CHAPTER IV.

SCHEMES OF DISTRIBUTION AND OF FREQUENCY.

Fraternities and Populations to be treated as Units.—Schemes of Distribution and their Grades.—The Shape of Schemes is independent of the number of observations.—Data for Eighteen Schemes.—Application of the method of Schemes to inexact Measures.—Schemes of Frequency.

Fraternities and Populations to be Treated as Units.—The science of heredity is concerned with Fraternities and large Populations rather than with individuals, and must treat them as units. A compendious method is therefore requisite by which we may express the distribution of each faculty among the members of any large group, whether it be a Fraternity or an entire Population.

The knowledge of an average value is a meagre piece of information. How little is conveyed by the bald statement that the average income of English families is 100*l.* a year, compared with what we should learn if we were told how English incomes were distributed; what proportion of our countrymen had just and only just enough means to ward off starvation, and what were the proportions of those who had incomes in each and every other degree, up to the huge annual receipts of a few great speculators, manufacturers, and landed proprietors. So in respect to the distribution of any human quality or faculty, a knowledge of mere averages tells but little; we want to learn how the quality is distributed among the various members of the Fraternity or of the Population, and to express what we know in so compact a form that it can be easily grasped and dealt with. A parade of great accuracy is foolish, because precision is unattainable in biological and social statistics; their results being never strictly constant. Over-minuteness is mischievous, because it overwhelms the mind with more details than can be compressed into a single view. We require no more than a fairly just and comprehensive method of expressing the way in which each measurable quality is distributed among the members of any group, whether the group consists of brothers or of members of any particular social, local, or other body of persons, or whether it is co-extensive with an entire nation or race.

A knowledge of the distribution of any quality enables us to ascertain the Rank that each man holds among his fellows, in respect to that quality. This is a valuable piece of knowledge in this struggling and competitive world, where success is to the foremost, and failure to the hindmost, irrespective of absolute efficiency. A blurred vision would be above all price to an individual man in a nation of blind men, though it would hardly enable him to earn his bread elsewhere. When the distribution of any faculty has been ascertained, we can tell from the measurement, say of our child, how he ranks among other children in respect to that faculty, whether it be a physical gift, or one of health, or of intellect, or of morals. As the years go by, we may learn by the same means whether he is making his way towards the front, whether he just holds his place, or whether he is falling back towards the rear. Similarly as regards the position of our class, or of our nation, among other classes and other nations.

Schemes of Distribution and their Grades.—I shall best explain my graphical method of expressing Distribution, which I like the more, the more I use it, and which I have latterly much developed, by showing how to determine the Grade of an individual among his fellows in respect to any particular faculty. Suppose that we have already put on record the measures of many men in respect to Strength, exerted as by an archer in pulling his bow, and tested by one of Salter's well-known dial instruments with a movable index. Some men will have been found strong and others weak; how can we picture in a compendious diagram, or how can we define by figures, the distribution of this faculty of Strength throughout the group? How shall we determine and specify the Grade that any particular person would occupy in the group? The first step is to marshal our measures in the orderly way familiar to statisticians, which is shown in Table I. I usually work to about twice its degree of minuteness, but enough has been entered in the Table for the purpose of illustration, while its small size makes it all the more intelligible.

The fourth column of the Table headed "Percentages" of "Sums from the beginning," is pictorially translated into Fig. 2, and the third column headed "Percentages" of "No. of cases observed," into Fig. 3. The scale of

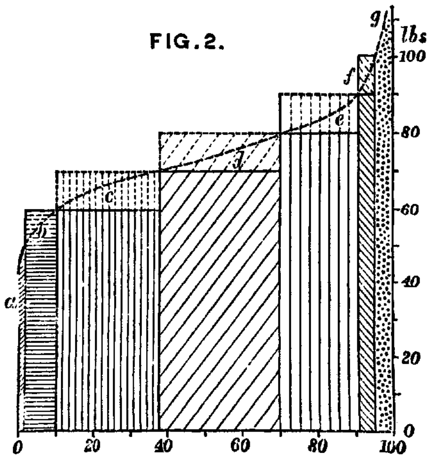


FIG. 2.

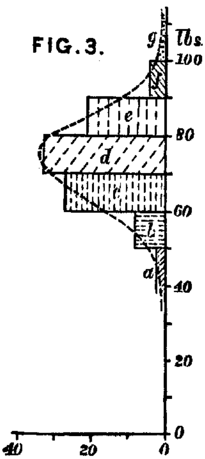


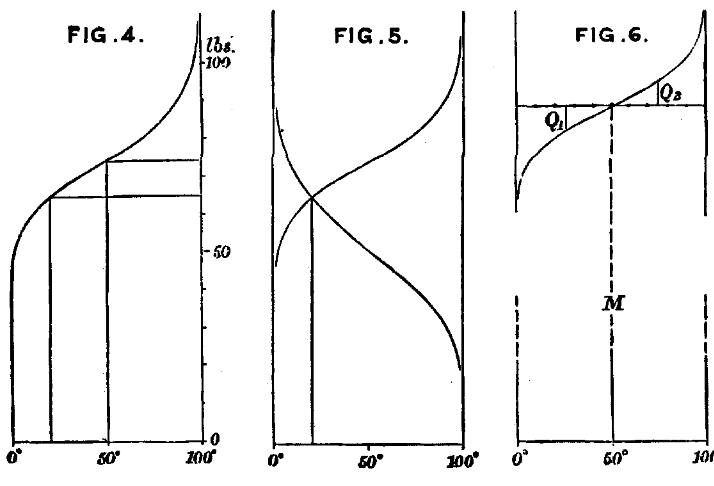
FIG. 3.

lbs. is given at the side of both Figs.: and the compartments *a* to *g*, that are shaded with *broken* lines, have the same meaning in both, but they are differently disposed in the two Figs. We will now consider Fig. 2 only, which is the one that principally concerns us. The percentages in the last column of Table I. have been marked off on the bottom line of Fig. 2, where they are called (centesimal) Grades. The number of lbs. found in the first column of the Table determines the height of the vertical lines to be erected at the corresponding Grades when we are engaged in constructing the Figure.

Let us begin with the third line in the Table for illustration: it tells us that 37 per cent. of the group had Strengths less than 70 lbs. Therefore, when drawing the figure, a perpendicular must be raised at the 37th grade to a height corresponding to that of 70 lbs. on the side scale. The fourth line in the Table tells us that 70 per cent. of the group had Strengths less than 80 lbs.; therefore a perpendicular must be raised at the 70th Grade to a height corresponding to 80 lbs. We proceed in the same way with respect to the remaining figures, then we join the tops of these perpendiculars by straight lines.

As these observations of Strength have been sorted into only 7 groups, the trace formed by the lines that connect the tops of the few perpendiculars differs sensibly from a flowing curve, but when working with double minuteness, as mentioned above, the connecting lines differ little to the eye from the dotted curve. The dotted curve may then be accepted as that which would result if a separate perpendicular had been drawn for every observation, and if permission had been given to slightly smooth their irregularities. I call the figure that is bounded by such a curve as this, a Scheme of Distribution; the perpendiculars that formed the scaffolding by which it was constructed having been first rubbed out. (See Fig 4, next page.)

A Scheme enables us in a moment to find the Grade of Rank (on a scale reckoned from 0° to 100°) of any person in the group to which he belongs. The measured strength of the person is to be looked for in the side scale of the Scheme; a horizontal line is thence drawn until it meets the curve; from the point of meeting a perpendicular is dropped upon the scale of Grades at the base; then the Grade on which it falls is



the one required. For example: let us suppose the Strength of Pull of a man to have been 74 lbs., and that we wish to determine his Rank in Strength among the large group of men who were measured at the Health Exhibition in 1884. We find by Fig. 4 that his centesimal Grade is 50°; in other words, that 50 per cent. of the group will be weaker than he is, and 50 per cent. will be stronger. His position will be exactly Middlemost, after the Strengths of all the men in the group have been marshalled in the order of their magnitudes. In other words, he is of mediocre strength. The accepted term to express the value that occupies the Middlemost position is "Median," which may be used either as an adjective or as a substantive, but it will be usually replaced in this book by the abbreviated form **M**. I also use the word "Mid" in a few combinations, such as "Mid-Fraternity," to express the same thing. The Median, **M**, has three properties. The first follows immediately from its construction, namely, that the chance is an equal one, of any previously unknown measure in the group exceeding or falling short of **M**. The second is, that the most probable value of any previously unknown measure in the group is **M**. Thus if N be any one of the measures, and u be the value of the unit in which the measure is recorded, such as an inch, tenth of an inch, &c., then the number of measures that fall between $(N - \frac{1}{2}u)$ and $(N + \frac{1}{2}u)$, is greatest when $N = \mathsf{M}$. Mediocrity is always the commonest condition, for reasons that will become apparent later on. The third property is that whenever the curve of the Scheme is symmetrically disposed on either side of **M**, except that one half of it is turned upwards, and the other half downwards, then **M** is identical with the ordinary Arithmetic Mean or Average. This is closely the condition of all the curves I have to discuss. The reader may look on the Median and on the Mean as being practically the same things, throughout this book.

It must be understood that M, like the Mean or the Average, is almost always an interpolated value, corresponding to no real measure. If the observations were infinitely numerous its position would not differ more than infinitesimally from that of some one of them; even in a series of one or two hundred in number, the difference is insignificant.

Now let us make our Scheme answer another question. Suppose we want *to know the percentage of men in the* group of which we have been speaking, whose Strength lies between any two specified limits, as between 74 lbs. and 64 lbs. We draw horizontal lines (Fig. 4) from points on the side scale corresponding to either limit, and drop perpendiculars upon the base, from the points where those lines meet the curve. Then the number of Grades in the intercept is the answer. The Fig. shows that the number in the present case is 30; therefore 30 per cent. of the group have Strengths of Pull ranging between 74 and 64 lbs.

We learn how to transmute female measures of any characteristic into male ones, by comparing their respective schemes, and devising a formula that will change the one into the other. In the case of Stature, the simple multiple of 1·08 was found to do this with sufficient precision.

If we wish to compare the average Strengths of two different groups of persons, say one consisting of men and the other of women, we have simply to compare the values at the 50th Grades in the two schemes. For even if the Medians differ considerably from the Means, both the ratios and the differences between either pair of values would be sensibly the same.

A different way of comparing two Schemes is sometimes useful. It is to draw them in opposed directions, as in Fig. 5, p. 40. Their curves will then cut each other at some point, whose Grade when referred to either of the two Schemes (whichever of them may be preferred), determines the point at which the same values are to be found. In Fig. 5, the Grade in the one Scheme is 20°; therefore in the other Scheme it is 100°−20°, or 80°. In respect to the Strength of Pull of men and women, it appears that the woman who occupies the Grade of 96° in her Scheme, has the same strength as the man who occupies the Grade of 4° in his Scheme.

I should add that this great inequality in Strength between the sexes, is confirmed by other measurements made at the same time in respect to the Strength of their Squeeze, as tested by another of Salter's instruments. Then the woman in the 93rd and the man in the 7th Grade of their resective Schemes, proved to be of equal strength. In my paper[1] on the results obtained at the laboratory, I remarked: "Very powerful women exist, but happily perhaps for the repose of the other sex such gifted women are rare. Out of 1,657 adult women of all ages measured at the laboratory, the strongest could only exert a squeeze of 86 lbs., or about that of a medium man."

[1] *Journ. Anthropol. Inst.* 1885. *Mem.:* There is a blunder in the paragraph, p. 23, headed "Height Sitting and Standing." The paragraph should be struck out.

The Shape of Schemes is Independent of the Number of Observations.—When Schemes are drawn from different samples of the same large group of measurements, though the number in the several samples may differ greatly, we can always so adjust the horizontal scales that the breadth of the several Schemes shall be uniform. Then the shapes of the Schemes drawn from different samples will be little affected by the number of observations used in each, supposing of course that the numbers are never too small for ordinary statistical purposes. The only recognisable differences between the Schemes will be, that, if the number of observations in the sample is very large, the upper margin of the Scheme will fall into a more regular curve, especially towards either of its limits. Some irregularity will be found in the above curve of the Strength of Pull; but if the observations had been ten times more numerous, it is probable, judging from much experience of such curves, that the irregularity would have been less conspicuous, and perhaps would have disappeared altogether.

However numerous the observations may be, the curve will always be uncertain and incomplete at its extreme ends, because the next value may happen to be greater or less than any one of those that preceded it. Again, the position of the first and the last observation, supposing each observation to have been laid down separately, can never coincide with the adjacent limit. The more numerous the observations, and therefore the closer the perpendiculars by which they are represented, the nearer will the two extreme perpendiculars approach the limits, but they will never actually touch them. A chess board has eight squares in a row, and eight pieces may be arranged in order on any one row, each piece occupying the centre of a square. Let the divisions in the row be graduated, calling the boundary to the extreme left, 0°. Then the successive divisions between the squares will be 1°, 2°, 3°, up to 7°, and the boundary to the extreme right will be 8°. It is clear that the position of the first piece lies half-way between the grades (in a scale of eight grades) of 0° and 1°; therefore the grade occupied by the first piece would be counted on that scale as 0·5°; also the grade of the last piece as 7·5°. Or again, if we had 800 pieces, and the same number of class-places, the grade of the first piece, in a scale of 800 grades, would exceed the grade 0°, by an amount equal to the width of one half-place on that scale, while the last of them would fall short of the 800th grade by an equal amount. This half-place has to be attended to and allowed for when schemes are constructed from comparatively few observations, and always when values that are very near to either of the centesimal grades 0° or 100° are under observation; but between the centesimal grades of 5° and 95° the influence of a half class-place upon the value of the corresponding observation is insignificant, and may be disregarded. It will not henceforth be necessary to repeat the word centesimal. It will be always implied when nothing is said to the contrary, and nothing henceforth will be said to the contrary. The word will be used for the last time in the next paragraph.

Data for Eighteen Schemes.—Sufficient data for reconstructing any Scheme, with much correctness, may be printed in a single line of a Table, and according to a uniform plan that is suitable for any kind of values. The measures to be recorded are those at a few definite Grades, beginning say at 5°, ending at 95°, and including every intermediate tenth Grade from 10° to 90°. It is convenient to add those at the Grades 25° and 75°, if space permits. The former values are given for eighteen different Schemes, in Table 2. In the memoir from which that table is reprinted, the values at what I now call (centesimal) Grades, were termed Percentiles. Thus the values at the Grades 5° and 10° would be respectively the 5th and the 10th percentile. It still seems to me that the word percentile is a useful and expressive abbreviation, but it will not be necessary to employ it in the present book. It is of course unadvisable to use more technical words than is absolutely necessary, and it will be possible to get on without it, by the help of the new and more important word "Grade."

A series of Schemes that express the distribution of various faculties, is valuable in an anthropometric laboratory, *for they enable every person who is measured to* find his Rank or Grade in each of them.

Diagrams may also be constructed by drawing parallel lines, each divided into 100 Grades, and entering each round number of inches, lbs., &c., at their proper places. A diagram of this kind is very convenient for reference, but it does not admit of being printed; it must be drawn or lithographed. I have constructed one of these from the 18 Schemes, and find it is easily understood and much used at my laboratory.

Application of Schemes to Inexact Measures.—Schemes of Distribution may be constructed from observations that are barely exact enough to deserve to be called measures.

I will illustrate the method of doing so by marshalling the data contained in a singularly interesting little memoir written by Sir James Paget, into the form of such a Scheme. The memoir is published in vol. v. of St. Bartholomew's Hospital Reports, and is entitled "What Becomes of Medical Students." He traced with great painstaking the career of no less than 1,000 pupils who had attended his classes at that Hospital during various periods and up to a date 15 years previous to that at which his memoir was written. He thus did for St. Bartholomew's Hospital what has never yet been done, so far as I am aware, for any University or Public School, whose historians count the successes and are silent as to the failures, giving to inquirers no adequate data for ascertaining the real value of those institutions in English Education. Sir J. Paget divides the successes of his pupils in their profession into five grades, all of which he carefully defines; they are *distinguished; considerable; moderate; very limited success;* and *failures.* Several of the students had left the profession either before or after taking their degrees, usually owing to their unfitness to succeed, so after analysing the accounts of them given in the memoir, I drafted several into the list of failures and distributed the rest, with the result that the number of cases in the successive classes, amounting now to the full total of 1,000, became 28, 80, 616, 151, and 125. This differs, I should say, a little from the inferences of the author, but the matter is here of small importance, so I need not go further into details.

If a Scheme is drawn from these figures, in the way described in page 39, it will be found to have the characteristic shape of our familiar curve of Distribution. If we wished to convey the utmost information that this Scheme is capable of giving, we might record in much detail the career of two or three of the men who are clustered about each of a few selected Grades, such as those that are used in Table II., or fewer of them. I adopted this method when estimating the variability of the Visualising Power (*Inquiries into Human Faculty*). My data were very lax, but this method of treatment got all the good out of them that they possessed. In the present case, it appears that towards the foremost of the successful men within fifteen years of taking their degrees, stood the three Professors of Anatomy at Oxford, Cambridge, and Edinburgh; that towards the bottom of the failures, lay two men who committed suicide under circumstances of great disgrace, and lowest of all Palmer, the Rugeley murderer, who was hanged.

We are able to compare any two such Schemes as the above, with numerical precision. The want of exactness in the data from which they are drawn, will of course cling to the result, but no new error will be introduced by the process of comparison. Suppose the second Scheme to refer to the successes of students from another hospital, we should draw the two Schemes in opposed directions, just as was done in the Strength of Pull of Males and Females, Fig. 5, and determine the Grade in either of the Schemes at which success was equal.

Schemes of Frequency.—The method of arranging observations in an orderly manner that is generally employed by statisticians, is shown in Fig. 3, page 38, which expresses the same facts as Fig. 2 under a different aspect, and so gives rise to the well-known Curve of "Frequency of Error," though in Fig. 3 the curve is turned at right angles to the position in which it is usually drawn. It is so placed in order to show more clearly its relation to the Curve of Distribution. The Curve of Frequency is far less convenient than that of Distribution, for the purposes just described and for most of those to be hereafter spoken of. But the Curve of Frequency has other uses, of which advantage will be taken later on, and to which it is unnecessary now to refer.

A Scheme as explained thus far, is nothing more than a compendium of a mass of observations which, on being marshalled in an orderly manner, fall into a diagram whose contour is so regular, simple, and bold, as to admit of being described by a few numerals (Table 2), from which it can at any time be drawn afresh. The regular distribution of the several faculties among a large population is little disturbed by the fact that its members are varieties of different types and sub-types. So the distribution of a heavy mass of foliage gives little indication of its growth from separate twigs, of separate branches, of separate trees.

The application of theory to Schemes, their approximate description by only two values, and the properties of their bounding Curves, will be described in the next chapter.