JIA 71 (1943) 0005-0067

TESTS OF A MORTALITY TABLE GRADUATION

By H. L. SEAL, B.Sc., F.F.A.
Statistician to the Director of Air Matériel, Admiralty

[Submitted to the Institute, 22 April 1940]

"Sangue di Bacco! ... les dés sont pipés!"
(Bertrand)

A FORMER President of the Institute once remarked that whereas the graduation of a mortality table might be considered a specialized technique not required by the practising actuary, the efficient testing of any graduation submitted to him should be an integral part of his actuarial ability. The student anxious to equip himself in this respect encounters considerable divergence of opinion regarding the best method of testing a graduation. There is complete agreement as to the duality of purpose of such a test: smoothness and fidelity to the data are essential factors. But whilst the criterion of smoothness is easily formulated—the run of the second or third differences of q_x must show no awkward breaks—the views relating to the tests of adherence to the data are divided between the application of a "mean deviation" rule at each age or group of ages and a purely "practical" judgment based on that elusive virtue, experience.

It is here proposed to conduct a critical examination of the various tests of fidelity to the data and to show that some generally used criteria are not statistically adequate.

HISTORICAL REVIEW

In the early days of actuarial science the success of a mortality table graduation was judged by means of comparisons of the observed and graduated values of q_x, e_x, and l_x. Apparently it was Thiele, writing in 1872, who first suggested that a comparison of l_x and l'_x left something to be desired: he says (*J.I.A.* Vol. XVI, p. 120), "In examining my adjusted table of mortality by a comparison of the adjusted and unadjusted numbers of living, the fact must not be overlooked that the unadjusted number living at any age is the result of a calculation on which the observations for all ages between that age and the youngest observed have influence; and that therefore a long series of numbers all differing on the same side is no proof against the correctness of the adjustment, as is the case where each of the numbers compared depends upon its own particular observation, as, for instance, the probabilities of living a year."

Up to this time the use of "expected deaths" had been limited to the comparison of two or more tables of mortality (e.g. Samuel Brown, *J.I.A.* Vol. VIII, p. 184) or to demonstrations of the effect of selection (e.g. Meikle, *J.I.A.* Vol. XIII, p. 261), and it was left to Thiele (*J.I.A.* Vol. XVI, p. 313) to provide the first test of a graduation based on a comparison of the actual deaths with those expected on the basis of the graduated values of q_x. In fact he went a step further and introduced a criterion to test the goodness of the graduation. He calculated

$$\sum_x \frac{(\theta_x - \mathrm{E}_x q_x)^2}{\mathrm{E}_x p_x q_x},$$

and compared it with its expectation which, he stated, is equal to the number of ages graduated less the number of constants used in the graduating formula. It will be seen below that this expression plays a central part in Pearson's (1900)* χ^2 test applied to mortality data. The method of "expected deaths" met with gradual acceptance but there seems to have been no application, or even mention, of Thiele's criterion since the date he proposed it.

The seventeenth volume of the *Journal* saw the first use of the "probable error" concept in connexion with mortality observations, although there Wittstein (*J.I.A.* Vol. XVII, pp. 178, 355, 417) is more concerned to test whether the mortality of an office diverges significantly from some standard table employed by it.

In 1879 McCay (*J.I.A.* Vol. XXII, p. 24) proposed his five empirical criteria to be used in the testing of a graduation: "The excellence of the adjustment would...be shown, first, by the general regularity in the rates, second, by the near agreement in the whole number of deaths, third, by the equality of the [respective sums of the] positive and negative [accumulated] deviations, fourth, by the smallness of the [individual] accumulated deviations, and fifth, by the frequency of the changes in their signs from positive to negative." These tests were very generally adopted, to the complete neglect of previous contributions, so that Makeham, writing in 1890 (*J.I.A.* Vol. XXVIII, p. 316), was

* References to *J.I.A.*, etc., are given in this paper as they occur, but other references have been collected in alphabetical order of authors at the end.

able to state that only one English writer had, up to that time, used the calculus of probabilities in connexion with mortality observations. He set out to remedy this defect in the simplest manner possible and introduced the use of the mean (average) deviation (approximately equal to four-fifths of the square root of the expected deaths) at each age. This test was systematically employed by G. F. Hardy in the British Offices, 1863/93, graduations and it is still in general use. No further contributions to this subject have appeared in the *Journal* since Makeham wrote.

The remainder of this paper is limited to a consideration of the graduation of q_x, but it is to be observed that Cramèr (1927) has developed a test applicable to a graduation of μ_x and analogous to the χ^2 test set forth below. Cramèr and Wold (1935) have used this criterion, viz. the value of

$$\sum_x \frac{(\theta_x - E_{x+\frac{1}{2}}\mu_{x+\frac{1}{2}})^2}{E_{x+\frac{1}{2}}\mu_{x+\frac{1}{2}}},$$

in their well-known paper on forecasting of mortality rates.

The effect of the heterogeneity of mortality observations upon probabilistic* tests of a graduation was mentioned by C. N. Jones in 1894 (*T.A.S.A.* Vol. III, p. 299) but little attention has been paid to his warnings, at least in this country. In order not to confuse the issue it will for the moment be assumed that there is an equal probability of dying within a year applicable to all persons aged precisely x, and that each person dies quite independently of any other. Deviations from these assumptions will be dealt with thereafter.

THE PROBLEM

Strictly, the graduated table of mortality must be regarded as an ideal smooth set of values of q_x from which the originally observed values could, with reasonable probability, have arisen in random sampling with the actually observed numbers of persons exposed. It is assumed that there is an infinite population at each age of the mortality table. Each constituent of these populations is marked "dead" or "alive" in such a way that at age x the proportion of "dead" to the total is q_x, where q_x is the graduated rate

* This word seems to be acquiring current usage in the works of a number of writers on statistical theory. It is preferable to "stochastic" if only because its meaning is plain.

of mortality at age x. A sample of E_x individuals is supposed drawn from the hypothetical population corresponding to age x and it is observed that θ_x of these individuals are "dead". The problem is to test the hypothesis relating to the various proportions of "dead" in these populations and, if possible, obtain a criterion for its rejection or acceptance. A satisfactory method lies in the calculation of the probability that the observed deaths, or a more improbable set of deaths as judged by the graduated table, would occur by chance in such a random sample and, from this probability, to form a judgment of the goodness of the graduation.

This rejection or acceptance of a hypothesis on the ground of the probability of its producing a given sample is a method peculiar to Statistics. It is a common principle of daily life to be surprised at the occurrence of an improbable event, and the more improbable the event, the greater the surprise. Bertrand (1889) records the case of a man who, in the presence of the Abbé Galiani, bet he would obtain three sixes in one throw with three dice. He succeeded ("Cette chance est possible, dit-on"), and then repeated the feat four more times in succession. Observing this the Abbé was surprised into making the remark at the head of this paper. Presumably his surprise was intuitive but, on the hypothesis that each die was unbiased, the probability of the observed events is $(1/6)^{15}$, i.e. about ·000,000,000,002, and this is so small that it seems reasonable to reject the hypothesis in favour of some other hypothesis, e.g. that the dice were loaded.

It is to be noted that if the probability of a certain set of observations turns out to be "reasonable", say ·2, on a certain hypothesis, this does not mean that the hypothesis is verified, for there may well exist other hypotheses leading to probabilities of ·3 or more. Whilst, therefore, a hypothesis may be rejected on the grounds that it produces a very small probability for the observed data, it can only be provisionally accepted if the probability of the data is not very small.

There can be no hard and fast rule to distinguish between a probability calling for rejection of a hypothesis and one allowing its provisional acceptance. Certain practical facilities accruing from the adoption of a certain graduation formula will lead to its preference over another formula which may produce a higher probability for the observations. Perhaps the data are intractable and will not yield a really successful graduation. Or perhaps speed in the production of the graduated table is the first consideration, and a low probability of the observations will not lead to a rejection of the graduation. Apart from these and other practical factors, however, it will be assumed in what follows that if the critical probability be less than ·001 or greater than ·999 the hypothesis (i.e. the graduation under test) would be rejected; in the first case because it leads to too small a probability of the data (i.e. the graduation is too far from the facts), and in the second instance because the data are apparently undergraduated (i.e. the graduation is too close to the facts). If these limits are considered too stringent, the limits ·01 and ·99 may be adopted. In the latter case about two graduations in every hundred will be rejected when in fact they represent the true rates from which the observed values of q_x were obtained by sampling, i.e. on the average two graduations in a hundred will be wrongly rejected when the hypothesis tested is true. In many agricultural and industrial applications of statistical theory the hypothesis is considered disproved if the critical probability falls outside the limits ·05 and ·95.

THE PROBABILISTIC ARGUMENT

If q is the probability of a certain event, namely "death", and if E independent observations are made in which the event either does or does not happen, the probabilities of 0, 1, 2, ..., E events (deaths) are given by the successive terms of the expansion of

$$(p+q)^{\mathrm{E}} = \sum_{r=0}^{\mathrm{E}} \binom{\mathrm{E}}{r} p^{\mathrm{E}-r} q^r, \qquad \ldots\ldots(1)$$

where $p+q=1$. The mean number of events is Eq and the variance of the distribution* is Epq. The mean deviation of the number of events from their mean is approximately four-fifths of the standard deviation [cp. G. F. Hardy (1909), Note A].

Applying this argument directly to mortality observations, if there are E_x persons exposed to risk at age x, each subject to a rate of mortality of q_x and each living and dying independently of any other, the expected number of deaths among the group is $\mathrm{E}_x q_x$ and the value of the numerical deviation to be "expected" from this

* In Appendix I there is a résumé of the statistical theory required in what follows.

average number of deaths is nearly $\cdot 8\sqrt{(E_x q_x p_x)}$. Hence the comparison of $|\theta_x - E_x q_x|$ with $\cdot 8\sqrt{(E_x p_x q_x)}$ at any particular age is analogous to the comparison of some observed value of a random variable (e.g. "height") with its mean. Such a procedure is of little value in the individual case, but the comparison of the actual numerical deviations with their expected values is useful when extended to all the ages in the table. For, when $E_x q_x$ exceeds 10, say, the mean and median of the distribution (1) nearly agree and there are thus, on the average, as many deviations greater than the corresponding mean deviation as there are below it. Nevertheless, even supposing the requirements of this test are satisfied, the graduation may still be very unsatisfactory; some of the deviations may be extremely improbable and hence the hypothesis that the observations occurred as a random sample from the graduated table would be untenable.

The opportunity is taken to mention that the common practice of grouping deaths, say, in fives and calculating the corresponding mean deviation (see C. W. Kenchington, *J.I.A.* Vol. XLIV, p. 105) is open to serious criticism. The effect of the grouping generally is to collect, perhaps, large positive and negative deviations together into one fairly small deviation and thus to produce an apparent improvement in the graduation. Table 1 is extracted from Kenchington's paper and shows a comparison of actual and expected deaths

Table 1. Summary of O^{JF} data and comparison of actual deaths with those expected by Kenchington's 27-term summation formula graduation

Age group	Exposed to risk	Actual deaths	Expected deaths	Deviation	Expected deviation
20–24	1,425	9	11·3	2·3	2·68
25–29	4,915	50	48·1	− 1·9	5·52
30–34	8,492	92	92·8	·8	7·66
35–39	10,603	135	136·2	1·2	9·27
40–44	11,644	150	147·7	− 2·3	9·66
45–49	11,579	135	135·2	·2	9·25
50–54	10,225	144	141·8	− 2·2	9·46
55–59	7,888	155	151·7	− 3·3	9·76
60–64	5,291	142	146·0	4·0	9·53
65–69	2,863	113	107·5	− 5·5	8·14
70–74	1,266	64	76·1	12·1	6·76
Total	76,191	1189	1194·4		

in quinquennial age groups for the O^{JF} Table. Ten deviations out of the eleven are below the corresponding average, in six cases very considerably. In Table 2 the O^{JF} data have been dealt with age by age with the result that only 29 out of 50 deviations in column (6) are below the corresponding average found in column (7) of the table. The first comparison would strongly suggest undergraduation, the second reveals no such circumstance.*

The test of mean deviations suffers from the defect that it does not distinguish between large and small differences from the mean deviation. Now it is clear that, by means of the distribution (1), the probability of any given deviation at a particular age may be calculated exactly. In view of the fact that deviations in excess of the expected number of deaths are just as important as deviations in defect, it is justifiable to consider the probability of a deviation at any age equal to, or in excess of, the numerical value of the deviation actually observed, viz. $|\theta - Eq| = d$, say. The probability of obtaining a deviation equal to or greater than d is equal to the sum of the probabilities of obtaining $Eq+d$ or more deaths and of observing $Eq-d$ or less deaths, viz.

$$\binom{E}{Eq+d} p^{Ep-d} q^{Eq+d} + \binom{E}{Eq+d+1} p^{Ep-d-1} q^{Eq+d+1} + \dots + q^{E}$$

$$+ p^{E} + \binom{E}{1} p^{E-1} q + \dots + \binom{E}{Eq-d} p^{Ep+d} q^{Eq-d}$$

$$= \sum_{r=Eq+d}^{E} \binom{E}{r} p^{E-r} q^{r} + \sum_{r=0}^{Eq-d} \binom{E}{r} p^{E-r} q^{r}. \qquad \dots\dots(2)$$

This probability may be calculated, although with considerable labour, for each age of the graduated table. Assume for the moment that this has been done or that some reasonably close approximation to the probabilities has been found. At any age the observations are quite independent of those at any other age so the calculated probabilities may be compounded by multiplication. The resulting probability, Q_0 say, is the chance of a set of deviations from the expected number of deaths as large as, or larger than, those actually observed; it is necessarily very small and decreases with every additional age included. It is required to have some standard whereby to judge this value Q_0. If the

* In connexion with the general advisability of "grouping" see Elderton (1938), p. 203.

Table 2. Tests of Kenchington's graduation of O^{JF}

(1) x	(2) $q_x^{O^{JF}}$	(3) E_x	(4) θ_x	(5) $E_x q_x$	(6) $\theta_x - E_x q_x$	(7) $\cdot 8\sqrt{(E_x p_x q_x)}$	(8) $\frac{\lvert\theta_x - E_x q_x\rvert - \frac{1}{2}}{\sqrt{(E_x p_x q_x)}} = z_x$	(9) $\operatorname{colog}_{10}(1 - \alpha_{z_x})$	(10) $\frac{(\theta_x - E_x q_x)^2}{E_x p_x q_x}$
20	·00622	105	1						
21	680	158	3						
22	735	288	—						
23	802	382	3						
24	872	492	2						
25	927	672	8	6·23	1·77	1·99	·511	·21513	·508
26	962	846	12	8·14	3·86	2·27	1·184	·62633	1·850
27	981	989	10	9·70	·30	2·48	—	·00000	·009
28	994	1127	13	11·20	1·80	2·66	·390	·15705	·293
29	·01005	1281	7	12·87	− 5·87	2·86	1·504	·87752	2·706
30	1020	1432	12	14·61	− 2·61	3·04	·555	·23740	·471
31	1045	1567	20	16·38	3·62	3·22	·775	·35819	·808
32	1080	1729	14	18·67	− 4·67	3·44	·970	·47880	1·182
33	1122	1831	20	20·54	− ·54	3·61	·031	·01088	·014
34	1170	1933	26	22·62	3·38	3·78	·609	·26558	·511
35	1217	2006	22	24·41	− 2·41	3·93	·389	·15659	·241
36	1259	2071	28	26·07	1·93	4·06	·282	·10906	·144
37	1294	2134	24	27·61	− 3·61	4·18	·596	·25871	·479
38	1319	2179	25	28·74	− 3·74	4·26	·608	·26505	·493
39	1328	2213	36	29·39	6·61	4·31	1·135	·59112	1·506
40	1321	2286	38	30·20	7·80	4·37	1·337	·74179	2·042
41	1300	2303	34	29·94	4·06	4·35	·655	·29033	·558
42	1271	2345	36	29·80	6·20	4·34	1·051	·53275	1·306
43	1241	2343	18	29·08	−11·08	4·29	1·974	1·31532	4·277
44	1214	2367	24	28·74	− 4·74	4·26	·796	·37055	·792
45	1187	2381	33	28·26	4·74	4·23	·802	·37412	·805
46	1165	2340	23	27·26	− 4·26	4·15	·724	·32876	·674

47	1152	2327	22	26·81	− 4·81	4·12	·837	·39514	·872
48	1154	2297	28	26·51	1·49	4·09	·193	·07214	·085
49	1177	2234	29	26·29	2·71	4·08	·434	·17764	·283
50	1227	2189	26	26·86	− ·86	4·12	·070	·02494	·028
51	1297	2120	25	27·51	− 2·51	4·17	·386	·15521	·232
52	1382	2075	38	28·68	9·32	4·25	1·659	1·01271	3·073
53	1479	1962	24	29·02	− 5·02	4·28	·845	·39999	·882
54	1580	1879	31	29·69	1·31	4·32	·150	·05514	·059
55	1685	1792	24	30·20	− 6·20	4·36	1·046	·52936	1·295
56	1799	1689	33	30·39	2·61	4·37	·386	·15521	·228
57	1923	1576	26	30·31	− 4·31	4·36	·699	·31466	·624
58	2066	1474	37	30·45	6·55	4·37	1·108	·57209	1·438
59	2234	1357	35	30·32	4·68	4·36	·768	·35410	·740
60	2423	1251	30	30·31	− ·31	4·35	—	·00000	·003
61	2619	1160	29	30·38	− 1·38	4·35	·162	·05983	·065
62	2810	1064	21	29·90	− 8·90	4·31	1·558	·92361	2·726
63	2980	964	27	28·73	− 1·73	4·22	·233	·08844	·108
64	3130	852	35	26·67	8·33	4·07	1·540	·90813	2·686
65	3282	743	26	24·39	1·61	4·39	·229	·08679	·110
66	3481	658	22	22·90	− ·90	3·76	·085	·03046	·036
67	3761	577	23	21·70	1·30	3·66	·175	·06496	·081
68	4136	481	24	19·89	4·11	3·49	·827	·38909	·885
69	4611	404	18	18·63	− ·63	3·37	·031	·01088	·022
70	5143	351	16	18·05	− 2·05	3·31	·375	·15017	·245
71	5685	295	14	16·77	− 2·77	3·18	·571	·24565	·484
72	6213	252	10	15·66	− 5·66	3·07	1·346	·74884	2·182
73	6700	207	12	13·87	− 1·87	2·88	·381	·15292	·270
74	7211	161	12	11·61	·39	2·63	—	·00000	·014
Totals								16·63913	41·425

deviations had actually been ruled by chance alone and the corresponding value of Q had been calculated repeatedly for various random samples, there would be an average value of Q, a standard deviation of Q and, in short, a probability distribution for Q. Writing 2 colog_e Q $= x$, E. S. Pearson (1938) has shown that

$$p(x) = \frac{1}{2^k \Gamma(k)} x^{k-1} e^{-\frac{1}{2}x},$$

where k is the number of probabilities (ages) entering into the product Q (cp. the remarks on the Type III law in Appendix I).

A single summary test of the adequacy of the graduation is thus provided by the calculation of the probability that, in random sampling, a value of Q would arise as small as, or smaller than, Q_0, the value actually calculated for the graduation in hand. This probability is

$$\frac{1}{2^k \Gamma(k)} \int_{2 \operatorname{colog}_e Q_0}^{\infty} x^{k-1} e^{-\frac{1}{2}x} \, dx = P_{Q_0} \text{ say.}$$

Table A provides values of x_0 corresponding to certain "critical" values of

$$P = \frac{1}{2^{\frac{1}{2}f} \Gamma(\frac{1}{2}f)} \int_{x_0}^{\infty} x^{\frac{1}{2}f-1} e^{-\frac{1}{2}x} \, dx$$

and values of f at quinquennial points up to 170. Hence if $x_0 = 2 \operatorname{colog}_e Q_0$, calculated for the graduation under test, falls outside the limits fixed, say, by the ·99 and ·01 values of x corresponding to $f = 2k$, then the graduation is to be rejected.

Provided, then, that approximations to the probabilities defined by (2) can be calculated with reasonable simplicity, a test of a graduation has thus been developed which is not open to the criticisms that damn the mean deviation method. As is well known (cf. Appendix I) the binomial distribution may, with certain restrictions, be replaced by the normal curve. This is equivalent to writing (2) in the form

$$\frac{2}{\sqrt{(2\pi)}} \int_{\frac{d-\frac{1}{2}}{\sqrt{(Epq)}}}^{\infty} e^{-\frac{1}{2}t^2} \, dt. \qquad \text{......(3)}$$

It appears that the degree of accuracy of this approximation has never been investigated for probabilities, q, and numbers of observations, E, comparable with those encountered in actuarial practice. A rule occasionally adopted in statistical applications is to assume that the normal curve approximation is valid provided that the smaller of Ep and Eq is at least equal to 10; a numerical appreciation of this criterion will now be provided. Glancing at recent (ultimate) mortality rates it is noticed that, very roughly, they approximate, at the ages shown in the first column, to the values given in the second column of the following table:

Central age	Mortality rate	Exposed to risk required for 10 expected deaths
30	·0025	4000
45	·005	2000
55	·01	1000
65	·03	333
70	·05	200
75	·10	100

In Table 3 the exact values of the expression (2) have been calculated for consecutive values of d and for the values of q and E specified in the preceding table; these values are headed "Binomial". The values appearing in the columns headed "Normal" correspond to the expression (3). In every case the largest error is found at $d=5$ and never exceeds ·0042, whilst the largest relative errors occur at the tails but may be very considerable. With each increase in Eq the agreement at the tails improves, as is shown by Table 4 where the "Binomial" and "Normal" comparisons have been made for $q=·0025$ and three different values of E.

In the test of a mortality table graduation just described it is the relative error in the approximation to the binomial which is of importance, for the probabilities at each age are multiplied rather than added together to produce a final criterion. At first sight this rules out the normal approximation except for very large values of E. However it will be noticed that the relative errors only become considerable for values of d exceeding 9, that the probability of obtaining such a deviation is about ·005, and hence that, in a mortality table graduation of about 80 ages, such values will be required only rarely.

It is of interest to attempt further approximations to the expression (2). One such is provided by the employment of the

Table 3. "Normal" and Type III approximations to binomial probabilities

d	q=·0025 E=4000			q=·005 E=2000			q=·01 E=1000			Poisson	d
	Binomial	Normal	Type III	Binomial	Normal	Type III	Binomial	Normal	Type III		
1	·8747	·8742	·8744	·8746	·8741	·8743	·8743	·8701	·8740	·8749	1
2	·6356	·6348	·6350	·6352	·6344	·6346	·6343	·6336	·6337	·6360	2
3	·4281	·4286	·4276	·4275	·4280	·4270	·4264	·4269	·4259	·4287	3
4	·2651	·2678	·2653	·2645	·2672	·2647	·2633	·2660	·2635	·2657	4
5	·1500	·1542	·1510	·1495	·1537	·1506	·1486	·1527	·1496	·1505	5
6	·0776	·0816	·0791	·0773	·0812	·0788	·0766	·0805	·0780	·0780	6
7	·0371	·0396	·0386	·0369	·0393	·0384	·0365	·0388	·0379	·0374	7
8	·0169	·0176	·0180	·0168	·0174	·0179	·0165	·0171	·0176	·0170	8
9	·0076	·0071	·0083	·0075	·0070	·0082	·0074	·0069	·0080	·0077	9
10	·0035	·0026	·0038	·0034	·0026	·0038	·0033	·0025	·0037	·0035	10
11	·0016	·0009	·0017	·0015	·0009	·0017	·0015	·0008	·0015	·0016	11
12	·0007	·0003	·0008	·0007	·0003	·0008	·0007	·0003	·0007	·0007	12
13	·0003	·0001	·0003	·0003	·0001	·0003	·0003	·0001	·0003	·0003	13
14	·0001	·0000	·0001	·0001	·0000	·0001	·0001	·0000	·0001	·0001	14
15	·0000		·0000	·0000		·0000	·0000		·0000	·0000	15

d	q=·03 E=333			q=·05 E=200			q=·1 E=100			Poisson	d
	Binomial	Normal	Type III	Binomial	Normal	Type III	Binomial	Normal	Type III		
1	·8730	·8724	·8726	·8716	·8711	·8713	·8681	·8676	·8678	·8749	1
2	·6308	·6299	·6300	·6272	·6265	·6266	·6178	·6171	·6172	·6360	2
3	·4216	·4219	·4210	·4168	·4173	·4164	·4042	·4047	·4038	·4287	3
4	·2584	·2609	·2585	·2536	·2562	·2539	·2410	·2434	·2414	·2657	4
5	·1444	·1483	·1454	·1405	·1443	·1416	·1301	·1336	·1314	·1505	5
6	·0735	·0773	·0751	·0708	·0744	·0724	·0636	·0668	·0652	·0780	6
7	·0345	·0368	·0360	·0328	·0350	·0343	·0284	·0303	·0299	·0374	7
8	·0154	·0160	·0165	·0144	·0150	·0155	·0119	·0124	·0130	·0170	8
9	·0067	·0063	·0074	·0062	·0058	·0068	·0049	·0046	·0055	·0077	9
10	·0030	·0023	·0033	·0027	·0021	·0030	·0020	·0015	·0023	·0035	10
11	·0013	·0007	·0015	·0012	·0007	·0013	·0008	·0005	·0009	·0016	11
12	·0006	·0002	·0006	·0005	·0002	·0006	·0003	·0001	·0004	·0007	12
13	·0002	·0001	·0003	·0002	·0001	·0002	·0001	·0000	·0002	·0003	13
14	·0001	·0000	·0001	·0001	·0000	·0001	·0000		·0001	·0001	14
15	·0000		·0000	·0000		·0000			·0000	·0000	15

Note. The binomial values were calculated by the method suggested by Montessus de Ballore in *Mémorial* no. 10 *de l'Office National Météorologique de France* (1925).
The normal values were obtained from Sheppard's tables reproduced as Table II of Pearson's *Tables for Statisticians and Biometricians*, Part I.
The Type III probabilities were taken from Salvosa's *Tables of Pearson's Type III Function* (reprinted from *Annals of Mathematical Statistics*, May 1930).
The Poisson values were extracted from Table LI of Pearson's *Tables for Statisticians and Biometricians*, Part I.

Table 4. "Normal" and "Poisson" approximations to binomial probabilities ($q=\cdot 0025$)

d	E=2000			E=4000			E=6000			d
	Binomial	Normal	Poisson	Binomial	Normal	Poisson	Binomial	Normal	Poisson	
1	·8243	·8228	·8245	·8747	·8742	·8749	·8974	·8972	·8976	1
2	·5023	·5018	·5028	·6356	·6348	·6360	·6987	·6982	·6991	2
3	·2574	·2630	·2580	·4281	·4286	·4287	·5182	·5181	·5188	3
4	·1081	·1171	·1085	·2651	·2678	·2657	·3647	·3656	·3653	4
5	·0383	·0439	·0386	·1500	·1542	·1505	·2427	·2447	·2432	5
6	·0136	·0138	·0137	·0776	·0816	·0780	·1523	·1551	·1528	6
7	·0054	·0036	·0055	·0371	·0396	·0374	·0902	·0929	·0906	7
8	·0020	·0008	·0020	·0169	·0176	·0170	·0505	·0525	·0507	8
9	·0007	·0001	·0007	·0076	·0071	·0077	·0269	·0280	·0271	9
10	·0002	·0000	·0002	·0035	·0026	·0035	·0138	·0141	·0140	10
11	·0001		·0001	·0016	·0009	·0016	·0070	·0066	·0070	11
12	·0000		·0000	·0007	·0003	·0007	·0035	·0029	·0035	12
13				·0003	·0001	·0003	·0017	·0012	·0018	13
14				·0001	·0000	·0001	·0009	·0005	·0009	14
15				·0000		·0000	·0004	·0002	·0004	15
16							·0002	·0001	·0002	16
17							·0001	·0000	·0001	17

asymptotic value for $\binom{E}{r} p^{E-r}q^r$ as $E\to\infty$ and $q\to 0$ in such a way that Eq remains constant (at 10 in the case under consideration). This leads to the replacement of (2) by

$$\sum_{r=Eq+d}^{\infty} \frac{e^{-Eq}\,(Eq)^r}{r!} + \sum_{r=0}^{Eq-d} \frac{e^{-Eq}\,(Eq)^r}{r!}, \qquad \ldots\ldots(4)$$

numerical values of which appear in Tables 3 and 4 under the heading "Poisson". A further approximation is obtained by fitting a Type III curve to the binomial by means of the method of moments (see Appendix I), viz.

$$s+1=\frac{4Epq}{1-4pq} \quad \text{and} \quad \gamma^2=\frac{4}{1-4pq}.$$

It is seen from Table 3 that the Type III approximation is a definite improvement on the Normal, particularly at the tails, but suffers from the disadvantage that a table of double entry must be used for the evaluation of the probabilities. The Poisson approximation is very good when q is as low as ·0025 or ·005 but then deteriorates until it is much worse than the Normal.

It is suggested, then, that the normal approximation to (2) may be used provided Eq is at least equal to 10 at every age of the mortality table tested and that $(|\theta - \mathrm{E}q| - \frac{1}{2})/\sqrt{(\mathrm{E}pq)}$ does not exceed 3 in any particular case. If the latter requirement is not satisfied the Type III approximation may be used for the two or three ages affected.

In Table 2 the $\mathrm{P_Q}$ test is applied to Kenchington's 27-term summation graduation of the $\mathrm{O^{JF}}$ Table. The values of

$$(d - \tfrac{1}{2})/\sqrt{(\mathrm{E}pq)}$$

are tabulated in column (8) and in the next column appear the cologarithms of the expression (3), the normal approximation to the true binomial probabilities, obtained from Table B which was calculated from Sheppard's sixteen-figure table of $\log_e \frac{1}{2}(1-\alpha_x)$ soon to be published by the British Association. The central second difference is entered against each value of $\mathrm{colog}_{10}(1-\alpha_x)$ appearing in Table B and interpolation is quite rapid if Everett's formula is used in conjunction with the coefficients published by Comrie in *Interpolation and Allied Tables* (1936). On summing column (9) of Table 2, changing the base of the logarithms from 10 to e, and doubling the result, the value 76·63 is obtained. Reference to Table A under $f = 2 \times 50 = 100$ indicates that the probability of obtaining, from an ideal mortality table with the $\mathrm{O^{JF}}$ values of q_x, a set of data as bad or worse than that actually observed is just more than ·95. The graduation is thus to be considered slightly unsatisfactory.

The application of the $\mathrm{P_Q}$ test could be much simplified by the calculation of a more extended table of colog $(1-\alpha_x)$ corresponding to an argument of x^2 instead of x. Owing, however, to the straightforward test about to be described such calculations have not been made.

The consideration that each value of $(\theta - \mathrm{E}q)/\sqrt{(\mathrm{E}pq)}$ is, apart from questions of continuity, approximately a standardized normal variate, i.e. has a mean value of zero, a standard deviation of unity, and is distributed normally (see Appendix I), suggests that it may be possible to construct a test of the graduation based on the various values of this function. Each of these values is, by hypothesis, a random sampling variate from an approximately normal universe of mean zero and unit standard deviation and hence the ensemble of values may be tested for agreement with this hypothesis in any of the usual ways [Geary and Pearson (1938)]. If there is a reasonably large number of ages under consideration, say 80, the observed values of $(\theta - Eq)/\sqrt{(Epq)}$ may be grouped to form a frequency distribution and this may then be compared with the "expected" distribution, the normal curve. As an illustration of the method the 50 values of $(\theta - Eq)^2/Epq$ produced by Kenchington's graduation of the O^{JF} Table have been grouped into six categories as specified below and compared with the corresponding "normal" frequency. A superficial inspection of this comparison

Values of $(\theta - Eq)^2/Epq$	Corresponding values of $\frac{\theta - Eq}{\sqrt{(Epq)}} = z$	$\frac{1}{2}(1-\alpha_z)$ (Table B)	"Normal" frequency T	Observed frequency O	$\frac{(O-T)^2}{T}$
∞–2·25	(−∞) −(−1·5)		3·34	3	·035
2·25 – ·5625	(−1·5) −(− ·75)		7·99	8	·000
·5625– ·0	(− ·75)− 0	·50000	13·67	15	·129
·0 – ·5625	0 – ·75	·22663	13·67	13	·033
·5625–2·25	·75 – 1·5	·06681	7·99	9	·128
2·25 –∞	1·5 –∞	·00000	3·34	2	·538
			50·00	50	$\chi^2 = ·863$

indicates a very close agreement of theory and observation and this is confirmed by the application of the χ^2 test [see e.g. Caradog Jones (1924), pp. 211–12, or Elderton (1938), Ch. XI].* As P_{χ^2} lies between ·95 and ·99 the "fit" is rather "too good" and there is a suggestion that the original O^{JF} graduation adhered too closely to the crude values of q_x. Owing to the relatively few groups necessarily adopted in this test—it is a fundamental rule for the application of the χ^2 test that every group in the theoretical distribution must possess several units of frequency—a certain degree of insensitivity is introduced.

The collection of the values of $(\theta - Eq)/\sqrt{(Epq)}$ into groups and subsequent comparison with a normal distribution with zero mean and unit variance is equivalent to the assumption that the ensemble of values is a sample from a normal universe and that, subject to

* This χ^2 test, although analogous, is not exactly equivalent to that hereafter described.

the errors of random sampling, it will possess moments of the first, second, third, etc. orders all agreeing with the "normal" moments. In view of the fact that at any one age

$$(\theta - \mathrm{E}q)/\sqrt{(\mathrm{E}pq)}$$

is only approximately distributed normally, it would seem reasonable to seek a less stringent test for the set of values. Such a test might consist in the discovery whether the observed series could be considered as having come from a universe of zero mean and unit variance. By the very nature of almost every graduation, $\sum\limits_x (\theta_x - \mathrm{E}_x q_x) \doteq 0$, so that the mean of the values under consideration is likely to be fairly close to zero.

The variance of the observed series of values of $(\theta - \mathrm{E}q)/\sqrt{(\mathrm{E}pq)}$, calculated about a mean of zero instead of about the observed mean which approximates to zero [cp. Neyman and Pearson (1933), Example (2), p. 304] is

$$\sum_x \frac{(\theta_x - \mathrm{E}_x q_x)^2}{\mathrm{E}_x p_x q_x} = \chi_0^2 \text{ say.}$$

If each value of $(\theta - \mathrm{E}q)/\sqrt{(\mathrm{E}pq)}$ were a standardized normal variate this sum would be distributed as mentioned in Appendix I, pp. 39–40, and the probability of obtaining a larger variance than χ_0^2, the one observed, can be calculated. [This test is identically the same as that proposed by Elderton on p. 209 of his book (1938).] Once again Table A proves useful as it provides the values of the above sum corresponding to certain critical values of the probability of obtaining values larger than those specified. The theorem of the Appendix clearly indicates the value of f to use when entering the Table. As an example suppose Perks's formula

$$q_x = \frac{\mathrm{A} + \mathrm{B}c^x}{1 + \mathrm{D}c^x}$$

had been fitted by means of the following four equations to determine A, B, D and c:

$$\Sigma x^r \theta_x + \mathrm{D}\Sigma x^r \theta_x c^x = \mathrm{A}\Sigma x^r \mathrm{E}_x + \mathrm{B}\Sigma x^r \mathrm{E}_x c^x \quad (r = 0, 1, 2, 3).$$

(These are, of course, equivalent to the usual "successive summations") The above equations might have been written

$$\Sigma x^r [(1 + \mathrm{D}c^x)\,\theta_x - (\mathrm{A} + \mathrm{B}c^x)\,\mathrm{E}_x] = 0 \quad (r = 0, 1, 2, 3),$$

i.e. $$\Sigma x^r (1+Dc^x)(\theta_x - E_x q_x) = 0,$$

or $$\Sigma x^r (1+Dc^x)\sqrt{(E_x p_x q_x)}\frac{\theta_x - E_x q_x}{\sqrt{(E_x p_x q_x)}} = 0 \quad (r=0, 1, 2, 3),$$

or $$\Sigma a_{xr}\frac{\theta_x - E_x q_x}{\sqrt{(E_x p_x q_x)}} = 0 \quad (r=0, 1, 2, 3),$$

and in this latter form it is plain that the l of the Appendix must here be replaced by 4.

This deduction from k allows for a slight modification in the original hypothesis introduced by the fact of "graduation". The universe specified by the hypothesis is no longer completely *a priori* (as would be the case were the probabilities determined by some law, e.g. Mendel's law in the case of the phenomena of segregation) but agrees with the sample in certain respects, namely, agreement of total expected and actual deaths, etc. Theory can only deal with the case where the universe is linked to the sample by linear relations; other types of restriction have yet to be investigated. No adjustment to the P_Q test to allow for graduation restrictions on the hypothesis tested has been discovered up to the present and to this extent the previous test is at a disadvantage in comparison with the χ^2 test.

In order to determine empirically the loss of freedom in χ^2 due to graduation by Spencer's 21-term summation formula, a random sampling experiment was carried out as follows. First a simple series of values of q_x was assumed. In view of the labour of the subsequent sampling, the linear series

$$q_x = x/100 \qquad (1 \leqslant x \leqslant 50)$$

was taken. Then the corresponding values of E_x were fixed by the simple relations

$$E_x = \begin{cases} 100 & (1 \leqslant x \leqslant 19 \text{ and } 30 \leqslant x \leqslant 50) \\ 200 & (20 \leqslant x \leqslant 29). \end{cases}$$

The actual sampling to obtain the deaths was performed as follows. Consider, for example, age 15; according to hypothesis the probability of death is ·15 so it is necessary to construct an infinite universe of individuals of whom ·15 are marked "dead" and then proceed to draw 100 of these individuals, the exposed to risk, noting those that are "dead". Tables of Random Sampling

Numbers, Tippett (1927) or Fisher and Yates (1938), enable this experiment to be carried out. Considering 100 random sampling numbers of two digits the true probability of any particular one of these numbers being less than, or equal to, 14 (i.e. being 00, 01, ..., 14) is ·15. Hence the number of numbers in this sample of 100 which are less than, or equal to, 14 represents the number of deaths at age 15. A similar argument is used for the other ages; for ages 20 to 29 inclusive the number in the sample is 200 instead of 100.

By this means the sample values of θ shown in Table 5 were obtained for nine "mortality" tables. As an illustration of the theory the corresponding rates of mortality were then graduated by the formula

$$q_x = a + bx,$$

in the usual manner. The resulting rates are shown in Table 6 under the heading "Straight Line". Finally, the rates were graduated by means of Spencer's 21-term summation formula and are shown under the heading "Spencer".*

The values of χ^2 calculated from the true universe of values have, by the theorem of the Appendix, 30 degrees of freedom whilst those obtained from the straight line graduation have 28 degrees of freedom. The sample values of χ^2 (Table 7 shows the full calculations for sample number V) are given below and indicate, on the average, a reasonable agreement of theory and observation. (The application of the t test to the difference between $30{\cdot}53 - 28{\cdot}97 = 1{\cdot}56$ and 2 shows no significance.) The

Sample number	"Universe" values of χ^2	"Straight line" values of χ^2	"Spencer" values of χ^2
I	23·45	22·38	19·51
II	20·63	19·08	11·39
III	39·19	35·25	35·53
IV	33·94	31·73	26·95
V	30·50	29·12	22·97
VI	30·58	29·03	23·18
VII	29·65	29·60	28·69
VIII	28·18	26·85	25·68
IX	38·67	37·70	36·02
Totals	274·79	260·74	229·92
Mean	30·53	28·97	25·55

* "Ages" 1–10 and 41–50 were only used in the Spencer graduations.

Table 5. Sampling experiment

Age	True q	E	Sample values of θ								
			I	II	III	IV	V	VI	VII	VIII	IX
1	·01	100	0	1	1	0	1	0	1	0	0
2	·02	100	0	1	0	0	4	3	0	0	2
3	·03	100	0	5	6	3	3	3	2	2	3
4	·04	100	5	8	2	4	3	5	3	3	5
5	·05	100	6	10	5	6	7	6	3	6	7
6	·06	100	6	6	7	5	9	9	12	8	3
7	·07	100	6	11	4	11	3	9	9	7	6
8	·08	100	6	8	5	8	11	10	11	13	4
9	·09	100	9	4	11	13	3	5	9	13	8
10	·10	100	11	14	5	11	11	13	15	12	5
11	·11	100	7	15	15	15	6	14	10	12	9
12	·12	100	13	14	12	12	15	9	7	7	17
13	·13	100	18	15	14	14	14	6	8	11	8
14	·14	100	13	16	21	13	14	15	21	19	14
15	·15	100	17	18	14	20	12	9	13	14	23
16	·16	100	12	20	15	16	21	14	10	10	18
17	·17	100	16	15	15	15	18	16	22	17	11
18	·18	100	18	19	19	15	24	9	19	21	22
19	·19	100	22	11	27	23	22	23	11	17	23
20	·20	200	53	36	37	40	45	41	37	36	40
21	·21	200	42	40	42	46	44	40	49	49	32
22	·22	200	53	44	61	46	41	48	49	40	46
23	·23	200	44	45	43	43	43	53	48	42	50
24	·24	200	45	47	38	48	45	46	48	44	45
25	·25	200	47	51	61	49	57	38	45	49	62
26	·26	200	54	49	51	45	48	57	62	49	39
27	·27	200	46	61	60	53	55	58	55	55	58
28	·28	200	65	55	50	66	49	56	50	51	54
29	·29	200	55	60	51	69	41	55	57	58	65
30	·30	100	30	39	38	25	23	26	27	31	25
31	·31	100	33	39	32	36	30	28	34	31	31
32	·32	100	30	36	33	48	34	33	28	23	36
33	·33	100	33	31	35	28	31	33	36	31	35
34	·34	100	39	34	38	31	41	36	35	29	41
35	·35	100	38	38	33	34	27	35	31	34	31
36	·36	100	31	35	36	40	35	37	37	38	43
37	·37	100	40	33	27	43	35	33	36	50	42
38	·38	100	42	38	45	39	44	33	36	36	32
39	·39	100	39	37	38	39	37	29	32	35	38
40	·40	100	35	38	42	32	39	39	41	33	38
41	·41	100	40	36	39	36	44	34	38	36	38
42	·42	100	37	50	46	40	44	36	43	42	45
43	·43	100	46	39	41	49	38	41	47	42	50
44	·44	100	39	40	40	37	42	56	47	47	49
45	·45	100	36	37	57	42	43	44	42	49	42
46	·46	100	46	53	48	49	57	51	46	51	40
47	·47	100	47	44	44	56	49	44	54	45	42
48	·48	100	53	45	43	51	39	43	47	43	56
49	·49	100	50	53	46	55	39	47	47	40	47
50	·50	100	44	48	52	56	49	45	53	58	55

Table 6. Graduation of samples

True q	Sample number: I Obs. q	St. line q	Spencer q	II Obs. q	St. line q	Spencer q	III Obs. q	St. line q	Spencer q	IV Obs. q	St. line q	Spencer q
·11	·07	·1185	·1096	·15	·1213	·1290	·15	·1278	·1134	·15	·1185	·1288
·12	·13	·1283	·1203	·14	·1308	·1418	·12	·1371	·1277	·12	·1285	·1363
·13	·18	·1380	·1307	·15	·1403	·1530	·14	·1464	·1403	·14	·1385	·1428
·14	·13	·1478	·1412	·16	·1499	·1611	·21	·1558	·1519	·13	·1485	·1495
·15	·17	·1575	·1524	·18	·1594	·1658	·14	·1651	·1626	·20	·1584	·1563
·16	·12	·1673	·1647	·20	·1690	·1677	·15	·1744	·1732	·16	·1684	·1646
·17	·16	·1770	·1780	·15	·1785	·1687	·15	·1838	·1836	·15	·1784	·1740
·18	·18	·1868	·1918	·19	·1881	·1703	·19	·1931	·1945	·15	·1884	·1840
·19	·22	·1966	·2052	·11	·1976	·1745	·27	·2024	·2058	·23	·1984	·1937
·20	·265	·2063	·2170	·18	·2071	·1815	·185	·2118	·2170	·20	·2083	·2033
·21	·21	·2161	·2265	·20	·2167	·1914	·21	·2211	·2267	·23	·2183	·2121
·22	·265	·2258	·2333	·22	·2262	·2038	·305	·2304	·2351	·23	·2283	·2202
·23	·22	·2356	·2387	·225	·2358	·2182	·215	·2398	·2425	·215	·2383	·2278
·24	·225	·2453	·2433	·235	·2453	·2343	·19	·2491	·2492	·24	·2483	·2369
·25	·235	·2551	·2486	·255	·2549	·2517	·305	·2584	·2560	·245	·2583	·2482
·26	·27	·2648	·2558	·245	·2644	·2699	·255	·2677	·2644	·225	·2682	·2622
·27	·23	·2746	·2655	·305	·2740	·2887	·30	·2771	·2753	·265	·2782	·2783
·28	·325	·2843	·2768	·275	·2835	·3071	·25	·2864	·2880	·33	·2882	·2953
·29	·275	·2941	·2895	·30	·2930	·3237	·255	·2957	·3014	·345	·2982	·3120
·30	·30	·3038	·3026	·39	·3026	·3371	·38	·3051	·3139	·25	·3082	·3268
·31	·33	·3136	·3157	·39	·3121	·3470	·32	·3144	·3252	·36	·3181	·3388
·32	·30	·3233	·3286	·36	·3217	·3525	·33	·3237	·3341	·48	·3281	·3484
·33	·33	·3331	·3408	·31	·3312	·3543	·35	·3331	·3409	·28	·3381	·3565
·34	·39	·3428	·3519	·34	·3408	·3537	·38	·3424	·3460	·31	·3481	·3627
·35	·38	·3526	·3622	·38	·3503	·3534	·33	·3517	·3515	·34	·3581	·3674
·36	·31	·3623	·3711	·35	·3598	·3548	·36	·3611	·3573	·40	·3681	·3710
·37	·40	·3721	·3775	·33	·3694	·3589	·27	·3704	·3651	·43	·3780	·3738
·38	·42	·3818	·3820	·38	·3789	·3654	·45	·3797	·3748	·39	·3880	·3761
·39	·39	·3916	·3848	·37	·3885	·3743	·38	·3891	·3875	·39	·3980	·3783
·40	·35	·4013	·3871	·38	·3980	·3840	·42	·3984	·4019	·32	·4080	·3816

Sample number: V			VI			VII			VIII			IX		
Obs. q	St. line q	Spencer q	Obs. q	St. line q	Spencer q	Obs. q	St. line q	Spencer q	Obs. q	St. line q	Spencer q	Obs. q	St. line q	Spencer q
·06	·1187	·1001	·14	·1049	·1018	·10	·1086	·1123	·12	·1038	·1181	·09	·1133	·0973
·15	·1278	·1135	·09	·1146	·1039	·07	·1184	·1167	·07	·1135	·1217	·17	·1234	·1142
·14	·1368	·1296	·06	·1242	·1067	·08	·1283	·1223	·11	·1233	·1260	·08	·1335	·1317
·14	·1458	·1467	·15	·1338	·1112	·21	·1381	·1293	·19	·1331	·1323	·14	·1436	·1478
·12	·1549	·1643	·09	·1435	·1191	·13	·1479	·1386	·14	·1428	·1410	·23	·1536	·1616
·21	·1639	·1802	·14	·1531	·1305	·10	·1577	·1489	·10	·1526	·1515	·18	·1637	·1732
·18	·1729	·1941	·16	·1627	·1451	·22	·1675	·1599	·17	·1624	·1627	·11	·1738	·1826
·24	·1820	·2049	·09	·1724	·1618	·19	·1773	·1717	·21	·1722	·1742	·22	·1839	·1900
·22	·1910	·2133	·23	·1820	·1790	·11	·1871	·1846	·17	·1819	·1851	·23	·1940	·1969
·225	·2001	·2196	·205	·1917	·1959	·185	·1970	·1978	·18	·1917	·1950	·20	·2041	·2044
·22	·2091	·2247	·20	·2013	·2114	·245	·2068	·2112	·245	·2015	·2038	·16	·2141	·2125
·205	·2181	·2288	·24	·2109	·2243	·245	·2166	·2245	·20	·2212	·2125	·23	·2242	·2215
·215	·2272	·2324	·265	·2206	·2349	·24	·2264	·2369	·21	·2210	·2214	·25	·2343	·2306
·225	·2362	·2352	·23	·2302	·2438	·24	·2362	·2477	·22	·2308	·2311	·225	·2444	·2400
·285	·2452	·2378	·19	·2398	·2512	·225	·2460	·2563	·245	·2406	·2411	·31	·2545	·2493
·24	·2543	·2408	·285	·2495	·2571	·31	·2559	·2634	·245	·2503	·2513	·195	·2646	·2587
·275	·2633	·2450	·29	·2591	·2634	·275	·2657	·2702	·275	·2601	·2604	·29	·2746	·2680
·245	·2724	·2509	·28	·2688	·2710	·25	·2755	·2774	·255	·2699	·2681	·27	·2847	·2785
·205	·2814	·2597	·275	·2784	·2806	·285	·2853	·2855	·29	·2796	·2742	·325	·2948	·2910
·23	·2904	·2709	·26	·2880	·2918	·27	·2951	·2950	·31	·2894	·2806	·25	·3049	·3055
·30	·2995	·2848	·28	·2977	·3040	·34	·3049	·3055	·31	·2992	·2886	·31	·3150	·3211
·34	·3085	·3002	·33	·3073	·3160	·28	·3148	·3160	·23	·3090	·2998	·36	·3251	·3371
·31	·3175	·3164	·33	·3169	·3270	·36	·3246	·3255	·31	·3187	·3136	·35	·3351	·3514
·41	·3266	·3319	·36	·3266	·3346	·35	·3344	·3335	·29	·3285	·3297	·41	·3452	·3626
·27	·3356	·3468	·35	·3362	·3385	·31	·3442	·3403	·34	·3383	·3452	·31	·3553	·3702
·35	·3447	·3594	·37	·3459	·3391	·37	·3540	·3466	·38	·3480	·3584	·43	·3654	·3749
·35	·3537	·3706	·33	·3555	·3388	·36	·3638	·3534	·50	·3578	·3679	·42	·3755	·3791
·44	·3627	·3806	·33	·3651	·3398	·36	·3737	·3619	·36	·3676	·3751	·32	·3856	·3844
·37	·3718	·3913	·29	·3748	·3459	·32	·3835	·3731	·35	·3774	·3811	·38	·3956	·3917
·39	·3808	·4019	·39	·3844	·3580	·41	·3933	·3869	·33	·3871	·3881	·38	·4057	·4011

Table 7. Calculation of χ^2 for sample V

Age	E	θ	"True"				"Straight line"				"Spencer"			
			q	Eq	$\theta-$Eq	$\frac{(\theta-\text{E}q)^2}{\text{E}pq}$	q	Eq	$\theta-$Eq	$\frac{(\theta-\text{E}q)^2}{\text{E}pq}$	q	Eq	$\theta-$Eq	$\frac{(\theta-\text{E}q)^2}{\text{E}pq}$
11	100	6	·11	11	− 5	2·554	·1187	11·87	− 5·87	3·294	·1001	10·01	− 4·01	1·785
12	100	15	·12	12	3	·852	·1278	12·78	2·22	·442	·1135	11·35	3·65	1·324
13	100	14	·13	13	1	·088	·1368	13·68	·32	·009	1296	12·96	1·04	·096
14	100	14	·14	14	0	·000	·1458	14·58	− ·58	·027	·1467	14·67	− ·67	·036
15	100	12	·15	15	− 3	·706	·1549	15·49	− 3·49	·930	·1643	16·43	− 4·43	1·429
16	100	21	·16	16	5	1·860	·1639	16·39	4·61	1·551	·1802	18·02	2·98	·601
17	100	18	·17	17	1	·071	·1729	17·29	·71	·035	·1941	19·41	− 1·41	·127
18	100	24	·18	18	6	2·439	·1820	18·20	5·80	2·259	·2049	20·49	3·51	·756
19	100	22	·19	19	3	·585	·1910	19·10	2·90	·544	·2133	21·33	·67	·027
20	200	45	·20	40	5	·781	·2001	40·02	4·98	·775	·2196	43·92	1·08	·034
21	200	44	·21	42	2	·121	·2091	41·82	2·18	·144	·2247	44·94	− ·94	·025
22	200	41	·22	44	− 3	·262	·2181	43·62	− 2·62	·201	·2288	45·76	− 4·76	·642
23	200	43	·23	46	− 3	·254	·2272	45·44	− 2·44	·170	·2324	46·48	− 3·48	·339
24	200	45	·24	48	− 3	·247	·2362	47·24	− 2·24	·139	·2352	47·04	− 2·04	·116
25	200	57	·25	50	7	1·307	·2452	49·04	7·96	1·712	·2378	47·56	9·44	2·458
26	200	48	·26	52	− 4	·416	·2543	50·86	− 2·86	·216	·2408	48·16	− ·16	·001
27	200	55	·27	54	1	·025	·2633	52·66	2·34	·141	·2450	49·00	6·00	·973
28	200	49	·28	56	− 7	1·215	·2724	54·48	− 5·48	·758	·2509	50·18	− 1·18	·037
29	200	41	·29	58	−17	7·018	·2814	56·28	−15·28	5·773	·2597	51·94	−10·94	3·113
30	100	23	·30	30	− 7	2·333	·2904	29·04	− 6·04	1·770	·2709	27·09	− 4·09	·847
31	100	30	·31	31	− 1	·047	·2995	29·95	·05	·000	·2848	28·48	1·52	·113
32	100	34	·32	32	2	·184	·3085	30·85	3·15	·465	·3002	30·02	3·98	·754
33	100	31	·33	33	− 2	·181	·3175	31·75	− ·75	·026	·3164	31·64	− ·64	·019
34	100	41	·34	34	7	2·184	·3266	32·66	8·34	3·163	·3319	33·19	7·81	2·751
35	100	27	·35	35	− 8	2·813	·3356	33·56	− 6·56	1·930	·3468	34·68	− 7·68	2·604
36	100	35	·36	36	− 1	·043	·3447	34·47	·53	·012	·3594	35·94	− ·94	·038
37	100	35	·37	37	− 2	·172	·3537	35·37	− ·37	·006	·3706	37·06	− 2·06	·182
38	100	44	·38	38	6	1·528	·3627	36·27	7·73	2·586	·3806	38·06	5·94	1·497
39	100	37	·39	39	− 2	·168	·3718	37·18	− ·18	·001	·3913	39·13	− 2·13	·190
40	100	39	·40	40	− 1	·042	·3808	38·08	·92	·036	·4019	40·19	− 1·19	·059
Totals	4000	990		1010	−20	30·496		990·02	− ·02	29·115		995·13	− 5·13	22·973

corresponding values of χ^2 for the Spencer graduation suggest a deduction of about 5 degrees of freedom. However, not only is this estimate uncertain owing to random sampling fluctuations—the actual deduction might well lie between 2·44 and 7·24 so variable is an estimate based only on nine sample values—but the very process of deduction of degrees of freedom may be illegitimate founded as it is on an analogy with a rather different type of graduation restraint.

Analogy once more suggests that the deduction from k applicable to Kenchington's 27-term formula would be about 6. Hence the value of χ^2 obtained from the O^{JF} graduation, namely 41·425 (Table 2), corresponds to a value of $f=50-6=44$ and Table A indicates that the graduation is excellent.

Apart from the deduction of degrees of freedom to allow for the "fitting" of a graduation, the accuracy of the substitution of a standardized binomial variate by a normal variate and consequent replacement of a discrete series by a continuous curve may be verified by the comparison of the true moments of the discrete distribution of $\Sigma\,[(\theta_x-\mathrm{E}_x q_x)^2/\mathrm{E}_x p_x q_x]$ with the moments of the continuous χ^2 distribution with $f=k$. The former [see Haldane (1937)] are given by

$$\mathrm{Mean}=k,$$

$$\mathrm{Variance}=2k+\Sigma\,\frac{1-6p_xq_x}{\mathrm{E}_xp_xq_x},$$

$$\mu_3=8k+2\Sigma\,\frac{11-56p_xq_x}{\mathrm{E}_xp_xq_x}+\Sigma\,\frac{1-30p_xq_x+120p_x^2q_x^2}{\mathrm{E}_x^2p_x^2q_x^2},$$

whereas the moments of the χ^2 distribution (see Appendix I) are

$$\mathrm{Mean}=k,$$

$$\mathrm{Variance}=2k,$$

$$\mu_3=8k.$$

The closeness of the approximation thus depends upon the value of the additional terms in the moments of the true distribution. The latter have been calculated for the O^{JF} data and are

$$\mathrm{Mean}=50,$$

$$\mathrm{Variance}=100+2{\cdot}1760=102{\cdot}1760,$$

$$\mu_3=400+48{\cdot}9892+{\cdot}0716=449{\cdot}0608,$$

$$\beta_1={\cdot}18904.$$

The two Type III distributions corresponding to these sets of moments have been calculated with 17 groups and are compared below. The last column is the generally accepted approximation on which Table A is based when $f > 30$. The use of the χ^2 distribution in this case would seem to lead to very little inaccuracy. With

Values of $\Sigma \frac{(\theta - \mathrm{E}q)^2}{\mathrm{E}pq}$	Probability of a value between limits specified		
	Type III with true moments	χ^2 distribution	R. A. Fisher's normal approx.
Up to 20	·0000	·0001	·0001
20– 25	·0010	·0011	·0019
25– 30	·0100	·0100	·0118
30– 35	·0429	·0420	·0429
35– 40	·1055	·1036	·1006
40– 45	·1704	·1694	·1644
45– 50	·1991	·2004	·1983
50– 55	·1799	·1824	·1848
55– 60	·1320	·1338	·1377
60– 65	·0814	·0819	·0843
65– 70	·0435	·0430	·0434
70– 75	·0205	·0198	·0191
75– 80	·0087	·0081	·0073
80– 85	·0034	·0030	·0025
85– 90	·0013	·0010	·0007
90– 95	·0004	·0003	·0002
95–100	·0001	·0001	·0000

the increase in the values of Eq the approximation improves until, with the values of the exposed as large as those of the A 1924–29 Table, it is "perfect".

SEQUENCES OF DEVIATIONS OF LIKE SIGN

Both P_Q and χ^2 tests treat the set of values of $(\theta - \mathrm{E}q)/\sqrt{(\mathrm{E}pq)}$ as unordered; that is, they make no distinction between a graduation in which all the positive deviations precede all the negative deviations and one in which the positive and negative deviations alternate fairly regularly. In the absence of a test which takes into account a consideration of the runs of positive and negative values of $\theta - \mathrm{E}q$, it would be difficult to decide categorically that some desirable graduation formula systematically distorts the facts.

The test devised, apparently, by Makeham (*J.I.A.* Vol. XXVIII,

p. 316) and used by Steffensen (1912) consists of a comparison of the expected number of runs of r similar signs among the k deviations, viz.

$$\sum_{s=1}^{k-r-1} (\tfrac{1}{2})^{r+2} + 2 \times (\tfrac{1}{2})^r = \frac{k-r+3}{2^{r+1}},$$

with the actually observed numbers of such runs, for values of $r=1, 2, 3, \ldots, k$. This procedure is rather insensitive owing to the difficulty of judging the significance of differences between the fractional expected numbers of sequences of high order and the frequencies actually observed.

In order to test for systematic deviations it would appear sufficient to consider the number of groups of positive or negative deviations—these numbers may differ by a unit at most. The method of graduation will have ensured that the numbers of positive and negative signs appearing in the column of $\theta - \mathrm{E}q$ are approximately equal and hence it is reasonable to regard the number, m, of positive signs as ancillary information. On this basis Stevens (1939) has developed an appropriate test which may be described as follows.

Suppose the m positive signs are distributed into t groups. Stevens's test is equivalent to that deduced under the following circumstances. An office has its mortality observations at one age divided into two categories (e.g. "male" and "female") containing m and $k-m$ exposed to risk respectively. In the course of a year, t and $k-m+1-t$ deaths, respectively, have been observed in these categories and it is desired to amalgamate the experience in the two classes. The χ^2 test already described provides an immediate answer to the question, is such an amalgamation justified? If the observations are homogeneous the total deaths, $k-m+1$, divided by the total exposed, k, is an estimate of the rate of mortality applicable to the two classes. The "expected deaths" are calculated on this basis for each of the categories and thence the appropriate value of χ^2 which, since the total of the expected and actual deaths agree, has only one degree of freedom.

As an example of the application of this test Lever's Makeham graduation of his Female Life Tenants data (*J.I.A.* Vol. LIII, p. 1) is taken. Considering ages 41 to 102 inclusive, $k=62$, $m=28$ and $t=17$, hence

E	θ	Eq	$\theta-$Eq	$\frac{(\theta-\mathrm{E}q)^2}{\mathrm{E}pq}$
28 34	17 18	15·806 19·194	1·194 $-$1·194	·207 ·171
62	35	35·000		$\chi^2=$·378
$q=$·564516				$f=1$

Contrary to certain opinions which have been expressed concerning the systematic distortion caused by this graduation, no fault may be found in the alternations of positive and negative deviations.

LEGITIMACY OF APPLICATION

The applicability of the P_Q and χ^2 tests in the form proposed depends essentially upon the satisfaction of the two fundamental assumptions mentioned on p. 7, viz. (i) that each person lives and dies independently of any other person, and (ii) that each person observed is as likely to die as any other person of the same age, or, in other words, there is a probability q_x that any person aged exactly x will die during the course of a year. The theory of Life Contingencies is implicitly based upon these assumptions which have been widely discussed on the continent since G. Bohlmann (1901) expressly formulated them. As recently as 1937 B. de Finetti (11th *T.I.C.A.* Vol. II, p. 435) has discussed these fundamental hypotheses which are the basis of the Theory of Risk.

It is obvious that a certain degree of interdependence between deaths is a fairly general phenomenon. Monozygotic twins have a notorious tendency to die at about the same age, husband and wife often follow each other to the grave in close succession, and, more intermittently but of greater importance, epidemics and pandemics are the cause of a close relationship between the deaths during a given epoch. It would seem that no attempt has been made to measure numerically this lack of independence between deaths but, in the ordinary course of events, it must be of small consequence. A mortality investigation is seldom limited to one year and seldom includes both husband and wife so that many of the forces that tend towards the causation of dependent rates of mortality are missing. But there has been a deliberate introduction of "dependence" into recent mortality investigations—the presence of duplicates. If every insured person took out two policies and a mortality investigation were made, based upon "policies", the exposed to risk and deaths would be automatically doubled and every death would have a further death as an invariable concomitant. Actually some people only take out one policy on their lives whilst others may effect five or six. This means that not only are policy-deaths linked in some way but that this "linkage" varies, perhaps considerably, in the different classes of policy and between the various ages.

The general effect of the inclusion of duplicates in a mortality experience is to increase the apparent variance in the number of deaths at any age as calculated by the usual "binomial" formula, Epq. In Appendix II an attempt has been made to estimate this increase when the total of duplicates is 40%. The conclusion to be drawn is that the presence of duplicates in a mortality experience may more than double the variance at any age but that a simple investigation would enable accurate adjustments to be made to the probabilistic tests already described.

The question of the constancy of q_x for all individuals appearing in the data at age x is more elusive. Intuitively, arguing on a Laplacean definition of probability, it is obvious that all persons aged x do not have the same probability of dying within a year; some of the exposed to risk will be on their death-beds as they enter on age x, others will be enjoying robust health, so of necessity the group of individuals appearing in E_x is heterogeneous in just those qualities which affect the rate of mortality. Adopting the "frequency" or "collective" definition of probability [see Mises (1939), in particular pp. 21–4] the appropriate collective reduces in this case to one individual and probability theory would seem to be ruled out altogether. Suppose, however, that certain broad categories of individuals who would appear to be liable to the same risk of death during a year can be distinguished among the E_x people exposed to risk at age x. If these group rates of mortality differ it can be shown [cp. e.g. Castelnuovo (1933), Ch. X, or Rietz (1927), Ch. VI] that the variance in the expected number of deaths is no longer Epq but is less than this quantity by a figure which depends on the differences between the group probabilities. Fortunately the distribution of deaths at any age about the mean Eq still tends towards the normal with the increase in E and hence the preceding tests only require alteration in respect of the factor Epq.

If this hypothetical "grouping" of the exposed to risk corresponds to reality—and this would certainly seem to be the case with data not distinguishing e.g. sex, class of policy, occupation, duration since entry and, maybe, a number of other variables—it becomes of importance to discover whether probabilistic tests are invalidated altogether.*

It is a simple matter in any given case to determine whether or not there are present groups of individuals whose mortality differs significantly from the main body of the exposed. Two such examples are given here. The first relates to the A 1924–9 (ultimate after three years) data and the second to C. F. Warren's Normal Pensioners (ultimate after one year) data (*J.I.A.* Vol. LIX, p. 221).

Table 8 shows a threefold test for the homogeneity of the data included in the values of q_x (A 1924–9 ultimate) at ages $36\frac{1}{2}$ and $46\frac{1}{2}$. Each class of policy is first tested for homogeneity between the six calendar years of the investigation (medical and non-medical sections treated separately). Only those classes providing sufficient data for a reliable test have been entered in the table. The value of q applicable to each class irrespective of year of exposure is obtained by dividing the deaths of that class by the corresponding exposed to risk; the expected deaths obtained for each of the years 1924–9 on the basis of this rate of mortality are then compared with the deaths actually observed and the χ^2 test is applied in the manner already described. The number of deaths "expected" having been arbitrarily made to agree with the total of deaths for the class in question, the number of degrees of

* J. H. Peek (1899), F. Esscher (1920) and A. Lange (1932) have carried out researches to determine if the variance of the deviations exhibited by secular mortality about a trend line is approximately "binomial". In addition to the fact that these investigations relate to population data it is difficult to see just how they may be utilized to form a judgment concerning the fundamental question whether observed rates of mortality possess variances determined by the binomial law of mortality. [See Cramèr (1927) and Riebesell (1933) for what seems to be a contrary opinion.] However Blaschke (1906) quotes some results from calculations made on the H^M (ultimate after 5 years) data which suggest that, in those days, neither class nor duration had any effect on mortality rates but I have been unable to check the example he gives on p. 142, *loc. cit.* The columns "Lebende" and "Tote" do not appear to be taken from *The Mortality Experience of Life Assurance Companies collected by the Institute of Actuaries* (1869) at all!

Table 8. Tests of homogeneity

Year	Medically examined lives															Lives not medically examined					Year
	Life with profits					Endowment assurance with profits					Endowment assurance without profits					Endowment assurance with profits					
	$E_{36\frac{1}{2}}$	$\theta_{36\frac{1}{2}}$	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	$E_{36\frac{1}{2}}$	$\theta_{36\frac{1}{2}}$	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	$E_{36\frac{1}{2}}$	$\theta_{36\frac{1}{2}}$	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	$E_{36\frac{1}{2}}$	$\theta_{36\frac{1}{2}}$	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	
1924	2480¾	4	6·63	−2·63	1·046	24648½	85	78·76	6·24	·496	2677	12	6·72	5·28	4·159	3515¾	9	10·99	−1·99	·361	1924
1925	2387	10	6·38	3·62	2·059	24262¼	82	77·53	4·47	·259	3072½	10	7·71	2·29	·682	3771	16	11·79	4·21	1·508	1925
1926	2249	3	6·01	−3·01	1·512	23562¾	82	75·29	6·71	·600	3236½	7	8·13	−1·13	·157	3961¾	16	12·38	3·62	1·062	1926
1927	2305	3	6·16	−3·16	1·625	24020½	76	76·76	−·76	·008	3446½	8	8·65	−·65	·049	4420	12	13·82	−1·82	·240	1927
1928	2200	4	5·88	−1·88	·603	23956½	57	76·55	−19·55	5·009	3350	6	8·41	−2·41	·692	5112½	16	15·98	·02	·000	1928
1929	2229½	13	5·95	7·05	8·376	24443¾	81	78·11	2·89	·107	3331¾	5	8·37	−3·37	1·360	5773¾	14	18·05	−4·05	·912	1929
Total	13851	37	37·01		15·221	144894¼	463	463·00		6·479	19114½	48	47·99		7·099	26554½	83	83·01		4·083	Total
	$q=·002671$				$f=5$	$q=·0031954$				$f=5$	$q=·002511$				$f=5$	$q=·0031256$				$f=5$	

Year	Medically examined lives																				Lives not medically examined										Year
	L.P.					L.N.P.					E.A.P.					E.A.N.P.					E.A.P.					E.A.N.P.					
	$E_{46\frac{1}{2}}$	$\theta_{46\frac{1}{2}}$	Eq	$(\theta - Eq)$	$\frac{(\theta - Eq)^2}{Epq}$	$E_{46\frac{1}{2}}$	$\theta_{46\frac{1}{2}}$	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	$E_{46\frac{1}{2}}$	$\theta_{46\frac{1}{2}}$	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	$E_{46\frac{1}{2}}$	$\theta_{46\frac{1}{2}}$	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	$E_{46\frac{1}{2}}$	$\theta_{46\frac{1}{2}}$	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	$E_{46\frac{1}{2}}$	$\theta_{46\frac{1}{2}}$	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	
1924	[illegible]312½	42	37·38	4·62	·575	1181½	9	6·29	2·71	1·174	25309¾	146	142·95	3·05	·065	3076	11	17·18	−6·18	2·235	2628¼	11	15·39	−4·39	1·260	1568¾	12	7·80	4·20	2·273	1924
1925	[illegible]098½	29	36·59	−7·59	1·585	1256½	6	6·69	−·69	·072	24958¾	113	140·97	−27·97	5·581	3504½	25	19·57	5·43	1·515	2886½	16	16·90	−·90	·048	1589¾	7	7·90	−·90	·103	1925
1926	[illegible]933¼	29	34·76	−5·76	·961	1350½	2	7·19	−5·19	3·766	24978	123	141·08	−18·08	2·330	3827½	17	21·38	−4·38	·902	3217¾	19	18·84	·16	·001	1547	7	7·69	−·69	·062	1926
1927	[illegible]335	39	36·15	2·85	·226	1604	9	8·54	·46	·025	26493¼	164	149·64	14·36	1·386	4258½	28	23·78	4·22	·753	3625½	18	21·23	−3·23	·494	1370¼	6	6·81	−·81	·097	1927
1928	[illegible]972¼	35	34·34	·66	·013	1605½	10	8·54	1·46	·251	26356½	155	148·86	6·14	·255	4146	14	23·16	−9·16	3·643	4206	28	24·63	3·37	·464	1078	3	5·36	−2·36	1·044	1928
1929	[illegible]046¾	38	32·78	5·22	·837	1646½	10	8·76	1·24	·177	26290¼	171	148·49	22·51	3·432	4106½	33	22·93	10·07	4·447	4784½	33	28·01	4·99	·894	892	5	4·43	·57	·074	1929
Total	3[illegible]98	212	212·00		4·197	8644	46	46·01		5·465	154387	872	871·99		13·049	22918½	128	128·00		13·495	21349	125	125·00		3·161	8045¾	40	39·99		3·653	Total
	$q=·0069060$				$f=5$	$q=0053216$				$f=5$	$q=·0056481$				$f=5$	$q=·0055850$				$f=5$	$q=·0058551$				$f=5$	$q=·0049716$				$f=5$	

Type of Assurance	Age 36½										Age 46½															Type of Assurance
	E.A.P.					E.A.N.P.					L.P.					E.A.P.					E.A.N.P.					
	E	θ	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	E	θ	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	E	θ	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	E	θ	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	E	θ	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	
Med.	144894¼	463	461·43	1·57	·005	19114½	48	52·20	−4·20	·339	30698	212	207·04	4·96	·120	154387	872	875·88	−3·88	·017	22918½	128	124·35	3·65	·108	Med.
Non. Med.	26554½	83	84·56	−1·56	·029	6516½	22	17·80	4·20	·994	2070¼	9	13·96	−4·96	1·774	21349	125	121·12	3·88	·125	8045¾	40	43·65	−3·65	·307	Non Med.
Total	171448¾	546	545·99		·034	25630¾	70	70·00		1·333	32768¼	221	221·00		1·894	175736	997	997·00		·142	30964¼	168	168·00		·415	Total
	$q=·0031846$				$f=1$	$q=·0027311$				$f=1$	$q=·0067443$				$f=1$	$q=·0056733$				$f=1$	$q=·0054256$				$f=1$	

Class of policy	Age 36½					Age 41½					Age 46½					Age 51½					Class of policy
	E	θ	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	E	θ	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	E	θ	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	E	θ	Eq	$\theta - Eq$	$\frac{(\theta - Eq)^2}{Epq}$	
L.P.	15117¼	38	46·07	−8·07	1·418	23070	125	103·67	21·33	4·408	32768¼	221	189·01	31·99	5·446	41444¼	404	347·86	56·14	9·137	L.P.
L.N.P.	5665¼	10	17·27	−7·27	3·070	7867	57	35·35	21·65	13·319	9307¾	49	53·69	−4·69	·412	9814¾	87	82·38	4·62	·261	L.N.P.
E.A.P.	171448¾	546	522·54	23·46	1·056	182696¾	769	821·00	−52·00	3·308	175736	997	1013·68	−16·68	·276	137610¼	1124	1155·03	−31·03	·841	E.A.P.
E.A.N.P.	25630¾	70	78·12	−8·12	·847	29147¼	140	130·98	9·02	·624	30964¼	168	178·61	−10·61	·634	27607½	202	231·73	−29·73	3·847	E.A.N.P.
Total	217862	664	664·00		6·391	242781	1091	1091·00		21·659	248776¼	1435	1434·99		6·768	216476¾	1817	1817·00		14·086	Total
	$q=·0030478$				$f=3$	$q=·0044938$				$f=3$	$q=·0057682$				$f=3$	$q=·0083935$				$f=3$	

Data extracted from *Continuous Investigation into the Mortality of Assured Lives, Statistics* 1924–1929 (Cambridge, 1933).

freedom is not 6 but 5. On the whole, i.e. adding the 10 different values of χ^2 and the corresponding degrees of freedom (see Appendix I), "time" is not a factor which introduces any significant degree of heterogeneity. In three classes, however, there is evidence of variations other than those ascribable to chance, but in each case they seem to be caused rather by the exceptional year 1929 than by any noticeable "trend" in the mortality rates.

The second part of the test relates to comparisons between "medical" and "non-medical" mortality and here, at the ages and classes in question, there appears to be no significant difference between the mortalities observed, neither are the differences uniformly in favour of one or the other forms of assurance. Finally there is the comparison between the four classes of policy included in the investigation. At ages $36\frac{1}{2}$ and $46\frac{1}{2}$ there did not appear to be any significant difference between the mortalities of these classes and this was such a surprising result that two other ages, $41\frac{1}{2}$ and $51\frac{1}{2}$, were taken and here the differences were strongly significant. In the interpretation of these figures it must not be forgotten that all the values of χ^2 obtained are over-estimated owing to the presence of duplicate policies.

Assuming that at age $41\frac{1}{2}$ the 242781 exposed to risk can be subdivided into the three groups of $182696\frac{3}{4}$, $52217\frac{1}{4}$, and 7867 exposed to risk, respectively, with "true" rates of mortality ·004, ·005 and ·007, constant throughout each group, the variance (apart from consideration of duplicates) becomes (cp. Rietz, *loc. cit.* p. 149)

$$242781 \times \cdot 00431 \times \cdot 99569 - \{182696\tfrac{3}{4} \times (\cdot 004 - \cdot 00431)^2$$
$$+ 52217\tfrac{1}{4} \times (\cdot 005 - \cdot 00431)^2 + 7867 \times (\cdot 007 - \cdot 00431)^2\} = 1041\cdot 8$$

instead of the binomial variance of 1041·9. The reduction is negligible and would nearly always appear to be so in the case of mortality statistics. This means that heterogeneity is not a question at issue in the testing of a graduation.

Lastly, Table 9 tests the propriety of the amalgamation of the Banks and Insurance Offices data effected by C. F. Warren in Table D of his paper. The agreement is satisfactory and, in particular, the apparent discrepancy at ages 70–74 may well be due to chance fluctuations and not worthy of the misgivings felt by Warren (p. 224, *loc. cit.*).

These tests of homogeneity are applicable generally. They may be used for the discovery of duration of selection, the determination whether "year of entry" has any effect on mortality and many other factors.

Table 9. Comparison of Banks and Insurance Offices (ultimate), normal pensioners' mortality

q	Age	E	θ	Eq	$\theta -$ Eq	$\frac{(\theta - \mathrm{E}q)^2}{\mathrm{E}pq}$
	60–64	2825	83	82·59	41	·002
		1143	33	33·41	− ·41	·005
·029234		3968	116	116·00		
	65–69	6852	251	248·28	2·72	·031
		1731	60	62·72	− 2·72	·122
·036234		8583	311	311·00		
	70–74	5239	279	295·43	−16·43	·968
		1411	96	79·57	16·43	3·595
·056391		6650	375	375·00		
	75–79	2909	224	222·93	1·07	·006
		823	62	63·07	− 1·07	·020
·076634		3732	286	286·00		
	80–84	1269	144	146·48	− 2·48	·047
		377	46	43·52	2·48	·160
·115431		1646	190	190·00		
	85–89	372	71	70·86	·14	·000
		132	25	25·14	− ·14	·001
·19048		504	96	96·00		
	90–94	71	15	15·06	− ·06	·000
		28	6	5·94	·06	·001
·21212		99	21	21·00		
	95–	13	5	3·37	1·63	1·065
		14	2	3·63	− 1·63	·988
·25926		27	7	7·00		
						$\chi^2 = 7{\cdot}011$ $f = 8$

CONCLUSIONS

The preceding considerations suggest that a satisfactory statistical test of a graduation may be effected provided duplicates are not present or are dealt with by the methods outlined in Appendix II. Computing convenience would give preference to the χ^2 test rather than to the P_Q test which has, however, been mentioned for

Table 10. Test of Warren's graphic graduation of normal pensioners' (ultimate) rates

x	(1) q_x	(2) E_x	(3) $E_x q_x$	(4) θ_x	(5) $\theta_x - E_x q_x$	(6) $(\theta_x - E_x q_x)^2$	(7) $E_x p_x q_x$	(8) (6) ÷ (7)	x
61	·0256	673	17·23	22	4·77	22·753	16·79	1·355	61
62	·0269	897	24·13	21	− 3·13	9·797	23·48	·417	62
63	·0284	1067	30·30	36	5·70	32·490	29·44	1·104	63
64	·0300	1159	34·77	37	2·23	4·973	33·73	·147	64
65	·0318	1309	41·63	42	·37	·137	40·31	·003	65
66	·0339	1895	64·24	57	− 7·24	52·418	62·06	·845	66
67	·0364	1892	68·87	60	− 8·87	78·677	66·36	1·186	67
68	·0393	1794	70·50	73	2·50	6·250	67·73	·092	68
69	·0427	1693	72·29	79	6·71	45·024	69·20	·651	69
70	·0466	1571	73·21	94	20·79	432·224	69·80	6·192	70
71	·0509	1448	73·70	66	− 7·70	59·290	69·95	·848	71
72	·0553	1344	74·32	75	·68	·462	70·21	·007	72
73	·0598	1207	72·18	63	− 9·18	84·272	67·86	1·242	73
74	·0645	1080	69·66	77	7·34	53·876	65·17	·827	74
75	·0695	951	66·09	50	−16·09	258·888	61·50	4·210	75
76	·0748	843	63·06	64	·94	·884	58·34	·015	76
77	·0805	735	59·17	57	− 2·17	4·709	54·41	·087	77
78	·0867	639	55·40	48	− 7·40	54·760	50·60	1·082	78
79	·0936	564	52·79	67	14·21	201·924	47·85	4·220	79
80	·1014	465	47·15	55	7·85	61·622	42·37	1·454	80
81	·1101	385	42·39	35	− 7·39	54·612	37·72	1·448	81
82	·1197	329	39·38	41	1·62	2·624	34·67	·076	82
83	·1301	256	33·31	26	− 7·31	53·436	28·98	1·844	83
84	·1413	211	29·81	33	3·19	10·176	25·60	·398	84
85	·1533	158	24·22	28	3·78	14·288	20·51	·697	85
86	·1661	123	20·43	20	− ·43	·185	17·04	·011	86
87	·1797	93	16·71	15	− 1·71	2·924	13·71	·213	87
88	·1942	73	14·18	13	− 1·18	1·392	11·43	·122	88
89	·2096	57	11·95	20	8·05	64·802	9·45	6·857	89
90	·2259	33	7·45	7	− ·45	·202	5·77	·035	90
			1370·52	1381	+90·73 −80·25			37·685 $=\chi^2$	

$f = 30$
$\cdot 05 < P < \cdot 5$

Mr J. N. Shine kindly provided me with the necessary values of E_x.

3-2

completeness and because its theoretical basis may appear more satisfactory to some actuaries. In Table 10 a complete test of Warren's graphical graduation of his combined (ultimate) data has been presented in a form suitable for general adoption; the value of χ^2 obtained indicates a good fit.

If practical considerations cause theoretical aims to assume a second place, the formulation of some definite practical criterion for rejection or acceptance of a graduation is of fundamental importance. In the absence of comparisons having a justifiable theoretical basis the very *raison d'être* of Thiele's arguments in favour of the calculation of expected deaths is in question and it would seem that the function to be graduated should be the function of greatest practical importance—the annuity value calculated at a representative rate of interest.

APPENDIX I

Résumé of Statistical Theory required in the preceding paper

Any non-negative function $p(x)$, defined for certain specified values of x, is said to be the probability law of the random variable x if

$$\int_a^b p(x)\,dx \quad \text{or} \quad \sum_{x=a}^{b} p(x),$$

according as $p(x)$ is continuous or discrete respectively, is equal to the probability that x lies in the interval $[a, b]$. This latter probability is written Pr. $\{a \leqslant x \leqslant b\}$ for short.

It follows from this definition that if a and b, respectively, are replaced by the smallest and largest values assumed by x, the integral or sum is equal to unity. A random variable x will be spoken of as being distributed according to the law $p(x)$.

Any probability law is characterized by certain constants defined as follows:

$$\text{Mean} = \xi = \int_{-\infty}^{\infty} xp(x)\,dx \quad \text{or} \quad \sum_{-\infty}^{\infty} xp(x);$$

$$\text{Variance} = \sigma^2 = \mu_2 = \int_{-\infty}^{\infty} (x-\xi)^2 \, p(x)\, dx \quad \text{or} \quad \sum_{-\infty}^{\infty} (x-\xi)^2 \, p(x)$$
$$= \int_{-\infty}^{\infty} x^2 \, p(x)\, dx - \xi^2 \quad \text{or} \quad \sum_{-\infty}^{\infty} x^2 \, p(x) - \xi^2;$$

*n*th moment

$$\text{about} \quad \text{mean} = \mu_n = \int_{-\infty}^{\infty} (x-\xi)^n \, p(x)\, dx \quad \text{or} \quad \sum_{-\infty}^{\infty} (x-\xi)^n \, p(x);$$
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

Here, $-\infty$ and ∞ have been inserted as the lower and upper limits of x, respectively, since $p(x)$ may be defined arbitrarily to be zero outside the actual lowest and highest values.

The values of some of these constants in the case of certain common probability laws are given below. It is a common practice to equate these theoretical values to "sample" moments obtained directly from observations and thence obtain estimates of the parameters of the universe from which the sample may be supposed to derive.

The binomial law.

$$p(x) = \binom{n}{x} p^{n-x} q^x, \quad p+q=1 \quad (0 \leqslant x \leqslant n),$$
$$\xi = nq,$$
$$\mu_2 = \sigma^2 = npq,$$
$$\mu_3 = npq\,(p-q),$$
$$\mu_4 = npq\,[3\,(n-2)\,pq+1]$$

[cp. Hardy (1909), Note A, p. 107],

$$\beta_1 = \frac{1-4pq}{npq} \quad \text{and} \quad \beta_2 = 3 + \frac{1-6pq}{npq}.$$

The normal law.

$$p(x) = \frac{1}{\sqrt{(2\pi)}\,\sigma} e^{-\frac{1}{2}(x-\xi)^2/\sigma^2} \qquad (-\infty < x < \infty).$$

The constants ξ and σ^2, respectively, are the mean and variance of the distribution.

$$\mu_3 = 0 \qquad \mu_4 = 3\sigma^4,$$
$$\beta_1 = 0 \quad \text{and} \quad \beta_2 = 3,$$

[Elderton (1938), p. 80].

The Type III law.

$$p(x)=\frac{\gamma^{s+1}}{\Gamma(s+1)}\,x^s e^{-\gamma x} \qquad (0 \leqslant x < \infty),$$

where
$$\Gamma(s+1)=\int_0^\infty t^s e^{-t}\,dt = s\Gamma(s),$$

$$\xi=(s+1)/\gamma,$$
$$\mu_2=(s+1)/\gamma^2,$$
$$\mu_3=2(s+1)/\gamma^3,$$
$$\beta_1=4/(s+1),$$

[Elderton (1938), p. 92].

In the particular case where $s=\frac{1}{2}f-1$ and $\gamma=\frac{1}{2}$ the distribution becomes

$$p(x)=\frac{1}{2^{\frac{1}{2}f}\,\Gamma(\frac{1}{2}f)}\,x^{\frac{1}{2}f-1}\,e^{-\frac{1}{2}x},$$

$$\xi=f,$$
$$\sigma^2=2f,$$
$$\mu_3=8f,$$
$$\beta_1=8/f.$$

When $f>30$ this last distribution may be fairly well represented by a normal curve in which the variable is not x but $\sqrt{(2x)}$ and the mean and variance are $\sqrt{(2f-1)}$ and unity respectively. Table A is constructed on this basis.

Just as the mean and variance completely specify a normal curve so do the mean, variance and skewness (measured by $\sqrt{\beta_1}$) specify any Type III curve.

A random variable, i.e. any variable which assumes values in accordance with some probability law, may be "standardized" by choosing its mean as origin and the square root of its variance (standard deviation) as unit. Hence the standardized variable corresponding to x is $(x-\xi)/\sigma=z$, say.

In particular, if a variable is distributed normally the probability law of its standardized form is

$$p(z)=\frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}z^2}$$

and $\quad \text{Pr.}\{-\infty < z \leqslant z_0\}=\dfrac{1}{\sqrt{(2\pi)}}\displaystyle\int_{-\infty}^{z_0} e^{-\frac{1}{2}z^2}\,dz=\tfrac{1}{2}(1+a_{z_0})$ say.

This last function is tabulated in Sheppard's Tables reproduced as Table II of Pearson (1914).

The application to "ideal" mortality observations is as follows. The number of deaths at age x, θ_x, is a random variable with a binomial probability law $(p_x+q_x)^{E_x}$ and, as mentioned on p. 11, *ante*,

$$\Pr.\{|\theta_x - E_x q_x| \geqslant d\} = \sum_{r=Eq+d}^{E} \binom{E}{r} p^{E-r} q^r + \sum_{r=0}^{Eq-d} \binom{E}{r} p^{E-r} q^r,$$

where the suffixes have been suppressed on the right. It has been shown in *J.I.A.S.S.* Vol. 1, no. 2, p. 44, that

$$\binom{E}{r} p^{E-r} q^r \sim \frac{1}{\sqrt{(2\pi E p q)}} e^{-(r-Eq)^2/2Epq},$$

but it is more convenient and more accurate to replace the right-hand side of this relation by

$$\frac{1}{\sqrt{(2\pi)}} \int_{(r-Eq-\frac{1}{2})/\sqrt{(Epq)}}^{(r-Eq+\frac{1}{2})/\sqrt{(Epq)}} e^{-\frac{1}{2}t^2}\, dt,$$

since, in this case,

$$\begin{aligned}\sum_{r=Eq+d}^{E} \binom{E}{r} p^{E-r} q^r + \sum_{r=0}^{Eq-d} \binom{E}{r} p^{E-r} q^r &= \frac{1}{\sqrt{(2\pi)}} \int_{(d-\frac{1}{2})/\sqrt{(Epq)}}^{\infty} e^{-\frac{1}{2}t^2}\, dt \\ &\quad + \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{-(d-\frac{1}{2})/\sqrt{(Epq)}} e^{-\frac{1}{2}t^2}\, dt \\ &= 2 \times \frac{1}{\sqrt{(2\pi)}} \int_{z}^{\infty} e^{-\frac{1}{2}t^2}\, dt,\end{aligned}$$

where $z=(d-\frac{1}{2})/\sqrt{(Epq)}$

$$\begin{aligned}&= 2\left[1 - \tfrac{1}{2}(1+a_z)\right] \\ &= 1 - a_z.\end{aligned}$$

The cologarithm of this function is tabulated in Table B.

Among the useful properties of the normal law perhaps the most prolific in its implications is provided by the following theorem due to R. A. Fisher (1922) which is stated here without proof.

Theorem.

The probability law of the random variable, χ^2, composed of the sum of the squares of k standardized normal variates x_i ($i=1, 2, 3, \ldots, k$), which are subject to l linear connecting relations

$$\sum_{i=1}^{k} a_{ij} x_i = 0 \quad (j=1, 2, \ldots, l),$$

is given by

$$p(\chi^2)=\frac{1}{2^{\frac{1}{2}f}\,\Gamma(\frac{1}{2}f)}(\chi^2)^{\frac{1}{2}f-1}\,e^{-\frac{1}{2}\chi^2},$$

where $f=k-l$ and is called the number of degrees of freedom. Hence

$$\text{Pr.}\,\{\chi^2\geqslant\chi_0^2\}=\frac{1}{2^{\frac{1}{2}f}\,\Gamma(\frac{1}{2}f)}\int_{\chi_0^2}^{\infty}x^{\frac{1}{2}f-1}\,e^{-\frac{1}{2}x}\,dx.$$

Corollary.

As an immediate corollary of this theorem it follows that if χ_1^2 be the sum of the squares of a number of standardized normal variates with f_1 degrees of freedom, χ_2^2 be a similar sum with f_2 degrees of freedom and so on, then $\chi^2=\chi_1^2+\chi_2^2+\chi_3^2+\ldots$ is a variable distributed as χ^2 with $f=f_1+f_2+f_3+\ldots\ldots$

APPENDIX II

Effect of Duplicate Policies on Binomial Variance

Let n_r $(r=1, 2, 3, \ldots)$ be the number of assured at age x who possess r policies included in E_x, then $\sum_{r=1}^{\infty} rn_r=E_x$. Write h_r $(r=1, 2, 3, \ldots)$ for the number of deaths observed at age x amongst the n_r policyholders, then $\sum_{r=1}^{\infty} rh_r=\theta_x$.

Now, on the usual assumptions of "independence" and "homogeneity", h_r is a random variable distributed according to the binomial law $(p_x+q_x)^{n_r}$ and hence possesses a variance of $n_r pq$, dropping the "age" suffixes. Therefore, since the variance of a random variable which is the sum of a number of independent random variables is equal to the sum of the separate variances [see Caradog Jones (1924), p. 158], the variance of θ_x is

$$\sum_{r=1}^{\infty} r^2 n_r pq=pq\sum_{r=1}^{\infty} r^2 n_r,$$

and this is necessarily in excess of Epq.*

* G. J. Lidstone points out that G. F. Hardy anticipated this result in *J.I.A.* Vol. XXIII, pp. 2–3. Hardy is there considering the ratio of the s.d. of q when duplicates are present to the s.d. of q *based on the same number of lives* with no duplicates.

Unfortunately I have no figures relating to the distribution of duplicates at any age but Sir William P. Elderton suggested to me that the numbers $n_1, n_2, n_3, \ldots$ might, at least approximately, form a geometric progression. On this assumption $n_r = na^{r-1}$ say,

$$E_x = \sum_{r=1}^{\infty} rna^{r-1} = n \sum_{r=1}^{\infty} ra^{r-1}$$

$$= n/(1-a)^2,$$

and the variance of θ_x is

$$pq \sum_{r=1}^{\infty} r^2 na^{r-1} = npq \frac{1+a}{(1-a)^3}.$$

As a "guess" the proportion of duplicates in the recent mortality investigation has been estimated at 40% (*J.I.A.* Vol. LXVIII, p. 62) and thus, on the assumption that there is no significance in the differences observed in this proportion from age to age,

$$\frac{2}{5} = \frac{\text{number of duplicates}}{\text{number of policies}} = \frac{E_x - \sum\limits_{r=1}^{\infty} n_r}{E_x} = 1 - \frac{\sum\limits_{r=1}^{\infty} na^{r-1}}{E_x}$$

$$= 1 - n/E_x\,(1-a) = 1-(1-a) = a,$$

$$n = (1-a)^2\, E_x = \tfrac{9}{25}\, E_x.$$

On this basis the variance of θ_x is

$$\tfrac{9}{25} E_x p_x q_x \frac{7/5}{(3/5)^3} = 2\tfrac{1}{3} E_x p_x q_x.$$

Before these tentative conclusions can be applied with confidence it would be necessary to investigate numerically the differences observable in the proportion of duplicates from age to age—these differences could be tested for "significance" by the χ^2 method—and to compare the actual distribution of duplicates, triplicates, etc. at any age with the geometric series suggested by Elderton.*

* The geometric series based on the above figures allows 3 policies per ten thousand of the exposed to risk to have been effected on the lives of persons with 12 or more policies in force. This would appear to be reasonable.

APPENDIX III

Approximations to the ordinates of the binomial distribution

Since the foregoing was circulated to members, Mr G. J. Lidstone has suggested that Table 3, showing the Normal, Poisson and Type III approximations to the sum of a number of terms of the binomial probability law, might be amplified to give a comparison of these approximations with the binomial expansion term by term. Although not strictly related to the proposed methods of testing a graduation, such a comparison may be of interest and was easily derived from the original calculations made to obtain Table 3.

In Table 3(a) the argument is θ, the number of deaths, instead of $d = |\,\theta - \mathrm{E}q\,|$. The first column of each group of three provides the exact probability of observing the number of deaths stated, viz. $\binom{\mathrm{E}}{\theta} p^{\mathrm{E}-\theta} q^{\theta}$, whilst the second and third columns give the Normal and Type III approximations respectively, viz.

$$\frac{1}{\sqrt{(2\pi)}}\int_{(\theta-\mathrm{E}q-\frac{1}{2})/\sqrt{(\mathrm{E}pq)}}^{(\theta-\mathrm{E}q+\frac{1}{2})/\sqrt{(\mathrm{E}pq)}} e^{-t^2/2}\,dt$$

and

$$\frac{2\,(4/\beta_1)^{4/\beta_1-1}}{\sqrt{\beta_1}\,e^{4/\beta_1}\,\Gamma(4/\beta_1)}\int_{(\theta-\mathrm{E}q-\frac{1}{2})/\sqrt{(\mathrm{E}pq)}}^{(\theta-\mathrm{E}q+\frac{1}{2})/\sqrt{(\mathrm{E}pq)}} (1+\tfrac{1}{2}\sqrt{\beta_1}t)^{4/\beta_1-1}e^{-2t/\sqrt{\beta_1}}\,dt,$$

where $\beta_1 = (1-4pq)/\mathrm{E}pq$. The column headed "Poisson" corresponds to the approximation

$$\binom{\mathrm{E}}{\theta} p^{\mathrm{E}-\theta} q^{\theta} \sim \frac{e^{-\mathrm{E}q}(\mathrm{E}q)^{\theta}}{\theta!},$$

and, as $\mathrm{E}q$ is constant at 10 throughout the table, there is only one series of Poisson ordinates. It is to be observed that whilst the latter is a discrete probability law, both Normal and Type III approximations are continuous and replace the binomial ordinates by areas on unit base.

Table 3(*a*). "Normal" and Type III approximations to binomial probabilities

θ	$q=\cdot 0025$ E = 4000			$q=\cdot 005$ E = 2000			Poisson	θ
	Binomial	Normal	Type III	Binomial	Normal	Type III		
−3		·0000			·0000			−3
−2		·0001			·0001			−2
−1		·0003	·0000		·0003	·0000		−1
0	·0000	·0009	·0001	·0000	·0009	·0001	·0000	0
1	·0004	·0022	·0007	·0004	·0022	·0007	·0005	1
2	·0022	·0052	·0026	·0022	·0052	·0026	·0023	2
3	·0075	·0110	·0079	·0075	·0110	·0078	·0076	3
4	·0188	·0210	·0189	·0188	·0209	·0188	·0189	4
5	·0377	·0363	·0374	·0376	·0362	·0373	·0378	5
6	·0630	·0568	·0624	·0629	·0567	·0623	·0631	6
7	·0901	·0804	·0896	·0900	·0804	·0895	·0901	7
8	·1127	·1031	·1125	·1127	·1032	·1126	·1126	8
9	·1252	·1197	·1254	·1254	·1198	·1255	·1251	9
10	·1253	·1258	·1256	·1254	·1259	·1257	·1251	10
11	·1139	·1197	·1141	·1140	·1198	·1142	·1137	11
12	·0949	·1031	·0948	·0950	·1032	·0949	·0948	12
13	·0729	·0804	·0728	·0730	·0804	·0728	·0729	13
14	·0521	·0568	·0519	·0521	·0567	·0519	·0521	14
15	·0347	·0363	·0345	·0346	·0362	·0345	·0347	15
16	·0216	·0210	·0216	·0216	·0209	·0216	·0217	16
17	·0127	·0110	·0127	·0127	·0110	·0127	·0128	17
18	·0071	·0052	·0071	·0070	·0052	·0071	·0071	18
19	·0037	·0022	·0038	·0037	·0022	·0038	·0037	19
20	·0018	·0009	·0019	·0018	·0009	·0019	·0019	20
21	·0009	·0003	·0009	·0009	·0003	·0009	·0009	21
22	·0004	·0001	·0004	·0004	·0001	·0004	·0004	22
23	·0002	·0000	·0002	·0002	·0000	·0002	·0002	23
24	·0001		·0001	·0001		·0001	·0001	24
25	·0000		·0000	·0000		·0000	·0000	25

θ	$q=\cdot 01$ E = 1000			$q=\cdot 03$ E = 333			Poisson	θ
	Binomial	Normal	Type III	Binomial	Normal	Type III		
−3		·0000			·0000			−3
−2		·0001			·0001			−2
−1		·0003	·0000		·0003	·0000		−1
0	·0000	·0009	·0001	·0000	·0008	·0001	·0000	0
1	·0004	·0022	·0007	·0004	·0020	·0006	·0005	1
2	·0022	·0051	·0026	·0021	·0048	·0024	·0023	2
3	·0074	·0108	·0077	·0071	·0104	·0074	·0076	3
4	·0186	·0208	·0187	·0181	·0202	·0180	·0189	4
5	·0375	·0361	·0371	·0369	·0355	·0363	·0378	5
6	·0627	·0567	·0621	·0623	·0563	·0615	·0631	6
7	·0900	·0804	·0895	·0901	·0805	·0893	·0901	7
8	·1128	·1033	·1127	·1135	·1040	·1132	·1126	8
9	·1256	·1183	·1258	·1268	·1212	·1269	·1251	9
10	·1257	·1299	·1260	·1270	·1276	·1274	·1251	10
11	·1143	·1183	·1145	·1154	·1212	·1157	·1137	11
12	·0952	·1033	·0951	·0957	·1040	·0959	·0948	12
13	·0731	·0804	·0729	·0731	·0805	·0732	·0729	13
14	·0520	·0567	·0518	·0517	·0563	·0517	·0521	14
15	·0345	·0361	·0344	·0340	·0355	·0340	·0347	15
16	·0215	·0208	·0214	·0209	·0202	·0210	·0217	16
17	·0126	·0108	·0126	·0121	·0104	·0122	·0128	17
18	·0069	·0051	·0070	·0065	·0048	·0067	·0071	18
19	·0036	·0022	·0037	·0034	·0020	·0035	·0037	19
20	·0018	·0009	·0020	·0016	·0008	·0017	·0019	20
21	·0008	·0003	·0008	·0008	·0003	·0008	·0009	21
22	·0004	·0001	·0004	·0003	·0001	·0004	·0004	22
23	·0002	·0000	·0002	·0001	·0000	·0002	·0002	23
24	·0001		·0001	·0001		·0001	·0001	24
25	·0000		·0000	·0000		·0000	·0000	25

Table 3(*a*) cont.

θ	$q=\cdot05$ E=200			$q=\cdot1$ E=100			Poisson	θ
	Binomial	Normal	Type III	Binomial	Normal	Type III		
−3		·0000						−3
−2		·0001			·0000			−2
−1		·0002	·0000		·0002	·0000		−1
0	·0000	·0007	·0001	·0000	·0005	·0001	·0000	0
1	·0004	·0019	·0006	·0003	·0015	·0005	·0005	1
2	·0019	·0046	·0023	·0016	·0039	·0020	·0023	2
3	·0067	·0100	·0071	·0059	·0089	·0063	·0076	3
4	·0174	·0197	·0175	·0159	·0182	·0160	·0189	4
5	·0359	·0348	·0356	·0339	·0334	·0336	·0378	5
6	·0614	·0559	·0608	·0596	·0549	·0590	·0631	6
7	·0896	·0806	·0890	·0889	·0807	·0883	·0901	7
8	·1137	·1046	·1135	·1148	·1062	·1146	·1126	8
9	·1277	·1223	·1279	·1304	·1253	·1306	·1251	9
10	·1284	·1289	·1287	·1319	·1324	·1322	·1251	10
11	·1167	·1223	·1169	·1199	·1253	·1201	·1137	11
12	·0967	·1046	·0967	·0988	·1062	·0987	·0948	12
13	·0736	·0806	·0734	·0743	·0807	·0741	·0729	13
14	·0518	·0559	·0515	·0513	·0549	·0511	·0521	14
15	·0338	·0348	·0336	·0327	·0334	·0326	·0347	15
16	·0206	·0197	·0205	·0193	·0182	·0193	·0217	16
17	·0117	·0100	·0118	·0106	·0089	·0107	·0128	17
18	·0063	·0046	·0064	·0054	·0039	·0055	·0071	18
19	·0032	·0019	·0033	·0026	·0015	·0027	·0037	19
20	·0015	·0007	·0016	·0011	·0005	·0013	·0019	20
21	·0007	·0002	·0008	·0005	·0002	·0006	·0009	21
22	·0003	·0001	·0003	·0002	·0000	·0002	·0004	22
23	·0001	·0000	·0001	·0001		·0001	·0002	23
24	·0000		·0001	·0000		·0001	·0001	24
25			·0000			·0000	·0000	25

The anomaly observed by Elderton in *J.I.A.S.S.* Vol. 1, No. 2, p. 47, is here repeated: the normal curve produces a non-zero probability for a negative number of deaths. But more serious than this is the poorness of the normal approximation for values of θ below the mean number of deaths and the evidence that the shape of the normal curve is quite different from that of the binomial distribution. In contrast the Type III approximation remains very close to the binomial values throughout the range of variation. It should, perhaps, be emphasized that this poor showing of the normal curve has no bearing on the tests proposed for a mortality table graduation and that, for this purpose, the normal curve theory may be applied subject only to the restrictions specifically mentioned.

Table A. Table of values of x_0 corresponding to critical values of

$$P = \frac{1}{2^{\frac{1}{2}f}\Gamma(\frac{1}{2}f)}\int_{x_0}^{\infty} x^{\frac{1}{2}f-1} e^{-\frac{1}{2}x}\,dx$$

Judgment	P	*f*							
		1	5	10	15	20	25	30	35
"Much too probable"	·999								13·61
"Too probable"	·99	·0002	·55	2·56	5·23	8·26	11·52	14·95	17·88
"Rather too probable"	·95	·004	1·14	3·94	7·26	10·85	14·61	18·49	22·19
	·5	·455	4·35	9·34	14·34	19·34	24·34	29·34	34·50
"Of doubtful improbability"	·05	3·84	11·07	18·31	25·00	31·41	37·65	43·77	49·52
"Improbable"	·01	6·64	15·09	23·21	30·58	37·57	44·31	50·89	56·53
"Very improbable"	·001	10·83	20·52	29·59	37·70	45·32	52·62	59·70	64·94
	P	40	45	50	55	60	65	70	75
"Much too probable"	·999	16·81	20·12	23·53	27·01	30·56	34·18	37·84	41·55
"Too probable"	·99	21·53	25·26	29·06	32·92	36·83	40·78	44·78	48·81
"Rather too probable"	·95	26·23	30·34	34·49	38·68	42·91	47·17	51·46	55·77
	·5	39·50	44·50	49·50	54·50	59·50	64·50	69·50	74·50
"Of doubtful improbability"	·05	55·47	61·37	67·22	73·03	78·80	84·54	90·25	95·93
"Improbable"	·01	62·88	69·15	75·35	81·49	87·58	93·63	99·63	105·60
"Very improbable"	·001	71·74	78·43	85·02	91·54	97·98	104·37	110·71	117·00
	P	80	85	90	95	100	105	110	115
"Much too probable"	·999	45·31	49·10	52·93	56·79	60·68	64·60	68·54	72·51
"Too probable"	·99	52·87	56·96	61·08	65·22	69·39	73·57	77·78	82·00
"Rather too probable"	·95	60·11	64·47	68·85	73·24	77·65	82·07	86·51	90·96
	·5	79·50	84·50	89·50	94·50	99·50	104·50	109·50	114·50
"Of doubtful improbability"	·05	101·59	107·24	112·86	118·47	124·06	129·63	135·19	140·74
"Improbable"	·01	111·54	117·45	123·33	129·19	135·02	140·84	146·63	152·41
"Very improbable"	·001	123·24	129·45	135·62	141·76	147·87	153·95	160·00	166·04
	P	120	125	130	135	140	145	150	155
"Much too probable"	·999	76·50	80·51	84·54	88·59	92·66	96·74	100·84	104·95
"Too probable"	·99	86·24	90·50	94·77	99·05	103·35	107·66	111·98	116·31
"Rather too probable"	·95	95·42	99·90	104·38	108·87	113·38	117·89	122·41	126·94
	·5	119·50	124·50	129·50	134·50	139·50	144·50	149·50	154·50
"Of doubtful improbability"	·05	146·28	151·81	157·33	162·83	168·33	173·82	179·30	184·77
"Improbable"	·01	158·17	163·91	169·64	175·36	181·06	186·75	192·43	198·10
"Very improbable"	·001	172·05	178·04	184·01	189·96	195·89	201·81	207·71	213·60
	P	160	165	170					
"Much too probable"	·999	109·08	113·22	117·38					
"Too probable"	·99	120·66	125·01	129·37					
"Rather too probable"	·95	131·47	136·02	140·57					
	·5	159·50	164·50	169·50					
"Of doubtful improbability"	·05	190·23	195·69	201·14					
"Improbable"	·01	203·76	209·40	215·04					
"Very improbable"	·001	219·47	225·33	231·17					

Table B. Table of $\operatorname{colog}_{10}(1-\alpha_x)$

x	$\operatorname{colog}_{10}(1-\alpha_x)$	δ^2	x	$\operatorname{colog}_{10}(1-\alpha_x)$	δ^2
·0	·00000	276	2·6	2·03047	398
·1	·03605	286	2·7	2·15902	399
·2	·07496	294	2·8	2·29156	400
·3	·11681	302	2·9	2·42810	403
·4	·16168	311	3·0	2·56867	403
·5	·20966	318	3·1	2·71327	406
·6	·26082	324	3·2	2·86193	405
·7	·31522	331	3·3	3·01464	408
·8	·37293	337	3·4	3·17143	409
·9	·43401	343	3·5	3·33231	409
1·0	·49852	347	3·6	3·49728	410
1·1	·56650	353	3·7	3·66635	412
1·2	·63801	357	3·8	3·83954	413
1·3	·71309	362	3·9	4·01686	412
1·4	·79179	366	4·0	4·19830	415
1·5	·87415	369	4·1	4·38389	415
1·6	·96020	372	4·2	4·57363	415
1·7	1·04997	377	4·3	4·76752	416
1·8	1·14351	379	4·4	4·96557	417
1·9	1·24084	382	4·5	5·16779	417
2·0	1·34199	384	4·6	5·37418	419
2·1	1·44698	388	4·7	5·58476	418
2·2	1·55585	389	4·8	5·79952	419
2·3	1·66861	392	4·9	6·01847	420
2·4	1·78529	393	5·0	6·24162	
2·5	1·90590	396			

REFERENCES

BERTRAND, J. (1889). *Calcul des Probabilités*. Paris: Gauthier-Villars.

BLASCHKE, E. (1906). *Vorlesungen über mathematische Statistik*. Berlin: Teubner.

BOHLMANN, G. (1900–4). *Lebensversicherungs-Mathematik*, Band I, Teil II, D4*b*, of *Enzyklopädie der mathematischen Wissenschaften*. Leipzig: Teubner.

CASTELNUOVO, G. (1933). *Calcolo delle Probabilità*, Vol. I. Bologna: Zanichelli.

CRAMÈR, H. (1927). *Sannolikhetskalylen*. Stockholm: Gjallarhornets Förlag.

CRAMÈR, H. and WOLD, H. (1935). "Mortality variations in Sweden." *Skand. Aktuar. Tidskrift*, p. 162.

ELDERTON, W. P. (1938). *Frequency Curves and Correlation*. Cambridge Univ. Press.

ESSCHER, F. (1920). "Über die Sterblichkeit in Schweden 1886–1914." *Meddel. från Lunds astronomiska Observatorium*, Ser. II, nr. 23, Lund.

FISHER, R. A. (1922). "On the interpretation of χ^2 from contingency tables and the calculation of P." *J. Roy. Statist. Soc.* Vol. LXXXV, p. 87.

FISHER, R. A. and YATES, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research*. London: Oliver and Boyd.

GEARY, R. C. and PEARSON, E. S. (1938). *Tests of Normality*. London: Biometrika Office.

HALDANE, J. B. S. (1937). "The exact value of the moments of the distribution of χ^2, etc." *Biometrika*, Vol. XXIX, p. 133.

HARDY, G. F. (1909). *The Theory of the Construction of Tables of Mortality*. London: C. and E. Layton.

JONES, D. CARADOG (1924). *A First Course in Statistics*. London: Bell.

LANGE, A. (1932). "Untersuchungen über die jährlichen Schwankungen der Schadensquotienten in der Lebensversicherung und in der Feuerversicherung." *Wirtschaft und Recht der Versicherung*, Nr. 2.

MISES, R. VON (1939). *Probability, Statistics and Truth*. London: Hodge.

NEYMAN, J. and PEARSON, E. S. (1933). "On the problem of the most efficient tests of statistical hypotheses." *Philos. Trans.* A, Vol. CCXXXI, p. 289.

PEARSON, E. S. (1938). "The probability integral transformation for testing goodness of fit and combining independent tests of significance." *Biometrika*, Vol. XXX, p. 134.

PEARSON, K. (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." *Phil. Mag.* Vol. L, p. 157.

PEARSON, K. edited by (1914). *Tables for Statisticians and Biometricians*, Part I. London: Biometrika Office.

PEEK, J. H. (1899). "Das Problem vom Risiko in der Lebensversicherung." *Zeitschrift für Versicherungsrecht und Wissenschaft*, Bd. V, p. 169.

RIEBESELL, P. (1933). "Was folgt aus dem Misesschen Wahrscheinlichkeitsbegriff für die Versicherungsmathematik?" *Blätter für Versicherungs-Mathematik*, Bd. II, p. 395.

RIETZ, H. L. (1927). *Mathematical Statistics*. Chicago, Ill.: Open Court.

STEFFENSEN, J. F. (1912). "On the fitting of Makeham's curve to mortality observations." *Proc. Int. Cong. Math.* Cambridge.

STEVENS, W. L. (1939). "Distribution of groups in a sequence of alternatives." *Ann. Eugen., Lond.*, Vol. IX, p. 10.

TIPPETT, L. H. C. (1927). *Random Sampling Numbers. Tracts for Computers*, No. XV. Camb. Univ. Press.

ABSTRACT OF THE DISCUSSION

Mr L. H. Longley-Cook, in opening the discussion, said that, as the author had observed, there was complete agreement as to the duality of purpose of the test of a graduation—smoothness, and fidelity to data. Sir George Hardy, in his lectures on the construction of tables of mortality, considered that it was easier to set up a criterion of adherence to data than a standard of smoothness. The author, however, took the opposite view, and his remarks were confined almost entirely to tests of adherence to data.

The author's fundamental assumption was that the graduation should be such that the deaths actually observed could with reasonable probability have arisen in random sampling if the actual observed exposed to risk were drawn from an infinite population liable to the graduated rates of mortality. That assumption was not new, but nevertheless it was pertinent to inquire whether it was either necessary or sufficient.

Concerning necessity, if a graduation were being made by a formula such as that proposed by Makeham in order to demonstrate that mortality followed a definite law, then the foregoing assumption was clearly valid. If, however, a graduation were made in order to provide a table for the humdrum purpose of calculating premiums, reserves, etc., the position was not so clear. In the latter case, the aim was to find a curve which was sufficiently smooth for all practical purposes and as near to the original data as possible. It might be that the only graduations which could be made would be classed "improbable" or "very improbable" by the author's test, or it might be that a curve could be found which was so near to the data as to be "too probable" or "much too probable" by the same standard. In the former case, an effort should be made to find a closer graduation; in the latter, the sufficiency of the smoothness of the graduated curve should be investigated; but, if a more satisfactory graduation could not be found, the original work should not, in his opinion, be rejected.

As regards sufficiency, the assumption provided a test of the magnitude of deviations between actual and expected deaths and of the number of changes in sign of deviations; but, as the author had stated, in almost every graduation the difference between the total actual and expected deaths should be approximately zero, and not the amount expected on the assumption of random sampling. That was an additional requirement in a graduation, and was taken into account in the author's χ^2 test. A further necessary test was the smallness and the frequency of the change of sign of the accumulated difference between the actual and the expected deaths.

There was, in his opinion, only one satisfactory method of criticizing a graduation, and that was to produce a better one; but how should it be proved that the new graduation was better? In the paper it was shown most clearly how it was possible to compare the relative fidelity to data of two graduations of the same experience. The smaller the value of χ^2, the closer was the curve to the original data. But how could a choice be made from two graduations of the same data, one smooth but not too close to the data and the other less smooth but closer to the data? He did not think that the smoother curve, if it satisfied the χ^2 test, must necessarily be accepted.

He would go further than the author had done on p. 8 and say that because on one hypothesis the probability of a certain set of observations was ·2, and on another hypothesis was ·3, it did not necessarily follow that the latter hypothesis was the more probable.

The point was illustrated by an example, in a lecture on Probability by Sir Arthur Eddington, of a man who took a penny from his pocket and, tossing it five times, noticed that it came down "heads" each time. The chance of a sequence of five alike throws with a normal penny was 1/16. With a double-headed penny the chance was unity, but it could not be argued that it was more probable that the penny was double-headed than that it was normal; the rejection of the argument that the penny was probably double-headed would be based on the secret information that double-headed pennies were rare. Setting aside that and all other information which was not openly stated, there was no particular reason to reject the suggestion that the probability was 16 to 1 that the penny was double-headed. On the other hand, there was no reason for accepting it. When the argument was examined in detail, it was found to assume that, prior to testing, it was equally likely that the penny in question was double-headed or normal.

In the same way, if, prior to testing by the author's methods, it were assumed that two smooth graduations were equally likely to represent the underlying experience, then the graduation which was more probable by the author's test was to be preferred. In many cases, however, such an assumption was not only unfounded but definitely false. For instance, previous knowledge of the shape of mortality curves might show that one of the graduations was definitely unlikely. It would thus be seen that, despite the most careful planning of scientific tests for a graduation, that elusive virtue, experience, crept in. Therefore before applying the author's test to compare two graduations, it was necessary to be satisfied that they were equally likely in the light of experience.

On pp. 10 and 11 the author had considered the graduation of the O^{JF} Table and had shown the effect of grouping on the comparison with the expected deviations of the deviation between the actual and expected deaths. On purely probabilistic arguments, the comparison in age groups could not be expected to show any marked difference from the comparison in individual ages, although the fact that the total of the expected deaths for the whole table had been made approximately equal to the total of the actual deaths affected the position slightly. The results of Tables 1 and 2 were therefore at first sight surprising; but when it was remembered that the O^{JF} Table had been graduated by a summation formula which gave not an absolutely smooth curve but a curve which tended to follow the waves in the original data, it would be seen that those results were not really unexpected. The same phenomenon would normally occur in any graduation dependent upon a summation formula, and the test of such a graduation in age groups should not, therefore, be considered sufficient. Thus, taking the A 1924–29 Table and ignoring the question of duplicates, a comparison of the deviations of actual and expected deaths in quinquennial age groups with the expected deviations suggested that the graduation was satisfactory, but from a similar comparison that he (the speaker) had made at individual ages over a portion of the data it would appear that the graduation was, by the author's standard, of "doubtful improbability". Allowing for developments on the lines suggested in Appendix II, however, the graduation of the A 1924–29 Table was seen to be more acceptable.

He was glad to see the author's consideration of the application of the normal curve approximation for probabilities and numbers observed usually encountered in actuarial work. That filled an important gap in actuarial statistics.

The use of the χ^2 test to investigate the homogeneity of the A 1924–29 Table was most interesting, although the results brought out were not startling. He felt that it was in that part of mortality investigations that the χ^2 test could be applied most usefully.

Mr H. W. Haycocks said that the first point that he would like to consider concerned the author's theoretical model. He could not reconcile the author's statement of the problem on pp. 7 and 8 with the subsequent calculations and with the references to Rietz and others. Those references suggested a model in which n lives were considered, all of the same age but having different *a priori* probabilities of death. For instance, it could be assumed that there corresponded to each life an urn containing black and white balls, the proportion of black balls representing the *a priori* probability of dying. The first sample consisted in drawing at random a ball from each urn, and the number of black balls in the selection represented the number of deaths. The balls were then returned to the urns, and the experiment was repeated indefinitely. It could be shown that the variance of the distribution of the number of deaths with such a model was npq, where q was the mean of the *a priori* probabilities less a small deduction equal to n multiplied by the variance of the *a priori* probabilities.

That model did not seem to involve any conception of a large hypothetical population; it was rather a matter of imagining a series of repeated experiments on the same group of lives. The model could be extended to other cases such as a large population of urns from which a selection of n was made at random before drawing the balls, and in that case the variance could be shown to be equal to the simple binomial form of npq, q being the mean of the *a priori* probabilities in the whole population. A more realistic model would be a large bag of tickets, some marked "Deaths" and some "Not-Deaths", from which repeated samples of n could be taken. In that case it could also be shown that the variance approximately equalled npq where q was the proportion of tickets marked "Deaths" in the original bag.

The strange thing about all three models was that in each case the variance equalled approximately the binomial form npq, and it was rather difficult to see exactly how the question of heterogeneity entered into the problem. It seemed to him that it entered simply in the following way. If there was evidence that the data had been selected from various classes of lives, and that the proportion of each class varied appreciably from age to age, then it must be assumed that the hypothetical population had the same characteristic. In that event it could not be assumed that the underlying true probabilities lay on a simple smooth curve; in fact if such a curve existed it might be so complicated and unknown in form that it became impossible to make a strictly accurate graduation. He suggested that in such circumstances the graduation of a sample would result in a graduation which failed to satisfy the χ^2 test, and that the actuary would disregard the χ^2 test and choose some other criterion, for instance, he might be content to use grouped data; and, having full regard to the purpose for which he required the mortality table, he would choose a simple smooth curve.

The second point which he would like to consider concerned the question of degrees of freedom. There were two hypotheses which could be assumed in the test of a graduation. The first was that the graduation represented the true underlying probabilities and that it remained invariable for each sample. In that case the probability was given by the author's Table A according to the full number of ages. That was the hypothesis which the author assumed in testing the graphic graduation, although a graphic graduation would "force" a curve to fit the data just as a summation formula would.

The second hypothesis was that a fresh graduation was made for each subsequent sample. It was clear in such circumstances that the mean χ^2 would be less than under the first hypothesis, and hence that the probability would be over-stated if Table A were used according to the full number of ages. The author had explained how that problem could be dealt with when fitting a mathematical formula under which the constants were calculated by linear relations; but in the case of the summation formula the problem was more difficult. The author had overcome that difficulty by making a very interesting but laborious experiment, deducing as a result that for the Spencer graduation covering 30 ages the deduction should be about 5. From that point the author jumped somewhat hastily to the conclusion that for the Kenchington formula the deduction should be about 6. In his opinion the deduction actually should be less. It was known that in the case of the author's experiment the true underlying curve was a straight line, and therefore that the summation formula would not disturb the underlying true values but merely reduce and smooth the random errors. Therefore the more powerful of the two summation formulae would produce a graduation curve nearer to the true curve, and a mean χ^2 nearer to the mean "universe" χ^2. The Kenchington formula would therefore produce a mean χ^2 nearer to the mean "universe" χ^2 than would the Spencer formula and he thought that the deduction should be about 3 or 4, as the author would find if he were to repeat his experiment, graduating by the Kenchington formula.

His third point concerned the Stevens test, i.e. the signs of the deviations. He had found difficulty in understanding the explanation given on p. 29, and the origin of the figures 18 and 35 in the table at the top of p. 30. On referring to Stevens's article in the *Annals of Eugenics*, however, he discovered that the matter could be explained fairly easily by the use of

Lever's data as an illustration. There was a long series of 62 signs, 28 positive and 34 negative. Those two figures were assumed to be reasonable, and it was desired to prove that the probability of a plus or a minus sign falling in a particular place was independent of what had happened at previous places. The total possible number of arrangements with the 62 signs was obviously $\frac{62!}{28!\,34!}$, and each arrangement was considered equally likely. The total number of arrangements containing 17 positive groups (a positive group being a group containing one or more consecutive positive signs) was then calculated, the result being $\frac{35!}{18!\,17!}$, which gave the origin of the figures 18 and 35. The probability of obtaining a series containing 17 positive groups was given by dividing the latter expression by the former. On the basis of that probability it was possible to construct a probability distribution from which could be obtained the chance of getting a series containing 17 or more positive groups; i.e. a worse series. Next it could be shown that, provided that the figures were fairly large, a reasonable approximation to that distribution was the χ^2-distribution for 1 degree of freedom, and that was what the author had used.

It should be noted that Stevens's test was concerned only with the signs of the deviations and not with the relative magnitude of the positive and negative deviations. To test the magnitude it was necessary to consider the signs of the accumulated deviations, and Stevens's test might be used in that connexion. In the case of Lever's graduation, as would be expected, the result was very bad; $\chi^2 = 13{\cdot}5$, giving a probability of considerably less than 1 in 1000; and, contrary to what the author had said, that result indicated distortion. He had applied the test to the Kenchington graduation, and had obtained surprisingly good results, the values of χ^2 being $\cdot 33$ and $1{\cdot}3$. In the case of the Warren graduation, however, the values of χ^2 were $1{\cdot}8$ and $4{\cdot}5$. Though not significant, $1{\cdot}8$ was not particularly good, and the result was rather surprising in view of the fact that a mere inspection of the graduation suggested that the signs were reasonable. The trouble with Warren's graduation was that the signs changed too frequently, and by Stevens's test that meant a large value of χ^2 and a low probability. That must sound strange to actuaries, because he could not imagine a graduation being rejected on the ground that the signs changed too frequently; and quite rightly, because although such an event was improbable it was of no practical importance.

The idea of the signs changing too frequently suggested an alternative test which he thought was simpler and easier to understand than that of Stevens, viz. a simple comparison of the number of changes of sign with the number of non-changes. For example, in the short series "$+ + - -$" there was one change of sign and two non-changes; the plus sign persisted once and the minus sign persisted once. In a long random series the two figures should be approximately equal, i.e. the discrepancy should not be more than that allowed by the theory of random error. He had tested all three graduations in that way and the results were precisely the same as were obtained by Stevens's test.

It seemed to him that from the actuary's point of view the χ^2 test was not a complete test of a graduation, and that any significance test fell short of a complete test; it was rather a piece of evidence the importance of which varied considerably from problem to problem. From the actuary's point of view the χ^2 test could break down as a result of heterogeneity of data or because in connexion with the signs of the deviations it gave the same weight to the existence of "too many changes" as to that of "too many non-changes", and only the latter was of practical importance; and, moreover, it disregarded the signs of the accumulated deviations. It must be remembered that the actuary was not concerned with representing the data as mere historical facts, but rather with using them as a means to an end, i.e. to obtain a mortality table which could be used as a financial measuring rod.

He would like to close with a quotation from one of Professor Steffensen's works which put the matter in a nutshell: "An actuary may, because he has chiefly the future in view, content himself with a few essential features of the past, provided he keeps on the safe side; but a statistician may perhaps be particularly interested in the features which the actuary feels justified in neglecting. Therefore a graduation formula which was of value to the actuary might have but little value from the statistician's point of view."

Mr F. M. Redington said that his remarks were in support and amplification of the first of Mr Haycocks's three points. On pp. 30–3 of the paper the author had considered the question of heterogeneity, had discussed its effect on the expression Epq and had concluded, personally he thought rightly, that the effect was immaterial. He could not agree with the author, however, in the statement that "This means that heterogeneity is not a question at issue in the testing of a graduation", because the author overlooked a much more important effect of heterogeneity.

He could best illustrate his case by taking an extreme example such as a peculiar mortality experience, in which at the even ages all the lives exposed were mine workers and at the odd ages were insurance clerks. Such an experience would produce a fundamentally zigzag set of rates of mortality. It was quite clear that if an attempt were made to graduate that experience by a smooth curve, the graduation when tested by the χ^2 test would fail. That was not because the χ^2 test was in any way invalid. It gave the incontrovertible but not very useful answer that the adherence of the smooth curve to the zigzag experienced rates was bad, but it did not bring out the fundamental fact that it was not possible to produce a smooth curve of any sort which would satisfy the χ^2 test. That was an extreme example, but the effect was always present. There was heterogeneity through a large number of different causes, and it was extremely unlikely that the heterogeneity would be operating in such a way as to produce results progressing smoothly from age to age.

The data should not be heterogeneous; it was strictly not permissible to use heterogeneous data in actuarial work, but that was a counsel of perfection. The experience available would always be heterogeneous, and the question was whether it was something which could be ignored or whether it was a large factor which invalidated the χ^2 test in practice. Mr R. L. Michaelson had joined him in writing a paper for the Congress which was to have been held in 1940 discussing the quantitative measurement of that type of heterogeneity, and they had applied their test to the A 1924–9 experience, both select and ultimate, with rather interesting results. The select experience was consistent with a smooth homogeneous set of probabilities of death, and they found that it should be possible to make a smooth graduation of the select data which would satisfy the χ^2 test. For the ultimate experience, however, the picture was very different; as might be expected, the data were very heterogeneous, and there was an extremely low probability that the rates of mortality could have arisen from an underlying smooth series of q_x.

It seemed certain that if the author had tested the A 1924–29 ultimate experience he would have pronounced it a failure, but by so doing he would have unjustly criticized the authors of the graduation because they were faced with an inescapable dilemma. What were actuaries to do? Were they to discard the A 1924–29 experience or to eschew the χ^2 test? To do their work properly, he suggested that they should apply an amenability test of the kind developed in the Congress paper to which he had referred, to see how far the experience was capable of being graduated reasonably, and then they could apply the χ^2 test to see how far there was agreement with that amenability test. He could not seriously, however, advocate that actuaries should adopt such a lengthy procedure because the whole question of heterogeneity of data and errors in graduation paled into insignificance in practice when they came to use their rates of mortality extending into far periods in the future. The heterogeneity in time, so to speak, was of far greater importance than the heterogeneity in data.

In conclusion, he would like to say that he was not criticizing the χ^2 test; it was, with the provisos which the author had stated, perfectly valid. He was criticizing its applicability for actuarial purposes. It was a spirit-level which would be useful in testing a billiard table but was of no use in testing a football field, and he was afraid that actuaries had more football fields than billiard tables with which to deal.

Mr H. G. Jones remarked that it was evident, as Mr Redington had said, that the χ^2 test would fail in the circumstances envisaged in his example of a very special table, but that was surely due to the highly artificial manner in which the data had been mixed. So far as the effect of heterogeneity on the χ^2 test was concerned, it seemed to him that it was more the nature of the heterogeneity that mattered than the extent to which it existed. If, for example, lives of a certain type were present up to age 50 in considerable quantities, and none at all thereafter, he could see that it would interfere with the test; but if those lives were present in all the groups, in numbers varying in successive age groups in a random manner, would not the test be perfectly valid? He thought that in most instances actuaries would be able to see by investigation where an artificial heterogeneity existed, such as the presence of a group of lives up to and not beyond a certain age; that, he imagined, would usually be known. They might not be able, perhaps, so clearly to distinguish the other kind, which was random. Did not the importance of heterogeneity lie in the danger that the same proportion of lives would not hold in any other case to which it was desired to apply the table, rather than in a distortion of the actual results experienced? In other words, the table might reflect the probabilities existing in the experience under examination, and might embody a perfect graduation of that experience regarded purely as a record of what had happened. If, however, the results were applied to another body of lives of a similar nature, but mixed in a different way, that difference in composition might itself be sufficient to prevent agreement between the original experience and the results in the new body.

He had not thought that the question of change of sign implied anything more than had already been tested by the probability test. If a test had been made for the chance of a certain result arising out of a given theoretical set of probabilities, did not that test include implicitly the question of the way in which the signs changed? If the curve at any point were so distorted as to cause a run of similar signs, the numerical value of the deviations in that region would be increased with corresponding effect on the result of the χ^2 test.

Sir William Elderton said that he wished to begin by saying "Thank you" to the author for the delightful quotation at the head of the paper, for the story which went with it, and for the demonstration that the members of the Institute, including himself, had for many years overlooked a paper in their own *Journal*, in 1872, of which they ought to have known and to have been proud—a paper which they should have used continuously, instead of leaving it to someone outside their profession to rediscover the method, to extend it and to make the method generally applicable over a much wider sphere.

Mr Redington made some remarks about a spirit-level, with a very nice analogy about using it with a football field when it was suitable for a billiard table; but was that really a criticism of the method of testing a graduation when, after all, the things that the author was trying to replace were rough and ready methods which aimed at the very spirit-levelling to which objection was raised? He agreed that a spirit-level could not be expected to do more than spirit-levelling, but the methods suggested by the author replaced something which called itself a spirit-level, and was not, by what was at any rate much better at spirit-levelling.

He would like to suggest that if a graduation had been made which was a very long way from the data, whether using a smooth curve or some other method, it was not a "graduation" of those data. People who talked a great deal about smoothness made him feel that in the end the trouble was their own, because instead of using for their graduations some mathematical formula, which, even if it were a sine curve, was at any rate smooth, they had followed many devices and were content with something which was not graduation in the "smooth" sense. Those who had suffered for many years under the H^M and Carlisle Tables would know exactly what he meant.

There was one suggestion which he would make to those who thought that the shape of a mortality curve was known. Perhaps in changed circumstances all their preconceived notions might be wrong. He always regarded the A 1924–29 Table as one of the greatest pieces of luck in graduation that he had ever seen. The data were terrible. Though it was not possible to go behind them, there was something definitely wrong with them. At the time that the graduation was made, it was not known what was wrong with them. The graduation produced something which could be used, and, as luck would have it, when at a later date they were able to find out that some ingenious people in insurance companies had written down information which was not exact fact, they discovered that the graduation was nearer the facts than the people who provided the data meant them to get!

Mr Haycocks had referred to the number of degrees of freedom which the author suggested should be deducted in the case of a Spencer graduation, and thought that with the Kenchington formula fewer should be deducted. He did not pretend to have considered that point in detail, but, *a priori*, he would have thought, with the author, that a 21-term formula would have fewer degrees of freedom to be deducted than a 27-term formula, just as a parabolic curve with three unknowns had fewer degrees of freedom than one with four.

He had enjoyed reading the paper more than he could say. He regarded it as a first-class piece of research and a paper of real importance, and the best thing he had read for a long time on the subject with which it dealt. He was grateful to the author, moreover, for showing on p. 20 that a lucky guess of his own was more defensible than he had realized.

The President (Colonel H. J. P. Oakley) said that Mr Haycocks's closing reference to a quotation from Dr Steffensen prompted him to announce that it had been hoped that it would be possible for Dr Steffensen to be present that evening, and also Professor Meidell, of Oslo. Both gentlemen had written regretting their inability to attend, due, no doubt, to very recent events in Denmark and in Norway, but Dr Steffensen had been able to send a written contribution to the discussion, which he would ask the Honorary Secretary to read to the meeting. The following contribution from **Dr Steffensen** was then read:

"By the courtesy of the President of the Institute I have been enabled to read in proof Mr H. L. Seal's paper entitled 'Tests of a Mortality Table Graduation'; and being prevented under the present circumstances from accepting the kind invitation to be present at the meeting when the paper is to be discussed, I avail myself with pleasure of the opportunity offered to me of presenting my observations in writing.

"I have found the paper very interesting; it deals with a subject which deserves more attention than it has hitherto received, and takes up several threads which have been somewhat neglected in the past. A distinguishing feature is the care with which the accuracy of various approximations to binomial probabilities ('normal', Type III and Poisson approximations) is examined; the author's Tables 3 and 4, incorporating the results of this investigation, may also in future serve to avoid inadequate approximations which frequently occur in this field. For it can hardly be denied that, in spite of Karl Pearson's work, a very human inclination persists to label a distribution 'normal' chiefly because a normal distribution is so easy to deal with. As a matter of fact, nearly every distribution is normal at the top, and only the tails can tell us whether the distribution is really normal or not.

"The author is aware that a single summary test, such as the P_Q or the χ^2 test, is not sufficient for accepting a graduation; also that the additional examination of the deviations should comprise not only their sign but also their magnitude. In this, the author is in entire agreement with the views expressed in my paper 'On the Fitting of Makeham's Curve to Mortality Observations', which he quotes (compare also my book *Forsikrings-matematik*, pp. 129–32). The comparison in tabular form on p. 19 between the actual and the theoretical distribution of the deviations is everything that could be desired, and it is, in principle, the same method which I have employed on p. 393 of my paper. But if the number of ages comprised in the graduation is sufficiently large for allowing a detailed examination of the distribution, then the value of the summary (χ^2) test, already doubtful, becomes so problematic that it seems preferable to leave it out altogether.

"The author states (p. 11) that 'at any age the observations are quite independent of those at any other age so the calculated probabilities may be compounded by multiplication'. This statement does not seem clear. Thus, for instance, in Table 2 we have $E_{42} = 2345$, $E_{43} = 2343$, and since E_{43} consists to a large extent of the same lives as E_{42}, the observed values of q_{42} and q_{43}, viz.:

$$q'_{42} = \frac{36}{2345} = \cdot 01535, \qquad q'_{43} = \frac{18}{2343} = \cdot 00768,$$

are not independent of one another. It seems fairly obvious that the fact that an unusually high proportion of E_{42} failed had a selective influence on E_{43}, of which, therefore, 'too few' died. The observations are thus interdependent at neighbouring ages, and this is a point which may influence methods of graduation resting on the assumption of independence. Be this as it may, the calculated *probabilities* may all the same be compounded by multiplication according to the ordinary rules, for the question of combination of probabilities has nothing to do with the observations from which the probabilities have arisen.

"To the list of references at the end of the paper might have been added T. N. Thiele, *Theory of Observations*, whose para. 45 is devoted to the test of fit; L. v. Bortkiewics, *Die Iterationen*, dealing with the question of sequences of same sign; and R. v. Mises, *Wahrscheinlichkeitsrechnung*, where his own contributions to the theory of sequences are mentioned in para. 4, art. 2."

Mr R. E. Beard said that what was required was a method suitable for testing a grouped graduation so that unnecessary arithmetic could be avoided. He had not been able to investigate the problem theoretically, but the results of some practical experiments using the χ^2 method, although not sufficiently extensive to provide conclusive results, might be of interest.

In the first place he had tested his graduation of the O^m data published in *J.I.A.* Vol. LXVII. The value of P calculated from the individual ages was ·037; from the quinquennial grouped graduation a value of ·030 was obtained, a not unreasonable approximation to the value obtained from the extended calculations. If the components of χ^2 were calculated by grouping the individual deviations, the value obtained for P was ·062, the increase in the probability being in accordance with the remarks in Sir William Elderton's book. The difference between ·062 and ·037 corresponded to approximately 2 degrees of freedom and that suggested a possible modification of the χ^2 test for use with grouped data.

To test the idea further he had collected the results of Table 5 in the present paper into quinquennial groups and had calculated the mean value of χ^2 from the nine samples. The value was 4·21 for six groups, indicating that compensation for the grouping could be made by a reduction in the number of degrees of freedom. As a final test he had utilized the Kenchington graduation given in Table 2. Using 44 degrees of freedom the P for $\chi^2 = 41{\cdot}425$ was ·39, whereas from Table 1, using 5 degrees of freedom, the value of P was ·66. By using only 3 degrees of freedom in the latter case a value of ·37 was obtained for P.

Mr C. F. Trustam, in closing the discussion, said that if the author had had any doubts as to the reception that his paper would encounter his mind would have been set at rest by the discussion that evening. The experts had spoken on the contents of the paper, and there was not very much that he could add in a constructive way, but perhaps he might say something on the presentation of the matter. His contact with γ functions and the like dated back many years, and he had approached the paper with a good deal of apprehension. It was, therefore, a revelation to him to find how gently the author took his readers by the hand and led them along, so that even if they had to take the analysis for granted, the problems and their solution stood out in a commendably clear way.

He confessed that he always read such a paper with one eye on the mathematical book-shelf and the other on the office desk; and he could not help wondering to what extent the nature of the data with which actuaries had to deal when handling mortality problems justified the elaboration of statistical analysis. It often seemed to him that they were very much between the devil and the deep sea. When dealing with the devil of a small experience, probably elaborate methods were neither justified theoretically nor warranted by the purpose of the particular problem in hand. When they came to the deep sea of a wider investigation they encountered the bugbear of heterogeneity, and he had found it difficult to escape from the thought that they were concentrating attention on the waves in the curve when the very direction of that curve might at some point or points be affected by disturbing factors from outside; or, to take Mr Redington's analogy of the spirit-level and the football field, was not it the case that not only were they trying to use a spirit-level on the surface of a football field, but they were trying to use it on a football field which was not even horizontal, and where their first attack ought to be a rough levelling with pieces of wood and string, before coming to the finer process of smoothing the surface.

One question struck him when reading the paper on which the author might be able to make some comment. It seemed to him that the primary usefulness of tests of the kind described was negative rather than positive. If the graduation passed the test, the result was very little more than an extra glow of warmth for the investigator; but if the test rejected it, then something specific had been achieved. It so happened that in illustrating the paper the tests described were all applied to graduations which passed as satisfactory, and he would be interested to know what would happen if they were applied to bad graduations. What he had particularly in mind was that if a graduation was so bad that the P_Q or the χ^2 test rejected it, was not it possible that it was so bad that the average deviation test, or something equally simple, would also reject it, so that there would be no need to have gone to the elaboration of tests of the kind in question?

In conclusion, it seemed to him that perhaps after all the prime importance of the paper lay not merely in its contents or in its presentation, but above all in the sheer fact that there were people like Mr Seal engaged in research. However practical the actuary might be, he had to admit that the gas fire arose out of the chemical laboratory and the telephone was born in the physical laboratory, and only by continuing research work of the kind represented by the paper was there any hope that the craft of the actuary would be saved from stagnation and decay.

The President remarked that the Institute had had a very bare session, owing first to the black-out and the long dark nights in the winter months, and secondly to the distance from the Institute of many members on account of evacuation. They would all agree, however, that the paper which had been discussed that evening was of outstanding merit, and would probably still have been outstanding even had the session been normal; in fact, if they had had a session full of papers and every paper had reached the standard achieved by Mr Seal, it would no longer have been a normal session but abnormal, and as a random sample might have proved misleading. The discussion also had been of high merit.

If he had any criticism, it was that the conclusions set out by the author might have been expressed more fully, and in a way which would have presented the results more clearly to the minds of those who had never really enjoyed the study of the long list of references which came at the end of the paper, and also of those who, although they might, with effort, have made such study, had been content to forget. That criticism, however, did not detract from the value of the work, especially in setting out the problem and then giving the probabilistic argument. Such papers enriched the *Journal* and stimulated thought by those whose bent was in similar directions; that stimulus in turn might have far-reaching results in the further development of actuarial science, as had been so delightfully expressed a few moments ago by Mr Trustam.

He knew that he could add the thanks of the meeting for the excellent piece of work that Mr Seal had submitted.

Mr H. L. Seal, in reply, regretted that limitations of space had necessitated the condensation of his paper to the exclusion of some fuller explanations which several speakers had desired.

Mr Longley-Cook had very courteously sent him a copy of his remarks in advance, and the President had sent him a copy of Professor Steffensen's contribution, so that he proposed to reply more fully to those at the moment, leaving other contributions to be dealt with later.

It seemed to him that Mr Longley-Cook's remarks were developed from a philosophy of graduation different from that which he himself held. Karl Pearson, in Chapter 3 of his *Grammar of Science*, described a scientific law as "a *résumé* in mental shorthand, which replaces for us a lengthy description of the sequences among our sense impressions". It might be said that it was an hypothesis which accounted for recorded observations until they accumulated and overthrew it. And so, speaking of mortality, he would take the simplest hypothesis of all, viz. a linear progression for q_x, and the fact that it would immediately prove hopelessly wrong would not discourage him because he would expect it. He would then proceed to consider the more complex hypotheses that Karl Pearson had categorized as "laws". They were not put forward in the belief that they were universally true but they synthesized the observations. They varied from the simple to the more complex, and the aim in a graduation was not to produce a law which gave the highest value of P_{χ^2}, as he thought Mr Longley-Cook had suggested. He would not say, with Mr Longley-Cook, that the way to test the graduation was to produce one better; he would produce the most simple graduation consistent with the observations; so that *a priori* probability hardly entered into the question of mortality graduations. He did not believe that it was possible to talk about an *a priori* probability for a graduation.

The equalization of the expected and the actual deaths and their accumulated deviations was quite an arbitrary procedure. R. A. Fisher had produced reasons for believing that the usual equalization of first, second, third and fourth moments was seldom theoretically appropriate. By "the method of moments" was meant, in that case, the method of comparison of the sums of the expected and actual deaths and their accumulated deviations. It was wrong to say that the successive sums of the deviations *must* be zero; it was only an arbitrary and practical method of fitting a formula to the data and of making sure that the graduated curve was not too far from the observations. Mr Longley-Cook appeared to think that the equalization of the expected and actual deaths was necessary, and had put into his mouth words that he did not use by saying their difference *should* be approximately zero, when, in fact, he had said that it was *likely* to be fairly close to zero.

He noticed that Professor Steffensen did not agree with the use of the χ^2 test, but preferred the collection of the values of $(\theta - Eq)/\sqrt{(Epq)}$ into groups. He had observed, in Steffensen's work quoted in the list of references and also in his text-book on insurance mathematics, that he preferred that test and did not use the χ^2 test at all. The reason that he (the speaker) used the χ^2 test and preferred it was that it provided a test that anyone could use. In the case of the table on p. 19, for example, it was necessary to know what the normal curve was, and more or less how to fit it. He was not saying that actuaries could not do it, but he thought that they would find a certain amount of trouble, whereas any actuary, and indeed anyone who had passed Part I, could carry out the χ^2 graduation test as set out in full on p. 35. It was then necessary to test for sequences of positive and negative signs, and the work was almost finished.

It would seem that Mr Longley-Cook was mistaken with regard to grouping. Mr Longley-Cook suggested that probabilistically speaking there was no reason why grouping should alter a test of a graduation. But in the case of a graduation, grouped quinquennially, with about 12 groups altogether there might be a sequence of 3 successive negative signs in the central groups. It could be said that such a sequence was not bad and that the graduation would be passed; but if those 3 negative groups were actually 3×5, i.e. 15, single ages each showing negative deviations, no one would deem the graduation satisfactory.

He had thought rather carefully about dependence between the ages mentioned by Professor Steffensen, because a long time ago Dr W. F. Sheppard, in a classical paper on linear compounding (*J.I.A.* Vol. XLVIII, 403, note A), said that much had to be done with regard to the correlation between the observations at the different ages. He was not convinced that Sheppard was right, although Professor Steffensen seemed to agree with him. Such a dependence would exist, he thought, when E_x was found by the census method; but in the usual case where a person was either "exposed" or not to risk of death at age x, it was in his opinion wrong to contend that the mere fact of observing a man might alter his mortality, and that if he managed to escape death at age x whilst under observation, he would be more likely to die at age $(x+1)$ than a contemporary of his who only came under observation at age $(x+1)$.

Sir William Elderton, interposing, said that in other words the question was whether, if three successive tossings of a penny had resulted in "heads", the next tossing would be expected to produce a "tail".

Mr Seal said he had looked up the reference to Thiele given by Professor Steffensen, but did not think it very suitable as it required the close study of a somewhat "dated" work for its comprehension. As regards the reference to von Mises, he found that von Mises actually produced the formula which was given on p. 29 of the paper, and on the discovery of which he had been congratulating himself, but here again a close study of a controversial theory had to be made before the paragraph cited became clear.

Sir William Elderton had mentioned the paper by Thiele in the *Journal*, and had said how glad he was to see that Thiele in a way anticipated Karl Pearson; but it was interesting to find that in 1876 de Forrest wrote a paper which in several respects anticipated his own paper. Mr H. H. Wolfenden had drawn his attention to it, and he found that it was an extraordinary paper for a man writing in 1876.

Mr Redington had said that as regards the A 1924–29 data he thought it "certain" that the χ^2 test would have rejected the official graduation. On the other hand Mr Longley-Cook thought that the official graduation was only of "doubtful improbability", at least over part of the data! He himself had tried the χ^2 test on the A 1924–29 data, and had noticed that, as Sir William Elderton had mentioned, it was possible to detect the ages at which the offices had misled the Mortality Committee; at certain ages the deviation between the actual and expected was over three times the standard deviation. The χ^2 test picked that up at once, and produced a value of P_{χ^2} in the region of ·000008; but when rough allowance was made for duplicates in the way suggested it would be found that the χ^2 test adjudged the Spencer graduation of the A 1924–29 ultimate table as very reasonable.

He had tried to produce a test for tyros in actuarial science. His remarks about experience on the first page were made because there was a tendency to disparage young actuarial students on account of their lack of experience. He had tried to produce a test which did not require a great deal of previous experience for its practical use. It could no longer be said "You need a great deal of experience to test a mortality table graduation". It must not be thought, however, that he had produced a test that fools could use; a certain amount of *le bon sens* was needed in applying any test. He thought that the χ^2 test was a useful, simple, summary test for any medium-sized office to use when comparing actual and expected deaths and in seeing whether the mortality of that office was above, or below, or very much deviated from, the standard table.

He agreed with Mr Trustam about the negative aspect of the test. That was, however, one of the hall-marks of statistical tests; it could never be said that an hypothesis was certainly verified, only that it was quite possibly true. When Mr Trustam asked for an example where his "old friend" the average deviation test could not give a definite answer concerning a graduation whilst the χ^2 test could do so, he would reply that he was unable to see how the mass of figures in column (7) of Table 2 could ever lead to any firm judgment concerning the success of a graduation. Even the figures "29 below, 21 above" scarcely led to any definite conclusion.

In reply to the question, asked by Mr Jones, whether the χ^2 test included the test of signs, he was not sure of the complete independence of the tests. The χ^2 test, owing to the fact that it squared every deviation, ignored the signs altogether, but they could be tested for sequence, and that seemed to give all that was required.

Mr Seal subsequently wrote as follows:

The probability "set-up". The theoretical model which I put forward on pp. 7–8 *ante* hypothecates as many "populations" as there are ages in the mortality table; each of these populations is, by hypothesis, subject to a uniform rate of mortality. The populations are not infinite of necessity for limited populations of balls in a bag could be substituted provided that, in drawing E times from the bag, the ball just drawn is replaced before the next drawing. It is plain that Mr Haycocks, in concentrating upon the boundlessness of my hypothetical population, has been led to construct, in his first model, a population which no longer gives rise to a simple binomial distribution. Mr Haycocks's third model is the standard case of "drawings of balls from a bag without replacement", and here again the binomial distribution is not directly applicable.

Preference should only be given to a complicated hypothesis when a simpler hypothesis has been shown to fail in practice. Until Mr Haycocks can do this his models must be discarded in favour of the simple binomial hypothesis. It is to be observed, by the way, that the acceptance of a model other than the one I suggest would necessitate the recalculation of the examples which I have based on the binomial distribution.

Heterogeneity of the data. In the paper I dealt with the heterogeneity of the observations at any one age. Messrs Haycocks, Redington and H. G. Jones, however, refer to possible differences in heterogeneity at different ages. Whereas I was concerned to show that the *variance* at any one age was only slightly less than the simple binomial variance when homogeneity was absent, the new problem is to consider the effect of varying heterogeneities at different ages on the *mean* rate of mortality at each age. As Mr Redington points out, this effect might be very serious; but in practice the position would be rather as outlined by Mr H. G. Jones, whose remarks I whole-heartedly endorse. I believe that the misuse of the χ^2 test which would happen in such circumstances could "never" occur in practice, and that the legitimate employment of mortality tables based on a "broad basis both in space and time" can now be substantiated from a theoretical as well as a practical standpoint.

At the time of writing the foregoing paper I had been unable to refer to a copy of Peek's (1899) article but had included it in my list of references because it has, for many years, been considered the fundamental paper on the legitimacy of application of probabilistic methods in life assurance. Furthermore, Esscher (1920) reproduces many of Peek's figures and it was on this ground that I bracketed together the three authors mentioned at the beginning of the footnote on p. 32 *ante*.

I have now had access to Peek's paper and find that I was mistaken in thinking that he had "graduated" the rates of mortality observed in successive years by means of a trend line and that he had restricted himself to population data. Owing to the central position which has been accorded to this article (see e.g. Bohlmann, *6th T.I.C.A.* Vol. 1, i, pp. 662 *et seq.*), I propose to summarize the results derived from Peek's figures by means of modern statistical methods.

Two separate sets of data were considered by Peek. The first related to the Dutch population statistics of the years 1880–9 and the second to the mortality statistics of Dutch civil servants from 1878 to 1894 inclusive. In the former case, in effect, the value of χ^2 for differences between calendar years was calculated for each age from 0 to 90 separately (cp. the first part of Table 8 *ante*). The result is 91 separate values of the random variable χ^2 based on $10-1=9$ degrees of freedom; the last 80 of these values have been collected below and compared with the expected frequency obtained from Elderton's Table XII in Pearson (1914).

Value of χ^2	Actual frequency	Expected frequency
0– 2	2	·68
2– 4	1	6·40
4– 6	9	13·72
6– 8	8	16·46
8–10	9	14·69
10–12	14	10·97
12–14	13	7·28
14–16	4	4·44
16–18	2	2·54
18–20	6	1·38
20–22	5	·72
22–24	2	·37
24–26	2	·18
26–∞	3	·17
	80	80·00

There can be no doubt that Czuber's dictum that the corresponding mortality rates "may quite well be regarded and treated as probabilities" must be viewed with considerable scepticism.

Peek's figures for the Dutch civil servants are repeated in the appended table, pp. 66, 67. The total number of expected deaths for each age group does not agree with the actual number of deaths in that group, so there are 17 degrees of freedom (one degree for each calendar year) in each age group. The twelve corresponding values of χ^2 are: 14·61, 15·04, 8·83, 14·90, 16·20, 11·40, 24·50, 18·28, 13·80, 14·13, 21·89 and 13·95, a total of 187·53 with $f=17\times 12=204$; thus $P_{\chi^2}=\cdot 79$ approximately. In this case "time" is not to be regarded as influencing the mortality.

It is of interest to collect the 204 constituent parts of the total χ^2 into a frequency distribution and compare this with the theoretical distribution corresponding to $f=1$. The result is given in the table opposite.

The agreement of observed values with those theoretically expected is, as Peek says, "completely satisfactory". On the basis of these figures the direct application of probability calculus to mortality statistics was considered to be substantiated.

The use of accumulated deviations. Two speakers, Mr Longley-Cook and Mr Haycocks, referred to the test of the changes of sign of the accumulated deviations between actual and expected deaths. The former spoke of it as a "necessary test" (presumably additional to the χ^2 test and the test for sequences) and the latter would possibly substitute it for the χ^2 test applied to the whole data. Unfortunately I have no proof that my suggested procedure is a "better" measuring rod than Mr Haycocks's, and can only express my personal dissatisfaction with the accumulated deviation test in any shape or form.

Value of χ^2	Actual frequency	Expected frequency
·00– ·25	78	78·12
·25– ·50	21	28·07
·50– ·75	28	18·98
·75–1·00	16	14·11
1·00–1·25	10	10·96
1·25–1·50	10	8·75
1·50–1·75	7	7·10
1·75–2·00	6	5·83
2 –3	14	15·10
3 –4	7	7·70
4 –5	5	4·11
5 –∞	2	5·17
	204	204·00

The alternative test for sequences of sign suggested by Mr Haycocks seems to be the same as that described by Bond on pp. 116–17 of his *Probability and Random Errors* (1935), to which I refer interested members. Lack of space prevented my including it in my paper.

As regards Mr Haycocks's statements concerning "too many changes" of sign in the progression of the differences $(\theta_x - E_x q_x)$, might not such an occurrence be taken as an indication that the data would repay further investigation? For example, Mr Redington's table of mine workers and insurance clerks would give rise to just such an excess of changes of sign.

Conclusions. The interesting and instructive remarks of the speakers in the discussion have not altered my principal thesis which I still maintain: Provided "duplicates" can be allowed for, the χ^2 test together with Stevens's test of sequences provide an essentially practical method of testing a graduation of any mortality statistics likely to be encountered by an actuary.

Data of the "Eerste Ambtenaarentafel"

Year	14–34			35–43			44–51		
	E	θ	Eq	E	θ	Eq	E	θ	Eq
1878	4356	33	24·19	3271	27	24·08	2446	35	28·80
1879	4526	26	25·12	3307	32	24·38	2463	25	28·84
1880	4619	31	25·54	3390	25	24·92	2548	32	29·89
1881	4787	24	26·41	3438	30	25·30	2573	29	30·22
1882	5058	29	27·94	3474	29	25·66	2559	36	29·89
1883	5330	30	29·58	3546	24	26·17	2588	32	30·04
1884	5503	32	30·61	3603	31	26·52	2682	36	31·17
1885	5530	29	30·81	3653	27	26·82	2775	37	32·26
1886	5535	31	30·64	3753	32	27·52	2864	29	33·33
1887	5640	40	30·92	3845	22	28·18	2923	35	34·12
1888	5480	23	29·87	3942	28	28·94	2929	32	34·33
1889	5382	21	29·20	4004	19	29·33	2988	28	35·03
1890	5319	25	28·84	4110	34	30·11	3030	31	35·77
1891	5408	32	29·62	4267	30	31·08	3041	34	35·79
1892	5488	33	30·13	4465	25	32·50	3108	34	36·50
1893	5553	29	30·60	4601	28	33·56	3168	38	37·16
1894	5580	23	31·03	4679	40	34·14	3295	44	38·49
Total	89094	491	491·05	65348	483	479·21	47980	567	561·63

Year	52–56			57–61			62–65		
	E	θ	Eq	E	θ	Eq	E	θ	Eq
1878	1347	23	24·05	1247	24	31·61	831	32	28·31
1879	1403	27	25·07	1239	28	29·92	827	31	28·09
1880	1407	15	25·25	1237	26	29·99	817	26	27·54
1881	1435	26	25·82	1216	31	29·44	872	28	29·39
1882	1454	26	26·05	1214	28	29·11	913	34	30·71
1883	1476	28	26·46	1226	39	29·35	922	23	31·25
1884	1444	33	25·93	1269	25	30·41	895	26	30·56
1885	1441	24	25·92	1283	32	30·80	873	36	29·79
1886	1444	26	25·98	1309	45	31·56	825	26	27·99
1887	1429	21	25·62	1317	30	31·60	856	23	28·84
1888	1451	23	25·92	1336	32	32·11	884	26	29·91
1889	1504	25	26·89	1303	26	31·33	911	40	30·75
1890	1546	29	27·67	1312	32	33·42	917	31	32·92
1891	1622	23	28·43	1324	34	32·25	908	30	30·99
1892	1684	35	29·55	1308	35	31·67	951	29	32·33
1893	1722	35	30·20	1339	25	32·33	949	36	32·20
1894	1770	20	31·09	1411	32	34·15	941	33	32·20
Total	25579	439	455·90	21890	524	531·05	15092	510	513·77

Data of the "Eerste Ambtenaarentafel"

Year	Age Group								
	66–69			70–72			73–75		
	E	θ	Eq	E	θ	Eq	E	θ	Eq
1878	729	45	34·43	463	33	30·67	348	28	28·34
1879	746	33	35·47	431	34	28·62	357	37	30·76
1880	781	36	37·29	403	33	26·72	364	32	31·49
1881	738	38	35·40	418	25	27·39	361	28	31·30
1882	716	33	34·25	465	36	30·53	331	24	28·94
1883	707	42	33·81	477	28	31·61	319	25	27·79
1884	705	29	33·45	471	22	31·16	339	33	28·97
1885	751	32	35·65	446	39	29·53	370	34	31·75
1886	789	39	37·38	434	25	28·78	372	27	32·33
1887	795	33	38·03	418	27	27·58	375	29	32·51
1888	770	34	36·99	456	25	30·05	347	33	30·20
1889	756	32	36·19	479	26	31·52	349	24	30·23
1890	721	47	34·44	516	27	34·05	344	29	29·74
1891	738	44	35·48	503	36	33·32	368	26	31·47
1892	748	34	35·98	464	31	30·73	398	45	33·92
1893	779	46	37·40	442	25	29·09	404	40	34·54
1894	779	22	37·65	432	20	28·42	401	29	34·54
Total	12748	619	609·29	7718	492	509·77	6147	523	528·82

Year	Age Group								
	76–78			79–83			84–100		
	E	θ	Eq	E	θ	Eq	E	θ	Eq
1878	226	21	24·85	228	46	33·99	98	27	24·05
1879	233	32	25·41	232	27	34·64	89	25	21·54
1880	251	24	27·69	225	29	34·28	84	19	20·27
1881	256	27	28·08	234	37	35·11	98	25	22·95
1882	267	31	29·32	231	36	34·53	104	22	24·12
1883	272	26	29·92	236	28	35·21	106	16	25·23
1884	278	27	30·62	246	39	36·49	121	33	28·99
1885	263	27	29·19	262	46	39·26	102	26	24·83
1886	244	22	26·95	269	35	39·93	105	21	25·33
1887	255	31	27·88	276	50	41·05	117	22	27·92
1888	278	40	30·40	256	34	38·62	121	32	29·18
1889	273	36	30·09	252	38	38·07	119	33	28·98
1890	278	35	30·62	247	38	37·37	116	21	27·84
1891	261	36	28·86	248	42	37·45	129	34	31·81
1892	263	30	28·96	238	43	35·68	127	28	30·31
1893	255	29	28·02	246	30	36·81	123	27	30·56
1894	265	25	29·25	256	29	38·77	116	30	30·06
Total	4418	499	486·11	4182	627	627·26	1875	441	453·97