On the Graduation of Frequency Distributions.

BY

H. C. CARVER.

The graduation of frequency distributions may not improperly be referred to as that branch of actuarial and statistical theory which is most neglected.

Elderton's "Frequency Curves and Correlation," which is recommended by the Educational Committee of this Society and is unquestionably the text best known to the English speaking actuaries, presents the Pearsonian methods which have dominated to a marked degree the English school of biometricians.

Pearson's method is empirical and is based upon the assumption that the differential equation of a unimodal distribution is of the form

$$\frac{dy}{dx} = \frac{y(a-x)}{f(x)} .$$

The following phenomena of such distributions suggest this equation:

(*a*) At the mode ($x = a$) the derivative of the curve is necessarily zero. The factor $(a - x)$ takes this fact into consideration.

(*b*) At the extremes of the distribution there is generally high contact, that is, the slope of the curve tends to approach zero as y likewise diminishes in value.

(*c*) The balance of the differential equation for any distribution may be represented by $f(x)$ appearing in the denominator. We assume that this function may be expanded in the power series $b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots$, which is generally so rapidly convergent that the terms which involve the third and higher powers of x may be neglected.

We thus arrive at

$$\frac{1}{y} \cdot \frac{dy}{dx} = \frac{a-x}{b_0 + b_1x + b_2x^2}$$

as the differential equation of unimodal distributions in general.

The integration of this equation may produce various types of frequency curves, depending upon the inter-related values of the constants b_0, b_1 and b_2. Thus we may have as solutions

1. $y = \kappa e^{\frac{-(x-b)^2}{2\sigma^2}}$ if $b_1 = b_2 = 0$.

2. $y = \kappa \left(1 + \frac{x}{\alpha}\right)^{\gamma a} e^{-\gamma x}$ if $b_2 = 0$ $b_1 \neq 0$.

3. $y = \kappa \left(1 + \frac{x}{\alpha_1}\right)^{m_1} \left(1 + \frac{x}{\alpha_2}\right)^{m_2}$
if the roots of $b_0 + b_1x + b_2x^2 = 0$ are real.

4. $y = \kappa \left(1 + \frac{x^2}{\alpha^2}\right)^{m_1} e^{m_2 \tan^{-1}\frac{x}{a}}$
if the roots of $b_0 + b_1x + b_2x^2 = 0$ are complex.

The above curves, with modifications, make up Pearson's system which, according to Mr. Elderton's recent supplement, comprises twelve distinct functions.

Although the practical sufficiency of Pearson's method leaves but little to be desired, still it is a highly unfortunate fact that a graduation involving such procedure can only be effected at the expense of a vast amount of labor. Only those who have completed a graduation of the transcendental type four (which is the one most frequently met in practice) will fully appreciate the truth of the preceding statement.

The purpose of this paper is to point out, and illustrate with examples, that if we slightly modify Pearson's hypothesis so that it will permit treatment by the finite, rather than the infinitesimal, calculus, we may eliminate a great deal of theoretical and laborious procedure, and what is equally, if not more important, treat all distributions which belong to Pearson's system (and certain others) by a single method regardless of "type."

Section II.

The reasoning which prompted Pearson to choose his *differential* equation also suggests

$$\frac{\Delta y_x}{\Delta x} = \frac{y_x(a - x)}{b_0 + b_1x + b_2x^2}$$

as the *difference* equation of a unimodal distribution, since if there be a maximum there must also be a value of $x = a$ for which

$\Delta y_x = 0$, and moreover, at the extremes where $y_x = 0$ (to quote Elderton) "the finite difference between two successive ordinates must be zero, or there will not be contact."

By arbitrarily allowing Δx to represent the difference in magnitudes of two successive classes, this element may be considered equal to unity for the particular distribution being graduated, and thus eliminated from further consideration.

It follows that our difference equation may now be written as

$$\frac{y_{x+1}}{y_x} = 1 + \frac{a - x}{b_0 + b_1 x + b_2 x^2} = \frac{x^2 + c_1 x + c_2}{x^2 + c_3 x + c_4}$$

It is important to note here that a knowledge of the values of the constants in the differential equation

$$\frac{1}{y} \cdot \frac{dy}{dx} = \frac{a - x}{b_0 + b_1 x + b_2 x^2}$$

for a particular distribution does not permit a calculation of the ordinates of the distribution until the equation has been integrated, producing a solution of the form $y = \kappa f(x)$. Our case is, however, essentially different, for as soon as c_1, c_2, c_3 and c_4 are known, the computed values of y_{x+1}/y_x (corresponding to the l_{x+1}/l_x or p_x in actuarial theory) absolutely determine without any integration the shape of the frequency curve. The condition that the sum of the graduated ordinates must equal that of the ungraduated ordinates will enable us to arrive at the proper radix.

Consequently, variations in type, an outgrowth of integration, will in no way concern us.

There remains but the problem of determining the constants for a particular distribution.

Section III.

Although it is possible to proceed, using the calculus of finite differences, along lines parallelling Elderton's, yet the constants may be determined more easily as follows:

We obtain by clearing our difference equation of fractions

$$c_1 x y_x + c_2 y_x - c_3 x y_{x+1} - c_4 y_{x+1} = x^2 y_{x+1} - x^2 y_x.$$

If we multiply this through by x^n, and sum with respect to x between the limits $x = r$ and $x = s - 1$, we have, giving n successively the values 0, 1, 2 and 3,

$$
\text{I.}\quad\begin{cases}
c_1\Sigma xy_x + c_2\Sigma y_x - c_3\Sigma xy_{x+1} - c_4\Sigma y_{x+1} = \Sigma x^2y_{x+1} - \Sigma x^2y_x \\
c_1\Sigma x^2y_x + c_2\Sigma xy_x - c_3\Sigma x^2y_{x+1} - c_4\Sigma xy_{x+1} = \Sigma x^3y_{x+1} - \Sigma x^3y_x \\
c_1\Sigma x^3y_x + c_2\Sigma x^2y_x - c_3\Sigma x^3y_{x+1} - c_4\Sigma x^2y_{x+1} = \Sigma x^4y_{x+1} - \Sigma x^4y_x \\
c_1\Sigma x^4y_x + c_2\Sigma x^3y_x - c_3\Sigma x^4y_{x+1} - c_4\Sigma x^3y_{x+1} = \Sigma x^5y_{x+1} - \Sigma x^5y_x
\end{cases}
$$

where Σ means $\sum\limits_{x=r}^{x=s-1}$.

If we desire to graduate that portion of a distribution lying between $x=r$ and $x=s$, we must first calculate from the known ungraduated frequencies $f_r, f_{r+1} \cdots f_{s-1}, f_s$ the numerical values of $\sum\limits_{x=r}^{x=s-1} x^nf_x$ and $\sum\limits_{x=r}^{x=s-1} x^nf_{x+1}$, corresponding to the $\sum\limits_{x=r}^{x=s-1} x^ny_x$ and $\sum\limits_{x=r}^{x=s-1} x^ny_{x+1}$ of equations I. Imposing the usual condition that the corresponding moments of the graduated and ungraduated frequencies must be identical, we obtain the numerical values of the coefficients of equations I. A simultaneous solution then yields the desired values of c_1, c_2, c_3 and c_4.

In the above the limits of summation are, as stated, $x=r$ and $x=s-1$. Clearly we could not sum to $x=s$, since such procedure would require a knowledge of the value of f_{s+1}, which is contrary to our assumption that the ungraduated frequencies from f_r to f_s, only, are known.

Although the values of Σx^nf_{s+1} can be computed in the same way as those of Σx^nf_s, yet this would be practically a duplication of work, since we may easily show that

$$
\sum_r^{s-1} x^nf_{x+1} = \overline{s-1}^{\,n} f_s - \overline{r-1}^{\,n} f_r + \sum_r^{s-1} x^nf_x - {}_nC_1\sum_r^{s-1} x^{n-1}f_x + {}_nC_2\sum_r^{s-1} x^{n-2}f_x - \text{etc.}
$$

In other words

$$
\text{II.}\quad\begin{cases}
\Sigma\ f_{x+1} = \underline{\quad\ } f_s - \underline{\quad\ } f_r + \Sigma\ f_x, \\
\Sigma x\, f_{x+1} = \overline{s-1}\, f_s - \overline{r-1}\, f_r + \Sigma x\, f_x - \Sigma\ f_x, \\
\Sigma x^2f_{x+1} = \overline{s-1}^{\,2} f_s - \overline{r-1}\, f_r + \Sigma x^2f_x - 2\Sigma\ xf_x + \Sigma\ f_x, \\
\Sigma x^3f_{x+1} = \overline{s-1}^{\,3} f_s - \overline{r-1}\, f_r + \Sigma x^3f_x - 3\Sigma x^2f_x + 3\Sigma xf_x - \Sigma f_x, \\
\Sigma x^4f_{x+1} = \overline{s-1}^{\,4} f_s - \overline{r-1}\, f_r + \Sigma x^4f_x - 4\Sigma x^3f_x + 6\Sigma x^2f_x - 4\Sigma xf_x + \Sigma f_x, \\
\Sigma x^5f_{x+1} - \Sigma x^5f_x = \overline{s-1}^{\,5} f_s - \overline{r-1}^{\,5} f_r - 5\Sigma x^4f_x + 10\Sigma x^3f_x - 10\Sigma x^2f_x + 5\Sigma xf_x - \Sigma f_x.
\end{cases}
$$

Formulae I and II enable us to graduate a section or "stump" of a distribution and are therefore more powerful than formulae that presuppose a complete distribution—that is, assume $y_r = y_s = 0$.

If we restrict ourselves likewise, and in the usual manner reduce the moments $\Sigma x^n f_x$ to unit frequency, we have, denoting

$$\frac{\Sigma x^n f_x}{\Sigma f_x} \text{ by } \nu_n'$$

$$v_1'c_1 + c_2 \quad - (v_1' - 1) c_3 - c_4 = -2v_1' + 1,$$
$$v_2'c_1 + v_1'c_2 - (v_2' - 2v_1' + 1) c_3 - (v_1' - 1) c_4 = -3v_2' + 3v_1' - 1,$$
$$v_3'c_1 + v_2'c_2 - (v_3' - 3v_2' + 3v_1' - 1) c_3 - (v_2' - 2v_1' + 1) c_4 = -4v_3' + 6v_2' - 4v_1' + 1,$$
$$v_4'c_1 + v_3'c_2 - (v_4' - 4v_3' + 6v_2' - 4v_1' + 1) c_3 - (v_3' - 3v_2' + 3v_1' - 1) c_4 = -5v_4' + 10v_3' - 10v_2' + 5v_1' - 1.$$

Lastly, if the moments v_n' be transferred to the mean by means of the relations

$$v_1 = 0$$
$$v_2 = v_2' - (v_1')^2,$$
$$v_3 = v_3' - 3v_2'v_1' + 2(v')^3,$$
$$v_4 = v_4' - 4v_3'v_1' + 6v_2'(v_1')^2 - 3(v_1')^4,$$

equations I become

$$\text{III.} \begin{cases} c_2 + c_3 - c_4 = 1, \\ v_2c_1 - (v_2 + 1) c_3 + c_4 = -3v_2 - 1, \\ v_3c_1 + v_2c_2 - (v_3 - 3v_2 - 1) c_3 - (v_2 + 1) c_4 = -4v_3 + 6v_2 + 1, \\ v_4c_1 + v_3c_2 - (v_4 - 4v_3 + 6v_2 + 1) c_3 - (v_3 - 3v_2 - 1) c_4 = -5v_4 + 10v_3 - 10v_2 - 1. \end{cases}$$

Solving III we have

$$\text{IV} \begin{cases} c_1 = \left(\frac{\nu_3}{\nu_2} - 1\right) \delta - 1, \\ c_2 = \nu_2(1 + 2\delta), \\ c_3 = \left(\frac{\nu_3}{\nu_2} + 1\right) \delta + 3, \\ c_4 = c_1 + c_2 + 3 + 2\delta = c_2 + c_3 - 1, \end{cases}$$

where we let

$$\beta_1 = \frac{\nu_3^2}{\nu_2^3},$$

$$\beta_2 = \frac{\nu_4}{\nu_2^2},$$

$$\delta = \frac{\beta_2 + 3 - \frac{1}{\nu_2}}{2\beta_2 - 3\beta_1 - 6 + \frac{1}{\nu_2}}.$$

It is interesting to compare these results with Pearson's.
If we change our difference equation

$$\frac{y_{x+1}}{y_x} = \frac{x^2 + c_1x + c_2}{x^2 + c_3x + c_4}$$

back to the form

$$\frac{\Delta y_x}{y_x} = \frac{a - x}{b_0 + b_1x + b_2x^2},$$

corresponding to Pearson's

$$\frac{1}{y}\frac{dy}{dx} = \frac{a - x}{b_0 + b_1x + b_2x^2}$$

we have, since

$$a = \frac{c_2 - c_4}{c_3 - c_1}, \qquad b_0 = \frac{c_4}{c_3 - c_1},$$

$$b_1 = \frac{c_3}{c_3 - c_1}, \qquad b_2 = \frac{1}{c_3 - c_1},$$

the following comparison:

TABLE I.

Finite Constants.	Const.	Pearson's Values.
$\dfrac{-\frac{\nu_3}{\nu_2}\left(\beta_2 + 3 - \frac{1}{\nu_2}\right)}{2\left(5\beta_2 - 6\beta_1 - 9 + \frac{1}{\nu_2}\right)} - \dfrac{1}{2}$	a	$\dfrac{-\frac{\nu_3}{\nu_2}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$
$-a + \dfrac{\nu_2\left(4\beta_2 - 3\beta_1 - \frac{1}{\nu_2}\right)}{2\left(5\beta_2 - 6\beta_1 - 9 + \frac{1}{\nu_2}\right)}$	b_0	$\dfrac{\nu_2(4\beta_2 - 3\beta_1)}{2(5\beta_2 - 6\beta_1 - 9)}$
$b_2 - a$	b_1	$-a$
$-\dfrac{\left(3\beta_1 - 2\beta_2 + 6 - \frac{1}{\nu_2}\right)}{2\left(5\beta_2 - 6\beta_1 - 9 + \frac{1}{\nu_2}\right)}$	b_2	$-\dfrac{(3\beta_1 - 2\beta_2 + 6)}{2(5\beta_2 - 6\beta_1 - 9)}$

SECTION IV.

The following problems will serve to illustrate various applications of the proposed method to statistical problems.

Example I.

The distribution of deaths due to old age in the U. S. Registration Area for the year 1910 is shown in Table II:

TABLE II.

Age Intervals.	Deaths Ungraduated.	x.	$\frac{y_{x+1}}{y_x}$.	y'_x.	Deaths y_x Graduated.
45–		−7.056	8.18444	100	
50–		−6.056	6.82875	818	4
55–		−5.056	5.32671	5,589	24
60–	189	−4.056	3.93885	29,777	127
65–	519	−3.056	2.80070	117,262	499
70–	1,379	−2.056	1.93457	328,416	1,396
75–	2,475	−1.056	1.30420	635,344	2,701
80–	3,716	− .056	.85728	828,618	3,523
85–	3,116	.944	.54555	710,359	3,020
90–	1,587	1.944	.33085	387,534	1,648
95–	443	2.944	.18502	128,216	545
100–	173*	3.944	.08792	23,723	101
105–		4.944	.02531	2,086	9
110–		5.944		53	
Total....	13,597			3,197,895	13,597

* Deaths 100 and over. In calculating moments treated as class 100—.

Taking the middle of class 80— as origin we have

$\Sigma f_x = 13{,}597,\quad v_1' = .056262,\quad v_2 = 2.34330,$

$\Sigma x f_x = 765,\quad v_2' = 2.34647,\quad v_3 = -.45240,$

$\Sigma x^2 f_x = 31{,}905,\quad v_3' = -.056704,\quad v_4 = 16.5051,$

$\Sigma x^3 f_x = -771,\quad v_4' = 16.4478,\quad \beta_1 = .0159,$

$\Sigma x^4 f_x = 223{,}641,\quad \beta_2 = 3.0058.$

$$c_1 = -18.041,$$
$$c_2 = 69.284,$$
$$c_3 = 14.526,$$
$$c_4 = 82.810.$$

Using these constants, the value of y_{x+1}/y_x, shown in Table II were calculated. An arbitrary radix of 100 at class 45— produces a total frequency of 3,197,895; hence each frequency, y_x', must be multiplied by the decimal equivalent of $\frac{13{,}597}{3{,}197{,}895}$. The final graduation, as shown in the last column, results.

Let us now consider the problem of calculating the number of deaths at each age, instead of within quinquennial groups.

Although an interpolation formula might be used, yet it is preferable to modify the original moments so that both the graduation and interpolation may be performed simultaneously. This may be done as follows:

Taking Δ_x as one year, it follows that the corresponding moments v_n may be obtained by multiplying the previously calculated ones by 5^n. This will, however, not alter the values of β_1 or β_2.

The new difference equation referred to the mean age of 82.28 as origin, becomes

$$\frac{y_{x+1}}{y_x} = \frac{x^2 + 618.46x - 36871}{x^2 - 8x - 36880}.$$

The results of this graduation are as follows:

TABLE III.

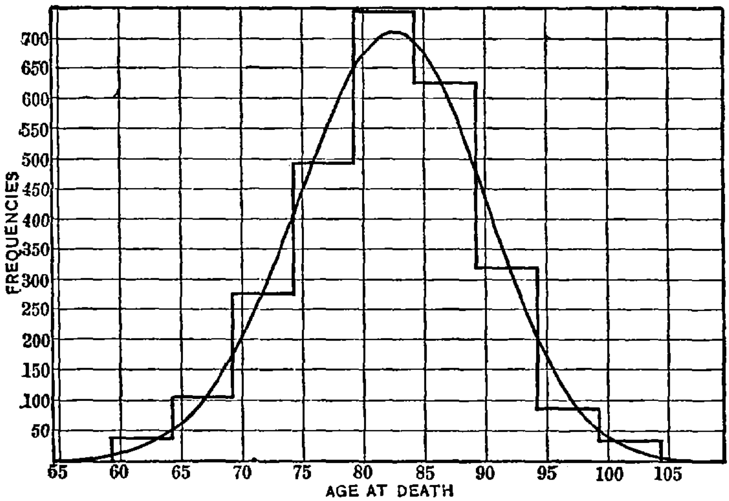
Age.	y_x.	Age.	y_x.	Age.	y_x.	Age.	y_x.	Age.	y_x.
52	1	64	46	76	437	88	557	100	42
53	1	65	60	77	539	89	503	101	30
54	1	66	78	78	587	90	446	102	20
55	2	67	100	79	629	91	387	103	13
56	3	68	127	80	664	92	330	104	9
57	5	69	158	81	690	93	275	105	5
58	7	70	193	82	705	94	225	106	3
59	10	71	234	83	708	95	180	107	2
60	13	72	279	84	699	96	141	108	1
61	19	73	328	85	679	97	108	109	1
62	26	74	380	86	647	98	81		
63	35	75	433	87	606	99	59	Total...	13,597

The results of the graduation, as shown graphically in Plate I, are entirely satisfactory.

A very important point which is always involved when the frequencies are associated with graduated variates, may be brought to light by comparing the grouped frequencies of Table III with the graduated frequencies of Table II. Column 5 of Table IV shows the group totals of Table III while 6 is merely the final column of Table II.

The discrepancies between grouped results are marked, and are due to the fact that we have been treating as ordinates these frequencies that in reality should be represented by areas. Thus, the frequency of class 80— should be represented by the area under

Plate I.



the curve from age 80 to age 85, instead of by the ordinate at age 82½.

In general we may state that a distribution of graduated variates should be represented by areas under a curve, while a distribution of integral variates should be considered as proportional to ordinates of a curve.*

Since approximately

$$\int_{x+\frac{1}{2}}^{x+\frac{1}{2}} y dx = \frac{1}{24} [y_{x-1} + 22y_x + y_{x+1}]$$

we may revise the computations of Table II as follows:

In the computing areas, a_x, the factor $\frac{1}{24}$ may be neglected, since the fraction $\frac{13,597}{76,749,327}$ automatically introduces it.

Columns 4 and 5 are now practically identical. The differences, which are slight when compared to the total frequency, may be attributed to the following causes:

TABLE IV.

Ages.	(1) $\frac{y_{x+1}}{y_x}$.	(2) y'_x.	(3) $y'_{x+1}+22y'_x+y'_{x+1}$.	(4). a_x.	(5) Grouped Frequencies from Table III	(6) y_x.
45–	8.18444	100	3,018	1		
50–	6.82875	818	23,635	4	3	4
55–	5.32671	5,589	153,553	27	27	24
60–	3.93885	29,777	777,945	138	139	127
65–	2.80070	117,262	2,937,957	520	523	499
70–	1.93457	328,416	7,977,758	1,413	1,414	1,396
75–	1.30420	635,344	15,134,602	2,681	2,675	2,701
80–	.85728	828,618	19,575,299	3,468	3,466	3,523
85–	.54555	710,359	16,844,050	2,984	2,992	3,020
90–	.33085	387,534	9,364,323	1,659	1,663	1,648
95–	.18502	128,216	3,232,009	573	569	545
100–	.08792	23,723	652,208	116	114	101
105–	.02531	2,086	69,638	12	12	9
110–		53	3,252	1		
			76,749,327	13,597	13,597	13,957

1. No quadrature formula was used in the calculation of the frequencies of Table III. If there be many classes, that is, if the class interval is small as compared with the visible range, the error involved by treating the areas as ordinates is slight. Thus, we treat the number of persons dying, as per a mortality table, as an ordinate (which practically assumes a uniform distribution of deaths throughout the year) without introducing an appreciable error. On the other hand, if our tables were based on a quinquennial, rather than an annual, basis, such an hypothesis would introduce a very considerable error, such as we just noted in comparing columns 5 and 6 of Table IV.

2. The quadrature formula used, though practically sufficient, is only approximate. Better ones, involving however a greater amount of labor, could be used if necessary; for example

$$\int_{x-\frac{1}{2}}^{x+\frac{1}{2}} y\,dx = \tfrac{1}{5760}\{5178y_x + 308(y_{x-1} + y_{x+1}) - 17(y_{x-2} + y_{x+2})\}$$

is somewhat better.

3. In changing from a quinquennial to an annual basis, the moment v_n should, strictly speaking, be modified. Sheppard's adjustments will, in general, slightly improve results, although their tendency in this case would be to "over adjust."

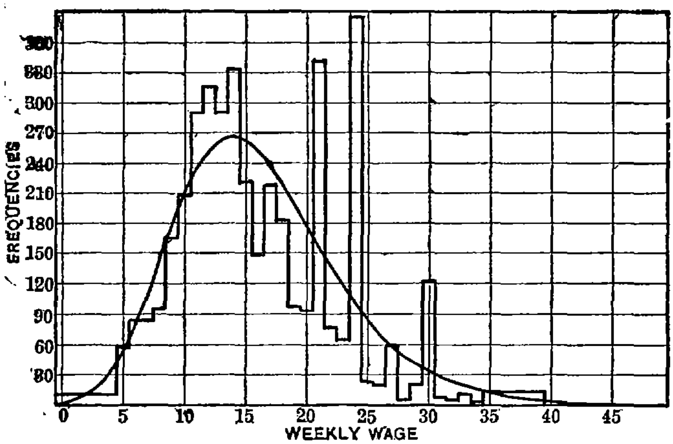
On the whole, if a quadrature formula must be used, the one we made use of is practically sufficient, and the modification of moments, moreover, may be entirely eliminated.

Example II.

WAGE DISTRIBUTION, ALL COMP. CLASSIFICATIONS, 1916.

Wage.	Ungraduated.	Graduated.	Wage.	Ungraduated.	Graduated.	Wage.	Ungraduated.	Graduated.
0–		2	18–	182	220	36–	(35–40)	10
1–	(Under 5)	4	19–	97	199	37–	68	9
2–	43	8	20–	94	178	38–		7
3–		16	21–	341	157	39–		6
4–		28	22–	77	137	40–	(40–50)	5
5–	59	47	23–	65	118	41–	9	4
6–	84	71	24–	384	101	42–		3
7–	84	101	25–	23	85	43–	(50–60)	2
8–	95	134	26–	21	72	44–	2	2
9–	160	168	27–	60	60	45–		2
10–	207	200	28–	6	50	46–	(60–)	1
11–	286	227	29–	20	42	47–	2	1
12–	311	248	30–	121	34	48–		1
13–	288	260	31–	8	28	49–		1
14–	332	265	32–	7	23			
15–	220	262	33–	11	19	Total . . .	4,138	4,138
16–	149	253	34–	4	16			
17–	218	238	35–		13			

Plate II.



$v_2 = 48.4321$, $\beta_1 = .8083$, $c_1 = 33.974$,

$v_3 = 303.0239$, $\beta_2 = 4.7857$, $c_2 = 692.899$,

$v_4 = 11225.6435$, $c_3 = 51.281$,

$c_4 = 743.180$.

Mean is .493 interval from central ordinate of class 16—.

Example III.

GRADUATION OF TYPE I DISTRIBUTION IN ELDERTON.

Central Age of Group.	Ungraduated Exposed to Risk.	Elderton's Graduation.	Present Graduation.	Elderton's $\frac{d^2}{y}$.	Present Grad. $\frac{d^2}{y}$.
17	34	44	44	2.27	2.27
22	145	137	142	.47	.06
27	156	149	151	.33	.17
32	145	142	142	.06	.06
37	123	127	126	.13	.07
42	103	108	106	.23	.08
47	86	88	87	.05	.01
52	71	69	68	.06	.13
57	55	51	50	.31	.50
62	37	36	36	.03	.03
67	21	24	23	.38	.17
72	13	14	14	.07	.07
77	7	7	7		
82	3	3	3		
87	1	1	1		
Total.......	1,000	1,000	1,000	$\chi^2 = 4.39$	$\chi^2 = 3.62$

Since there is not high contact for this distribution at the y_r end, we cannot properly assume $y_{r-1} = 0$. We should use, therefore, equations I and II.

Taking our provisional origin at class 37 we obtain $r-1=-5$, $f_r = 34$, $s-1=9$, $f_s = 1$, $\Sigma f_x = 999$, $\Sigma x f_x = 165$, $\Sigma x^2 f_x = 7593$, $\Sigma x^3 f_x = 18{,}135$ and $\Sigma x^4 f_x = 174{,}309$.

$$\begin{aligned}
165c_1 + 999c_2 + 655c_3 - 966c_4 &= -100,\\
7{,}593c_1 + 165c_2 - 7{,}493c_3 + 655c_4 &= -18{,}304,\\
18{,}135c_1 + 7{,}593c_2 + 169c_3 - 7{,}493c_4 &= -41{,}332,\\
174{,}309c_1 + 18{,}135c_2 - 132{,}977c_3 + 169c_4 &= -601{,}000.
\end{aligned}$$

In solving simultaneous equations of this type it is advisable to divide each equation through by the coefficient c_1 and then eliminate this unknown from the set by subtraction. This process should be repeated for c_2 and c_3.

The solution is

$$c_1 = -5.180514, \quad c_3 = -8.163185,$$

$$c_2 = -42.734192, \quad c_4 = -50.510486.$$

If we use χ^2 as the criterion for the goodness of fit, it is seen that a comparison somewhat favors the method of the difference equation.

Had we treated y_8 as $y_{11} = 0$, the results would have been practically the same: the graduated frequencies for classes 17 and 18 would have been, to the nearest integers 45 and 141 instead of 44 and 142, all others remaining unchanged.

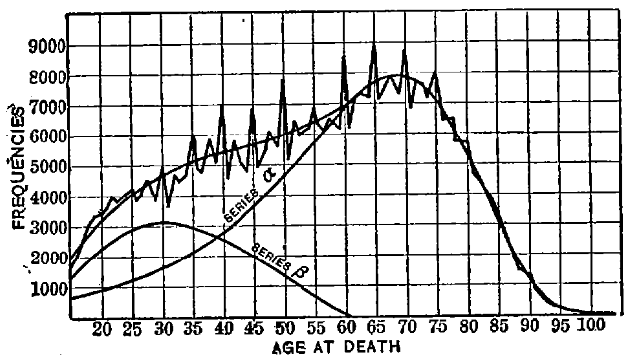
Example IV.

Example IV. Distribution of Deaths of Males in the O. R. S. for the Years 1909, 1910 and 1911.

The recorded deaths, as shown in Plate III, together with the estimated population as of July, 1910, were used in the construction of the U. S. Life Table, smoothing being done by grouping the frequencies within quinquennial intervals and redistributing with the aid of fifth-degree osculatory interpolation.

A glance at Plate III clearly indicates that we are dealing with

Plate III.



a compound frequency distribution. We may assume that there exists a distribution of deaths due to causes that produce the recorded deaths at the higher ages, Series α, and another which is responsible for additional deaths at the earlier ages, Series β. Without attempting to philosophize, I may point out that the range of this curve, ages 11 to 60 odd, may reflect on industrial and social conditions.

In our attempt to smooth these statistics, we shall further assume that all deaths after the age of 62 may be taken to belong to Series α. This is, of course, rather arbitrary. A glance at Plate III shows that the end of Series β is somewhere between 60 and

TABLE V.

Distribution of Deaths of Males in the O. R. S. for the Years 1909, 1910, 1911.

Age Last Birthday.	Ungraduated Deaths.	Graduated Deaths.		
		Total.	Series α.	Series β.
15	1,679	1,928.25	708.10	1,220.15
6	1,995	2,202.18	747.97	1,454.21
7	2,579	2,461.75	790.16	1,671.59
8	2,990	2,707.14	834.81	1,872.33
9	3,338	2,938.58	882.04	2,056.54
20	3,429	3,156.41	932.01	2,224.40
1	3,645	3,360.98	984.87	2,376.11
2	3,963	3,552.70	1,040.77	2,511.93
3	3,871	3,732.01	1,099.89	2,632.12
4	3,960	3,899.36	1,162.39	2,736.97
5	4,170	4,055.23	1,228.44	2,826.79
6	3,892	4,200.16	1,298.25	2,901.91
7	4,100	4,334.65	1,371.98	2,962.67
8	4,566	4,459.26	1,449.84	3,009.42
9	3,926	4,574.58	1,532.03	3,042.55
30	4,998	4,681.17	1,618.75	3,062.42
1	3,772	4,779.64	1,710.21	3,069.43
2	4,719	4,870.62	1,806.62	3,064.00
3	4,467	4,954.74	1,908.19	3,046.55
4	4,700	5,032.63	2,015.12	3,017.51
5	6,050	5,104.96	2,127.62	2,977.34
6	4,932	5,172.39	2,245.89	2,926.50
7	4,835	5,235.58	2,370.12	2,865.46
8	5,873	5,295.23	2,500.50	2,794.73
9	5,091	5,351.99	2,637.18	2,714.81
40	6,929	5,406.56	2,780.32	2,626.24
1	4,565	5,459.60	2,930.03	2,529.57
2	5,848	5,511.77	3,086.42	2,425.35
3	5,182	5,563.73	3,249.54	2,314.19
4	4,869	5,616.10	3,419.41	2,196.69
5	6,801	5,669.50	3,596.00	2,073.50
6	4,937	5,724.52	3,779.22	1,945.30
7	5,332	5,781.69	3,968.91	1,812.78
8	6,077	5,841.53	4,164.85	1,676.68
9	5,700	5,904.51	4,366.72	1,537.79
50	7,696	5,971.03	4,574.10	1,396.93
1	5,205	6,041.47	4,786.48	1,254.99
2	6,461	6,116.12	5,003.19	1,112.93
3	5,964	6,195.23	5,223.47	971.76
4	6,156	6,278.98	5,446.36	832.62
5	6,810	6,367.54	5,670.79	696.75
6	6,304	6,461.02	5,895.47	565.55
7	6,060	6,559.58	6,118.96	440.62
8	6,512	6,663.44	6,339.60	323.84
9	6,224	6,773.07	6,555.53	217.54
60	8,504	6,889.39	6,764.68	124.71
1	6,177	7,014.45	6,964.77	49.68
2	7,153	7,153.44	7,153.34	
3	7,284	7,327.68	7,327.68	
4	7,241	7,484.96	7,484.96	
5	9,000	7,622.17	7,622.17	

TABLE V.—(*Continued.*)

DISTRIBUTION OF DEATHS OF MALES IN THE O. R. S. FOR THE YEARS 1909, 1910, 1911.

Age Last Birthday.	Ungraduated Deaths.	Graduated Deaths. Total.	Series α.	Series β.
6	7,072	7,736.19	7,736.19	
7	7,504	7,823.85	7,823.85	
8	7,902	7,881.95	7,881.95	
9	7,436	7,907.41	7,907.41	
70	8,767	7,897.28	7,897.28	
1	6,861	7,848.88	7,848.88	
2	7,769	7,759.92	7,759.92	
3	7,659	7,628.61	7,628.61	
4	7,224	7,453.79	7,453.79	
5	7,911	7,235.06	7,235.06	
6	7,001	6,972.91	6,972.91	
7	6,427	6,668.83	6,668.83	
8	6,469	6,325.33	6,325.38	
9	5,814	5,946.27	5,946.27	
80	5,777	5,536.33	5,536.33	
1	4,750	5,101.48	5,101.48	
2	4,667	4,648.60	4,648.60	
3	4,319	4,185.37	4,185.37	
4	3,911	3,719.96	3,712.96	
5	3,417	3,260.79	3,260.79	
6	2,925	2,816.12	2,816.12	
7	2,379	2,393.67	2,393.67	
8	1,972	2,000.25	2,000.25	
9	1,588	1,641.39	1,641.39	
90	1,360	1,321.10	1,321.10	
1	972	1,041.67	1,041.67	
2	741	803.663	803.663	
3	549	609.950	609.950	
4	378	445.970	445.970	
5	284	320.028	320.028	
6	188	223.679	223.679	
7	155	152.127	152.127	
8	91	100.597	100.597	
9	81	64.6394	64.6394	
100	49	40.3431	40.3431	
1	25	24.4539	24.4539	
2	22	14.3984	14.3984	
3	17	8.23918	8.23918	
4	12	4.58610	4.58610	
5	10	2.48616	2.48616	
6	6	1.31478	1.31478	
7	3	.679630	.679630	
8	5	.344192	.344192	
9	3	.171226	.171226	
110	5	.0839097	.0839097	

70, assuming that the balance of the distribution belongs to a single curve. Since the ordinates at 62 and 63 appear to be closer to this curve than the others, because of the well-known systematic errors in statements regarding ages, this limiting section was chosen. I have found that it is possible to choose any point between 50 and 70, which is approximately on Series α; what is gained or lost on one series is made up on the other.

Taking the number of deaths at age 62–63 as f_r, and choosing age 94– as one provisional origin, we obtain by means of equations I and II the following difference equation

$$\frac{y_{x+1}}{y_x}=\frac{x^2-30.616x+589.440}{x^2-21.439x+821.405}.$$

The result of the graduation is the portion of Series α falling after age 62.

By calculating values of y_x/y_{x+1} for ages less than 62, the remainder of Series α is easily obtained. Taking the provisional origin at age 35—, the difference equation of the rough residual Series β must now be taken as

$$\frac{y_{x+1}}{y_x}=\frac{(x+c_1)(x-26)}{x^2+c_2x+c_3}$$

in order to effect a union of the two curves at the proper point. Where systematic errors are not present, the data is generally able to provide this point of union itself, but in problems where such large variations occur (the last three terms of the rough residual series are -332, $+1{,}739$ and -788) this must be provided for by the introduction of the proper factor. Incidentally, we are able to proceed now with but three moments instead of four. The elimination of this higher moment, $\Sigma x^4 f_x$, with its high probable error, is somewhat in our favor.

We proceed as follows:

$$\frac{y_{x+1}}{y_x}=\frac{x^2-26x+c_1(x-26)}{x^2+c_2x+c_3},$$

$$c_1(\Sigma xy_x-26\Sigma y_x)-c_2\Sigma xy_{x+1}-c_3\Sigma y_{x+1}$$
$$=\Sigma x^2y_{x+1}-\Sigma x^2y_x+26\Sigma y_x,$$

$$c_1(\Sigma x^2y_x-26\Sigma xy_x)-c_2\Sigma x^2y_{x+1}-c_3\Sigma xy_{x+1}$$
$$=\Sigma x^3y_{x+1}-\Sigma x^3y_x+26\Sigma xy_x,$$

$$c_1(\Sigma x^3y_x-26\Sigma x^2y_x)-c_2\Sigma x^3y_{x+1}-c_3\Sigma x^2y_{x+1}$$
$$=\Sigma x^4y_{x+1}-\Sigma x^4y_x+26\Sigma x^2y_x.$$

TABLE VI.

REGRADUATION OF THE U. S. LIFE TABLE FOR MALES IN THE O. R. S. BY THE METHOD OF COMPOUND CURVES.

Age	Fisher's Graduation—10 Curves (Modified Radix).	Life Table Ungraduated.	Present Graduation—2 Curves.	Graduated Deaths. Total d_x.	Series α.	Series β.
20	78,792	78,792	78,792	3,967	209	3,758
1	78,445	78,396	78,395	4,079	232	3,847
2	78,082	77,974	77,987	4,189	257	3,932
3	77,704	77,543	77,568	4,305	286	4,019
4	77,313	77,110	77,138	4,422	317	4,105
5	76,906	76,675	76,696	4,544	352	4,192
6	76,484	76,237	76,241	4,671	391	4,280
7	76,046	75,794	75,774	4,800	435	4,365
8	75,590	75,339	75,294	4,933	483	4,450
9	75,116	74,867	74,801	5,071	537	4,534
30	74,627	74,378	74,294	5,213	596	4,617
1	74,119	73,872	73,773	5,359	662	4,697
2	73,592	73,344	73,237	5,511	736	4,775
3	73,045	72,792	72,686	5,671	818	4,853
4	72,477	72,215	72,119	5,835	909	4,926
5	71,646	71,614	71,535	6,005	1,010	4,995
6	71,280	70,988	70,935	6,183	1,122	5,061
7	70,653	70,341	70,316	6,369	1,246	5,123
8	70,006	69,676	69,679	6,562	1,383	5,179
9	69,340	68,995	69,023	6,766	1,535	5,231
40	68,652	68,297	68,347	6,980	1,703	5,277
1	67,944	67,583	67,648	7,205	1,889	5,316
2	67,216	66,850	66,928	7,441	2,094	5,347
3	66,464	66,096	66,184	7,692	2,321	5,371
4	65,689	65,319	65,415	7,955	5,385	5,385
5	64,891	64,518	64,619	8,233	2,843	5,390
6	64,065	63,689	63,796	8,529	3,144	5,385
7	63,208	62,833	62,943	8,841	3,472	5,369
8	62,322	61,951	62,059	9,173	3,832	5,341
9	61,400	61,046	61,142	9,524	4,224	5,300
50	60,443	60,118	60,189	9,896	4,651	5,245
1	59,441	59,167	59,200	10,288	5,113	5,175
2	58,397	58,189	58,171	10,704	5,614	5,090
3	57,306	57,170	57,100	11,143	6,154	4,989
4	56,161	56,099	55,986	11,604	6,734	4,870
5	54,962	54,970	54,826	12,087	7,354	4,733
6	53,706	53,773	53,617	12,593	8,015	4,578
7	52,394	52,505	52,358	13,120	8,715	4,405
8	51,016	51,173	51,046	13,663	9,453	4,210
9	49,581	49,787	49,679	14,224	10,226	3,998
60	48,086	48,343	48,257	14,793	11,030	3,763
1	46,528	46,842	46,778	15,368	11,860	3,508
2	44,913	45,285	45,241	15,943	12,709	3,234
3	43,243	43,669	43,647	16,512	13,569	2,943
4	41,524	41,993	41,995	17,064	14,429	2,635
5	39,756	40,264	40,289	17,592	15,278	2,314
6	37,946	38,490	38,530	18,083	16,102	1,981
7	36,102	36,676	36,722	18,528	16,886	1,642
8	34,229	34,824	34,869	18,918	17,614	1,304
9	32,333	32,938	32,977	19,238	18,266	972

TABLE VI.—(*Concluded.*)

REGRADUATION OF THE U. S. LIFE TABLE FOR MALES IN THE O. R. S. BY THE METHOD OF COMPOUND CURVES.

Age.	Fisher's Graduation—10 Curves (Modified Radix).	Life Table Ungraduated.	Present Graduation—2 Curves.	Graduated Deaths.		
				Total d_x.	Series a.	Series β.
70	30,425	31,023	31,053	19,485	18,827	658
1	28,516	29,087	29,105	19,653	19,276	377
2	26,609	27,134	27,139	19,743	19,596	147
3	24,711	25,165	25,165	19,769	19,769	
4	22,838	23,188	23,188	19,781	19,781	
5	20,996	21,213	21,210	19,621	19,621	
6	19,197	19,246	19,248	19,283	19,283	
7	17,452	17,311	17,320	18,761	18,761	
8	15,766	15,438	15,443	18,060	18,060	
9	14,157	13,648	13,638	17,191	17,191	
80	12,620	11,942	11,918	16,169	16,169	
1	11,168	10,322	10,301	15,015	15,015	
2	9,808	8,804	8,800	13,757	13,757	
3	8,550	7,413	7,424	12,425	12,425	
4	7,389	6,165	6,182	11,057	11,057	
5	6,331	5,059	5,076	9,684	9,684	
6	5,374	4,093	4,103	8,343	8,343	
7	4,517	3,263	3,273	7,064	7,064	
8	3,756	2,562	2,567	5,873	5,873	
9	3,092	1,978	1,980	4,792	4,792	
90	2,515	1,502	1,501	3,835	3,835	
1	2,022	1,121	1,117	3,007	3,007	
2	1,605	821	816	2,310	2,310	
3	1,253	591	585	1,737	1,737	
4	964	417	412	1,278	1,278	
5	730	289	284	920	920	
6	541	196	192	647	647	
7	393	130	127	446	446	
8	280	84	83	300	300	
9	193	53	53	198	198	
100	129	33	33	127	127	
1	83	19	20	80	80	
2	51	11	12	50	50	
3	30	6	7	30	30	
4	17	3	4	18	18	
5	9	2	2	10	10	
6	4	1	1	6	6	
7	1			3	3	
8				2	2	
9				1	1	

The difference equation,

$$\frac{y_{x+1}}{y_x} = \frac{x^2 - .56149x - 661.40128}{x^2 - 3.14933x - 672.89232}$$

is the result of a simultaneous solution.

The lack of red tape involved resulting through application of the finite calculus is brought out by the above.

I regret that I cannot agree with Mr. Fisher's "Note on the Construction of Mortality Tables," *Proc.*, Vol. IV, to the extent that it is possible to construct mortality tables from only records of deaths.

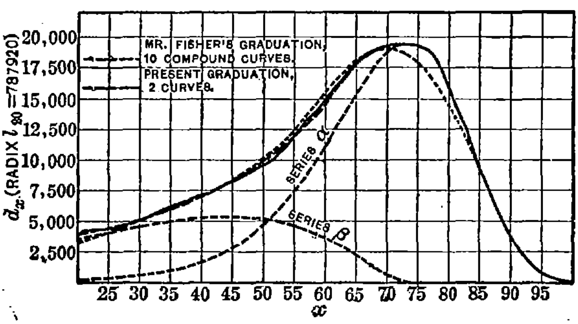
If we add the graduated deaths of Table V backwards, we will produce a series that approximates the l_x column of a mortality table. I believe this sort of thing could be done for a community which enjoyed a stationary population, and also not affected by immigration or emigration, but only in such an event.

However, we can graduate the populations as we just did the deaths and compute values of q_x by means of the formula

$$q_x = \frac{d_x}{L_x + \frac{1}{2}d_x}.$$

In order to bring out the fact that a graduation of the d_x column may be performed by breaking it up into but two curves, Table VI and Plate IV are added—showing the results obtained by the pro-

Plate IV.



posed method compared to the actual values as shown in the Life Tables prepared by Professor Glover and the regraduated values as computed by Mr. Fisher by means of ten compound curves.

SECTION V.

In conclusion I wish to bring out certain points bearing upon the systems devised by Pearson and Charlier which are, unfortunately and incorrectly, considered by many to be radically different, both as regards philosophic basis and effectiveness.

If we integrate Pearson's differential equation in the modified form

$$\frac{1}{y}\cdot\frac{dy}{dx} = \frac{a-x}{b_0}$$

we obtain as a solution the Gaussian or normal curve of error

$$y = \varphi_x = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-b)^2}{\sigma 2^2}}.$$

This curve, obviously symmetrical, graduates with a considerable degree of satisfaction, many distributions possessing but slight skewness.

To take care of skew distributions a function $f(x)$ is introduced *in the differential equation,* as before stated, giving

$$\frac{1}{y}\cdot\frac{dy}{dx} = \frac{a-x}{b_0\cdot f(x)} \quad \text{or}$$

$$\frac{1}{y}\cdot\frac{dy}{dx} = \frac{a-x}{b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots}.$$

Charlier's Type A curve is given, on the other hand, as

$$y = A_0\varphi_{(x)} + A_3\varphi_{(x)}^{\text{III}} + A_4\varphi_{(x)}^{\text{IV}} + A_5\varphi_{(x)}^{\text{V}} + \cdots,$$

where $\phi(x)$ is the same symmetric function that we obtained above from Pearson's differential equation

$$\frac{1}{y}\cdot\frac{dy}{dx} = \frac{a-x}{b_0}.$$

The type A curve is identically equivalent to

$$y = \phi_{(x)}\{1 + a_3[(x-b)^3 - 3c(x-b)] + a_4[(x-b)^4 - 6c(x-b)^2 + 3c^2] + \cdots\},$$

which may in turn be written as

$$y = \phi_{(x)} f(x) = \phi_{(x)}[b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots].$$

In effect, then, the basis of Pearson's system, and Charlier's Type A curve is the same symmetrical function $\phi_{(x)}$. Skewness is taken care of in *each* case by the introduction of an UNKNOWN function $f(x)$ which is represented as the converging power series $b_0 + b_1 x + b_2 x^2 + \cdots$. In one case this is introduced in the differential equation of the graduating curve—in the other, in the curve itself.

The basis of Charlier's Type B curve

$$y = B_0\psi(x) + B_1\Delta\psi(x) + B_2\Delta^2\psi(x) + \cdots$$

is

$$\psi(x) = \frac{e^{-m}m^x}{\lfloor x},$$

whose difference equation, from a fixed origin is

$$\frac{y_x}{y_{x+1}} = \frac{x+1}{m} \qquad \text{or} \qquad \frac{\Delta y_x}{y_{x+1}} = \frac{(m-1)-x}{m}.$$

this is of the type

$$\frac{\Delta y_x}{y_{x+1}} = \frac{a-x}{b_0},$$

which is quite similar to the corresponding difference equation of this paper.

Again, we might generalize the above by introducing a function $f(x)$ in the difference equation, giving

$$\frac{\Delta y_x}{y_{x+1}} = \frac{a-x}{b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots}.$$

Neglecting the terms involving the third and higher powers of x we have our difference equation

$$\frac{y_{x+1}}{y_x} = \frac{x^2 + c_1x + c_2}{x^2 + c_3x + c_4}.$$

Therefore we have in effect modified the basic curve $y = \psi(x)$ by introducing a function $f(x)$ in the *difference equation.* Charlier's Type B curve introduces $f(x)$ in the *curve* (or rather series) *itself* since the Type B curve can be written as

$$y = \psi(x)f(x) = \psi(x)\left[b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots\right].$$