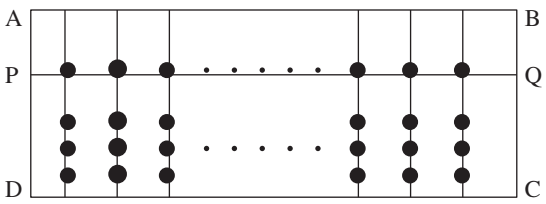
A

abacus– A simple mechanical device to facilitate arithmetical operations such as addition, subtraction, multiplication, and division by sliding counters along rods or ingrooves. There are several such instruments currently in use. The one most commonly used at present is depicted below.



The form of an abacus currently in common use

abortion rate– The number of abortions actually observed during a calendar year divided by the total number of women of childbearing age (expressed per 1000). The childbearing or reproductive age is normally defined as between 15 and 44 years.

abscissa– The **horizontal axis** or ***x* axis** on a graph using the **Cartesian coordinate** system. Generally, it refers to the baseline of most of the graphs used in **statistics.** Compare *ordinate.*

absolute central moment– See *absolute moment.*

absolute class frequency– The actual number of items or **observations** that belong to a particular class as opposed to **relative frequency** or **proportion** of items, namely the ratio of the **frequency** to the total frequency. See also *cumulative class frequency, cumulative frequency.*

absolute deviation– In general, absolute difference between any two quantities. In **statistics,** the term is normally used to denote the **absolute values** of the differences of the observed **scores** from their **mean.** See also *deviation from the mean.*

absolute frequency distribution– A **tabular representation** of a **data set** showing the actual class frequencies in each of several mutually exclusive and exhaustive classes, as opposed to **relative frequencies.** See also *frequency distribution.*

absolute moment– The **moment** of the **absolute value** of a **random variable.** The rth absolute moment about the origin of a random variable X is defined as $E[|X|^r]$. The rth absolute **central moment** of X is given by the quantity $E[|X - E(X)|^r]$.

absolute risk– Same as *incidence rate.*

absolute risk difference– In epidemiological studies, the absolute risk difference (ARD) is defined as the difference of the **risks** of an event, such as a disease or death, between two groups of subjects; for example, between the **intervention** or **exposure group** and the **control.** Algebraically, $\text{ARD} = R_i - R_c$, where R_i and R_c denote the risks in the intervention and **control group** respectively. See also *attributable risk.*

absolute value– The numerical value of a mathematical expression, regardless of its algebraic sign. Thus, the absolute value of a positive number is the number itself and the absolute value of a negative number is the positive of that number. Absolute value marks are two vertical lines (| |), one on each side of the expression under consideration.

acceptable quality level– In **quality control,** the **proportion** of nondefective items in a lot that is considered to be at an acceptable quality level by the consumer. See also *acceptance sampling.*

acceptance error– Same as *type I error.*

acceptance number– In **acceptance sampling,** the number of defective items such that the decision to accept or reject the lot depends on this number.

acceptance region– Same as *region of acceptance.*

acceptance–rejection algorithm– The name of an **algorithm** commonly employed for generating **random numbers** from a **probability distribution.**

acceptance sampling– A type of **sampling** used in **quality control** where a **sample** is taken from a batch of items and the decision to accept or reject the batch is based on the **proportion** of defective items in the sample. See also *acceptance number.*

accrual rate– The rate at which eligible patients are enrolled in a **clinical trial.** It is measured as the number of persons per unit of time.

accuracy– A term used to denote the tendency of an observed **score** to cluster around the true value being measured. The **reliability** of a measuring method depends among other things on its accuracy. In **statistical estimation,** it refers to the **deviation** of an **estimate** from the true **parameter** value. The term is not synonymous to **precision,** though sometimes they are used interchangeably. In general, the term is used for the quality of a **measurement** that is both correct and precise.

acquiescence bias– A term used in **public opinion surveys** to designate a type of **bias** caused by the tendency of certain respondents to give affirmative responses (yes, true, certainly, etc.) to a question.

action branches– In a **decision tree** diagram, branches emanating from an **action point** are called action branches. They represent the possible actions available to the decision maker.

action lines– See *control charts.*

action point– In a **decision tree** diagram, a point of choice represented by a square. This is the point at which the decision maker is in control. It is also called decision fork, decision node, and decision point.

actions– In **decision theory,** the mutually exclusive choices of decision alternatives available to a decision maker.

active controlled trial– A **clinical trial** in which **experimental treatment** is compared with some other active drug rather than an inert substance or **placebo.**

active treatment– Same as *experimental treatment.*

actuarial analysis– See *life table analysis.*

actuarial statistics– The statistical methods and techniques used in the calculation of **risks,** liabilities, insurance premium rates, policy dividends, and many other situations that arise in the insurance business. In addition to **mortality** and **morbidity** data, actuarial methods make use of **statistics** relating to the rates of return on investments and to the rates of expense involved in implementing life, health insurance, or pension programs. In many business applications, actuarial methods are employed to determine the annual retirement of plants and equipment and to provide an **estimate** of the average life facilities on the basis of detailed company records of each unit of plant and equipment.

actuary– A person, often an official of an insurance company, who is trained in the applications of mathematical and statistical procedures in the scientific study of insurance risks and premiums. In Europe, the term is sometimes used to refer to a clerk, especially one employed by a large corporation

acute angle– An angle whose magnitude is less than 90°.

adaptive sampling– Sampling designs in which the procedure for selecting units in the **sample** depends on the values of certain **variables** of interest observed during the **survey.** For example, in a survey designed to estimate the **population** of certain rare species, neighboring sites may be added to **sampling units** whenever the species is encountered in the survey.

addition of matrices– The **matrix** obtained by adding two or more matrices of the same dimension. Given two matrices $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij})$, let $\mathbf{C} = (c_{ij})$ be the matrix obtained by adding the matrices **A** and **B**, then $c_{ij} = a_{ij} + b_{ij}$, for all i and j.

addition rule for probability– A **probability law** used to calculate the **probability** for the occurrence of a **union of two** or more **events**. For any two arbitrary events A and B, it is expressed as $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. For two **mutually exclusive events,** when $P(A \cap B) = 0$, it reduces to $P(A \cup B) = P(A) + P(B)$. The rule can be generalized for more than two events.

additive effect– A term used to represent the condition when the **effect** of administering two **treatments** together is the sum of their separate effects. See also *additive model.*

additive model– A **mathematical** or **statistical model** in which the **explanatory variables** have an **additive effect** on the **response measure** of interest. For example, if a **treatment** A has an effect α on some response measure and another treatment B has an effect β on the same measure, then an additive model for A and B has a combined effect of $\alpha + \beta$.

An additive model excludes the possibility of any **interaction** between the two treatments. See also *multiplicative model.*

additive time-series model– A **classical time-series model** that represents the actual value (Y) of a **time series** as the sum of its components comprising **trend** (T), **cycle** (C), **seasonality** (S), and **irregular variation** (I); i.e., $Y = T + C + S + I$. See also *mixed time-series model, multiplicative time-series model.*

additivity– The term used to indicate the property of an **additive model.**

adherence– Same as *compliance.*

adjoint of a matrix– Given a **matrix** A, the matrix obtained by transposing A to obtain A' and then replacing each element a_{ij} by its cofactor A_{ij}. It is written as adj A. The adjoint of a matrix is useful in the evaluation of the **inverse of a matrix.** See also *cofactor of a matrix, transpose of a matrix.*

adjusted death rates– Death rates that provide an overview of the general well-being of a community or population when various demographic factors such as age, sex, and education are held constant.

adjusted means– Same as *adjusted treatment means.*

adjusted rate– A **rate** adjusted so that it is independent of the distribution of a possible **confounding variable.** In comparative studies, it refers to rate computed after taking into account a **confounding factor,** which may possibly explain the event. For example, when computing **death rates** in two populations, it may be necessary to take into account any age differences between the two populations so that age-adjusted rates are independent of the age distribution in the population to which they apply. There are a number of methods such as **stratification, standardization,** and **multiple regression** that are used to obtain adjusted rates.

adjusted sample coefficient of multiple determination– A measure, denoted by $\overline{R}^2$ or R^2_{adj}, that is the value of R^2 adjusted for the number of **independent variables** (**degrees of freedom**). It provides an **unbiased estimator** of the corresponding **population coefficient of determination.** See also *coefficient of multiple determination, sample coefficient of multiple determination.*

adjusted treatment means– A term used for **estimates** of the **treatment means** after adjusting them to the **mean** level of any **covariate**(s) that may act as **confounder**(s). Adjusted means are frequently used in **experimental design** when an increase in **precision** is desired and a concomitant **observation** is used (rather than **blocking**). The overall objective is to adjust the **average** response so that it reflects the true **effect** of the **treatment.**

adjusting for baseline– In a **longitudinal study,** the term is used to denote the process of adjusting for the **effects** of **baseline characteristics** on the **response measure** of interest.

adjustment– The process of accounting for the **effects** of **prognostic factors** or **baseline characteristics** when estimating differences attributable to **treatments** or other prognostic factors. Two primary tools for adjustment are **multiple regression** and **stratified analysis.**

age distribution– A **frequency distribution** based on **measurements** of the chronological age of the **population** under study and grouped according to **class intervals** selected to best describe the age profile of the population.

age-specific death rate– The **death rate** in a specified period of time for a given age or age group. See also *cause-specific death rate, standardized death rate.*

age-specific fertility rate– Number of live births per woman in a specified period of time for a given age or age group. See also *fertility rate.*

age-specific incidence rate– The **incidence rate** calculated for a given age or age group.

age-specific mortality rate– Same as *age-specific death rate.*

age-specific rate– The **rate** or **frequency** of occurrence of an event in a given age group.

aggregate index number– An **index number** obtained by calculating the sum of the figures applicable to each period of time under consideration, assigning index number 100 to the period chosen as the base, and determining for each of the other periods a figure that bears the same relation to 100 that the sum of the figures for that period bears to the sum of the **base period.**

aggregative model– A **statistical model** involving **variables** whose individual **observations** represent aggregates.

aleatory variable– Same as *random variable.*

algebra of events– The algebra of events defines rules for some basic operations on **events,** similar to the algebraic operations on real numbers. Some basic operations on events are the so-called union, intersection, and complementation.

algorithm– A set of well-defined rules or a formula that, when applied step by step, permits the solution of any mathematical or computational problem in a finite number of steps, for example, calculation of the roots of an equation through an **iterative procedure** or computing a rate of return on an investment.

alias– In **experimental design,** a **treatment effect** that is confounded with another **effect.** The term is especially associated with **fractional factorial designs,** in the analysis of which **estimates** of certain **contrasts** have **sums of squares** and **distributions** that reflect the existence of any one, or some, or a number of different effects. See also *confounding.*

alignment chart– Same as *nomogram.*

allocation of a sample– In **stratified random sampling,** the assignment of parts of a **sample** to different **strata** of subpopulations. See also *optimum allocation, proportional allocation.*

allometry– A field of study dealing with the quantitative relationship between size and shape of an organism. An important problem in allometry is concerned with whether one group of individuals or species represents an allometric extension of another.

all subsets regression– A type of **regression analysis** in which **models** with all possible subsets of **predictors** are fitted and the "best" one, selected by comparing the values of some appropriate criterion such as R^2 or **Mallow's** C_p **statistic.**

alpha (α)– Same as *significance level.*

alpha (α) error– Same as *type 1 error.*

alpha (α) level– Same as *alpha (α).*

alphanumeric– The term is generally used in reference to a computer statement pertaining to a character set that includes alphabetic letters, digits, and special characters such as asterisk (*), dollar sign ($), etc.

alpha (α) risk– Same as *alpha* (α).

alpha (α) value– The level of alpha (α) selected by the researcher in a **test of hypothesis.**

alternating logistic regression– A form of **logistic regression** used in the analysis of **longitudinal data** involving a **binary response** variable.

alternative hypothesis– In **hypothesis testing,** the proposition about an unknown **parameter**(s) that the researcher proposes to establish. It is also called the research hypothesis. It always states that the **population parameters**(s) have value(s) different from that specified by the **null hypothesis.** Thus, it is a complement of the null hypothesis to be concluded if the null hypothesis is rejected. In general, it is any admissible hypothesis, alternative to the null hypothesis, that is tentatively assumed to be false. It is usually denoted by H_A or H_I.

amplitude– In **time-series analysis,** the value of the series at its peak or trough measured from some **mean** value or **trend** line.

analysis of covariance– A statistical procedure for comparing the **means** of a quantitative **response variable** while taking into account the **measurements** made on one or more other quantitative **independent variables** that may act as **cofounders.** It is a special type of **analysis of variance** (or **regression**) used to control for the linear effect of a possible **confounding variate.** The confounding variate is often referred to as a **covariate.** The procedure consists of the combined application of the analysis of variance and **linear regression** techniques by using **dummy variables** to represent the groups being compared. In performing analysis of covariance, it is assumed that the covariates are unaffected by **treatments** and are linearly related to the response variable. If this assumption holds, the use of covariates decreases the **error mean square** and thus increases the **power** of the ***F* test** in testing treatment differences. The use of analysis of covariance allows the researcher to remove the effect of covariates as the source of possible explanations of **variation** in the **dependent variable.** Nowadays the term is used to describe almost any analysis seeking to assess the relationship between a response variable and a set of **explanatory variables.**

analysis of dispersion– A term sometimes used as a synonym for the **multivariate analysis of variance.**

analysis of regression– Same as *regression analysis.*

analysis of repeated measure– Same as *repeated measures analysis.*

analysis of residuals– In an **analysis of variance** or **regression,** an analysis of the differences between the observed and the expected values, known as **residuals,** in order to evaluate the validity of the **assumptions** of the **model.**

analysis of variance– Analysis of variance is a statistical procedure devised by Sir Ronald A. Fisher to analyze the results of complex **experiments** involving several **factors.** It involves a method of comparing any number of **group means** simultaneously, for determining whether or not the **means** of several **populations** are equal, by the use of one or more ***F* tests.** The ***F* statistics** are based on **sums of squares** obtained by partitioning the **total sum of squares,** calculated as the sum of squares of **deviations** of response **measurements**

about their mean, into parts on the basis of particular factors. It is an extension of the **two-sample *t* test** for comparing the means of a **quantitative variable** between two or more than two groups. The results of an analysis of variance procedure can be obtained rather conveniently by **regression** methods, by using a **dummy** or **indicator variable** to represent the groups. Like the ***t* test,** analysis of variance is based on a **model** that requires certain **assumptions** for its validity. Three main assumptions of analysis of variance are that: (1) each **treatment group** is selected randomly, with each **observation** independent of all other observations, and the treatment groups independent of each other; (2) the **samples** emanate from populations in which the observations are normally distributed; and (3) the treatment group **variances** all are assumed to be equal to a common variance σ^2. See also *analysis of variance table, multiway analysis of variance, one-way analysis of variance, three-way analysis of variance, two-way analysis of variance.*

analysis of variance *F* test– Same as *F test for analysis of variance.*

analysis of variance table– In an **analysis of variance,** a table used to summarize the results of analysis of variance calculations. It contains columns showing the sources of variation, the **degrees of freedom,** the **sums of squares,** the **mean squares,** and the values of ***F* statistics.**

analytical statistics– Same as *inferential statistics.*

ANCOVA– Acronym for *analysis of covariance.*

Anderson–Darling test– A **test procedure** for testing the **hypothesis** that a given **sample** of **observations** comes from some specified theoretical population. In particular, it is useful for testing the **normality** of a **data set.** It is based on a modified version of the **Cramér–von Mises statistic.** It is an omnibus test in the sense that it is sensitive to all types of deviations from **normality.** In addition, it is somewhat more sensitive to deviations in the tails of the **distribution,** which is frequently the way **nonnormality** makes itself known. The test is competitive with the better known **Shapiro–Wilk *W* test.** Although other tests are sometimes more powerful, they are often more difficult to calculate. The combination of ease of computation and good **power** makes it an attractive procedure for a **goodness-of-fit test.** See also *Cramér–von Mises test, D'Agostino test, Michael's test, Shapiro–Francia test.*

Andrews' plot– These are **graphical representations** of **multivariate data** in which all dimensions of the **data** are displayed. Each **data point** is depicted as a line or function running across the detecting groups of similar **observations** and assessing **outliers** in multivariate data. Statistical properties of the plots enable the **tests of significance** to be made directly from the plot. See also *Chernoff's faces.*

angular transformation– Same as *arc-sine transformation.*

annual rate– A quantity determined to reflect relative annual change for demographic or economic **data.**

annual rate of population increase (growth)– Relative change in a population size per year.

ANOVA– Acronym for *analysis of variance.*

ANOVA *F* test– Same as *F test for analysis of variance.*

Ansari–Bradley test– A **nonparametric procedure** for testing the equality of **variances** of two **populations** having the common **median.** It is assumed that the two populations being compared are of identical **shape** and differing at most in their **scale parameters.** See also *Barton–David test, Conover test, F test for two population variances, Klotz test, Mood test, Rosenbaum test, Siegel–Tukey test.*

antagonistic effect– See *interaction.*

antimode– A term sometimes used to denote the opposite of a **mode** in the sense that it corresponds to a (local) minimum **frequency.**

AOQ– Acronym for *average outgoing quality.*

a posteriori comparison– Same as *post-hoc comparison.*

a posteriori distribution– Same as *posterior distribution.*

a posteriori probabilities– Same as *posterior probabilities.*

apparent limits– The lower and upper limits actually shown for the **class intervals** of **a frequency distribution.**

applied economics– The application of principles and methods of **economics** to the solution of economic problems of a country or region.

approximate test– Often, it is not possible to obtain a test with a **level of significance** exactly equal to α, and then the test is referred to as an approximate test. See also *conservative test, exact test, liberal test.*

approximation– A mathematical result that is not exact but is sufficiently close to the exact value and can be recommended for practical use in many scientific and research applications.

a priori comparison– Same as *planned comparison.*

a priori distribution– Same as *prior distribution.*

a priori probabilities– Same as *prior probabilities.*

arc–sine transformation– The **transformation** of the form $y = \sin^{-1}(x/n)$ designed to stabilize the **variance.** The arc–sine transformation is normally used on **data** in the form of **proportion** and produces values that satisfy the assumption of **homogeneity** required in the application of **analysis of variance** and **regression** techniques. It is also called angular transformation because arc–sine is angle. A modified form of the transformation given as arc–sin $\left(\sqrt{x+\frac{3}{8}}\Big/\sqrt{n+\frac{3}{4}}\right)$ is somewhat more effective in equalizing variances. See also *logarithmic transformation, power transformation, reciprocal transformation, square-root transformation, square transformation.*

ARE– Acronym for *asymptotic relative efficiency.*

area sample– See *area sampling.*

area sampling– A type of **sampling design** employed when a complete **frame** of **reference population** is not available. The total area under the study is divided into a small number of subareas (e.g., counties, towns, blocks) that are sampled **at random** or by some

restricted **random** device. Each of the chosen subareas is then enumerated and may constitute a frame for further **sampling** in the subarea. For example, suppose a **sample** of households within a state is desired, and there does not exist a comprehensive list from which such a sample might be selected. The state in such a case might be divided into a certain geographical units, say counties, and a certain number of counties selected for the sample. Each county included in the sample might then be divided into municipalities and a certain number of municipalities selected for the sample. Each municipality included in the sample might then be divided into blocks and a certain number of blocks selected for the sample. Finally, from each block included in the sample, a certain number of households might be identified and selected in the sample. Area sampling is usually less costly and less reliable than alternative procedures, such as **stratified** or **simple random sampling.** See also *cluster sampling, multistage sampling.*

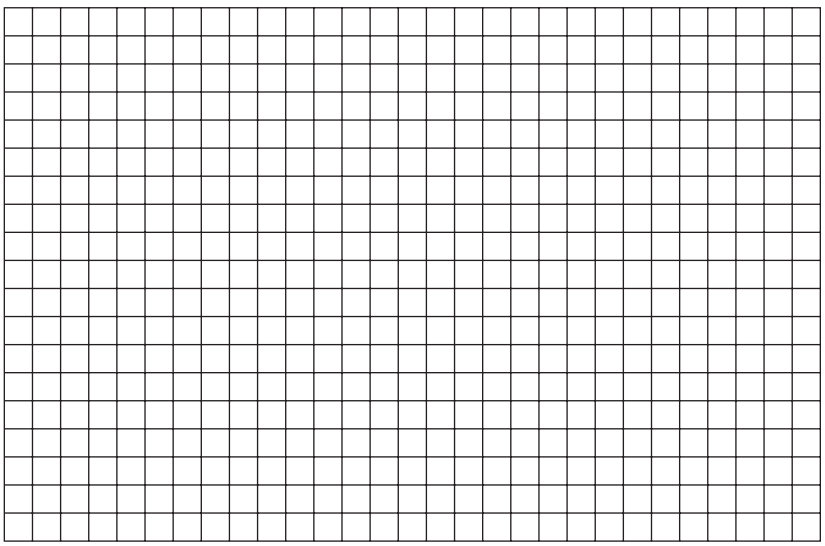
area under the curve– In pharmacokinetic studies, the term is used to describe the estimated area under a time–concentration curve. It may indicate a predictor of biological or clinical effects such as efficacy or toxicity.

ARIMA– Acronym for *autoregressive integrated moving average.*

arithmetic chart– Same as *arithmetic paper.*

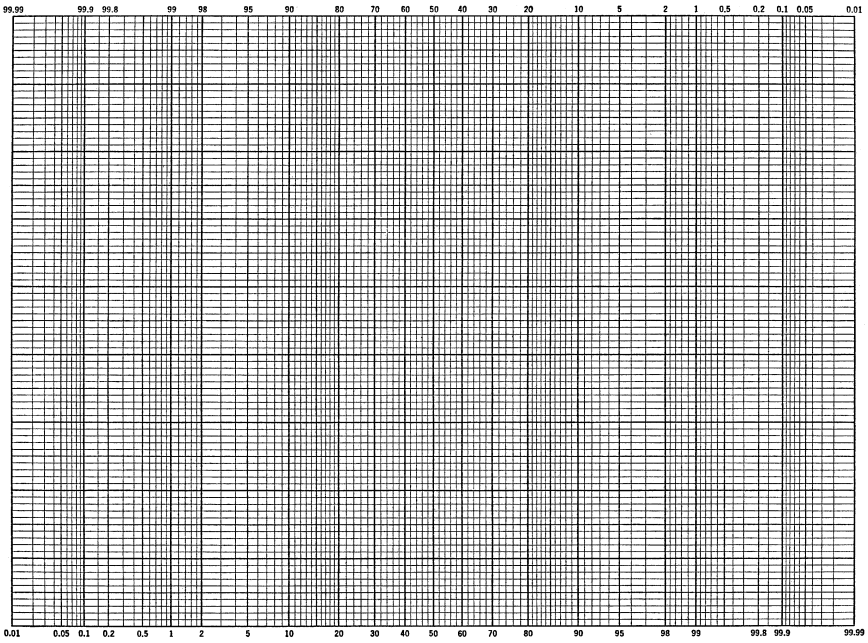
arithmetic mean– Same as *mean.*

arithmetic paper– A graph paper having uniform subdivisions for both ***x*** and ***y*** **axes.**



Arithmetic paper

arithmetic probability paper– A graph paper that has uniform subdivisions for the ***x*** **axis** but the ***y*** **axis** is ruled in such a way that a plot of the cumulative **normal distribution** appears as a straight line.



Arithmetic probability paper

arithmetic progression or series– A series of numbers is said to form an arithmetic progression when the difference between any two adjacent numbers is the same. For example, the series 3, 5, 7, 9, 11, 13, . . . is an arithmetic progression. Population sizes over a period of time are in arithmetic progression if the size of the population changes by a **constant** amount each year.

array– A simple arrangement of the individual **observations** or values of a **data set** arranged in order of magnitude, from the smallest to the largest value. For example, for the data set {2, 7, 5, 9, 3, 4, 6}, an ordered array is {2, 3, 4, 5, 6, 7, 9}.

artificial intelligence– A term coined to designate a scientific discipline concerned with investigating the intelligent behavior, i.e., reasoning, thinking, learning, and **decision making,** of machines by means of **computer simulation.**

ascertainment bias– A type of **bias** that arises from a relationship between the **exposure** to a certain **risk factor** and the **probability** of detecting an event of interest. It commonly occurs in many epidemiological studies, particularly in **retrospective case-control studies**.

ASN– Acronym for *average sample number.*

association– The term is more or less synonymous with **correlation**. It is more commonly used to describe the relationship between a pair of **nominal** or **qualitative variables**. See also *measures of association.*

association analysis– Same as *correlation analysis.*

assumptions– The term is most commonly used to refer to certain specific conditions that should be satisfied for the application of certain statistical procedures in order to produce

valid statistical results. For example, the usual assumptions for the application of an **analysis of variance** procedure are **normality** of distribution, **homogeneity of variance,** and **independence** of **observations.**

asymmetrical distribution– A **frequency (probability) distribution** that is not symmetrical. A **univariate distribution** is said to be asymmetrical if a vertical centerline divides it into two parts, which are different in **shape** and area. Some examples of an asymmetrical distribution are **exponential distribution** and **lognormal distribution.** Compare *symmetrical distribution.* See also *skewed distribution.*

asymmetrical population– A **population** that is not symmetrical.

asymmetric measure of association– A **measure of association** that is based on conceptual and computational distinctions between the **independent** and **dependent variables.** Compare *symmetric measure of association.* See also *Somer's D.*

asymmetry– The property of the **shape** of a **frequency distribution** that exhibits **skewness.** Compare *symmetry.*

asymptotic– Said of a line on a graph that continually approaches but never reaches the ***x* axis.** For example, the tails of a **normal curve** are asymptotic to the x axis. Moreover, lines or curves may be asymptotes to things other than just the x axis. In general, any line such that, for any given curve, the shortest distance from a point on the curve to the line approaches zero as the point moves to infinity from the origin. The term is also commonly used as a prefix to denote a large **sample** or a limiting property (as $n \to \infty$) in expressions such as **asymptotic test** and **asymptotic variance.**

asymptotically efficient estimator– An **estimator** of a **parameter** with a **variance** achieving the **Cramér–Rao lower bound** as the **sample size** approaches infinity.

asymptotically unbiased estimator– An **estimator** of a **parameter** that is biased but tends to become unbiased as the **sample size** increases and becomes infinitely large. For example, the usual **sample variance** (with divisor n) is a **biased estimator** of the **population variance** σ^2, but is asymptotically unbiased.

asymptotic distribution– The limiting form of the **probability distribution** of a **random variable** as the **sample size** approaches infinity.

asymptotic efficiency– The **efficiency** of an **estimator** in the limit as the **sample size** approaches infinity.

asymptotic method– Same as *large sample method.*

asymptotic normality– The exact **distribution** of a **statistic** is usually complicated and difficult to work with. The distribution is said to possess asymptotic normality if its limiting form approaches a **normal distribution.** The **central limit theorem** can often be used to approximate the distribution of a statistic by a normal distribution.

asymptotic relative efficiency– The **relative efficiency** of two **estimators** of a **parameter** as the **sample size** approaches infinity. The term is also used as an asymptotic measure of relative test efficiency as the sample size (n) increases against alternatives that approach the **null hypothesis** as n increases.

asymptotic technique– See *large sample method.*

asymptotic test– See *large sample method.*

asymptotic variance– The **variance** of a **statistic** as the **sample size** becomes infinitely large.

at random– In a **random** fashion.

attenuation– A term applied to denote the **correlation** between two **variables,** when both variables are subject to **measurement error,** to indicate that the value of the correlation between the true values is likely to be underestimated if both variables were measured with perfect **reliability.**

attributable fraction– See *attributable risk.*

attributable risk– The term is often used as a synonym to **absolute risk difference.** The attributable risk is often expressed as the fraction or **proportion** of the **risk** in the **intervention** or **exposed group,** and then it is known as proportional attributable risk (PAR) or attributable fraction. The PAR is defined as $\text{PAR} = (R_i - R_c)/R_i$, where R_i and R_c denote the risks in the **intervention** and **control group,** respectively.

attribute– The qualitatively distinct characteristics such as healthy or diseased, positive or negative. The term is often applied to designate characteristics that are not easily expressed in numerical terms.

attribute sampling– A **sampling procedure** in which the characteristic being measured is simply a quality or **attribute** of the items or individuals included in the **sample.** For example, an item may be classified as defective or nondefective. Compare *variable sampling.*

attrition– A term used to describe the loss of **study subjects** that may occur in a **clinical trial** or any **longitudinal study.**

autocorrelation– In a **time-series analysis,** it is the internal **correlation** between **observations** often expressed as a function of the lag time between them. For example, given the observed values $x_1, x_2, \ldots, x_n$ of a series, the sample autocorrelation of lag ℓ is defined as

$$\frac{\sum_{i=1}^{n-\ell}(X_i - \overline{X})(X_{i+\ell} - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

More generally, autocorrelation can occur when **residual error terms** from observations of the same **variable** at different time intervals are correlated. In **regression analysis,** autocorrelations can be reduced by using generalized rather than **ordinary least squares.** When the **variance** term from the denominator is omitted it is called autocovariance.

autocovariance– See *autocorrelation.*

automation– A term often used to refer to the use of advanced machinery and other modern equipment, especially in combination with high-speed computers.

autoregression– A term used to indicate the possibility that the **error term** in a **regression model** may be correlated with one or more lagged **endogenous variables.**

autoregressive integrated moving average– Same as *Box–Jenkins method.*

average– A medial numerical figure describing the typical or characteristic value of a group of numbers. It is a general term used for all types of averages, variously described as **measures of location** or **measures of central tendency.** When used unqualified, the term can be taken to refer to the **arithmetic mean.** Otherwise, it is a figure describing any **statistical measure** of the center of a **data set,** including arithmetic mean, **median,** or **mode,** among others.

average absolute deviation– A measure of **variability** or **dispersion** obtained by averaging **the absolute values** of the **deviations** about the **mean, median,** or **mode** for a particular set of **data.** Average absolute deviation is called mean deviation from the mean, the median, or the mode, according to the point about which the deviations have been measured. It is also called mean deviation, mean variation, average error, average departure, average variation, mean absolute error, and sometimes mean error.

average departure– Same as *average absolute deviation.*

average deviation– Same as *average absolute deviation.*

average error– Same as *average absolute deviation.*

average outgoing quality (AOQ)– The expected quality of the outgoing product following the use of an **acceptance sampling** plan for a given value of incoming product quality. It is calculated as the ratio of defective items to total items, i.e., the total number of defectives in the lots accepted divided by the total number of items in these lots. The AOQ serves as an index of performance measure associated with an acceptance sampling when the sampling plan is used repeatedly.

average rank– Suppose that x is one of a set of n **observations** that has the same value as (is tied with) some of the other observations. The **average** rank of x, in the **ranking** of the n observations, is the **mean** of those **ranks** that would be assigned to x and the other observations having the same value as x, if these tied observations could be distinguished.

average run length– In **statistical process control,** the length of time the process, on the average, must run before a **control chart** is capable of detecting a shift in the process level. It is usually measured in terms of the number of consecutive points plotted on the control chart.

average sample number– In a **sequential sampling** procedure, the **average** or **expected value** of the **sample size** required to reach a decision to accept or reject the **null hypothesis** and thereby to terminate **sampling.**

average variation– Same as *average absolute deviation.*