Decomposition - Class 2

Serena Vigezzi and José Manuel Aburto

Recap

- We decompose to better understand aggregate measures changes/differences
- Decomposition methods improve on existing standardisation methods

Kitagawa decomposition

Decomposing a weighted mean into

- Difference in rates
- Difference in population composition (weight)
  $\rightarrow$ can be along any categorical dimension
  $\rightarrow$ age, marital status, gender, education

Quiz time!

Kitagawa decomposition: extensions

(Kitagawa-)Oaxaca-Blinder decomposition method

- Separately proposed by Ronald Oaxaca and Alan Blinder in 1973
- Generalisation of Kitagawa method (Oaxaca, 2025)
- Regression-based: it can include multiple independent variables

Kitagawa decomposition: extensions

Das Gupta decomposition method

- Proposed by Prithwis Das Gupta in 1978
- Kitagawa can be used with more than one compositional factor, but this creates interaction terms
- Das Gupta eliminates these terms by equally redistributing them across the main terms (symmetric approach)

Kitagawa decomposition: extensions

Chevan and Sutherland decomposition method (extension of Das Gupta's method)

- Proposed by Albert Chevan and Micheal Sutherland in 2009
- Estimates contributions of specific categories vs of overall variables (e.g. married individuals vs marital status)
- Apply decomposition to polytomous dependent variable

Kitagawa decomposition and extensions

Decomposing a crude rate into

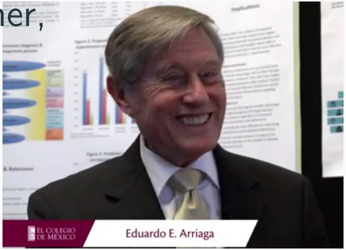
- Difference in rates
- Difference in population composition
  $\rightarrow$ can be along any categorical dimension
  $\rightarrow$ age, marital status, gender, education

Various extensions

What about other measures?

- Kitagawa: analytical decomposition of a crude rate
- One can do that with more aggregate measures
- For example, life expectancy

Eduardo E. Arriaga



- Spanish-born, Argentinian and US demographer, worked in US and Argentinian universities
- Mostly worked on mortality, also fertility, population change and urbanisation
- Focus on Latin America
- *Measuring and explaining the change in life expectancies*, 1984 $\rightarrow$ similar approaches developed by other researchers at the same time

Arriaga decomposition

Decompose difference in life expectancy by age

Total change is the sum of age-specific changes:

$$\Delta LE = \sum_x ({}_n\Delta_x)$$

Arriaga decomposition

$$
{}_n\Delta_x = \underbrace{\frac{l_x^1}{l_0^1}\left(\frac{{}_nL_x^2}{l_x^2} - \frac{{}_nL_x^1}{l_x^1}\right)}_{\text{Direct effect}} + \underbrace{\frac{T_{x+n}^2}{l_0^1}\left(\frac{l_x^1}{l_x^2} - \frac{l_{x+n}^1}{l_{x+n}^2}\right)}_{\text{Indirect and interaction effects}}
$$

Change in mortality rates between ages x and $x+n$ affects

- Number of years lived between ages x and $x+n$ $\rightarrow$ direct effect
- Number of years lived after age $x+n$ by new survivors $\rightarrow$ indirect and interaction effect

Arriaga decomposition

$$
{}_n\Delta_x = \underbrace{\frac{l_x^1}{l_0^1}\left(\frac{{}_nL_x^2}{l_x^2} - \frac{{}_nL_x^1}{l_x^1}\right)}_{\text{Direct effect}} + \underbrace{\frac{T_{x+n}^2}{l_0^1}\left(\frac{l_x^1}{l_x^2} - \frac{l_{x+n}^1}{l_{x+n}^2}\right)}_{\text{Indirect and interaction effects}}
$$

Change in mortality rates between ages x and $x+n$ affects

- Number of years lived between ages x and $x+n \rightarrow$ direct effect
- Number of years lived after age $x+n$ by new survivors $\rightarrow$ indirect and interaction effect

For open-ended age-group there is only a direct effect:

$${}_\infty\Delta_x = \frac{l_x^1}{l_0^1}\left(\frac{T_x^2}{l_x^2} - \frac{T_x^1}{l_x^1}\right)$$

An example

PNAS RESEARCH ARTICLE | DEMOGRAPHY OPEN ACCESS

Check for updates

Significant impacts of the COVID-19 pandemic on race/ethnic differences in US mortality

José Manuel Aburto[a,b,c,1], Andrea M. Tilstra[a,d,1], Ginevra Floridi[a], and Jennifer Beam Dowd[a]

Edited by Douglas Massey, Princeton University, Princeton, NJ; received April 2, 2022; accepted June 20, 2022

An example

Let's try ourselves

Arriaga decomposition by causes of death

We can extend this method to account for contributions from different causes of death i:

$$
\begin{aligned}
{}_n\Delta_x^i &= {}_n\Delta_x \frac{{}_nm_{x,i}^2 - {}_nm_{x,i}^1}{{}_nm_x^2 - {}_nm_x^1} \\
&= {}_n\Delta_x \frac{{}_nR_{x,i}^2({}_nm_x^2) - {}_nR_{x,i}^1({}_nm_x^1)}{{}_nm_x^2 - {}_nm_x^1}
\end{aligned}
$$

Where

- ${}_nR_{x,i}^t$ is the proportion of deaths from cause i in ages x to $x+n$ and population t
- ${}_n\Delta_x$ is the contribution of all-cause mortality differences in the same age group
- ${}_nm_x^t$ is the mortality rate in the same age group of population t

Arriaga decomposition by causes of death

All-cause contribution, weighted by the change in cause-specific mortality as a proportion of change in all-cause mortality

An example

Back to R!

Kitagawa-Arriaga combination

New method to decompose differences in life expectancy by population composition

- Proposed by Riffe, Tursun-Zade and Trias Llimós in 2024 (working paper)
- Kitagawa decomposition to distinguish e_x (rate) vs composition effects
- Arriaga decomposition of differences in e_x
- Rescaling to make everything match

References

R. Oaxaca (1973). "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, 14(3): 693—709.

A.S. Blinder (1973). "Wage Discrimination. Reduced Form and Structural Estimates", *The Journal of Humand Resources*, 8(4): 436—455. P. Das Gupta (1978). "A general method of decomposing a difference between two rates into several components", *Demography*, 15: 99—112. A. Chevan and M. Sutherland (2009). "Revisiting Das Gupta: Refinement and extension of standardization and decomposition", *Demography*, 46: 429—449.

E.E. Arriaga (1984). "Measuring and explaining the change in life expectancies", *Demography*, 21(1): 83—96.

T. Riffe, R. Tursun-Zade, S. Trias Llimós (2024). "Arriaga meets Kitagawa. Life expectancy decomposition with population subgroups", MPIDR Working Paper 2024-029.