GRADES AND DEVIATES

(INCLUDING A TABLE OF NORMAL DEVIATES CORRESPONDING TO EACH MILLESIMAL GRADE IN THE LENGTH OF AN ARRAY, AND A FIGURE).

BY FRANCIS GALTON, F.R.S.

THE Table is an amplification of one that I published a long time ago* under the title of "Per Centiles," and has been calculated for me (see p. 405) by Mr W. F. Sheppard. I should often have been glad to possess this enlarged Table for my own use to save the trouble of interpolation, and trust that it may be serviceable to others.

The values to be dealt with are supposed to be replaced by vertical lines of proportionate lengths, standing in a row, and arranged in order of their lengths beginning with the lowest, and they are set upon a horizontal base, at equal distances apart, between two termini. They thus form what is known as an "Array." Their upper outline, drawn with a free hand, forms what is called a Curve of Distribution, whose Axis is that horizontal line bounded by the termini, which passes through the top of the middlemost, or median line. When the variability of the array is normal, the median is also the mean.

The portion of the above lines, prolonged where necessary, that is intercepted between the curve and its axis, forms a system by itself; it represents an array of deviates from the median, and it is to these that the Table and the Figure refer. Where the line originally reached above the axis, and is therefore of more than median height, the corresponding deviate is positive: this is the case throughout the right half of the array. On the other hand, when the line originally reached only to the curve, and had to be prolonged in order to reach its axis, the prolongation shows its deficiency from the median, and corresponds to a negative deviate: this is the case throughout the left half of the array, and of the Figure.

* *Natural Inheritance*, by Francis Galton (Macmillan, 1889), pp. 202, 203.

In an array of n deviates, the point on the axis at which each of them stands, may be described in either of two ways:

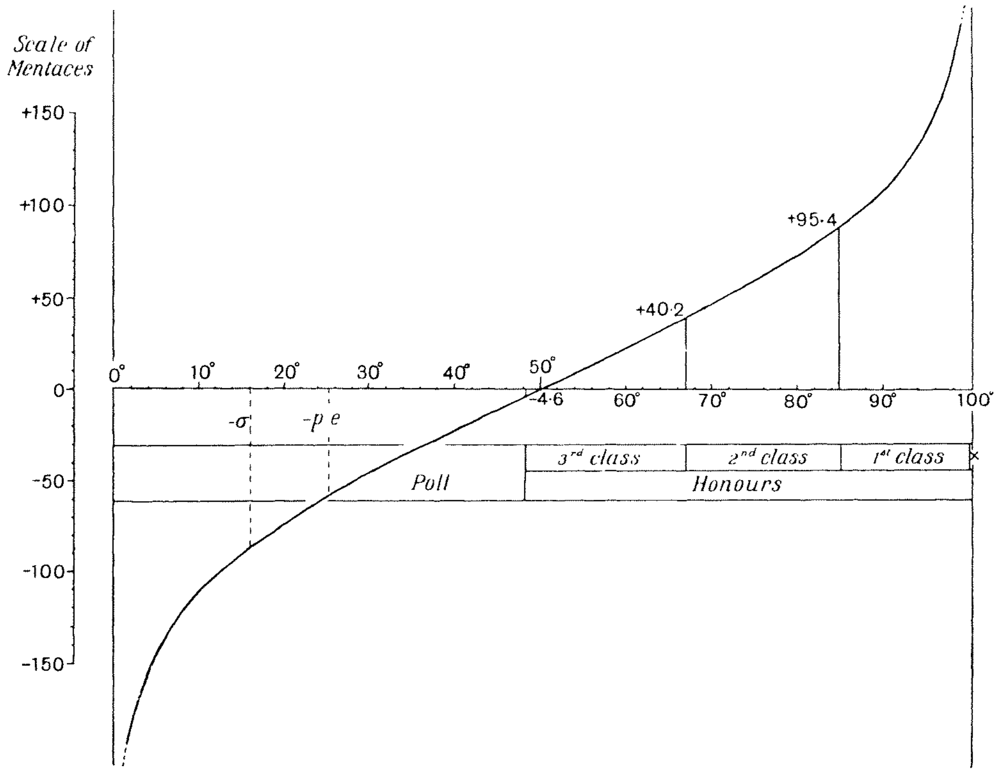
(1) by its order in the array; first (or lowest), second, third, ... rth ... nth,

(2) by the fraction of the length of the axis at which it stands. If the axis be divided into n equal parts, by grades, the first of which corresponds to the lower terminus and counts as $0°N$, and the last to the higher terminus, counting as $n°N$, then each deviate will stand midway between two grades; the first between $0°N$ and $1°N$, the second between $1°N$ and $2°N$, and the nth between $\overline{n-1}°N$ and $n°N$. Therefore the rth deviate in an array of n lines is equivalent to the grade of $\left(r-\frac{1}{2}\right)^{\circ}N$. This is translatable into a centesimal system as $\left\{\frac{100}{n}\left(r-\frac{1}{2}\right)\right\}^{\circ}C$, which is more conveniently treated as $\left(\frac{100r-50}{n}\right)^{\circ}C$. Thus the third deviate in an array of 25 deviates is $\frac{250}{25}=10°C$.

The difference is considerable between the number that describes the order and that which describes the grade, when n is small; it is less so when n is large, though the ratio between n and r may be the same in both cases. Thus the 2nd in an array of 10, the 20th in an array of 100, and the 200th in an array of 1000, stand respectively at the C grades 15°·0, 19°·50, and 19°·95 respectively.

The length of the axis of the curve of distribution is a *linear* representative of the *area* of the well known limpet-shaped curve of the Probability Integral, and the deviates are equal to the abscissæ, $\frac{x}{\sigma}$, of that curve. But as the length of the axis is taken as running from 0° to 100°, and the integral as running from 0 to ± 5·0, or else from 0 to 1, the values in the present Table are one hundred times those of the corresponding values in the Probability Table. The one is an inversion of the other, with the foregoing and the following modification, namely that the grades on the axis run from 0° to 100°, and not as in the ordinary tables (if multiplied by 100) from + 50 to − 50. As the − probable error cuts off the lowest quarter of the deviates, it stands at the grade of 25°·0 C, the + P.E. at 75°·0C, the median at 50°C. Also $-\sigma$ stands at 15°·86C, and $+\sigma$ at 84°·16C. This is partially but sufficiently shown in the Figure where, to avoid confusion of lines, the $+\sigma$ and + P.E. are left out.

The deviates in this Table are calculated on the usual basis of x/σ; those in *Natural Inheritance* above referred to, were calculated on that of $x/(\text{P.E.})$. This makes no difference in their internal proportions, or in the general form of the curve derived from them, the latter entries being identical with the former ones multiplied by 0·6745. The curve shown in the Figure is exact for *any* array of normal deviates, if it be measured by an appropriate vertical scale, namely such that if a be any specified deviate in the observed array and b be the corresponding deviate in the Figure, the multiplier of the latter must be b/a. It is of no importance that the deviates so compared should be the σ or the P.E., for *any* deviate serves the purpose equally well when the array is normal.



An array that is not normal can be adequately described by measuring a few of its deviates at well selected points. Deciles have been used for the purpose, that is the deviates at $10°C$, $20°C, \ldots 90°C$, but it is better as a rule to take more observations towards either end of the array where the changes are rapid and fewer about the middle, but the peculiarities of each several array may deserve special treatment.

The method of Grades and centesimal Deviates (or as they were formerly called "per centiles") is very convenient in dealing with groups of qualities that admit of a fairly good classification in order of merit by the judgement, though not by any numerical system of measurement. The values are arranged, and a few cases at and about each of a few well selected centesimal grades are described as fully as practicable.

Limits of classes have often to be determined; the question being of this form: "In an array of n values (normal or otherwise), those up to the rth inclusive are of the quality A, those of the $r+1$th and upwards are of quality B, what is the centesimal grade that separates A from B? In other words, what is the upper limit of the one and the lower limit of the other?"

First suppose a system, N, of n grades, then the rth deviate stands at $(r-\frac{1}{2})^\circ N$, and the $r+1$th at $(r+\frac{1}{2})^\circ N$. Consequently in the N system the desired value is $r^\circ N$, which translated into the centesimal system becomes $\frac{100r^\circ}{n}C$. It may be well to work out a simple example as follows.

In the memoir by Prof. Karl Pearson on "Relationship of Intelligence, &c.," *Biometrika,* Vol. v., the number of Cambridge graduates is given, p. 137, who passed their Examination in the Poll, in the third class, the second class, and the first class in recent years, as 487, 189, 182, and 153 respectively. He also takes the range of intelligence in the third and second classes collectively, as a convenient unit and equivalent to 100 "mentaces." Then, supposing the distribution of intelligence to be normal, it is required to find the limiting values of intelligence for the above several divisions. A slight addition to the conditions has been made here, which is justified by remarks in the memoir, namely by cutting off the extreme top of the first class and calling it "extra-first," or X, to include at the rate of about 1 in 1000 undergraduates. This gives a fourth limit to be appraised. The little problem is worked out in the accompanying table with sufficient fulness to make detailed explanation unnecessary, especially if it be studied in connection with the Figure. It will be seen in the fourth column under the heading t that the range of intelligence of the third and second classes collectively, extends from $-0{\cdot}050$ to $+1{\cdot}030 = 1{\cdot}08$ tabular units. This being taken equal to 100 mentaces, the second half of the fourth column, which is headed m, is derived from the entries in the first half under t, by multiplying them into $\frac{100}{108}$ or by 0·926 approximately.

TABLE I.

Classes	Number in each Class Total $= n$	Sums from lowest — Values of r	Centesimal Grades of r — $\frac{100r-50}{n}$	Normal Deviates corresponding to the Specified Grades of r		Range of the Classes in Mentaces (1·08 Tab. Units = 100 Mentaces)	
				Tabular Units t	Mentaces m	Classes	Mentaces m
Poll ...	487	487	48°·12	−0·050	− 4·6	3rd and 2nd	100·0
Third Class ...	189	676	66°·87	+0·434	+ 40·2	3rd	44·8
Second Class ...	182	858	84°·86	+1·030	+ 95·4	2nd	54·2
First, less X ...	152	1010	99°·88	+3·002	+277·0	1st, less X	181·6
X	1	1011	100°·00	+infinity	+infinity	X	above

It becomes easy now to deal with the series in any further way, always supposing it to be normal; the value of σ being equal to that of the deviate (irrespective of sign) at grade $15^\circ{\cdot}86\,C$ or at $84^\circ{\cdot}16\,C$, and so on.

The Figure is on a scale amply sufficient to explain the way in which the graphical solution of problems is to be effected, but it is much too small for careful work. It indicates, but does not clearly show the position of the deviate that marks off X, and it is not high enough to include the whole length of that deviate. Neither was there room without creating confusion, to insert the deviates $+\sigma$ and + P.E.

It would be convenient for occasional preliminary work, to possess a large slate or slab, on which the permanent features of the diagram were permanently inscribed.

To sum up the merits of the method of Grades and Deviates:

It establishes a centesimal scale of precedence, into which the order of any individual, in an array of any number of individuals and of any length, can be easily translated, and it gives the normal deviate at the grade which the individual occupies.

It easily defines the limiting values of successive classes of given numbers in a normal array.

It classifies objects that can be arrayed by judgement, though not by actual measurement.

It gives by inspection the value of σ in a normal series, and that of P.E. in any series, whether normal or not.

It exhibits processes under their real forms, and so is free from the danger of errors in principle, to which those unpractised in statistics are liable.

It affords an excellent criterion whether an observed array is or is not normal, and of the degree of its departure from normality.

Table of Deviates of the Normal Curve.

BY W. F. SHEPPARD.

This table gives to 4 decimal places the values of the deviation of the normal curve, the standard deviation being taken as the unit, for the ordinates which divide the area into 1000 equal parts or the "permilles" of frequency; i.e. it gives the values of x for $A = {\cdot}000, {\cdot}001, \ldots 1{\cdot}000$, where

$$A = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2}\, dx.$$

The values up to $A = {\cdot}800$ were obtained by interpolation from Table III. in *Biometrika*, Vol. II, pt. 2, pp. 189, 190. This table gives x in terms of α, where $A = \frac{1}{2}(1+\alpha)$. The remaining values were obtained from tables not yet published. In the latter part of the table, where the regularity could not be checked by inspection of differences, each value was calculated twice, by different methods.

Table of Deviates of the Normal Curve for each Permille of Frequency.

Permille	*·000*	*·001*	*·002*	*·003*	*·004*	*·005*	*·006*	*·007*	*·008*	*·009*	*·010*	
·00	∞	3·0902	2·8782	2·7478	2·6521	2·5758	2·5121	2·4573	2·4089	2·3656	2·3263	*·99*
·01	2·3263	2·2904	2·2571	2·2262	2·1973	2·1701	2·1444	2·1201	2·0969	2·0749	2·0537	*·98*
·02	2·0537	2·0335	2·0141	1·9954	1·9774	1·9600	1·9431	1·9268	1·9110	1·8957	1·8808	*·97*
·03	1·8808	1·8663	1·8522	1·8384	1·8250	1·8119	1·7991	1·7866	1·7744	1·7624	1·7507	*·96*
·04	1·7507	1·7392	1·7279	1·7169	1·7060	1·6954	1·6849	1·6747	1·6646	1·6546	1·6449	*·95*
·05	1·6449	1·6352	1·6258	1·6164	1·6072	1·5982	1·5893	1·5805	1·5718	1·5632	1·5548	*·94*
·06	1·5548	1·5464	1·5382	1·5301	1·5220	1·5141	1·5063	1·4985	1·4909	1·4833	1·4758	*·93*
·07	1·4758	1·4684	1·4611	1·4538	1·4466	1·4395	1·4325	1·4255	1·4187	1·4118	1·4051	*·92*
·08	1·4051	1·3984	1·3917	1·3852	1·3787	1·3722	1·3658	1·3595	1·3532	1·3469	1·3408	*·91*
·09	1·3408	1·3346	1·3285	1·3225	1·3165	1·3106	1·3047	1·2988	1·2930	1·2873	1·2816	*·90*
·10	1·2816	1·2759	1·2702	1·2646	1·2591	1·2536	1·2481	1·2426	1·2372	1·2319	1·2265	*·89*
·11	1·2265	1·2212	1·2160	1·2107	1·2055	1·2004	1·1952	1·1901	1·1850	1·1800	1·1750	*·88*
·12	1·1750	1·1700	1·1650	1·1601	1·1552	1·1503	1·1455	1·1407	1·1359	1·1311	1·1264	*·87*
·13	1·1264	1·1217	1·1170	1·1123	1·1077	1·1031	1·0985	1·0939	1·0893	1·0848	1·0803	*·86*
·14	1·0803	1·0758	1·0714	1·0669	1·0625	1·0581	1·0537	1·0494	1·0450	1·0407	1·0364	*·85*
·15	1·0364	1·0322	1·0279	1·0237	1·0194	1·0152	1·0110	1·0069	1·0027	0·9986	0·9945	*·84*
·16	0·9945	0·9904	0·9863	0·9822	0·9782	0·9741	0·9701	0·9661	0·9621	0·9581	0·9542	*·83*
·17	0·9542	0·9502	0·9463	0·9424	0·9385	0·9346	0·9307	0·9269	0·9230	0·9192	0·9154	*·82*
·18	0·9154	0·9116	0·9078	0·9040	0·9002	0·8965	0·8927	0·8890	0·8853	0·8816	0·8779	*·81*
·19	0·8779	0·8742	0·8705	0·8669	0·8633	0·8596	0·8560	0·8524	0·8488	0·8452	0·8416	*·80*
·20	0·8416	0·8381	0·8345	0·8310	0·8274	0·8239	0·8204	0·8169	0·8134	0·8099	0·8064	*·79*
·21	0·8064	0·8030	0·7995	0·7961	0·7926	0·7892	0·7858	0·7824	0·7790	0·7756	0·7722	*·78*
·22	0·7722	0·7688	0·7655	0·7621	0·7588	0·7554	0·7521	0·7488	0·7454	0·7421	0·7388	*·77*
·23	0·7388	0·7356	0·7323	0·7290	0·7257	0·7225	0·7192	0·7160	0·7128	0·7095	0·7063	*·76*
·24	0·7063	0·7031	0·6999	0·6967	0·6935	0·6903	0·6871	0·6840	0·6808	0·6776	0·6745	*·75*
·25	0·6745	0·6713	0·6682	0·6651	0·6620	0·6588	0·6557	0·6526	0·6495	0·6464	0·6433	*·74*
·26	0·6433	0·6403	0·6372	0·6341	0·6311	0·6280	0·6250	0·6219	0·6189	0·6158	0·6128	*·73*
·27	0·6128	0·6098	0·6068	0·6038	0·6008	0·5978	0·5948	0·5918	0·5888	0·5858	0·5828	*·72*
·28	0·5828	0·5799	0·5769	0·5740	0·5710	0·5681	0·5651	0·5622	0·5592	0·5563	0·5534	*·71*
·29	0·5534	0·5505	0·5476	0·5446	0·5417	0·5388	0·5359	0·5330	0·5302	0·5273	0·5244	*·70*
·30	0·5244	0·5215	0·5187	0·5158	0·5129	0·5101	0·5072	0·5044	0·5015	0·4987	0·4959	*·69*
·31	0·4959	0·4930	0·4902	0·4874	0·4845	0·4817	0·4789	0·4761	0·4733	0·4705	0·4677	*·68*
·32	0·4677	0·4649	0·4621	0·4593	0·4565	0·4538	0·4510	0·4482	0·4454	0·4427	0·4399	*·67*
·33	0·4399	0·4372	0·4344	0·4316	0·4289	0·4261	0·4234	0·4207	0·4179	0·4152	0·4125	*·66*
·34	0·4125	0·4097	0·4070	0·4043	0·4016	0·3989	0·3961	0·3934	0·3907	0·3880	0·3853	*·65*
·35	0·3853	0·3826	0·3799	0·3772	0·3745	0·3719	0·3692	0·3665	0·3638	0·3611	0·3585	*·64*
·36	0·3585	0·3558	0·3531	0·3505	0·3478	0·3451	0·3425	0·3398	0·3372	0·3345	0·3319	*·63*
·37	0·3319	0·3292	0·3266	0·3239	0·3213	0·3186	0·3160	0·3134	0·3107	0·3081	0·3055	*·62*
·38	0·3055	0·3029	0·3002	0·2976	0·2950	0·2924	0·2898	0·2871	0·2845	0·2819	0·2793	*·61*
·39	0·2793	0·2767	0·2741	0·2715	0·2689	0·2663	0·2637	0·2611	0·2585	0·2559	0·2533	*·60*
·40	0·2533	0·2508	0·2482	0·2456	0·2430	0·2404	0·2378	0·2353	0·2327	0·2301	0·2275	*·59*
·41	0·2275	0·2250	0·2224	0·2198	0·2173	0·2147	0·2121	0·2096	0·2070	0·2045	0·2019	*·58*
·42	0·2019	0·1993	0·1968	0·1942	0·1917	0·1891	0·1866	0·1840	0·1815	0·1789	0·1764	*·57*
·43	0·1764	0·1738	0·1713	0·1687	0·1662	0·1637	0·1611	0·1586	0·1560	0·1535	0·1510	*·56*
·44	0·1510	0·1484	0·1459	0·1434	0·1408	0·1383	0·1358	0·1332	0·1307	0·1282	0·1257	*·55*
·45	0·1257	0·1231	0·1206	0·1181	0·1156	0·1130	0·1105	0·1080	0·1055	0·1030	0·1004	*·54*
·46	0·1004	0·0979	0·0954	0·0929	0·0904	0·0878	0·0853	0·0828	0·0803	0·0778	0·0753	*·53*
·47	0·0753	0·0728	0·0702	0·0677	0·0652	0·0627	0·0602	0·0577	0·0552	0·0527	0·0502	*·52*
·48	0·0502	0·0476	0·0451	0·0426	0·0401	0·0376	0·0351	0·0326	0·0301	0·0276	0·0251	*·51*
·49	0·0251	0·0226	0·0201	0·0175	0·0150	0·0125	0·0100	0·0075	0·0050	0·0025	0·0000	*·50*
	·010	*·009*	*·008*	*·007*	*·006*	*·005*	*·004*	*·003*	*·002*	*·001*	*·000*	Permille

The final figure is "corrected" in all cases, doubtful values having been specially checked.

In entering the table we enter from the left-hand column and top row, if the permille is less than 500. For example, if the frequency below a particular value were 387 per thousand, the corresponding deviate would be $-0{\cdot}2871$, the number placed at the intersection of the ·38 row from left and ·007 column from top. The negative sign is always to be given when reading permilles below 500, because the deviate will be in defect of the mean, supposing increasing variates to be plotted as usual from left to right.

On the other hand if the permille be greater than 500 we enter the table from the right-hand column and bottom row. For example, if the permille be 748, the deviate is $+0{\cdot}6682$, the number placed at the intersection of the ·74 row from right and ·008 column from bottom of the table. The plus sign must be given, as the deviation is in excess of the mean, if the convention as to plotting variables has been observed.

If M be the median value, and $M-q_1$ and $M+q_2$ the upper and lower quartile values, the standard-deviation may be expressed in terms of the difference between the two quartiles by the formula:

$$\sigma = \tfrac{1}{2}(q_1+q_2) \div {\cdot}67449 = 1{\cdot}48260 \times \tfrac{1}{2}(q_1+q_2).$$

This is on the assumption that the distribution is actually normal, and that the difference between q_1 and q_2, which are the deviations of the two quartiles from the median, is due to errors of random sampling. To express the deviations from the median in terms of the quartile deviation, the numerical values of the deviate as given by the table must be divided by ·67449, or multiplied by 1·48260.

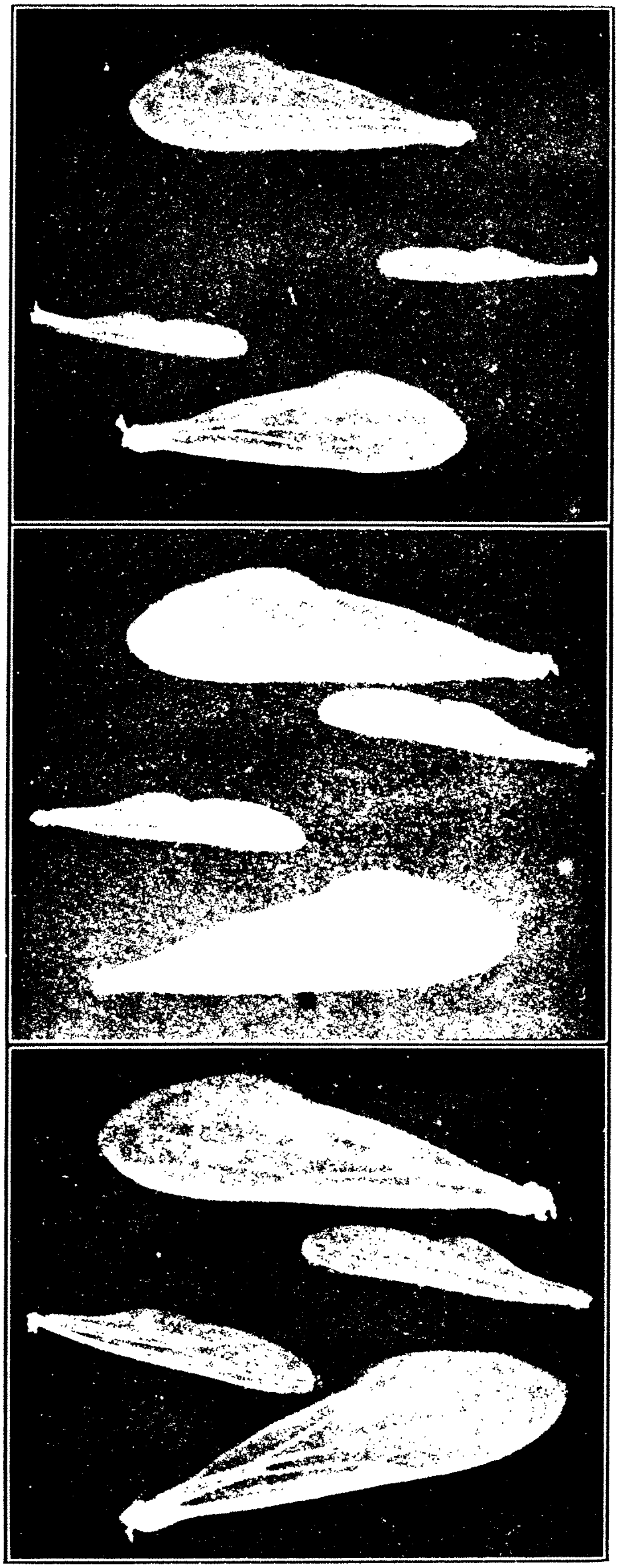
Note. The deviate x as given in the Table may be regarded as the ordinate of a graph whose abscissa is A. Since

$$\int_\alpha^\beta x\,dA = \int_{x_\alpha}^{x_\beta} x\frac{dA}{dx}\,dx = \int_{x_\alpha}^{x_\beta} xz\,dx = -\int_{x_\alpha}^{x_\beta}\frac{dz}{dx}\,dx = z_\alpha - z_\beta,$$

the area of this graph, comprised between any two ordinates, is equal to the difference of the corresponding ordinates of the normal curve. In applying this, it must be observed that the ordinate of the graph, and therefore also its area, is negative from $A=0$ to $A=\frac{1}{2}$, and that ordinates at equal distances from the centre (where the ordinate is zero) are numerically equal but of opposite sign, so that the total area of the graph is algebraically zero. Numerically, the area of either half of the graph is equal to $z_{A=\frac{1}{2}}$, i.e. to $1/\sqrt{2\pi}$, so that the total area of the graph, counting all ordinates as positive, is $\sqrt{2/\pi} = {\cdot}7979$. If α and β are both less than $\frac{1}{2}$ or both greater than $\frac{1}{2}$, the area of the graph between $A=\alpha$ and $A=\beta$ is $z_\alpha - z_\beta$; but, if α is less than $\frac{1}{2}$ and β greater than $\frac{1}{2}$, the area, taking all ordinates as positive, is

$$(1/\sqrt{2\pi} - z_\alpha) + (1/\sqrt{2\pi} - z_\beta) = \sqrt{2/\pi} - (z_\alpha + z_\beta).$$

The value of z corresponding to any value of x may be found from Table II. on pages 182 to 188 of *Biometrika*, Vol. II. Part 2. Thus, to find how much of the area of the graph lies outside the limits $A={\cdot}001$ and $A={\cdot}999$, we have $x=3{\cdot}0902$ for $A={\cdot}999$; and this gives $z={\cdot}00337$. The ratio of the portion of the graph outside the above limits to the total area is therefore ${\cdot}00337 \div {\cdot}39894 = {\cdot}0084$.



Worker

Drone

Queen

WASP WINGS.