S. Krishnamoorthy

A Note on Sibling Survival as an Estimator of Adult Survival

SIBLING survival as an estimator of mortality has been studied by Hill and Trussell (1977) and they find that the proportion of surviving siblings for women of age x and $x + 5$ needs be only trivially altered to yield an estimate of life table $l(x)$ value. This finding as arrived at through a regression analysis of computed values of the proportions of surviving siblings by varying the parameters of the fertility and mortality schedules. *In this note, an attempt is made to show analytically the relationship between the proportion of surviving siblings and the life table $l(x)$ values.*

Model and Derivation

Starting point in this work is from Goodman, Keyfitz and Pullum (1974) who show that in a stable population the expected number of older sisters to a random girl aged a is

$$S(a) = \int_{\alpha}^{\beta} \left[\int_{\alpha}^{x} m(y)\, dy \right] e^{-ra}\, l(x)\, m(x)\, dx, \qquad (1)$$

where α and β are the lowest and highest ages of reproduction respectively, $m(x)\, dx$ is the chance that a woman of age x has a child in the next dx years, $l(x)$ is the probability that a new born girl survives to age x, and r is the intrinsic rate of natural increase. The expected number of younger sisters for a random girl aged a is shown to be

$$S'(a) = \int_{\alpha}^{\beta} \left[\int_{0}^{a} \{l(x + y)/l(x)\}\, m(x + y)\, dy \right] e^{-r\alpha}\, (x)\, m(x)\, dx. \qquad (2)$$

Given the above, it is not difficult to write the sum of ages of older sisters ever born, if they all survive, and it is

$$0(a) = \int_{\alpha}^{\beta} \left[\int_{\alpha}^{x} (x - y + a)\, m(y)\, dy \right] e^{-r\alpha}\, l(x)\, m(x)\, dx, \qquad (3)$$

and similarly the sum of ages of younger sisters ever born if they all survive is

$$Y(a) = \int_{\alpha}^{\beta} \left[\int_{0}^{a} \{ (a - y)\, l(x + y)\, m(x + y)/l(x)\}\, dy \right] e^{-r\alpha}\, l(x)\, m(x)\, dx. \qquad (4)$$

Now changing the order of integration in (3) from $\alpha < x < \beta$ and $\alpha < y < x$ to $\alpha < y < \beta$ and $y < x < \beta$, and exchanging the variables x and y, we have

$$0(a) = \int_{\alpha}^{\beta} \left[\int_{x}^{\beta} (a + y - x) e^{-r(y-\alpha)}\, l(y)\, m(y)\, dy \right] e^{-r\alpha}\, m(x)\, dx. \qquad (5)$$

For $a > \beta - \alpha$, we can write (4) as

$$Y(a) = \int_{\alpha}^{\beta} \left[\int_{x}^{\beta} (a - y + x)\, l(y)\, m(y)\, dy \right] e^{-r\alpha}\, m(x)\, dx. \qquad (6)$$

Applying the same procedure used to derive expression (5) from (3), we can write differently $S(a)$, as Goldman (1978) shows

$$S(a) = \int_{\alpha}^{\beta} \left[\int_{x}^{\beta} e^{-r(y-\alpha)}\, l(y)\, m(y)\, dy \right] e^{-r\alpha}\, m(x)\, dx, \qquad (7)$$

and for $a > \beta - \alpha$, expression (2) becomes

$$S'(a) = \int_{\alpha}^{\beta} \left[\int_{x}^{\beta} l(y)\, m(y)\, dy \right] e^{-r\alpha}\, m(x)\, dx. \qquad (8)$$

Therefore, the mean age of sisters ever born to a random girl aged a, if all her sisters had survived is

$$M(a) = \frac{0(a) + Y(a)}{S(a) + S'(a)}$$

which simplifies to

$$M(a) = a + \frac{\int_{\alpha}^{\beta} \left[\int_{x}^{\beta} (y - x) \{e^{-r(y-x)} - \quad l(y)\, m(y)\, dy \right] e^{-rx}\, m(x)\, dx}{\int_{\alpha}^{\beta} \left[\int_{x}^{\beta} \{e^{-r(y-x)} + 1\}\, l(y)\, m(y)\, dy \right] e^{-rx}\, m(x)\, dx}. \tag{9}$$

From (9) it can be seen that when $r = 0$ $M(a)$ equals a. That is for a stationary population the mean age of the siblings of a random respondent equals the age of the respondent, whose age is greater than $\beta - \alpha$.

Let us now call the second term on the right hand side of (9) as the correction factor (CF) and rewrite it as

$$CF = \frac{\int_{\alpha}^{\beta} \left[\int_{x}^{\beta} (y - x) \{e^{-ry} - e^{-rx}\}\, l(y)\, m(y)\, dy \right] m(x)\, dx}{\int_{\alpha}^{\beta} \left[\int_{x}^{\beta} \{e^{-ry} + e^{-rx}\}\, l(y)\, m(y)\, dy \right] m(x)\, d(x)}. \tag{10}$$

In order to obtain an approximate value of the CF, let us make the reasonable assumption that $l(x)$ is constant within the reproductive age interval and, therefore, drop it from (10) and write CF as below.

$$CF \approx \frac{\int_{\alpha}^{\beta} \int_{x}^{\beta} (y - x) \{e^{-ry} - e^{-rx}\}\, m(y)\, m(x)\, dy\, dx}{\int_{\alpha}^{\beta} \int_{x}^{\beta} \{e^{-ry} + e^{-rx}\}\, m(y)\, m(x)\, dy\, dx}. \tag{11}$$

The value of this is not affected by the assumption that $\int_{\alpha}^{\beta} \int_{x}^{\beta} m(y)\, m(x)\, dy\, dx$ equals unity. Expanding the exponentials in the numerator in Taylor's series and omitting the second and higher powers of r, we have the numerator approximated to

$$-r \int_{\alpha}^{\beta} \int_{x}^{\beta} (y - x)^2 \, m(y) \, m(x) \, dy \, dx,$$

which equals $-2r\sigma^2$, where σ^2 is the variance of $m(x)$. Now expanding the exponential in Taylor's series we can write the denominator as

$$1 + \exp\left[\ln\left\{1 - r\int_{\alpha}^{\beta}\int_{x}^{\beta}(y + x)\, m(y)\, m(x)\, dy\, dx + \ldots\right\}\right].$$

Now expanding the logarithm and omitting the higher powers of r, we get

$$1 + \exp\left[-r \int_{\alpha}^{\beta}\int_{x}^{\beta}(y + x)\, m(y)\, m(x)\, dy\, dx\right].$$

Since the double integral equals 2μ, where μ is the mean of $m(x)$, the denominator of (11) is approximated by $1 + \exp(-2r\mu)$ and hence

$$CF \approx -2r\sigma^2/\{1 + \exp(-2r\mu)\}.$$

Therefore, we have the approximation

$$M(a) \approx a - 2r\sigma^2/\{1 + \exp(-2r\mu)\}. \tag{12}$$

The correction factor is negative for positive r, and zero for $r = 0$.

Hill and Trussell find a simple regression of the form $1(x) = A + B_5S_x$ to fit extremely well, where $_5S_x$ is the proportion of surviving siblings for women in the age group x and $x + 5$. They also observe that B is close to 1.0 and A is small and hence conclude that the "proportion of surviving siblings need be only trivially altered to yield an estimate of 1_N $(1(x))$." But our analytical work suggest that $_5S_x$ is very closely approximated to 1_z, where $z = (x + 2.5) - 2r\,\sigma^2/\{1 + \exp(-2r\mu)\}$. Approximate knowledge of r, μ and σ^2 will help to determine z closely.

Conclusion

The approximation to the mean age of the siblings of a random girl in a stable population developed here may be improved by including the higher order terms in the Taylor's expansion. When the rate of natural increase is zero the analysis shows clearly that the mean age of the siblings is equal to the age of the respondent. However, the difficulty in applying the technique mentioned by Hill and Trussell is recalled before concluding this note.

Acknowledgement

This study was supported by funds from Rockefeller Foundation (Contract No. RF 78058) awarded to the University of Pennsylvania. The author has benefited greatly from correspondence with Dr. Noreen Goldman.

References

1. Hill.K. and Trussell, J., 1977, Further developments in indirect mortality estimation, *Population Studies,* 31(2), 313-334.
2. Goodman, Leo A., Keyfitz, Nathan and Pullum, Thomas W., 1974, Family formation and the frequency of various kinship relationships, *Theoretical Population Biology,5* (1), 1-27.
3. Goldman, Noreen, 1978, Estimating the intrinsic rate of increase of a population from the average numbers of younger and older sisters, *Demography* 15(4), 499-508,