MPIDR-NES Training Programme

Moscow, New Economic School, 14th January - 1st February 2013

Population and Health

Лекция 8. Методы декомпозиции. Lecture 8. Methods of decomposition.

MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG

РЭШ Российская экономическая школа

Outline of Lecture 8: Methods of decomposition

❖ The problem of decomposition and how to approach it

❖ First example: a difference between crude death rates and the Kitagawa's decomposition by mortality and population age structure

❖ Second example: a difference between life expectancies and its decomposition by age-specific mortality

❖ The general algorithm of stepwise replacement and how it applies to the age decomposition of a difference between LEs

❖ Life expectancy: continuous versions of the method for decomposition.

❖A general algorithm of stepwise replacement decomposition for the multidimensional case.

❖How many replacements should we do?

❖Decomposition by age based on a stepwise replacement: algorithm running from young to old ages.

❖Some problems of decomposition base on stepwise replacement.

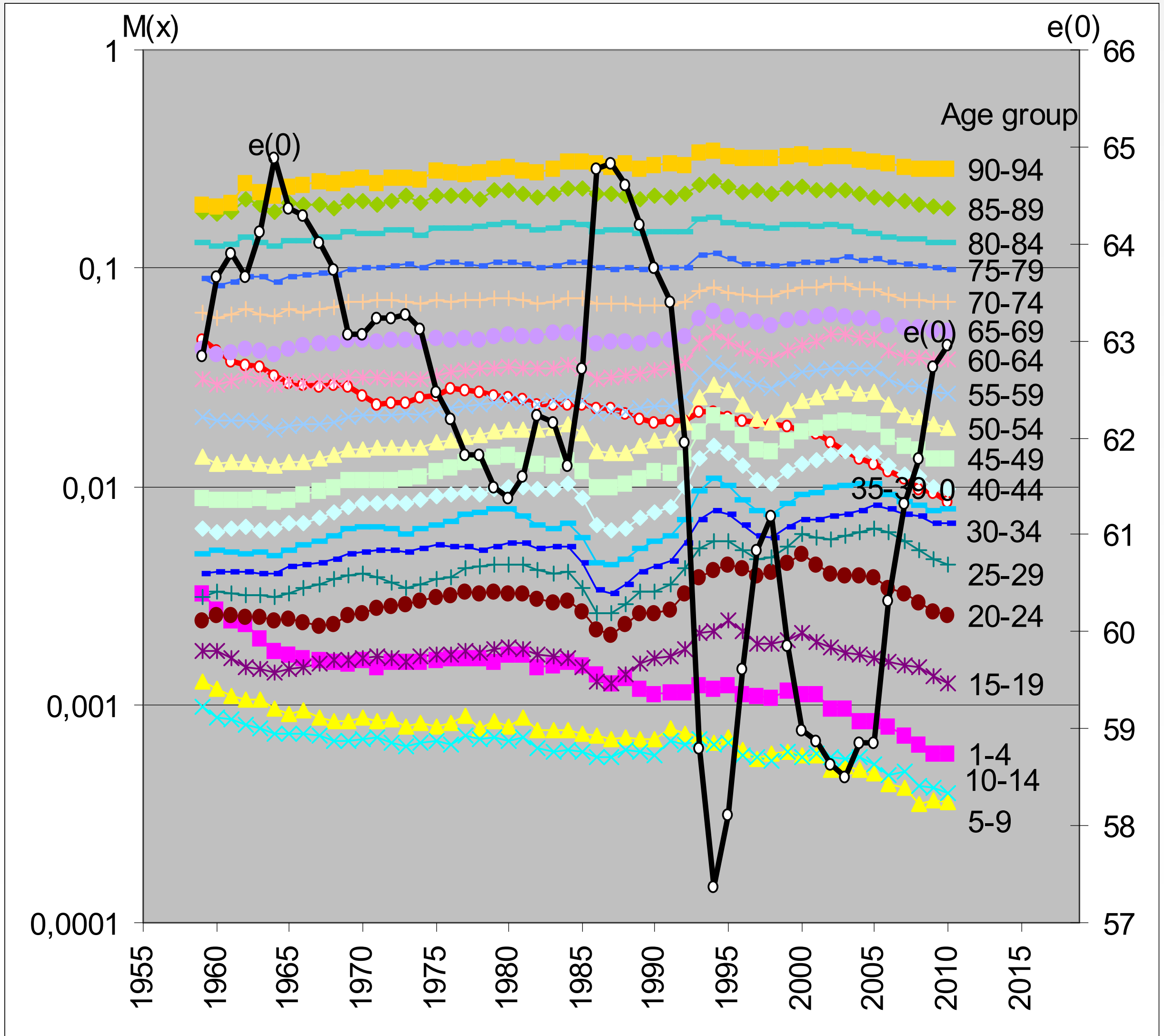
❖Third example: a difference between two health expectancies and its decomposition by age-specific mortality and age-specific health.

❖ Forth example: a difference between two life expectancies and its decomposition by age-cause-specific mortality

❖ Fifth example: a difference between life expectancies and its decomposition by group-specific mortality and population structure

Introduction

Dynamics age-specific death rates and life expectancy at birth in Russia (Male)



Each aggregate demographic measure combines a vector or a matrix of elementary rates of demographic events into one number. When analyzing changes in an aggregate demographic measure in time or its variations across countries, it is useful to be able to decompose observed changes or differences by age and other demographic dimensions such as cause of death, or population group. Decomposition aims at estimating contributions of differences between elementary rates of demographic events to the overall difference between two values of the aggregate measure.

The decomposition problem

Let *T* be an aggregate health and/or mortality measure depending on *n* variables

$$T(\theta_1, \theta_2, \ \dots \ , \theta_n)$$

Consider two values of *T* corresponding to two points in *n*-dimensional space

$$T^1 = T(\theta_1^1, \theta_2^1, \ \dots \ , \theta_n^1)$$
$$T^2 = T(\theta_1^2, \theta_2^2, \ \dots \ , \theta_n^2)$$

These two positions can be related to a change in the *teta*-values over time and/or space, and/or sex, and/or population group etc. A researcher might want to know to what extent the total difference between T^1 and T^2 is determined by each of the *n*-changes in each of the *n* variables

$$T^2 - T^1 = \sum_i \delta_i^{2-1}(\theta_i^{2-1})$$

, where each component $\delta_i^{2-1}(\theta_i^{2-1})$ corresponds to a contribution of the elementary change $\theta_i^1 \rightarrow \theta_i^2$

How to approach the problem. Just an idea.

The sum in the right hand side of the last equation is a decomposition of the difference T^1-T^2

The most common approach to this problem is very similar to the method of *standardization* (or *replacement*).

Imagine that T depends on only two dimensions. Then a transition from T^1 to T^2 can be decomposed as follows

$$T^2 - T^1 = T(\theta_1^2, \theta_2^2) - T(\theta_1^1, \theta_2^1) = [T(\theta_1^2, \theta_2^1) - T(\theta_1^1, \theta_2^1)] +$$

$$+ [T(\theta_1^2, \theta_2^2) - T(\theta_1^2, \theta_2^1)]$$

I only added and subtracted terms $T(\theta_1^2, \theta_2^1)$

The first additive term is a contribution of change in *teta1*, the second term is a contribution of the change in *teta2*.

How to approach the problem. Just an idea.

From other hand it is possible added and subtracted terms $T(\theta_1^1, \theta_2^2)$

$$T^2 - T^1 = T(\theta_1^2, \theta_2^2) - T(\theta_1^1, \theta_2^1) = [T(\theta_1^1, \theta_2^2) - T(\theta_1^1, \theta_2^1)] +$$

$$+ [T(\theta_1^2, \theta_2^2) - T(\theta_1^1, \theta_2^2)]$$

In this case the first additive term is a contribution of change in *teta2*, the second term is a contribution of the change in *teta1*. It is very possible that

$$[T(\theta_1^1, \theta_2^2) - T(\theta_1^1, \theta_2^1)] \neq [T(\theta_1^2, \theta_2^2) - T(\theta_1^2, \theta_2^1)]$$

$$[T(\theta_1^2, \theta_2^1) - T(\theta_1^1, \theta_2^1)] \neq [T(\theta_1^2, \theta_2^2) - T(\theta_1^1, \theta_2^2)]$$

First example: Crude Death Rate (CDR)

In 1991-2004 according HMD CDR in Germany declined from 11.39 to 9.93 per 1000. In general

$$CDR^{1991} = \sum_x M_x^{1991} \cdot \theta_x^{1991} \qquad CDR^{2004} = \sum_x M_x^{2004} \cdot \theta_x^{2004}$$

where x – age, M_x are age-sex specific death rates, and θ_x are age-sex specific population weights. Between 1991 and 2004 both mortality and age-sex structure changed. Formally, we can make 4 calculations of CDRs based on 4 different combinations of mortality (in 1991 or 2004) and age-sex structure (in 1991 or 2004):

Changes only in population structure

Structure \ Mortality	1991	2004	Difference
1991	**11.39**	8.41	-2.99
2004	13.35	**9.93**	-3.42
Difference	1.96	1.52	**-1.46**

Changes only in mortality

Hypothetical variants are marked with blue. Thus we can see that dynamics in the population structure (ageing) led to the increase in CDR, while changes in mortality led to the CDR decrease.

The Kitagawa's decomposition formula for CDR

There are two possible paths for estimation of the role of the two factors: the 'green path' yields -2.99+1.52=-1.46, whereas the 'red path' yealds -3.42+1.96=-1.46. The values of the first (mortality) component and the second (structural) component slightly differ between the two paths.

Mortality / Structure	1991	2004	Difference
1991	11.39	8.41	-2.99
2004	13.35	9.93	-3.42
Difference	1.96	1.52	-1.46

Evelyn M. Kitagawa (1955) proposed a good compromise: to use for calculation average mortality and average population structure

$$CDR^2 - CDR^1 = \sum_x (M_x^2 \cdot \theta_x^2 - M_x^1 \cdot \theta_x^1) =$$

$$= \underbrace{\sum_x (M_x^2 - M_x^1) \cdot \frac{(\theta_x^1 + \theta_x^2)}{2}} + \underbrace{\sum_x \frac{(M_x^2 + M_x^1)}{2} \cdot (\theta_x^2 - \theta_x^1)}$$

The mortality component
In our example it equals -3.20

The structural component
In our example it equals 1.74

Kitagawa's formula is equivalent to averaging values of each component over the two paths.

Second example: Life expectancy

In 1967, a Ukrainian researcher Yuri Korchak-Chepurkovskhiy proposed a general idea for decomposition of life expectancy.

The real method for decomposition of a difference between two life expectancies was independently developed in the 1980s by three different researchers from Russia (Andreev, 1982), the USA (Arriaga, 1984), and France (Pressat, 1985). The formulae for decomposition by Andreev and Pressat are exactly equivalent.

Arriaga's formula is written in a slightly different form, but it is essentially equivalent to the formulae by Andreev and Pressat but Arriaga did not transform his formula to symmetrical form.

A continuous version of the method for decomposition of differences between life expectancies by age was developed by Pollard (1982).

Second example: Life expectancy (LE). (2)

LE at birth in cohort of Swedish women born at 1870 was 51.3 years. In daughters' cohort born at 1900, LE was 10 years greater and equaled 61.3 years. It would be good to estimate contribution different ages to this growth. If $l_0 = 1$ then LE at birth can be written as $e_0 = {}_xL_0 + l_x e_x$ where ${}_xL_0 = T_0 - T_x$. If mortality at ages < x does not change, then LE in the second cohort would be $e^2_{0|x} = {}_xL^2_0 + l^2_x e^1_x$. Thus, contribution of age interval $(x,\ \omega)$ to LE growth can be expressed as

$$\delta^{2-1}_{x+} = e^2_0 - e^2_{0|x} = l^2_x (e^2_x - e^1_x)$$

The contribution of elementary age interval $[x, x+1)$ can be expressed as

$$\delta^{2-1}_x = l^2_x (e^2_x - e^1_x) - l^2_{x+1} (e^2_{x+1} - e^1_{x+1}) \quad \textbf{(1a)}$$

Thus, the overall difference between two life expectancies is $e^2_0 - e^1_0 = \sum_{x=0}^{\omega-1} \delta^{2-1}_x$

In a similar way, one can decompose the difference $e^1_0 - e^2_0$

$$\delta^{1-2}_x = l^1_x (e^1_x - e^2_x) - l^1_{x+1} (e^1_{x+1} - e^2_{x+1}) \quad \textbf{(1b)}$$

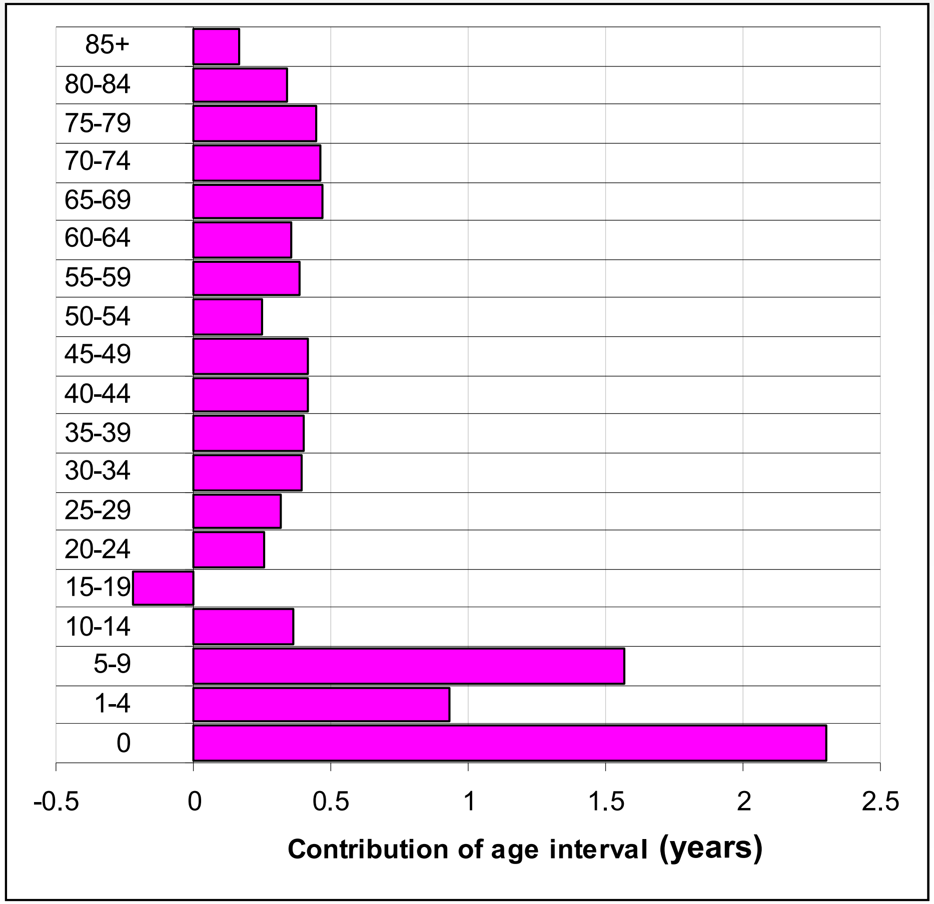
Formulas 1a and 1b are the ones derived by Andreev, Arriaga, and Pressat.

Second example: Life expectancy. Continuation.

Usually $\delta_x^{1-2} \neq \delta_x^{1-2}$ and (as in the case of CDR) it is natural to use their average to obtain symmetric age components $\delta_x = \frac{1}{2} \cdot (\delta_x^{2-1} - \delta_x^{1-2})$

The calculation result is presented on the left panels of this slide. More than half of the LE growth is concentrated at ages 0-14.
The impacts of every 5-year age group from 20 to 80 are less than 0,5 years.

Mortality at age 15-19 even grew that reduced LE on 0.2 years. This growth is probably connected with maternal mortality.



Decomposition_LE

Life expectancy: continuous versions of the method for decomposition.

In case of infinitesimal age interval the formulae (1a) can be written Andreev (1982)

$$\varepsilon_{x,x+\Delta x} = l_x^2 \cdot (\mu_x^1 - \mu_x^2) \cdot e_x^1$$

Let us show that this is exactly the same that the Pollard (1982) formula.
The main **Pollard's** formula is

$$e_0^2 - e_0^1 = \int_0^\infty (\mu_x^1 - \mu_x^2) \cdot \exp\left\{\int_0^x (\mu_u^1 - \mu_u^2) du\right\} \cdot l_x^1 \cdot e_x^1 dx,$$

or using the exponential formula for l_x

$$e_0^2 - e_0^1 = \int_0^\infty (\mu_x^1 - \mu_x^2) \cdot \frac{l_x^2}{l_x^1} \cdot l_x^1 \cdot e_x^1 dx = \int_0^\infty (\mu_x^1 - \mu_x^2) \cdot l_x^2 \cdot e_x^1 dx.$$

It means that contribution of a small age interval is

$$\varepsilon_{x,x+\Delta x} = (\mu_x^1 - \mu_x^2) \cdot l_x^2 \cdot e_x^1$$

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

Life expectancy: continuous versions of the method for decomposition.

The elegant formula by Vaupel & Canudas Romo also can be received on the basis of the same concerns but for infinitesimal age and time intervals.

$$\frac{\partial e(0,t)}{\partial t} = \bar{\rho}(t) \cdot e^{\dagger}(0,t) + Cov(\rho(a,t) \cdot e(a,t))$$

where $\rho(a,t) = -\dfrac{\partial \mu(a,t)}{\partial t} \Big/ \mu(a,t)$ is the relative change of the force of mortality at age *a* and at time *t*,

$$e^{\dagger}(0,t) = \int_0^{\infty} -\frac{\partial l(a,t)}{\partial a} \cdot e(a,t)da$$

and

$$Cov(\rho(a,t) \cdot e(a,t)) = \int_0^{\infty} (\rho(a,t) - \bar{\rho}(t)) \cdot (e(a,t) - e^{\dagger}(0,t)) \cdot \mu(a,t) \cdot l(a,t)da$$

By authors opinion the formula separates the change in an average into a level-1 term involving the average of age-specific changes and a level-2 covariance term that captures the effect of heterogeneity in age-specific or subpopulation-specific changes. Authors supposed that the second terms by in absolute magnitude is smaller that the fist one. This is correct for changes in Swedish life expectancy. Our calculation for about 6000 male and female life tables from the HMD showed that that the first term more than the second one in absolute magnitude only for ~73% life tables, and it is one order less only for ~ 15% life tables

Vaupel J.W., Canudas Romo V.C. Decomposing Change in Life Expectancy: A Bouquet of Formulas in Honor of Nathan Keyfitz's 90th Birthday. Demography, 2003. Vol. 40, No. 2, pp. 201-216.

A general algorithm of stepwise replacement decomposition for the multidimensional case.

Let $\mathrm{T}(\theta_1,...,\theta_n)$ is some aggregate demographic measure that can be presented as function of n variables $\theta_1,...,\theta_n$, which may be scalars, vectors, matrices, etc.

We want to decompose the difference $\delta\mathrm{T}=\mathrm{T}(\theta_1^2,...,\theta_n^2)-\mathrm{T}(\theta_1^1,...,\theta_n^1)$ to n components corresponded to n variables. It can be done as follows

$$\delta\mathrm{T}=\left[\mathrm{T}(\theta_1^2,\theta_2^2,\theta_3^2,...,\theta_n^2)-\mathrm{T}(\theta_1^1,\theta_2^2,\theta_3^2,...,\theta_n^2)\right]+\left[\mathrm{T}(\theta_1^1,\theta_2^2,\theta_3^2,...,\theta_n^2)-\mathrm{T}(\theta_1^1,\theta_2^1,\theta_3^2,...,\theta_n^2)\right]+$$
$$+\left[\mathrm{T}(\theta_1^1,\theta_2^1,\theta_3^2,...,\theta_n^2)-\mathrm{T}(\theta_1^1,\theta_2^1,\theta_3^1,...,\theta_n^2)\right]+...+\left[\mathrm{T}(\theta_1^1,\theta_2^1,...,\theta_{n-1}^1,\theta_n^2)-\mathrm{T}(\theta_1^1,\theta_2^1,...,\theta_{n-1}^1,\theta_n^1)\right]$$

Each square bracket is a component corresponded to the changed variable. In the presented chain the first bracket corresponds to the first variable, the second one corresponds to the second variable, etc.

Unfortunately (as we saw earlier in the simplest example of CDR), the ranking of variables influences the result. So, theoretically one needs to try all possible paths ($n!$)and then to calculate average components.

General principle of decomposition. How many replacements should we do?

Let $\mathrm{T}(\theta_1,...,\theta_n)$ is some aggregate demographic measure that can be presented as function of n variables $\theta_1,...,\theta_n$. We want decompose the difference $\delta T = \mathrm{T}(\theta_1^1,...,\theta_n^1) - \mathrm{T}(\theta_1^0,...,\theta_n^0)$ to n components corresponded to n variables. There are 2^{n-1} estimations of contribution of θ_k to this difference:

$$\delta^B T(\theta_k) = \mathrm{T}(\theta_1^{b_1},...,\theta_k^2,...,\theta_n^{b_n}) - \mathrm{T}(\theta_1^{b_1},...,\theta_k^1,...,\theta_n^{b_n}),$$

where $B = (b_1,...,b_{k-1},b_{k+1},...,b_n)$ is *n*-dimensional binary vectors ($b_i = 0,1$). Simultaneously B is binary representation a number less 2^{n-1}. Thus we should do for each variable 2^{n-1} replacements and calculate after that calculate its average contribution. Total number of replacements is $n \cdot 2^{n-1}$. Unfortunately this approach does not guarantee that summa n average contributions is equal the differences.

The summa of contribution of all variables is equal whole changes for certain if replacements are organized as a chain of coherent replacements which begins from $\theta_1^2,...,\theta_n^2$ and ends in $\theta_1^1,...,\theta_n^1$. Thus it is necessary to make $n!$ replacements that this condition “the sum of components is equal the whole change” was satisfied.

General principle of decomposition. How many replacements should we do?

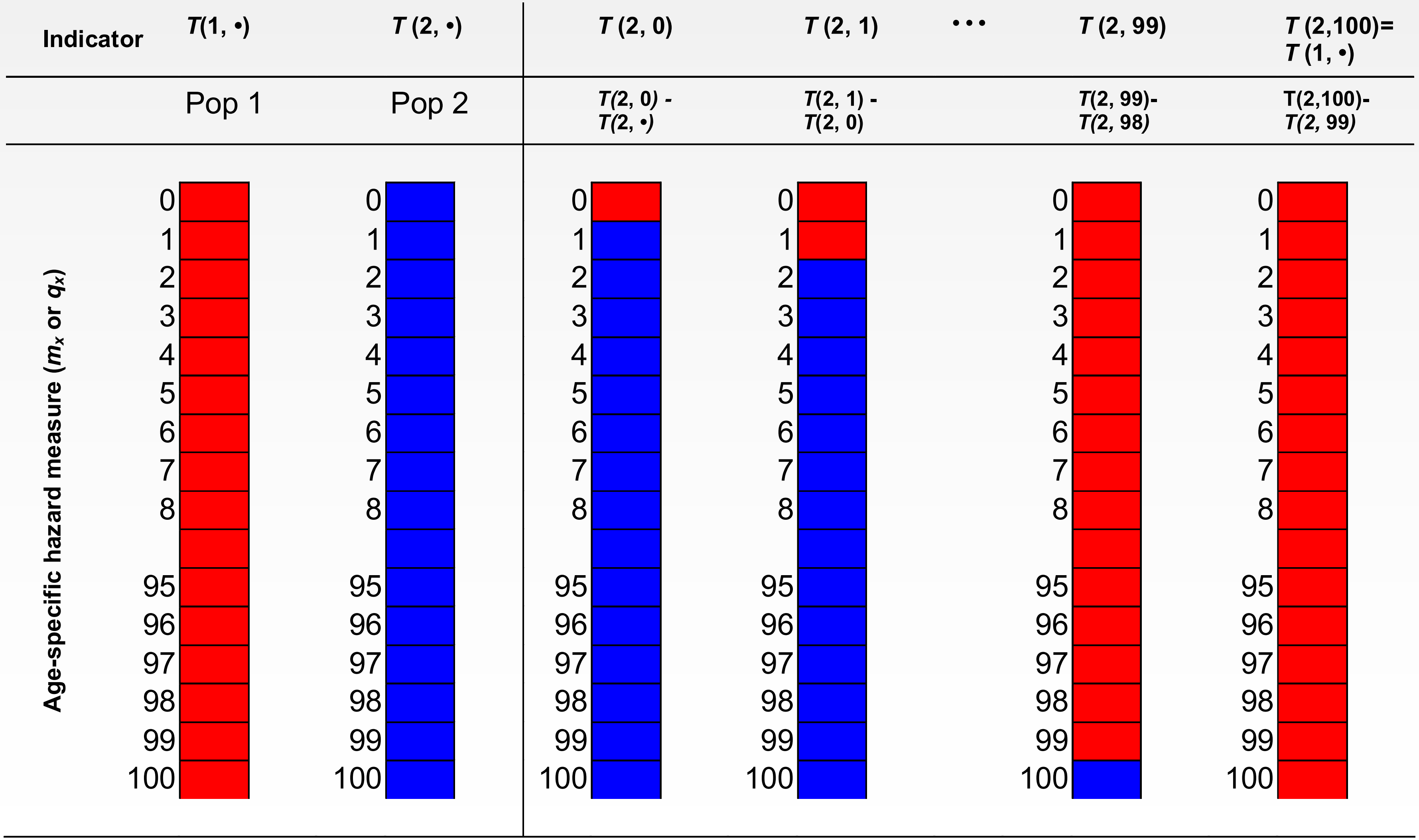
If n=2 then $n!$=2 and we can consider all possible chains (Kitagava's formulae for CDR). If n<11 then a standard modern PC can check all variants during reasonable period.

n	*n*!	*n*	*n*!
3	6	8	40 320
4	24	9	362 880
5	120	10	3628 800
6	720	11	39 916 800
7	5 040	12	479 001 600

In the case of decomposition by one-year age groups $n > 100$, $n!>10^{160}$ and it is impossible to check all variants.

Existing decomposition formulae for LE corresponds to the replacement running from young to old ages. It looks natural and meaningful.

Decomposition by age based on a stepwise replacement: algorithm running from young to old ages. (1)



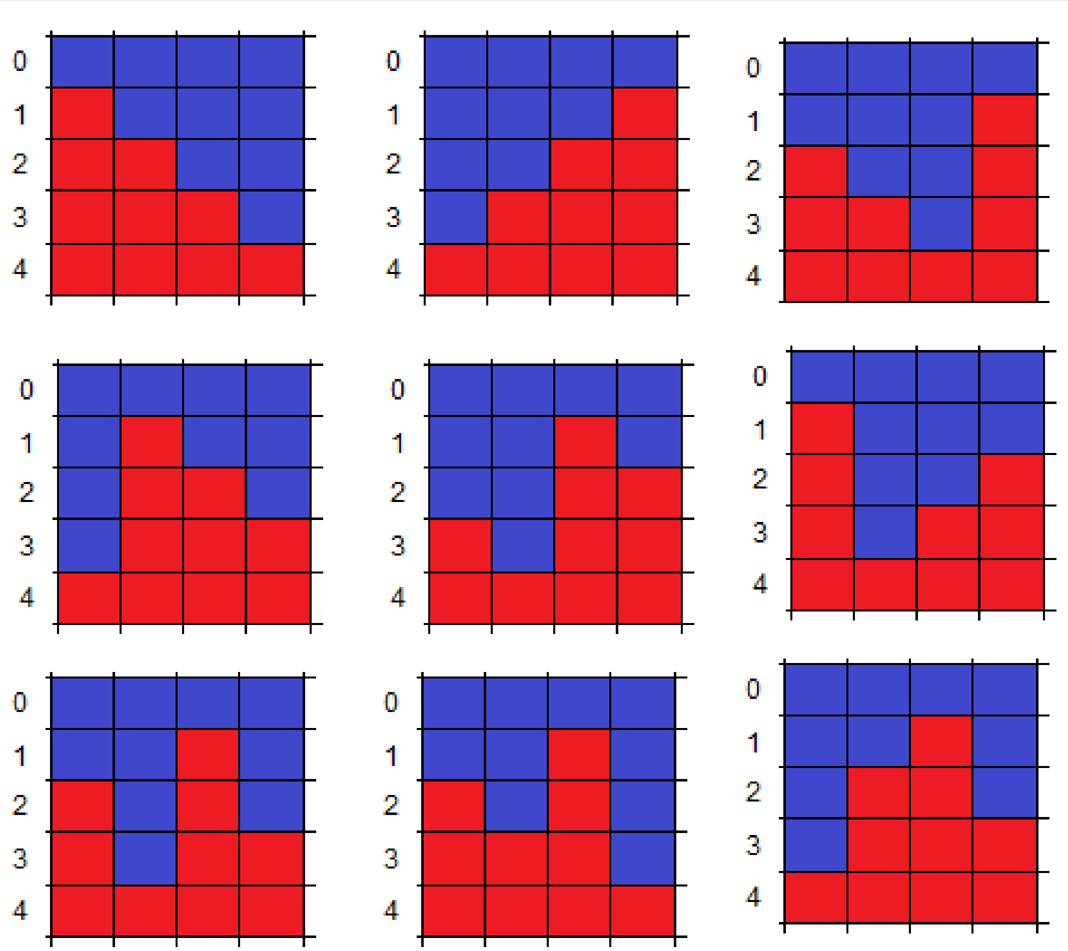
$$T(2,\cdot) - T(1,\cdot) = (T(2,0) - T(2,\cdot)) + (T(2,1) - T(2,2)) + \ldots + (T(2,99) - T(2,100)) + (T(2,100) - T(1,\cdot))$$

Decomposition by age based on a stepwise replacement: algorithm running from young to old ages. (2)

Generally speaking, it could be organized differently. For example, it could run from old to young ages (Pollard, 1988) or in a random manner.
Fortunately it is not difficult to show that if we use symmetrical form then replacement running from young to old ages and from old to young ages gives the same results. We made a lot of experiments with random replacement and they showed that in case of LE decomposition by age the order of replacement is not important.

Decomposition by age based on a stepwise replacement: replacement by causes at some age group.

In case of decomposition by age and some anther demographic category we should combine two approaches. We may use algorithm running from young to old ages for decomposition by age (number of replacements is 2·number of age group). However at any age group we should realize all sequences of replacement. If number of categories is N we should made N· N! replacements.



Total number of sequences is 24

Some problems of decomposition base on stepwise replacement. (1)

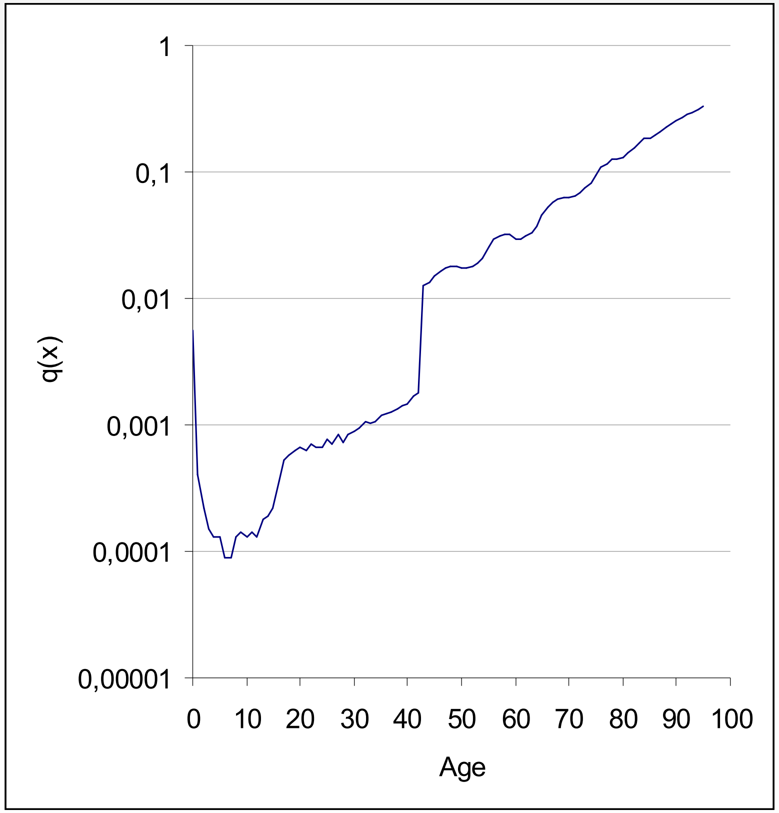
Let $\overline{\delta \mathrm{T}}^{1-2} = (\delta \mathrm{T}_1^{1-2}, \delta \mathrm{T}_2^{1-2}, ..., \delta \mathrm{T}_n^{1-2})$ is *n*-dimensional vector where $\delta \mathrm{T}_i^{1-2}$ equals a contribution of variable θ_i to difference $\delta \mathrm{T}^{1-2}$, $\delta \mathrm{T}^{1-2} = \sum_i \delta \mathrm{T}_i^{1-2}$. Thanks to symmetrical form we can be sure that $\overline{\delta \mathrm{T}}^{1-2} = -\overline{\delta \mathrm{T}}^{2-1}$. This feature in mathematics is called **commutativity**.

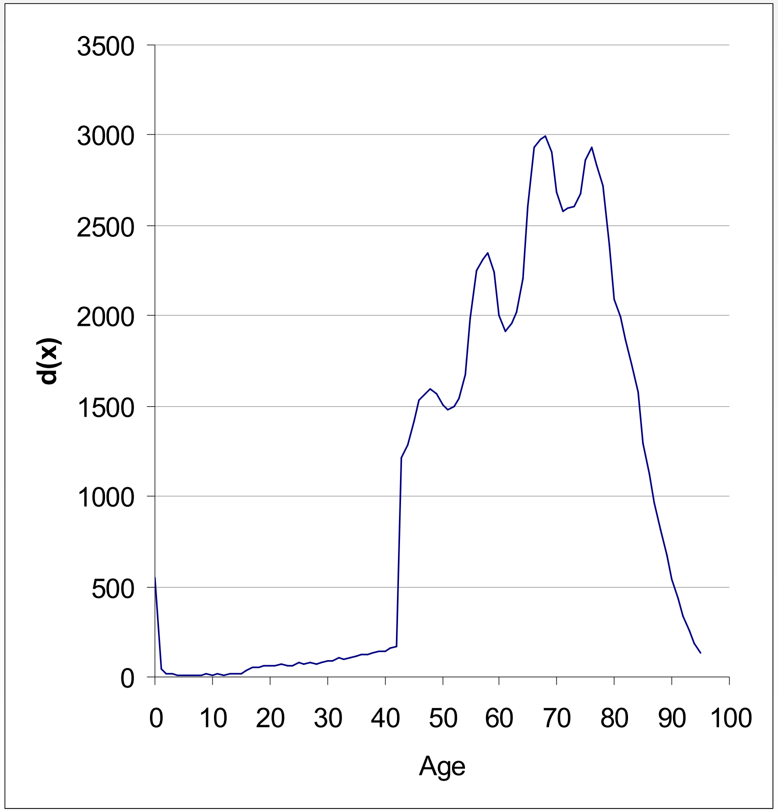
Unfortunately stepwise decomposition did not have second significant feature named **transitivity**. In general case $\overline{\delta \mathrm{T}}^{1-2} + \overline{\delta \mathrm{T}}^{2-3} \neq \overline{\delta \mathrm{T}}^{1-3}$. Non-transitivity is common defect of majority formulae for decomposition including Andreev-Arriaga-Pressat's formula. If we analyze for example dynamics of life expectancy at some population it is better initially to compare the nearest points and to receive decomposition for other periods by means of summing.

Formally we can compare any two life tables. However if these table is too differ then interim tables can be too exotic and calculation of same aggregate measures may be problematical. For example we compare life tables for E&W in 1861 and 2005 (for male).

Some problems of decomposition base on stepwise replacement. (2)

Example of function q(x) and d(x) of some interim table during stepwise replacements from young to old ages





However stepwise replacement technique is a general method which can be used directly in many cases when a formula for decomposition is unknown or too complicate. For this purpose it is necessary only write a program for replacement that is simpler than to derive a new formula.

The formulae for decomposition AID and $e^{\dagger}$.

Using the scheme of stepwise replacements it is possible to deduce the formula for any aggregate demographic measure.

$$AID_0(\mathrm{M}^{[x]}) - AID_0(\mathrm{M}^{[x+\tau]}) = (e'_{0|x} - e_{0|x}) - (e'_{0|x+\tau} - e_{0|x+\tau}) +$$
$$+ (l'_x - l_x)\cdot e_x - (l'_{x+\tau} - l_{x+\tau})\cdot e_{x+\tau}(\theta'_{0|x} - \theta_{0|x}) + (\theta'_{0|x+\tau} - \theta_{0|x+\tau}) -$$
$$- (l'^2_x - l^2_x)\cdot \theta_x + (l'^2_{x+\tau} - l^2_{x+\tau})\cdot \theta_{x+\tau},$$

where $\theta_x = \frac{1}{(l_x)^2}\int_x^{\infty}[l(t)]^2\,dt$

$$e^{\dagger}_0(\mathrm{M}^{[x+\tau]}) - e^{\dagger}_0(\mathrm{M}^{[x]}) = \left(\frac{l'_{x+\tau}}{l_{x+\tau}} - \frac{l'_x}{l_x}\right)\cdot l_{x+\tau}\cdot e^{\dagger}_{x+\tau} + \frac{1}{2}\delta_{x,x+\tau}(0)\cdot \sum_{y=0}^{x-\tau}\left[\frac{{}_\tau d'_y}{l'_y} + \frac{{}_\tau d'_y}{l'_{y+\tau}}\right] +$$
$$+ \frac{1}{2}\left[{}_\tau d'_x\left(e_x + \frac{1}{l'_x}\delta_{x,x+\tau}(0) + e_{x+\tau}\right) - {}_\tau d_x \frac{l'_x}{l_x}(e_x + e_{x+\tau})\right],$$

where $\delta_{x,x+\tau}(0) = l'_y(e'_y - e_y) - l'_{y+\tau}(e'_{y+\tau} - e_{y+\tau})$

These formulas are too complicate for practical calculations. It is much easier to use numerical algorithm stepwise replacement.

Third example: decomposition of differences between two healthy life expectancies.

According to D.Sullivan (1964) method, health expectancy is defined as

$$eH_0 = \sum_{x=0}^{\omega} {}_1L_x \cdot \eta_x \text{ ,}$$

where η_x is the share of person-years lived in "good" health within the elementary age interval [*x*, *x*+1). Usually the health-weights $\eta_x = 1 - \pi_x$ are obtained from nationally representative surveys including questions on self-perceived health.

According to this formula, two vectors are needed for calculating the health expectancy. These are the vector of age-specific mortality rates M and the vector of age-specific health-weights Π.

According to the algorithm of stepwise replacement, the component of the overall difference in eH_0 due to the difference between mortality rates at age *x* is

$$\lambda_x^{2-1} = \frac{1}{2}\{[(eH_0(M^{[x+1]},\Pi^{[x]}) - eH_0(M^{[x]},\Pi^{[x]})] + [(eH_0(M^{[x+1]},\Pi^{[x+1]}) - eH_0(M^{[x]},\Pi^{[x+1]})]\}$$

The component of the overall difference in eH_0 due to the difference in health-weights at age *x* is

$$\gamma_x^{2-1} = \frac{1}{2}\{[eH_0(M^{[x]},\Pi^{[x+1]}) - eH_0(M^{[x]},\Pi^{[x]})] + [eH_0(M^{[x+1]},\Pi^{[x+1]}) - eH_0(M^{[x+1]},\Pi^{[x]})]\}$$

Third example: healthy life expectancy

After simple but fatiguing transformations (for detail look Andreev, Shkolnikov, Begun, 2002) we found formulae for 1→2 replacement. The mortality component is

$$\lambda_x^{2-1} = \frac{1}{2} l_x^2 ({}_nP_x^2 - {}_nP_x^1)\, {}_n\eta_x^1 + eH_{x+1}^1 l_x^2 ({}_nq_x^1 - {}_nq_x^2),$$

where ${}_1P_x^i = {}_1L_x^i / l_x^i$ and the health component is

$$\gamma_x^{2-1} = \frac{1}{2} l_x^2 ({}_1P_x^1 + {}_1P_x^2)(\eta_x^2 - \eta_x^1).$$

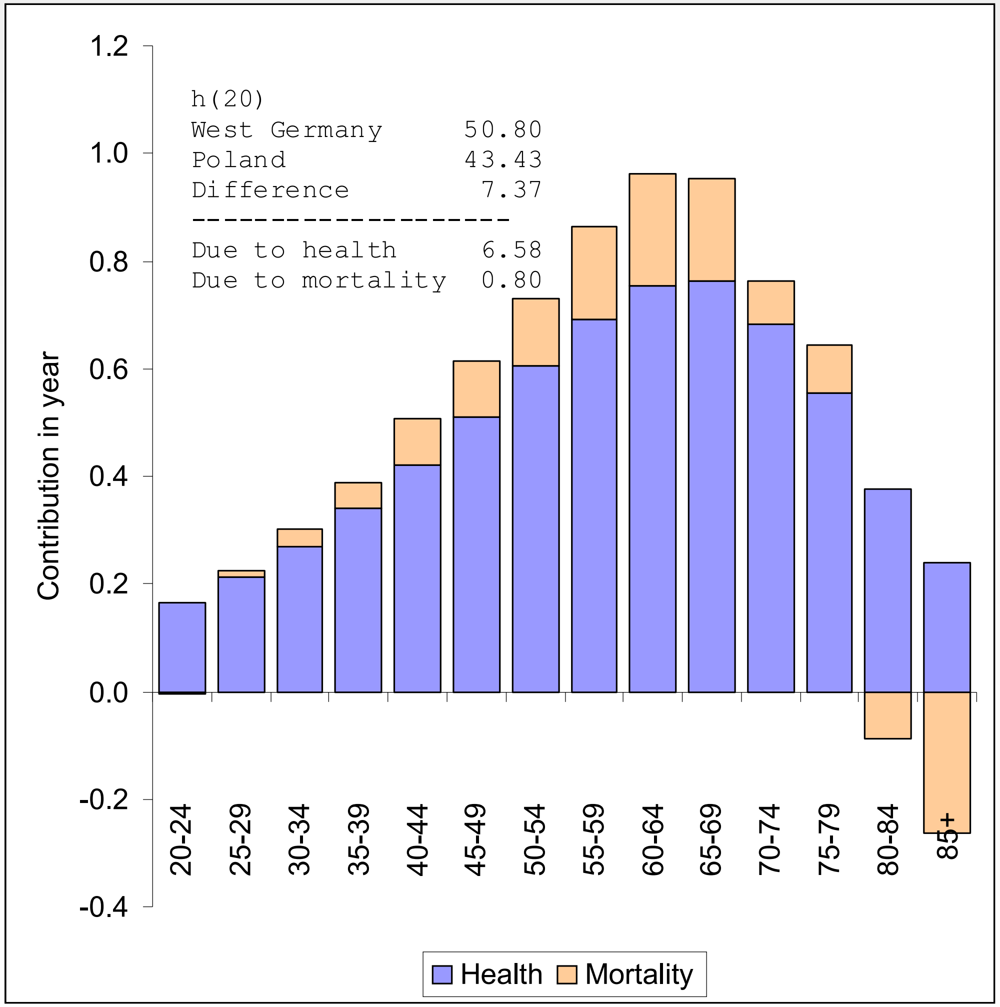
The components $\lambda_x^{2-1}, \gamma_x^{2-1}$, produced by 1→2 replacement, should be averaged with $\lambda_x^{1-2}, \gamma_x^{1-2}$ produced by 2→1 replacement. The final formulae are

$$\lambda_x = \frac{1}{4}(l_x^1 + l_x^2)({}_1P_x^2 - {}_1P_x^1)(\eta_x^1 + \eta_x^2) + \frac{1}{2}(eH_{x+1}^1 l_x^2 + eH_{x+1}^2 l_x^1)({}_1q_x^1 - {}_1q_x^2)$$

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

If $\pi_x^1 = \pi_x^2 = 1$ then the age-components λ_x of the difference between health expectancies due to mortality became equal to the conventional age-components of the difference between life expectancies δ_x.

Decomposition of the difference between female health expectancies at age 20 between West Germany and Poland for the 1990s.



This Figure suggests that contributions due to differences in self-reported health are much greater than those due to differences in mortality.

6.6 years of the overall difference of 7.4 years are attributable to differences in health.

Decomposition_HE

Health-weights are calculated from the data on self-perceived health, extracted from the second and third wave of the World Value Surveys For each five-year age group weights are the sums of the original proportions of women with "fair", "good" and "very good" self-perceived health.

Fourth example: life expectancy, ages, and causes of death.

For small age interval [*x*, *x*+Δ*x*] $\mu_x = \frac{l_x - l_{x+\Delta x}}{l_x \Delta x} \Rightarrow l_{x+\Delta x} = l_x(1-\mu_x \Delta x)$, and $\mu_x = \frac{l_x - l_{x+\Delta x}}{l_x e_x - l_{x+\Delta x} e_{x+\Delta x}} \Rightarrow e_{x+\Delta x} = e_x - (1-\mu_x e_x)\Delta x$, $L_{0|x+\Delta x} = L_{0|x} + l_x \cdot \Delta x$. Thus the values $\delta_{x,x+\Delta x}$ may be presented as

$$\delta_{x,x+\Delta x} = \alpha_x \cdot (\mu_x - \mu'_x) \cdot \Delta x = \sum_j \alpha_x \cdot ({\mu^{[j]}}_x - {\mu'^{[j]}}_x) \cdot \Delta x$$

where α_x are rational functions of life table indicators and $\mu_x^{[j]}$ force of mortality from cause [*j*]. $\mu_x^{[j]} = \frac{d_x^{[j]}}{d_x} \mu_x$

If ${}_\tau M_x$ and ${}_\tau M_x^{[j]}$ are age-specific mortality rates from all causes and cause [*j*] respectively and $\left| {}_\tau M_x - {}_\tau M'_x \right| \neq 0$, then the following approximate formula can be used

$$\delta^{[j]}_{x,x+\tau} = \frac{M^{[j]}_{x,x+\tau} - M'^{[j]}_{x,x+\tau}}{M_{x,x+\tau} - M'_{x,x+\tau}} \delta_{x,x+\tau} .$$

This is an approximate formula. It does not work if its denominator is close to 0. It is possible that some age component is zero but several age-cause components for the same age are not zero.

Life expectancy (LE) decomposition using the stepwise replacement by age and cause

For each of the two populations, life expectancy is presented as function of a matrix $\left\| {}_{\tau}M_x^{[j]} \right\|$ of death rates by age and cause of death. The rows of this matrix correspond to age groups and the columns correspond to causes of death. Calculation of LE is based on the conventional LT methods with

$M_x = \sum_j M_x^{[j]}$ and $M'_x = \sum_j M'^{[j]}_x$

According to the general algorithm, the elements of the first matrix $\left\| {}_{\tau}M_x^{[j]} \right\|$ should be replaced by respective elements of the second matrix $\left\| M'^{[j]}_x \right\|$ and vice versa.

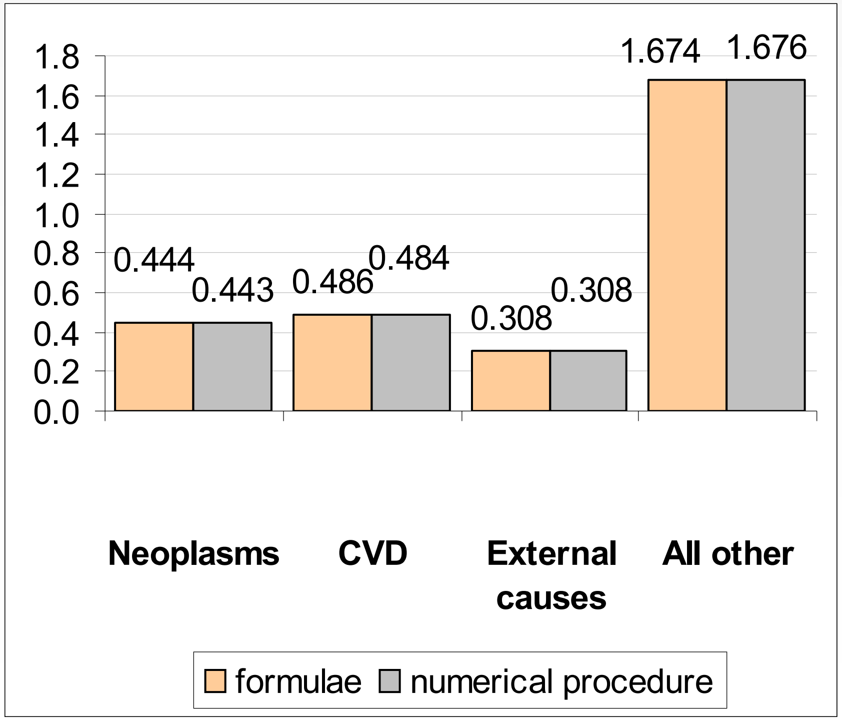
Replacement by age runs from young to old ages and in each age group *x*, all possible ways of replacements of $M_x^{[j]}$ by $\left\| M'^{[j]}_x \right\|$ should be performed.

Similarly one can decompose the difference between two average ages at death for a given cause of death. The organization of replacement you can see in the following Excel file.

Decomposition of life expectancy: examples

For example we calculated two variants of decomposition of the differences in life expectancy by age and causes death between women in Italy and the USA in 2000 using the formula and the stepwise replacement procedure. The difference is very small.

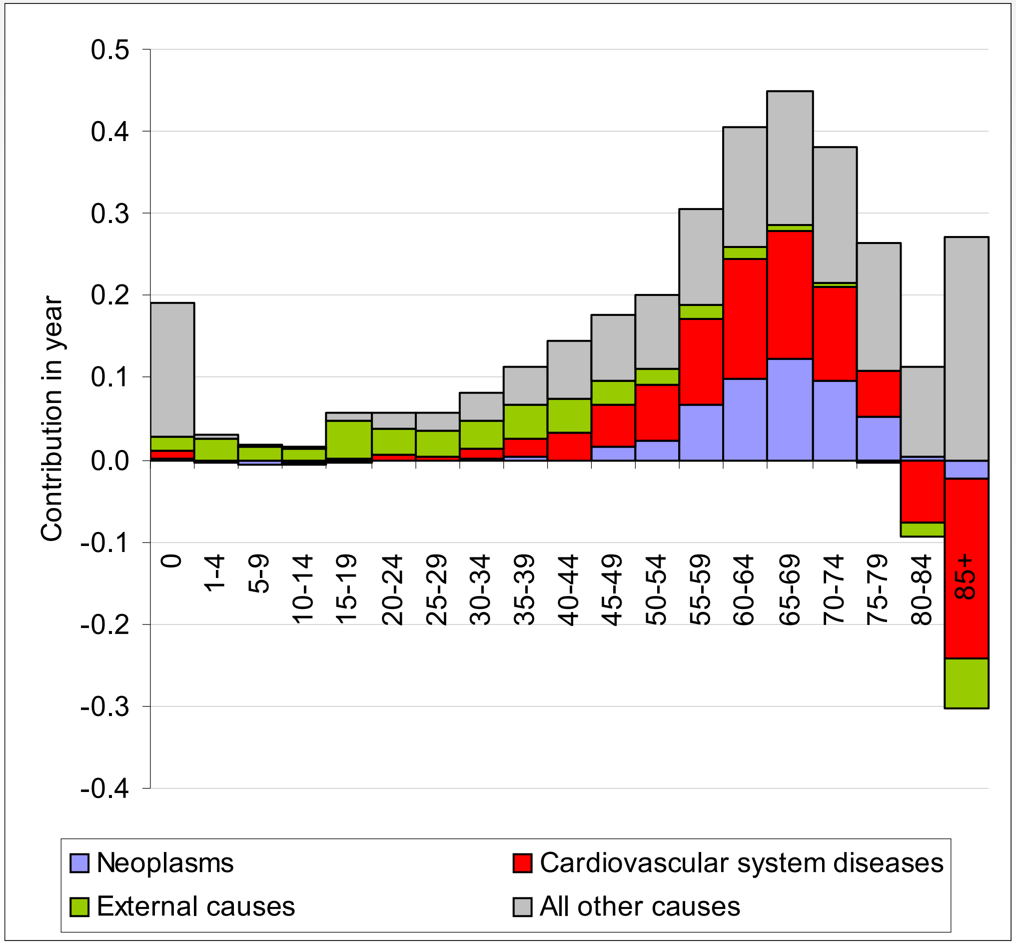
Difference of the two calculations



Decomposition_USA_Italy_by_formulae

Decomposition_replacement_USA-Italy

Age and cause of death decomposition of the LE difference between the USA and Italy females in 2000



e(0)	
USA	79.60
Italy	82.51

Differnece	2.91
Neoplasms	0.44
Cardiovascular diseases	0.48
External causes	0.31
All other causes	1.68

Fifth example: life expectancy, group-specific mortality, and population composition

The procedure of stepwise replacement can be used for a very different decomposition where an aggregate demographic measure is a function of elements from two matrices of the population weights $\|\theta_x^i\|$ and mortality rates $\|M_x^i\|$ for the group *i*, where $M_x^1 = \sum_i \theta_x^{1i} M_x^{1i}$ and $M_x^2 = \sum_i \theta_x^{2i} M_x^{2i}$, .

For example, we can decompose an increase in LE of Finnish men from 1988-1989 to 1998-1999 by components connected to mortality in the three educational groups: high education, secondary education, and low education group (all lower levels of education combined).

Mortality is described by the group-specific death rates in five-year age groups 30-34,35-39,…,85+. Population structure is described by the two age specific variables: age-specific shares of people with high and secondary education among people with low and secondary education.

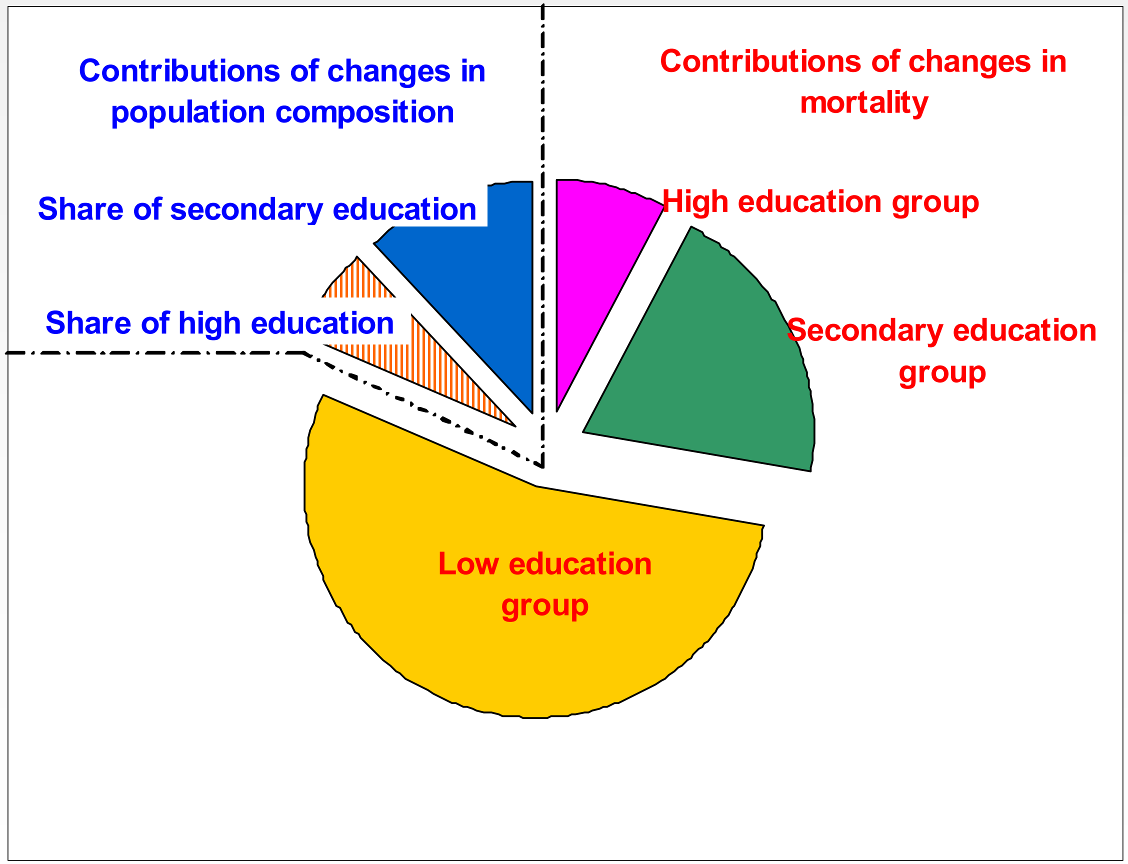
Layout of input data

Decomposition_replacement_Finland

	M(x) H	M(x) S	M(x) L	Share H	Share S in S+L
30-34	0.000642	0.001874	0.003183	0.132889	0.684849
35-39	0.001011	0.002505	0.003783	0.13812	0.598154
40-44	0.002056	0.003685	0.005022	0.147158	0.503257
45-49	0.003047	0.004836	0.006924	0.136577	0.417186
…………					
75-79	0.057418	0.074896	0.084376	0.062652	0.136192
80-84	0.091281	0.106962	0.129463	0.063253	0.116056
85+	0.192566	0.188272	0.215025	0.056559	0.104833

Decomposition of the life expectancy growth in Finland 1988-89 – 1998-99

Total life expectancy increase	2,58
of them	
due to mortality decline	**2,10**
high educated;	0,20
secondary educated	0,52
low educated	1,38
due to changes of population composition	**0,48**
share of high educated	0,17
share of secondary educated among non-high educated	0,31



The most influential component of the total mortality improvement is the decline in mortality of people with low education.

Decomposition_replacement_Finland(2)

References

Andreev, E.M., Shkolnikov, V.M., Begun, Z. (2002). Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity progression ratios and total fertility rates. *Demographic Research*, Vol. 7, Article 14, pp. 500-526

Arriaga E.E. (1994) Measuring and explaining the change in life expectancies. *Demography*, Vol. 21, No. 1, pp. 83-96

Kitagawa E.M. (1955) Components of a difference between two rates. *Journal of the American Statistical Association*, Vol. 50, No. 272, pp. 1168-1194

Pollard J.H. (1998) On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography*, Vol. 25, No.2, pp. 265-276

Pressat R. (1985) Contribution des écarts de mortalité par âge à la différence des vies moyennes. *Population*, Vol. 4, No.5, pp. 766-801

Shkolnikov V.M., Valkonen T., Begun A., Andreev E.M. (2001) Measuring inter-group inequalities in length of life. *Genus*, Vol. LVII, No. 3-4, pp. 33-62

Vaupel J.W., Romo V.C. (2003) Decomposing Change in Life Expectancy: A Bouquet of Formulas in Honor of Nathan Keyfitz's 90th Birthday. *Demography*, Vol. 40, No. 2, pp. 201-216.

Андреев Е.М. (1982) Метод компонент в анализе продолжительности жизни *Вестник статистики*, № 9. С. 42-47. http://demoscope.ru/weekly/knigi/andreev/andreev.html

Andreev, E.M., Shkolnikov, V.M. (2012) An Excel spreadsheet for the decomposition of a difference between two values of an aggregate demographic measure by stepwise replacement running from young to old ages. *MPIDR Technical Report*.
http://www.demogr.mpg.de/en/projects_publications/publications_1904/mpidr_technical_reports/an_excel_spreadsheet_for_the_decomposition_of_a_difference_between_two_values_of_an_aggregate_4591.htm

Two tasks to Lecture 8 are on the server
Please solve them up to 2 p.m. 23 January 2013

The end