Decomposition techniques in population health research

José Manuel Aburto

EUROPEAN DOCTORAL SCHOOL OF DEMOGRAPHY

CPop SDU

UNIVERSITY OF OXFORD DEPARTMENT OF SOCIOLOGY

LEVERHULME CENTRE FOR DEMOGRAPHIC SCIENCE

The British Academy

Preliminaries

- Introduction
- Course materials
  https://github.com/jmaburto/EDSD-Decomposition-Course-2021
- Teams
- Grading: Exercises given in class and 2 challenges from day 1, 1 from day 2 and 1 from day 3 in groups. I expect an Rmarkdown file and PDF or HTML.

Outline

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

Origins of decomposition

- We love rates.

Origins of decomposition

- We love rates.
- But crude rates are summary measures of population change

Origins of decomposition

- We love rates.
- But crude rates are summary measures of population change
- due to births, deaths or migration.
- These can differ due to underlying rates and population structure

Origins of decomposition

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

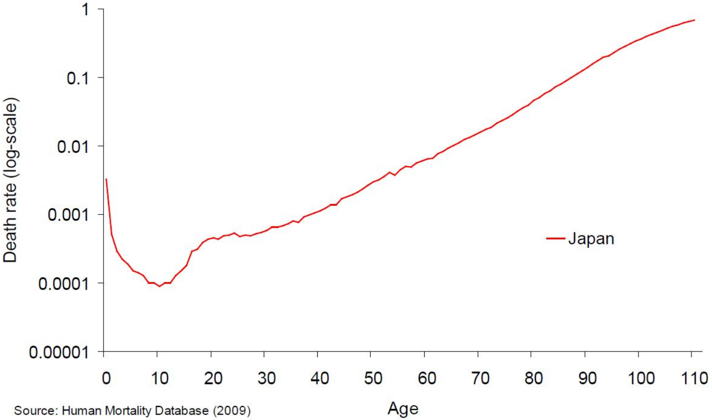
Origins of decomposition

- Methods of standardization
  Aim: Eliminate compositional effect from overall rates of some phenomenon.
  - Indirect standardization $\longrightarrow$ 1876
  - Direct standardization $\longrightarrow$ 1844

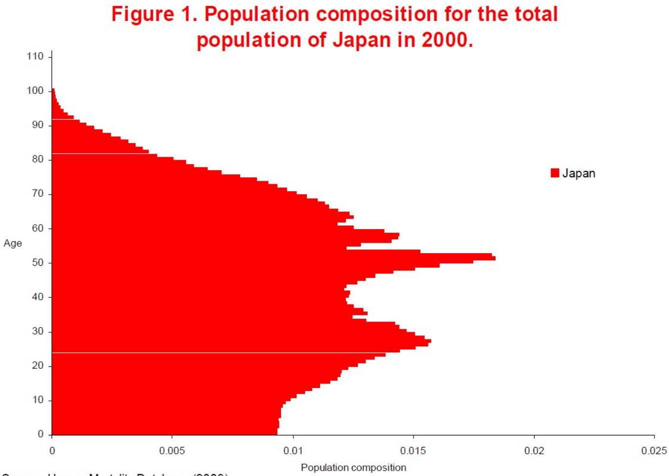
Origins of decomposition

- 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

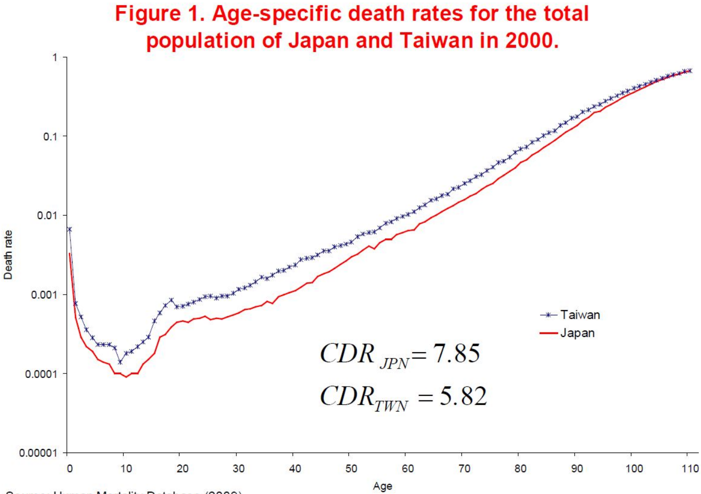
Figure 1. Age-specific death rates for the total population of Japan in 2000.



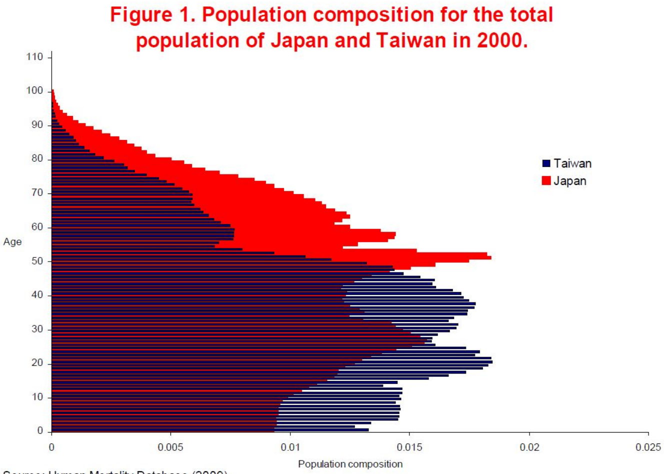
Source: Human Mortality Database (2009)



Source: Human Mortality Database (2009)



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)

t_1 is the initial period and t_2 is the final period
D_x = number of deaths at age x
M_x = death rate
N_x = is the mid-year population
N = total population over ages
Note that $D_x = M_x * N_x$

Kitagawa (1955)

$$\Delta CDR = \sum_x M_x(t_2)\frac{N_x(t_2)}{N(t_2)} - \sum_x M_x(t_1)\frac{N_x(t_1)}{N(t_1)} \tag{1}$$

Kitagawa (1955)

The difference between the crude rates can be expressed as

$$\Delta CDR = \underbrace{\sum_x M_x(t_1)\left[\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)}\right]}_{\text{Changes in x-composition}} + \underbrace{\sum_x \frac{N_x(t_1)}{N(t_1)}\left[M_x(t_2) - M_x(t_1)\right]}_{\text{Change in rates with pop 1 as standard}} + \underbrace{\sum_x (M_x(t_2) - M_x(t_1))\left[\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)}\right]}_{\text{Interaction of rates and compositions}} \tag{2}$$

Kitagawa (1955)

To avoid the interaction term, Kitagawa suggests

$$\Delta CDR = \underbrace{\sum_x \left(\frac{M_x(t_2)+M_x(t_1)}{2}\right)\left(\frac{N_x(t_2)}{N(t_2)}-\frac{N_x(t_1)}{N(t_1)}\right)}_{\text{Changes in x-composition}} + \underbrace{\sum_x \left(\frac{\frac{N_x(t_2)}{N(t_2)}+\frac{N_x(t_1)}{N(t_1)}}{2}\right)(M_x(t_2)-M_x(t_1))}_{\text{Changes in rates}} \quad (3)$$

Challenge 1: show that (1) can be expressed as (3) (step by step)

Kitagawa (1955)

Exercise 1: After class, follow Exercise 1 in the repository and apply the decomposition to two countries that are not the ones in the exercise. (you can use data from HMD, UN or your country's stats office.

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 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.
- 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.

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)

Kitagawa (1955)

Challenge 2: With data on fertility (e.g. HFD) select 3 countries and analyze the change in their crude fertility rate (CFR) in a recent period (10 years) and decompose these changes following Kitagawa's decomposition and describe your results. Then for the most recent period select the two countries (among the 3) with the highest and lowest CFR and decompose their difference and describe your results.

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

$$E(v) = \bar{v}(y) = \frac{\int_0^\infty v(x,y)w(x,y)dx}{\int_0^\infty w(x,y)dx}, x \text{ continuous}$$
$$= \frac{\sum_x v(x,y)w(x,y)dx}{\sum_x w(x,y)dx}, x \text{ discrete} \qquad (4)$$

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

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

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

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

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

$$\dot{\bar{v}} = \underbrace{\bar{\dot{v}}}_{\text{Direct component}} + \underbrace{Cov(v, \acute{w})}_{\text{Structural or compositional component}} \tag{5}$$

$$\dot{\bar{v}} = \underbrace{\bar{\dot{v}}}_{\text{Direct component}} + \underbrace{Cov(v, \acute{w})}_{\text{Structural or compositional component}} \tag{5}$$

Main result in Vaupel & Canudas-Romo (2002)

After class, follow Exercise 2 in the repository and apply the decomposition to a country different from Denmark and not in Exercise 1. (you can use data from HMD, UN or your country's stats office.)

Challenge 3: Following Vaupel and Canudas-Romo (2002) (Table 1), perform the decomposition of the change in the average age at childbearing for some country (not in table 1) for a recent period.

Preliminaries:

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

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

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

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

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

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

The exponential: a Rock-star

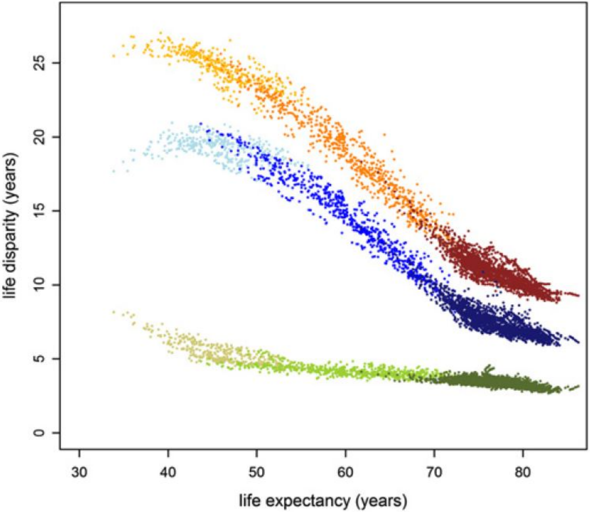
Following Vaupel & Canudas-Romo (2002)

$$\dot{e}_o(t) = \int_0^\infty \rho(x)e(x)f(x)dx \qquad (6)$$

can be written as:

$$\dot{e}_o(t) = \bar{\rho}(t)e^\dagger(t) + Cov(\rho, e_x) \qquad (7)$$

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

Challenge 4: Decompose the life expectancy for USA in the most recent decade of data available in HMD following Vaupel & Canudas-Romo (2003). Describe and plot your results.