Edited by Howard Eves, University of Maine, Orono, Maine

*The development of modern statistics**

by Dale E. Varberg, Hamline University, St. Paul, Minnesota

That area of study which we now call statistics has only recently come of age. While its origins may be traced back to the eighteenth century, or perhaps earlier, the first really significant developments in the theory of statistics did not occur until the late nineteenth and early twentieth centuries, and it is only during the last thirty years or so that it has reached a full measure of respectability. It was antedated by the theory of probability and has its roots embedded in this subject. In fact, any serious study of statistics must of necessity be preceded by a study of probability theory, for it is in the latter subject that the theory of statistics finds its foundation and fountainhead.

The word *statistik* was first used by Gottfried Achenwall (1719–72), a lecturer at the University of Göttingen [1].** He is sometimes referred to as the "Father of Statistics"—perhaps mistakenly, since he was mainly concerned with the description of interesting facts about his country.

Our English word "statistic" means different things to different people. To the man on the street, statistics is the mass of figures that the expert on any subject uses to support his contentions—it's "what you use to prove anything by." To the more sophisticated person, the word may evoke some notion of the procedures which are used to condense and interpret a collection of data, such as the computing of means and standard deviations. But to the practitioner of the craft, statistics is the art of making inferences from a body of data, or, more generally, the science of making decisions in the face of uncertainty.

Statisticians concern themselves with answering such questions as: Is this particular lot of manufactured items defective? Is there a connection between smoking and cancer? Will Kennedy win the next election? In answering these questions, it is necessary to reason from the specific to the general, from the sample to the population. Therefore, any conclusions reached by the statistician are not to be accepted as absolute certainties. It is, in fact, one of the jobs of the statistician to give some measure of the certainty of the conclusions he has drawn.

It should not be inferred from this lack of certainty that the mathematics of statistics is nonrigorous. The mathematics that forms the basis of statistics stems from probability theory and has a firm axiomatic foundation and rigorously proved theorems.

If we conceive of statistics as the science of drawing inferences and making decisions, it is appropriate to date its beginnings with the work of Sir Francis Galton (1822–1911) and Karl Pearson (1857–1936) in the late nineteenth century. Starting here, modern statistical theory

* This is the first of two lectures on the history of statistics given by Professor Varberg at a National Science Foundation Summer Institute for High School Mathematics Teachers held at Bowdoin College during the summer of 1962. Notes were taken by Alvin K. Funderburg.

** Numerals in brackets refer to the notes at the end of this article.

has developed in four great waves of ideas, in four periods, each of which was introduced by a pioneering work of a great statistician [2].

The first period was inaugurated by the publication of Galton's *Natural Inheritance* in 1889. If for no other reason, this book is justly famous because it sparked the interest of Karl Pearson in statistics. Until this time, Pearson had been an obscure mathematician teaching at University College in London. Now the idea that all knowledge is based on statistical foundations captivated his mind. Moving to Gresham College in 1890 with the chance to lecture on any subject that he wished, Pearson chose the topic: "the scope and concepts of modern science." In his lectures he placed increasingly stronger emphasis on the statistical foundation of scientific laws and soon was devoting most of his energy to promoting the study of statistical theory. Before long, his laboratory became a center in which men from all over the world studied and went back home to light statistical fires. Largely through his enthusiasm, the scientific world was moved from a state of disinterest in statistical studies to a situation where large numbers of people were eagerly at work developing new theory and gathering and studying data from all fields of knowledge. The conviction grew that the analysis of statistical data could provide answers to a host of important questions.

An anecdote, related by Helen Walker [3], of Pearson's childhood illustrates in a vivid way the characteristics which marked his adult career. Pearson was asked what was the first thing he could remember. "Well," he said, "I do not know how old I was, but I was sitting in a high chair and I was sucking my thumb. Someone told me to stop sucking it and said that unless I do so the thumb would wither away. I put my two thumbs together and looked at them a long time. 'They look alike to me,' I said to myself, 'I can't see that the thumb I suck is any smaller than the other. I wonder if she could be lying to me.' "

We have here in this simple story, as Helen Walker points out, "rejection of constituted authority, appeal to empirical evidence, faith in his own interpretation of the meaning of observed data, and finally imputation of moral obliquity to a person whose judgment differed from his own." These were to be prominent characteristics throughout Pearson's whole life.

This first period, then, was marked by a change in attitude toward statistics, a recognition of its importance by the scientific world. But, in addition to this, many advances were made in statistical technique. Among the technical tools invented and studied by Galton, Pearson, and their followers were the standard deviation, correlation coefficient, and the chi square test.

About 1915, a new name appeared on the statistical horizon, R. A. Fisher (1890——). His paper of that year on the exact distribution of the sample correlation coefficients ushered in the second period of statistical history and was followed by a whole series of papers and books which gave a new impetus to statistical inquiry. One author has gone so far as to credit Fisher with half of the statistical theory that we use today. Among the significant contributions of Fisher and his associates were the development of methods appropriate for small samples, the discovery of the exact distributions of many sample statistics, the formulation of logical principles for testing hypotheses, the invention of the technique known as analysis of variance, and the introduction of criteria for choice among various possible estimators for a population parameter.

The third period began about 1928 with the publication of certain joint papers by Jerzy Neyman and Egon Pearson, the latter a son of Karl Pearson. These papers introduced and emphasized such concepts as "Type II" error, power of a test, and confidence intervals. It was during this period that industry began to make widespread application of statistical techniques, especially in connection with quality control. There was increasing interest in taking of surveys with consequent attention to the theory and technique of taking samples.

We date the beginning of the fourth period with the first paper of Abraham Wald (1902–50) on the now often used statistical procedure-sequential sampling. This paper of 1939 initiated a deluge of papers by Wald, ended only by his untimely death in an airplane crash when at the height of his powers. Perhaps Wald's most significant contribution was his introduction of a new way of looking at statistical problems, what is known as statistical decision theory. From this point of view, statistics is regarded as the art of playing a game, with nature as the opponent. This is a very general theory, and, while it does lead to formidable mathematical complications, it is fair to say that a large share of present-day research statisticians have found it advantageous to adopt this new approach.

Having given this brief bird's-eye view of statistical history, we move t● a discussion of some of the most basic of statistical concepts. For this purpose it will be convenient to refer to a table showing the heights and weights of twelve people (Fig. 1). The height X is shown in inches; the weight Y is shown in pounds.

To get some feeling for such a collection of data, it is clearly desirable to display

TABLE OF HEIGHTS AND WEIGHTS

INDIVIDUAL	X	Y
1	60	110
2	60	135
3	60	120
4	62	120
5	62	140
6	62	130
7	62	135
8	64	150
9	64	145
10	70	170
11	70	185
12	70	160

Figure 1

FREQUENCY DIAGRAM

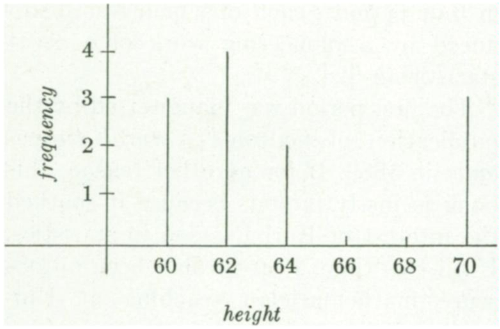

Figure 2

the data pictorially. William Playfair (1759–1823) of England is usually given credit for introducing the idea of graphical representation into statistics. His writings, mostly on economics, were illustrated with extremely good graphs, histograms, bar diagrams, etc. In our problem, the data is most simply represented by means of what is called a frequency diagram.

We have shown such a diagram for the height X (Fig. 2). A similar diagram for Y would be easy to construct. While such pictures do help our intuition, we need more than this if we are to treat the data mathematically. We need mathematical measures which describe the data precisely.

Among the most important of such measures are the measures of central tendency. The earliest of these, actually dating back to the Greeks, is the arithmetic mean μ, which for a discrete variable X, such as we have in our example, is defined by

$$\mu_X = (1/n) \sum_{i=1}^{n} x_i.$$

Here x_i denotes a value of the variable X, and n is the size of the population. In our example, the mean μ_X of the heights is 63.83; the mean μ_Y of the weights is 141.67.

To understand the significance of this concept, we rewrite the definition in the form

$$\mu_X = (1/n) \sum x_j f_j.$$

Here f_j stands for the frequency of occurrence of the value x_j and the summation extends over the distinct values of the variable X. Consider now a weightless rod on which there is a scale running through the range of the variable X, and suppose that at x_j is attached a mass of size f_j/n. This gives a system of total mass 1, which will have μ_X as its center of mass, that is, the system will balance on a fulcrum placed at μ_X. In the case of the heights, the system would look as in Figure 3. This interpretation of the mean will be helpful later when we consider the notion of a continuously distributed variable.

While the concept is probably quite old, it was not until 1883 that the median was introduced into statistics by Galton as a second measure of central tendency [4]. The median is simply the middle value of the distribution in the case of an odd number of values and is the average of the two middle values otherwise. The median height in our example is 62.

Another measure of central tendency is the mode, introduced by Karl Pearson around 1894. The mode is the most frequently occurring value, if there is one. In the case where two or more values occur with equal frequency, the mode is not well defined. In the example, the mode of heights is again 62.

If the distribution of a variable X is exactly symmetrical, i.e., if its frequency diagram is exactly symmetrical about a vertical line, then the mean, median, and mode (if there is a mode) will agree. The reader should be able to convince himself that the converse is false by constructing a nonsymmetrical distribution for which the mean, median, and mode agree.

For most purposes, certainly for theoretical purposes, the mean is the most useful measure of central tendency, although admittedly it may take much calculation to get it. The median does, however, have a property which is sometimes advantageous. It is not as subject to distortion due to a few extreme values. For example, if in the table of heights of twelve persons, one of the 70-inch persons were exchanged for a 90-inch person, the mean would be changed considerably while the median would be unaffected.

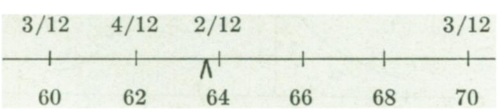

***Figure* 3**

We next consider measures of dispersion, i.e., measures of how the data spreads out about the mean. Perhaps the first such measure was the probable error introduced by Bessel in 1815 in connection with problems in astronomy. Most commonly used today is the standard deviation σ, this terminology due to Karl Pearson (1894). It is defined for a discrete variable X by

$$\sigma_X = \left[(1/n) \sum_{i=1}^{n} (x_i - \mu_X)^2 \right]^{1/2}.$$

Inspection of this formula reveals that σ tends to be large when the data is widely dispersed, small when the data clusters about the mean.

To introduce the next notion, which is correlation, we refer back to the table of heights and weights (Fig. 1). Inspection of the data reveals that these two variables are somehow related. Even common sense tells us that tall people should generally weigh more than short people. Graphically, this relationship can be portrayed by means of what is called a scatter diagram, this being merely a plot of the data in the Cartesian plane (see Fig. 4). The relationship, if linear, will be indicated by a tendency of the points to simulate a straight line.

In the late nineteenth century, Sir Francis Galton asked whether such a relationship between two sets of data could be measured, and he introduced the notion of correlation. It was Karl Pearson, however, who gave us our present coefficient of correlation ρ defined by

$$\rho = (1/n\sigma_X\sigma_Y) \sum_{i=1}^{n} (x_i - \mu_X)(y_i - \mu_Y).$$

Scatter diagram

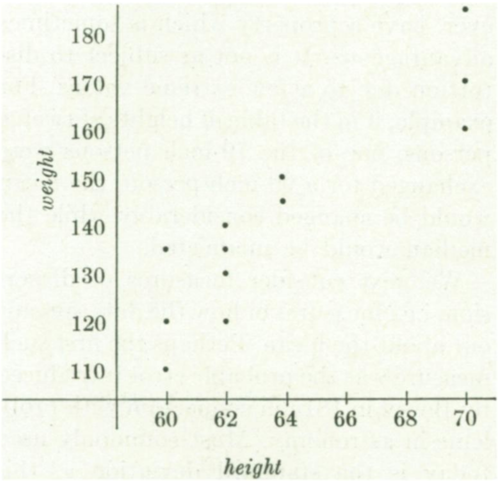

Figure 4

It is a matter of simple algebra to show that ρ ranges between -1 and $+1$. A value of zero indicates no linear relationship; $+1$ indicates that the data lies on a straight line of positive slope; -1 means that the data lies on a line of negative slope; values near ± 1 suggest a strong linear relationship, while values near zero are characteristic of little such relationship. In our example, ρ is about 0.9. It should be emphasized that ρ is a measure of *linear* relationship. The data may lie on a circle, in which case $|\rho|$ will be very small. However, in this case the variables would certainly be related, albeit not linearly related.

Sir Francis Galton, prominent in our discussion thus far, was a cousin of Charles Darwin and did some statistical work for him. His interest in correlation has already been mentioned—it was his writings on this subject which turned the brilliant Karl Pearson toward the study of statistics. But Galton will be remembered most vividly by teachers because he was the first to suggest the use of the normal curve in connection with problems of grading.

The normal curve, which actually dates back at least to Abraham De Moivre in 1733, is a highly useful concept to statistics. It is determined by the equation

$$f(x) = (1/\sqrt{2\pi}\,\sigma) \exp\left[-(x-\mu)^2/2\sigma^2\right].$$

Here μ and σ are parameters which turn out to be the mean and standard deviation. The normal curve is often thought of roughly as any "bell shaped" curve. However, this is inaccurate, for other functions, such as $g(x) = [\pi(1+x^2)]^{-1}$, also have graphs which are bell shaped and yet lack completely the qualities which make the normal curve so useful. While the definition of the normal curve given above may appear complicated, from the point of view of the mathematician it is one of the simplest and best behaved of all curves. Figure 5 pictures a special normal curve.

If the area under the normal curve from $-\infty$ to $+\infty$ were to be calculated by integration, it would be found to be 1. Approximately two-thirds of this area lies between points one standard deviation to the left and one standard deviation to the right of the mean. The probability that a normal variable assumes values on any interval $a \leq x \leq b$ is equal to the area above this interval and under the corresponding normal curve. Areas under the normal curve for various intervals are tabulated in any standard book of mathematical tables.

Earlier, in the discussion of discrete distribution, it was shown that the mean could be interpreted as the center of mass of a system of discrete masses of total mass 1. The normal distribution described above is an example of a continuous distribution. Reasoning by analogy, we may associate with the normal distribution an

The normal curve ($\mu = 1$, $\sigma = 1$)

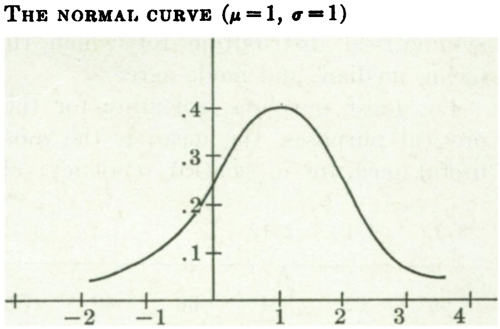

Figure 5

idealized continuous rod of mass 1 running indefinitely far in both directions with density varying according to the function f which determines the normal curve. From calculus, the center of mass μ of such a rod would be given by

$$\mu=\int_{-\infty}^{\infty} xf(x)dx.$$

This is, in fact, the formula which we use to define the mean of a continuous distribution. Perhaps surprisingly, not every continuous distribution has a mean, for the above integral may fail to converge. This is the case, for example, for the Cauchy distribution, determined by the equation $g(x)=[\pi(1+x^2)]^{-1}$, as the reader may verify.

Recalling the formula in the discrete case, it is natural to define the standard deviation for a continuous distribution by

$$\sigma=\left[\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx\right]^{1/2}.$$

It is a matter of a fairly simple integration to check that if these formulas are used to calculate the mean and standard deviation for the normal distribution, they turn out to be the two parameters μ and σ respectively.

Some important uses of the normal curve in connection with statistical problems will be described in the second of these lectures, which is to be reprinted in the next issue of THE MATHEMATICS TEACHER.

NOTES

1. WALKER, HELEN M., *Studies in the History of Statistical Method* (Baltimore: The William and Wilkins Company, 1929). This book has been of great value in the preparation of this lecture, especially in connection with statistical history before 1900.
2. In dividing the history of statistics into four periods, we are following Helen M. Walker, "The Contributions of Karl Pearson," *Journal of the American Statistical Association*, LIII (1958), 11–22.
3. *Ibid.*, p. 13.
4. Actually, Gustav Fechner had employed this measure under the name *der Centralwerth* in 1874 and had given a description of its properties. Galton's use of the concept appears to go back as early as 1869, but the name *median* is first used by him in 1883.

The origin of L'Hôpital's rule

by D. J. Struik, Massachusetts Institute of Technology, Cambridge, Massachusetts

The so-called rule of L'Hôpital, which states that

$$\lim_{x\to a}\frac{f(x)}{g(x)}=\frac{f'(a)}{g'(a)}$$

when $f(a)=g(a)=0$, $g'(a)\neq 0$, was published for the first time by the French mathematician G. F. A. de l'Hôpital (or De Lhospital) in his *Analyse des infiniment petits* (Paris, 1696) [1].* The Marquis de l'Hôpital was an amateur mathematician who had become deeply interested in the new calculus presented to the learned world by Leibniz in two short papers, one of 1684 and the other of 1686. Not quite convinced that he could master the new and exciting branch of mathematics all by himself, L'Hôpital engaged, during some months of 1691–92, the services of the brilliant young Swiss physician and mathematician, Johann Bernoulli, first at his Paris home and later at his château in the

* Numerals in brackets refer to the notes at the end of this article.