L

lambda– An **asymmetric measure of association** between the two **variables** forming a **contingency table.** The measure is designed for the situation in which one variable is considered explanatory and the other the response. It is more fully known as Goodman–Kruskal lambda. It ranges in value from zero to one.

Laplace criterion– Same as *equal-likelihood criterion.*

Laplace distribution– Same as *double exponential distribution.*

large sample method– A statistical procedure that makes the assumption of a large sample for its validity; that is, its **sampling distribution** is derived under the assumption of large sample theory. The procedure is based on an approximation to a **normal** or other **probability distribution** whose accuracy increases as the **sample size** increases.

large-scale trial– A **multicenter clinical trial** that enrolls a larger number of patients than the typical trial. The term is more or less synonymous with multicenter clinical trial.

Laspeyres' index number– A weighted aggregative price index named after a German economist named Etienne Laspeyres which is based on a combination of several items, with base period quantities employed as weights. If p_0^i, q_0^i $(i = 1, 2, \ldots, n)$ denote the prices and quantities sold of a set of n commodities in a **base period** and p_1^i $(i = 1, 2, \ldots, n)$ denote the corresponding prices in a given period, then the Laspeyres' index is defined as

$$L_{01} = \frac{\sum_{i=1}^{n} p_1^i q_0^i}{\sum_{i=1}^{n} p_0^i q_0^i}$$

The formula assigns to each current price a quantity weight that is appropriate for the base year. The quantity weight for each commodity is held constant for a number of years' computations. Laspeyres' index number is the most widely used throughout the world for making **price index numbers.** It is based on the basket of goods principle; that is, if a basket of goods costs \$20 in the based period and if the same basket costs \$25 in the given period, then the price index in the given period compared to the base is $25/20 = 1.25$. Price indices derived by this method usually have an upward **bias** because they allow for shifts in quantity in response to price increases.

latent factor– Same as *latent variable.*

latent variable– A **variable** representing a theoretical construct that cannot be measured directly. A latent variable is also called a true or unobserved variable. Many of the variables used in social and behavioral sciences are latent variables, for example, ambition, anxiety, aspiration, attitude, motivation, intelligence, and so forth.

latent variable modeling– See *structural equation model.*

Latin square– An **experimental design** involving the allocation of p **treatments** in a $p \times p$ square array such that each treatment occurs exactly once in each row or column. A Latin square is used to control for two sources of **variation** that may be identified with rows and columns. The design is also useful for investigating simultaneous **effects** of three **factors**: rows, columns, and Latin letters in a single **experiment.** The following is an example of a 5 × 5 Latin square. See also *Graeco–Latin square, hyper-Graeco–Latin square, hyper square.*

Layout of a 5 × 5 Latin square design

A	B	C	D	E
B	A	E	C	D
C	D	A	E	B
D	E	B	A	C
E	C	D	B	A

lattice design– A type of **incomplete block design** used in agricultural experimentation in order to increase the precision of treatment comparisons. It is also sometimes called a quasi-factorial design because of its analogy to **confounding** in a **factorial experiment.**

law of error– An **empirical rule** that states that **frequencies** with which **errors** of **measurement** and differences between actual values and **estimates** occur tend to form a **symmetrical distribution** approaching a **normal curve.**

law of large numbers– The law that states that the **probability** of a **deviation** of an **empirical probability** value from a theoretical one tends to zero as the number of repetitions of the **random experiment** in question increases to infinity.

LC50– Acronym for *lethal concentration 50.*

LD50– Acronym for *lethal dose 50.*

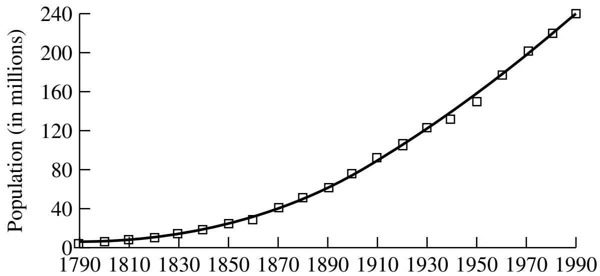
least absolute deviation estimation– In **regression analysis,** a method of fitting a **regression line** to **data values** so that the sum of the **absolute values** of the vertical **deviations** between the line and the individual data points is minimized. The method is more robust to usual violations of **assumptions** than the ordinary **least squares estimation.** See also *weighted least squares estimation.*

least significant difference test– In **analysis of variance,** a procedure for comparing a set of **means** that controls the overall **error rate** at some predetermined value, say α. The procedure consists of making an overall ***F* test** of the **hypothesis** of the equality of means at the α **level of significance.** If this test is significant, then the **pairwise comparisons** among the **treatments** are performed by using an α-level **two-sample *t* test**; otherwise, the procedure is terminated without making any further **inferences** on pairwise differences. See also *multiple comparison.*

least squares– Same as *least squares estimation.*

least squares estimation– In **regression analysis,** a method of fitting a **regression line** to **data values** in a **scatter diagram** in such a way that the sum of the squares of the vertical **deviations** between the line and the individual data plots is minimized. The method of least squares is a very general method of curve fitting that selects as the best-fitting curve the one that minimizes the sum of squares of the data points from the fitted curve. The least squares method is used extensively in many economic applications, for example, in estimating **secular trend** and for calculating the relationship between two or more **variables** for comparison purposes. It is also referred to as ordinary least squares to distinguish it from the method of **weighted least squares.** See also *least absolute deviation estimation.*

least squares estimate/estimator– An **estimate/estimator** of a **parameter** using the **method of least squares.** A least squares estimator has a smaller **variance** than any other **linear estimator** and is unbiased. See also **Gauss–Markov theorem.**



Least squares curve fitting of the population of the United States, 1790–1990, showing a quadratic trend

least squares method– Same as *least squares estimation.*

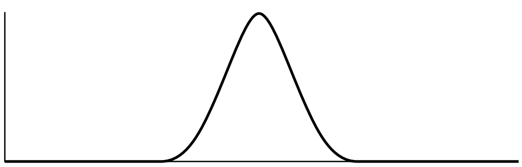
least squares regression– See *least squares estimation.*

least squares theory– See *least squares estimation.*

left-skewed distribution– Same as *negatively skewed distribution.*

left-tailed test– Same as *lower-tailed test.*

leptokurtic– A **distribution** is said to be leptokurtic when data points tend to accumulate more around the **mean** and in the tails than they do in a **normal curve.** Thus, a leptokurtic distribution is more sharply peaked and has larger tail areas than the **normal distribution.** Compare *mesokurtic, platykurtic.*



A leptokurtic distribution ($\beta_2 > 3$)

leptokurtic curve– See *leptokurtic.*

leptokurtic distribution– See *leptokurtic.*

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

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

level– In an **experiment** or **study,** a general term referring to the characteristic or amount that defines or designates a particular level, category or classification of a **factor** or **variable.**

level of measurement– Same as *scale of measurement.*

level of significance– Same as *significance level.*

leverage point– In **regression diagnostics,** a leverage point is used to refer to an **observation** that has an extreme value on one or more **explanatory variables,** and therefore a potentially large **effect** on the **regression equation.** See also *Cook's distance, influence statistics, influential observation.*

liberal test– An approximate **statistical test** with the **level of significance** greater than or equal to the nominal value. If it is known that the actual level of significance of a liberal test is not much greater than α (the nominal value), the liberal test can be recommended. See also *approximate test, conservative test, exact test.*

life expectancy– The expected life at a given age, that is, the **average** length of subsequent life remained to be lived. In other words, the number of years a person of a particular age group can hope to live.

life table– A table showing **life expectancy** at various periods of time and/or for different age/sex groups. It shows the number of persons who, out of a given number of persons born and living during a given age group, live to reach successive higher age groups, as well as the number of persons who die in those groups. The life table provides useful indices of **mortality** experience which are unaffected by the age structure of the population concerned. The important elements of a life tables are:

1. ${}_nq_x$: The **probability** of dying between any two ages x and $x + n$. This is obtained by the **ratio** of total deaths between two ages to the number alive at the beginning of the first age.
2. ${}_np_x$: The probability of surviving between any two ages x and $x + n$. This is obtained by the ratio of those who are alive between two ages to the number alive at the beginning of the first age. Note that ${}_np_x + {}_nq_x = 1$.
3. ℓ_x: The number alive at age x out of those starting at age 0.
4. ${}_nd_x$: The number of deaths between ages x and $x + n$.
5. ${}_nL_x$: The number alive in the age interval x to $x + n$.
6. T_x: The number alive in this and the subsequent age interval.
7. e_x^0: The expectation of life at age x, that is, the average length of subsequent life lived by those who have reached age x.

The table on the next page gives an abridged life table of the United States for the year 1980.

life table analysis– A technique for analyzing **survival data** containing **censored observations** that have been grouped into intervals. The technique can be applied to the study of not only death, but also any endpoint of interest such as the onset or remission of a

Abridged life table of the United States, 1980

Age interval	Proportion dying	Of 100,000 born alive		Stationary population		Average remaining lifetime
Period of life between two exact ages stated (in years) (1) x to $x + n$	Proportion of persons alive at beginning of age interval (dying during interval) (2) ${}_nq_x$	Number living at beginning of age interval (3) ℓ_x	Number dying during age interval (4) ${}_nd_x$	The number alive in the age interval (5) ${}_nL_x$	The number alive in this and subsequent age interval (6) T_x	Average number of years of life remaining at beginning of age interval (7) e_x
All races						
0–1	0.0127	100,000	1,266	98,901	7,371,986	73.7
1–5	0.0025	98,734	250	394,355	7,273,085	73.7
5–10	0.0015	98,484	150	492,017	6,878,730	69.8
10–15	0.0015	98,334	152	491,349	6,386,713	64.9
15–20	0.0049	98,182	482	489,817	5,895,364	60.0
20–25	0.0066	97,700	648	486,901	5,405,547	55.3
25–30	0.0066	97,052	638	483,665	4,918,646	50.7
30–35	0.0070	96,414	672	480,463	4,434,981	46.0
35–40	0.0091	95,742	875	476,663	3,954,518	41.3
40–45	0.0139	94,867	1,321	471,250	3,477,855	36.7
45–50	0.0222	93,546	2,079	462,857	3,006,605	32.1
50–55	0.0351	91,467	3,209	449,811	2,543,748	27.8
55–60	0.0530	88,258	4,676	430,230	2,093,937	23.7
60–65	0.0794	83,582	6,638	402,081	1,663,707	19.9
65–70	0.1165	76,944	8,965	363,181	1,261,626	16.4
70–75	0.1694	67,979	11,517	312,015	898,445	13.2
75–80	0.2427	56,462	13,702	248,534	586,430	10.4
80–85	0.3554	42,760	15,197	175,192	337,896	7.9
85 and over	1.0000	27,563	27,563	162,704	162,704	5.9

disease. For example, the technique is often applied in **cohort studies** to examine the **distribution** of **mortality** and/or **morbidity** due to one or more diseases over a fixed period of time.

likelihood function– A mathematical function that gives the **probability** of obtaining observed **data,** given the values of **parameters** of a **probability distribution.** In other words, a likelihood function measures the probability of observing a given set of data, given that certain values are assigned to the parameters. Thus the likelihood function combines data with a given **probability model** and parameters of interest.

likelihood ratio statistic– The **statistic** obtained as the **ratio** of the **likelihood function** calculated under the **null** and the **alternative hypotheses.** In large samples, a function of the likelihood ratio, i.e., $-2 \log_e(L_{H0}/L_{H_1})$ has approximately a **chi-square distribution** with the **degrees of freedom** equal to the difference in the number of **parameters** in the two hypotheses. See also *chi-square statistic, G^2 statistic, goodness-of-fit statistic, likelihood function, likelihood ratio test.*

likelihood ratio test– A **statistical test** based on the **likelihood ratio statistic.** The test was originally proposed by J. Neyman and E. S. Pearson in 1928.

Likert scale– A widely used scale to measure attitudes and opinions originally developed by Rensis Likert. In developing a Likert scale, **raw scores** are obtained as graded alternative responses to a questionnaire. For example, the respondents are given a series of statements relevant to the construction of scale and are asked to indicate their degree of agreement by stating "strongly agree," "agree," "disagree," "strongly disagree." A number is attached to each possible responses, e.g., 1 for "strongly agree," 2 for "agree," etc. The final scale is constructed as the composite score obtained as the sum of these numbers. Likert scales and Likert-like scales are easy to construct and are widely used in studies of opinions and attitudes in many areas of social and behavioral sciences. A widely used Likert-type scale in medicine is the Apgar scale employed to measure the health status of newly born babies.

linear association– Same as *linear relationship.*

linear combination– A linear combination of a set of k **variables,** $x_1, x_2, \ldots, x_k$, is an expression of the form $\ell_1 x_1 + \ell_2 x_2 + \cdots + \ell_k x_k$ where $\ell_1, \ell_2, \ldots, \ell_k$ are **constants.** An example of a linear combination is a **weighted average** of a set of variables or measures. The **prediction equation** in a **multiple regression analysis** can be considered as a linear combination of the **predictor variables.**

linear contrast– Same as *contrast.*

linear correlation– Same as *linear relationship.*

linear estimator– A **sample statistic** that is a **linear function** of **observations. Sample mean** is an example of a linear estimator.

linear function– Same as *linear combination.*

linear logistic regression– Same as *logistic regression.*

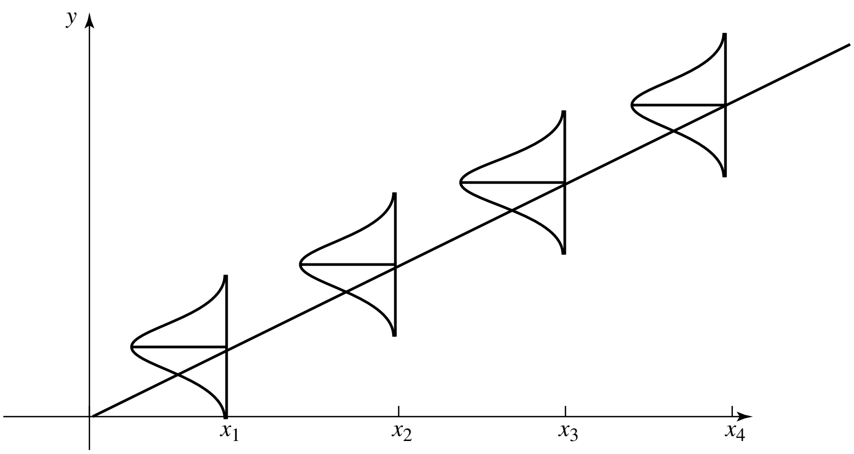
linear model– A **model** in which the equations relating the **random variables** and **parameters** are linear. More precisely, a relationship of the form

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + e_i \qquad i = 1, 2, \ldots, n$$

where Y is a random variable; $X_1, X_2, \ldots, X_p$ are fixed variables; $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are parameters to be estimated; and the **errors** e_i are usually independent normally distributed random variables with **mean** zero and **variance** σ^2. Note that linearity applies to the parameters and not to the variables. Thus, $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^2 + e_i$ is a linear model, but $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2^2 X_{2i}$ is not a linear model. See also *generalized linear model, linear regression, nonlinear model.*

linear programming– A mathematical technique of optimizing (i.e., maximizing or minimizing) a linear objective function subject to constraints in the form of linear inequalities. It is designed to select from a number of alternative courses of actions the one that is most likely to yield a desired result. The technique provides a decision-making tool for business management and has been employed on a variety of problems ranging from the selection of the ingredients appropriate to producing the most economical cattle feed of a given nutritional value to the determination of the safest site for a nuclear plant.

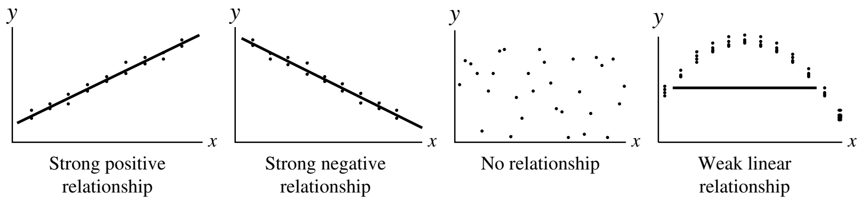
linear regression– The method of determining a **regression** or **prediction equation** to predict the value of a **dependent variable** from the given value of an **independent variable** by calculating a "best-fitting" straight line on a graph. A linear regression is represented by the model $Y_i = \beta_0 + \beta_1 X_i + e_i$, where Y is a continuous dependent or **response variable,** X is a continuous independent or **explanatory variable,** and e is the **random** or **residual term.** Compare *nonlinear regression.* See also *least squares estimation, linear model, multiple regression.*



Graphical illustration of a linear regression

linear regression analysis– Same as *linear regression.*

linear relationship– When correlated **data** exhibit only one kind of relationship, either direct or inverse, but not both, the two **variables** involved are said to have a linear or straight-line relationship. When plotted on a graph paper, a linear relationship forms a straight-line.



Graphical illustration of a linear relationship between *X* and *Y*

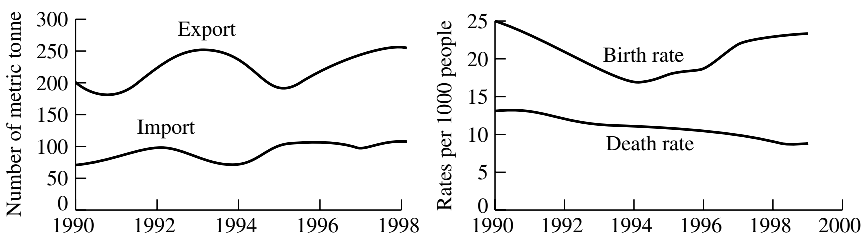
linear transformation– A mathematical **transformation** involving a **linear function** of a set of **variables.** The transformation consists of adding, subtracting, multiplying, or dividing the variables by a **constant.**

linear trend– A relationship between two **variables** such that a unit change in one variable produces a unit change in the other variable. The **trend** is expressed as the **linear function** of the time variable.

line chart– Same as *line graph.*

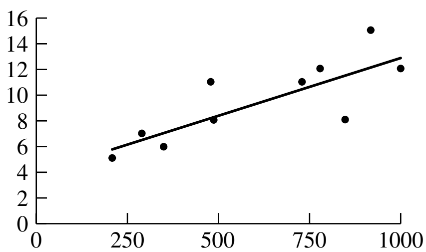
line diagram– Same as *line graph.*

line graph– A graph constructed by locating the points representing the observed values of the two variable magnitudes, and then connecting these points by either straight lines or smooth lines. It is also called line chart and line diagram.



Line graphs for hypothetical data

line of best fit– The line that best fits or averages the **data points** in a **scatter diagram** of a set of **bivariate data.** A plot of a **regression equation** obtained by using the **least squares method** is an example of a line of best fit. The line of best fit may also be drawn freehand by personal judgment as shown in the figure below.



Simple regression line drawn freehand in a scatter diagram

link relative– Same as *trend ratio.*

LISREL– Acronym for LInear Structural RELation, a name given to a **computer program** for fitting **structural equation models** involving **latent variables.** It is a highly versatile program, originally developed by K. Joreskog, to analyze **covariance** structures by the method of **maximum likelihood estimation.** It has gone into numerous versions. It also allows the researcher to perform **exploratory** and **confirmatory factor analyses** as well as **path analyses.** The program has been so popular that it has become synonymous with the methods of analysis as well as the **software** for analyzing the data.

LISREL model– Same as *LISREL.*

local odds ratio– The **odds ratio** computed from a **2 × 2 contingency table** obtained by taking two adjacent rows and columns of an $r \times c$ **contingency table.** It can be shown that $(r - 1)(c - 1)$ local odds ratios determine all $\binom{r}{2}\binom{c}{2}$ odds ratios that can be formed from pairs of rows and pairs of columns. The local odds ratios treat row and column **variables** alike, and their values describe the relative magnitudes of local **associations** in the table. The **independence** of the two variables is equivalent to the condition that the local odds ratios are identically equal to one.

location– See *central tendency.*

location parameter– A **parameter** which describes the central or middle point, or the most typical value of a **distribution,** such as **mean, median,** or **mode.** A location parameter has the property that if a **constant** is added to each value of a **random variable** having the given distribution, then the same constant must be added to the parameter.

logarithmic chart– A graph in which one or more axes are expressed in terms of logarithmic scales. Where only the vertical scale is so designed, the graph is known as a semilogarithmic chart. Where both axes are scaled in terms of logarithms, the graph is known as a double-logarithmic chart. In both cases, the natural numbers are plotted on the logarithmic grids. To construct a logarithmic grid, all one needs to do is measure the required range of logarithms on a normal scale, insert the logarithms of whole numbers at appropriate fractions, at intervals, and the corresponding natural numbers, then erase the logarithms. A geometric series plotted on a semilogarithmic chart would appear a straight line, whereas on a rectilinear graph, it would represent a curve. A double logarithmic chart is used for graphing the series of two **variables** when there is a logarithmic relationship between the two.

logarithmic transformation– A **transformation** of a **variable** to a new variable obtained by using a mathematical operation on a logarithmic scale. A logarithmic transformation is frequently applied in a number of situations in order to achieve **normality** and/or **homogeneity of variances** and to reduce a **nonlinear model** to a **linear model.** For example, large to moderately skewed **data** are sometimes subjected to logarithmic transformation to achieve normality, and methods of **estimation** and **hypothesis testing** are applied to log values, and the results are back-transformed to the original scale. Similarly, logarithmic transformations are employed in **regression analysis** to reduce a **curvilinear relationship** to a **linear relationship.** See also *arc-sine transformation, power transformation, reciprocal transformation, square-root transformation, square transformation.*

logistic model– Same as *logistic regression model.*

logistic regression– A kind of **regression** technique used when the **dependent variable** is a **binary** or **dichotomous measure.** If X is an **independent variable** and Y is a **binary**

response variable with **probability** of success equal to p, then the logistic regression model is given by

$$p = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} = \frac{1}{1+e^{-(\alpha+\beta x)}}$$

where e is the (natural) exponential function. The functional form given above is the logistic function, and hence the term logistic model. This model has the desirable range for p, i.e., between 0 and 1, and has many other useful statistical properties. See also *multiple logistic* re*gression.*

logistic regression model– See *logistic regression.*

logit method– A method for constructing **confidence interval** of the **odds ratio** in a **2 × 2 contingency table.** The upper and lower limits of the confidence interval are given by the formula

$$\log_e\left(\frac{ad}{bc}\right) \pm \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

where a, b, c, and d are four **cell counts.** It is also known as the Taylor series method.

log–likelihood function– The **transformation** of a **likelihood function** using natural logarithms. It is generally employed for mathematical simplicity in performing partial derivatives.

log–linear analysis– A statistical method for analyzing the relationships among three or more **nominal variables.** It may be used similar to a **regression analysis** to predict a dependent nominal outcome from nominal **independent variables.**

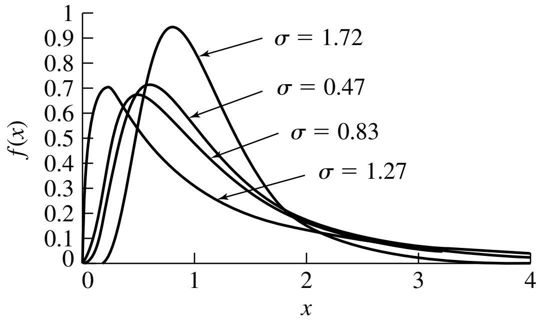
log–linear models– **Statistical models** for analyzing **count data.** These models are similar to **analysis of variance** models for **continuous data** except that the interest is now focused on **parameters** representing **interactions** rather than those for **main effects.** Log–linear models are so called because they use equations that are transformed to linear forms by taking their natural logarithms. The analysis of log–linear models is based on **odds** rather than **proportions** as is done in the chi-square analysis. The models can handle count data from several **categorical variables** and can be analyzed either by the **likelihood ratio test** or the usual **chi-square test** for **goodness of fit.**

log–log paper– Same as *double-logarithmic chart.*

lognormal distribution– If $\log_e(X)$ is normally distributed with **mean** μ and **variance** σ^2, then X is said to have a lognormal distribution. The **density function** of the lognormal distribution is given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log_e x - \mu)^2\right] \qquad x > 0, \sigma > 0$$

The lognormal distribution is especially useful in modeling **data** from a **positively skewed distribution.** For example, in clinical studies, triglycerides data may sometimes be approximated by a lognormal distribution. See also *logarithmic transformation.*



Probability density curves for lognormal distribution for various values of σ

log paper– See *logarithmic chart.*

logrank test– A **nonparametric method** for comparing two **survival curves** when there are **censored observations.** The principle of the logrank test is to divide the survival time scale into intervals according to the distinct observed survival times, ignoring censored survival times. It then uses the relative **death rate** in intervals to form a test for comparing the overall survival curves for different **treatment groups.** The **test statistic** essentially involves a comparison of the observed number of deaths occurring at each time period with the expected number of deaths if the two survival curves were the same. It is a special application of the **Mantel–Haenszel chi-square test,** where an overall comparison of the groups is performed by summarizing the significance of the differences in survival rates in each one of the time intervals which constitute the **follow-up period.** See also **stratified logrank test.**

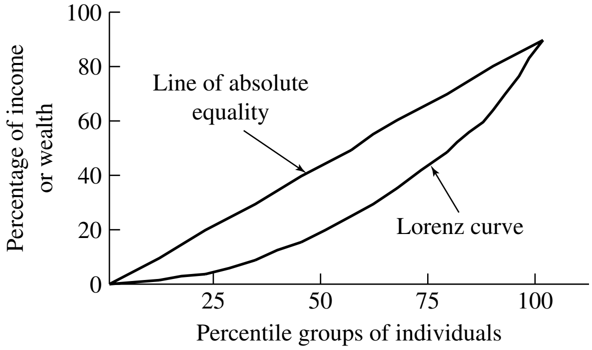
LOGXACT– See *STATXACT.*

longitudinal data– **Data** arising from a **longitudinal study.** A characteristic of this type of data is a **correlation** between pairs of **measurements** on the same subject, the magnitude of which usually depends on the time lag between the measurements. Typically the correlation becomes weaker as the time lag increases. This correlation needs to be properly accounted for if appropriate **inferences** are to be made. Special methods of analysis are often needed to take into account the correlation structure.

longitudinal study– A study involving a group of subjects that takes place over an extended period of time. A cohort of individuals is identified and followed through with **observations** made at several points in time. A longitudinal study can be carried out prospectively, and is known as **prospective study,** or retrospectively, and is then known as **retrospective study.** See also *cohort study.*

long-term forecast– A business **forecast** extending at least 5 years ahead of the current period, although such forecasts are often made for a period that may extend as far ahead as 15 or 20 years.

Lorenz curve– A curve used to display the nature of any **distribution,** particularly, the income distribution of a country. The curve is obtained by plotting the cumulative **proportion** of people against the cumulative share of total income that they receive. If there were a perfect equality in the distribution of income, with every one receiving the same amount of money, the Lorenz curve would be a 45° straight line. On the other hand, for the hypothetical situation of absolute inequality, with only one person receiving all the money, the curve would form the bottom right side of the square. In any practical situation, income distribution lies between these two hypothetical extremes and is thus represented by a sagging line. It is commonly used in many economic studies to display the extent of equality or inequality in the distribution of money income in an economy.



A diagram depicting the Lorenz curve

loss function– In **decision theory,** a mathematical function that assumes numerical values representing a gain or penalty for making correct or incorrect decision. Two popular loss functions are quadratic and absolute deviation.

loss to follow-up– In a **longitudinal study,** the term is applied to subjects who for a variety of reasons cannot be contacted to determine **outcome** measures or other characteristics of interest. Loss to follow-up often leads to censoring since the outcomes remain unknown. See also *censored observations.*

lower confidence limit– See *confidence limits.*

lower hinge– See *five-number summary.*

lower *p*th percentile– Same as *pth percentile.*

lower real limit– See *real limits.*

lower-tailed test– A **one-tailed hypothesis** test in which the entire **rejection region** is located in the lower tail of the **sampling distribution** of the **test statistic.** See also *one-tailed test, two-tailed test, upper-tailed test.*

LSD– Acronym for *least significant difference.*