Econometrics, branch of *economics* that uses mathematical methods and models. *Calculus, probability, statistics, linear programming*, and *game theory*, as well as other areas of *mathematics*, are used to analyze, interpret, and predict various economic factors and systems, such as price and market action, production cost, *business* trends, and economic policy.

Calculus (mathematics), branch of mathematics concerned with the study of such concepts as the rate of change of one variable quantity with respect to another, the slope of a curve at a prescribed point, the computation of the maximum and minimum values of functions, and the calculation of the area bounded by curves. Evolved from *algebra*, arithmetic, and geometry, it is the basis of that part of *mathematics* called analysis.

Calculus is widely employed in the physical, biological, and social sciences. It is used, for example, in the physical sciences to study the speed of a falling body, the rates of change in a chemical reaction, or the rate of decay of a radioactive material. In the biological sciences a problem such as the rate of growth of a colony of bacteria as a function of time is easily solved using calculus. In the social sciences calculus is widely used in the study of *statistics* and *probability*.

Calculus can be applied to many problems involving the notion of extreme amounts, such as the fastest, the most, the slowest, or the least. These maximum or minimum amounts may be described as values for which a certain rate of change (increase or decrease) is zero. By using calculus it is possible to determine how high a projectile will go by finding the point at which its change of altitude with respect to time, that is, its velocity, is equal to zero. Many general principles governing the behavior of physical processes are formulated almost invariably in terms of rates of change. It is also possible, through the insights provided by the methods of calculus, to resolve such problems in logic as the famous paradoxes posed by the Greek philosopher *Zeno*.

The fundamental concept of calculus, which distinguishes it from other branches of mathematics and is the source from which all its theory and applications are developed, is the theory of limits of functions of variables.

Let f be a function of the real variable x, which is denoted $f(x)$, defined on some set of real numbers surrounding the number x_0. It is not required that the function be defined at the point x_0 itself. Let L be a real number. The expression

$$\lim_{x \to x_0} f(x) = L$$

is read: “The limit of the function $f(x)$, as x approaches x_0, is equal to the number L.” The notation is designed to convey the idea that $f(x)$ can be made as “close” to L as desired simply by choosing an x sufficiently close to x_0. For example, if the function $f(x)$ is defined as $f(x) = x^2 + 3x + 2$, and if $x_0 = 3$, then from the definition above it is true that

$$\lim_{x \to 3} f(x) = 20$$

This is because, as x approaches 3 in value, x^2 approaches 9, $3x$ approaches 9, and 2 does not change, so their sum approaches 9 + 9 + 2, or 20.

Another type of limit important in the study of calculus can be illustrated as follows. Let the domain of a function $f(x)$ include all of the numbers greater than some fixed number m. L is said to be the limit of the function $f(x)$ as x becomes positively infinite, if, corresponding to a given positive number ϵ, no matter how small, there exists a number M such that the numerical difference between $f(x)$ and L (the absolute value $|f(x) - L|$) is less than ϵ whenever x is greater than M. In this case the limit is written as

$$\lim_{x \to \infty} f(x) = L$$

For example, the function $f(x) = 1/x$ approaches the number 0 as x becomes positively infinite.

It is important to note that a limit, as just presented, is a two-way, or bilateral, concept: A dependent variable approaches a limit as an independent variable approaches a number or becomes infinite. The limit concept can be extended to a variable that is dependent on several independent variables. The statement "*u* is an infinitesimal" meaning "*u* is a variable approaching 0 as a limit," found in a few present-day and in many older texts on calculus, is confusing and should be avoided. Further, it is essential to distinguish between the limit of $f(x)$ as x approaches x_0 and the value of $f(x)$ when x is x_0, that is, the correspondent of x_0. For example, if $f(x) = \sin x/x$, then

$$\lim_{x \to 0} f(x) = 1$$

however, no value of $f(x)$ corresponding to $x = 0$ exists, because division by 0 is undefined in mathematics.

The two branches into which elementary calculus is usually divided are differential calculus, based on the consideration of the limit of a certain ratio, and integral calculus, based on the consideration of the limit of a certain sum.

Differential Calculus

Let the dependent variable y be a function of the independent variable x, expressed by $y = f(x)$. If x_0 is a value of x in its domain of definition, then $y_0 = f(x_0)$ is the corresponding value of y. Let h and k be real numbers, and let $y_0 + k = f(x_0 + h)$. (Δx, read "delta x," is used quite frequently in place of h.) When Δx is used in place of h, Δy is used in place of k. Then clearly

$$k = f(x_0 + h) - f(x_0)$$

and

$$\frac{k}{h} = \frac{f(x_0 + h) - f(x_0)}{h}$$

This ratio is called a difference quotient. Its intuitive meaning can be grasped from the geometrical interpretation of the graph of $y = f(x)$. Let A and B be the points (x_0, y_0), $(x_0 + h, y_0 + k)$, respectively, as in the Derivatives illustration. Draw the secant AB and the lines AC and CB, parallel to the x and y axes, respectively, so that $h = AC$, $k = CB$. Then the difference quotient k/h equals the tangent of angle BAC and is therefore, by definition, the slope of the secant AB. It is evident that if an insect were crawling along the curve from A to B, the abscissa x would always increase along its path but the ordinate y would first increase, slow down, then decrease. Thus, y varies with respect to x at different rates between A and B. If a second insect crawled from A to B along the secant, the ordinate y would vary at a constant rate, equal to the difference quotient k/h, with respect to the abscissa x. As the two insects start and end at the same points, the difference quotient may be regarded as the average rate of change of $y = f(x)$ with respect to x in the interval AC.

If the limit of the ratio k/h exists as h approaches 0, this limit is called the derivative of y with respect to x, evaluated at $x = x_0$. For example, let $y = x^2$ and $x = 3$, so that $y = 9$. Then $9 + k = (3 + h)^2$; $k = (3 + h)^2 - 9 = 6h + h^2$; $k/h = 6 + h$; and $\lim_{h \to 0} k/h = 6.$ Referring back to the Derivatives illustration, the secant AB pivots around A and approaches a limiting position, the tangent AT, as h approaches 0. The derivative of y with respect to x, at $x = x_0$, may be interpreted as the slope of the tangent AT, and this slope is defined as the slope of the curve $y = f(x)$ at $x = x_0$. Further, the derivative of y with respect to x, at $x = x_0$, may be interpreted as the instantaneous rate of change of y with respect to x at x_0.

If the derivative of y with respect to x is found for all values of x (in its domain) for which the derivative is defined, a new function is obtained, the derivative of y with respect to x. If $y = f(x)$, the new function is written as y' or $f'(x)$, $D_x y$ or $D_x f(x)$, $(dy)/(dx)$ or $df(x)/dx$. Thus, if $y = x^2$, $y + k = (x + h)^2$; $k = (x + h)^2 - x^2 = 2xh + h^2$; $k/h = 2x + h$, whence $D_x x^2 = \lim_{h \to 0} k/h = 2x.$ Thus, as before, $y' = f'(x) = 6$ at $x = 3$, or $f'(3) = 6$; also, $f'(2) = 4$, $f'(0) = 0$, and $f'(-2) = -4$.

As the derivative $f'(x)$ of a function $f(x)$ of x is itself a function of x, its derivative with respect to x can be found; it is called the second (order) derivative of y with respect to x, and is designated by any one of the symbols y'' or $f''(x)$, $D_x^2 y$ or $D_x^2 f(x)$, $(d^2y)/(dx^2)$ or $(d^2f(x))/(dx^2)$. Third- and higher-order derivatives are similarly designated.

Every application of differential calculus stems directly or indirectly from one or both of the two interpretations of the derivative as the slope of the tangent to the curve and as the rate of change of the dependent variable with respect to the independent variable. In a detailed study of the subject, rules and methods developed by the limit process are provided for rapid calculation of the derivatives of various functions directly by means of various known formulas. Differentiation is the name given to the process of finding a derivative.

Differential calculus provides a method of finding the slope of the tangent to a curve at a certain point; related rates of change, such as the rate at which the area of a circle increases (in square feet per minute) in terms of the radius (in feet) and the rate at which the radius increases (in feet per minute); velocities (rates of change of distance with respect to time) and accelerations (rates of change of velocities with respect to time, therefore represented as second derivatives of distance with respect to time) of points moving on straight lines or other curves; and absolute and relative maxima and minima.

Integral Calculus

Let $y = f(x)$ be a function defined for all x's in the interval $[a, b]$, that is, the set of x's from $x = a$ to $x = b$, including a and b, where $a < b$ (suitable modifications can be made in the definitions to follow for more restricted ranges or domains). Let $x_0, x_1, \ldots, x_n$ be a sequence of values of x such that $a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$, and let $h_1 = x_1 - x_0$, $h_2 = x_2 - x_1$, …, $h_n = x_n - x_{n-1}$, in brief, $h_i = x_i - x_{i-1}$, where $i = 1, 2, \ldots, n$. The x's form a partition of the interval $[a,$

b]; an h with a value not exceeded by any other h is called the norm of the partition. Let n values of x, for example, $X_1, X_2, \ldots, X_n$, be chosen so that $x_{i-1} < X_i < x_i$, where $i = 1, 2, \ldots, n$. The sum of the area of the rectangles is given by

$$f(X_1)h_1 + f(X_2)h_2 + \ldots. + f(X_n)h_n$$

usually abbreviated to $\sum_{i=1}^{n} f(x_i)h_i$. (Σ is the Greek capital letter *sigma.*) Aside from the given function $f(x)$ and the given a and b, the value of the sum clearly depends on n and on the choices of the x_i's and X_i's. In particular, if, after the x_i's are chosen, the X_i's are chosen so that $f(X_i)$, for each i, is a maximum in the interval $[x_{i-1}, x_i]$ (that is, no ordinate from x_{i-1} to x_i exceeds the ordinate at X_i), the sum is called an upper sum; similarly, if, after the x_i's are chosen, the X_i's are chosen so that $f(X_i)$, for each i, is a minimum in the interval $[x_{i-1}, x_i]$, the sum is called a lower sum. It can be proved that the upper and lower sums will have limits, $\overline{S}$ and $\underline{S}$, respectively, as the norm approaches 0. If $\overline{S}$ and $\underline{S}$, are equal and have the common value S, S is called the definite integral of $f(x)$ from a to b and is written

$$S = \int_a^b f(x)\,dx$$

The symbol $\int$ is an elongated S (for sum); the $f(x)\,dx$ is suggested by a term $f(X_i)h_i = f(X_i)\,\Delta x_i$ of the sum which is used in defining the definite integral.

If $y = g(x)$, then by differentiation $y' = g'(x)$. Let $g'(x) = f(x)$, and C be any constant. Then $f(x)$ is also the derivative of $g(x) + C$. The expression $g(x) + C$ is called the antiderivative of $f(x)$, or the indefinite integral of $f(x)$, and it is represented by

$$\int f(x)\,dx = g(x) + c$$

The dual use of the term *integral* is justified by one of the fundamental theorems of calculus, namely, if $g(x)$ is an antiderivative of $f(x)$, then, under suitable restrictions on $f(x)$ and $g(x)$,

$$\int_a^b f(x)\,dx = g(b) - g(a)$$

The process of finding either an indefinite or a definite integral of a function $f(x)$ is called integration; the fundamental theorem relates differentiation and integration.

If the antiderivative, $g(x)$, of $f(x)$ is not readily obtainable or is not known, the definite integral $^{b}{}_{a}$ □ $f(x)dx$ can be approximated by the trapezoidal rule, $(b - a)\,[f(a) + f(b)]/2$ or by the more accurate Simpson's rule:

$$\frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right]$$

If $|b - a|$ is small, Simpson's rule gives a fairly close result. If $|b - a|$ is large, a good approximation can be obtained by dividing the interval from *a* to *b* into a number of small intervals and applying Simpson's rule to the subintervals.

Integral calculus involves the inverse process of finding the derivative of a function, that is, it is the process of finding the function itself when its derivative is known. For example, integral calculus makes it possible to find the equation of a curve if the slope of the tangent is known at an arbitrary point; to find distance in terms of time if the velocity (or acceleration) is known; and to find the equation of a curve if its curvature is known. Integral calculus can also be used to find the lengths of curves, the areas of plane and curved surfaces, volumes of solids of revolution, centroids, moments of inertia, and total mass and total force.

Differential Equations

Calculus leads directly to the branch of mathematics called differential equations, which is extremely useful in engineering and in the physical sciences. An ordinary differential equation is an equation involving an independent variable, a dependent variable (one or both of these two may be missing), and one or more derivatives (at least one derivative must be present). Many physical laws or statements are initially expressed as differential equations. For example, the law that the acceleration of gravity is a constant *g* can be expressed mathematically by the differential equation $d^2x/dt^2 = g$; the principle that the rate of disintegration of radium is proportional to the amount present is expressed as $dR/dt = -kR$. A differential equation is solved if an equivalent equation is found involving only the independent and dependent variables.

This article has considered functions of a single independent variable only. Partial derivatives, multiple integrals, and partial differential equations are defined and studied in investigating functions of two or more independent variables.

Development of Calculus

The English and German mathematicians, respectively, Isaac Newton and Gottfried Wilhelm Leibniz invented calculus in the 17th century, but isolated results about its fundamental problems had been known for thousands of years. For example, the Egyptians discovered the rule for the volume of a pyramid as well as an approximation of the area of a circle. In ancient Greece, Archimedes proved that if *c* is the circumference and *d* the diameter of a circle, then 3□ *d* < *c* < 3□ *d*. His proof extended the method of inscribed and circumscribed figures developed by the Greek astronomer and mathematician Eudoxus. Archimedes used the same technique for his other results on areas and volumes. Archimedes discovered his results by means of heuristic arguments involving parallel slices of the figures and the law of the lever. Unfortunately, his treatise *The Method* was only rediscovered in the 19th century, so later mathematicians believed that the Greeks deliberately concealed their secret methods.

During the late middle ages in Europe, mathematicians studied translations of Archimedes' treatises from Arabic. At the same time, philosophers were studying problems of change and the infinite, such as the addition of infinitely many quantities. Greek thinkers had seen only contradictions there, but medieval thinkers aided mathematics by making the infinite philosophically respectable.

By the early 17th century, mathematicians had developed methods for finding areas and volumes of a great variety of figures. In his *Geometry by Indivisibles,* the Italian mathematician

F. B. Cavalieri, a student of the Italian physicist and astronomer Galileo, expanded on the work of the German astronomer Johannes Kepler on measuring volumes. He used what he called “indivisible magnitudes” to investigate areas under the curves $y = x^n$, $n = 1 \ldots 9$. Also, his theorem on the volumes of figures contained between parallel planes (now called Cavalieri’s theorem) was known all over Europe. At about the same time, the French mathematician René Descartes’ *La Géométrie* appeared. In this important work, Descartes showed how to use algebra to describe curves and obtain an algebraic analysis of geometric problems. A codiscoverer of this analytic geometry was the French mathematician Pierre de Fermat, who also discovered a method of finding the greatest or least value of some algebraic expressions—a method close to those now used in differential calculus.

About 20 years later, the English mathematician John Wallis published *The Arithmetic of Infinites,* in which he extrapolated from patterns that held for finite processes to get formulas for infinite processes. His colleague at the University of Cambridge was Newton’s teacher, the English mathematician Isaac Barrow, who published a book that stated geometrically the inverse relationship between problems of finding tangents and areas, a relationship known today as the fundamental theorem of calculus.

Although many other mathematicians of the time came close to discovering calculus, the real founders were Newton and Leibniz. Newton’s discovery (1665-66) combined infinite sums (infinite series), the *binomial* theorem for fractional exponents, and the algebraic expression of the inverse relation between tangents and areas into methods we know today as calculus. Newton, however, was reluctant to publish, so Leibniz became recognized as a codiscoverer because he published his discovery of differential calculus in 1684 and of integral calculus in 1686. It was Leibniz, also, who replaced Newton’s symbols with those familiar today.

In the following years, one problem that led to new results and concepts was that of describing mathematically the motion of a vibrating string. Leibniz’s students, the Bernoulli family of Swiss mathematicians (*Bernoulli, Daniel* etc), used calculus to solve this and other problems, such as finding the curve of quickest descent connecting two given points in a vertical plane. In the 18th century, the great Swiss-Russian mathematician Leonhard Euler, who had studied with Johann Bernoulli, wrote his *Introduction to the Analysis of Infinites,* which summarized known results and also contained much new material, such as a strictly analytic treatment of trigonometric and exponential functions.

Despite these advances in technique, calculus remained without logical foundations. Only in 1821 did the French mathematician A. L. Cauchy succeed in giving a secure foundation to the subject by his theory of limits, a purely arithmetic theory that did not depend on geometric intuition or infinitesimals. Cauchy then showed how this could be used to give a logical account of the ideas of continuity, derivatives, integrals, and infinite series. In the next decade, the Russian mathematician N. I. Lobachevsky and German mathematician P. G. L. Dirichlet both gave the definition of a function as a correspondence between two sets of real numbers, and the logical foundations of calculus were completed by the German mathematician J. W. R. Dedekind in his theory of real numbers, in 1872.

Probability, also theory of probability, branch of mathematics that deals with measuring or determining quantitatively the likelihood that an event or experiment will have a particular outcome. Probability is based on the study of *permutations and combinations* and is the necessary foundation for *statistics*.

The foundation of probability is usually ascribed to the 17th-century French mathematicians Blaise Pascal and Pierre de Fermat, but mathematicians as early as Gerolamo Cardano had made important contributions to its development. Mathematical probability began in an attempt to answer certain questions arising in games of chance, such as how many times a pair of dice must be thrown before the chance that a six will appear is 50-50. Or, in another example, if two players of equal ability, in a match to be won by the first to win ten games, are obliged to suspend play when one player has won five games, and the other seven, how should the stakes be divided?

The probability of an outcome is represented by a number between 0 and 1, inclusive, with "probability 0" indicating certainty that an event will not occur and "probability 1" indicating certainty that it will occur. The simplest problems are concerned with the probability of a specified "favorable" result of an event that has a finite number of equally likely outcomes. If an event has n equally likely outcomes and f of them are termed favorable, the probability, p, of a favorable outcome is f/n. For example, a fair die can be cast in six equally likely ways; therefore, the probability of throwing a 5 or a 6 is 2/6. More involved problems are concerned with events in which the various possible outcomes are not equally likely. For example, in finding the probability of throwing a 5 or 6 with a pair of dice, the various outcomes (2, 3, … 12) are not all equally likely. Some events may have infinitely many outcomes, such as the probability that a chord drawn at random in a circle will be longer than the radius.

Problems involving repeated trials form one of the connections between probability and statistics. To illustrate, what is the probability that exactly five 3s and at least four 6s will occur in 50 tosses of a fair die? Or, a person, tossing a fair coin twice, takes a step to the north, east, south, or west, according to whether the coin falls head, head; head, tail; tail, head; or tail, tail. What is the probability that at the end of 50 steps the person will be within 10 steps of the starting point?

In probability problems, two outcomes of an event are mutually exclusive if the probability of their joint occurrence is zero; two outcomes are independent if the probability of their joint occurrence is given as the product of the probability of their separate occurrences. Two outcomes are mutually exclusive if the occurrence of one precludes the occurrence of the other; two outcomes are independent if the occurrence or nonoccurrence of one does not alter the probability that the other will or will not occur. Compound probability is the probability of all outcomes of a certain set occurring jointly; total probability is the probability that at least one of a certain set of outcomes will occur. Conditional probability is the probability of an outcome when it is known that some other outcome has occurred or will occur.

If the probability that an outcome will occur is p, the probability that it will not occur is $q = 1 - p$. The odds in favor of the occurrence are given by the ratio $p:q$, and the odds against the occurrence are given by the ratio $q:p$. If the probabilities of two mutually exclusive outcomes X and Y are p and P, respectively, the odds in favor of X and against Y are p to P. If an event must result in one of the mutually exclusive outcomes $O_1, O_2, \ldots, O_n$, with probabilities $p_1, p_2, \ldots, p_n$, respectively, and if $v_1, v_2, \ldots v_n$ are numerical values attached to the respective outcomes, the expectation of the event is $E = p_1v_1 + p_2v_2 + \ldots p_nv_n$. For example, a person throws a die and wins 40 cents if it falls 1, 2, or 3; 30 cents for 4 or 5; but loses $1.20 if it falls 6. The expectation on a single throw is $3/6 \times .40 + 2/6 \times .30 - 1/6 \times 1.20 = .10$.

The most common interpretation of probability is used in statistical analysis. For example, the probability of throwing a 7 in one throw of two dice is 1/6, and this answer is interpreted to mean that if two fair dice are randomly thrown a very large number of times, about one-sixth of the throws will be 7s. This concept is frequently used to statistically determine the probability of an outcome that cannot readily be tested or is impossible to obtain. Thus, if long-range statistics show that out of every 100 people between 20 and 30 years of age, 42 will be alive at age 70, the assumption is that a person between those ages has a 42 percent probability of surviving to the age of 70.

Mathematical probability is widely used in the physical, biological, and social sciences and in industry and commerce. It is applied in such diverse areas as genetics, quantum mechanics, and insurance. It also involves deep and important theoretical problems in pure mathematics and has strong connections with the theory, known as mathematical analysis, that developed out of calculus.

Statistics, branch of mathematics that deals with the collection, organization, and analysis of numerical data and with such problems as experiment design and decision making.

History

Simple forms of statistics have been used since the beginning of civilization, when pictorial representations or other symbols were used to record numbers of people, animals, and inanimate objects on skins, slabs, or sticks of wood and the walls of caves. Before 3000 BC the Babylonians used small clay tablets to record tabulations of agricultural yields and of commodities bartered or sold. The Egyptians analyzed the population and material wealth of their country before beginning to build the pyramids in the 31st century BC. The biblical books of Numbers and 1 Chronicles are primarily statistical works, the former containing two separate censuses of the Israelites and the latter describing the material wealth of various Jewish tribes. Similar numerical records existed in China before 2000 BC. The ancient Greeks held censuses to be used as bases for taxation as early as 594 BC.

The Roman Empire was the first government to gather extensive data about the population, area, and wealth of the territories that it controlled. During the Middle Ages in Europe few comprehensive censuses were made. The Carolingian kings Pepin the Short and Charlemagne ordered surveys of ecclesiastical holdings: Pepin in 758 and Charlemagne in 762. Following the Norman Conquest of England in 1066, William I, king of England, ordered a census to be taken; the information gathered in this census, conducted in 1086, was recorded in the *Domesday Book*. Registration of deaths and births was begun in England in the early 16th century, and in 1662 the first noteworthy statistical study of population, *Observations on the London Bills of Mortality,* was written. A similar study of mortality made in Breslau, Germany, in 1691 was used by the English astronomer Edmund Halley as a basis for the earliest mortality table. In the 19th century, with the application of the scientific method to all phenomena in the natural and social sciences, investigators recognized the need to reduce information to numerical values to avoid the ambiguity of verbal description.

At present, statistics is a reliable means of describing accurately the values of economic, political, social, psychological, biological, and physical data and serves as a tool to correlate and analyze such data. The work of the statistician is no longer confined to gathering and tabulating data, but is chiefly a process of interpreting the information. The development of the theory of *probability* increased the scope of statistical applications. Much data can be approximated accurately by certain probability distributions, and the results of probability distributions can be used in analyzing statistical data. Probability can be used to test the reliability of statistical inferences and to indicate the kind and amount of data required for a particular problem.

Statistical Methods

The raw materials of statistics are sets of numbers obtained from enumerations or measurements. In collecting statistical data, adequate precautions must be taken to secure complete and accurate information.

The first problem of the statistician is to determine what and how much data to collect. Actually, the problem of the census taker in obtaining an accurate and complete count of the population, like the problem of the physicist who wishes to count the number of molecule collisions per second in a given volume of gas under given conditions, is to decide the precise nature of the items to be counted. The statistician faces a complex problem when, for example, he or she wishes to take a sample poll or straw vote. It is no simple matter to gauge the size and constitution of the sample that will yield reasonably accurate predictions concerning the action of the total population.

In protracted studies to establish a physical, biological, or social law, the statistician may start with one set of data and gradually modify it in light of experience. For example, in early studies of the growth of populations, future change in size of population was predicted by calculating the excess of births over deaths in any given period. Population statisticians soon recognized that rate of increase ultimately depends on the number of births, regardless of the number of deaths, so they began to calculate future population growth on the basis of the number of births each year per 1000 population. When predictions based on this method yielded inaccurate results, statisticians realized that other limiting factors exist in population growth. Because the number of births possible depends on the number of women rather than the total population, and because women bear children during only part of their total lifetime, the basic datum used to calculate future population size is now the number of live births per 1000 females of childbearing age. The predictive value of this basic datum can be further refined by combining it with other data on the percentage of women who remain childless because of choice or circumstance, sterility, contraception, death before the end of the childbearing period, and other limiting factors. The excess of births over deaths, therefore, is meaningful only as an indication of gross population growth over a definite period in the past; the number of births per 1000 population is meaningful only as an expression of the proportion of increase during a similar period; and the number of live births per 1000 women of childbearing age is meaningful for predicting future size of populations.

Tabulation and Presentation of Data

The collected data must be arranged, tabulated, and presented to permit ready and meaningful analysis and interpretation. To study and interpret the examination-grade distribution in a class of 30 pupils, for instance, the grades are arranged in ascending order: 30, 35, 43, 52, 61, 65, 65, 65, 68, 70, 72, 72, 73, 75, 75, 76, 77, 78, 78, 80, 83, 85, 88, 88, 90, 91, 96, 97, 100, 100. This progression shows at a glance that the maximum is 100, the minimum 30, and the range, or difference, between the maximum and minimum is 70.

In a cumulative-frequency graph, such as Fig. 1, the grades are marked on the horizontal axis and double marked on the vertical axis with the cumulative number of the grades on the left and the corresponding percentage of the total number on the right. Each dot represents the accumulated number of students who have attained a particular grade or less. For example, the dot *A* corresponds to the second 72; reading on the vertical axis, it is evident that there are 12, or 40 percent, of the grades equal to or less than 72.

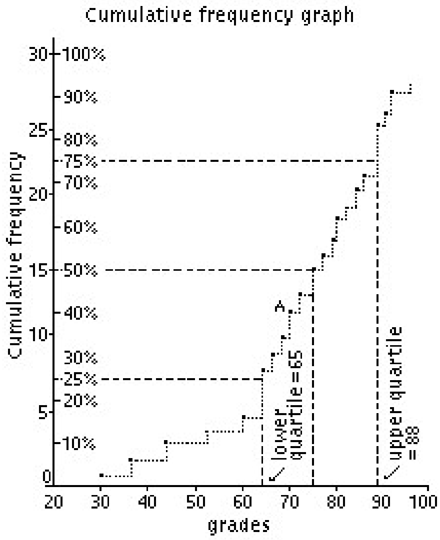


Figure 1

In analyzing the grades received by 10 sections of 30 pupils each on four examinations, a total of 1200 grades, the amount of data is too large to be exhibited conveniently as in Fig. 1. The statistician separates the data into suitably chosen groups, or intervals. For example, ten intervals might be used to tabulate the 1200 grades, as in column (a) of the accompanying frequency-distribution table; the actual number in an interval, called the frequency of the interval, is entered in column (c). The numbers that define the interval range are called the interval boundaries. It is convenient to choose the interval boundaries so that the interval ranges are equal to each other; the interval midpoints, half the sum of the interval boundaries, are simple numbers, because they are used in many calculations. A grade such as 87 will be tallied in the 80-90 interval; a boundary grade such as 90 may be tallied uniformly throughout the groups in either the lower or upper intervals. The relative frequency, column (d), is the ratio of the frequency of an interval to the total count; the relative frequency is multiplied by 100 to obtain the percent relative frequency. The cumulative frequency, column (e), represents the number of students receiving grades equal to or less than the range in each succeeding interval; thus, the number of students with grades of 30 or less is obtained by adding the frequencies in column (c) for the first three intervals, which total 53. The cumulative relative frequency, column (f), is the ratio of the cumulative frequency to the total number of grades.

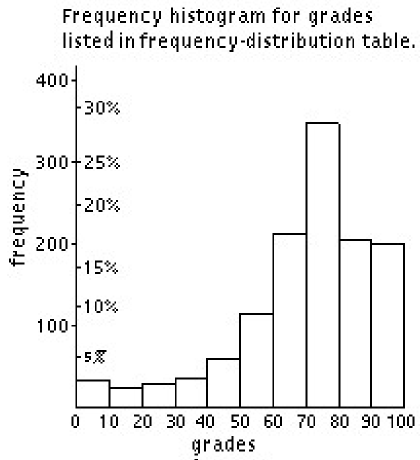


Figure 2

The data of a frequency-distribution table can be presented graphically in a frequency histogram, as in Fig. 2, or a cumulative-frequency polygon, as in Fig. 3. The histogram is a series of rectangles with bases equal to the interval ranges and areas proportional to the frequencies. The polygon in Fig. 3 is drawn by connecting with straight lines the interval midpoints of a cumulative frequency histogram.

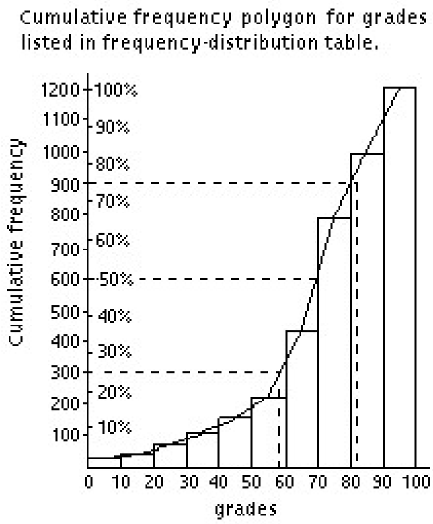


Figure 3

Newspapers and other printed media frequently present statistical data pictorially by using different lengths or sizes of various symbols to indicate different values.

Measures of Central Tendency

After data have been collected and tabulated, analysis begins with the calculation of a single number, which will summarize or represent all the data. Because data often exhibit a cluster or central point, this number is called a measure of central tendency.

Let x_1, x_2, …, x_n be the n tabulated (but ungrouped) numbers of some statistic; the most frequently used measure is the simple arithmetic average, or mean, written x, which is the sum of the numbers divided by n:

$$\bar{x} = \frac{\Sigma x}{n}$$

If the x's are grouped into k intervals, with midpoints m_1, m_2, …, m_k and frequencies f_1, f_2, …, f_k, respectively, the simple arithmetic average is given by

$$\frac{\Sigma f_i m_i}{\Sigma f_i}$$

with $i = 1, 2, \ldots, k$.

The median and the mode are two other measures of central tendency. Let the x's be arranged in numerical order; if n is odd, the median is the middle x; if n is even, the median is the average of the two middle x's. The mode is the x that occurs most frequently. If two or more distinct x's occur with equal frequencies, but none with greater frequency, the set of x's may be said not to have a mode or to be bimodal, with modes at the two most frequent x's, or trimodal, with modes at the three most frequent x's.

Measures of Variability

The investigator frequently is concerned with the variability of the distribution, that is, whether the measurements are clustered tightly around the mean or spread over the range. One measure of this variability is the difference between two percentiles, usually the 25th and the 75th percentiles. The pth percentile is a number such that p percent of the measurements are less than or equal to it; in particular, the 25th and the 75th percentiles are called the lower and upper quartiles, respectively. The pth percentile is readily found from the cumulative-frequency graph, (Fig. 1) by running a horizontal line through the p percent mark on the vertical axis on the graph, then a vertical line from this point on the graph to the horizontal axis; the abscissa of the intersection is the value of the pth percentile.

The standard deviation is a measure of variability that is more convenient than percentile differences for further investigation and analysis of statistical data. The standard deviation of a set of measurements x_1, x_2, …, x_n, with the mean x is defined as the square root of the mean of the squares of the deviations; it is usually designated by the Greek letter sigma (s). In symbols

$$\sigma = \sqrt{\frac{1}{n}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2\right]} = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2}$$

The square, s^2, of the standard deviation is called the variance. If the standard deviation is small, the measurements are tightly clustered around the mean; if it is large, they are widely scattered.

Correlation

When two social, physical, or biological phenomena increase or decrease proportionately and simultaneously because of identical external factors, the phenomena are correlated positively; under the same conditions, if one increases in the same proportion that the other decreases, the two phenomena are negatively correlated. Investigators calculate the degree of correlation by applying a coefficient of correlation to data concerning the two phenomena. The most common correlation coefficient is expressed as

$$\frac{\Sigma\left(\frac{x}{\sigma^x}\cdot\frac{y}{\sigma^y}\right)}{N}$$

in which *x* is the deviation of one variable from its mean, *y* is the deviation of the other variable from its mean, and *N* is the total number of cases in the series. A perfect positive correlation between the two variables results in a coefficient of +1, a perfect negative correlation in a coefficient of -1, and a total absence of correlation in a coefficient of 0. Intermediate values between +1 and 0 or -1 are interpreted by degree of correlation. Thus, .89 indicates high positive correlation, -.76 high negative correlation, and .13 low positive correlation.

Mathematical Models

A mathematical model is a mathematical idealization in the form of a system, proposition, formula, or equation of a physical, biological, or social phenomenon. Thus, a theoretical, perfectly balanced die that can be tossed in a purely random fashion is a mathematical model for an actual physical die. The probability that in *n* throws of a mathematical die a throw of 6 will occur *k* times is

$$P(k)=\binom{n}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k}$$

in which (,) is the symbol for the binomial coefficient

$$\frac{n(n-1)...(n-k+1)}{1\cdot 2\cdot \;...\; \cdot k}\cdot\left(\binom{n}{0}=1\right)$$

The statistician confronted with a real physical die will devise an experiment, such as tossing the die *n* times repeatedly, for a total of *Nn* tosses, and then determine from the observed throws the likelihood that the die is balanced and that it was thrown in a random way.

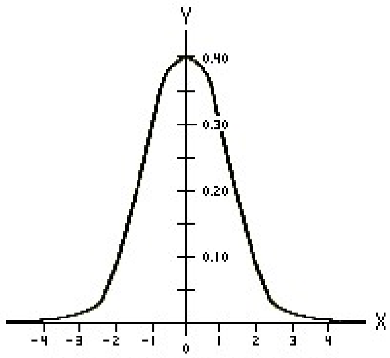
In a related but more involved example of a mathematical model, many sets of measurements have been found to have the same type of frequency distribution. For example, let $x_1, x_2, \ldots, x_N$ be the number of 6's cast in the *N* respective runs of *n* tosses of a die and assume *N* to be moderately large. Let $y_1, y_2, \ldots, y_N$ be the weights, correct to the nearest 1/100 g, of *N* lima beans chosen haphazardly from a 100-kg bag of lima beans. Let $z_1, z_2, \ldots, z_N$ be the barometric pressures recorded to the nearest 1/1000 cm by *N* students in succession, reading the same barometer. It will be observed that the *x*'s, *y*'s, and *z*'s have amazingly similar frequency patterns. The statistician adopts a model that is a mathematical prototype or idealization of all these patterns or distributions. One form of the mathematical model is an equation for the frequency distribution, in which *N* is assumed to be infinite:

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

in which *e* (approximately 2.7) is the base for natural logarithms (*see* Logarithm). The graph of this equation (Fig. 4) is the bell-shaped curve called the normal, or Gaussian, probability curve. If a variate *x* is normally distributed, the probability that its value lies between *a* and *b* is given by

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{(-(x^2)/2)} dx$$

The mean of the *x*'s is 0, and the standard deviation is 1. In practice, if *N* is large, the error is exceedingly small.



The normal probability curve.

Figure 4

Tests of Reliability

The statistician is often called upon to decide whether an assumed hypothesis for some phenomenon is valid or not. The assumed hypothesis leads to a mathematical model; the model, in turn, yields certain predicted or expected values, for example, 10, 15, 25. The corresponding actually observed values are 12, 16, 21. To determine whether the hypothesis is to be kept or rejected, these deviations must be judged as normal fluctuations caused by sampling techniques or as significant discrepancies. Statisticians have devised several tests for the significance or reliability of data. One is the chi-square (c^2) test. The deviations (observed values minus expected values) are squared, divided by the expected values, and summed:

$$\chi^2 = \frac{(12-10)^2}{10} + \frac{(16-15)^2}{15} + \frac{(21-25)^2}{25} = 1.11$$

The value of c^2 is then compared with values in a statistical table to determine the significance of the deviations.

Higher Statistics

The statistical methods described above are the simpler, more commonly used methods in the physical, biological, and social sciences. More advanced methods, often involving advanced mathematics, are used in further statistical studies, such as sampling theory, inference and estimation theory, and design of experiments.

Linear Programming, mathematical and operations-research technique, used in administrative and economic planning to maximize the linear functions of a large number of variables, subject to certain constraints. The development of high-speed electronic computers and data-processing techniques has brought about many recent advances in linear programming, and the technique is now widely used in industrial and military operations.

Linear programming is basically used to find a set of values, chosen from a prescribed set of numbers, that will maximize or minimize a given polynomial form and this is illustrated by the following example of a particular kind of problem and a method of solution. A manufacturer makes two varieties, $V_1$and V_2, of an article having parts that must be cut, assembled, and finished; the manufacturer knows that as many articles as are produced can be sold. Variety V_1 takes 25 min to cut, 60 min to assemble, and 68 min to finish; it yields \$30 profit. Variety V_2 takes 75 min to cut, 60 min to assemble, and 34 min to finish, and yields a \$40 profit. Not more than 450 min of cutting time, 480 min of assembly time, and 476 min of finishing time are available each day. How many articles of each variety should be manufactured each day to maximize profit?

Let *x* and *y* be the numbers of articles of varieties V_1 and V_2, respectively, that should be manufactured each day to maximize profit. Because *x* and *y* cannot be negative numbers,

$$x \geq 0 \qquad (1)$$
$$y \geq 0 \qquad (2)$$

The cutting, assembly, and finishing data determine the following inequalities and equalities:

$$25x + 75y = 450 \qquad (3)$$
$$60x + 60y = 480 \text{ and} \qquad (4)$$
$$68x + 34y = 476 \qquad (5)$$

The profit is given by

$$p(x,y) = 30x + 40y \qquad (6)$$

The problem is to find the values of *x* and *y,* if any, subject to restrictions (1) through (5), that will maximize the linear polynomial or linear form (6).

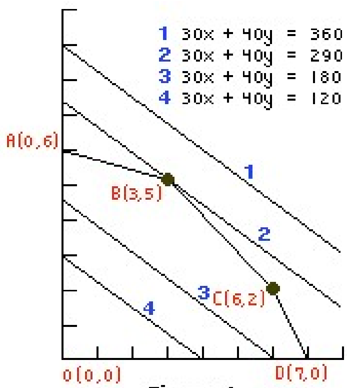


Figure 1

The equation $25x + 75y = 450$ represents a straight line in the Cartesian plane (*see* Geometry); if point *P* has coordinates (*r, s*), *P* is above the line, on the line, or below the line, as $25r + 75s$ is greater than, equal to, or less than 450. Therefore, condition (3) is satisfied by the coordinates of any point in the half plane determined by the line $25x + 75y = 450$ and all points below it. Similarly, each of the conditions (1), (2), (4), and (5) is satisfied by the coordinates of a point in a certain half plane. To satisfy all five conditions, the point must lie on the boundary or interior of the convex, polygonal region OABCD in Figure 1. The region is convex because if *R* and *S* are any two points of the region, so is every point of the line segment *RS;* it is polygonal because its boundary consists of line segments.

The equation $30x + 40y = p$, indicating the profit, also represents a straight line; the equation $30x + 40y = p'$ represents a parallel line that is above, coincides with, or is below the first line, as p' is greater than, equal to, or less than *p.* The profit will be maximized by choosing the line, of the family of parallel lines, that just touches the region OABCD above, namely, the line through the vertex B(3,5). The manufacturer will earn a maximum profit (of $290) if 3 articles of variety V_1 and 5 of variety V_2 are made per day. Any other quantities of the two varieties, within the constraints of the time limitations, will yield a smaller profit.

Linear programming is applied to many other kinds of problems, and many other methods of solution exist, but the above example is generally illustrative.

Game Theory, mathematical analysis of any situation involving a conflict of interest, with the intent of indicating the optimal choices that, under given conditions, will lead to a desired outcome. Although game theory has roots in the study of such well-known amusements as checkers, ticktacktoe, and poker—hence the name—it also involves much more serious conflicts of interest arising in such fields as sociology, economics, and political and military science.

Aspects of game theory were first explored by the French mathematician Émile Borel, who wrote several papers on games of chance and theories of play. The acknowledged father of game theory, however, is the Hungarian-American mathematician John von Neumann, who in a series of papers in the 1920s and '30s established the mathematical framework for all subsequent theoretical developments. During World War II military strategists in such areas as logistics, submarine warfare, and air defense drew on ideas that were directly related to game theory. Game theory thereafter developed within the context of the social sciences. Despite such empirically related interests, however, it is essentially a product of mathematicians.

Basic Concepts

In game theory, the term *game* means a particular sort of conflict in which *n* of individuals or groups (known as players) participate. A list of rules stipulates the conditions under which the game begins, the possible legal "moves" at each stage of play, the total number of moves constituting the entirety of the game, and the terms of the outcome at the end of play.

Move

In game theory, a move is the way in which the game progresses from one stage to another, beginning with an initial state of the game through the final move. Moves may alternate between players in a specified fashion or may occur simultaneously. Moves are made either by personal choice or by chance; in the latter case an object such as a die, instruction card, or number wheel determines a given move, the probabilities of which are calculable.

Payoff

Payoff, or *outcome,* is a game-theory term referring to what happens at the end of a game. In such games as chess or checkers, payoff may be as simple as declaring a winner or a loser. In poker or other gambling situations the payoff is usually money; its amount is predetermined by antes and bets amassed during the course of play, by percentages or by other fixed amounts calculated on the odds of winning, and so on.

Extensive and Normal Form

One of the most important distinctions made in characterizing different forms of games is that between extensive and normal. A game is said to be in extensive form if it is characterized by a set of rules that determines the possible moves at each step, indicating which player is to move, the probabilities at each point if a move is to be made by a chance determination, and the set of outcomes assigning a particular payoff or result to each possible conclusion of the game. The assumption is also made that each player has a set of preferences at each move in anticipation of possible outcomes that will maximize the player's own payoff or minimize losses. A game in extensive form contains not only a list of rules governing the activity of each player, but also the preference patterns of each player. Common parlor games such as checkers and ticktacktoe and games employing playing cards such as “go fish” and gin rummy are all examples.

Because of the enormous numbers of strategies involved in even the simplest extensive games, game theorists have developed so-called normalized forms of games for which computations can be carried out completely. A game is said to be in normal form if the list of all expected outcomes or payoffs to each player for every possible combination of strategies is given for any sequence of choices in the game. This kind of theoretical game could be played by any neutral observer and does not depend on player choice of strategy.

Perfect Information

A game is said to have perfect information if all moves are known to each of the players involved. Checkers and chess are two examples of games with perfect information; poker and bridge are games in which players have only partial information at their disposal.

Strategy

A strategy is a list of the optimal choices for each player at every stage of a given game. A strategy, taking into account all possible moves, is a plan that cannot be upset, regardless of what may occur in the game.

Kinds of Games

Game theory distinguishes different varieties of games, depending on the number of players and the circumstances of play in the game itself.

One-Person Games

Games such as solitaire are one-person, or singular, games in which no real conflict of interest exists; the only interest involved is that of the single player. In solitaire only the chance structure of the shuffled deck and the deal of cards come into play. Single-person games, although they may be complex and interesting from a probabilistic view, are not rewarding from a game-theory perspective, for no adversary is making independent strategic choices with which another must contend.

Two-Person Games

Two-person, or dual, games include the largest category of familiar games such as chess, backgammon, and checkers or two-team games such as bridge. (More complex conflicts—*n*-person, or plural, games—include poker, Monopoly, Parcheesi, and any game in which multiple players or teams are involved.) Two-person games have been extensively analyzed by game theorists. A major difficulty that exists, however, in extending the results of two-person theory to *n*-person games is predicting the interaction possible among various players. In most two-party games the choices and expected payoffs at the end of the game are generally well-known, but when three or more players are involved, many interesting but complicating opportunities arise for coalitions, cooperation, and collusion.

Zero-Sum Games

A game is said to be a zero-sum game if the total amount of payoffs at the end of the game is zero. Thus, in a zero-sum game the total amount won is exactly equal to the amount lost. In economic contexts, zero-sum games are equivalent to saying that no production or destruction of goods takes place within the "game economy" in question. Von Neumann and Oskar Morgenstern showed in 1944 that any *n*-person non-zero-sum game can be reduced to an $n + 1$ zero-sum game, and that such $n + 1$ person games can be generalized from the special case of the two-person zero-sum game. Consequently, such games constitute a major part of mathematical game theory. One of the most important theorems in this field establishes that the various aspects of maximal-minimal strategy apply to all two-person zero-sum games. Known as the minimax theorem, it was first proven by von Neumann in 1928; others later succeeded in proving the theorem with a variety of methods in more general terms.

Applications

Applications of game theory are wide-ranging and account for steadily growing interest in the subject. Von Neumann and Morgenstern indicated the immediate utility of their work on mathematical game theory by linking it with economic behavior. Models can be developed, in fact, for markets of various commodities with differing numbers of buyers and sellers, fluctuating values of supply and demand, and seasonal and cyclical variations, as well as significant structural differences in the economies concerned. Here game theory is especially relevant to the analysis of conflicts of interest in maximizing profits and promoting the widest distribution of goods and services. Equitable division of property and of inheritance is another area of legal and economic concern that can be studied with the techniques of game theory.

In the social sciences, *n*-person game theory has interesting uses in studying, for example, the distribution of power in legislative procedures. This problem can be interpreted as a three-person game at the congressional level involving vetoes of the president and votes of representatives and senators, analyzed in terms of successful or failed coalitions to pass a given bill. Problems of majority rule and individual decision making are also amenable to such study.

Sociologists have developed an entire branch of game theory devoted to the study of issues involving group decision making. Epidemiologists also make use of game theory, especially with respect to immunization procedures and methods of testing a vaccine or other medication. Military strategists turn to game theory to study conflicts of interest resolved through "battles" where the outcome or payoff of a given war game is either victory or defeat. Usually, such games are not examples of zero-sum games, for what one player loses in terms of lives and injuries is not won by the victor. Some uses of game theory in analyses of political and military events have been criticized as a dehumanizing and potentially dangerous oversimplification of necessarily complicating factors. Analysis of economic situations is also usually more complicated than zero-sum games because of the production of goods and services within the play of a given “game.”

Decision Theory, formal study of the making of decisions. Real-life studies, which use surveys and experimentation, are called descriptive decision theory; studies of rational decision making, which employ logic and statistics, are called prescriptive decision theory. Such studies grow progressively more complex when more than one person is involved, when outcomes of various options are not known with certainty, and when even the probabilities of outcomes are unknown. Decision theory shares characteristics with *game theory*, but in decision theory the “opponent” is reality, not another player or players.