Arriaga meets Kitagawa. life expectancy decompositions including population subgroups

Dr. Timothy Riffe[1,2,3*]

[1*]Department of Sociology and Social Work, University of the Basque Country (UPV/EHU), Leioa, Spain, 48940, Bizkaia, Barrio Sarriena s/n.
[2]Ikerbasque (Basque Foundation for Science).
[3]Laboratory of Population Health, Max Planck Institute for Demographic Research.

Corresponding author(s). E-mail(s): tim.riffe@ehu.eus;

Abstract

Background: An Arriaga (1984) decomposition allows us to decompose differences in life expectancy into contributions due to mortality rate differences in each age. A Kitawaga (1955) decomposition allows us to decompose differences in a weighted mean into an effect from differences in structure and effects from differences in each element of the weighted value.

Objective: Sometimes we would like to decompose a difference between two populations that are each composed of like-defined subpopulations. Said decomposition would produce effects for each age of each subpopulation, as well as a marginal effect of composition differences. I propose a straightforward analytic method to do this.

Methods: In short, within-subpopulation life expectancy differences can be handled by the Arriaga method. The case of weighting together the life expectancies of subpopulations by way of a composed radix can be handled using the well-known Kitagawa method. The elements of the value component of the Kitagawa method tell us how to rescale the Arriaga results specific to each subpopulation.

Results: I show the proposed analytic method to be equivalent to a Horiuchi et al (2008) reframing of the same problem. Some mentionable properties: (i) There is no limit to the number of subpopulations, (ii) it is straightforward to incorporate cause-of-death information, (iii) composition is here only considered in the radix age. I currently have results for simulated mortality rates, but I promise to wrangle up an empirical application to demonstrate the method.

Conclusions: The analytic decomposition method I propose is advantageous compared to a Horiuchi method for this problem purely for reasons of computational efficiency. This method could help further disentangle the effects of

mortality and composition differences in explaining or clarifying paradoxes or secular change. I promise to think further about subpopulation weighting that might occur over all ages (using prevalence information).}

1 Background

An Arriaga (1984) decomposition allows us to decompose changes in life expectancy into contributions due to mortality changes in different ages. The method was designed to be practical, and framed in terms of age lifetable columns expressed in discrete age groups. A well-known property of the method is that mortality changes in different ages need not be proportional. Derived contributions also sum exactly to the observed life expectancy difference. A not-well-known property of the method is that it is asymmetrical, in the sense that the absolute values of age-specific contributions depend on whether we compare population 1 with 2, or population 2 with 1. A further property of the method is that it is designed to work with homogeneous populations, meaning that populations 1 and 2 are each homogeneous, in the sense that each is expressed with only one lifetable.

A Kitagawa (1955) decomposition allows us to decompose differences in a weighted mean due to differences in weights (structure) and differences in the value being weighted (often rates, but in our case life expectancies). This widely-used decomposition method is well-known to be exact in that the resulting structure and value components sum exactly to the observed difference in weighted means. The individual elements (ages, or life expectancies for us) of the value being weighted have identifiable effects. It is not well-known that the individual elements of the structure component do not have identifiable effects. Rather, the structure effects should be summed to a a marginal effect due to differences in structure.

2 Introduction

A demographer might like to decompose a difference in life expectancy between two populations (or one population at two times) that are each composed of like-defined subpopulations. For example, life expectancy in France versus Spain, each with education-specific subpopulations, or a national life expectancy calculated as the weighted average of its regions at two points in time. In such situations, a decomposition would ideally tell us the contribution to the difference in overall life expectancy due to rate differences in each age in each subpopulation, and also separate an effect due to overall compositional change. In this paper, I would like to propose a straightforward analytic method to decompose in this way. In short, within-subpopulation life expectancy differences can be handled by the Arriaga method. The case of weighting together the life expectancies of subpopulations by way of a composed radix can be

handled using the well-known Kitagawa method. The elements of the value component of the Kitagawa method tell us how to rescale the Arriaga results specific to each subpopulation so as to isolate the age-subpopulation-specific effects on the overall life expectancy difference. I will justify this rescaling, and show that the results of this procedure are fully consistent with the results of a Horiuchi, Wilmoth, and Pletcher (2008) reframing of the problem.

3 Method

3.1 Notation

In formulas, we use the following variables and scripting, most of which are lifetable columns:

- $\ell(a)$ lifetable survivorship at exact age a.
- $_nL_a$ lifetable person years lived in the interval $[a, a+n)$.
- T_a total lifetable person years lived beyond age a.
- e_a remaining life expectancy at exact age a.
- π^s the initial population fraction for subgroup s at the initial age.

We also use the superscript s to index subpopulations comprising the total population, and the superscript t to index time points, or some other way of differentiating the total population. The index $e_0^{s,t}$ reads as "life expectancy at birth for subpopulation s at time t", e.g. $e_0^{low,1990}$ could be the life expectancy in 1990 of a low education group.

3.2 Averaging life expectancy

There are two major approaches to producing average life expectancy for a total population composed of subgroups, as Vaupel (2002) points out. One could derive the average mortality rate in each age by summing the respective deaths and exposures from subpopulations and then deriving the rate, as the Human Mortality Database Wilmoth et al. (2021) does, then recalculate the total population lifetable from these. This is what Vaupel called *current rates* approach, which ignores even controllable heterogeneity. The other approach derives subgroup-specific lifetables, then weights life expectancies together according to an initial mixing composition in order to arrive at the total life expectancy. This is Vaupel's *current conditions* approach, at least in part. This second approach is what is often done in multistate models (cite), but also sometimes with standard lifetables, for example, when doing between-within decompositions of variance or other similar summary measures. Our decomposition method deals with this second case, specifically, we have:

$$e_0 = \sum_s \pi^s \cdot e_0^s \quad, \tag{1}$$

where the S total subgroup fractions comprise a mixing distribution, $1 = \sum_s \pi^s$.

3.3 Kitagawa decomposition

Eq (1) treats the total life expectancy as a weighted average, which means we can decompose a change over time in e_0 ($\Delta = e_0^2 - e_o^1$) exactly using the formulas from Kitagawa (1955). Specifically, eq (2) gives an overall effect of changes in composition:

$$\Delta^{\text{composition}} = \sum_s \left(\pi^{s,t2} - \pi^{s,t1} \right) \cdot \frac{e_0^{s,t1} + e_0^{s,t2}}{2} \quad , \tag{2}$$

This result is widely known, but many are unaware that the composition effect should be summed like this, since the group-specific composition (structure) effects are not well-identified. Eq (5) gives the subgroup-specific effects from changes in life expectancy:

$$\Delta^{e_0^s \text{ change}} = \left(e_0^{s,t2} - e_0^{s,t1} \right) \cdot \frac{\pi^{s,t1} + \pi^{s,t2}}{2} \quad . \tag{3}$$

Eq (4) says that the overall change in life expectancy is the sum of (i) a single component that captures the effect of composition change and (ii) a set of components describing the effect on overall life expectancy from each group's specific life expectancy change. This second effect could just as well be called the *rate* effect, since each life expectancy is fully determined by mortality rates.

$$\Delta = \Delta^{\text{composition}} + \sum_s \Delta^{e_0^s \text{ change}} \tag{4}$$

3.4 Symmetrical Arriaga decomposition

The above-mentioned *rate* effect can be conceived of as the *net* effect on total life expectancy change of the age-specific effects within subgroups, as isolated by the many varieties of life expectancy decompositions. In this case, it does not matter much which method we choose to derive age-specific effects for subgroup-specific changes in life expectancy. The Arriaga (1984) decomposition approach is popular in part because it is efficient, analytic, and sums exactly to the observed life expectancy difference. This method can be implemented in a number of relevant ways, per Riffe et. al. (2024), from which we here consider only *symmetrical* Arriaga decomposition. This implies following the original formulas twice, once in each direction, and averaging the sign-adjusted results. We repeat this exercise for each subgroup to decompose group-specific changes in life expectancy (hence the ^{s} on everything):

$$\begin{aligned} \Delta^s &= e_0^{s,2} - e_0^{s,1} \\ &= \sum_x {}_n\Delta_x^s \quad , \end{aligned} \tag{5}$$

where Δ^s is the subgroup-specific (s) difference in life expectancy being decomposed, and it is composed of age-specific contributions, $\overrightarrow{{}_n\Delta_x^s}$ or $\overleftarrow{{}_n\Delta_x^s}$ depending on whether we decompose forward in time or backward, respectively (Riffe et. al., 2024). The

forward age-specific values $\overrightarrow{_n\Delta_x^{\acute{s}}}$ can be calculated following Arriaga's decomposition method following eq (6), consistent with Arriaga (1984) or the presentation in Preston, Heuveline, and Guillot (2000). I use a lifetable radix of 1, meaning $\ell_0 = 1$, to reduce the formula slightly.

$$\overrightarrow{_n\Delta_x^{\acute{s}}} = \begin{cases} \ell_x^{s,1} \cdot \left(\frac{_nL_x^{s,2}}{\ell_x^{s,2}} - \frac{_nL_x^{s,1}}{\ell_x^{s,1}} \right) + T_{x+n}^{s,2} \cdot \left(\frac{\ell_x^{s,1}}{\ell_x^{s,2}} - \frac{\ell_{x+n}^{s,1}}{\ell_{x+n}^{s,2}} \right) & \forall x < \omega \\ \ell_\omega^{s,1} \cdot (e_\omega^{s,2} - e_\omega^{s,1}) & \forall x = \omega \end{cases} \quad (6)$$

Equation (6) is the first pass of our symmetrical decomposition, and (7) is the second pass, identical except all time superscripts are swapped.

$$\overleftarrow{_n\Delta_x^{s}} = \begin{cases} \ell_x^{s,2} \cdot \left(\frac{_nL_x^{s,1}}{\ell_x^{s,1}} - \frac{_nL_x^{s,2}}{\ell_x^{s,2}} \right) + T_{x+n}^{s,1} \cdot \left(\frac{\ell_x^{s,2}}{\ell_x^{s,1}} - \frac{\ell_{x+n}^{s,2}}{\ell_{x+n}^{s,1}} \right) & \forall x < \omega \\ \ell_\omega^{s,2} \cdot (e_\omega^{s,1} - e_\omega^{s,2}) & \forall x = \omega \end{cases} \quad (7)$$

Importantly,

$$-\Delta^s = \sum_x \overleftarrow{_n\Delta_x^{s}} \quad (8)$$
$$= e_0^1 - e_0^2$$

Then our symmetrical estimate of $_n\Delta_x^s$ is just the average of these:

$$_n\Delta_x^s = \frac{\left(\overrightarrow{_n\Delta_x^{\acute{s}}} - \overleftarrow{_n\Delta_x^{s}} \right)}{2} \quad (9)$$

Repeat the above to derive $_n\Delta_x^s$ for each within-subpopulation comparison.

3.5 Rescale Arriaga results

The age-specific quantities $_n\Delta_x^s$ derived from eq (9) are on the scale of subgroup-specific life expectancy changes. To know the net impact of $_n\Delta_x^s$ on overall life expectancy change we rescale per eq (10) to match the effect sizes from eq (5)

$$_n\Delta_x^{s,\text{net}} = \Delta^{e_0^s \text{ change}} \cdot \frac{_n\Delta_x^s}{\Delta^s} \quad . \quad (10)$$

With eqs (2) and (10), we have all elements of the present decomposition:

$$\Delta = \Delta^{\text{composition}} + \sum_s \sum_x {_n\Delta_x^{s,\text{net}}} \quad (11)$$

Supplementary information. If your article has accompanying supplementary file/s please state so here.

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading

3.6 Funding

3.7 Conflict of interest

The author declares no conflict of interest

3.8 Availability of data and materials

All data are ___

3.9 Code availability

All code used here is available in an open `GitHub` repository https://github.com/timriffe/arriagakitagawa

Appendix A Section title of first appendix

An appendix contains supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem or it is information that is too cumbersome to be included in the body of the paper.

For submissions to Nature Portfolio Journals please use the heading "Extended Data".

References

Arriaga, E.E. (1984). Measuring and explaining the change in life expectancies. *Demography*, *21*, 83–96, https://doi.org/10.2307/2061029

Horiuchi, S., Wilmoth, J.R., Pletcher, S.D. (2008). A decomposition method based on a model of continuous change. *Demography*, *45*, 785–801, https://doi.org/10.1353/dem.0.0033

Kitagawa, E.M. (1955). Components of a difference between two rates. *Journal of the american statistical association*, *50*(272), 1168–1194, https://doi.org/10.1080/01621459.1955.10501299

Preston, S., Heuveline, P., Guillot, M. (2000). *Demography: measuring and modeling population processes.* Wiley-Blackwell.

Vaupel, J.W. (2002). Life expectancy at current rates vs. current conditions: A reflexion stimulated by bongaarts and feeney's "how long do we live?". *Demographic Research*, *7*, 365–378,

Wilmoth, J., Andreev, K., Jdanov, D., Glei, D., Riffe, T., Boe, C., . . . M, B. (2021). Methods protocol for the human mortality database. version 6. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality.org [version 26/01/2021], ,*