Demographic Research a free, expedited, online journal
of peer-reviewed research and commentary
in the population sciences published by the
Max Planck Institute for Demographic Research
Konrad-Zuse Str. 1, D-18057 Rostock · GERMANY
www.demographic-research.org

DEMOGRAPHIC RESEARCH

VOLUME 27, ARTICLE 15, PAGES 419-428
PUBLISHED 12 SEPTEMBER 2012
http://www.demographic-research.org/Volumes/Vol27/15/
DOI: 10.4054/DemRes.2012.27.15

Formal Relationships 19

The difference between alternative averages

James W. Vaupel

Zhen Zhang

This article is part of the Special Collection “Formal Relationships”.
Guest Editors are Joshua R. Goldstein and James W. Vaupel.

© *2012 James W. Vaupel & Zhen Zhang.*

This open-access work is published under the terms of the Creative Commons Attribution NonCommercial License 2.0 Germany, which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit.
See http://creativecommons.org/licenses/by-nc/2.0/de/

Table of Contents

Demographic Research: Volume 27, Article 15
Formal Relationships 19

The difference between alternative averages

James W. Vaupel [1, 3]

Zhen Zhang [2, 3]

Abstract

BACKGROUND
Demographers have long been interested in how compositional change, e.g., change in age structure, affects population averages.

OBJECTIVE
We want to deepen understanding of how compositional change affects population averages.

RESULTS
The difference between two averages of a variable, calculated using alternative weighting functions, equals the covariance between the variable and the ratio of the weighting functions, divided by the average of the ratio. We compare weighted and unweighted averages and also provide examples of use of the relationship in analyses of fertility and mortality.

COMMENTS
Other uses of covariances in formal demography are worth exploring.

[1] Max Planck Institute for Demographic Research. E-mail: jwv@demogr.mpg.de.
[2] Corresponding author: Max Planck Institute for Demographic Research. E-mail: zhang@demogr.mpg.de.
[3] The authors would like to thank the two anonymous reviewers for pointing out missing references and making constructive comments which improved the quality of the paper.

1. Relationship

Let $\overline{v}$ denote the mean average of $v(x)$ over x,

(1a) $$\overline{v} = \frac{\int v(x)w(x)dx}{\int w(x)dx}$$

if x is continuous and

(1b) $$\overline{v} = \frac{\sum_x v(x)w(x)}{\sum_x w(x)}$$

if x is discrete, where $v(x)$ is some demographic function, e.g., a birth rate or death rate, and $w(x)$ is some weighting function, e.g., population size or number of children. The variable x could denote, e.g., age, time, individual, family or country.

Let $w(x)$ and $w^*(x)$ be two alternative weighting functions with

(2) $$\phi(x) = \frac{w^*(x)}{w(x)}.$$

Let $\overline{v}$ and $\overline{v}^*$ be the mean values of $v(x)$ weighted by $w(x)$ and $w^*(x)$, respectively. Then

(3) $$\overline{v}^* - \overline{v} = \frac{v \cap \varphi}{\overline{\varphi}},$$

where $v \cap \varphi$, spoken "v covφ", is a simpler notation than the usual $Cov(v, \varphi)$ to denote the covariance between v and φ (Vaupel 1992). Because the covariance of two functions is the average value of the product of the functions minus the product of their average values, this covariance is defined by

(4a) $$v \cap \varphi = \overline{v\varphi} - \overline{v}\,\overline{\varphi}.$$

In more detail and in the continuous case, the formula is

(4b) $$v \cap \varphi = \frac{\int v(x)\varphi(x)w(x)dx}{\int w(x)dx} - \frac{\int v(x)w(x)dx}{\int w(x)dx}\frac{\int \varphi(x)w(x)dx}{\int w(x)dx}.$$

Note that in calculating the covariance the weight that is used is $w(x)$.

2. Proof

It follows from (1) and (2) that

$$\begin{aligned} \overline{v}^* &= \frac{\int v(x)w^*(x)dx}{\int w^*(x)dx} \\ &= \frac{\int v(x)w(x)\varphi(x)dx}{\int w(x)\varphi(x)dx} \\ &= \frac{\int v(x)w(x)\varphi(x)dx}{\int w(x)dx}\frac{\int w(x)dx}{\int w(x)\varphi(x)dx} \\ &= \frac{\overline{v\varphi}}{\overline{\varphi}}. \end{aligned} \tag{5}$$

This implies

$$\overline{v}^* - \overline{v} = \frac{\overline{v\varphi}}{\overline{\varphi}} - \overline{v} = \frac{\overline{v\varphi} - \overline{v}\,\overline{\varphi}}{\overline{\varphi}} = \frac{v \cap \varphi}{\overline{\varphi}}. \tag{6}$$

Q.E.D.

3. History and related results

The age structure of a population is often the weighting function used to compute demographic averages. Analysis of how change in age structure affects averages was first addressed by the method of indirect standardization, which, as discussed by Keiding (1987), can be traced back to Dale (1777). Equation (3) can be used to study how much difference it makes if indirect standardization is done using one set of weights vs. another set of weights. Preston, Himes and Eggers (1989), following a suggestion from Ansley Coale, use the covariance in their study of change in a population's average age; they find that the change in average age is equal to the covariance between age and age-specific growth rates. This result was generalized by Vaupel (1992), Schoen and Kim (1992) and Vaupel and Canudas-Romo (2002). Our main result, equation (3), can be viewed either as an elaboration of the Preston et al. formula or an application of the Vaupel formula. Although the covariance is widely used in statistics and in biology, we do not know of any other uses of the covariance in formal demography.

The covariance is closely related to Pearson's correlation coefficient ρ :

$$v \cap u = \sigma(v)\sigma(u)\rho(v, u), \tag{7}$$

where $\sigma(v)$ and $\sigma(u)$ are the standard deviations of the functions v and u. The covariance has the following noteworthy properties:

(8) $$v \cap v = \sigma^2(v),$$

where σ^2 denotes the variance,

(9) $$v \cap u = u \cap v,$$

(10) $$kv \cap u = k(v \cap u)$$

and

(11) $$(v_1 + v_2) \cap u = v_1 \cap u + v_2 \cap u.$$

4. Applications

4.1 Average family sizes weighted by mothers and by children

Preston (1976) considers the relationship between the number of children mothers have and the number of siblings children have. The average number of children in a family is given by

(12) $$\bar{i} = \sum iN(i) / \sum N(i),$$

where $N(i)$ is the number of mothers with children. For a child's perspective, the average number of children per family is

(13) $$\bar{i}^* = \sum iiN(i) / \sum iN(i),$$

where the denominator is the total number of children. Hence $\varphi = i$ and $\overline{\varphi} = \bar{i}$. Applying equation (3) leads to Preston's elegant result that

(14) $$\bar{i}^* - \bar{i} = \frac{{\sigma_i}^2}{\bar{i}}$$

where ${\sigma_i}^2 = i \cap \varphi = i \cap i$ is the variance of the number of children per mother. This is an example of length-biased sampling: Equation (3.3) in Cox (1969) is identical to equation (14).

4.2 Effect of changes in fertility

Preston, Heuveline and Guillot (2001; pp.157-8) derive an equation to account for the effect of a fertility-induced increase in r on the death rate $\overline{\mu} = \int_0^\omega c(a)\mu(a)da$ in a stable population:

$$\frac{d \ln \overline{\mu}}{dr} = A_P - A_D, \tag{15}$$

where A_P is the mean age of the stable population:

$$A_P = \frac{\int_0^\omega c(a)ada}{\int_0^\omega c(a)da}$$

and A_D is the mean age at death in the stable population:

$$A_P = \frac{\int_0^\omega c(a)\mu(a)ada}{\int_0^\omega c(a)\mu(a)da}$$

The function $c(a) = be^{-ra}p(a)$ gives the age distribution of the population with $p(a)$ denoting the probability of surviving to age a and $\mu(a)$ denoting the force of mortality at age a .

Note that

$$\varphi(a) = \frac{c(a)\mu(a)}{c(a)} = \mu(a).$$

Applying equation (3) to the right side of equation (15) gives

$$A_D - A_P = \frac{a \cap \varphi}{\bar{\varphi}} = \frac{a \cap \mu}{\bar{\mu}}, \tag{16}$$

which implies

(17) $$\frac{d\ln\bar{\mu}}{dr} = -\frac{a \cap \mu}{\bar{\mu}}.$$

Note that the left side of the above equation is the relative change in $\bar{\mu}$. Hence the absolute change in $\bar{\mu}$ will be

(18) $$\frac{d\bar{\mu}}{dr} = a \cap \mu.$$

If the correlation between age and the force of mortality is positive, as it generally is in modern human populations, then the death rate will fall when fertility rises. In human populations with high infant and childhood mortality but with such high fertility that the population is growing, the covariance can be negative. Furthermore, for many trees, fish and other non-human populations mortality tends to fall with age (Vaupel et al., 2004): for such species the death rate will rise when fertility rises.

In a stationary or lifetable population, the inverse of the death rate equals life expectancy e_o and the average age at death also equals life expectancy. Consequently, (16) implies

(19) $$\frac{A_P}{e_o} = 1 - a \cap \mu.$$

An extreme case is when no one dies until an age X when everyone dies. In this case, life expectancy is X and the average age of the living is $X/2$, so the covariance is 0.5. If mortality is constant over age, then the covariance is zero: the average age of the living and the average age at death are equal. If age is negatively correlated with the force of mortality, then the living are, on average, older than life expectancy.

4.3 Change in number of births due to population growth

Schoen and Kim (1992) derive a formula which is equivalent to the discrete form of equation (3). Schoen and Kim consider the number of births at time t, given by

$$B_t = \sum_{a-0}^{\infty} N_{a,t} m_{a,t},$$

where $N_{a,t}$ is the population size at age a and time t and $m_{a,t}$ is age-specific fertility. They define population growth over a period of unit time by

$$\exp(r_{a,t}) = \frac{N_{a,t}}{N_{a,t-1}}. \tag{20}$$

If age-specific fertility is constant over time, $m_{a,t} = m_{a,t-1}$ for all t, the change in the number of births from time $t-1$ to t is

$$B_t - B_{t-1} = \frac{N_{t-1}}{N_t} m \cap \exp(r) \tag{21}$$

where $N_t = \sum_{a=0}^{\infty} N_{a,t}$ is the total population at time t. This equation is equation (23) in the Schoen and Kim article. It can be seen that the equation is equivalent to equation (3), by noting an interesting property of the growth rate defined in equation (20):

$$\overline{\exp(r)} = \frac{\sum_{a=0}^{\infty} \exp(r_{a,t}) N_{a,t-1}}{\sum_{a=0}^{\infty} N_{a,t-1}} = \frac{\sum_{a=0}^{\infty} \left(\frac{N_{a,t}}{N_{a,t-1}}\right) N_{a,t-1}}{\sum_{a=0}^{\infty} N_{a,t-1}} = \frac{N_t}{N_{t-1}}.$$

Substituting this result in equation (21) leads to

$$B_t - B_{t-1} = \frac{m \cap \exp(r)}{\overline{\exp(r)}} = \frac{m \cap \varphi}{\bar{\varphi}}. \tag{22}$$

with $\varphi = \exp(r_{a,t}) = N_{a,t}/N_{a,t-1}$.

4.4 Weighted vs. unweighted averages

Consider some demographic function $v(i)$ that is defined over populations designated by i, e.g., life expectancy for various countries or average income for different household sizes. Then the unweighted average is

$$\bar{\nu} = \frac{\sum_{i=1}^{n} v(i)}{n}, \tag{23}$$

whereas the average weighted by population size $N(i)$ is

(24)
$$\bar{\nu}^* = \frac{\sum_{i=1}^{n} N(i)\nu(i)}{\sum_{i=1}^{n} N(i)}.$$

Hence (3) implies

(25)
$$\bar{\nu}^* - \bar{\nu} = \frac{\nu \cap N}{\overline{N}},$$

where $\overline{N} = \sum N/n$ is the average population size. The weighted average will exceed the unweighted average if the variable of interest is positively correlated with population size.

As an example, consider average household income. The unweighted average was about $57 thousand in total income per household in the US in 2010. The corresponding average weighted by the number of persons in a household was $61 thousand (US Census Bureau 2011). The $57 thousand figure gives the average income per household whereas the $61 thousand figure gives the average household income per person. The covariance is positive because bigger households tend to have more total income. In particular, the total income of households with two members was $56 thousand vs. $31 thousand for households with one member.

References

Cox, D. (1969). Some sampling problems in technology. In: H. Johnson N.L. Smith (ed.) *New Developments in Survey Sampling*. New York: Wiley.

Dale, W. (1777). *A Supplement to Calculations of the Value of Annuities, Published for the Use of Societies Instituted for Benefit of Age Containing Various Illustration of the Doctrine of Annuities, and Compleat Tables of the Value of 1£. Immediate Annuity. (Being the Only Ones Extant by Half-Yearly Interest and Payments). Together with Investigations of the State of the Laudable Society of Annuitants; Showing What Annuity Each Member Hath Purchased, and Real Mortality Therein, from its Institution Compared with Dr. Halley's Table. Also Several publications, Letters, and Anecdotes Relative to that Society. And Explanatory of Proceedings to the Present year*. London: Ridley.

Keiding, N. (1987). The method of expected number of deaths 1786-1886-1986. *International Statistical Review* 55(1): 1–20. doi:10.2307/1403267.

Preston, S. (1976). Family sizes of children and family sizes of women. *Demography* 13(1): 105–114. doi:10.2307/2060423.

Preston, S., Heuveline, P., and Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers.

Preston, S., Himes, C., and Eggers, M. (1989). Demographic conditions responsible for population aging. *Demography* 26(4): 691–703. doi:10.2307/2061266.

Schoen, R. and Kim, Y. (1992). Covariances, roots, and the dynamics of age-specific growth. *Population Index* 58(1): 4–17. doi:10.2307/3643798.

US Census Bureau (2011). Current population survey, annual social and economic supplement. Washington: US Census Bureau. Http://www.census.gov/cps/. (Retrieved 2012-07-15).

Vaupel, J.W. (1992). Analysis of population changes and differences: Methods for demographers, statisticians, biologists, epidemiologists, and reliability engineers. (Paper (107 pp.) presented at the PAA Annual Meeting held in Denver, Colorado, April 30 - May 2 1992).

Vaupel, J.W., Baudisch, A., Dölling, M., Roach, D., and Gampe, J. (2004). The case for negative senescence. *Theoretical Population Biology* 65(4): 339–351. doi:10.1016/j.tpb.2003.12.003.

Vaupel, J.W. and Canudos-Romos, V. (2002). Decomposing demographic change into direct vs. compositional components. *Demographic Research* 7(1): 1–14. doi:10.4054/DemRes.2002.7.1.