подобни таблици се състои в това, че трябва текущо да се събират данни за това поколение в продължение на сто и повече години. Кохортни таблици за смъртност обикновено се пресмятат по данни от църковните регистри и имат предимно историческа стойност.

Освен за отделни населености таблици за смъртност като правило се изготвят отделно за двата пола, за да се отрази спецификата на смъртността при мъжете и жените. Най-напред детската смъртност при момчетата е по-висока, отколкото при момичетата. При следващите възрасти тези различия намаляват и чак след възраст 30 – 40 години отново излизат наяве. Смъртността при мъжете в по-старша възраст е по-висока в сравнение с жените и тази разлика се задълбочава с напредването във възрастта. При българското население средната продължителност на живота на мъжете в края на XX век е около 67 години, докато при жените е към 75 години. При това тази разлика показва тенденция към непрекъснато увеличение през последните десетилетия. В исторически план обаче дисбаланс между двата пола трудно би се получил, тъй като от 1000 живородени средно 515 са момчета и така се компенсират по-високата смъртност при мъжете.

Срещат се и таблици за смъртност за отделни региони или групи от населението по национален, етнически, религиозен, расов или друг признак. В застраховането често се използват таблици за смъртност, извлечени по данни само за застрахованите или за отделни рискови групи (*select-and-ultimate life tables*). Важно място заемат таблиците за смъртност по различни здравословни признаци в зависимост от епидемиологичния профил на населението, както и таблици, при които диференцирано се отразяват различните причини за смъртта.

Може да изглежда изненадващо, но идеята за таблиците за смъртност се е зародила в дълбока древност. През III в. римският адвокат Улпиан е изработил таблица, в една от колоните на която се намира показател, поразително съвпадащ със средната продължителност на предстоящия живот за епохата. Някои демографи и историци са склонни да виждат в числата дори стойността на застраховка „Живот“. Последното обаче е доста съмнително, тъй като по онова време понятието за лихвата е било съвсем в ембрионален период. Остава хипотезата, че този показател е действително средната продължителност на предстоящия живот, което е съвсем правдоподобно. Има факти, които подсказват, че още при император Юстиниан данните на Улпиан са били използвани в гражданското право за уреждане на наследствата (*Lex Falcidia*). Таблицата на Улпиан (с някои по-късни допълнения) е била използвана за застрахователни цели в Северна Италия до XIX век (Keyfitz, 1978).

Много плътно до идеята за показателите на таблиците за смъртност се доближава Джон Граунт при изучаването на смъртността в Лондон (1662) след го-

лямата чума, а също така и бащата на „политическата аритметика“ Уилям Петти. Но като първа по-пълна таблица за смъртност се счита тази на астронома Едмунд Халей за смъртността сред населението на Бреслау (сега Вроцлав), публикувана през 1693 г. В таблицата се дават и застрахователно-техническите величини, в т.ч. стойността на пожизнени ренти, както и правила за ползването на различните показатели. През 1776 г. Пер Варгентин публикува първата таблица за смъртност за цяло население, това на Швеция, като използва данни (за периода 1755-1757 г.) от регистъра на населението.

По-съвременна трактовка на тази проблематика намираме у Джоша Майлс (1815), при който има по-строга систематизация на показателите на таблиците за смъртност, но в най-точен и завършен вид този метод се открива при Джордж Кинг (1902), от когото произлиза и съвременното означаване на показателите. През XX век значително се развиват методите за получаване на таблици за смъртност с отчитане на качеството на изходния статистически материал. Многочислени приложения на метода има и в други предметни области.

В търсенето на систематика на научното познание за възрастовата смъртност не липсват и съответни математически модели. Като класически в това направление се считат работите на Бенджамен Гомперц (Gompertz, 1825) и Уилиям Мейкхем (Makeham, 1860). Много интересна е една хипотеза на известния от по-далечното минало лекар и натурфилософ Джироламо Кардано<sup>1</sup>, лансирана от него през 1570 г., съгласно която средната продължителност на предстоящия живот<sup>2</sup> прогресивно и равномерно спада с напредването на живите във възрастта. Равномерността на такова намаление е повече от спорна, но идеята да се търси модел на средната продължителност на предстоящия живот може да се окаже интересна (Чолаков, 2003а).

Използването на математически модели за изучаването на смъртността позволява на Лауел Рийд и Маргарет Меръл (Reed, Merrell, 1939), както и на Томас Гревил (Greville, 1943) да разработят методи за получаване на таблици за смъртност от непълна или ненадеждна статистическа информация. Добре известни са още и таблиците за смъртност на Енсли Коул и Пол Дементи (1966), в основата на които се намират четири математически модела за основните региони на света. Всички приложения на метода на таблиците за смъртност в други области – примерно за изучаването на трудовата мобилност (Tabach, 1968), са характерни със съответен математически модел. Важна роля за развитието на тази методология имат и статистическата служба и демографското отделение на ООН и специализирани институти на Европейския съюз.

<sup>1</sup> Познат повече като откривател на решението на кубичните уравнения. Джироламо Кардано (1501-1576) е бил предимно практикуващ лекар, непоколебим привърженик на астрологията и тълкувател на сънища. Имал е сериозни проблеми с инквизицията заради направения хороскоп на Христос. Негови са произведенията „Practica arithmetica“ (1539), „Artis magnaе sive de regulis algebraicis liber unus“ (1545), както и автобиографичната „De vita propria liber“ (1560).

<sup>2</sup> Този показател е основен за потенциалната демография, разглеждаща продължителността на живота като жизнен потенциал, като ресурс, който постепенно се изчерпва с напредването във възрастта (Dublin, Lotka, 1936).

До Втората световна война таблици за смъртност за българското население са изготвяни от БАН по данни за смъртността и възрастовата структура на живите от преброяванията. В по-ново време Националният статистически институт периодично публикува таблици за смъртност в „Статистически годишник на Р България“ и в демографския бюлетин „Население“, както и в някои специализирани издания на института, където е дадена и използваната методология.

Рефлексията в методологията на таблиците за смъртност е сравнително ново направление в демографията и се появява през втората половина на XX век с работите на Chiang (1960, 1961, 1968, 1978), Keyfitz (1968), Ноем (1969, 1976), Feichtinger (1972) и др. Основната идея е с помощта на вече изградените модели или поне в техните рамки да се оцени точността на получените резултати и изводи, като се отчита естествената вариация в движението на наблюдаваните от статистиката величини, както и многообразието от външни за процеса фактори, които биха могли да повлияят на адекватността на демографската информация за смъртността в една или друга посока.

## 2. ПОДХОДИ КЪМ ОПРЕДЕЛЯНЕТО НА ПОКАЗАТЕЛИТЕ НА ТАБЛИЦИТЕ ЗА СМЪРТНОСТ

Таблиците за смъртност представляват система от взаимно свързани показатели (често наричани **биометрични функции**), разкриващи отделни аспекти в стопяването на едно поколение в хода на неговото стареене. В началото размерът на поколението се избира госта условно като  $l_0 = 100,000$  живородени<sup>1</sup> и в съответната колона на таблицата по отделните редове се показва колко от тази съвкупност стигат до поредните възрасти. Този показател представлява **първа главна съвкупност на живите** и се бележи с  $l_x$  за отделните възрасти  $x$ . Размерът на отпадналите в хода на напредването във възрастта от  $x$  до  $x+n$  представлява **първа главна съвкупност на умрелите**  ${}_n d_x$  и по построение за нея важи:  ${}_n d_x = l_x - l_{x+n}$ , което означава, че с прехода от възраст  $x$  до  $x+n$  съвкупността  $l_x$  се разпада на две части: една част –  $l_{x+n}$ , стигат до следващата възраст  $x+n$ , а останалите представляват умрелите –  ${}_n d_x$ . В таблиците за смъртност обикновено не се предвижда друга възможност за отпадане (като миграция например) освен смърт, в резултат на което е изпълнено  $l_x = l_{x+n} + {}_n d_x$ . Много често паралелно с показателите  $l_x$  и  ${}_n d_x$  в таблиците за смъртност се разполагат в отделни колони и вероятностите за тези две възможности:

$$(1) \quad {}_n p_x = \frac{l_{x+n}}{l_x} \text{ и } {}_n q_x = 1 - {}_n p_x = \frac{{}_n d_x}{l_x}.$$

<sup>1</sup> Кое то показва, че тези и подобни показатели на декрементните таблици трябва да се схващат повече като относителни величини.

Тези релации лесно се откриват в табл. 1.

Тук е мястото да се отбележи, че с левия голен индекс  $n$  се означава ширината на възрастовия интервал, както това е възприето в демографията и актюерството преди повече от сто години. Това означение обикновено се изпуска при единичните възрастови интервали, т.е. когато  $n=1$ . При съкратено представяне на таблици за смъртност, каквато е табл. 1, единични възрастови интервали се използват само за детските възрасти с оглед на това по-подробно да се представят особеностите при детската смъртност, докато за следващите възрасти данните обикновено се представят на по-широки възрастови интервали.

**Таблица 1**

Таблица за смъртност на българското население – мъже, 1995-1997 г.

| $x$ | $l_x$   | ${}_n d_x$ | ${}_n q_x$ | ${}_n p_x$ | ${}_n L_x$ | ${}_n a_x$ | $T_x$     | $e_x$  |
|-----|---------|------------|------------|------------|------------|------------|-----------|--------|
| 0   | 100 000 | 1 771      | 0.01771    | 0.98229    | 98 620     | 0.221      | 6 708 290 | 67.083 |
| 1   | 98 229  | 439        | 0.00447    | 0.99553    | 391 832    | 1.530      | 6 609 670 | 67.288 |
| 5   | 97 790  | 236        | 0.00241    | 0.99759    | 488 339    | 2.409      | 6 217 838 | 63.584 |
| 10  | 97 554  | 213        | 0.00218    | 0.99782    | 487 273    | 2.667      | 5 729 499 | 58.732 |
| 15  | 97 341  | 435        | 0.00447    | 0.99553    | 485 713    | 2.720      | 5 242 226 | 53.854 |
| 20  | 96 906  | 626        | 0.00646    | 0.99354    | 483 027    | 2.599      | 4 756 513 | 49.084 |
| 25  | 96 280  | 696        | 0.00723    | 0.99277    | 479 707    | 2.568      | 4 273 486 | 44.386 |
| 30  | 95 584  | 952        | 0.00996    | 0.99004    | 475 708    | 2.676      | 3 793 779 | 39.691 |
| 35  | 94 632  | 1 516      | 0.01602    | 0.98398    | 469 661    | 2.692      | 3 318 071 | 35.063 |
| 40  | 93 116  | 2 514      | 0.02700    | 0.97300    | 459 768    | 2.688      | 2 848 410 | 30.590 |
| 45  | 90 602  | 3 857      | 0.04257    | 0.95743    | 443 954    | 2.652      | 2 388 642 | 26.364 |
| 50  | 86 745  | 5 495      | 0.06335    | 0.93665    | 420 751    | 2.639      | 1 944 689 | 22.418 |
| 55  | 81 250  | 7 498      | 0.09228    | 0.90772    | 388 517    | 2.635      | 1 523 937 | 18.756 |
| 60  | 73 752  | 9 691      | 0.13140    | 0.86860    | 345 211    | 2.570      | 1 135 420 | 15.395 |
| 65  | 64 061  | 11 657     | 0.18197    | 0.81803    | 292 363    | 2.603      | 790 209   | 12.335 |
| 70  | 52 404  | 14 168     | 0.27036    | 0.72964    | 227 153    | 2.539      | 497 846   | 9.500  |
| 75  | 38 236  | 15 552     | 0.40674    | 0.59326    | 152 471    | 2.511      | 270 694   | 7.080  |
| 80  | 22 684  | 12 665     | 0.55832    | 0.44168    | 81 162     | 2.453      | 118 223   | 5.212  |
| 85  | 10 019  | 10 019     | 1.00000    | 0.00000    | 37 060     | 3.699      | 37 060    | 3.699  |

**Забележка:** по данни от „Население“, НСИ, София, 1998 с някои допълнения от автора.

Следващата група показатели е свързана с **втора главна съвкупност на живите**  ${}_n L_x$ , или т.нар. **стационарно население**. Величината на този показател разкрива колко човекогодини притежават лицата от първа главна съвкупност на жи-

вите  $l_x$  при преминаването им от възраст  $x$  до  $x+n$ . Този фонд се образува от преживелите този възрастов интервал  $l_{x+n}$ , всеки от които допринася по  $n$  години плюс човекогодините, идващи от умрелите  ${}_n d_x$ , чийто размер средно на лице се дава с показателя<sup>1</sup>  ${}_n a_x$ . Последното показва, че

$${}_n L_x = n \cdot l_{x+n} + {}_n a_x \cdot {}_n d_x.$$

Общият фонд предстоящи човекогодини за лицата  $l_x$  без оглед на достигане или не на следващите възрасти се дава с показателя

$$T_x = \sum_{t \geq 0} {}_n L_{x+t},$$

представляващ кумулативните сборове на числата от колоната  ${}_n L_x$  в посока на по-високите възрасти. Привеждането на този фонд средно на лице дава **предстоящата продължителност на живота** като средна величина

$$e_x = \frac{T_x}{l_x}.$$

По-съвременното виждане (Ноев, 2002) за показателите на таблиците за смъртност е свързано със схващането за продължителността на живота като варираща (случайна) величина  $X$  с определена функция на разпределение  $F(x)$ , отразяваща възрастовия профил на смъртността

$$P\{X < x\} = F(x).$$

Тъй като продължителността на живота е неотрицателна величина, често се оказва по-удобно да се работи с допълнителната вероятност

$$1 - F(x) = P\{X \geq x\},$$

представляваща шанса да се стигне до възраст  $x$ .

Структурата на таблиците за смъртност и верижният характер на връзката между показателите за последователните възрасти насочват към разглеждане на условната вероятност за достигане на възраст  $x+n$ , при условие че вече е достигната  $x$ ,

<sup>1</sup> Този показател е въведен в демографията от Чанг (Chiang, 1968). Може само да се съжалева, че НСИ не използват този показател дори и само за верификация на публикуваните таблици за смъртност.

$$(2) \quad P\{X \geq x+n \mid X \geq x\} = \frac{P\{X \geq x+n\}}{P\{X \geq x\}} = {}_n p_x.$$

В последната формула може лесно да се открие ролята на  ${}_n p_x$  от (1), разкриваща тясната връзка между разпределението на продължителността на живота като случайна величина и първа главна съвкупност на живите и потвърждаваща тезата, че показателят  $l_x$  трябва да се схваща повече като относителна величина, информативно еквивалентна на вероятностите от (2).

Важно концептуално значение при таблиците за смъртност има **интензитетът на смъртта**<sup>1</sup>

$$(3) \quad \mu(x) = -\frac{1}{P\{X \geq x\}} \frac{d}{dx} P\{X \geq x\} \\ = -\frac{d}{dx} \ln P\{X \geq x\},$$

като показател за риска за точните възрасти. Обратната на формула (3) в известен смисъл показва как става акумулирането на риска за отделните възрастови интервали:

$$\exp\left(-\int_0^n \mu(x+t) dt\right) = \frac{P\{X \geq x+n\}}{P\{X \geq x\}} = \frac{l_{x+n}}{l_x} = {}_n p_x,$$

което представлява условната вероятност за достигане на възраст  $x+n$ , при условие че вече е достигната  $x$  (вж. (2)).

Както се вижда, разпределението на продължителността на живота и заедно с това и интензитетът на смъртта са в основата на таблиците за смъртност и представляват важен методологичен инструмент за получаването на техните показатели.

### 3. НАПРЕДВАНЕТО ВЪВ ВЪЗРАСТТА КАТО СЛУЧАЕН ПРОЦЕС

Старееенето на условното поколение от таблиците за смъртност се съпровожда с непрекъснато отпадане на умрелите и придвижване на останалите в следващата възрастова група. Така например още съвсем в началото на процеса от първо-

<sup>1</sup> Въведен в демографията от Смит (Smith, 1948), но познат в естествените науки още от времето на Нютон, откъдето е дошъл и съответният термин на английски: *force of mortality*.



началните  $l_0 = 100,000$  живородени стигащите до възраст 1 представляват случаен брой, който би варираше около средната си стойност  $l_1$ . Обикновено се счита, че продължителностите на живота на единиците от условното поколение протичат независимо една от друга<sup>1</sup>, в резултат на което достигащите възраст 1 формират съвкупност със случаен размер  $K_1$ , имащ биномно разпределение

$$(4) \quad P\{K_1 = k | l_0\} = \binom{l_0}{k} p_0^k (1 - p_0)^{l_0 - k}.$$

Математическото очакване на случайната величина  $K_1$  е

$$(5) \quad E(K_1) = l_0 p_0,$$

което, като се има предвид (2), кореспондира на показателя  $l_1$  от таблиците за смъртност като средна стойност, като очакван брой стигащи до възраст 1. Случайният брой  $K_1$  фактически варира около математическото си очакване и големината на вариацията може да се представи с дисперсията<sup>2</sup>

$$(6) \quad \text{Var}(K_1) = l_0 p_0 (1 - p_0)$$

или със стандартното отклонение

$$(7) \quad \sqrt{\text{Var}(K_1)} = \sqrt{l_0 p_0 (1 - p_0)}.$$

Тъй като вероятността да се стигне до първия рожден ген (т.е. до възраст 1) кореспондира на съответната от таблицата за смъртност

$$p_0 = \frac{l_1}{l_0},$$

може да се получи количествена оценка на показателите за разпределението на случайната величина  $K_1$ . Формула (5) придобива вида

$$(8) \quad E(K_1) = l_1,$$

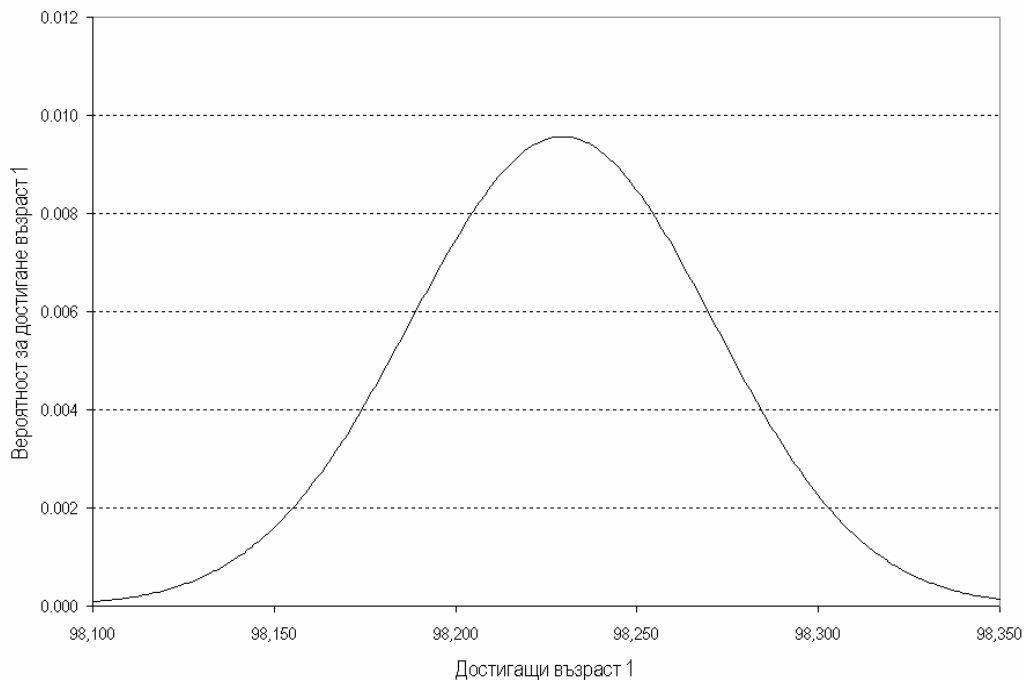
<sup>1</sup> Подобна хипотеза е обсъждана от много изследователи – вж. Пресмон (Preston, S., N. Keyfitz, R. Schoen, 1972).

<sup>2</sup> За този показател ще използваме разпространеното означение  $\text{Var}(\cdot)$ , като в скоби се посочва вариращата величина.

респ. (6) –  $v_{uda}$

$$(9) \quad Var(K_1) = l_0 \frac{l_1}{l_0} \frac{l_0 - l_1}{l_0} = \frac{l_1 d_0}{l_0} = l_1 q_0,$$

и аналогично се получава със стандартното отклонение (7).



**Фиг. 1.** Разпределение на достигащите възраст 1 от първоначална съвкупност 100 000 живородени

На фиг. 1 (по данните от табл. 1) е показано как варира броят на достигащите възраст 1, представляващ случайната величина  $K_1$ . Ясно се вижда, че би било изключително съвпадение този брой да се окаже равен на математическото си очакване  $l_1 = 98,229$  – вероятността за това е под 1%. Съседните на математическото очакване възможности са дори още по-малко вероятни, но въпреки това величината на показателя  $l_1$  като център на привличане задържа около себе си броя на достигащите възраст 1. От фигурата става ясно, че както и да варира величината  $K_1$  в мнозинството случаи би се оказала в границите от 98,150 до 98,350, което никак не е далече от математическото си очакване. Това е резултат от действието на законите за големите числа и камбановидната плътност на разпределението на фиг. 1 е израз на закономерностите, познати в статистиката като закон на Гаус-Лаплас и централни гранични теореми. В табл. 2 са дадени по-подробни числови данни за размера на вариацията в броя на доживяващите за различните възрасти на условното поколение в табл. 1.

Може да се очаква, че с напредването в следващите възрасти би се натрупала дисперсия, подобна на тази от фиг. 1, и би се получило разпиляване на възможните траектории на движение на условното поколение. Всъщност законите на големите числа фокусират процеса чрез взаимно погасяване на отклоненията в противоположните посоки и го канализират в една сравнително добре очертана тенденция в стареенето на условното поколение, с която се определя познавателната стойност на таблиците за смъртност.

Националният център по здравна статистика на САЩ използва подобен подход (US Department of Health & Human Services 1997), за да оцени доколко са лабилни показателите на таблиците за смъртност на американското население в резултат на вариацията в броя на доживяващите до поредните възрасти<sup>1</sup>. Идентичен подход е възприет и от канадската статистика (Statistics Canada, 2002) за оценка на стохастичния елемент в продължителността на живота. Така стигналите до поредните възрасти се разглеждат като случаен процес  $\{K_x, x \geq 0\}$ , представляващ верига на Марков (Чолаков, 2003b). Фактически редът  $K_x$  образува една от множеството възможни траектории на движение на условното поколение при един и същ възрастов профил на смъртността (т.е. при едно и също разпределение на продължителността на живота и една и съща таблица за смъртност). Някои от тези траектории се оказват по-вероятни и по-често срещани, други – по-малко, и интерес представлява вероятността да се случи една или друга траектория и да се определят съответните стохастични характеристики на този процес. Фиг. 1 показва как тези траектории може да се различават една от друга още в първата фаза от стареенето на условното поколение в таблиците за смъртност. Въпреки възможните отклонения в едната или другата посока процесът на стареене на условното поколение и на постепенното му стопяване все пак остава еднопосочен и подчинен на основната тенденция, представена с показателя  $I_x$ .

**Таблица 2**

Вариация в размера на първа главна съвкупност на живите

| x | $I_x$   | Дисперсия  | Стандартно отклонение | 95 %-ов интервал на вариране |            | 99 %-ов интервал на вариране |            |
|---|---------|------------|-----------------------|------------------------------|------------|------------------------------|------------|
|   |         |            |                       | ляв край                     | десен край | ляв край                     | десен край |
| 1 | 2       | 3          | 4                     | 5                            | 6          | 7                            | 8          |
| 0 | 100 000 | 0          | 0                     | 100,000                      | 100,000    | 100,000                      | 100,000    |
| 1 | 98 229  | 1 739.6356 | 41.708939             | 98 147                       | 98 311     | 98 122                       | 98 336     |
| 5 | 97 790  | 2 161.1590 | 46.488267             | 97 699                       | 97 881     | 97 670                       | 97 910     |

<sup>1</sup> Т.е. в резултат на напредването във възрастта като случаен процес, без да се включва стохастичната грешка на самите показатели в хода на тяхното получаване.

Продължение

| 1  | 2      | 3           | 4          | 5      | 6      | 7      | 8      |
|----|--------|-------------|------------|--------|--------|--------|--------|
| 10 | 97 554 | 2 386.1708  | 48.848448  | 97 458 | 97 650 | 97 428 | 97 680 |
| 15 | 97 341 | 2 588.2972  | 50.875310  | 97 241 | 97 441 | 97 210 | 97 472 |
| 20 | 96 906 | 2 998.2716  | 54.756476  | 96 799 | 97 013 | 96 765 | 97 047 |
| 25 | 96 280 | 3 581.6160  | 59.846604  | 96 163 | 96 397 | 96 126 | 96 434 |
| 30 | 95 584 | 4 220.9894  | 64.969142  | 95 457 | 95 711 | 95 417 | 95 751 |
| 35 | 94 632 | 5 079.8458  | 71.273037  | 94 492 | 94 772 | 94 448 | 94 816 |
| 40 | 93 116 | 6 410.1054  | 80.063134  | 92 959 | 93 273 | 92 910 | 93 322 |
| 45 | 90 602 | 8 514.7760  | 92.275544  | 90 421 | 90 783 | 90 364 | 90 840 |
| 50 | 86 745 | 11 498.0498 | 107.228959 | 86 535 | 86 955 | 86 469 | 87 021 |
| 55 | 81 250 | 15 234.3750 | 123.427610 | 81 008 | 81 492 | 80 932 | 81 568 |
| 60 | 73 752 | 19 358.4250 | 139.134557 | 73 479 | 74 025 | 73 394 | 74 110 |
| 65 | 64 061 | 23 022.8828 | 151.732932 | 63 764 | 64 358 | 63 670 | 64 452 |
| 70 | 52 404 | 24 942.2078 | 157.931022 | 52 094 | 52 714 | 51 997 | 52 811 |
| 75 | 38 236 | 23 616.0830 | 153.675252 | 37 935 | 38 537 | 37 840 | 38 632 |
| 80 | 22 684 | 17 538.3614 | 132.432479 | 22 424 | 22 944 | 22 343 | 23 025 |
| 85 | 10 019 | 9 015.1964  | 94.948388  | 9 833  | 10 205 | 9 774  | 10 264 |

**Забележка:** по данни от табл. 1.

Разсъжденията за вариацията в броя на достигащите възраст 1 могат да продължат нататък за всяка следваща възраст и като се използва моделът на биномното разпределение, да се получи така, че математическото очакване на този брой да е:

$$(10) \quad E(K_x) = l_x,$$

с дисперсия:

$$(11) \quad \text{Var}(K_x) = \frac{l_x(l_0 - l_x)}{l_0}.$$

На табл. 2 е показано в какви граници броят на доживяващите до поредните възрасти би могъл да варира при една и съща смъртност. Границите на интервалите на вариране (в последните четири колони на таблицата) в съответствие с (10) и (11) са получени по формулата

$$(12) \quad l_x \pm z \sqrt{\frac{l_x(l_0 - l_x)}{l_0}},$$

където  $z = 1.96$  за 95 %-те интервали и  $z = 2.576$  – за 99 %-те<sup>1</sup> интервали на вариране.

Ефектът от вариацията в броя на доживяващите може да се открие и при съответните пропорции в размера на първа главна съвкупност на живите. На табл. 3 може да се види в какви граници се движат тези пропорции около съответните вероятности  ${}_n p_x$ , представляващи техните средни величини. Задачата е малко по-комплицирана в сравнение с предходната, тъй като дисперсията на пропорциите е равна на

$$\text{Var}\left(\frac{K_{x+n}}{K_x}\right) = E\left(\frac{1}{K_x}\right) \frac{l_{x+n}}{l_x} \left(1 - \frac{l_{x+n}}{l_x}\right)$$

и за пресмятането ѝ се налага допълнително приближение като

$$E\left(\frac{1}{K_x}\right) = \frac{1}{l_x},$$

върху точността на което ще се спрем в следващата част. Като се изключи тази особеност, останалото е подобно на (12) и размахът на вариацията е получен по следната формула:

$$E\left(\frac{K_{x+n}}{K_x}\right) \pm z \sqrt{\text{Var}\left(\frac{K_{x+n}}{K_x}\right)},$$

където  $z = 1.96$  за 95 %-те интервали и  $z = 2.576$  – за 99 %-те интервали на вариране.

**Таблица 3**

Вариация в последователните пропорции на преживяващите

| x | ${}_n p_x$ | 95 %-ов интервал на вариране |            |                     | 99 %-ов интервал на вариране |            |                     |
|---|------------|------------------------------|------------|---------------------|------------------------------|------------|---------------------|
|   |            | ляв край                     | десен край | ширина на интервала | ляв край                     | десен край | ширина на интервала |
| 1 | 2          | 3                            | 4          | 5                   | 6                            | 7          | 8                   |
| 0 | 0.982290   | 0.981473                     | 0.983107   | 0.001635            | 0.981216                     | 0.983364   | 0.002149            |
| 1 | 0.995531   | 0.995114                     | 0.995948   | 0.000834            | 0.994983                     | 0.996079   | 0.001096            |
| 5 | 0.997587   | 0.997279                     | 0.997894   | 0.000615            | 0.997182                     | 0.997991   | 0.000808            |

<sup>1</sup> Представляващи съответните перцентили на нормалното разпределение като апроксимация на биномното. Поради това, че се борава със сравнително многобройни съвкупности, не се налага привличането на разпределението на Стюдънт или някоя друга точна апроксимация.

Продължение

| 1  | 2        | 3        | 4        | 5        | 6        | 7        | 8        |
|----|----------|----------|----------|----------|----------|----------|----------|
| 10 | 0.997817 | 0.997524 | 0.998109 | 0.000586 | 0.997432 | 0.998202 | 0.000770 |
| 15 | 0.995531 | 0.995112 | 0.995950 | 0.000838 | 0.994980 | 0.996082 | 0.001101 |
| 20 | 0.993540 | 0.993036 | 0.994045 | 0.001009 | 0.992877 | 0.994203 | 0.001326 |
| 25 | 0.992771 | 0.992236 | 0.993306 | 0.001070 | 0.992068 | 0.993474 | 0.001407 |
| 30 | 0.990040 | 0.989411 | 0.990670 | 0.001259 | 0.989213 | 0.990868 | 0.001655 |
| 35 | 0.983980 | 0.983180 | 0.984780 | 0.001600 | 0.982929 | 0.985031 | 0.002103 |
| 40 | 0.973001 | 0.971960 | 0.974042 | 0.002082 | 0.971633 | 0.974370 | 0.002736 |
| 45 | 0.957429 | 0.956115 | 0.958744 | 0.002629 | 0.955701 | 0.959157 | 0.003456 |
| 50 | 0.936653 | 0.935032 | 0.938274 | 0.003242 | 0.934523 | 0.938784 | 0.004261 |
| 55 | 0.907717 | 0.905727 | 0.909707 | 0.003980 | 0.905101 | 0.910333 | 0.005231 |
| 60 | 0.868600 | 0.866162 | 0.871038 | 0.004876 | 0.865396 | 0.871805 | 0.006409 |
| 65 | 0.818033 | 0.815045 | 0.821021 | 0.005975 | 0.814106 | 0.821960 | 0.007853 |
| 70 | 0.729639 | 0.725836 | 0.733442 | 0.007606 | 0.724641 | 0.734637 | 0.009996 |
| 75 | 0.593263 | 0.588339 | 0.598187 | 0.009848 | 0.586792 | 0.599734 | 0.012943 |
| 80 | 0.441677 | 0.435215 | 0.448139 | 0.012925 | 0.433184 | 0.450170 | 0.016987 |
| 85 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |

**Забележка:** По данни от табл. 1 и 2.

Вижда се, че статистически най-значими са предимно първите два знака след десетичната точка, докато при третия и особено при четвъртия информативността значително спада, а следващите са направо случайни цифри (особено при по-старшите възрасти). Това насочва мисълта към началото на таблиците за смъртност, изходния брой живородени, който би бил достатъчно статистически репрезентативен при  $l_0 = 10\ 000$  или дори 1000. Намаляването на мащаба на първа главна съвкупност на живите обаче би довело до разширяване размаха на вариацията в пропорцията на преживяващите. Така например при  $l_0 = 1000$  ширината на интервалите на вариране би се увеличила десет пъти, което означава загуба на един надежден знак в съответното десетично представяне на тези пропорции. В демографската статистика е възприето показателите относителни величини да се представят „на хиляда“, т.е. с три точни знака след десетичната точка. Така че изборът  $l_0 = 100\ 000$  изглежда подходящ, като се изключат най-старшите възрасти, при които репрезентативността малко се губи. За съжаление подобна проблематика е останала малко встрани от научните търсения около таблиците за смъртност. Изборът на начална стойност за изходния брой живородени може да изглежда безобиден, но, както ще се покаже по-нататък, този избор може да има и по-далечни импликации за смисъла на показателите от таблиците за смъртност.

**Таблица 4**

Вариація в реда на умрелите

| $x$ | $l_x$   | ${}_n d_x$ | Дисперсия   | Стандартно отклонение |
|-----|---------|------------|-------------|-----------------------|
| 0   | 100 000 | 1 771      | 1 739.6356  | 41.71                 |
| 1   | 98 229  | 439        | 437.0728    | 20.91                 |
| 5   | 97 790  | 236        | 235.4430    | 15.34                 |
| 10  | 97 554  | 213        | 212.5463    | 14.58                 |
| 15  | 97 341  | 435        | 433.1078    | 20.81                 |
| 20  | 96 906  | 626        | 622.0812    | 24.94                 |
| 25  | 96 280  | 696        | 691.1558    | 26.29                 |
| 30  | 95 584  | 952        | 942.9370    | 30.71                 |
| 35  | 94 632  | 1 516      | 1 493.0174  | 38.64                 |
| 40  | 93 116  | 2 514      | 2 450.7980  | 49.51                 |
| 45  | 90 602  | 3 857      | 3 708.2355  | 60.90                 |
| 50  | 86 745  | 5 495      | 5 193.0498  | 72.06                 |
| 55  | 81 250  | 7 498      | 6 935.8000  | 83.28                 |
| 60  | 73 752  | 9 691      | 8 751.8452  | 93.55                 |
| 65  | 64 061  | 11 657     | 10 298.1435 | 101.48                |
| 70  | 52 404  | 14 168     | 12 160.6778 | 110.28                |
| 75  | 38 236  | 15 552     | 13 133.3530 | 114.60                |
| 80  | 22 684  | 12 665     | 11 060.9778 | 105.17                |
| 85  | 10 019  | 10 019     | 9 015.1964  | 94.95                 |

**Забележка:** по данни от табл. 1.

Съществува възможност за глобално определяне на вариацията в броя на стигащите до поредните възрасти. За целта може да се разгледа първоначалният брой живородени  $l_0$ , който с течение на времето се разпада на отделните съвкупности на умрелите,

$$(13) \quad l_0 = {}_1d_0 + {}_4d_1 + {}_5d_5 + {}_5d_{10} + \dots$$

Вече бе показано, че редът на стигащите до поредните възрасти  $K_x$  образува случен процес, чиято траектория варира около реда на първа главна съвкупност на живите в таблиците за смъртност. В резултат на това и броят на умрелите  ${}_nD_x = K_x - K_{x+n}$  би трябвало също да се колебае около съответните  ${}_nd_x$ , образувайки полиномно разпределение с вероятности за отделните категории, равни

на  $\frac{l_x}{l_0}(1 - \frac{l_{x+n}}{l_x})$  (т.е. вероятността да се достигне възраст  $x$ , но не и следващата  $x+n$ ) и математическо очакване, равно на  ${}_n d_x = l_x - l_{x+n}$ . Дисперсията (респ. стандартното отклонение) на умрелите  ${}_n D_x$  като случайна величина може да се получи по формулата

$$\text{Var}({}_n D_x) = \frac{(l_x - l_{x+n})(l_0 - l_x + l_{x+n})}{l_0}.$$

В табл. 4 са дадени тези показатели по данни от табл. 1.

При такова положение сборът от квадрати на отклоненията на броя на умрелите от математическото им очакване с отчитане на стандартното отклонение би имал разпределение, близко до  $\chi^2$ , със степени на свобода, равни на броя на възрастите в таблицата минус едно (поради релацията (13)). В резултат на това елипсоидът

$$(14) \quad \sum_x \left( \frac{{}_n D_x - (l_x - l_{x+n})}{\sqrt{\frac{(l_x - l_{x+n})(l_0 - l_x + l_{x+n})}{l_0}}} \right)^2 \leq \chi_{0.95}^2$$

би определял областта на вероятните стойности за реда  $K_x$ , ако това поколение е изложено на същия риск, както е представен в таблицата за смъртност (лимитната стойност  $\chi_{0.95}^2$  в случая е избрана като десен 5 %-ен квантил на  $\chi^2$  – разпределението, съответна на едностранния критерий откъм високите стойности). В противен случай, ако неравенството (14) се окаже нарушено, би трябвало да се наложи изводът, че смъртността в условното поколение  $K_x$  е различна от тази в таблицата за смъртност и се подчинява на друг възрастов профил на риска.

Формула (14) представлява интегрален показател за различието между фактическото разпределение на умрелите и доживяващите до отделните възрасти, от една страна, и съответните показатели от таблиците за смъртност като модел на възрастовата смъртност. Би могло този резултат да се представи с помощта на показателите на таблиците за смъртност и така

$$(15) \quad l_0 \sum_x \frac{({}_n K_x - K_{x+n} - {}_n d_x)^2}{{}_n d_x (l_0 - {}_n d_x)} \leq \chi_{0.95}^2.$$

Методът, свързан с формули (14) и (15), има определени предимства пред съответните от известния метод  $\chi^2$ , тъй като в знаменател по-точно се отчита



вариацията в броя на умрелите  ${}_nD_x$ . Същевременно наличието на общ мащабен множител в тези формули (най-добре това се вижда при (15)) разкрива известна неопределеност, доколкото такъв би могъл да бъде размерът на изходната съвкупност  $l_0$ . Впрочем такава зависимост от мащаба е характерна въобще за методите от групата на  $\chi^2$ .

#### 4. МЕТОДИ ЗА СРАВНИТЕЛЕН АНАЛИЗ НА СРЕДНАТА ПРОДЪЛЖИТЕЛНОСТ НА ЖИВОТА

Средната продължителност на предстоящия живот е вероятно най-често използваният показател от таблиците за смъртност и широко се прилага като инструмент на сравнителния анализ. Съответните доверителни интервали може да се получат, като преди това се намери формула за дисперсията на тази величина. В научната литература и в редица практически приложения се среща формулата<sup>1</sup> на Чанг за дисперсията на средната продължителност на предстоящия живот, извеждането на която почива на няколко резонни, но хипотетични положения най-вече по отношение на корелацията между различните събития, възникващи в хода на стареенето на условното поколение от таблиците за смъртност.

Тази формула е<sup>2</sup>

$$(16) \quad \text{Var}(E_x) = \frac{1}{l_x^2} \sum_{t=0}^{\omega-x} (\dot{e}_{x+t+1} + 1 - a_{x+t})^2 l_{x+t} p_{x+t} q_{x+t} + l_{\omega} \text{Var}(E_{\omega}),$$

където с  $\omega$  е означена последната възраст в таблиците за смъртност.

В тази формула има два съществени момента. Първият е свързан с присъствието на дисперсията на броя на умрелите за отделните възрасти, оценката на която за най-старшите възрасти в таблиците за смъртност (и по-специално величината на  $\text{Var}(E_{\omega})$  за последната възраст) изглежда свързана с доста трудности. Националният център по здравна статистика на САЩ използва за целта помощни данни от социалното осигуряване за възрастните над 85 години (по данни от Medicare, US Department of Health & Human Services 1997) с оглед преодоляването на известни непълноти в текущата статистическа отчетност, породени от неточното отчитане на възрастта на умрелите. Особено остро стои този проблем при последната възраст на таблиците за смъртност, от която всъщност започва неограничен, отворен възрастов интервал, както и при оценката на средната продължителност на оставащия живот след тази възраст. Ста-

<sup>1</sup> Вж. Chiang (1968). Еквивалентна формула се намира у Keyfitz (1968). Същият автор отбелязва, че този резултат е получен от Уилсън още през 1938 г.

<sup>2</sup> Означенията съответстват на показателите от табл. 1, като  $E_x$  представлява средната продължителност на предстоящия живот като случайна величина с математическо очакване  $\dot{e}_x$ . Използването на формулата се препоръчва само при таблици за смъртност по единични възрасти, към което ще се придържа в тази част.

мистиката на смъртността на българското население разкрива изненадващ спад на риска след възраст 80-85 години – факт, на който вече е обръщано внимание (Големанов, 2001, Golemanov et al., 2001). Същевременно данните на нашата статистика за средната продължителност на предстоящия живот на лицата, стигнали до пределна възраст, примерно сто години, показват доста ниски стойности. Трудно може да се приеме, че тези лица живеят средно още само около половин година.

Другият съществен момент във формулата на Чанг е хипотезата за независимост в доживяването с напредването във възрастта, некорелираност на отделните фази от стареенето. Такова положение е съвсем логично (особено при липса на емпиричен материал, говорещ за обратното) и е в съответствие с модела на биномното разпределение на доживяващите (и умрелите) през отделните възрасти, но не е съвсем пълно. Последното се открива в (16) при показателя на Чанг  $a_{x+t}$ , който е също варираща величина и зависи от равнището на възрастовата смъртност, т.е. свързан е с останалите показатели от таблиците за смъртност като  $l_{x+t}$ ,  $p_{x+t}$  и  $q_{x+t}$  за същата възраст.

Тези особености навеждат на мисълта, че намирането на по-точна количествена оценка на вариацията в средната продължителност на предстоящия живот е все пак проблематично. Във всеки случай логиката на хипотезите, залегнали във формула (16), изглежда издържана, което може да се потвърди от едно друго разглеждане, при което не се налага хипотеза за независимост на отделните фази от стареенето, макар че се привличат други опростяващи модела положения. Такава верификация на формулата на Чанг би могла да се постигне по следния начин.

Точната формула за средната продължителност на предстоящия живот на лицата, достигнали възраст  $x$ , е

$$e_x^o = \frac{1}{l_x} \int_0^\infty l_{x+t} dt = \frac{1}{l_x} (L_x + L_{x+1} + L_{x+2} + \dots),$$

което, представено с показателите  $a_x$ , става

$$(17) \quad e_x^o = a_x + \frac{1}{l_x} \sum_{t=0}^{\infty} (1 - a_{x+t} + a_{x+t+1}) l_{x+t+1}.$$

В актюерството често се полага  $a_x = 0.5$ , в резултат на което последната формула се превръща в следния съкратен вариант (м.нар. Curtate Life Expectancy):

$$(18) \quad e_x = \frac{1}{2} + \frac{1}{l_x} \sum_{t=1}^{\infty} l_{x+t}.$$

Резултатите, получени от (18) по единични възрасти, обикновено са доста близки до тези по точната формула (17), което обяснява популярността на съкратения вариант. Строго погледнато обаче, формула (18) притежава редица ге-

фекти. За възраст нула хипотезата  $a_0 = 0.5$  е доста далече от действителността. Поради високата, но бързо спадаща детска смъртност, величината на  $a_0$  се движи в граници от 0,15 до 0,25 в зависимост от това, в какво съчетание се намират двата непосредствени фактора: дали височината на детската смъртност или скоростта на нейното намаление доминира през първите дни и седмици от живота. И за следващите няколко възрасти този показател е под 0,5, но разликата не е толкова голяма. Хипотезата  $a_x = 0.5$  приляга най-вече за средните възрасти от десет до тридесет години, където смъртността е най-ниска и сравнително равномерно разпределена в рамките на отделните възрастови интервали. След тези възрасти рискът непрекъснато нараства, при това с ускоряващ се темп, в резултат на което величината на показателя  $a_x$  започва да надхвърля междинната стойност 0,5 и с нарастване на възрастта разликата прогресивно се увеличава.

Тези отклонения в двете посоки обаче взаимно се погасяват в (18) и по тази причина формулата се оказва напълно приемлива.

Предвид това, интерес представлява вариацията във величината

$$(19) \quad E_x = \frac{1}{2} + \frac{1}{K_x} \sum_{t=1}^{\infty} K_{x+t},$$

представляваща емпиричният еквивалент на (18), чиято стойност се очаква да се отклонява от  $e_x$  в определена степен, което може да послужи като изходен пункт за оценка на дисперсията на средната продължителност на предстоящия живот.

Математическото очакване на  $E_x$  съвпада с  $e_x$ , което лесно се установява, като се има предвид следното:

$$(20) \quad E(E_x | K_x) = 0.5 + \frac{1}{K_x} \sum_{t=1}^{\infty} \frac{l_{x+t}}{l_x} K_x = e_x.$$

Последното осигурява неизместеност (при хипотеза, че до възраст  $x$  стигат  $K_x$ ) поне по отношение на съкратената формула за средната продължителност на предстоящия живот.

За дисперсията на величината (19) може да се изведе следната сравнително проста формула:

$$(21) \quad \text{Var}(E_x | K_x) = \frac{1}{K_x^2} \text{Var}\left(\sum_{t=1}^{\infty} K_{x+t} | K_x\right)$$

$$\begin{aligned}
&= \frac{1}{K_x} \left[ \sum_{t=1}^{\infty} K_x \frac{l_{x+t}}{l_x} \left(1 - \frac{l_{x+t}}{l_x}\right) + 2 \sum_{t=1}^{\infty} \sum_{s=1}^{\infty} K_x \frac{l_{x+t+s}}{l_x} \left(1 - \frac{l_{x+t}}{l_x}\right) \right] \\
&= \frac{1}{K_x} \sum_{t=1}^{\infty} \left(1 - \frac{l_{x+t}}{l_x}\right) \left( \frac{l_{x+t}}{l_x} + 2 \sum_{s=1}^{\infty} \frac{l_{x+t+s}}{l_x} \right) \\
&= \frac{1}{K_x} \sum_{t=1}^{\infty} \left(1 - \frac{l_{x+t}}{l_x}\right) \left( \frac{l_{x+t}}{l_x} + 2 \frac{l_{x+t}}{l_x} \sum_{s=1}^{\infty} \frac{l_{x+t+s}}{l_{x+t}} \right) \\
&= \frac{1}{K_x} \sum_{t=1}^{\infty} \left(1 - \frac{l_{x+t}}{l_x}\right) \frac{l_{x+t}}{l_x} (1 + 2(e_{x+t} - 0.5)) \\
&= \frac{2}{K_x} \sum_{t=1}^{\infty} \frac{l_{x+t}}{l_x} \left(1 - \frac{l_{x+t}}{l_x}\right) e_{x+t} .
\end{aligned}$$

Очакванията са, че (21) би давала малко по-ниски стойности за търсената дисперсия, тъй като при тази формула се игнорира вариацията в показателя на Чанг  $a_x$  след полагането  $a_x = 0.5$  при съкратената формула за средната продължителност на предстоящия живот (18). От друга страна обаче, при формула (21) не се предполага никаква хипотеза за некорелираност (и независимост) на събитията за различните възрасти, без което формулата на Чанг просто не би била валидна.

**Таблица 5**

Вариация в средната продължителност на предстоящия живот

| x  | $l_x$   | $e_x$  | Дисперсия на средната продължителност на предстоящия живот |                   | Стандартно отклонение на средната продължителност на предстоящия живот |                   |
|----|---------|--------|------------------------------------------------------------|-------------------|------------------------------------------------------------------------|-------------------|
|    |         |        | по съкращения метод                                        | по метода на Чанг | по съкращения метод                                                    | по метода на Чанг |
| 1  | 2       | 3      | 4                                                          | 5                 | 6                                                                      | 7                 |
| 0  | 100 000 | 67.072 | 0.003192                                                   | 0.003198          | 0.056498                                                               | 0.056554          |
| 1  | 98 229  | 67.277 | 0.002480                                                   | 0.002480          | 0.049800                                                               | 0.049797          |
| 5  | 97 790  | 63.571 | 0.002304                                                   | 0.002304          | 0.047997                                                               | 0.047997          |
| 10 | 97 554  | 58.719 | 0.002222                                                   | 0.002222          | 0.047134                                                               | 0.047133          |
| 15 | 97 341  | 53.842 | 0.002161                                                   | 0.002160          | 0.046481                                                               | 0.046481          |
| 20 | 96 906  | 49.071 | 0.002059                                                   | 0.002059          | 0.045378                                                               | 0.045378          |

Продължение

| 1  | 2      | 3      | 4        | 5        | 6        | 7        |
|----|--------|--------|----------|----------|----------|----------|
| 25 | 96 280 | 44.373 | 0.001939 | 0.001939 | 0.044035 | 0.044035 |
| 30 | 95 584 | 39.677 | 0.001833 | 0.001833 | 0.042815 | 0.042816 |
| 35 | 94 632 | 35.049 | 0.001723 | 0.001723 | 0.041509 | 0.041510 |
| 40 | 93 116 | 30.576 | 0.001593 | 0.001593 | 0.039914 | 0.039915 |
| 45 | 90 602 | 26.350 | 0.001437 | 0.001437 | 0.037910 | 0.037914 |
| 50 | 86 745 | 22.404 | 0.001266 | 0.001266 | 0.035582 | 0.035587 |
| 55 | 81 250 | 18.741 | 0.001094 | 0.001095 | 0.033078 | 0.033087 |
| 60 | 73 752 | 15.378 | 0.000931 | 0.000932 | 0.030514 | 0.030528 |
| 65 | 64 061 | 12.315 | 0.000783 | 0.000785 | 0.027987 | 0.028010 |
| 70 | 52 404 | 9.476  | 0.000672 | 0.000674 | 0.025928 | 0.025969 |
| 75 | 38 236 | 7.046  | 0.000604 | 0.000609 | 0.024583 | 0.024668 |
| 80 | 22 684 | 5.156  | 0.000610 | 0.000621 | 0.024698 | 0.024927 |
| 85 | 10 019 | 3.699  | 0.000661 | 0.000693 | 0.025702 | 0.026326 |

**Забележка:** по данни от табл. 1.

Данните в табл. 5 показват обаче минимални разлики между двата метода, които са без никакво практическо значение, като за ред възрасти е налице пълно съвпадение. Ефектът от хипотезата  $a_x = 0.5$  може да се забележи само при възраст нула и при най-старшите възрасти, тъй като тъкмо за тези възрасти величината на показателя на Чанг по-съществено се отклонява от средата на възрастовия интервал.

Формула (21) има това предимство пред тази на Чанг, че насочва вниманието към ефекта върху дисперсията от броя на стигналите до възраст  $x$  и позволява да се прецени това влияние. На пръв поглед присъствието на  $K_x$  в знаменател на-вежда на мисълта за възможност за значителни отклонения на величината на дисперсията в зависимост от вариацията в броя на доживяващите. Фактически както при метода на Чанг, така и при съкратения вариант (21) се полага  $K_x = l_x$  (на базата на което са пресметнати величините в табл. 5), което трябва да се разглежда като особен случай, затъмняващ действителното положение с вариацията на  $K_x$  и ефекта ѝ върху  $E_x$  от (19). Отклонения в едната посока в броя на доживяващите  $K_x$  биха могли да доведат до значителни разлики в големината на дисперсията, получена по (21) или (16). Но тъй като в реда  $K_x, K_{x+1}, K_{x+2}, \dots$  е налице ярка позитивна корелация, по скоро флуктуации в  $K_x$  биха породили аналогични и еднопосочни отклонения при всички следващи възрасти, които взаимно биха се погасили във формула (19). Този въпрос може да се осветли, като се има предвид, че формула (21) има много по-широка валидност при сравнително лекото условие  $K_x > 0$  (при мащаб на таблиците за смъртност от порядъка на

$K_0 = 100,000$  събитието  $K_x = 0$  е съвсем невероятнo за повечето възрасти), в резултат на което може да се получи пълната дисперсия на средната продължителност на предстоящия живот:

$$\text{Var}(E_x) = E(\text{Var}(E_x | K_x)) + \text{Var}(E(E_x | K_x)).$$

Второто събираемо в тази формула е нула предвид (20), в резултат на което

$$(22) \quad \text{Var}(E_x) = 2E\left(\frac{1}{K_x} | K_x > 0\right) \sum_{t=1}^{\infty} \frac{l_{x+t}}{l_x} \left(1 - \frac{l_{x+t}}{l_x}\right) e_{x+t}.$$

Данните в табл. 6 показват, че въпросният ефект е практически нулев. Във втора колона на таблицата са пресметнати  $E(\frac{1}{K_x} | K_x > 0)$ , а в трета колона – реципрочните стойности на първа главна съвкупност на живите. Поразителното съвпадение на числата показва, че при такъв голям мащаб на таблиците за смъртност от порядъка на  $l_0 = 100,000$  законите за големите числа надделяват над случайността във величината на  $K_x$ , изравняват я с  $l_x$  и фактически правят условието  $K_x > 0$  съвсем нищожно.

**Таблица 6**

Корекция в дисперсията на средната продължителност на предстоящия живот

|    | $E(\frac{1}{K_x}   K_x > 0)$ | $\frac{1}{l_x}$ | $l_x E(\frac{1}{K_x}   K_x > 0)$ |
|----|------------------------------|-----------------|----------------------------------|
| 1  | 2                            | 3               | 4                                |
| 0  | 0.00001000                   | 0.00001000      | 1.00000000                       |
| 1  | 0.00001018                   | 0.00001018      | 0.99997122                       |
| 5  | 0.00001023                   | 0.00001023      | 1.00039170                       |
| 10 | 0.00001025                   | 0.00001025      | 0.99992850                       |
| 15 | 0.00001027                   | 0.00001027      | 0.99969207                       |
| 20 | 0.00001032                   | 0.00001032      | 1.00006992                       |
| 25 | 0.00001039                   | 0.00001039      | 1.00034920                       |
| 30 | 0.00001046                   | 0.00001046      | 0.99980864                       |
| 35 | 0.00001057                   | 0.00001057      | 1.00026024                       |
| 40 | 0.00001074                   | 0.00001074      | 1.00006584                       |
| 45 | 0.00001104                   | 0.00001104      | 1.00024608                       |
| 50 | 0.00001153                   | 0.00001153      | 1.00016985                       |
| 55 | 0.00001231                   | 0.00001231      | 1.00018750                       |

Продължение

| 1  | 2          | 3          | 4          |
|----|------------|------------|------------|
| 60 | 0.00001356 | 0.00001356 | 1.00007712 |
| 65 | 0.00001561 | 0.00001561 | 0.99999221 |
| 70 | 0.00001908 | 0.00001908 | 0.99986832 |
| 75 | 0.00002615 | 0.00002615 | 0.99987140 |
| 80 | 0.00004409 | 0.00004408 | 1.00013756 |
| 85 | 0.00009982 | 0.00009981 | 1.00009658 |

**Забележка:** по данни от табл. 1.

По такъв начин формула (22) се оказва практически равносилна на (21) и може да получи следния окончателен вариант:

$$(23) \quad \text{Var}(E_x) = \frac{2}{l_x} \sum_{t=1}^{\infty} \frac{l_{x+t}}{l_x} \left(1 - \frac{l_{x+t}}{l_x}\right) e_{x+t}.$$

В последната колона на табл. 6 е показано в какво съотношение се променя дисперсията на средната продължителност на предстоящия живот, получена по формули (22) и (23) в сравнение с (21).

Разгледаните формули за дисперсията на средната продължителност на предстоящия живот позволяват намирането на доверителни интервали за този показател. При най-често използваните нива на доверителност 95 % и 99 % такива интервали се получават приблизително равни на плюс и минус две стандартни отклонения от получената величина. Данните от табл. 5 показват, че този диапазон е равен на около един месец за младшите възрасти и постепенно се стеснява към най-високите възрасти, като спада до около половин месец. Това потвърждава, че средната продължителност на предстоящия живот е подчертано стабилен показател, слабо влияещ се от случайните отклонения в хода на стареенето на условното поколение от таблиците за смъртност. Той остава такъв и при намаляване на мащаба на таблиците до около  $l_0 = 1000$ .

Подобна техника е приложима и при сравняване на данните от две или повече таблици за смъртност. Пропорциите на доживяващите например може да се сравняват за отделните възрасти, като се използва формула, аналогична на (12),

$$(24) \quad \frac{\frac{l'_{x+t}}{l'_x} - \frac{l''_{x+t}}{l''_x}}{\sqrt{\frac{l'_{x+t} + l''_{x+t}}{l'_x l''_x} \left(1 - \frac{l'_{x+t} + l''_{x+t}}{l'_x + l''_x}\right)}}.$$

Тази величина би имала приблизително стандартно нормално разпределение, ако две таблици за смъртност със съответна първа главна съвкупност на живите

$l'_x$  и  $l''_x$  отразяват един и същ профил на смъртността във възрастовия интервал от  $x$  до  $x+t$ . При такава хипотеза, ако се констатира някаква разлика в съответните числа, то това би трябвало да се отнесе изцяло към случайните колебания в стареенето на едно и също условно поколение, просто представено с два екземпляра. В противен случай, ако величината на (24) надхвърля съответните перцентилни граници (аналогично на формула (12)), би трябвало да се счита, че двете таблици представят различна смъртност за съответните възрасти.

Формула (24) обаче дава ефикасен отговор само за отделни фрагменти от възрастовия профил на смъртността и по-интересен и по-полезен би бил един цялостен показател за различията в смъртността, общ за всички възрасти. Такъв показател би могъл да се получи аналогично на (15)

$$(25) \quad 2l_0 \sum_x \frac{(d'_x - d''_x)^2}{(d'_x + d''_x)(2l_0 - d'_x - d''_x)}.$$

При еднакъв възрастов профил на смъртността от две таблици със съответна първа главна съвкупност на умрелите  $d'_x$  и  $d''_x$  тази величина би имала разпределение  $\chi^2$  с определени степени на свобода. И съответно на това не би надхвърляла примерно 95 %-ия перцентил на разпределението. В противен случай би трябвало да се стигне до извода, че двете таблици представят различен възрастов профил на смъртността и разликата между тях, изразяваща се в различен размер на първа главна съвкупност на умрелите за отделните възрасти, не може да се обясни само със случайността.

При формули (24) и (25) се изхожда от хипотезата, че двете условни поколения стареят независимо едно от друго, което е съвсем приемливо в случай на сравняване на данни, извлечени от текущата демографска статистика, отнасящи се за различни години или различни територии и населения. При клиничните изследвания или пък при изучаване на застрахователните съвкупности подобна хипотеза може да не изглежда съвсем убедителна.

Такава хипотеза дава възможност двете условни поколения да се разглеждат обединено, като че ли произлизат от една и съща таблица за смъртност с двойно по-голям мащаб, ако, разбира се, възрастовият профил на смъртността е еднакъв. Общата дисперсия от така окрупнените показатели би била просто сбор от дисперсиите на отделните съставляващи, тъй като поради хипотезата за независимост корелацията между тях би била нула.

По подобен начин би се получил и показател за сравняване на продължителността на живота по данни от две или повече таблици за смъртност. Дисперсията, съответна на разликата  $e'_x - e''_x$ , би била сбор от съответните дисперсии и при еднаква смъртност тази величина би могла да се получи по формула (16) и (21) след комбиниране на двете условни поколения в едно. Тъй като средната продължителност на живота е доста обобщаващ и синтезиращ показател, този инструмент за сравняване би заемал междинно положение между интегралния показател (25) и възрастовите сравнения с помощта на (24).

От това разглеждане става ясно в каква насока може да се развие методът и как би могъл да се получи интегрален показател за разликите в средната



продължителност на предстоящия живот, общ за всички възрасти. Той би имал следната форма:

$$(26) \quad \frac{1}{2} \sum_x \frac{(e'_x - e''_x)^2}{Var(E_x)},$$

където величината на  $Var(E_x)$  би могла да се получи по формули (16) и (21) след комбиниране на двете условни поколения в едно. Ако възрастовата смъртност би била еднаква в двете таблици, величината (26) би имала разпределение хи-квадрат с определени степени на свобода и в повечето случаи не би надхвърляла критичните перцентилни нива.

Сравняването на данните от таблиците за смъртност за мъжете в България за годините 1995-1997 г. и 1999-2001 г. разкрива съществено различие във възрастовия профил на смъртността. Величината  $\chi^2$  по формула (25) има стойност 962.31, далече надхвърляща критичните нива ( $\chi^2_{0.95} = 113.15$  и  $\chi^2_{0.99} = 124.12$  при степени на свобода 90, съответни на единичните възрасти от нула до деветдесет).

Подобна разлика се получава и при сравняване на средната продължителност на предстоящия живот за отделните възрасти и общо по формула (26). Дори и простото сравняване на средната продължителност на целия живот да показва нарастване с около година и половина, което е доста изненадващо, трудно би се постигнало в действителността за един сравнително кратък период от около пет години и изглежда малко вероятно.

Подобни разлики може да произлизат не само от различна смъртност, а и от естествената вариация в изходните статистически данни за смъртността, неизбежно пренесена в показателите на таблиците за смъртност, а също така и от евентуална промяна в методологията на получаването на таблиците или дори от допускането на някакви систематични грешки както в отчетните данни, така и в използваните методи. На табл. 7 са дадени първите няколко възрасти от таблицата за смъртност, публикувана от НСИ в „Население и демографски процеси 2001“. С използване на показателя на Чанг  $a_x$  лесно се откриват някои неблагоприятния при началните възрасти, където се забелязват необичайно ниски и нелогични числа. По отношение на средната продължителност на предстоящия живот би могла да се направи и една проверка с използване на формула (18), при което може да се установи, че последната дава стойности, по-ниски от тези в табл. 7 за всички възрасти без изключение. Това хвърля известна светлина върху произхода на разликите между двете таблици за смъртност.

Таблица 7

Таблица за смъртност на българското население – мъже, 1999-2001 г.

| $x$ | $l_x$   | $d_x$ | $p_x$    | $q_x$    | $L_x$  | $a_x$ | $T_x$     | $e_x$  |
|-----|---------|-------|----------|----------|--------|-------|-----------|--------|
| 0   | 100 000 | 1 456 | 0.985440 | 0.014560 | 98 654 | 0.076 | 6 853 416 | 68.534 |
| 1   | 98 544  | 118   | 0.998803 | 0.001197 | 98 443 | 0.144 | 6 754 762 | 68.546 |
| 2   | 98 426  | 60    | 0.999390 | 0.000610 | 98 375 | 0.150 | 6 656 319 | 67.628 |
| 3   | 98 366  | 67    | 0.999319 | 0.000681 | 98 308 | 0.134 | 6 557 944 | 66.669 |
| 4   | 98 299  | 45    | 0.999542 | 0.000458 | 98 260 | 0.133 | 6 459 636 | 65.714 |
| 5   | 98 254  | 37    | 0.999623 | 0.000377 | 98 234 | 0.459 | 6 361 376 | 64.744 |
| 6   | 98 217  | 35    | 0.999644 | 0.000356 | 98 199 | 0.486 | 6 263 142 | 63.768 |
| 7   | 98 182  | 46    | 0.999531 | 0.000469 | 98 158 | 0.478 | 6 164 943 | 62.791 |
| 8   | 98 136  | 38    | 0.999613 | 0.000387 | 98 115 | 0.447 | 6 066 785 | 61.820 |
| 9   | 98 098  | 22    | 0.999776 | 0.000224 | 98 086 | 0.455 | 5 968 670 | 60.844 |
| 10  | 98 076  | 25    | 0.999745 | 0.000255 | 98 063 | 0.480 | 5 870 584 | 59.857 |

Източник: „Население и демографски процеси 2001“, НСИ, София, 2003.

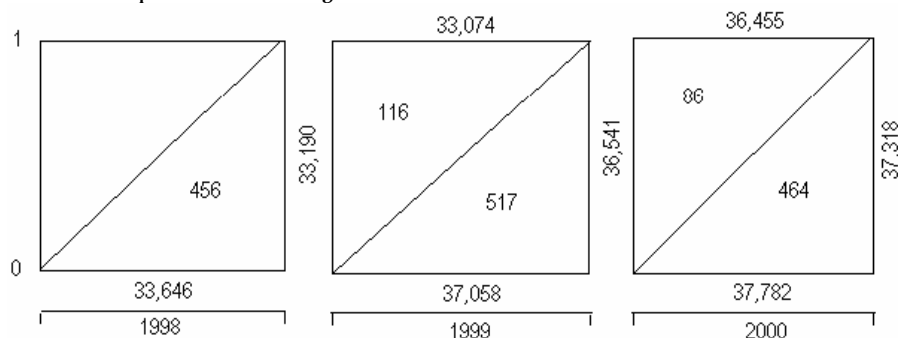
Чисто технически, но доста важен е въпросът за степените на свобода на разпределенията, свързани с формули (25) и (26). Условните поколения от двете таблици за смъртност с основание може да се считат за независими, но в сборовете в тези формули фигурират събираеми от едни и същи поколения (но за различни възрасти), което би могло да редуцира степените на свобода. Тъй като проблемът е повече математически и статистически, само ще скицираме решението му. Става дума за отклоненията  $u_x = e'_x - e''_x$ , които при една и съща възрастова смъртност биха варирали около нулата, така че по формула (26) в повечето случаи не би се получила величина, надхвърляща съответната критична стойност на разпределението  $\chi^2$ . Самата формула (26) представлява квадратична форма на отклоненията  $u_x$  с пълен ранг, при това ковариациите (корелациите) на отклоненията също образуват матрица с пълен ранг. Фактически коефициентите на корелация може да се получат по следната формула:

$$Cor(u_x, u_{x+t}) = \frac{l_{x+t}}{l_x} \sqrt{\frac{Var(E_{x+t})}{Var(E_x)}} \text{ при } t \geq 0.$$

В резултат на това разпределението на величината (26) би имала разпределение  $\chi^2$  със степени на свобода, равни точно на броя на събираемите в съответната формула, т.е. на броя на възрастовите групи. От това може да се направи и изводът, че величината (25) би се оказала с единица по-малко степени на свобода поради релацията (13).

## 5. ЕНДОГЕННИ И ЕКЗОГЕННИ ФАКТОРИ ЗА ВАРИАЦИЯТА В СТАТИСТИЧЕСКИТЕ ДАННИ

Общоприето е показателите на таблиците за смъртност да се получават по данни от преброяванията на населението и от текущата статистическа отчетност или от регистрите за населението по единични възрасти и отделно за двата пола и за различните поколения с отразяване на най-малките подробности (вж. базата данни за смъртността HMD (Human Mortality Data) на Университета на Калифорния в Бъркли ([www.mortality.org](http://www.mortality.org)) и на Института за демографски изследвания „Макс Планк“ в Рочок ([www.humanmortality.de](http://www.humanmortality.de))). Съществен момент в съответните методи е свързан с използването на данни за умрелите по елементарните съвкупности и диаграмата на Лексис<sup>1</sup>. Така например за началната възраст вероятността за умирање, респ. за стигане до следващата възраст, се получава чрез съпоставяне на умрелите на възраст под една година със съответния брой живородени, от които фактически води началото си това поколение.



Фиг. 2. Диаграма на Лексис

На фиг. 2 е показано как може да се определят тези пропорции по информация<sup>2</sup> за три съседни поколения. По абсцисата са разположени годишният брой живородени в последователността на съответните години. По ординатата са дадени възрастите, като на фигурата е показан само началният възрастов интервал от нула до една година, който представлява интерес в случая. Рождената дата на всяко живородено може да се изобрази на абсцисата като отделна точка, от която започва „линията на живота“ – диагонална права в посока нагоре и надясно под ъгъл 45°, продължаваща, докато лицето е живо. Краят на линията има две съществени координати: абсцисата дава датата на смъртта, а ординатата – точната възраст. Така се формират и елементарните съвкупности на умрелите, представляващи отделните триъгълници в диаграмата на Лексис, образувани от съчетанието на три признака: възраст на умрелите, година на смъртта и от кое поколение произлизат.

<sup>1</sup> Тук се използва англосаксонският вариант на диаграмата на Лексис. Немският вариант може да се намери у Сугарев (1975).

<sup>2</sup> Данни за населението на България за мъжкия пол от публикацията „Население и демографски процеси“ на НСИ за съответните години.

На графичното изображение се вижда, че от живородените 33 646 през 1998 г. 456 умират същата година и до края на годината стигат 33 190. Така се формира долната елементарна съвкупност на умрелите от това поколение за възраст нула и 1998 г. По-нататък, но вече през следващата 1999 г., от същото поколение отпадат 116 и в резултат до възраст 1 стигат 33 074. Това представлява горната елементарна съвкупност на умрелите от поколението 1998 г. за възраст нула. Движението на поколенията по диаграмата на Лексис разкрива как може да се намери и вероятността живородено през 1998 г. да стигне до следващата възраст, а именно

$$(27) \quad p_0 = \frac{33.074}{33,646} = 0.98300,$$

което представлява основен показател на таблиците за смъртност. Допълнителната вероятност директно отразява риска за умиране и е също основен показател на таблиците за смъртност

$$(28) \quad q_0 = 1 - p_0 = 0.01700.$$

Такива величини често се представят на хиляда, в случая 17 на хиляда би могло да се интерпретира и като средна величина: от хиляда живородени средно 17 умират преди възраст 1, респ. 983 стигат до тази възраст. Да видим на какво може да се опре подобна интерпретация.

Сигналите до възраст една година 33 074 в сравнение с първоначалната съвкупност живородени 33 646 са намалели с 572, което представлява първа главна съвкупност на умрелите и е равна на сбора от разгледаните вече две елементарни съвкупности. Вероятностите  $p_0$  и  $q_0$  са повече свързани с първа главна съвкупност на умрелите, отколкото с двете ѝ елементарни съвкупности.

Живородените през дадена година може да се разглеждат като база за биномно разпределение както на умрелите от първоначалната съвкупност, така и за доживяващите до следващата възраст 1. Ключовата работна хипотеза в такова построение се състои в независимост в развитието на отделните членове на поколенията, което обсъдихме в т. 3. Друг важен момент е свързан с въпроса, дали живородените от едно поколение имат еднакъв шанс да стигнат до следващите възрасти. Строго погледнато, това едва ли е така, но в рамките на едно поколение подобно положение може да се приеме. Този проблем заслужава по-дълбоко изследване, което би довело и до получаване на така липсващите у нас таблици за смъртност за новородените по дни, седмици и месеци на възрастта.

В рамките на модела на биномното разпределение математическото очакване за броя на умрелите на възраст до една година действително е равен на произведението от броя на живородените и вероятността за умиране, което обяснява терминологията „на хиляда“, използвана за вероятностите  $p_0$  и  $q_0$ .

Формули (27) и (28) фактически представляват емпирични еквиваленти за тези вероятности и неизбежно са обременени с вариацията от статистическите данни. Оценка за големината на тази вариация може да се получи във формата на

дисперсията (или стандартното отклонение) на пропорциите на доживяващите, респ. умрелите. Както може да се очаква, дисперсията на тези показатели е сравнително ниска за повечето възрасти главно по две причини. Първо, съвкупностите на живите са достатъчно обширни, така че законите за големите числа не оставят много място за стохастични отклонения. И второ, рискът за умирање е сравнително нисък, което прави смъртните случаи изключително редки събития. Чак при най-старшите възрасти може да се забележи по-значително увеличение на дисперсията пак по същите причини, но действащи вече в обратна посока: съвкупностите на достигналите високите възрасти са сравнително малобройни и рискът за умирање е значително по-висок. Ректификацията и изглаждането на данните за смъртността в старшите възрасти са добре известен проблем в демографията и демографската статистика и бе засегнат в т. 4.

Една възможност за смекчаване на ефекта от такива случайности е чрез агрегиране и осредняване на данните за няколко съседни поколения. Това е широко разпространена практика, към която се придържа и НСИ. Получените по такъв начин средни величини може да отговарят на фактическото положение, ако не се наблюдава динамика във възрастовия профил на смъртността. В противен случай подобни показатели биха се оказали на едно по-високо ниво на абстрактност.

Така например би могло да се провери доколко съседните поколения мъже, родени през 1998 и 1999 г., се различават по детска смъртност. Формула, аналогична на (24), но с отчитане на фактическите величини (по данните от фиг. 2), би дала следния резултат:

$$\frac{\frac{33,074}{33,646} - \frac{36,455}{37,058}}{\sqrt{\frac{33,074 + 36,455}{(33,646)(37,058)} \left(1 - \frac{33,074 + 36,455}{33,646 + 37,058}\right)}} = -0.757013.$$

Получената стойност не превишава критичните нива и показва, че не се наблюдава различие в детската смъртност между двете поколения. С други думи, аритметичната разлика между пропорции

$$\frac{33,074}{33,646} = 0.98300 \text{ и } \frac{36,455}{37,058} = 0.98373$$

се дължи само на случайността и няма достатъчна статистическа тежест, за да докаже такава разлика и в смъртността. Следователно средната пропорция на доживяващите

$$\frac{33,074 + 36,455}{33,646 + 37,058} = 0.98338$$

е не само достатъчно репрезентативна за търсената вероятност, но е и на същото ниво на непосредственост, на каквато са и изходните статистически данни.

Ситуацията е аналогична на разглежданата вече стохастика в стареенето на условното поколение в таблиците за смъртност. Разликата е само в това, че тук мащабът се определя от самите демографски процеси, от размера на конкретното население през разглеждания период и е извън всякакъв избор на изследователя.

Подобно е положението и при следващите възрасти, но проблемът с оценката на вариацията се усложнява от факта, че трябва умрелите от отделните поколения да се съпоставят с живите в началото на годината, като данните за последните обикновено са получени с придвижване на възрастовите групи през предходните периоди. Така се получава натрупване на дисперсия в течение на няколко години, което неизбежно увеличава размаха в пропорциите на доживяващите. Същото е налице и при началната възраст, макар и в по-малка степен, тъй като годишният брой живородени е също варираща величина *ceteris paribus* – при едно и също ниво на раждаемост и една и съща възрастова плодovitост на родилния контингент.

Освен стохастичните отклонения, генетично свързани със стареенето на поколенията, случват се разнообразни разлики и дори систематични грешки в резултат на външни за демографските процеси фактори, което може да увеличи значително дисперсията в показателите на таблиците за смъртност, както и в останалите раздели на демографската статистика. Такива фактори могат да бъдат: непрецизност в обработката на информацията, смяна на методологията за получаване на показателите с недостатъчен сравнителен анализ и стиковане на динамичните редове, непълнота и пропуски в регистрацията и преброяванията, междуведомствена разпокъсаност и други организационни неудачи при движението на информацията. Сред специалистите е разпространено мнението (US Department of Health & Human Services, 1997), че тъкмо тези екзогенни фактори привнасят много повече вариация и несигурност в статистическите показатели, отколкото вътрешната стохастика на самите демографски процеси. За българската демографска статистика – като теория и практика, от особено значение с негативен знак е нерелевантната, а често и съвсем липсващата статистическа информация за размера и структурата на миграционните движения по възраст, пол и диференцирано за отделните поколения, което хвърля сянка и върху резултатите от различни съпоставяния и пропорции в данните за естественото движение на населението.

**Таблица 8**

Деца на възраст под една година в края на 2001 г.

| Общо                             |         |          | В градовете |         |          | В селата |         |          |
|----------------------------------|---------|----------|-------------|---------|----------|----------|---------|----------|
| общо                             | момчета | момичета | общо        | момчета | момичета | общо     | момчета | момичета |
| Живородени през 2001 г.          |         |          |             |         |          |          |         |          |
| 68 180                           | 35 125  | 33 055   | 48 567      | 25 068  | 23 499   | 19 613   | 10 057  | 9 556    |
| От тях умрели през същата година |         |          |             |         |          |          |         |          |
| 826                              | 468     | 358      | 540         | 305     | 235      | 286      | 163     | 123      |

Продължение

| Достигнали края на годината                                |        |        |        |        |        |        |       |       |
|------------------------------------------------------------|--------|--------|--------|--------|--------|--------|-------|-------|
| 67 354                                                     | 34 657 | 32 697 | 48 027 | 24 763 | 23 264 | 19 327 | 9 894 | 9 433 |
| По отчетни данни на НСИ за населението към края на 2001 г. |        |        |        |        |        |        |       |       |
| 66 358                                                     | 34 136 | 32 222 | 47 345 | 24 401 | 22 944 | 19 013 | 9 735 | 9 278 |
| Разлика достигнали края на годината минус отчет            |        |        |        |        |        |        |       |       |
| 996                                                        | 521    | 475    | 682    | 362    | 320    | 314    | 159   | 155   |

**Източник:** „Население и демографски процеси 2001“, НСИ, София, 2003.

Във връзка с това може да се отбележи неприятната тенденция данните от придвижването на възрастовите групи<sup>1</sup> до поредното преброяване на населението да се разминават с данните от преброяването с няколко десетки хиляди. Така например през 1975 и 1985 г. преброяването показва първоначално около 120 000 лица по-малко от съответното изчислено население, която разлика впоследствие бе сведена без много шум до около 30 000. На пръв поглед такива разминавания може да се обяснят с необхващане, пропуски в регистрацията и разпадане в движението на информацията за външната миграция, но само ако за отделните възрасти и отделно за двата пола не се наблюдават разлики в двете посоки, и то от по няколко хиляди. Подобна разнопосочност показва, че такива различия не би могло да се обяснят само с неадекватността или просто с липсата на статистически данни за механичните движения и по-специално за външната миграция. В тази посока може да се намерят много примери и от по-ново време. Таблица 8 показва какви разлики се откриват между живородените през 2001 г., придвижени до края на годината, и отчетните данни на НСИ за населението на страната на 31.XII.2001 г., които се предполага, че са съпоставени и в някаква степен са взаимно верифицирани с резултатите от преброяването през март 2001 г.

Вижда се, че отчетните данни дават около хиляда деца по-малко, отколкото са живородените през годината, намалени със съответната долна елементарна съвкупност на умрелите от тях. Трудно може да се обясни къде са „изчезнали“ тези 996 деца, съвсем невероятно е това да се дължи на външната миграция и тази бройка да представлява отчетено по някакъв начин от статистиката салдо между емигрантите и имигрантите за тази възраст.

Констатираните разлики подсказват за някакъв дълбок проблем в информацията от самото преброяване, още повече, че две и половина години по-късно нашата статистика още не е публикувала точните факти за възрастово-половата структура на българското население към критичния момент на преброяването. Не е изключено да има проблем и при придвижването на възрастовите групи и

<sup>1</sup> Т.нар. изчислено население представлява основен показател в българската демографска статистика и главен продукт на текущата статистическа отчетност на населението на държавната статистика у нас през втората половина на миналия век.

проектирането им до края на м. декември за отделните години<sup>1</sup>, особено ако се срещат трудности при своевременното движение на информацията от първичните актове през регистъра на населението до Националния статистически институт. Не е изключено също преброяването на населението да не е спомогнало за верификацията на данните от текущата демографска статистика. Възможно е също да има значително закъснение в движението на информацията, което налага статистическият орган да изчаква или да публикува корекции, както това е направено с данните за населението за 2001 г.

**Таблица 8а**

Деца на възраст под една година в края на 2001 г.

| Общо                                                          |         |          | В градовете |         |          | В селата |         |          |
|---------------------------------------------------------------|---------|----------|-------------|---------|----------|----------|---------|----------|
| общо                                                          | момчета | момичета | общо        | момчета | момичета | общо     | момчета | момичета |
| С придвижване на живородените до края на годината             |         |          |             |         |          |          |         |          |
| 67 354                                                        | 34 657  | 32 697   | 48 027      | 24 763  | 23 264   | 19 327   | 9 894   | 9 433    |
| По коригирани данни на НСИ за населението към края на 2001 г. |         |          |             |         |          |          |         |          |
| 66 552                                                        | 34 249  | 32 303   | 47 478      | 24 488  | 22 990   | 19 074   | 9 761   | 9 313    |
| Разлика – достигнали края на годината минус отчет             |         |          |             |         |          |          |         |          |
| 802                                                           | 408     | 394      | 549         | 275     | 274      | 253      | 133     | 120      |

**Източник:** „Население и демографски процеси 2001“, НСИ, София, 2003 и „Население 2002“, НСИ, София, 2003

Таблица 8а представлява известно допълнение на данните от табл. 8 и показва сравнението с подобни коригирани данни, като резултатът не разкрива особено подобрене в статистическата отчетност по населението.

**Таблица 9**

Проекция на поколението, родени през 2002 г., към края на годината

|                                                                        | Общо   | Момчета | Момичета |
|------------------------------------------------------------------------|--------|---------|----------|
| Живородени през 2002 г.                                                | 66 499 | 34 497  | 32 002   |
| От тях живи в края на 2002 г. на възраст 0 навършени години            | 65 782 | 34 079  | 31 703   |
| Отчетни данни за населението на възраст под 1 година на 31.XII.2002 г. | 64 914 | 33 642  | 31 272   |
| Разлика:                                                               | 868    | 437     | 431      |

**Източник:** „Население 2002“, НСИ, София, 2003.

<sup>1</sup> Дефекти в придвижването на възрастовите групи, респ. в данните за изчисленото население, се забелязват на много места в продукцията на държавната ни статистика през изминалите десетилетия, вж. Чолаков, Малешков (2003).



Подобна картина се получава и в по-общ план при придвижване на генерациите, родени през периода 1998 – 2002 г. На табл. 9 се вижда, че от 66 499 живородени през 2002 г. умират 717 (образуващи долната елементарна съвкупност на умрелите за тази година от това поколение за възраст под една година), така че до края на 2002 г. достигат 65 782. В бюлетина „Население 2002“ се дават отчетни данни за тази възрастова група с размер 64 914, което прави липса на 868 деца, като разликата идва поравно от момчетата и момичетата. Не е ясно дали НСИ са залагали някакви макар и хипотетични данни за външната миграция, но такива отклонения, по-големи дори от съответните елементарни съвкупности на умрелите, не би могло да се обяснят само с миграционните движения. Впрочем констатираните разлики с отчетните данни са от подобен размер както и при придвижването на предходното поколение до края на 2001 г. (табл. 8 и 8а).

**Таблица 9а**

Проекция на генерациите, родени през 1998 – 2001 г., към края на 2002 г.

|                                                                 | Общо    | Момчета | Момичета |
|-----------------------------------------------------------------|---------|---------|----------|
| Живородени през 1998-2001 г.                                    | 279 511 | 143 611 | 135 900  |
| От тях живи в края на 2002 г. на възраст 1 – 4 навършени години | 275 080 | 141 094 | 133 986  |
| Отчетни данни за населението на възраст 1 – 4 на 31.XII.2002 г. | 262 691 | 134 757 | 127 934  |
| Разлика:                                                        | 12 389  | 6 337   | 6 052    |

**Източник:** „Население и демографски процеси“ и „Население“, НСИ, София, за съответните години

На табл. 9а е показана проекцията на генерациите, родени през периода 1998 – 2001 г., формиращи възрастовата група от една до четири навършени години към 31.XII.2002 г.. При сравняване на проектираните данни с отчетните цифри на НСИ се забелязва дефицит от 12 389, свидетелстващ за сериозно разминаване на демографската ни статистика с фактическото положение.

Тези факти потвърждават, че екзогенната вариация, привнесена отвън демографските процеси и дължаща се на разнородни причини, би могла да окаже значително по-голямо влияние върху основните статистически показатели, отколкото вътрешната стохастика на самите демографски процеси.

## 6. ЕФЕКТ ОТ ВАРИАЦИЯТА В СТАТИСТИЧЕСКИТЕ ДАННИ ВЪРХУ ПОКАЗАТЕЛИТЕ НА ТАБЛИЦИТЕ ЗА СМЪРТНОСТ

Измерването, претеглянето и количественото определяне на влиянието на екзогенните фактори като част от общия ефект върху вариацията на статистическите данни, а оттук и върху демографските показатели са затруднени по ред причини. Първо, външните източници за различните отклонения в данните представляват широка гама от причини и механизми, чието проследяване бързо излиза от сферата на демографската статистика. Второ, методиките и алгоритмите

за получаване на различните демографски показатели и особено тези от таблиците за смъртност са разнообразни и разнородни, а и строежът на показателите е доста различен, така че е трудно да се намери единна рецепта за оценка на вариацията, особено на онази част от нея, която идва от екзогенните фактори. Известна светлина по проблема може да се хвърли с разглеждането на вариацията в различните съотношения и пропорции между демографските съвкупности, представляващи широко разпространен градивен елемент за коефициентите и вероятностите, а заедно с това и за показателите на таблиците за смъртност.

Тези съотношения и пропорции имат следната сравнително обща форма:

$$(29) \quad M = \frac{D}{K},$$

където  $D$  представлява примерно размерът на определена съвкупност на умрелите, а  $K$  – съответно размерът на съвкупността на живите, от която произлизат умрелите на брой  $D$ , така че  $M$  е показател от рода на коефициентите за смъртност и вероятностите за умирање. По такъв начин този показател е в определен смисъл измерител на риска за умирање, на интензитета на смъртността и в идеалния случай (като се пренебрегне дори и неизбежната ендогенна вариация на самите демографски процеси) не би бил обременен с нищо случайно и би бил стопроцентово отражение на смъртността.

Статистическата информация за подобни показатели обикновено е от вида  $D + \Delta D$  и  $K + \Delta K$ , където  $\Delta D$  и  $\Delta K$  представляват отклоненията от действителните стойности на  $D$ , респ.  $K$ , дължащи се на различни ендогенни и екзогенни фактори. Тези отклонения биха могли да бъдат в двете посоки – в плюс или минус, в зависимост от ефекта на действащите върху показателите фактори. Но може и тенденциозно да отклоняват показателите в едната посока при наличието на систематични непълноти, грешки и пропуски в статистическата информация. Подобна конструкция е добре известна в теорията на репрезентативните проучвания (Hansen et al., 1953, Kish 1965) и при липса на систематични отклонения математическото очакване на  $\Delta D$  и  $\Delta K$  би трябвало да е нула. Такъв би бил случаят, ако върху показателя (29) действат само ендогенни фактори, свързани със стохастиката на стареенето на поколенията, което бе разгледано в т. 3 и 4. Такава ситуация, разбира се, не изключва част от екзогенните фактори също да действат разнопосочно и тяхното влияние да се самопогасява в математическото очакване. В такъв случай резултатът от влиянието им би се свеждал само до допълнителен принос в големината на дисперсията на показателите, без да се предизвиква изместване в математическото очакване.

Отклоненията в числителя и знаменателя на (29) независимо от своя произход неизбежно биха индуцирали подобни отклонения в самия показател и техният размер би могъл да се получи по правилата на математическото диференциране макар и само като приближение:

$$(30) \quad \Delta M = \Delta \frac{D}{K} = \frac{D}{K} \left( \frac{\Delta D}{D} - \frac{\Delta K}{K} \right) = M \left( \frac{\Delta D}{D} - \frac{\Delta K}{K} \right).$$

От тази формула се вижда, че ако  $E(\Delta D) = E(\Delta K) = 0$ , не би трябвало да има систематично отклонение и в производния показател:

$$E(\Delta M) = 0.$$

В случай че някое от отклоненията  $\Delta D$  и  $\Delta K$  има ненулево математическо очакване, това би означавало, че в статистическите данни е налице систематично изместване на стойностите и системно, тенденциозно отклоняване на показателите  $D + \Delta D$  и  $K + \Delta K$  от действителните им стойности  $D$  и  $M$ .

За дисперсията на отклоненията (30) би могла да се получи следната формула (с  $Cov(.,.)$  е означен коефициентът на ковариация на две вариращи величини, посочени в скоби):

$$(31) \quad Var(\Delta M) = M^2 \left( \frac{1}{D^2} Var(\Delta D) - \frac{2}{DK} Cov(\Delta D, \Delta K) + \frac{1}{K^2} Var(\Delta K) \right).$$

Оценката на отделните елементи на (31) е проблематична поради спецификата на информационната база на демографската статистика, имаща по-други цели. Действително многообразието от производни показатели в демографската статистика включва предимно различни коефициенти или величини с подобна структура, което стои доста далече от възможността да се извлече адекватна и достоверна информация за вариацията на съответните величини. Във връзка с това неизбежно се налага привличането на различни хипотези за тези величини. Логично изглежда например положението (Keyfitz, 1968), че ковариацията (и корелацията) между данните за умрелите и изходните съвкупности на живите е нула, тъй като информацията идва от различни и несвързани помежду си източници. Освен това не изглежда лесно да се обоснове знак (плюс или минус) на съответната корелационна зависимост, дори и да се допусне, че такава съществува.

По-лесно изглежда да се оцени дисперсията на броя на умрелите, макар че движението на информацията за тези събития не изглежда съвсем перфектно, на което се спряхме в началото. Ако се възприеме гледната точка, че броят на умрелите варира предимно под действието на ендогенните фактори на демографските процеси, големината на дисперсията би се получила по модела на биномното разпределение и би била равна приблизително на самия брой на умрелите. Това може да се демонстрира по следния начин. Нека  $N$  представлява изходната съвкупност живи, от които произлизат умрелите  $D$ , и вероятността за умирање е  $q$ . Ако се възприеме моделът на биномното разпределение за броя на умрелите, неговата дисперсия би била  $Nq(1-q)$ . Множителят  $1-q$  представлява вероятността за преживяване и неговата величина е близо до сто процента, така че може да се изостави. Остава другата част от формулата  $Nq$ , която е точно равна на  $D$ .

При такова положение може да се допусне, че екзогенните фактори биха могли само да увеличат вариацията, без систематично да изместват показателите за смъртността в една или друга посока. В резултат на това би могло да се окаже,

че така се компенсира отпадането на множителя  $1 - q$  във формулата за дисперсията и дори нещо повече – получената по този метод оценка да остане под действителната стойност на дисперсията.

Най-трудният проблем е свързан с оценката на дисперсията на големината на съвкупностите на живите  $Var(\Delta K)$  като статистически показател. Данните от последните няколко преброявания на населението без изключение даваха пониски стойности за отделните възрастово-полови групи в сравнение с придвижването от държавния статистически орган население на база естествения и механичния прираст (т.нар. изчислено население). Такова разминаване е съвсем обяснимо, като се имат предвид пропуските на държавната статистика в обхвата на външната миграция, чието салдо изглежда е значителна величина с отрицателен знак през последните няколко десетилетия. Подобен извод в същата посока може да се направи, като се сравнява процентът съгласно социологическите проучвания на гласувалите на изборите през последните години и съответната величина, получена от съпоставянето на изборните протоколи с избирателните списъци.

Никак не е лесно да се обоснове съвсем приемлива хипотеза за разпределението на величината на  $\Delta K$ . В предходната част бе показано, че разлики от хиляда души често се срещат при сравняването дори само на единични възрастови групи от един и същ пол. Ако се предположи нормално разпределение (т.е. разпределение на Гаус-Лаплас) на такива отклонения и че вероятността за възникване на подобни и по-големи отклонения е само 0.05, то стандартното отклонение на  $\Delta K$  би трябвало да е около 500. Може би подобни отклонения се срещат по-често и ако се възприеме съответната вероятност да бъде 0.10, то стандартното отклонение би било малко над 600. Данните от табл. 9а разкриват разлики от по около 3000 души на генерация, съдейки по отклоненията за възрастовата група 1 до 4 навършени години, което подсказва за много по-голяма дисперсия и за стандартно отклонение близо до 2000.

Ситуацията обаче е малко по-сложна, тъй като констатираните отклонения в размера на възрастовите групи са повече в едната посока, което подсказва за асиметрично разпределение, доста по-различно от нормалното на Гаус-Лаплас на случайните грешки. В това отношение заслужава да се разгледа като хипотеза равномерното разпределение на отклоненията с математическо очакване, различно от нула, с което се допускат едноточност и тенденциозност в съответните величини. Трябва да се има предвид също, че равномерното разпределение в най-голяма степен толерира големи отклонения и особено в сравнение с нормалното на Гаус-Лаплас. При последното големи отклонения се срещат много рядко, докато при равномерното разпределение големи и малки величини се срещат еднакво често.

Вероятността да се получат отклонения, близки до двете крайни стойности на интервала на вариране на равномерното разпределение, е равна на относителния дял на съответните остатъци от този интервал. Ако се приеме, че 1000 е 95 %-ово отклонение в едната посока, ширината на интервала на вариране би трябвало да бъде малко над 1050. При 90 %-ови отклонения тази ширина би била над 1100. При възможност за двупосочни отклонения съответната ширина би била 1025.64 и 1052.63. Като се има предвид, че дисперсията на равномерно разпределение в интервал с ширина  $h$  е  $h^2/12$ , стандартното отклонение на тези разлики в

размера на възрастовите групи може да се очаква да бъде с големина  $h/\sqrt{12}$ , или малко над 300.

Привличането на равномерното разпределение като хипотетичен вариант в това разглеждане позволява да се определи и някаква минимална стойност за търсената големина на дисперсията в размера на възрастовите групи. При наблюдавани разлики от 1000 може да се очаква този лимит да бъде 300, но при разлики от 3000 стандартното отклонение би се оказало около 1000. В общи линии обаче, съдейки по разпределението на Гаус-Лаплас, би трябвало големината на стандартното отклонение да се очаква в границите между 1000 и 2000 за единичните възрастови групи. При размер на последните около 70 000 последното означава относителна грешка (по-точно – коефициент на вариация) от порядъка на 1 % до 3 %. Данните от табл. 9а сочат, че подобен процент може да се възприеме и за по-широки 4 и 5-годишни възрастови интервали, въпреки че по принцип може да се очаква, че при по-широки възрастови интервали стандартното отклонение не би било толкова пъти пропорционално по-голямо, тъй като отклонението (и най-вече систематичната грешка) за една възраст може да бъде за сметка на някоя от съседните (примерно при възрастова акумулация). Такава отрицателна корелация показва, че дисперсията на сбора на отделните единични възрастови групи би могла да се окаже по-малка от сбора на съответните дисперсии.

**Таблица 10**

Стандартно отклонение на коефициентите за смъртност за различни нива на коефициента на вариация при възрастовите групи на населението, България, 2001-2002 г.

| Възраст | Коефициенти за смъртност |         |         | Стандартно отклонение на коефициентите за смъртност при коефициент на вариация на възрастовите групи: |          |          |          |          |          |
|---------|--------------------------|---------|---------|-------------------------------------------------------------------------------------------------------|----------|----------|----------|----------|----------|
|         |                          |         |         | 1 %                                                                                                   |          |          | 3 %      |          |          |
|         | общо                     | мъже    | жени    | общо                                                                                                  | мъже     | жени     | общо     | мъже     | жени     |
| 1       | 2                        | 3       | 4       | 5                                                                                                     | 6        | 7        | 8        | 9        | 10       |
| 0       | 0.01404                  | 0.01523 | 0.01279 | 0.000480                                                                                              | 0.000684 | 0.000642 | 0.000623 | 0.000808 | 0.000737 |
| 1- 4    | 0.00069                  | 0.00076 | 0.00063 | 0.000053                                                                                              | 0.000076 | 0.000072 | 0.000056 | 0.000079 | 0.000074 |
| 5- 9    | 0.00028                  | 0.00030 | 0.00025 | 0.000028                                                                                              | 0.000040 | 0.000038 | 0.000029 | 0.000041 | 0.000038 |
| 10-14   | 0.00026                  | 0.00033 | 0.00018 | 0.000023                                                                                              | 0.000036 | 0.000028 | 0.000024 | 0.000038 | 0.000028 |
| 15-19   | 0.00051                  | 0.00069 | 0.00033 | 0.000031                                                                                              | 0.000051 | 0.000036 | 0.000035 | 0.000054 | 0.000037 |
| 20-24   | 0.00081                  | 0.00112 | 0.00048 | 0.000039                                                                                              | 0.000063 | 0.000042 | 0.000045 | 0.000070 | 0.000044 |
| 25-29   | 0.00089                  | 0.00126 | 0.00050 | 0.000040                                                                                              | 0.000067 | 0.000042 | 0.000047 | 0.000076 | 0.000044 |
| 30-34   | 0.00125                  | 0.00170 | 0.00079 | 0.000049                                                                                              | 0.000080 | 0.000055 | 0.000061 | 0.000093 | 0.000059 |
| 35-39   | 0.00195                  | 0.00261 | 0.00128 | 0.000065                                                                                              | 0.000104 | 0.000072 | 0.000085 | 0.000127 | 0.000080 |
| 40-44   | 0.00321                  | 0.00457 | 0.00185 | 0.000083                                                                                              | 0.000137 | 0.000084 | 0.000123 | 0.000189 | 0.000099 |
| 45-49   | 0.00524                  | 0.00750 | 0.00303 | 0.000110                                                                                              | 0.000180 | 0.000107 | 0.000184 | 0.000279 | 0.000137 |
| 50-54   | 0.00831                  | 0.01228 | 0.00458 | 0.000146                                                                                              | 0.000242 | 0.000132 | 0.000277 | 0.000423 | 0.000185 |
| 55-59   | 0.01213                  | 0.01787 | 0.00695 | 0.000197                                                                                              | 0.000327 | 0.000177 | 0.000396 | 0.000602 | 0.000264 |

## Продължение

| 1     | 2       | 3       | 4       | 5        | 6        | 7        | 8        | 9        | 10       |
|-------|---------|---------|---------|----------|----------|----------|----------|----------|----------|
| 60-64 | 0.01856 | 0.02688 | 0.01143 | 0.000278 | 0.000455 | 0.000249 | 0.000594 | 0.000886 | 0.000408 |
| 65-69 | 0.02724 | 0.03711 | 0.01924 | 0.000366 | 0.000566 | 0.000337 | 0.000853 | 0.001192 | 0.000640 |
| 70-74 | 0.04503 | 0.05752 | 0.03557 | 0.000567 | 0.000826 | 0.000540 | 0.001394 | 0.001825 | 0.001142 |
| 75-79 | 0.07436 | 0.08789 | 0.06518 | 0.000896 | 0.001227 | 0.000890 | 0.002286 | 0.002772 | 0.002047 |
| 80-84 | 0.12113 | 0.13524 | 0.11225 | 0.001542 | 0.002113 | 0.001624 | 0.003757 | 0.004370 | 0.003566 |
| 85+   | 0.23715 | 0.25029 | 0.22983 | 0.023785 | 0.025225 | 0.023092 | 0.041116 | 0.043466 | 0.039871 |
| Общо  | 0.01425 | 0.01563 | 0.01295 |          |          |          |          |          |          |

**Забележка:** по данни от публикациите „Население и демографски процеси“ и „Население“, НСИ, София, за съответните години.

В табл. 10 е показано до какви величини би могло да стигне стандартното отклонение при възрастовите коефициенти за смъртност в зависимост от дисперсията на двете му компоненти съобразно (29) и (31). Вижда се, че по-големи стойности има при началната възраст и при по-старшите възрасти, при които смъртността е по-висока. Ако се приеме, че вариацията на коефициентите за смъртност е подобна на нормалното разпределение на Гаус-Лаплас, може да се каже, че при началната възраст този показател би се движил в границите около плюс/минус един до един и половина пункта на хиляда (по две стандартни отклонения в двете посоки, съответно на 95 %-ите доверителни интервали) в повече или по-малко от стойността, получена по статистическите данни, или размах общо до три пункта.

## Таблица 10а

Вариация на коефициентите за смъртност за различни нива на коефициента на вариация при възрастовите групи на населението, България, 2001-2002 г.

| Възраст | Коефициенти на вариация на коефициентите за смъртност при коефициент на вариация на възрастовите групи: |          |          |          |          |          |
|---------|---------------------------------------------------------------------------------------------------------|----------|----------|----------|----------|----------|
|         | 1 %                                                                                                     |          |          | 3 %      |          |          |
|         | общо                                                                                                    | мъже     | жени     | общо     | мъже     | жени     |
| 1       | 2                                                                                                       | 3        | 4        | 5        | 6        | 7        |
| 0       | 0.034207                                                                                                | 0.044917 | 0.050213 | 0.044386 | 0.053081 | 0.057631 |
| 1-4     | 0.075827                                                                                                | 0.101000 | 0.113668 | 0.080930 | 0.104886 | 0.117135 |
| 5-9     | 0.099518                                                                                                | 0.132830 | 0.149406 | 0.103460 | 0.135808 | 0.152060 |
| 10-14   | 0.089124                                                                                                | 0.109566 | 0.151949 | 0.093505 | 0.113158 | 0.154559 |
| 15-19   | 0.061126                                                                                                | 0.073519 | 0.107985 | 0.067353 | 0.078772 | 0.111627 |
| 20-24   | 0.047735                                                                                                | 0.056239 | 0.087448 | 0.055486 | 0.062951 | 0.091908 |
| 25-29   | 0.045311                                                                                                | 0.052837 | 0.084956 | 0.053415 | 0.059932 | 0.089540 |
| 30-34   | 0.039348                                                                                                | 0.046890 | 0.069086 | 0.048458 | 0.054760 | 0.074652 |

Продължение

| 1     | 2        | 3        | 4        | 5        | 6        | 7        |
|-------|----------|----------|----------|----------|----------|----------|
| 35-39 | 0.033061 | 0.039658 | 0.056114 | 0.043509 | 0.048711 | 0.062839 |
| 40-44 | 0.025902 | 0.030048 | 0.045566 | 0.038353 | 0.041266 | 0.053631 |
| 45-49 | 0.020927 | 0.024056 | 0.035347 | 0.035185 | 0.037131 | 0.045271 |
| 50-54 | 0.017513 | 0.019716 | 0.028772 | 0.033267 | 0.034478 | 0.040346 |
| 55-59 | 0.016267 | 0.018316 | 0.025434 | 0.032628 | 0.033697 | 0.038038 |
| 60-64 | 0.014987 | 0.016926 | 0.021805 | 0.032009 | 0.032962 | 0.035713 |
| 65-69 | 0.013441 | 0.015240 | 0.017513 | 0.031315 | 0.032129 | 0.033267 |
| 70-74 | 0.012589 | 0.014361 | 0.015171 | 0.030960 | 0.031721 | 0.032096 |
| 75-79 | 0.012052 | 0.013956 | 0.013661 | 0.030745 | 0.031540 | 0.031411 |
| 80-84 | 0.012734 | 0.015625 | 0.014465 | 0.031018 | 0.032313 | 0.031769 |
| 85+   | 0.012620 | 0.016032 | 0.013970 | 0.030972 | 0.032512 | 0.031546 |

**Забележка:** по данни от табл. 10.

Подобни по размер отклонения са възможни и за възрастите след шестдесет години и особено при най-старшите възрасти. Много особено положение се наблюдава при последния възрастов интервал 85+, при който стандартното отклонение е толкова голямо, че съответният коефициент за смъртност се оказва с максимум две значещи цифри след десетичната точка.

Стандартното отклонение при възрастовите коефициенти за смъртност е малко по-ниско при жените в сравнение с мъжете, но тъй като и смъртността е по-ниска при женския пол, на един пункт смъртност при жените се пада малко по-висок дял вариация (коефициент на вариация) в сравнение с мъжете (табл. 10а). Тази разлика е по-ярка при 3 % коефициент на вариация на възрастовите групи в сравнение с 1 %. Както може да се очаква, при двата пола общо стандартното отклонение е по-ниско в сравнение с двата пола отделно поради приблизително двукратно по-големия размер на съответните съвкупности.

Може да се направи интересно сравнение с ендегенната вариация в самите демографски процеси и ефекта ѝ върху съответните показатели. В Чанг (1961) е показано, че между дисперсията на коефициентите за смъртност и пропорциите на доживяващите (или умрелите) съществува следната проста (макар и приблизителна като апроксимация, но съвсем приемлива) връзка, която може да се използва в случая:

$$(32) \quad \text{Var}(q) = n^2 \text{Var}(M),$$

като  $n$  представлява ширината на възрастов интервал.

Таблица 3 показва, че за началната възраст под една година 95 %-ят интервал на вариране е с ширина около 1.6 пункта на хиляда (при 99 %-я интервал на вариране е с ширина около 2.1 пункта на хиляда), което представлява т.нар. естествена дисперсия на този показател. Данните от табл. 10 разкриват, че екзогенните фактори за вариацията в статистическите данни може да генерират приблизи-

телно още толкова допълнителна неопределеност и вариация, колкото е „естествената“ дисперсия за тази възраст.

По-различна е ситуацията при следващите възрасти, тъй като липсва непосредствената „котва“ на живородените, данните за които може да служат за сравняване и верификация на демографските показатели. За възраст 60 например пропорцията на доживяващите се движи в границите плюс/минус 2.5 до 3.2 пункта на хиляда (табл. 3), което отговаря на стандартно отклонение с размер 1.2 до 1.6 пункта. Дисперсията на съответния коефициент за смъртност би трябвало (съгласно формула (32)) да бъде двадесет и пет пъти по-ниска ( $n = 5$ ), т.е. стандартното отклонение да бъде пет пъти по-малка величина, което прави около 0.3 пункта на хиляда. Таблица 10 показва обаче (за същата възраст 60), че с добавяне на влиянието от външните фактори стандартното отклонение достига 0.5 до 0.9 пункта на хиляда, което прави размах на вариацията плюс/минус 1 до 2 пункта. Така че с включването на ефекта на екзогенните фактори увеличението представлява два до три пъти в сравнение с „естественото“ стандартно отклонение. Тези сравнения са представени в табл. 11 (формулата на Чанг (32) не е приложена за последния възрастов интервал, който е неограничен и величината на  $n$  е неопределена).

**Таблица 11**

Ширина на 95 %-ови интервали на вариране  
на вероятностите за умиране и пропорциите на доживяващите

| Възраст | При коефициент на вариация на възрастовите групи: |          |          |          |          |          |
|---------|---------------------------------------------------|----------|----------|----------|----------|----------|
|         | 1%                                                |          |          | 3%       |          |          |
|         | общо                                              | мъже     | жени     | общо     | мъже     | жени     |
| 1       | 2                                                 | 3        | 4        | 5        | 6        | 7        |
| 0       | 0.001883                                          | 0.002681 | 0.002517 | 0.002443 | 0.003168 | 0.002888 |
| 1       | 0.000826                                          | 0.001197 | 0.001122 | 0.000881 | 0.001243 | 0.001156 |
| 5       | 0.000542                                          | 0.000788 | 0.000736 | 0.000563 | 0.000806 | 0.000749 |
| 10      | 0.000452                                          | 0.000714 | 0.000540 | 0.000474 | 0.000737 | 0.000549 |
| 15      | 0.000616                                          | 0.000990 | 0.000704 | 0.000679 | 0.001061 | 0.000727 |
| 20      | 0.000755                                          | 0.001233 | 0.000820 | 0.000878 | 0.001380 | 0.000862 |
| 25      | 0.000787                                          | 0.001305 | 0.000827 | 0.000928 | 0.001480 | 0.000871 |
| 30      | 0.000966                                          | 0.001565 | 0.001069 | 0.001190 | 0.001827 | 0.001155 |
| 35      | 0.001264                                          | 0.002032 | 0.001408 | 0.001664 | 0.002495 | 0.001576 |
| 40      | 0.001629                                          | 0.002691 | 0.001653 | 0.002412 | 0.003696 | 0.001946 |
| 45      | 0.002148                                          | 0.003537 | 0.002102 | 0.003611 | 0.005460 | 0.002693 |
| 50      | 0.002854                                          | 0.004744 | 0.002585 | 0.005420 | 0.008295 | 0.003625 |
| 55      | 0.003868                                          | 0.006414 | 0.003466 | 0.007759 | 0.011800 | 0.005183 |
| 60      | 0.005451                                          | 0.008918 | 0.004884 | 0.011641 | 0.017368 | 0.008000 |
| 65      | 0.007176                                          | 0.011085 | 0.006605 | 0.016720 | 0.023369 | 0.012546 |



Продължение

| 1  | 2        | 3        | 4        | 5        | 6        | 7        |
|----|----------|----------|----------|----------|----------|----------|
| 70 | 0.011111 | 0.016190 | 0.010576 | 0.027324 | 0.035761 | 0.022374 |
| 75 | 0.017565 | 0.024040 | 0.017453 | 0.044808 | 0.054329 | 0.040130 |
| 80 | 0.030231 | 0.041417 | 0.031825 | 0.073641 | 0.085653 | 0.069893 |
| 85 | ---      | ---      | ---      | ---      | ---      | ---      |

**Забележка:** по данни от табл. 10.

Интересен въпрос е как всичко това би могло да се отрази върху величината на средната продължителност на предстоящия живот. Формулата на Чанг (16) не би помогнала много в случая, защото екзогенно предизвиканите отклонения във величината на този показател едва ли са независими като случайни величини и едва ли отделните фази в стареенето на поколенията остават независими една от друга. Същото се отнася и за броя на достигащите до отделните възрасти. Методите на изглаждане, изравняване и попъване на статистическите редове в общи линии привнасят положителна корелация в данните. Непрецизно прилагане на различни аналитични методи би могло да доведе до натрупване на грешки и заедно с това да се възбуди също положителна корелация в резултатите за отделните възрасти. Други корекции на данните, особено за преодоляване на възрастовата акумулация, биха могли да действат в обратна посока с пораждаване на отрицателна корелация поради компенсациите за съседни възрастови групи. Формула (23) също не е много обещаваща по аналогични причини, но би могло с нейна помощ да се хвърли известна светлина върху въпроса, макар и с много условности и допълнителни хипотези.

Лесно се открива, че основният множител в тази формула е дисперсията на пропорциите на доживяващите, в резултат на което формулата придобива следния вид:

$$(33) \quad \text{Var}(E_x) = 2 \sum_{t=1}^{\infty} \text{Var}\left(\frac{K_{x+t}}{K_x}\right) e_{x+t}.$$

**Таблица 12**

Дисперсия и стандартно отклонение на средната продължителност на предстоящия живот, 1999-2001 г.

| Взраст | Мъже                                        |           |                       | Жени                                        |           |                       |
|--------|---------------------------------------------|-----------|-----------------------|---------------------------------------------|-----------|-----------------------|
|        | средна продължителност на предстоящия живот | дисперсия | стандартно отклонение | средна продължителност на предстоящия живот | дисперсия | стандартно отклонение |
| 1      | 2                                           | 3         | 4                     | 5                                           | 6         | 7                     |
| 0      | 68.534                                      | 0.009433  | 0.097122              | 75.231                                      | 0.008141  | 0.090226              |
| 1      | 68.546                                      | 0.006507  | 0.080663              | 75.142                                      | 0.005157  | 0.071815              |

## Продължение

| 1  | 2      | 3        | 4        | 5      | 6        | 7        |
|----|--------|----------|----------|--------|----------|----------|
| 5  | 64.744 | 0.006129 | 0.078289 | 71.345 | 0.004744 | 0.068877 |
| 10 | 59.857 | 0.005989 | 0.077388 | 66.439 | 0.004583 | 0.067694 |
| 15 | 54.950 | 0.005899 | 0.076805 | 61.509 | 0.004518 | 0.067214 |
| 20 | 50.160 | 0.005770 | 0.075963 | 56.636 | 0.004425 | 0.066520 |
| 25 | 45.432 | 0.005601 | 0.074838 | 51.778 | 0.004317 | 0.065707 |
| 30 | 40.705 | 0.005459 | 0.073885 | 46.909 | 0.004230 | 0.065037 |
| 35 | 36.060 | 0.005315 | 0.072904 | 42.098 | 0.004115 | 0.064152 |
| 40 | 31.510 | 0.005135 | 0.071662 | 37.341 | 0.003960 | 0.062932 |
| 45 | 27.186 | 0.004927 | 0.070194 | 32.695 | 0.003813 | 0.061753 |
| 50 | 23.155 | 0.004710 | 0.068627 | 28.167 | 0.003646 | 0.060386 |
| 55 | 19.439 | 0.004447 | 0.066687 | 23.757 | 0.003485 | 0.059033 |
| 60 | 16.022 | 0.004129 | 0.064256 | 19.516 | 0.003306 | 0.057495 |
| 65 | 12.941 | 0.003726 | 0.061042 | 15.556 | 0.003128 | 0.055926 |
| 70 | 10.142 | 0.003595 | 0.059958 | 11.915 | 0.003055 | 0.055272 |
| 75 | 7.740  | 0.003586 | 0.059885 | 8.806  | 0.003077 | 0.055467 |
| 80 | 5.738  | 0.004112 | 0.064124 | 6.283  | 0.003389 | 0.058216 |
| 85 | 3.997  | 0.005091 | 0.071353 | 4.320  | 0.003827 | 0.061861 |

**Забележка:** по данни от „Население и демографски процеси – 2001“, НСИ, 2003. Дисперсията и стандартното отклонение са получени при коефициент на вариация на възрастовите групи на населението 3 %.

Дисперсията на пропорцията на доживяващите може да се получи от съответната на възрастовите коефициенти за смъртност по формула (32) (по единични възрасти за по-голяма точност на апроксимацията), като се използва формула (31). На табл. 12 са показани съответните резултати отделно за двата пола по данни от таблица за смъртност на българското население за периода 1999-2001 г. (при коефициент на вариация на възрастовите групи 3 %) Сравнението с данните от табл. 5 (за мъжкия пол) показва почти двойно увеличение на стандартното отклонение (а при по-старшите възрасти и повече) след отразяване на екзогенните фактори за вариацията. Границите, в които се очаква (с 95 %-ова вероятност) да се движи показателят средна продължителност на предстоящия живот, се оказват два-три месеца в повече или по-малко от онова, което би могло да се получи по статистическите данни, без това да означава фактическа разлика във възрастовата смъртност.

Дори и при нулев коефициент на вариация на възрастовите групи стандартното отклонение остава с размер около един месец за най-младшите възрасти, което показва, че общата продължителност на живота би могла да варира в плюс/минус два месеца. Обратно, при по-висок коефициент на вариация на възрастовите групи размахът на вариацията в общата продължителност на живота може да достигне значителни размери. Така например при коефициент на вариация

ция на възрастовите групи от 13 % общата продължителност на живота би могла да варира в граници плюс/минус шест месеца, т.е. общ размах от една година.

## 7. ЗАКЛЮЧЕНИЕ

В разглежданата проблематика се очертават три основни направления, в които може да се търсят причините и механизъмът на действие и въздействие върху подвижността на показателите на таблиците за смъртност и размера на вариацията им. Първото направление е в рамките на самите таблици за смъртност, изглежда като ендемичен демографски феномен и се проявява в хода на стареенето на условното поколение, в неговия постепенно стопяващ се размер, а оттам и в останалите показатели на таблиците за смъртност. В тази вариация обаче има нещо доста механично, породено от модела на таблиците за смъртност, при който фактически се имитира стигането до поредните възрасти. В подобно построение ролята на условното поколение е повече от скромна, нещо като строителен материал за пресъздаване на възрастовия профил на смъртността и ефекта ѝ върху продължителността на живота. При такова положение нищо чудно няма във факта, че големината на дисперсията на различните показатели зависи от първоначалния брой живородени, определящ мащаба на условното поколение, а с това и статистическата тежест и представителност на таблиците за смъртност като комплекс от взаимно свързани величини. Както бе показано, изборът на първоначалния брой живородени обаче не е никак безобиден, тъй като мащабът на таблиците за смъртност би могъл да маскира и дори да изопачи влиянието на фактори, идващи по другите две направления.

С повече основание може да се счита, че второто направление е ендемично от демографска гледна точка: това е ефектът от вариацията в действителните населения и техните съвкупности, пренесен върху показателите на таблиците за смъртност. В това отношение мащабът на действителните населения и големината на техните съвкупности и структури като източник на демографска информация имат значение за стабилността на специфичните коефициенти за смъртност и съответните показатели от таблиците за смъртност и този мащаб не е подвластен на изследователя и не може да се манипулира по начина, по който се определя началото на таблиците за смъртност.

Третото направление е може би най-същественото за релевантността, информативността и полезността на българската демографска статистика. То е в известен смисъл екзогенно по отношение на населението и е свързано с действието на различни причини и фактори, в т.ч. и систематични пропуски и грешки, между които непрецизност в обработката на информацията, смяна на методологията за получаване на показателите с недостатъчен сравнителен анализ и стикване на динамичните редове, непълнота и пропуски в регистрацията и преброяванията, междуведомствена разпокъсаност и други организационни неудачи при движението на информацията.

Количествената оценка на резултата от въздействието по това направление е затруднена по ред причини. Най-напред заслужава да се отбележи принципната непригодност на държавната статистика (въобще, не само у нас) да осъществява рефлексия и самооценка на качеството, информативността и полезността на

произвежданата информация. През последните десетилетия обаче се забелязват известно раздвижване в демографската област (американската, канадската и скандинавските статистики) в това отношение и опит да се разчупят традиционните схеми, което заслужава да бъде внимателно изучено.

Механизмът на въздействие по това направление е съвсем разнообразен и не е лесен за идентифициране и квалифициране, още по-малко и за квантифициране. Някои от тези ефекти може да имат само случаен характер и да се отразяват предимно в големината на дисперсията на демографските показатели, други пък съвсем системно да водят до деформация в статистическата информация – например най-малкото до изместване на средните величини и математическото очакване на показателите. Подобна неизвестност и неопределеност налагат привличане на допълнителни хипотези и построения и така разглеждането се натовазва с нови условности.

Друга трудност произтича от преплитането на въздействието по това направление с предходните две направления. Особено когато се търси отпечатъкът на този ефект като краен резултат върху показателите на таблиците за смъртност, преобразуван, пречупен и лимитиран от мащаба на условното поколение. Това налага различни гледни точки и към това сме се придържали в изложението, за да може проблемите да се осветлят от различни страни. Макар че получените резултати трябва да се възприемат повече като ориентировъчни, констатира се, че екзогенните фактори водят до увеличение два до четири пъти на вариацията на възрастовите показатели на таблиците за смъртност в сравнение с ендегенния процес на стареене на условното поколение. Аналогичен по размер е размахът на вариацията и при средната продължителност на живота, въпреки че като агрегатна величина този показател би трябвало да бъде по-стабилен и по-устойчив на външни влияния. Изглежда, че натрупването на отклонения и грешки в тази величина надделява над взаимното погасяване на такива с противоположни знаци.

Някои несъответствия и диспропорции в статистическите данни може да е трудно да се приемат като случайни отклонения. Подобен феномен очевидно е встрани от статистиката и неговото изучаване се нуждае от по-друг инструментариум.

## Литература

1. Chiang, C. (1960), A Stochastic Study of the Life Table and its Applications: II. Sample Variance of the Observed Expectation of Life and other Biometric Functions, *Human Biology*, XXXII: 221-238.
2. Chiang, C. (1961), Standard Error of the Age-Adjusted Death Rate. *Vital Statistics, Special Reports*, Vol. 47, No 9:271-285, Washington, DC : National Office of Vital Statistics.
3. Chiang, C. (1968), *Introduction to Stochastic Processes in Biostatistics*, New York: John Wiley and Sons.
4. Chiang, C. (1978), *Life Table and Mortality Analysis*, Geneva: World Health Organization.
5. Dublin, L., A. Lotka, (1936), *Length of Life*, New York: Ronald Press Company.
6. Feichtinger, G. (1972), *Stochastische Modelle demographischer processen*, Berlin: Springer.
7. Golemanov, N., E. Christov (2001), Bulgaria: Mortality of the Elderly (Trends and Perspectives). In: *Workshop on Ageing, Intergenerational Solidarity, and Age-Specific Vulnerabilities*, Rome.
8. Gompertz, B. (1825), On the Nature of Function Expressive of the Law of Human Mortality, *Philosophical Transactions* 27.

9. Hansen, M., W. Hurwitz, W. Madow (1953), *Sample Survey Methods and Theory* New York: Wiley.
10. Hoem, J. (2002), *Demographic Analysis: Probabilistic Approach* 5: 3428-32; Life Table 13: 8832-6, *International Encyclopedia of the Social & Behavioral Sciences*, под редакция на N. Smelser, P. Baltes, Oxford: Elsevier Science.
11. Hoem, J. (1969), Markov chain models in life insurance. *Blätter Deutsche Gesellschaft für Versicherung math*, 9 (2), 91-107.
12. Hoem, J. (1976), The statistical theory of demographic rates: A review of current developments (with discussion), *Scandinavian Journal Statistics* 3 (4), 169-185.
13. Keyfitz, N. (1968), *Introduction to the Mathematics of Population*, Reading, Mass: Addison-Wesley.
14. Keyfitz, N. (1978), *Applied Mathematical Demography*, New York: Wiley & Sons.
15. Kish, L. (1965), *Survey Sampling*, New York: Wiley.
16. Makeham, W. (1867), On the Law of Mortality, *Journal of the Institute of Actuaries* 13.
17. Preston, S., N. Keyfitz, R. Schoen (1972), *Causes of Death. Life Tables for National Population*, New York: Seminar Press.
18. Reed, L., Merrell, M. (1939), A Shot Method for Constructing an Abridged Life Table, *American Journal of Hygiene*, 30.
19. Smith, F. (1948), The force of mortality function, *The American Mathematical Monthly* 55(5), 277-284.
20. Statistics Canada (2002), *Life tables Canada, Provinces and Territories 1995-1997*, Health Statistics Division, Cat. No. 84-537-XIE, Ottawa, Ontario.
21. Tabah, L. (1968), Matrix Representation of Changes in the Active Population, *Population* 23, 437-470.
22. US Department of Health & Human Services (1997). *U.S. Decennial Life Tables for 1989-1991, Volume 1, Number 1 & 2*, National Center for Health Statistics, Hyattsville, Maryland.
23. Големанов, Н. (2001), *Пол и смъртност (България, 1971-2000)*, Академично издателство „Проф. М. Дринов“, София.
24. Сугарев, З. (1975), *Демографска статистика*, „Наука и изкуство“, София.
25. Чолаков, Н. (1991), Методологически и методически проблеми при статистическото изучаване на смъртността, *Научни трудове „Демография“*, кн. 1, УИ „Стопанство“, София.
26. Чолаков, Н. (2003а), Върху една хипотеза на Кардано за средната продължителност на предстоящия живот, *Статистика* № 2, Национален статистически институт, София.
27. Чолаков, Н. (2003б), Математически методи и модели в животозастраховането и пенсионното осигуряване, *Тракия-М*, София.
28. Чолаков, Н., Хр. Малешков (2003), Статистическата информация за естественото движение на населението и миграцията в България и някои насоки за нейното усъвършенстване, В сб. *Научни трудове на УНСС*, № 2, София.

## **СЛУЧАЙНИ ОТКЛОНЕНИЯ И СИСТЕМАТИЧНИ ГРЕШКИ И ВЪЗДЕЙСТВИЕТО ИМ ВЪРХУ ПОКАЗАТЕЛИТЕ НА ТАБЛИЦИТЕ ЗА СМЪРТНОСТ**

### **Резюме**

Движението на демографската статистическа информация у нас изглежда излишно удължено и усложнено – от съобщенията за смърт, през различни общински и централни административни звена, гражданския регистър за административно обслужване до националния статистически орган. Подобно прехвърляне на информацията на различни носители и през различни ведомства би могло да се окаже предпоставка за пропуски и допускане на грешки и неточности от разнообразно естество, което в крайна сметка би могло да се отрази по някакъв начин и на най-аналитичните показатели като тези в таблиците за смъртност. Не бива да се подценява и рискът от невнимателно, неточно или неправилно прилагане на съответните методики за обработка на информацията.

В студиума се разглеждат проблемите по случайни отклонения и систематични грешки и се търсят решения в по-широк план за въздействието върху показателите на таблиците за смъртност, което се определя не само от стохастиката в стареенето на условното поколение и въобще на демографските процеси, но и от разнообразни външни за демографията фактори.

В хода на изследването е показано как някои междинни резултати биха могли да се използват при сравнителния анализ на смъртността по данни от таблици за смъртност с отчитане на вариацията във величината на показателите, породена от разнообразните фактори.

## **RANDOM DEVIATIONS AND SYSTEMATIC ERRORS AND THEIR IMPACT ON LIFE TABLE INDICATORS**

### **Summary**

Statistical data processing of demographic information in Bulgaria seems to be excessively elongated and complicated - from the death reports through different local and central administrative units, the population register up to the national statistical authority. Such a transportation and exchange of information on different carriers through different departments might be a premise for errors and discrepancies of diverse nature which might affect even the most analytical indicators like those of the life tables. The risk of accidental, inaccurate or incorrect application of the corresponding methodologies of data processing should not be underestimated too.

The problems of random deviations and systematic errors are examined in the study and a wide range of solutions are sought for the impact on the life table indicators both within the stochastics of the life table generation and the demographic processes in general as well as external to demography factors.

In the process it is shown how some intermediate results could be used for comparative analysis of mortality using life table data bearing in mind the variability of the indicators due to different factors.