G

gambler's fallacy– The belief that if a certain event has not occurred for a long period of time, it is sure to occur sometime very soon.

game theory– A mathematical theory involving analysis of decisions that deals with the theory of contests between two or more players involving **random** strategies in which each player wants to play the best way under the rules of the game. The game strategy usually involves a series of **events,** each of which may have a finite number of distinct results. For each event, it is known which player is to make the decision and how much that player knows about the results of the earlier events at the time of the decision. Game theory has applications in diverse fields such as systems analysis, war gaming, disease surveillance and control, and clinical decision analysis.

gamma– A **symmetric measure of association** for **observations** measured on an **ordinal scale.** The measure ranges from -1 to $+1$ and takes into account only the number of untied pairs. It is denoted by the Greek letter Γ. It is more fully known as Goodman–Kruskal gamma.

gamma distribution– A **probability distribution** with **parameters** α and β given by a **density function** of the form

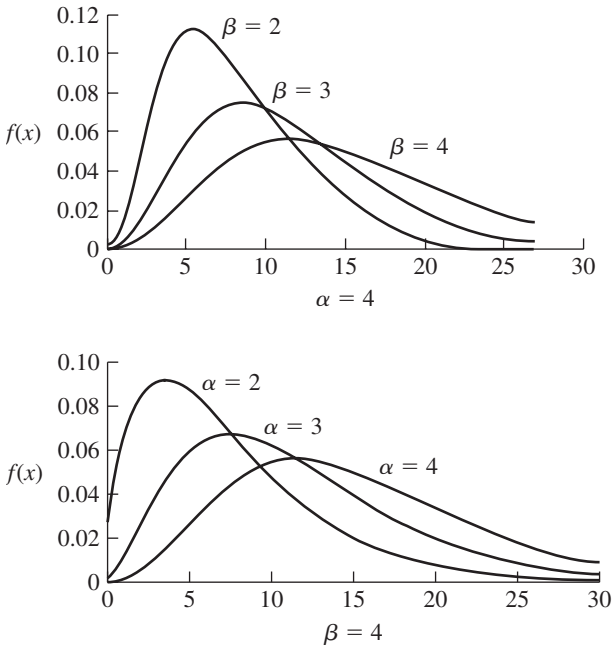
$$f(x;\alpha,\beta)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-\alpha/\beta} & x>0\\ 0 & \text{elsewhere}\end{cases}$$

The distribution has many important applications and includes the **chi-square distribution** and the **exponential distribution** as special cases.

gamma function– The gamma function (Γ) is defined by

$$\Gamma(p)=\int_0^{\infty}e^{-x}x^{p-1}\,dx \qquad p>0$$

A gamma function satisfies the recursive relationship $\Gamma(p+1)=p\Gamma(p)$. If p is any integer, it follows that $\Gamma(p+1)=p!$.



Probability density curves for gamma distribution for various values of α and β

GAUSS– A high-level programming language popular for writing programs in mathematical and scientific computations.

Gaussian distribution– Same as *normal distribution.*

Gaussian quadrature– An **algorithm** for performing numerical integration by approximating the function via a series expansion.

Gauss–Markov theorem– A theorem in mathematical statistics that states that the **least squares estimators** of the **parameters** in a **linear model** have uniformly smaller **variance** than any other **unbiased linear estimator.**

Geary's ratio– Same as *Geary's test.*

Geary's test– A test of **kurtosis** of a **distribution** based on the **statistic** G, defined as

$$G = \frac{\textbf{mean deviation}}{\textbf{standard deviation}}$$

In **samples** from a **normal population,** the value of G, when determined for the whole **population,** is 0.7979. Positive kurtosis yields higher values and negative kurtosis yields lower values of G.

Gehan's generalized Wilcoxon test– Same as *Gehan's test.*

Gehan's test– A **nonparametric statistical test** for comparing two **survival curves.** It is a version of the **Wilcoxon rank-sum test** applicable to **survival data** containing **censored observations.**

general fertility rate– Same as *fertility rate.*

generalized linear model– A class of **linear models** that allows the theory and methodology to be applicable to a much more general class of linear models, of which the normal theory is a special case. Such models allow the use of **sample data** that follow a **non-normal probability distribution** such as **Bernoulli** and **Poisson distributions. Estimates** of **parameters** in such models are generally determined by the method of **maximum likelihood estimation.**

generalized *p* value– A procedure for determining ***p* value** in the presence of **nuisance parameters.**

generalized Wilcoxon test– Same as *Gehan's test.*

general linear model– A class of **linear models** that includes both **regression** and **analysis of variance** models. Thus, a general linear model is used to study the effect of a continuous **dependent variable** on one or more **independent variables** whether continuous or categorical.

geographic correlation– The **correlation** between quantities determined as **averages** over a geographic region, such as state, country, or continent. These correlations generally give values that are very different from those that would be obtained from an analysis of unit level **data.** The phenomenon is often referred to as the ecological fallacy. It is also known as ecological correlation.

geometric distribution– The **probability distribution** of the number of **trials** required to obtain the first success in a series of **Bernoulli trials.** The **probability** of conducting n trials up to and including the first success is determined by the formula

$$P(n) = p\,(1-p)^{n-1} \qquad n = 1,\ 2, \ldots$$

where p is the probability of success at each trial.

geometric mean– The geometric mean, symbolized as GM or G, is the nth root of the product of n **observations.** Given $x_1, x_2, \ldots, x_n$, a set of n numbers, it is defined by the formula $\text{GM} = (x_1 \cdot x_2 \cdots x_n)^{1/n}$. It is generally used with characteristics measured on a logarithmic scale or with **skewed distributions.** It is calculated as the antilog of the **mean** from observations that have been transformed to logarithmic scale. The geometric mean lies between the **harmonic mean** and the **arithmetic mean.** It is not very useful as a **measure of location** and has a downward **bias** compared with the arithmetic mean. It is more suitable for averaging **ratios** and is therefore frequently used in the computation of **index numbers** that measure ratios of change in prices and other **data.**

geometric progression– A series of ordered numbers is said to form a geometric progression if the **ratio** of any two adjacent numbers is the same. For example, the series 2, 4, 8, 16, . . . is in geometric progression. Population size over a period of years is said to follow a geometrical pattern of growth if the change within a particular year is proportional to the population size at the beginning of that year.

gold standard– In medical diagnosis, the term is used to refer a **diagnostic procedure** that is highly accurate and reliable and generally gives correct diagnosis. Such procedures are generally expensive and are used in studies to assess the performance of a screening procedure. In **clinical trials** the term is applied to a randomized double-blind control clinical trial.

Goodman–Kruskal gamma– Same as *gamma.*

Goodman–Kruskal lambda– Same as *lambda.*

Goodman–Kruskal measures of association– Measures of association between two **qualitative variables** measured on **nominal scale.** Two such measures in common use are the so-called **gamma** and **lambda.**

goodness of fit– A term used to refer to the quality of a **model** or a theoretical **distribution** fitted to a given set of **data.**

goodness-of-fit statistic– An index or number that indicates how well a specified **model** or a theoretical **distribution** fits a given set of **data.** It is usually based on the comparison between the **observed** and **expected frequencies.** See also *chi-square statistic,* G^2 *statistic, likelihood ratio statistic.*

goodness-of-fit test– A **statistical procedure** performed to test whether to accept or reject a hypothesized **probability distribution** describing the characteristics of a **population.** It is designed to ascertain how well the **sample data** conform to expected theoretical values. It involves testing the fit between an observed **distribution** of **events** and a hypothetical distribution based on a theoretical principle, research findings, or other evidence by means of a **Pearson chi-square statistic** or any other **test statistic.** See also *chi-square test, goodness-of-fit statistic, Kolmogorov-Smirnov test.*

Graeco–Latin square– An **experimental design** involving the allocation of p **treatments** in $p \times p$ square array of Roman and Greek letters where each Roman and Greek letter appears once in each row and in each column, and each Roman letter appears once in combination with each Greek letter. A Graeco–Latin square is used to control three sources of **variation** which may be identified with rows, columns, and Greek letters. The design is also useful for investigating simultaneous **effects** of four **factors:** rows, columns, Latin letters and Greek letters in a single **experiment.** The following is an example of a 4×4 Graeco–Latin square. See also *hyper-Graeco–Latin square, hyper square, Latin square.*

$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
$B\gamma$	$A\delta$	$D\alpha$	$C\beta$
$C\delta$	$D\gamma$	$A\beta$	$B\alpha$
$D\beta$	$C\alpha$	$B\delta$	$A\gamma$

Layout of a 4×4 Graeco–Latin square

grand mean– The overall **mean** of all the **observations** in all the groups involved in an **analysis of variance** procedure.

graphical device– See *graphical methods.*

graphical display– See *graphical methods.*

graphical methods– A class of methods and techniques that make use of graphs and visual displays to represent the **data** or the results of an analysis. Some examples of graphical methods include **histograms, bivariate plots,** and **residual plots,** among others.

graphical presentation– See *graphical methods.*

graphical procedures– Same as *graphical methods.*

graphical representation– See *graphical methods.*

graphical techniques– Same as *graphical methods.*

graphing– A general term for plotting numbers and fitting a graph to the **scatter** of **data values.**

Greenwood's formula– In **survival analysis,** an algebraic formula for calculating the **variance** of the **Kaplan–Meier estimator.**

gross reproduction rate– Average number of female children that a synthetic cohort of women would have at the end of child-bearing years, assuming the absence of **mortality.** This **rate** gives a measure of replacement of **fertility** in the absence of mortality. See also *net reproduction rate.*

grouped data– Data values that have been sorted and grouped into **class intervals,** in order to reduce the number of scoring categories to a manageable level when the data range very widely. Data available in class intervals are then summarized by a **frequency distribution.** Individual values of the original data are not retained. Thus, with grouped data, one may not know the exact values of the **observations** falling within the class intervals. Compare *ungrouped data.* See also *grouped frequency distribution.*

grouped frequency distribution– A **frequency distribution** that lists **frequencies** for **class intervals** rather than individual **scores.** The **data** are grouped in intervals of equal range and each frequency represents the number of **data values** in one of the intervals. Compare *ungrouped frequency distribution.*

grouping– Same as *classification.*

group mean– The **mean** of all the **observations** in a particular group in an **analysis of variance** design.

group sequential trial– A **clinical trial** in which comparisons are made every time a group of patients has been enrolled in the study. These trials terminate early when **treatment** differences are large. See also *sequential sampling.*

growth curve analysis– The study of correlated **measurements** over time in individuals and groups. For example, in a study of height and weight of a group of children at a particular age, a graph of the height against the weight gives the individual's growth curve. The analysis of growth curve usually involves the problems of **repeated measures designs.**

G^2 **statistic–** A **statistic** based on the likelihood ratio used to test the **goodness of fit** of a specified model or a theoretical **distribution.** It is based on the comparison between the **observed** and **expected frequencies** and is calculated by the formula $G^2 = 2\sum_i O_i \log_e(O_i/E_i)$ where O_i and E_i are observed and expected frequencies in the *i*th class. See also *chi-square statistic, deviance, goodness-of-fit statistic, likelihood ratio statistic.*