THE VARIANCE OF THE PRODUCT OF TWO INDEPENDENT VARIABLES AND ITS APPLICATION TO AN INVESTIGATION BASED ON SAMPLE DATA

By H. A. R. BARNETT, F.I.A., A.S.A.

If u and v are independent random variables, the variance of their product can be obtained as follows:

$$\sigma^2_{uv} = \mathscr{E}(u^2v^2) - \bar{u}^2\bar{v}^2 \quad \text{(see Johnson and Tetley,* 7·20)}$$
$$= \mathscr{E}(u^2)\,\mathscr{E}(v^2) - \bar{u}^2\bar{v}^2,$$

since u^2 and v^2 are independent (see Johnson and Tetley,* 7·7, Theorem B)

$$= (\bar{u}^2 + \sigma^2_u)(\bar{v}^2 + \sigma^2_v) - \bar{u}^2\bar{v}^2$$
$$= \sigma^2_u\sigma^2_v + \bar{v}^2\sigma^2_u + \bar{u}^2\sigma^2_v. \qquad (a)$$

Let u represent the population central exposed to risk at age x as found from observations of a sample of $1/k$ of the data, and let this observed u be taken as an approximation to $\bar{u}$.

Let the total of the sample data at all ages be N.

Let v represent the graduated value of m_x, taken as an approximation to $\bar{v}$.

Required to find the variance of the expected deaths, σ^2_{uv}.

Now $$\sigma^2_{u/k} \doteqdot N \times \frac{u/k}{N} \times \frac{N - u/k}{N} = \frac{u}{k}\left(1 - \frac{u/k}{N}\right),$$

whence $$\sigma^2_u = k^2\sigma^2_{u/k} \doteqdot ku - u^2/N$$

and $$\bar{v}^2\sigma^2_u \doteqdot kum^2_x - u^2m^2_x/N$$
$$= \text{expected deaths} \times (km_x - \text{expected deaths}/N), \qquad (b)$$
$$\sigma^2_v \doteqdot m_x(1 - m_x)/u \quad \text{and} \quad u \doteqdot \bar{u},$$

whence $$\bar{u}^2\sigma^2_v \doteqdot \text{expected deaths} \times (1 - m_x) \qquad (c)$$

and $$\sigma^2_u\sigma^2_v \doteqdot (ku - u^2/N)\, m_x(1 - m_x)/u$$
$$= \text{expected deaths} \times \{k(1 - m_x)/u - (1 - m_x)/N\}. \qquad (d)$$

Substituting expressions (b), (c) and (d) in (a) we obtain

Variance of expected deaths

$$\doteqdot \text{expected deaths} \times \{1 + (k-1)\,m_x - (1 - m_x + m_x u)/N + k(1 - m_x)/u\}. \qquad (e)$$

In the experiments based on the 1% sample of the 1951 census data, which were described in §X of *The Components of Mortality*,† it was found that the relative sizes of k (which was of course 100), N and the various values of u were such that the third and fourth terms inside the bracket in expression (e) were negligible. It was therefore possible to use the approximation

$$\text{Variance of expected deaths} = \text{expected deaths} \times (1 + 99m_x), \qquad (f)$$

which may be generally used in a 1% sample investigation so long as u and N are sufficiently large to justify ignoring the two terms in question.

* Johnson, N. L. and Tetley, H., *Statistics*, vol. 1. Cambridge University Press.
† *J.I.A.* **81**, 105.