Decomposition techniques in population health research

José Manuel Aburto

SDU

EUROPEAN DOCTORAL SCHOOL OF DEMOGRAPHY

Preliminaries

- Introduction
- Course materials `github.com/jmaburto/EDSD-Decomposition-Course-2018`
- Assignment: 3 challenges in groups of 3 or 4.

Outline

- The first decomposition method: Kitagawa (1955)
- Direct vs Compositional effects: Vaupel & Canudas-Romo (2002)
- Change in life expectancy

Origins of decomposition

- Methods of standardization

  Aim: Eliminate compositional effect from overall rates of some phenomenon.

Origins of decomposition

- Methods of standardization

  Aim: Eliminate compositional effect from overall rates of some phenomenon.

  - Indirect standardization $\longrightarrow$ 1876
  - Direct standardization $\longrightarrow$ 1844

Origins of decomposition

- Methods of standardization

  Aim: Eliminate compositional effect from overall rates of some phenomenon.

  - Indirect standardization $\longrightarrow$ 1876
  - Direct standardization $\longrightarrow$ 1844

- Unreliable due to their dependence on an arbitrary standard

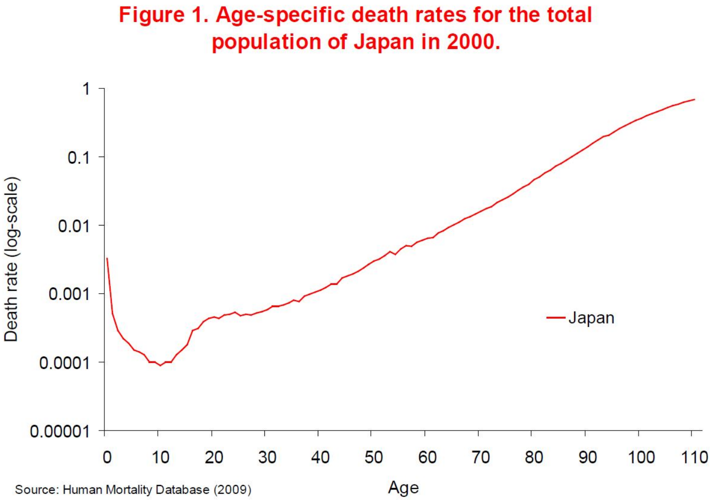
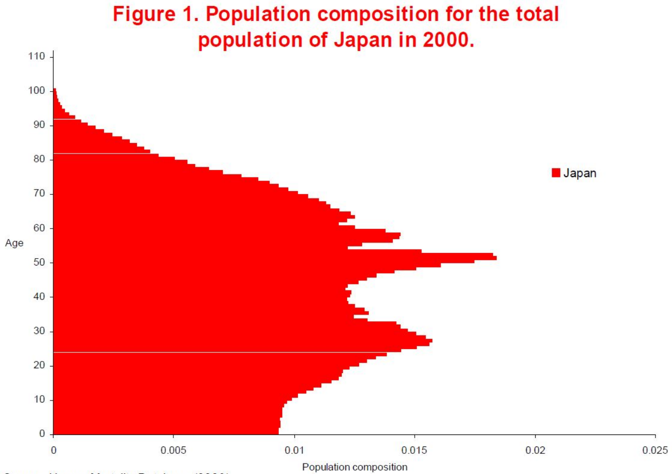
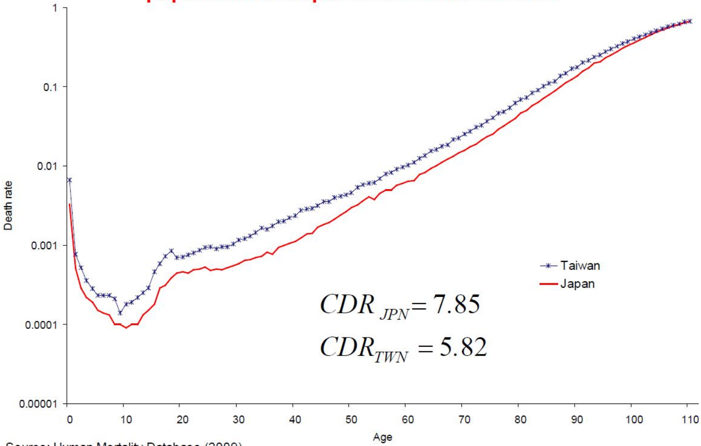


Figure 1. Age-specific death rates for the total population of Japan in 2000.
Death rate (log-scale)
1
0.1
0.01
0.001
0.0001
0.00001
0
10
20
30
40
50
60
70
80
90
100
110
Age
Japan
Source: Human Mortality Database (2009)

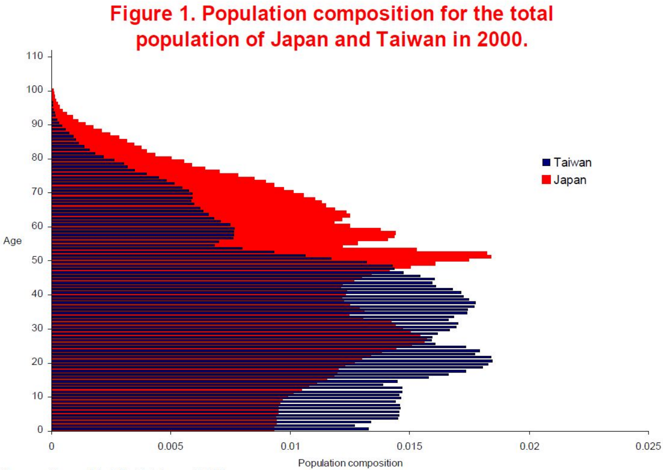


Source: Human Mortality Database (2009)

Figure 1. Age-specific death rates for the total population of Japan and Taiwan in 2000.



Source: Human Mortality Database (2009)



Source: Human Mortality Database (2009)

Crude Death rate (CDR)

	JAPAN	TAIWAN
CDR	7.85	5.82
SCDR Direct Stand.	5.79	8.72
SCDR Indirect Stand.	9.73	4.78

Motivation to develop further methods of comparison: Decomposition

Kitagawa (1955)

Notation for population 1 (population 2 same but lower cases):

$N_i =$ number of persons in the i-th category of I

$E_i =$ number of events (e.g. births, deaths) in the i-th category of I

$T_i = E_i/N_i$ rate of persons in the i-th category of I

$N = \sum_i N_i$ total number of persons

$E = \sum_i E_i$ total number of events

$T = E/N$ crude rate

Kitagawa (1955)

The difference between the crude rates can be expressed as

$$t - T = \underbrace{\sum_i T_i \left[\frac{n_i}{n} - \frac{N_i}{N}\right]}_{\text{Changes in I-composition}} + \underbrace{\sum_i \frac{N_i}{N}(t_i - T_i)}_{\text{Dif with pop 1 as standard}} + \underbrace{\sum_i (t_i - T_i)\left[\frac{n_i}{n} - \frac{N_i}{N}\right]}_{\text{Interaction of rates and compositions}} \tag{1}$$

Optional exercise: show that $t - T$ can be expressed as (1)

Kitagawa (1955)

To avoid the interaction term, Kitagawa suggests

$$t - T = \underbrace{\sum_i \frac{t_i + T_i}{2}\left[\frac{n_i}{n} - \frac{N_i}{N}\right]}_{\text{Changes in I-composition}} + \underbrace{\sum_i \frac{\frac{n_i}{n} + \frac{N_i}{N}}{2}(t_i - T_i)}_{\text{Changes in rates}} \quad (2)$$

Challenge 1: show that (1) can be expressed as (2)

Example: Berrington et al 2015

Aim: To investigate the relative contributions of childlessness, timing, and quantum to educational differences in completed fertility within cohorts born between 1940 and 1969.

Data: General Household Survey (GHS) in Britain.

Method: Completed family size (C) is equivalent to completed family size for mothers (C_m) times the proportion of women who are mothers (p_m) at the end of the reproductive period.For each 10-year birth cohort, we want to estimate the proportion of the total fertility differential between degree-educated (subscript H) and least-educated (subscript L) women that can be attributed to difference in childlessness.

$$C_H - C_L = \underbrace{\frac{p_{mH} + p_{mL}}{2}(C_{mH} - C_{mL})}_{\text{Cm weighted by avg. motherhood share}} + \underbrace{\underbrace{\frac{C_{mH} + C_{mL}}{2}(p_{mH} - p_{mL})}_{\text{Motherhood-share weighted by avg. completed fam size}}}_{\text{Childlessness cotribution}}$$

$$C_{mH} - C_{mL} = \sum_i \left(\frac{\frac{N_{iH}}{N_H} + \frac{N_{iL}}{N_L}}{2} \right) (C_{mHi} - C_{mLi})$$
$$\sum_i \left(\frac{C_{mHi} + C_{mLi}}{2} \right) \left(\frac{N_{iH}}{N_H} - \frac{N_{iL}}{N_l} \right)$$

$$C_{mH} - C_{mL} = \sum_i \left(\frac{\frac{N_{iH}}{N_H} + \frac{N_{iL}}{N_L}}{2} \right) (C_{mHi} - C_{mLi})$$

$$\sum_i \left(\frac{C_{mHi} + C_{mLi}}{2} \right) \left(\frac{N_{iH}}{N_H} - \frac{N_{iL}}{N_l} \right)$$

- The second component reflects the extent to which Cm would change if age-specific fertility rates changed but the distribution of women at entry into motherhood remained constant.
- The composition effect addresses the extent to which Cm would change if the distribution of age at entry into motherhood changed but the fertility rates conditional upon age at first birth remained constant.

Table 4: **Childlessness and completed family size by educational attainment and relative contributions of childlessness, rate effect, and composition effect to educational differences in completed family size, by cohort**

Cohort	% Childless		CFSM		CFST		Contribution to educational differences in CFST (%)		
	<O Level	*Degree*	*<O Level*	*Degree*	*<O Level*	*Degree*	*Childlessness*	*Rate*	*Composition*
1940-49	8.4	18.5	2.58	2.28	2.36	1.86	48.9 (47.6;50.2)	-4.4 (-13.5;4.3)	55.5 (48.1;63.2)
1950-59	10.0	20.6	2.51	2.23	2.26	1.77	51.8 (51.3;52.3)	-16.6 (-24.9;-8.5)	64.8 (57.2;72.6)
1960-69	10.2	22.0	2.62	2.15	2.35	1.68	41.7 (41.2;42.2)	1.3 (-4.5; 6.9)	57.0 (51.9;62.3)

Further reading

- Gupta, Prithwis Das. "A general method of decomposing a difference between two rates into several components." Demography 15.1 (1978): 99-112.
- Cho, L. J., & Retherford, R. D. (1973). Comparative analysis of recent fertility trends in East Asia.
- Gonalons-Pons, P., & Schwartz, C. R. (2017). "Trends in Economic Homogamy: Changes in Assortative Mating or the Division of Labor in Marriage?." Demography, 54(3), 985-1005.

Direct vs Compositional

Let $\bar{v}(y)$ denote the mean value of $v(x, y)$ over x as

$$E(v) = \bar{v}(y) = \frac{\int_0^\infty v(x, y)w(x, y)dx}{\int_0^\infty w(x, y)dx}, x \text{ continuous}$$
$$= \frac{\sum_x v(x, y)w(x, y)dx}{\sum_x w(x, y)dx}, x \text{ discrete} \tag{3}$$

where $v(x, y)$ is some demographic function and $w(x, y)$ is some weighting function.

A dot over a variable denotes the derivative with respect to y (usually time)

$$\dot{v} = \frac{\partial}{\partial y} v(x, y)$$

and an acute accent denotes the relative derivative or intensity with respect to y

$$\acute{v} = \frac{\frac{\partial}{\partial y} v(x, y)}{v(x, y)} = \frac{\partial}{\partial y} \ln[v(x, y)]$$

We want to decompose the derivative of $\bar{v}$ (e.g. mean age at childbearing, CDR) with respect to y (time) into direct and compositional effects

$$\bar{\dot{v}} = \frac{\partial}{\partial y} \frac{\int_0^\infty v(x,y) w(x,y) dx}{\int_0^\infty w(x,y) dx}$$

$$\dot{\bar{v}} = \frac{\partial}{\partial y} \frac{\int_0^\infty v(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx}$$

$$= \bar{\dot{v}} + \frac{\int_0^\infty v(x,y)\acute{w}(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx}$$

$$- \frac{\int_0^\infty v(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} \frac{\int_0^\infty \acute{w}(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx}$$

$$
\begin{aligned}
\dot{\bar{v}} &= \frac{\partial}{\partial y} \frac{\int_0^\infty v(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} \\
&= \bar{\dot{v}} + \frac{\int_0^\infty v(x,y)\acute{w}(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} \\
&\quad - \frac{\int_0^\infty v(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} \frac{\int_0^\infty \acute{w}(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} \\
&= \bar{\dot{v}} + E(v\acute{w}) - E(v)E(\acute{w})
\end{aligned}
$$

$$
\begin{aligned}
\dot{\bar{v}} &= \frac{\partial}{\partial y} \frac{\int_0^\infty v(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} \\
&= \bar{\dot{v}} + \frac{\int_0^\infty v(x,y)\acute{w}(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} \\
&\quad - \frac{\int_0^\infty v(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} \frac{\int_0^\infty \acute{w}(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx} \\
&= \bar{\dot{v}} + E(v\acute{w}) - E(v)E(\acute{w}) \\
&= \bar{\dot{v}} + Cov(v,\acute{w})
\end{aligned}
$$

$$\bar{\dot{v}} = \underbrace{\bar{\dot{v}}}_{\text{Direct component}} + \underbrace{Cov(v, \acute{w})}_{\text{Structural or compositional component}} \tag{4}$$

$$\dot{\bar{v}} = \underbrace{\bar{\dot{v}}}_{\text{Direct component}} + \underbrace{Cov(v, \acute{w})}_{\text{Structural or compositional component}} \tag{4}$$

Main result in Vaupel & Canudas-Romo (2002)

Open Exercise 1 in `R`

Preliminaries:

$$\mu(a), \text{force of mortality at age } a$$

$$\ell(x) = \exp(-\int_0^x \mu(a)\,da), \text{survival function}$$

$$e_o(a) = e(a,t) = \frac{\int_a^\infty \ell(a,t)\,da}{\ell(a)}, \text{life expectancy at age a}$$

$$\rho = -\acute{\mu(a)}, \text{rate of mortality improvement}$$

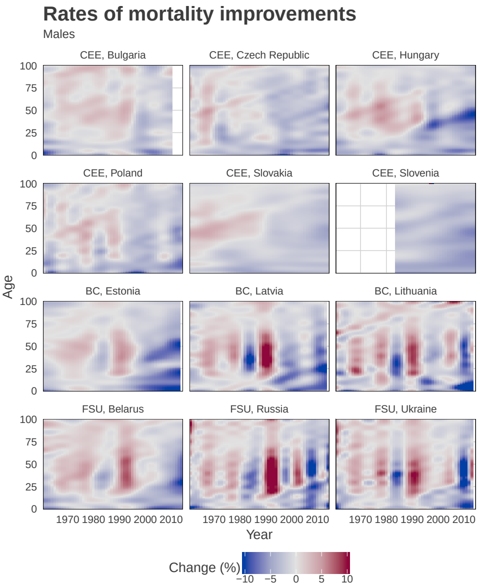
$$\frac{\partial}{\partial t} e_o(t) = \dot{e}_o(t) = \int_0^{\infty} \frac{\partial}{\partial t} \ell(x, t)\, dx$$

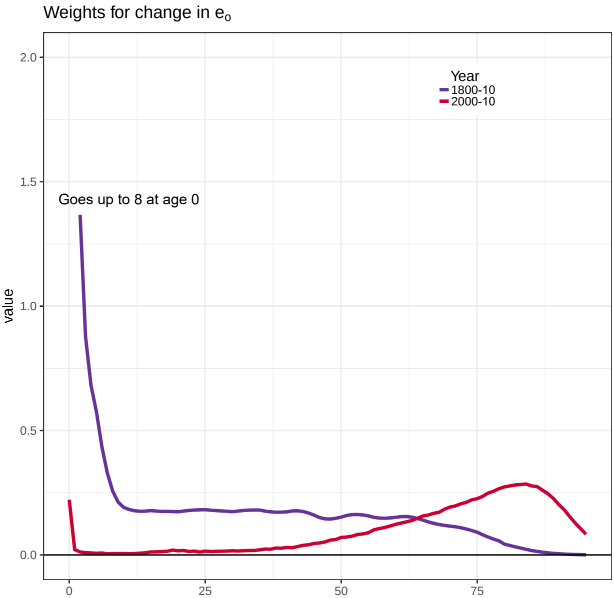
$$\frac{\partial}{\partial t}\, e_o(t) = \dot{e}_o(t) = \int_0^{\infty} \frac{\partial}{\partial t}\, \ell(x,t)\, dx$$
$$= -\int_0^{\infty} \ell(x,t) \int_0^{x} \frac{\partial}{\partial t}\, \mu(a,t)\, da\, dx$$

$$
\begin{aligned}
\frac{\partial}{\partial t}\, e_o(t) = \dot{e}_o(t) &= \int_0^\infty \frac{\partial}{\partial t}\, \ell(x,t)\, dx \\
&= -\int_0^\infty \ell(x,t) \int_0^x \frac{\partial}{\partial t}\, \mu(a,t)\, da\, dx \\
&= -\int_0^\infty \frac{\partial}{\partial t}\, \mu(x,t) \int_x^\infty \ell(a,t)\, da\, dx
\end{aligned}
$$

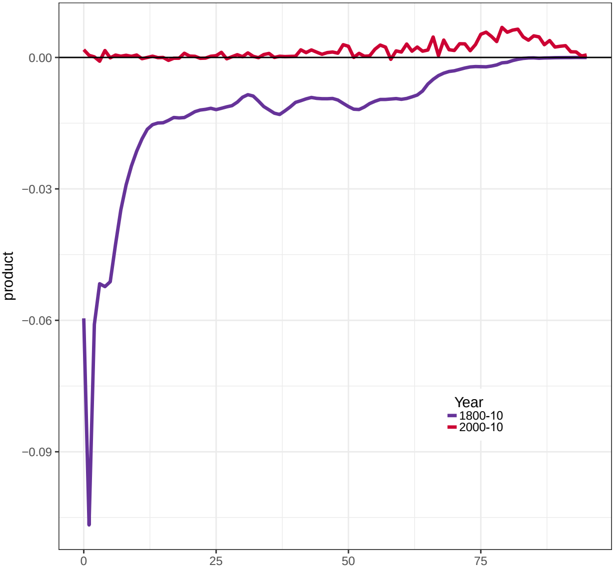
$$
\begin{aligned}
\frac{\partial}{\partial t} e_o(t) = \dot{e}_o(t) &= \int_0^{\infty} \frac{\partial}{\partial t} \ell(x,t)\, dx \\
&= -\int_0^{\infty} \ell(x,t) \int_0^{x} \frac{\partial}{\partial t} \mu(a,t)\, da\, dx \\
&= -\int_0^{\infty} \frac{\partial}{\partial t} \mu(x,t) \int_x^{\infty} \ell(a,t)\, da\, dx \\
&= \int_0^{\infty} \rho(x) e(x) f(x) dx
\end{aligned}
$$

where $f(x)$ is the age-at-distribution weighting function.


Rates of mortality improvements
Males
CEE, Bulgaria
CEE, Czech Republic
CEE, Hungary
CEE, Poland
CEE, Slovakia
CEE, Slovenia
BC, Estonia
BC, Latvia
BC, Lithuania
FSU, Belarus
FSU, Russia
FSU, Ukraine
Age
0
25
50
75
100
1970 1980 1990 2000 2010
Year
Change (%)
−10 −5 0 5 10


Weights for change in e_o
Year
1800-10
2000-10
Goes up to 8 at age 0
value
2.0
1.5
1.0
0.5
0.0
0
25
50
75

Product ρw



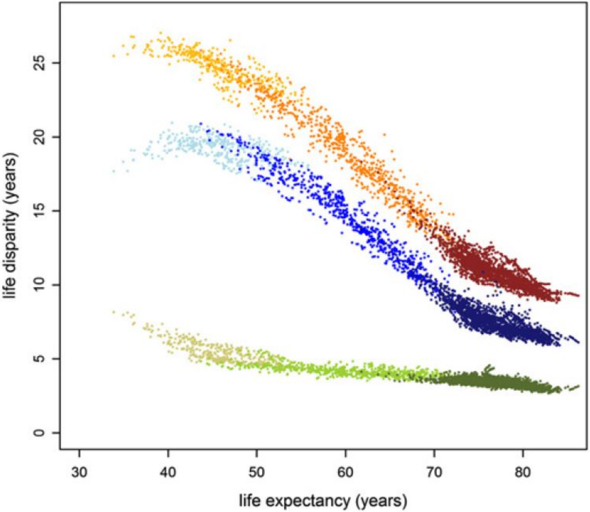
Following Vaupel & Canudas-Romo (2002)

$$\dot{e}_o(t) = \int_0^\infty \rho(x)e(x)f(x)dx \qquad (5)$$

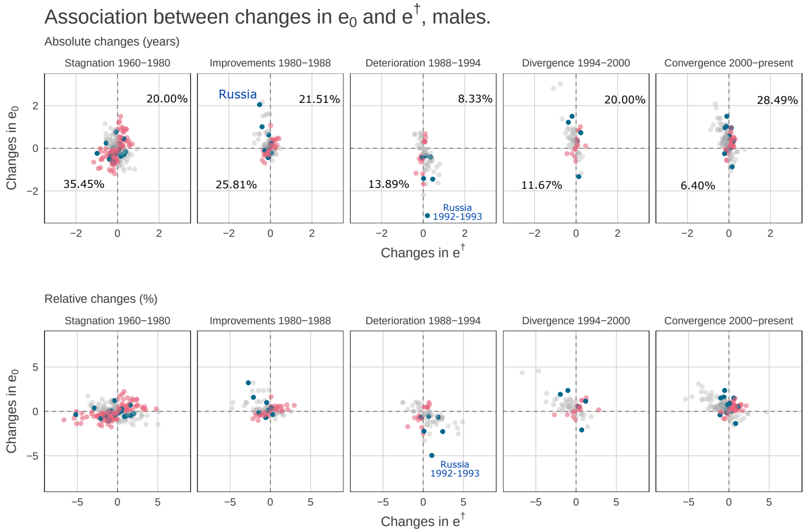
can be written as:

$$\dot{e}_o(t) = \bar{\rho}(t)e^\dagger(t) + Cov(\rho, e_x) \qquad (6)$$

where $e^\dagger = \int_0^\infty e(x)f(x)da$ is the average life lost at time of death (Vaupel & Canudas-Romo, 2003).



Source: Vaupel et al 2011



Source: Aburto & van Raalte 2018

Challenge 2: Using data from the UN give a descriptive (no more than 300 words with max 2 figures) answer to the question: Is there a female advantage in life disparity as there is in longevity?