JIA 74 (1948) 0094-0112

THE ANALYSIS OF HETEROGENEOUS MORTALITY DATA

BY LEON SOLOMON, M.A., PH.D., F.F.A.
of the Statistical Branch, Admiralty

I. INTRODUCTION

IN this paper, everyday actuarial data have been examined by means of standard statistical techniques. The data in question have been drawn from the publications of the Continuous Mortality Investigation Committee, and the methods of statistical analysis have been applied to investigate the nature and extent of the heterogeneity which may exist in those data.

It will be recalled that the returns of the Continuous Mortality Investigation Committee contain the exposures to risk and corresponding deaths subdivided into various groups. Twelve such groups have been distinguished in this paper, viz.:

Class of assurance: Whole-life with profits (abbreviated to L.W. hereafter),
Whole-life without profits (L.N.),
Endowment assurance with profits (E.W.),
Endowment assurance without profits (E.N.).

Period of experience: 1924–28 inclusive,
1929–33 inclusive,
1934–38 inclusive.

The following section of the total experience of these groups has been analysed:

Ages: 46–55 years (nearest birthday) inclusive.
Type of lives: Medically examined.
Duration of policies: 5 years and over.

Two types of heterogeneity in particular are of interest and, possibly, of practical importance. The first is concerned with the equality or non-equality of the mortality rates at a given age for the separate types of policy and quinquennia. The investigation of such a problem is, of course, a standard actuarial task, and the common methods of analysis are based on the theory of the binomial probability distribution. According to this theory, the probability that exactly θ lives will die in a year out of E lives exposed to risk is $\binom{E}{\theta} q^{\theta}(1-q)^{E-\theta}$, where q denotes the true rate of mortality. The particular result is easily deduced that the number of deaths in a year, considered as a random variable, has a variance (or square of the standard deviation) of $Eq(1-q)$; and on this result, essentially, is erected the series of tests of uniformity of mortality statistics. The vital point in the practical utility of the theory is, of course, that E should refer to a number of *lives*. The data now under examination are, on the contrary, based on policies. Because of the holding of multiple policies by some individuals, the break-down of the simple 'binomial' theory is possible (cf. Seal [1940], Daw [1945]).* An exact treatment of the problem of 'duplicate' policies has been carried out by H. L. Seal [1947], but for the purpose of this paper a simpler approach has been adopted. We shall say that the variance of

* A list of references appears at the end of this note.

the number of deaths (i.e., strictly, of the number of policies becoming claims by death) out of E policy years' exposure to risk is $kEq(1-q)$, where k is independent of the mortality rate q but depends, *inter alia*, on the proportions of persons who hold one, two, three, ..., policies. $k=1$ corresponds, of course, to the straightforward binomial distribution, and if on investigation we find that k is sensibly equal to unity, we shall describe the data as 'binomially homogeneous'.

It might be argued that, were k to differ appreciably from 1 in present-day life office data, it would be of some practical importance to realize the fact. The comparison of one mortality experience with another or with a standard table, the fluctuation loading which should be imposed upon a net premium, and any necessary calculations associated with the theory of risk—all might be affected if the basic mortality data departed materially from binomial homogeneity.

In the light of the foregoing, the objectives of the paper will now be more precisely defined. They are:

(*a*) *An investigation of heterogeneity* of the above-described data. The term heterogeneity will be reserved to denote any departure of the variance of an estimated mortality rate from its theoretical binomial value. The aim is thus to test whether k is significantly different from unity; and, if it is, to estimate its value.

(*b*) *An investigation of uniformity*. The term non-uniformity will be used to describe real differences in the mortality experiences of the several class and period sub-groups. This study must, of course, take account of the conclusions of investigation (*a*).

In principle, both investigations are simple. With full access to the records of an office or offices, one would examine (*b*) by taking the individual life, rather than the individual policy, as the unit of enumeration, and the simple binomial theory would apply. With regard to (*a*), one could allow for the effect of duplicate policies—probably the major source of any heterogeneity which exists—by ascertaining the distribution of lives holding one, two, three, ..., policies; or one could estimate k directly by splitting the data by some random process into groups which could be accepted as approximately uniform and equally heterogeneous. The published returns, however, permit no such direct approaches. The distribution of policies with respect to lives is not generally available, and the subdivisions according to class of policy and period are almost certainly non-uniform. Since this investigation is based purely on published statistics, the indirect and somewhat complex methods of analysis described in the sequel have had to be employed. The conclusions may possess at least the following point of interest: they indicate the type of intelligence concerning an aspect of British life office operations which could be evaluated by any sufficiently curious member of the public.

II. THE NATURE OF THE STATISTICAL ARGUMENTS

The two investigations carried through below employ standard statistical methods, the theory and field of application of which are fully discussed in the text-books. Recapitulation of the details would be out of place. However, since the processes may not yet be among the commonplace tools of the actuarial kit-bag, a brief summary of the nature of the underlying arguments may be helpful. Essentially, a hypothesis is postulated, and a test made to judge whether the observed data are compatible therewith. The initial choice of hypothesis is often made on non-statistical grounds. It may, for instance, be dictated by some *a priori* theory, or by the form in which a problem is presented to the statistician. In the absence of such guidance, however, it is customary to adopt first the simplest reasonable hypothesis, and if the data necessitate its rejection, to replace it by another which retains as large a degree of simplicity as possible. For example, we shall postulate below that all the mortality data are binomially homogeneous, and when this hypothesis is found—as we shall find it—untenable, we shall test next whether the existence of a uniform degree of heterogeneity throughout the data is compatible with the facts.

Tests of compatibility are generally made on a probability basis. The principle is to accept as true the hypothesis under test, and to calculate the chance that the observed results, or results more unfavourable to the hypothesis, would have occurred merely through random effects. If this probability is less than some agreed figure, such as 5 % or 1 %, the hypothesis is rejected as untrue. If it equals or exceeds the chosen level, the hypothesis is accepted. Suppose, for instance, that 5 % were adopted as the 'significance level', that the data comprised 20 deaths out of 1000 lives exposed to risk for a year, and that we wished to test whether $q = \cdot 01$ could be regarded as the true mortality rate. Since the expected number of deaths is 10, the procedure is to calculate the probability that 20 *or more* deaths would occur in a year out of 1000 lives who experienced a mortality rate of ·01. It is found that the probability in question is less than 5 %. Accordingly we reject the hypothesis. Clearly, if many such hypotheses were thus tested, and if it happened that all of them were true, 95 % (in the long run) would be accepted; but in the remaining 5 % of cases such true hypotheses would be rejected, and the decision of the test would be wrong. It is the consequences of an error of this nature which largely dictate the choice of significance level. If they are serious, a level of 1 %, say, might be preferred to 5 %—at the expense, of course, of a loss in discriminating power, since in the former case more false hypotheses will perforce be accepted as true. Throughout this paper, 5 % has been adopted as the significance level.

III. THE INVESTIGATION OF HETEROGENEITY

Let θ denote the number of deaths arising from E years of life exposed to risk, and let q denote the mortality rate to which the lives are subject. The value of θ is not, of course, completely fixed by a knowledge of E and q, but involves also a random element. The latter, it is well known, can be described by the statement that the probability distribution of θ is approximately normal, provided the expected number of deaths Eq is reasonably large; $Eq = 10$ is an adequate lower limit. The mean of this normal distribution is, of course, Eq, its variance $Eq(1-q)$. To describe the suspected heterogeneity in the actual data, in which E is based on policies, we generalize by postulating that θ is normally distributed with mean Eq and variance $kEq(1-q)$. (It may be remarked that Seal's researches [1947] indicate that the assumption of a normal distribution is valid for large experiences, though the approach to normality is much slower when there is a high proportion of duplicates than when binomially homogeneous data are in question. Furthermore, by the operations described below, the random element $\theta - Eq$ in the recorded numbers of deaths can be approximately estimated for various ages, and classes and periods of assurance; and the usual study of the higher moments of these random variates gives no cause to suspect this postulate of normality.) The problem of estimating the value of k, and of testing whether it differs significantly from unity, is thus associated with the estimation of the variance of a normal distribution.

The form of the available data does not permit a direct estimate of the required variance. For instance, the twelve class-period sub-groups at a given age are almost certainly non-uniform. Accordingly they do not provide a sample of observations from a common normal population. We are therefore led to consider separately each sub-group which comprises the data for ages 46–55 inclusive. Since the mortality rate certainly varies with age over this range, a direct calculation of the sample variance in a sub-group would again be illegitimate. We proceed to describe a method of regression analysis which overcomes such difficulties.

Consider, first, a series of observations $y'_1, y'_2, \ldots, y'_n$ dependent on exactly known quantities $x_1, x_2, \ldots, x_n$ respectively. Suppose it is known that, corresponding to a particular value x, the true value of y is of the form

$$y = a_0 + a_1 x + \ldots + a_p x^p,$$

but that any observed value y' of the latter quantity is subject to random fluctuations. More precisely we shall consider y' as normally distributed about its mean $a_0 + a_1 x + \ldots + a_p x^p$ with variance σ^2, where σ^2 is independent of x and where the values of $a_0, a_1, \ldots, a_p$ are not known to the investigator. Under these conditions, the text-books prove that the best estimates of the 'regression constants' $a_0, a_1, \ldots, a_p$ are given by the method of least squares, i.e. they are those values, $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p$ say, which minimize

$$\sum_{i=1}^{n} (y'_i - a_0 - a_1 x_i - \ldots - a_p x_i^p)^2.$$

Further, it is shown that if σ^2 is unknown the best estimate thereof is given by

$$\hat{\sigma}^2 = \frac{1}{n-p-1} \sum_{i=1}^{n} (y'_i - \hat{a}_0 - \hat{a}_1 x_i - \ldots - \hat{a}_p x_i^p)^2 = \frac{S_{p+1}}{n-p-1}, \quad \text{say.}$$

S_{p+1} denotes the sum of squares of the deviations of the observed y'_i from the corresponding estimated means $y_i = \hat{a}_0 + \hat{a}_1 x_i + \ldots + \hat{a}_p x_i^p$. The denominator of $\hat{\sigma}^2$ represents the 'degrees of freedom' on which the estimate is based, and is equal to the number of observations n less the number of regression coefficients $(p+1)$ derived from the data.

It may happen that the value of $\hat{a}_p$, say, is small, and that the question presents itself of assessing whether it is significantly different from zero. The accepted method of examination is as follows. Postulate that $a_p = 0$, i.e. that the true regression equation is $y = a_0 + a_1 x + \ldots + a_{p-1} x^{p-1}$. By the foregoing procedure, calculate S_p, the sum of squares of the deviations of the observed y'_i from the corresponding estimated means $\hat{b}_0 + \hat{b}_1 x + \ldots + \hat{b}_{p-1} x^{p-1}$. (Note that S_p is always greater than S_{p+1}, and that the estimates $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_{p-1}$ will differ in general from the $\hat{a}_0, \hat{a}_1, \ldots,$ of the earlier regression formula.) If the hypothesis $a_p = 0$ is true, $S_p/(n-p)$ will serve as an estimate of σ^2, and it can be shown, moreover, that

$$\sigma'^2 = \frac{S_p - S_{p+1}}{n-p-(n-p-1)} = \frac{S_p - S_{p+1}}{1}$$

will be another valid estimate possessing 1 degree of freedom. The latter, indeed, is the more useful, since it can be shown to be *statistically independent* of the first estimate $\hat{\sigma}^2 = S_{p+1}/(n-p-1)$. Under the condition of statistical independence, the well-known 'variance ratio test' (cf., for example, Wishart [1947]) can be applied to show whether the estimates $\hat{\sigma}^2$ and σ'^2 are compatible in the probability sense of section II. In order to use the published tables (e.g. Fisher and Yates [1942]) of the distribution of the variance ratio, we form

$$F = \frac{\text{greater estimate of } \sigma^2}{\text{lesser estimate of } \sigma^2},$$

and associate with numerator and denominator the respective number of degrees of freedom. If F is found to be equal to or less than the 5 % value given in the tables for the appropriate pair of degrees of freedom, we accept the hypothesis that $a_p = 0$. If F exceeds the critical 5 % value, we reject it. In the former event, the subsequent problem might arise of testing whether a_{p-1} (or any other constant or group of constants) could be regarded as zero. The procedure would be identical.

Application to mortality data

The foregoing general theory will be applied now to each of the twelve class-period groups under investigation. We first form a function of the deaths and exposures, and postulate that, over the limited range of ages in question, the dependence of the mean (or 'true value') of this function on age can be represented by a polynomial of high degree. An estimate of the variance of the observed values about the calculated regression means, and hence an estimate of k, follow. Next, the coefficient of the highest power in the regression formula is hypothetically equated to zero, and the above-described test invoked. If the hypothesis is found acceptable, the effect of omitting the next highest term is similarly investigated. In this way, the simplest representation will be reached of the dependence of the chosen actuarial function on age, and this representation will lead to an estimate of k. The process is exemplified below.

The particular actuarial function must satisfy the condition of the preceding paragraph, namely that the variance must be constant for all values of x, the age. The estimated mortality rate at age x, say

$$q'_x = \theta_x / E_x \quad (x = 46, 47, \ldots, 55),$$

has a variance

$$V(q'_x) = k q_x (1 - q_x)/E_x,$$

where q_x denotes the true value of the mortality rate.

Now over the ages in question, q_x varies from the order of ·005 to the order of ·01, i.e. by a factor of 2; and $1 - q_x$ is sensibly equal to unity. Hence $E_x V(q'_x)$ varies by about twofold between the youngest and the oldest ages, and the stipulated condition is not satisfied. A more satisfactory function is the square root of the estimated mortality rate

$$y'_x = \sqrt{(\theta_x / E_x)},$$

for its variance is approximately

$$V(y'_x) = V(\sqrt{q'_x}) \doteqdot V(q'_x)\left(\frac{\partial}{\partial q_x}\sqrt{q_x}\right)^2$$

$$= \frac{k q_x (1 - q_x)}{E_x} \frac{1}{4 q_x} \doteqdot \frac{k}{4 E_x},$$

since $1 - q_x$ is practically unity. The variance is thus virtually independent of the true rate of mortality, and y'_x has a weight proportional to E_x, i.e. $\sqrt{E_x}.y'_x$ (or $\sqrt{\theta_x}$) can be taken as possessing a constant variance at all ages from 46 to 55. (It may be interpolated that the question of the dependence of k on age will be discussed later. At present, the problem does not arise, since our first objective will be to ascertain whether k can be universally equated to unity.)

Each class-period group in turn has been analysed thus: a weight of E_x/E_{46} has been associated with y'_x, and a polynomial regression on age of degree three has been postulated of the form

$$y_x = a_0 + a_1 x + a_2 x^2 + a_3 x^3.$$

As the constants $a_i (i = 0, 1, 2, 3)$ are unknown, estimates $\hat{a}_i (i = 0, 1, 2, 3)$ of their respective values can be made by the 'least squares' criterion. The weighted sum of squares of the errors of estimate is given by

$$S_4 = \sum_{x=46}^{55} \frac{E_x}{E_{46}} (y'_x - \hat{a}_0 - \hat{a}_1 x - \hat{a}_2 x^2 - \hat{a}_3 x^3)^2.$$

This sum, when divided by $10 - 4 = 6$, the number of degrees of freedom, gives an estimate of the common variance $\frac{1}{4}k/E_{46}$ of $\sqrt{(E_x/E_{46})}\, y'_x$. The quadratic and linear regressions of y'_x on x have been similarly studied for each of the twelve class-period groups, and the corresponding estimates of k abstracted. The

Table 1

Source of estimate of variance	Weighted sum of squares of errors of estimate $\times 4E_{46}$	Degrees of freedom	Mean square = estimate of k	Weighted sum of squares of errors of estimate $\times 4E_{46}$	Degrees of freedom	Mean square = estimate of k
	L.W. 1924–28			L.N. 1924–28		
Linear regression	9·45	8	1·18	13·96	8	1·75
Quadratic regression	9·29	7	1·33	9·70	7	1·39
Cubic regression	9·25	6	1·54	9·61	6	1·60
	L.W. 1929–33			L.N. 1929–33		
Linear regression	16·45	8	2·06	15·54	8	1·94
Quadratic regression	15·94	7	2·28	11·62	7	1·66
Cubic regression	14·70	6	2·45	10·99	6	1·83
	L.W. 1934–38			L.N. 1934–38		
Linear regression	19·64	8	2·45	10·63	8	1·33
Quadratic regression	15·82	7	2·26	7·57	7	1·08
Cubic regression	12·71	6	2·12	7·35	6	1·23
	E.W. 1924–28			E.N. 1924–28		
Linear regression	8·07	8	1·01	13·75	8	1·72
Quadratic regression	7·72	7	1·10	13·37	7	1·91
Cubic regression	6·56	6	1·09	13·35	6	2·22
	E.W. 1929–33			E.N. 1929–33		
Linear regression	21·05	8	2·63	8·36	8	1·05
Quadratic regression	20·55	7	2·94	7·22	7	1·03
Cubic regression	13·23	6	2·21	5·36	6	·89
	E.W. 1934–38			E.N. 1934–38		
Linear regression	10·77	8	1·35	4·42	8	·55
Quadratic regression	5·28	7	·75	3·24	7	·46
Cubic regression	5·12	6	·85	2·96	6	·49

results of the investigation are summarized in Table 1. The method of computation is set out in the Appendix.

Before discussing the figures in Table 1, a few points may be made. First, it might reasonably be argued that higher polynomials than those of the third degree should have been considered. They have been, and it was rapidly obvious that they would in no case materially reduce the weighted sum of squares of errors of estimate. The detailed results are accordingly of little interest. Secondly, it might be urged that non-polynomial regression curves should also be envisaged. There is no evidence, however, that such functions would give a materially better representation of the variation of y'_x with age, *over the short range* of ages in question, than would a polynomial regression. Thirdly, another interpretation of the analysis may be indicated. Regard the linear, quadratic, cubic, ..., regressions as a series of graduations of increasing complexity. From each graduation in turn estimate the residual variance, and hence k. The process stops when a more complicated graduation fails to reduce significantly the estimate of k. It will be recalled that, in the discussion on Daw's paper [1945], serious objections were levied against the assumption that one particular graduation of crude data, or one standard mortality table, could provide 'true' rates of mortality from which k could be estimated. The performance of the *series* of graduations overcomes such difficulties.

Analysis of the variances

Consider the group L.W. 1924–28. The weighted sum of squares, postulating a cubic regression, is 9·25 (Table 1) with 6 degrees of freedom. Assuming a quadratic regression, the sum is 9·29 with 7 degrees of freedom. The cubic term itself therefore contributes $9{\cdot}29 - 9{\cdot}25 = {\cdot}04$ to the sum of squares, and 1 degree of freedom attaches to this estimate. Hence the hypothetical non-reality of the cubic term may be tested by the ratio

$$F = \frac{\text{greater mean square}}{\text{lesser mean square}} = \frac{1{\cdot}54}{{\cdot}04} = 39,$$

with 6 and 1 degrees of freedom for numerator and denominator respectively. For this pair of degrees of freedom, the published tables give $F = 234$ as the critical 5 % value for the variance ratio test. Since this exceeds the observed ratio, we accept the hypothesis that the cubic term is zero.

Similarly, it is found that the quadratic term is also non-significant. Therefore, in the L.W. 1924–28 experience, y'_x can be accepted as varying linearly with x ($x = 46, 47, \ldots, 55$), and the corresponding estimate of k is 1·18 (Table 1), i.e. the variance of a mortality rate estimated from this sample of data is 1·18 times greater than the homogeneous binomial value. (In this example of the L.W. 1924–28 group, it is immediately evident from Table 1 that the use of quadratic or cubic regressions is unnecessary, since the estimate of k is thereby *increased*. For the present purpose, therefore, the formal test could have been dispensed with. The L.W. 1934–38 group provides a more interesting example of the use of the test.)

Testing in like manner the other eleven groups detailed in Table 1, it appears that a linear regression is in every case adequate. No significant reduction of the estimate of variance ensues by postulating a quadratic, cubic, or higher regression. It is, of course, possible that with the limited number of degrees of freedom in a single group, we may have failed to establish statistically an essential feature of the experience. Consider therefore the twelve groups *as* *a whole* by adding the sums of squares in Table 1 from the individual linear, quadratic, and cubic regressions respectively. The degrees of freedom attaching to these grand totals are the sums of those of the separate components, and the pooling procedure is justified, according to a well-known theorem, by the statistical independence of the sums of squares in the several groups. The figures thus obtained permit us to test whether the cumulative evidence supports the assertion at the beginning of this paragraph. Tables 2A and 2B give the necessary details.

Table 2A

Source of estimate of variance	Weighted sum of squares $\times 4E_{46}$	Degrees of freedom	Mean square
Individual linear regressions	152	96	1·58
Individual quadratic regressions	127	84	1·52
Individual cubic regressions	111	72	1·54
Quadratic terms in individual regressions	152 − 127 = 25	96 − 84 = 12	2·08
Cubic terms in individual regressions	127 − 111 = 16	12	1·33
Quadratic and cubic terms in individual regressions	152 − 111 = 41	24	1·71

Table 2B

Hypothesis tested	Appropriate variance ratio (F)	Degrees of freedom	Critical 5 % value of F	Conclusion
Non-reality of cubic terms in the individual regressions	$\frac{1·54}{1·33} = 1·2$	72 and 12	2·38	Accept hypothesis
Non-reality of quadratic terms in the individual regressions	$\frac{2·08}{1·52} = 1·4$	12 and 84	1·87	Accept hypothesis
Non-reality of quadratic and cubic terms in the individual regressions	$\frac{1·71}{1·54} = 1·1$	24 and 72	1·68	Accept hypothesis

Our previous conclusion is thus sustained. For the data as a whole, the variation of y'_x with x is linear. We do not significantly alter the estimate of k by introducing quadratic or higher-order regression coefficients.

Consistency of the individual estimates

We now examine more closely the values of k (based on linear regressions) quoted in Table 1 for the class-period sub-groups. The lowest estimate is ·55 (E.N. 1934–38), the highest 2·63 (E.W. 1929–33). Are the twelve numbers sensibly equal, differing only through random fluctuations; or are there significant differences between them? To answer this question, postulate that there exists a single value of k appropriate to all the groups. Applying Bartlett's test [1934], in the manner shown in the Appendix, to the twelve independent estimates of this common value, we find that the hypothesis is acceptable. Alternatively, we may combine the data for all three periods, and examine whether the resulting variances for each class separately are mutually consistent.

An affirmative answer is again found acceptable (see Appendix). A similar result is obtained if the data for all classes are pooled so that the differences between quinquennia may be examined.

Heterogeneity of the data

By virtue of the preceding paragraph, we may assert that k—whatever its true value may be—is the same for each group. It is accordingly legitimate to pool all the relevant information in order to make the best single estimate of k, and to consider whether it is sensibly equal to unity. With regard to the former task, Table 2A provides the value of $k = 1{\cdot}58$, based on 96 degrees of freedom. If the data were binomially homogeneous, $k = 1$ would be the corresponding true value. Test their compatibility by forming the variance ratio $F = 1{\cdot}58/1$ with 96 and ∞ degrees of freedom for numerator and denominator respectively. The critical 5% value of F is 1·25. We therefore reject the postulate that $k = 1$. The data, we conclude, exhibit significant heterogeneity, and the variances of the estimated mortality rates are about 1·6 times greater than the homogeneous binomial values.

Dependence of k on age

The possible variation of heterogeneity with age has yet to be discussed. Over the whole range of ages of assured lives, one would hardly expect a uniform degree of heterogeneity. Whether marked variation occurs over the limited range of 46–55 years in the data with which we deal is, however, much less certain. Erratic alterations between consecutive ages are at any rate unlikely if the heterogeneity is mainly due to the presence of duplicate policies. The distribution of the latter with respect to lives probably varies regularly and slowly with age attained—a view supported by the only publicly available evidence, which comprises some small-scale data from an actual life office (Seal [1947]).

Let us test, then, whether there is any significant difference between the value of k appropriate to the younger ages alone, and that pertaining to the older ages alone. Operating upon the twelve class-period groups, separate—and statistically independent—estimates of k have been derived (a) for ages 46–50 years, and (b) for ages 51–55. (The regressions in question are, of course, distinct from those discussed earlier. The former are each based on the data of only five ages, whereas the latter were calculated on the experience of ten ages.) The results are as shown in Tables 3A and 3B.

Table 3A

Age-group	Weighted sum of squares based on individual linear regressions	Degrees of freedom	Mean square = estimate of k
46–50	46·0	36	1·28
51–55	58·6	36	1·63

Table 3B

Hypothesis tested:	That a common value of k applies to both age-groups
Variance ratio for test:	$F = 1{\cdot}63/1{\cdot}28 = 1{\cdot}27$
Degrees of freedom:	36 and 36
Critical 5% value of F:	1·75
Conclusion:	Accept hypothesis

There is thus no evidence to contradict the view that the degree of heterogeneity is uniform between the younger and older ages in the range 46–55 years. A more detailed examination, dealing with all ten ages individually, might be desirable but would be difficult with the present data. Thus the contributions from the separate ages to the total sums of squares of Table 2A can be evaluated, but—representing as they do squared deviations from a regression curve—they are not statistically independent, and tests such as Bartlett's are inapplicable. (It may be mentioned that the actual figures, for what they are worth, do not suggest any need for modifying the foregoing conclusion.) Again, if the experience in the several class-period groups had been uniform, one could estimate directly a value of k for each age. Unfortunately for this purpose, the mortality was markedly non-uniform, as we shall establish in the next section. In the absence of contrary evidence, we adhere to the statement that there was no marked variation in the degree of heterogeneity at the different ages (46–55) of the experience.

Fiducial limits for k

The value of $k = 1{\cdot}58$, applicable to all class-period groups and ages, has already been quoted (see Table 2A). This is, of course, merely an estimate, and it is unlikely that the true 'population' value of k is precisely 1·58. It is accordingly desirable to quote limits within which we can reasonably assert the true value lies, and for this purpose the fiducial argument (see, for example, Wishart [1947]) provides an appropriate technique. It can be briefly described in terms of the variance-ratio test which we have already employed.

Suppose k^* is the lowest value which we could 'reasonably' accept, and consider the consequence of postulating that this is the true value. We would then have $k = 1{\cdot}58$ as an estimate of k^*, and the test ratio would be $F = 1{\cdot}58/k^*$ with 96 and ∞ degrees of freedom for numerator and denominator respectively. (A 'true' value is equivalent to an estimate based on an infinity of degrees of freedom.) The critical 5% value of F for 96 and ∞ degrees of freedom is 1·25. Writing $1{\cdot}58/k^* = 1{\cdot}25$, we derive $k^* = 1{\cdot}3$ as the lowest acceptable value.

Similarly, the highest acceptable value is found to be $1{\cdot}58 \times 1{\cdot}29 = 2{\cdot}0$, the latter factor representing the critical 5% value for ∞ and 96 degrees of freedom respectively. We assert consequently that the true value of k lies between 1·3 and 2·0 (these limits forming the so-called '90% fiducial limits'). The sense of this assertion is that, if any value within this range were postulated as true and were tested on the basis of the statistical evidence and of the 5% level of significance, it would be accepted; and that any proposed value outside these limits would be rejected. We note that the lower limit exceeds unity—as we would expect, since the hypothesis $k = 1$ has already been examined and discredited.

IV. THE INVESTIGATION OF UNIFORMITY

The problem whether the mortality experience at a given age in each of several sub-groups was sensibly equal forms one aspect of an investigation of uniformity. Another aspect, independent of the foregoing, and of relevance when the data of more than one age are in question, concerns the *rate of change* of q_x (or y_x) with the age x. We proceed to consider the uniformity of such rates of change among the different classes of assurance and periods of experience.

Uniformity with respect to age variation

Let the suffices (i, j) indicate the ith class and jth period, thus:

First suffix	Class	Second suffix	Period
$i=1$	L.W.	$j=1$	1924–28
$i=2$	L.N.	$j=2$	1929–33
$i=3$	E.W.	$j=3$	1934–38
$i=4$	E.N.		

(The class abbreviations are explained on p. 94.) As an example, $\theta_{ij(x)}$, $E_{ij(x)}$ denote respectively the deaths and exposure in the ith class and jth period at age x nearest birthday.

The individual linear regressions of section III could have been written as

$$y_{ij(x)} = a_{ij} + b_{ij}x \quad (x = 46, 47, \ldots, 55), \quad (y_{ij(x)} = \sqrt{\{\theta_{ij(x)}/E_{ij(x)}\}}), \tag{1}$$

where the constants a_{ij}, b_{ij} differ for the various groups. A uniform rate of increase of $y_{ij(x)}$ with respect to age implies that $b_{1,1} = b_{1,2} = \ldots = b_{4,3} = b$, say. Let us therefore hypothesize that

$$y_{ij(x)} = a_{ij} + bx \quad (i = 1, 2, 3, 4; j = 1, 2, 3). \tag{2}$$

By the usual procedure, we may calculate the weighted sum of squares of the 120 deviations of the observed $y'_{ij(x)}$ from the corresponding means $y_{ij(x)}$—the latter being obtained from the regression (2) and the 'least squares' estimates of the a_{ij} and of b. This sum of squares must exceed the figure which arises when the regression (1) is adopted instead of (2), since the former set of equations contains 24 disposable constants, compared with the 13 of the latter. The actual excess provides a measure of the validity of (2) which may be tested in the customary manner. The details are summarized in Tables 4A and 4B. The weight of $y'_{ij(x)}$ has been taken as $4E_{ij(x)}$ for all i, j and x.

Table 4A

Basis of estimate	Weighted sum of squares of errors of estimate	Degrees of freedom	Mean square
Uniform age variation and linear regressions (equations (2))	182	107	1·70
Individual linear regressions (equations (1)), see also Table 2A	152	96	1·58
Difference	30	11	2·73

Table 4B

Hypothesis tested:	That a common value of b applies to all class-period groups
Variance ratio for test:	$F = 2{\cdot}73/1{\cdot}58 = 1{\cdot}7$
Degrees of freedom:	11 and 96
Critical 5 % value of F:	1·89
Conclusion:	Accept hypothesis

The variation of $y_{ij(x)}$ with age may thus be accepted as uniform for all classes of assurance and periods. This implies that the straight-line graphs of $y_{ij(x)}$ against x are sensibly parallel for all values of i and j. They do not necessarily coincide, since the individual values of the a_{ij} may differ. This point we examine below.

Uniformity between groups

In view of the preceding conclusion, uniformity between the groups need not be examined for each separate age, but merely for the ten ages as a whole. To this end, the quantity

$$Y'_{ij} = \sum_{x=46}^{55} y'_{ij(x)} \quad (i=1,2,3,4;\ j=1,2,3)$$

has been calculated as an 'over-all' measure of mortality between ages 46 and 55. The variance of Y'_{ij} is $\frac{1}{4}k \sum_{x=46}^{55} (1/E_{ij(x)})$. This measure is not, of course, unique. Other possible indices would be the group rate of mortality

$$\sum_x \theta_{ij(x)} \Big/ \sum_x E_{ij(x)}$$

or its square root; or such a quantity as $\sum_x q'_{ij(x)}$. However, if any non-uniformity were detected in the group rate (or the square root thereof), it would not be immediately clear whether it were due to differences in the mortality experience at all ages in the group, or in the average group age, or to differences in both respects. Non-uniformity in Y'_{ij}, when coupled with the knowledge of uniform age variation, can, on the other hand, be ascribed at once to essential mortality differences at all ages. The reason for choosing the square roots of the mortality rates, rather than the q'_x's, as components of Y'_{ij}, is, of course, the same as that developed in section III. Table 5 contains the values of Y'_{ij} and the respective weights (which are inversely proportional to the variances).

Table 5

	L.W. $i=1$		L.N. $i=2$		E.W. $i=3$		E.N. $i=4$		All classes	
	Y	Wt.	Y	Wt.	Y	Wt.	Y	Wt.	Y	Wt.
1924–28 $j=1$	·9277	1·26	·8847	·28	·8738	4·37	·8420	·58	3·5282	6·49
1929–33 $j=2$	·9311	1·09	·9737	·35	·8724	4·63	·8755	·78	3·6527	6·85
1934–38 $j=3$	·8590	·90	·9038	·35	·8445	4·46	·8226	·61	3·4299	6·32
All periods	2·7178	3·25	2·7622	·98	2·5907	13·46	2·5401	1·97	10·6108	19·66

The statistical technique appropriate to testing differences in such a double-entry table is the 'analysis of variance'. The usual methods are, however, inapplicable, because the variances of the individual Y'_{ij} are unequal. A modified procedure, due to Yates [1934], has therefore been adopted. It may be remarked that an alternative method of analysis of variance of contingency tables has been given by Vajda [1945], which, with necessary modifications to allow for the binomial heterogeneity, could be applied to the present data in respect of any one age.

As in the usual procedure of the analysis of variance, Yates's method is to postulate, in the first instance, that Y_{ij} (the mean value of Y'_{ij}) can be expressed in the form

$$Y_{ij} = m + l_i + n_j. \tag{3}$$

Here l_i denotes a quantity dependent on the class of assurance; n_j a quantity dependent on the period of experience; and m a 'universal' constant or grand mean. The explicit introduction of m in this sense implies that $\sum_{i=1}^{4} l_i = \sum_{j=1}^{3} n_j = 0$. It is important to recognize that the equations (3) do embody a postulate—viz. that Y_{ij} is expressible linearly in terms of three components—the validity of which must be tested. If it is invalid, *interaction* is said to be present. Interaction may thus be described as a differential period variation among the different classes of assurance; for instance, if whole-life assurance mortality decreased with time, and endowment assurance mortality increased with time, equations (3) would be inapplicable.

Yates's test for the presence of interaction is similar to those already described. Assume the truth of (3). Estimate the values of the 6 independent constants contained therein by the criterion of least squares, and calculate the weighted sum of squares of all 12 deviations of actual from estimated values of Y. That sum is associated with $12-6=6$ degrees of freedom, and leads, if the hypothesis is true, to an estimate of the variance factor k. Its compatibility with the estimate of k from section III, and hence the validity of our postulate, may be tested by means of the variance-ratio distribution.

The existence of differential mortality between classes of assurance (or of a 'class effect' in the usual terminology) can be similarly examined. Hypothesize that the mortality between classes is uniform, i.e. that $l_1 = l_2 = l_3 = l_4 = 0$ in (3), and on this basis repeat the least-squares calculation referred to in the last paragraph. The increase in the weighted sum of squares of errors of estimate over the sum there obtained gives, if the hypothesis is true, another estimate of k based on 3 degrees of freedom (corresponding to the three *independent* class constants l which have been set equal to zero). As before, this estimate may be tested against the statistically independent value of $k = 1.58$ previously derived. If the ratio of the two estimates surpasses the critical 5 % value, we reject the hypothesis, admit the existence of different mortality rates in the different classes of assurance, and examine the figures more closely for the source of the differences. Such a conclusion, it may be emphasized, would refer to the mortality experience of the classes over all periods combined. The existence of a 'period effect', or variation in the experiences in the three quinquennia when all classes of policy are combined, can be tested similarly.

The results of the relevant calculations and tests are summarized in Table 6. The mean square is in each case compared with the estimate $k = 1.58$ based on 96 degrees of freedom.

Table 6

Hypothesis tested	Appropriate weighted sum of squares	Degrees of freedom	Mean square	F = mean square ÷ 1·58	Critical 5 % value of F
Absence of class effect	81·3	3	27·1	17	2·7
Absence of period effect	52·0	2	26·0	16	3·1
Absence of class-period interaction	28·4	6	4·73	3·0	2·2

On the 5% level of significance, therefore, all three hypotheses fail. We must admit the reality of a class effect, of a period effect, and of a class-period interaction. Closer inspection of Table 5 (invoking 'Student's' t-test where appropriate) reveals the following—the principal results of our investigation of uniformity:

(*a*) For all three periods combined, the two whole-life classes exhibited uniform mortality. So, too, did the two classes of endowment assurances. The mortality experience of the latter was, however, significantly lighter than that of the former groups.

(*b*) For all classes of assurance combined, there was a significant difference of mortality in each quinquennium. The heaviest rates were experienced in 1929–33 and the lightest in 1934–38.

(*c*) The whole-life with profits and endowment assurance without profits mortality experience followed the 'over-all' secular trend described in (*b*). That of the other two classes of assured lives did not, however, conform to this pattern. These differences in secular variation explain the 'interaction effect' isolated above.

These particular results could, of course, have been arrived at by standard elementary methods—if not by mere inspection of the original figures. A critic could fairly ask what advantage or additional knowledge accrued through the use of the more complicated techniques. First, they led us to recognize the uniformity of the rate of mortality variation with age. It was not perhaps immediately obvious that the $y'_{ij(x)}$ could be represented as varying linearly with x, with sensibly equal gradients for all values of i and j. Secondly, the use of the measures Y'_{ij}, coupled with the last conclusion, enabled us to investigate compactly the over-all mortality differences between groups, free of the complications—unequal average ages, and so on—associated with 'group' rates of mortality. Thirdly, the 'analysis of variance' procedure permitted several aspects of non-uniformity to be examined at once. The conventional actuarial methods require any one rate to be tested against every other rate—a laborious procedure which in any event does not answer directly the natural question whether the data *as a whole* are uniform. Finally, the present methods have made explicit allowance for the heterogeneity of the material. The use of 1·58 (with 96 degrees of freedom) as the measure of random variation in Table 6 may be recalled—the justification for which rests on the whole investigation of section III. If, as not uncommonly happens, the heterogeneity had been ignored, the corresponding measure in the uniformity tests would have been taken as unity (with an infinity of degrees of freedom). It is easy to see the possible mistakes of judgment which could ensue. Suppose, as an example, that the sum of squares appropriate to testing class effects in Table 6 had been 4·0. The variance ratio would be $4{\cdot}0/1{\cdot}58=2{\cdot}53$ with 3 and 96 degrees of freedom. Noting that the critical 5% value is 2·7, we should accept the hypothesis that there was no difference between the mortality under the various classes of policy. But by ignoring the heterogeneity, we would take a variance ratio of $4{\cdot}0/1$, with 3 and ∞ degrees of freedom—the critical 5% value is 2·6—and would unjustifiably conclude that a real class distinction existed.

Fiducial limits for the mortality rates

As in the discussion of heterogeneity, it may be useful to indicate how limits may be assigned within which a rate of mortality is almost sure to lie. The fiducial argument can be readily applied. If θ is the number of claims by death out of E policy years of exposure to risk, if q_0 is the 'population' mortality rate, and if Eq_0 is sufficiently large, $z=(\theta-Eq_0)/\sqrt{\{kEq_0(1-q_0)\}}$ is distributed approximately normally with a mean of zero and a variance of unity. Given an actual pair of values of θ and E, we would test any hypothesis concerning q_0 by means of the tables of the normal distribution function. On the 5% level of significance, we would accept any hypothetical mortality rate q_0 if it led to a value of z within the range $-1\cdot96\leqslant z\leqslant 1\cdot96$. Accordingly, we can fix limits by solving for q_0 the equations

$$(\theta-Eq_0)/\sqrt{\{kEq_0(1-q_0)\}}=\pm 1\cdot96. \tag{4}$$

The roots are the '95% fiducial limits' of the rate of mortality corresponding to given θ and E, and it can be asserted—in the usual sense—that the true value of q_0 lies somewhere between these limits.

Numerical results are quoted in Table 7 for the particular case of whole-life with-profit policies during 1934–38. A value of $k=1\cdot5$ has been used in equation (4).

Table 7. Whole-life with profits experience, 1934–38, medically examined lives, duration 5 and over

Age x	E	θ	θ/E	95% fiducial limits of q_x
46	15,638½	85	·00544	·0042–·0070
47	16,621	82	493	40 64
48	17,714	110	621	49 78
49	18,755	92	491	38 63
50	19,910	140	703	57 86
51	21,064½	138	655	53 80
52	22,249½	209	939	80 111
53	23,587	210	890	76 105
54	25,054½	279	1114	97 128
55	26,682	292	1094	95 126

The fiducial limits, it may be remarked, differ appreciably from those which would be quoted on the basis of elementary theory, viz. $(\theta\pm 1\cdot96\sqrt{\theta})/E$.

V. SUMMARY OF CONCLUSIONS

(1) The investigation was concerned with the mortality experience of medically examined assured lives of ages 46–55 inclusive, over the period 1924–38, at durations 5 years and over. Four types of policy, whole-life and endowment assurances, each with and without profits, have been distinguished. All data were abstracted from information made available to the public by the Continuous Mortality Investigation Committee—a fact which has partly dictated the scope and methods of the analysis.

(2) For all twelve class-period groups which were examined, the square root of the crude mortality rate, $\sqrt{(\theta_x/E_x)}$, varied linearly with age over the range 46–55 years.

(3) The data were not homogeneous in the binomial sense. The degree of heterogeneity was sensibly the same for all twelve class-period groups, and can be expressed by the statement that the variance of an estimated rate of mortality was about 1·6 times the homogeneous binomial value of $q(1-q)/E$.

(4) The true value of the measure of heterogeneity, of which 1·6 was the best single estimate, almost certainly lay between 1·3 and 2·0.

(5) The degree of heterogeneity was sensibly the same for the younger and the older ages in the range 46–55 years.

(6) The experience of the twelve groups was uniform with respect to the rate of variation of mortality with age.

(7) Over the whole period 1924–38, the mortality experience of the two whole-life classes was uniform. So, too, was that of the two endowment assurance classes. However, the mortality of the latter pair was significantly lighter than that of the former.

(8) The mortality of the four classes combined was lightest in the period 1934–38, and heaviest in 1929–33, the experience in 1924–28 having been intermediate between these extremes.

(9) The secular variations of mortality among the four individual classes of policyholder were materially different from one another.

Finally, it should be stressed that the foregoing is no more than a summary of acceptable hypotheses. None of the conclusions has been rigorously established as true. The hypotheses which have been put forward are believed to be reasonable; they have aimed at describing the observed facts as simply as possible with the least number of statistical parameters; and it has been shown that they are not inconsistent with the recorded data. That is as much as one can assert. With additional *a priori* knowledge, one might be led to suggest other hypotheses, and they too might prove compatible with the facts. If that occurred, it would illustrate a weakness inherent in the methods of analysis; for the power accurately to distinguish a true from every false theory is beyond the scope of statistics.

ACKNOWLEDGMENTS

It is a pleasure to the author to acknowledge his debt and gratitude to Mr H. L. Seal, B.Sc., F.F.A., Dr S. Vajda, Dr. Phil., and Mr L. Ginsburg, M.A., F.F.A., for their helpful advice and stimulating criticism.

APPENDIX

1. Calculation of the weighted sums of squares of errors of estimate

In section III, it was postulated that the quantity y_x could be expressed as a polynomial of degree 3 in x, and it was required to find those values $\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{a}_3$ which would make

$$S_4 = \sum_{x=46}^{55} w_x (y'_x - \hat{a}_0 - \hat{a}_1 x - \hat{a}_2 x^2 - \hat{a}_3 x^3)^2$$

a minimum. ($w_x = E_x/E_{46}$ denotes the weight of y'_x.)

The required values satisfy the equations

$$\partial S_4 / \partial \hat{a}_i = 0 \quad (i = 0, 1, 2, 3)$$

or
$$\sum_{x=46}^{55} w_x x^i y'_x = \sum_{x=46}^{55} w_x x^i (\hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2 + \hat{a}_3 x^3) \quad (i = 0, 1, 2, 3).$$

Solving this set of linear equations for the $\hat{a}_i$ and substituting in S_4, the required minimum weighted sum of squares of errors of estimate is obtained. It can be shown that the sum is equal to

$$S_4\,(\text{min.}) = \begin{vmatrix} \Sigma w_x y_x'^2 & \Sigma w_x y_x' & \Sigma w_x x y_x' & \Sigma w_x x^2 y_x' & \Sigma w_x x^3 y_x' \\ \Sigma w_x y_x' & \Sigma w_x & \Sigma w_x x & \Sigma w_x x^2 & \Sigma w_x x^3 \\ \Sigma w_x x y_x' & \Sigma w_x x & \Sigma w_x x^2 & \Sigma w_x x^3 & \Sigma w_x x^4 \\ \Sigma w_x x^2 y_x' & \Sigma w_x x^2 & \Sigma w_x x^3 & \Sigma w_x x^4 & \Sigma w_x x^5 \\ \Sigma w_x x^3 y_x' & \Sigma w_x x^3 & \Sigma w_x x^4 & \Sigma w_x x^5 & \Sigma w_x x^6 \end{vmatrix} \div \begin{vmatrix} \Sigma w_x & \Sigma w_x x & \Sigma w_x x^2 & \Sigma w_x x^3 \\ \Sigma w_x x & \Sigma w_x x^2 & \Sigma w_x x^3 & \Sigma w_x x^4 \\ \Sigma w_x x^2 & \Sigma w_x x^3 & \Sigma w_x x^4 & \Sigma w_x x^5 \\ \Sigma w_x x^3 & \Sigma w_x x^4 & \Sigma w_x x^5 & \Sigma w_x x^6 \end{vmatrix}$$

This formula enables one to calculate S_4 (min.) without solving the 'least squares' equations explicitly for the $\hat{a}_i$'s. The determinant in the denominator is, of course, derived by deleting the first row and first column from the determinant in the numerator.

The 5×5 determinant (Δ, say) in the numerator of S_4 (min.) was evaluated by the operation of pivotal condensation repeated four times in succession. In each case the element in the second row and second column was taken as the pivotal element (e.g. at the first operation the second row was divided throughout by Σw_x, whereby an element of unity was produced in the position (2, 2), and the usual condensation formula was then applied). At the last stage a 2×2 determinant is reduced to a single element, and this element is readily seen to be the quantity S_4 (min.) which we seek. The process is essentially one of expressing Δ in the form $\alpha_1 \alpha_2 \alpha_3 \alpha_4 S_4$ (min.), where the α's are the factors arising in the condensation operations (e.g. $\alpha_1 = \Sigma w_x$). Because of the particular choice of pivotal elements the product $\alpha_1 \alpha_2 \alpha_3 \alpha_4$ is equal to the determinant in the denominator of S_4 (min.).

It is also readily seen that the leading element (i.e. that in position (1, 1)) in the 4×4 determinant which results from the first condensation operation on Δ is equal to S_0 (min.), the weighted sum of squares of errors of estimate if it is postulated that $y_x = a_0$. The next step in the reduction of Δ gives a 3×3 determinant, the leading element of which equals S_1 (min.), the weighted sum of squares of errors of estimate corresponding to a linear regression $y_x = a_0 + a_1 x$. Similarly, after the next condensation the leading element in the 2×2 determinant is S_2 (min.), corresponding to a quadratic regression of y_x on x. Thus, by the calculation of only one determinant in the manner indicated, the weighted sums of squares of errors of estimate were derived for the cubic and all lower-order regressions.

2. Consistency of the individual estimates of variance

The estimated variances (or estimates of k) for each of the twelve class-period groups, based on individual linear regressions, are contained in Table 1, and are quoted for convenience below.

Table 8. Estimates of k, based on 8 degrees of freedom in each case

Period	Class of assurance				
	L.W.	L.N.	E.W.	E.N.	Total
1924–28	1·18	1·75	1·01	1·72	5·66
1929–33	2·06	1·94	2·63	1·05	7·68
1934–38	2·45	1·33	1·35	·55	5·68
Total	5·69	5·02	4·99	3·32	19·02

Bartlett's test of the mutual consistency of the twelve estimates proceeds by the following steps:

(i) $\lambda = \dfrac{\text{geometric mean of the twelve estimates}}{\text{arithmetic mean of the twelve estimates}}$

$= 1{\cdot}46/1{\cdot}58.$

(ii) $x = -96 \log_e (1{\cdot}46/1{\cdot}58) = 7{\cdot}67.$

(96, or 12×8, represents the total number of degrees of freedom involved in all twelve estimates.)

(iii) $c_1 = \Sigma \{1/(\text{degrees of freedom in each estimate})\}$
$\quad - 1/(\text{total number of degrees of freedom})$
$= 12/8 - 1/96 = 1{\cdot}49.$

Let n = number of separate estimates = 12.

Then the quantity

$$\chi_0^2 = x/\{1 + \tfrac{1}{3}c_1/(n-1)\} = 7{\cdot}6$$

is distributed as a χ^2 variate with $n-1=11$ degrees of freedom, and may be tested accordingly. The critical 5 % value of χ^2 for 11 degrees of freedom is 19·7. Since the actual value of 7·6 does not exceed this figure, the hypothesis is acceptable that a single value of k applies to all classes and periods. This conclusion may be checked by the following tests.

Consistency of the variances for the separate classes

Consider merely the greatest and least estimates of k for the various classes of assurance, when the experience of all three periods is combined. If these two are found to be mutually consistent, Bartlett's test when applied to all four estimates would yield the same conclusion.

The relevant details are as follows.

$$F = \frac{\text{greatest variance estimate}}{\text{least variance estimate}}$$

$$= \frac{5{\cdot}69 \text{ (for L.W.)}}{3{\cdot}32 \text{ (for E.N.)}} = 1{\cdot}7, \text{ with 24 and 24 degrees of freedom.}$$

Critical 5 % value of $F = 2{\cdot}0$.

It is concluded that a single value of k is applicable to all four classes of assurance (over the whole period 1924–38).

Consistency of the variances for the separate periods

Comparing again merely the greatest and least values,

$$F=\frac{7\cdot68 \text{ (for 1929–33, all classes combined)}}{5\cdot66 \text{ (for 1924–28, all classes combined)}}$$

$= 1\cdot4$ with 32 and 32 degrees of freedom.

Critical 5 % value of $F=1\cdot8$.

It is concluded therefore that a single value of k is applicable to all three periods (for all classes combined).

Comparison of variance for the E.N. class with that for all other classes

Inspection of Table 8 might suggest that one value of k might apply to the L.W., L.N. and E.W. classes, but that a significantly smaller value might possibly be relevant to the E.N. class. Comparing these two groupings, on the hypothesis that there is no such difference in k,

$$F=\frac{15\cdot70}{3\times3\cdot32}=1\cdot6 \text{ with 72 and 24 degrees of freedom.}$$

Critical 5 % value of $F=1\cdot8$.

The foregoing suggestion cannot therefore be sustained.

REFERENCES

BARTLETT, M. S. The problem in statistics of testing several variances. *Proc. Camb. Phil. Soc.* Vol. XXX [1934].

DAW, R. H. On the validity of the statistical tests of the graduation of a mortality table. *J.I.A.* Vol. LXXII [1945].

FISHER, R. A. and YATES, F. *Statistical Tables for Use in Biological, Agricultural and Medical Research.* (Oliver and Boyd, Edinburgh) [1942].

SEAL, H. L. Tests of a mortality table graduation. *J.I.A.* Vol. LXXI [1940].

SEAL, H. L. Probability distribution of deaths at age x when policies are counted instead of lives. *Skandinavisk Aktuarietidskrift* [1947].

VAJDA, S. The analysis of variance of mortality rates. *J.I.A.* Vol. LXXII [1945].

WISHART, J. The variance ratio test in statistics. *J.S.S.* Vol. VI [1947].

YATES, F. The analysis of multiple classifications with unequal numbers in the different classes. *J. Amer. Statist. Ass.* Vol. XXIX [1934].