



EC50– Acronym for *effective concentration 50*.

ecological correlation– Same as *geographic correlation*.

ecological fallacy– See *geographic correlation*.

ecological statistics– Statistical methods and techniques used in the study of the dynamics of natural habitats and their interaction to environment. See also *environmental statistics*.

econometrics– A branch of **economics** concerned with the **empirical** study of economic laws by application of mathematical and statistical techniques, usually modeling economic phenomena involving stochastic elements. It expresses economic principles and theories in mathematical terms in order to verify them by statistical methods. It is mainly concerned with the empirical **measurement** and testing of economic relations that are expressible in mathematical form.

economic model– A set of mathematical equations designed to provide a quantitative explanation of the behavior of economic **variables**.

economics– The field of study concerned with production, distribution, and consumption of goods and services. Theoretical economics is concerned with the study of economic principles and laws while **applied economics** employs theoretical principles in developing economic programs and policies.

EDA– Acronym for *exploratory data analysis*.

effect– In a **factorial experiment**, a quantity representing a change in response caused by a change in **level** of one or more of the **factors**. In **analysis of variance** and **regression**, a change in **response variable** caused by a change in one or more **explanatory variables**.

effective concentration 50– Same as *median lethal dose*.

effective sample size– The final **sample size** after taking into account losses due to **attrition, nonresponse, drop-outs**, and any other causes.

effect size– The magnitude of a difference or relationship between two **treatments** or **variables** considered of importance to be detected in a study. To obtain the effect size, the magnitude of the difference is generally divided by the **standard deviation** of the **measurement**. It is the basis for statistical methods used in **meta-analysis** and the computation of **power** and **sample size**.

efficiency– A term most commonly used in the context of comparing **variances** of two **unbiased estimators**; an **estimator** being regarded more efficient than another if it has smaller variance. If T_1 and T_2 are two unbiased estimators of the same **parameter** with variances V_1 and V_2 , then the efficiency of T_1 with respect to T_2 is defined by the **ratio** V_2/V_1 . In an **experimental design**, a design is said to be more efficient if its **error mean square** is less than that of another design applied to the same number of experimental units. The term is also used for **power efficiency** of one test with respect to the other.

efficient estimator– The **sample estimator** or **statistic**, among the entire class of **unbiased estimators**, that has the smallest **variance** for a given **sample size**.

elementary event– In **probability theory**, the single elementary result of an **experiment** or **trial** that rules out the occurrence of all the alternative results. Observing a 7 for the total of face values when throwing a pair of dice and drawing an ace from a deck of cards are examples of an elementary event.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|---|---|---|
| A | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | J | Q | K |
| ♣ | ♣ | ♣ | ♣ | ♣ | ♣ | ♣ | ♣ | ♣ | ♣ | ♣ | ♣ | ♣ |
| A | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | J | Q | K |
| ♦ | ♦ | ♦ | ♦ | ♦ | ♦ | ♦ | ♦ | ♦ | ♦ | ♦ | ♦ | ♦ |
| A | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | J | Q | K |
| ♥ | ♥ | ♥ | ♥ | ♥ | ♥ | ♥ | ♥ | ♥ | ♥ | ♥ | ♥ | ♥ |
| A | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | J | Q | K |
| ♠ | ♠ | ♠ | ♠ | ♠ | ♠ | ♠ | ♠ | ♠ | ♠ | ♠ | ♠ | ♠ |

Elementary events for an experiment consisting of drawing a card from a deck of cards

elementary unit– A person or object possessing a certain characteristic of interest to an investigator. It is the smallest unit yielding information concerning the characteristic under investigation.

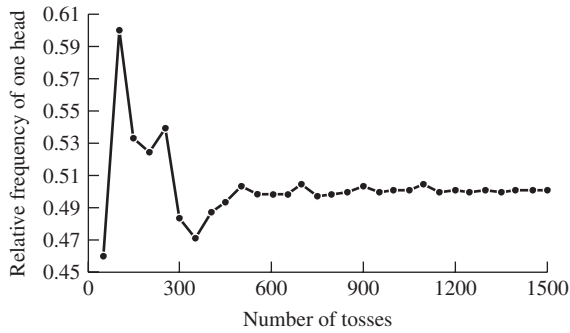
eligibility– A term used in a **clinical trial** to describe the criteria each patient must satisfy before entering a study.

empirical– A term commonly used to denote results based on **experimental data** rather than deduced from theoretical considerations.

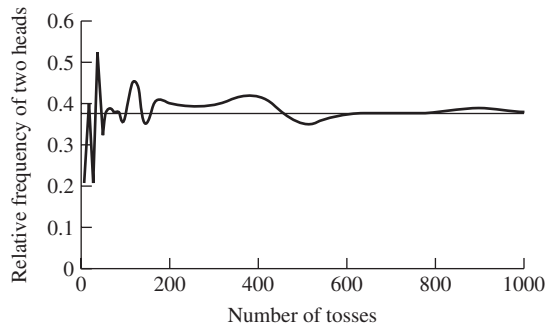
empirical Bayes method– A form of **Bayesian inference** in which the **prior distribution** is determined from some **empirical** evidence rather than the investigator's prior knowledge about the **parameters**. The empirical evidence is generally derived from the use of **data** previously collected by the same selection procedure as now proposed for use in a new **experiment** to be conducted on the same **study population**.

empirical probability– An **estimate** of **probability** based on past **experimental data** with the **outcomes** of an **experiment** or some other phenomenon. It is equal to the number

of times an **event** did occur in a large number of experimental **trials** divided by the maximum number of times the event could have occurred during these trials. See also *classical probability*, *objective probability*, *subjective probability*.



The relative frequency approaches the true probability of 0.5
Empirical probability as the relative frequency of obtaining a head; from a computer simulation



The relative frequency approaches the true probability of 0.375
Empirical probability as the relative frequency of obtaining two heads in many tosses of four coins; from a computer simulation

empirical rule– A rule that is useful in interpreting the **variability** of a **bell-shaped distribution**. The rule states that approximately 68% of the **observations** in a **data set** will be within one **standard deviation** of their **mean**, 95% of the observations will be within two standard deviations of the mean, and 99.7% of the observations will be within three standard deviations of the mean. See also *normal distribution*.

endogenous variable– A **variable** whose **variability** is assumed to be determined by variables in the causal system. In other words, an endogenous variable is a variable that is caused by variables internal to a causal system. For example, price and demand are considered endogenous to an **economic model**. Similarly, consumption, savings, investment, private wage payments, and profit are generally considered endogenous variables in studies of a nation's aggregate economic activity. Compare *exogenous variable*.

endpoint– A term commonly used in medical and health science investigation to describe a well-defined **event** or **outcome**, such as infection, myocardial arrest, death, or relapse.

Sometimes **surrogate outcomes** are used as endpoints because of their strong relationship with more definitive outcomes of interest.

enumerator– The person who carries out the enumeration of the households in a **census** or **survey** operation.

environmental statistics– Statistical methods and techniques used in the study of environment, especially in environmental pollution and monitoring involving soil, air, water, solid wastes, and hazardous substances. See also *ecological statistics*.

epidemiology– The study of the distribution and causes of disease in a **population** and the methods and techniques for acquiring such knowledge.

equal ignorance principle– Same as *equal-likelihood criterion*.

equal-likelihood criterion– In **decision theory**, the assignment of equal **prior probabilities** to all possible outcomes in the absence of any information about the likelihood of occurrence of any of these **outcomes**.

error– A general term used to describe any mistake associated with any action such as transcribing error, judgment error, or **observation** error. In statistics the term is used in a very limited context to describe the difference between the “true” or “expected” value and the observed value with no implication of any mistake. See also *random error*, *unexplained variation*.

error effect– In a **statistical model** the **effect** attributable to the **error term**.

error mean square– The **mean square** used in the denominator of an F test in an **analysis of variance** procedure. It is obtained by dividing the **error sum of squares** by its **degrees of freedom**. It provides an **unbiased estimator** of the common **error variance**.

error of acceptance– Same as *type II error*.

error of estimation– Same as *estimation error*.

error of rejection– Same as *type I error*.

error of the first kind– Same as *type I error*.

error of the second kind– Same as *type II error*.

error rate– A term sometimes used to designate the **rate** at which the **error of the first kind** will be allowed to occur. In general the rate at which any type of error can occur, for example, the **proportion** of cases misclassified by a classification rule derived from a **discriminant analysis**.

errors of classification– Same as *classification errors*.

error sum of squares– In an **analysis of variance**, the sum of squared deviations of all individual **observations** from the **sample means** of their respective **treatment groups**. It is algebraically equal to the **total sum of squares** minus **treatment sum of squares**. It is also called residual sum of squares or **within** group sum of squares.

error term– In a **statistical model** the term representing the contribution from various other variables, known or unknown, which are omitted from the model.

error variance– In an **analysis of variance**, the contribution to the **variance** that is not ascribed to **treatment** or **block** effects. It may be due to many causes such as individual

differences between subjects, inconsistencies in the experimental conditions, **measurement errors**, or any uncontrolled or **unexplained variation**. It is also referred to as the error variance of the **error term**.

establishment survey– A **survey** of business and commercial enterprises usually conducted monthly by a government agency of a country. In the United States such surveys are carried out by the **Bureau of the Census**.

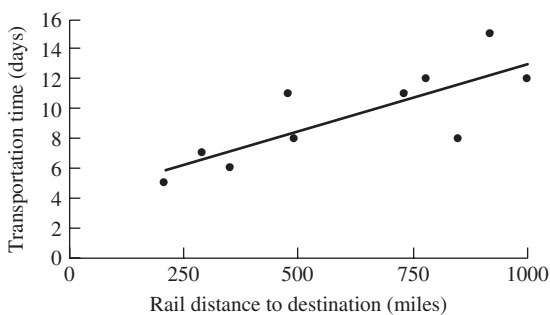
estimate– An estimate is the particular numerical value yielded by an **estimator** for a given **sample data**. An estimate can be a **mean, proportion, correlation coefficient** or any other **parameter value** derived from a **sample**. An estimate is used to make **inference** about a **target population** whose true parameter value is unknown. See also *estimation*.

estimated partial regression coefficient– The **estimated regression coefficient** of an **independent variable in a multiple regression model**; it is interpreted as an **estimate** of the net change in the **dependent variable** for a unit change in the independent variable, while other independent variables are kept **constant**. See also *partial regression coefficient*.

estimated regression coefficient– An **estimate** of the **regression coefficient** obtained by **sample data**, using the **least squares** or any other method of **estimation**.

estimated regression equation– An **estimate** of the **regression equation** obtained by **sample data** using the **least squares** or any other method of **estimation**.

estimated regression line– An **estimate** of the **regression line** obtained by **sample data**, using the **least squares** or any other method of **estimation**. A **graphical representation** of the estimated regression line drawn through the pattern of points on a **scatter diagram** summarizes and averages out the relationship between the **dependent** and **independent variables**.



An estimated regression line

estimated regression model– Same as *estimated regression equation*.

estimation– The process of using information from **sample data** in order to estimate the numerical values of unknown **parameters** in a **population**. If a single value is calculated to estimate a parameter, the process is called **point estimation**. If an interval is calculated, the process is called **interval estimation**. See also *confidence interval, inferential statistics, least squares estimation, maximum likelihood estimation*.

estimation error– The difference between an **estimate** and the true value of the **parameter** being estimated.

estimation of parameter– Same as *estimation*.

estimator– The **sample statistic** used to make **inferences** about an unknown **parameter**. For example, one might use **sample mean** to estimate the value of the **population mean**. An estimator is usually given as algebraic formula. See also *estimation*.

eta– The **correlation** between two **variables** measured on an **interval scale** that is an index of the nonlinear relationship between the variables. It is also known as correlation ratio.

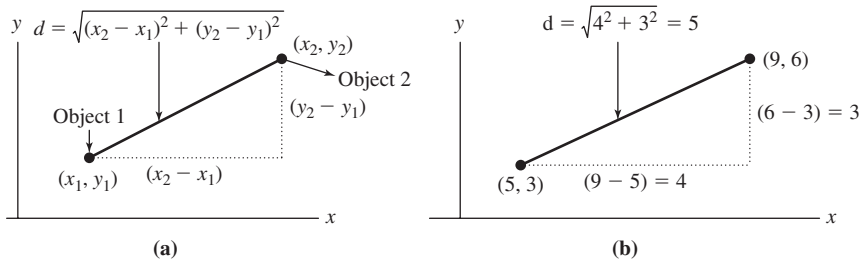
etiologic fraction– Same as *attributable risk*.

etiologic factor– Same as *risk factor*.

Euclidean distance– A measure of the distance between two points as determined by the location of their coordinates. The Euclidean distance between two points: (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_n) is determined by the formula

$$\sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2 + \dots + |x_n - y_n|^2}$$

where $|x_i - y_i|$ represents the **absolute value** of $(x_i - y_i)$ for $i = 1, 2, \dots, n$. Note that $|x_i - y_i|^2 = (x_i - y_i)^2$ when the coordinates are real numbers. It is a special case of the **Minkowski distance**.



The Euclidean distance between two objects measured on two variables:
(a) general definition and **(b)** specific example

event– A set consisting of a collection of **sample points** or **outcomes** of an **experiment**. In **probability theory**, an event is a subset of the **sample space**. In general, the term is used to represent any outcome, condition, or eventuality.

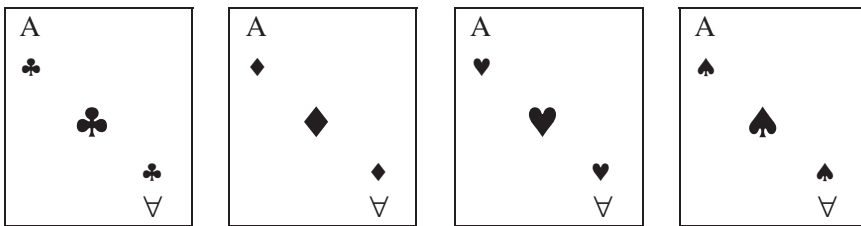


Illustration of the event “an ace card is selected”

event branches– In a **decision tree** diagram, branches emanating from an **event point** and representing the possible **outcomes** confronting the decision maker.

event point– In a **decision tree** diagram, a point representing a **random event** over which the decision maker has no control. It is usually symbolized by a circle.

exact hypothesis– A **hypothesis** that specifies a single value for an unknown **parameter**. See also *simple hypothesis*.

exact test– A **statistical test** is called exact if its **level of significance** is exactly equal to the nominal or stated level. See also *approximate test, conservative test, liberal test*.

exhaustive– A set of conditions, **events**, or values is said to be exhaustive if, taken together, the components account for all the possible **outcomes**.

exhaustive events– A set of **events** is said to be exhaustive if the components jointly contain all the **outcomes** in the **sample space**, that is, there are no other possible outcomes. The sum of the **probabilities** of the exhaustive events equals 1.

exogenous variable– A **variable** whose **variability** is assumed to be determined by causes outside the causal system. In other words, an exogenous variable is a variable that is caused by variables external to a causal system. For example, rainfall and natural disasters are considered exogenous to an **economic model**. Similarly, export statistics are usually considered an exogenous variable in a study of a nation's aggregate economic activity. Compare *endogenous variable*.

expanded safety trial– A type of surveillance trial designed primarily for estimating the frequency of unusual side effects as a consequence of administering a treatment. In drug development studies, the term is synonymous with **phase II trial**.

expectation– Same as *expected value*.

expectation of life at birth– The number of years a newborn child is expected to live, under the prevailing social, economic, and health conditions in the population. See also *life table*.

expected frequency– The **frequency** expected for an **event** if certain **probability laws** were exactly followed as distinct from the actual frequency that may be observed in a **sample**. In a **contingency table**, the frequency expected for each **cell** if the **null hypothesis** of **independence** or **homogeneity** were true. The expected frequency for a cell is obtained by multiplying its row total by its column total and dividing the result by the grand total. Compare *observed frequency*.

expected mean square– In an **analysis of variance**, the **expected value** of a **mean square** derived under a given set of **assumptions** of the **model** being postulated.

expected monetary gain– Same as *expected monetary value*.

expected monetary return– Same as *expected monetary value*.

expected monetary value– The **weighted average** of the **payoffs** associated with an action, the weights being the **probabilities** of the alternative **outcomes** that give rise to the various possible payoffs.

expected monetary value criterion– One of several probabilistic criteria for making decisions under **uncertainty**. According to this criterion, a decision maker determines an

expected monetary value for each possible action and selects the action that maximizes expected monetary value.

expected opportunity loss– The **weighted average** of the **opportunity loss** values associated with an action, the weights being the **probabilities** of the alternative **outcomes** that give rise to the various possible opportunity losses.

expected opportunity loss criterion– One of several probabilistic criteria for making decisions under **uncertainty**. According to this criterion, a decision maker determines an **expected opportunity loss** for each possible action and selects the action with the smallest of these values.

expected regret value– Same as *expected opportunity loss*.

expected utility– The **weighted average** of the **utilities** associated with an action, the weights being the **probabilities** of the alternative **outcomes** that give rise to the various possible utility **payoffs**.

expected utility criterion– One of several probabilistic criteria for making decisions under **uncertainty**. According to this criterion, a decision maker determines the **expected utility** for each possible action and selects the action that maximizes the expected utility.

expected value– The expected value of a **random variable** is the **weighted mean** of its **probability distribution**. It can be interpreted as the value of the random variable one can expect to obtain, on the **average**, in successive repetitions of the **random experiment** that generates the values of the random variable.

expected value of perfect information– The maximum amount a decision maker can be expected to pay for obtaining complete information about future **outcomes** and, thus, for eliminating **uncertainty** entirely.

expected value of sample information– The maximum amount a decision maker can be expected to pay for obtaining supposedly incomplete information about future **outcomes** and, thus, for reducing, rather than eliminating, **uncertainty**.

experiment– In **probability theory**, any process or operation that generates well-defined **elementary events** or **outcomes**; for example, tossing a coin or casting a die. In **statistics**, any study undertaken in which the researcher has control over some of the experimental conditions under which the study is undertaken and **measurements** or **observations** of possible outcomes are obtained. In particular, the investigator controls the conditions applied to the subjects and then carefully records the observations on outcomes of interest. The experiment is one of the distinctive tools of the scientist. It enables the scientist to put questions to nature and test **hypotheses** under controlled conditions.

experimental data– **Data** obtained from an **experiment**.

experimental design– Same as *design of experiment*.

experimental error– A term used to refer to the **errors** introduced into an **experiment** by the lack of uniformity in the conduct of the experiment and failure to standardize the use of materials and techniques. Results of an experiment are affected not only by the **treatments** (experimental procedures whose effects are being evaluated and compared) but also by the presence of experimental errors. The presence and cause of experimental error need not concern the investigator provided the results are sufficiently accurate to permit definite

conclusions. Often, however, the results of an experiment can be greatly influenced by presence of large experimental errors, making it difficult to draw any valid **inferences**.

experimental group– A group that receives a **treatment** or **intervention** and is compared to a **control group**.

experimental observations– Same as *experimental data*.

experimental planning– A term used to refer to the details of the proposal for and objectives of an **experiment**, including definitions of **treatments**, experimental materials and techniques, and other related **variables** and procedures. See also *experimental design*.

experimental study– A **comparative study** involving an **intervention** or manipulation of experimental conditions by the investigator. It is called a **clinical trial** when human subjects are used. See also *experiment, observational study*.

experimental treatment– The active **treatment** or **intervention** being assigned to the **treatment group**.

experimental unit– The object or item of interest in an **experiment**. It is the smallest independent unit of study assigned to a particular **treatment**.

experimentwise error rate– The experimentwise error rate is the **probability** that at least one (i.e., one or more) of the inferences to be drawn from the same set of **data** will be wrong. It is equivalent to the probability of incorrectly rejecting at least one of the **null hypotheses** in an **experiment** involving one or more tests or comparisons. In a **multiple comparison** procedure, it is the **significance level** associated with the entire set of comparisons of interest to the investigator. See also *comparisonwise error rate*.

explained deviation– In a **regression analysis**, the difference between the regression estimate of an individual **observation** and the **mean** of all the observations of the **dependent variable**.

explained variable– Same as *dependent variable*.

explained variance– Same as *explained variation*.

explained variation– The amount of shared **variation** between two correlated **variables**. In a **regression analysis**, it is the sum of the squares of all the **explained deviations**, also called the **regression sum of squares**. It is obtained by subtracting the **mean** of a set of **observations** from the value predicted by the **linear regression** and squaring and summing these values. It is interpreted as the variation in the **dependent variable** that can be accounted for by variation in the **independent variables**. See also *coefficient of multiple determination*.

explanatory analysis– In a **clinical trial**, a term used to refer to the analysis performed to compare two **treatments** under the assumption that patients remain on their treatment to which they were initially randomized. See also *intention-to-treat analysis*.

explanatory trial– A **clinical trial** designed to explain the process of a **treatment**.

explanatory variable– See *independent variable, predictor variable*.

exploratory analysis– Same as *exploratory data analysis*.

exploratory data analysis– Any of several modern **graphical techniques**, pioneered by John W. Tukey, often by presenting **quantitative data** visually by the use of simple

arithmetic and easy-to-draw **diagrams** with a view to examine the **data** more effectively and to discover unanticipated patterns and relationships. It emphasizes the use of informal graphical procedures rather than formal **models** based on prior assumptions. The development of exploratory data analysis has been greatly aided by the widespread availability of modern electronic computers for calculation and for efficient **graphical display**. The **stem-and-leaf plot** and **box-and-whisker diagram** are two well-known examples of exploratory data analysis. Compare *confirmatory data analysis*.

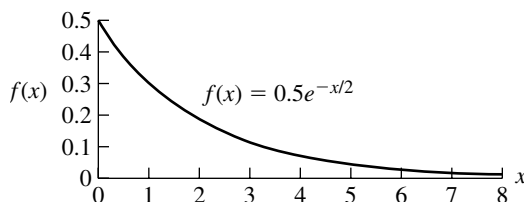
exploratory factor analysis— See *factor analysis*.

exponent— The power to which a number is raised.

exponential distribution— A **continuous probability distribution** defined by the **probability density function** of the form

$$f(x) = \theta e^{-\theta x} \quad \text{for } x \geq 0 \quad \text{and } \theta > 0$$

The **parameter** θ determines the **shape** of the **distribution** and is related to the **mean** and **standard deviation**, which are both equal to $1/\theta$. The distribution has important applications in life testing, **reliability** studies, **queuing theory**, and in many other areas of scientific investigation.



Exponential density function when $\theta = 0.5$

exponential family of distributions— A family of **probability distributions** that includes the **normal**, **binomial**, **Poisson**, and **gamma distributions** as special cases. The general form of the **density function** of an exponential family is given by

$$f(x) = \exp[p(\theta)k(x) + q(\theta) + s(x)]$$

where θ is a real **parameter** and $p(\theta)$, $q(\theta)$, $k(x)$, and $s(x)$ are known functions.

exponential regression— A type of **regression** in which the **prediction equation** between a continuous **dependent variable** y and a continuous **independent variable** x is represented by a **model** of the form $y_i = \alpha + \beta e^{\gamma x_i} + \epsilon_i$ where α , β , and γ are **constants** and ϵ_i is a random **error term**.

exponential smoothing— A **forecasting** technique that generates self-correcting **forecasts** by means of a built-in mechanism that adjusts for earlier forecasting errors. It makes use of exponentially weighted **moving averages** and continuously corrects for the amount by which the actual and estimated forecasts for a given period fail to conform. An essential feature of exponential smoothing for forecasting is that later periods are given greater weights than earlier periods.

exponential trend— A **trend** in **time-series data** that can be expressed as an exponential function. It is expressed by an equation of the form $y = ab^t$, where a and b are **constants** and t is time.

exposed– Same as *exposure group*.

exposed group– Same as *exposure group*.

exposure– Same as *exposure factor*.

exposure condition– Same as *risk condition*.

exposure factor– Same as *risk factor*.

exposure group– In epidemiology, a group of individuals who have been exposed to a certain **risk factor**, or possess a characteristic that is a determinant of certain health outcomes of interest.

external validity– The extent to which the findings of a study can be generalized to some **target population** of potential subjects beyond the **study population**. The external validity depends on the composition of the **study sample** and usually involves subject-matter judgment and nonstatistical considerations. Compare *internal validity*.

extraneous variable– A loosely used synonym for **confounding variable**.

extrapolation– The technique of estimating or predicting a value that falls outside the range of a series of known values. For example, in **regression analysis**, the value of the **response variable** may be estimated for a value of the **predictor variable** beyond the range of values used in estimating the **regression equation**. It is a practice that theoretical **statisticians** do not favor, and, indeed, it should be used with utmost care and discretion. The further the relationship is extended beyond the observed range, the more risky the procedure becomes. Compare *interpolation*.

extreme observations– Same as *extreme values*.

extreme values– The minimum and maximum values of a **data set**. See also *range*.