JIA 79 (1953) 0213-0220

SOME PROBABILITY RESULTS FOR MORTALITY RATES BASED ON INSURANCE DATA

BY JOHN E. WALSH

of the U.S. Naval Ordnance Test Station

INTRODUCTION

IT is not usual in mortality investigations of insurance data to associate exactly one unit with each life. It is easier to use policies or sums assured or some variation of policies and sums assured. But a proper statistical test of graduated rates of mortality derived from such an investigation cannot be made unless the effect of the variable number of units associated with each life is known. Beard and Perks (1949, *J.I.A.* LXXV, 75) gave, for four different sampling processes, formulae for the variance of the distribution of deaths for a mortality study based on policies, and Daw (1951, *J.I.A.* LXXVII, 261) illustrated the formulae by numerical examples on various hypotheses. In particular, he showed that the sampling process employed had little effect on the numerical results, but that the frequency distribution of the number of policies held was very important. The formulae derived by Beard and Perks assumed that the frequency distribution of the universe of number of policies held was known. In a discussion on a paper on graduation, Perks (1951, *J.I.A.* LXXVII, 427) suggested that to deal with the duplicates problem Offices should be asked to write a card, not for each policy in force, but for each policy included in the experience on becoming a claim by death. The duplicates distribution could be obtained from the claims. If the name of the life assured, the date of birth, the date of death, and the class of assurance etc., were put on the card, it would be possible to bring together all claims in all Offices on the same life.

His intention was clearly to presume the distribution of the universe of duplicates from the distribution of the claims duplicates. Probably a process of graduation would be employed, itself needing to be tested. In the present note a statistic is put forward based directly on the rates of mortality to be tested and the distribution of claims duplicates; from this statistic significance tests and confidence intervals for the true mortality rate may be obtained. The lives of the investigation are not assumed to have the same mortality rate or to be observed during the same period. The resulting tests and confidence intervals are nearly 100% efficient.

If there are no duplicates the statistic of this note [formula (1) *seq.*] becomes $\frac{q'-q}{\sqrt{[q'(1-q)/N]}}$, where N is the number of persons (or policies since there are no duplicates) under observation, q' is the observed rate of mortality (number of deaths divided by N), and q is the true value of the rate of mortality (expected value of q'). This statistic differs, by the substitution in one place of q for q', from the statistic $\frac{q'-q}{\sqrt{[q'(1-q')/N]}}$, which I showed in my paper *On the large sample distribution of mortality rates based on statistically independent lives* (1950, *T.S.A.* II, 228) is nearly standard normal for large N. The proof assumed that each of the N persons of the investigation represented statistically independent observations and that the probability of death for the ith person was q_i. The rate of mortality q was given by $q=\sum_{i=1}^{N}\frac{q_i}{N}$, and it was assumed that all the q_i were small. In the discussion on the paper Aditya Prakash pointed out that the theorem also held if all the q_i were close to unity.

The slight variation which has been made to the statistic in the present note has also slightly modified the condition for which the main theorem of the note holds. Instead of the q_i having to be all small or all large it is assumed that none of the q_i differ from each other widely.

The sampling process envisaged is the repeated exposure of batches of lives who each time hold the same number of policies (differing between the lives in one sample) and are each time subject to the same chance of dying (again differing between the lives in one sample). This may seem somewhat restrictive but as Daw shows (*J.I.A.* LXXVII, 261) the sampling process employed is not very important. If all the q_i are equal $(=q)$ then the sampling process is similar to stratified sampling where the respective proportions of lives holding the same number of policies are fixed by the proportions that occurred in the sample being examined.

STATEMENT OF RESULTS

As a first step in the presentation of results, let us consider some notation:

E = number of units exposed to risk,

d = number of deaths,

u_i = number of units associated with ith death $(i=1, \ldots, d)$,

q' = observed mortality rate

$$=\sum_{i=1}^{d}\frac{u_i}{E},$$

q = 'true' mortality rate (expected value of q').

The principal result of this paper is that the probability distribution of the statistic

$$\frac{q'-q}{\sqrt{[(1-q)\sum_{1}^{d}u_i^2]/E}} \tag{1}$$

is very nearly normal with zero mean and unit standard deviation (i.e. very nearly standard normal) if the number of units exposed to risk is large enough.

An exact specification of a lower bound for the values of E which are sufficiently large for application of (1) is difficult to obtain. However, it is possible to state some approximate rules which are believed to be on the conservative side. Let

$\hat{u}$ = estimate of average number of units per life for the lives considered (obtained without knowledge of d and the u_i),

$\hat{q}$ = estimate of q (obtained without knowledge of d and the u_i),

where $\hat{u}$ tends to exceed the quantity it estimates while $\hat{q}$ tends to underestimate q. Rough estimates with these properties can often be obtained from previous mortality studies. The values of $\hat{u}$ and $\hat{q}$ are important only in an order of magnitude sense. An error of over 25% in both values could be tolerated. It is important, however, that $\hat{u}$ and $\hat{q}$ be specified without any knowledge of the value of d and the values of the u_i. Otherwise the probability distribution of (1) might be appreciably biased by the additional conditions imposed on the data.

The approximate rules concerning the accuracy of (1) are

Questionable validity if $E < 20\hat{u}/\hat{q}(1-\hat{q})$,
Rough accuracy if $20\hat{u}/\hat{q}(1-\hat{q}) \leqq E < 40\hat{u}/\hat{q}(1-\hat{q})$,
Moderate accuracy if $40\hat{u}/\hat{q}(1-\hat{q}) \leqq E < 100\hat{u}/\hat{q}(1-\hat{q})$,
Good accuracy if $100\hat{u}/\hat{q}(1-\hat{q}) \leqq E < 200\hat{u}/\hat{q}(1-\hat{q})$,
Excellent accuracy if $E \geqq 200\hat{u}/\hat{q}(1-\hat{q})$.

It is believed that the accuracy is usually at least as good as these rules indicate. The rules are of a heuristic nature, and no attempt to furnish a rigorous proof of their validity will be made. However, an outline of the theoretical considerations which motivated their selection is presented in the Derivations section.

For the ordinary type of mortality investigation, data are obtained for several periods (e.g. age groupings). If the value of E for such a subdivision is less than $20\hat{u}/\hat{q}(1-\hat{q})$ for that subdivision, it is usually possible to combine the data for this period with those for one or more adjoining periods so that E for the combined periods exceeds $20\hat{u}/\hat{q}(1-\hat{q})$ for the combined periods. This should be done without any knowledge of the values of d and the u_i. The procedure of combining adjoining periods leaves $\hat{u}$ about the same, increases $\hat{q}$, and increases E.

Confidence intervals for q can be derived on the basis of (1) and its properties. Significance tests can be obtained from these confidence intervals. Let us consider the situation for one-sided and symmetrical confidence intervals and significance tests. In the derivations the quantity K_ϵ is used to denote the deviate of the standard normal distribution which is exceeded with probability ϵ.

If the probability distribution of (1) is standard normal, the following relations hold:

$$\Pr\left\{\frac{q'-q}{\sqrt{[(1-q)\Sigma u_i^2]/E}} \geqq K_\epsilon\right\} = \Pr\left\{\frac{q'-q}{\sqrt{[(1-q)\Sigma u_i^2]/E}} \leqq -K_\epsilon\right\} = \epsilon.$$

Use of monotonicity properties, combined with the solution for q of the quadratic equation defined by

$$\left\{\frac{q'-q}{\sqrt{[(1-q)\Sigma u_i^2]/E}}\right\}^2 = K_\epsilon^2$$

and selection of the appropriate root, shows that

$$\left.\begin{aligned} &\Pr\left\{\frac{q'-q}{\sqrt{[(1-q)\Sigma u_i^2]/E}} \geqq K_\epsilon\right\} \\ &\qquad = \Pr\{q' - s_\epsilon\sqrt{[1+s_\epsilon^2/4(1-q')^2]} - s_\epsilon^2/2(1-q') \geqq q\}, \\ &\Pr\left\{\frac{q'-q}{\sqrt{[(1-q)\Sigma u_i^2]/E}} \leqq -K_\epsilon\right\} \\ &\qquad = \Pr\{q' + s_\epsilon\sqrt{[1+s_\epsilon^2/4(1-q')^2]} - s_\epsilon^2/2(1-q') \leqq q\}. \end{aligned}\right\} \quad (2)$$

Here the quantity s_ϵ is defined by

$$s_\epsilon = K_\epsilon\sqrt{[(1-q')\Sigma u_i^2]/E}.$$

The probability relations (2) define one-sided confidence intervals for q (with confidence coefficient equal to ϵ).

Two-sided confidence intervals for q can be obtained as the complement of combinations of two non-overlapping one-sided confidence intervals. As an example, if $\epsilon < \frac{1}{2}$,

$$\{q' - s_\epsilon \sqrt{[1 + s_\epsilon^2/4(1-q')^2]} - s_\epsilon^2/2(1-q'), \quad q' + s_\epsilon \sqrt{[1 + s_\epsilon^2/4(1-q')^2]} - s_\epsilon^2/2(1-q)\}$$

is a symmetrical confidence interval for q with confidence coefficient equal to $1-2\epsilon$.

If, as is often the case, the value of s_ϵ is so small that terms involving s_ϵ^2 can be neglected, the forms of the confidence intervals are greatly simplified. For example, the symmetrical confidence interval with coefficient $1-2\epsilon$ ($\epsilon < \frac{1}{2}$) becomes

$$(q' - s_\epsilon, \quad q' + s_\epsilon).$$

An analysis which shows that these tests and confidence intervals are nearly 100% efficient is outlined in the Derivations section.

Experience indicates that the standard normal tends to furnish an accurate representation of the distribution of quantities of the type (1) near the mean value; also that this accuracy decreases as the deviation from the mean increases. This suggests that values of ϵ which are near 0 or 1 be used only when the accuracy of the standard normal as an approximation to the distribution of (1) is good. For example, use of a value of ϵ which deviates from 0 or 1 by as little as ·005 would probably not be warranted unless $E \geqq 100\hat{u}/\hat{q}(1-\hat{q})$.

This paper shows that asymptotically ($E \to \infty$) the probability distribution of (1) is normal with zero mean and standard deviation very nearly equal to unity. Some restrictions concerning the probabilities of death and the number of units associated with a life are used in the proof. However, these restrictions do not appear to have much practical significance. This analysis is presented in the Derivations section.

DERIVATIONS

Asymptotic probability distribution of (1). First let us consider some additional notation. Let

N = number of lives,

U_j = number of units for jth life ($j = 1, \ldots, N$),

Z_j = random variable which assumes the value 1 if the jth life dies during its period of observation and the value 0 otherwise ($j = 1, \ldots, N$),

q_j = probability that jth life dies during its period of observation ($j = 1, \ldots, N$).

Then

$$E = \sum_{j=1}^{N} U_j, \quad q' = \sum_{j=1}^{N} \frac{U_j Z_j}{E}, \quad q = \sum_{j=1}^{N} \frac{U_j q_j}{E}, \quad \sum_{i=1}^{d} u_i^2 = \sum_{j=1}^{N} U_j^2 Z_j.$$

Thus the statistic (1) can be written in the form

$$\left(\sum_1^N U_j Z_j - \sum_1^N U_j q_j\right) \Big/ \sqrt{\left\{(1-q)\sum_1^N U_j^2 Z_j\right\}}. \tag{3}$$

In the derivations it is assumed that the Z_j are statistically independent, that the q_j are such that $q_j(1-q_j)$ is bounded away from 0, and that

$$\lim_{N\to\infty}\sum_{j=1}^{N} U_j^2/(\max_j U_j^2)=\infty. \tag{4}$$

These restrictions do not seem to be very important from a practical point of view. They should be approximated for any reasonable type of mortality investigation and in particular relation (4) is obviously true if U_j and therefore $\max_j U_j^2$ is bounded. The U_j, q_j, N, and E are fixed quantities which do not have probability distributions. It may be noted that $N\to\infty$ as $E\to\infty$, for if on the contrary N were bounded by K, $E\leq\sum_{j=1}^{K} U_j$ and would also be bounded.

The quantity Z_j represents a sample value from a binomial population for which

$$\text{probability of value } 1=q_j,$$

$$\text{probability of value } 0=1-q_j \quad (j=1,\ldots,N),$$

Consequently, for $j=1,\ldots,N$,

$$\begin{aligned}&\text{expected value of } U_jZ_j=U_jq_j,\\ &\text{variance of } U_jZ_j=U_j^2q_j(1-q_j),\\ &\text{expected value of } |U_jZ_j-U_jq_j|^3=U_j^3q_j(1-q_j)[q_j^2+(1-q_j)^2].\end{aligned} \tag{5}$$

The statistic (3) can also be expressed in the form

$$\frac{\Sigma U_jZ_j-\Sigma U_jq_j}{\sqrt{[\Sigma U_j^2q_j(1-q_j)]}}\Big/\sqrt{\left[\frac{(1-q)\Sigma U_j^2Z_j}{\Sigma U_j^2q_j(1-q_j)}\right]}.$$

The method of proof consists in showing that the asymptotic probability distribution of

$$\frac{\Sigma U_jZ_j-\Sigma U_jq_j}{\sqrt{[\Sigma U_j{}^2q_j(1-q_j)]}} \tag{6}$$

is standard normal, while the quantity

$$\sqrt{\left[\frac{(1-q)\Sigma U_j^2Z_j}{\Sigma U_j^2q_j(1-q_j)}\right]} \tag{7}$$

converges in probability to a value which is very near unity. Then, on the basis of Cramér's Convergence Theorem (see reference (1), p. 254), the asymptotic probability distribution of (1) is normal with zero mean and standard deviation very near unity.

First let us show that the asymptotic distribution of (6) is standard normal. This is done by using the Liapounoff version of the Central Limit Theorem (see reference (1), pp. 216–17) combined with the restrictions on the U_j, q_j, N, E. Let $b>0$ be a fixed lower bound for the value of $q_j(1-q_j)$. Then from (5) and the statistical independence of the Z_j,

$$\sigma^2=\text{variance of }\sum_1^N U_jZ_j=\sum_1^N U_j^2q_j(1-q_j),$$

$$\rho^3=\text{expected value of }\sum_1^N|U_jZ_j-U_jq_j|^3=\sum_1^N U_j^3q_j(1-q_j)[q_j^2+(1-q_j)^2].$$

Thus
$$\frac{\rho}{\sigma} \leqq \frac{[\Sigma U_j^3 q_j(1-q_j)]^{\frac{1}{3}}}{[\Sigma U_j^2 q_j(1-q_j)]^{\frac{1}{2}}}$$
$$\leqq \frac{(\frac{1}{4})^{\frac{1}{3}}[\Sigma U_j^3]^{\frac{1}{3}}}{b^{\frac{1}{6}}[\Sigma U_j^2]^{\frac{1}{2}}}$$
$$= \frac{(\frac{1}{4})^{\frac{1}{3}}[\Sigma U_j^3/(\max U_j^3)]^{\frac{1}{3}}}{b^{\frac{1}{6}}[\Sigma U_j^2/(\max U_j^2)]^{\frac{1}{2}}}$$
$$\leqq \frac{(\frac{1}{4})^{\frac{1}{3}}}{b^{\frac{1}{6}}[\Sigma U_j^2/(\max U_j^2)]^{\frac{1}{6}}}.$$

On the basis of (4), $\lim_{N\to\infty}(\rho/\sigma)=0$. Consequently the Central Limit Theorem is applicable and the asymptotic distribution of (6) is standard normal.

Finally, let us show that (7) converges in probability to a value which is near unity. It is sufficient to show that this is the case for the quantity
$$\frac{(1-q)\Sigma U_j^2 Z_j}{\Sigma U_j^2 q_j(1-q_j)}. \tag{8}$$
Taking the square root of (8) only tends to make the value nearer unity but has no effect on the convergence in probability. The expected value of (8) is
$$\frac{(1-q)\Sigma U_j^2 q_j}{\Sigma U_j^2 q_j(1-q_j)} = 1 + \frac{\Sigma U_j^2 q_j(q_j-q)}{\Sigma U_j^2 q_j(1-q_j)}. \tag{9}$$
For the usual type of mortality study, the value of $(q_j-q)/(1-q_j)$ should be very small for almost all cases. Consequently the expected value of (8) should be very near unity. The variance of (8) has the value
$$\frac{(1-q)^2\Sigma U_j^4 q_j(1-q_j)}{[\Sigma U_j^2 q_j(1-q_j)]^2}$$
$$\leqq \frac{\frac{1}{4}(1-q)^2}{b^2\Sigma U_j^2/(\max U_j^2)}.$$
Thus, on the basis of (4), the variance of (8) has the limiting value 0 as $N\to\infty$. Use of Tchebycheff's Theorem (see reference (1), p. 253) shows that (8) converges in probability to its expected value (9) as $N\to\infty$.

Efficiency analysis. Let us consider the efficiency of tests and confidence intervals based on (1). If the value of the standard deviation of q' divided by $\sqrt{(1-q)}$ were known (denote this quantity by R), the tests and confidence intervals for q based on the statistic
$$(q'-q)/\sqrt{(1-q)}R, \tag{10}$$
would be at least as efficient as those obtainable without this additional knowledge. The quantity (10) has a probability distribution which is very nearly standard normal for the situation considered. However, the distribution of (1) is also nearly standard normal and (1) differs from (10) only in that $\Sigma u_{ij}^2/E^2$ replaces R^2. This implies that the probability distribution of the endpoint of a one-sided confidence interval based on (1) is approximately the same as the distribution of the endpoint of the corresponding one-sided confidence interval based on (10). A two-sided confidence interval is the complement of a combination of non-overlapping one-sided confidence intervals. Thus the prob-

ability distribution of the endpoints of a confidence interval based on (1) is nearly the same as the probability distribution of the endpoints of the corresponding confidence interval based on (10). This implies that confidence intervals obtained from (1) have approximately the same efficiency as those obtained from (10). Consequently significance tests based on these confidence intervals have about the same efficiency for the two procedures. However, the results based on (10) are at least as efficient as the best results obtainable on the basis of the data alone. Thus the efficiency of tests and confidence intervals based on (1) is nearly 100%.

Motivation for rules. In the Statement of Results section, some approximate rules for deciding when E is sufficiently large for (1) to be applicable were presented. It was conjectured that the accuracy is at least as good as the rules indicate. As a partial substantiation of this conjecture, the heuristic reasoning which led to the approximate rules is outlined here.

The basis is Bernstein's Theorem for the binomial distribution (see reference (2)). This result states that if n is the sample size and p the probability of 'success', then the total number of 'successes' has a probability distribution which is very nearly normal if $n \geqq 62{\cdot}5/p\,(1-p)$. If each life of the subdivision of data considered had the same probability of death and the same number of units, the corresponding equation using the notation of this paper would be $N \geqq 62{\cdot}5/q\,(1-q)$. The accuracy required in this paper is nowhere near that sought by Bernstein. Consequently, for a straight binomial situation it is likely that the factor 62·5 could be reduced to the factor of approximately 5 used in reference (3). On the other hand, the 'standard deviation' of q' used in (1) is not fixed but has a probability distribution itself. Also the probabilities of death and number of units vary with each life while the values of N and q are estimated from the relations $N = E/\hat{u}$, and $q = \hat{q}$, which may be noticeably in error (although usually on the conservative side). To allow for the inaccuracies and the additional variation, the minimum value of the factor was increased from 5 to 20 while the factor for excellent accuracy was increased from 62·5 to 200. The intermediate factors represent interpolation for intermediate accuracies.

APPENDIX

In the analysis several well-known but non-elementary theorems of mathematical statistics were used. For convenience of reference, these theorems are stated in this section.

Cramér's Convergence Theorem. Let $\xi_1, \xi_2, \ldots$ be a sequence of random variables whose cumulative distribution functions converge to a cumulative distribution function $F(x)$. Let $\eta_1, \eta_2, \ldots$ be another sequence of random variables which converge in probability to a value $\beta > 0$. Then the cumulative distribution functions for the sequence $\xi_1/\eta_1, \xi_2/\eta_2, \ldots$ converge to the cumulative distribution function $F(\beta x)$.

The material for this theorem is given by Cramér (reference (1) p. 254).

Central Limit Theorem (Liapounoff). Let $\xi_1, \xi_2, \ldots, \xi_k$ be k independent random variables and denote by m_ν and σ_ν^2 the expected value and variance of ξ_ν. Let

$$\rho_\nu^3 = \text{expected value of } |\xi_\nu - m_\nu|^3$$

be finite for all v. Also let

$$\sigma^2 = \sigma_1^2 + \ldots + \sigma_k^2,$$

$$\rho^3 = \rho_1^3 + \ldots + \rho_k^3.$$

If
$$\lim_{k\to\infty} (\rho/\sigma) = 0,$$

then the asymptotic ($k \to \infty$) distribution of

$$\left(\sum_1^k \xi_\nu - \sum_1^k m_\nu\right) \Big/ \sqrt{\left(\sum_1^k \sigma_\nu^2\right)}$$

is standard normal. See Cramér (reference (1), pp. 216–17).

Tchebycheff's Theorem. Let $\xi_1, \xi_2, \ldots$ be random variables and let m_ν and σ_ν^2 denote the expected value and variance of ξ_ν. If $\sigma_\nu^2 \to 0$ as $\nu \to \infty$, then $\xi_\nu - m_\nu$ converges in probability to zero. See Cramér (reference (1), p. 253).

REFERENCES

(1) CRAMÉR, HARALD (1946). *Mathematical Methods of Statistics*. Princeton University Press.

(2) FELLER, W. (1945). On the normal approximation to the binomial distribution. *Ann. Math. Statist.* XVI, 321.

(3) HOEL, P. G. (1947). *Introduction to Mathematical Statistics*, p. 45. John Wiley and Sons.