



R

random— A term used to denote the quality of something that is unpredictable, nondeterministic, or occurring simply by chance.

random allocation— Same as *random assignment*.

random assignment— The use of a **random** device to assign different **treatments** to subjects or vice versa. The random assignment should not be confused with haphazard assignment. Random assignment increases **internal validity** of a study. See also *block randomization*, *randomization*.

random-digit dialing— A method of **sampling** households through the selection of telephone numbers by a **random** choice of digits in the telephone numbers. If the households being surveyed have high levels of telephone coverage, the technique can provide a **representative sample** of the households. Random-digit dialing provides the advantage of low cost of conducting a **survey** and it is now considered a useful procedure in many social and health sciences investigation. See also *telephone sampling*.

random effects— A term used to denote effects attributable to a large collection of **levels** of a **factor** or **treatment** (usually infinite) of which only a small **sample** are included in a given study. Random effects are frequently used in the context of **linear models** and **meta-analysis**. Compare *fixed effects*.

random effects analysis of variance— See *random effects model*.

random effects model— An **analysis of variance** or **regression model** in which the **treatment levels** associated with a **factor** are randomly selected and are considered to have **random effects**. Random effects are usually assumed to follow a **normal distribution**. This model is also referred to as model II. In the context of **meta-analysis**, the term is used to describe a model that assumes that the studies being summarized constitute a **random sample** from a larger population of similar studies. See also *fixed effects model*, *mixed effects model*.

random error— The **variation** in **measurements** that can be expected to occur entirely by chance. Random errors represent **deviations** of an observed value from a true value that

are due to chance rather than to one of the other factors being studied. Random errors on the **average** tend to cancel out in the sense of having a **mean** that tends to zero. See also *systematic error*.

random event– An **event** or phenomenon that is unpredictable and whose occurrence is governed purely by chance. A random event may or may not occur at a given **trial** or moment of time, but does possess some degree of statistical regularity, with a **probability** of occurrence determined by some **probability distribution**.

random experiment– Any activity or **trial** that will result in one and only one of several possible **outcomes**, but it cannot be predicted in advance which of these will occur in any particular trial.

random factors– **Factors** in an **analysis of variance** or **regression model** thought to have a **random effect**. Some examples of factors that are usually considered random are days, subjects, and plots. Compare *fixed factors*. See also *fixed effects*.

randomization– The process of assigning subjects or other **experimental units** to different **treatments** (or vice versa) by using **random numbers** or any other **random** device. The purpose of randomization is to produce comparable **treatment groups** in terms of important **prognostic factors**. The randomization ensures that, within the limits of **chance** variation, the **experimental** and **control groups** are similar at the beginning of the investigation. The randomization eliminates **bias** in the assignment of treatments and provides the sound basis for statistical analysis. The **random assignment**, however, frequently gives rise treatment groups with unequal **sample sizes**. This problem can be overcome by using **block randomization**. See also *cluster randomization*.

randomization test– A **nonparametric test** for **quantitative variables** in which certain aspects of a **sample** are studied by enumerating all possible arrangements of its elements. In a randomization test, the **test statistic** is derived directly from the **data** and does not require the use of a **sampling distribution**.

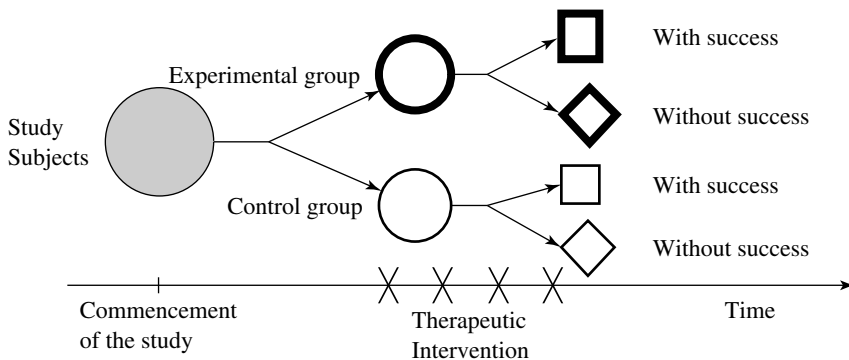
randomized block design– An **experimental design** employing **blocking** to control for individual differences among **experimental units**. This is a **two-factor analysis of variance** design in which each **block** consists of a set of fairly homogenous experimental units, and **treatments** are allocated to the various units within the blocks in a **random** manner. See also *block design*, *completely randomized design*, *randomized group design*.

		Block				
		1	2	3	...	<i>b</i>
	T_1	T_1	T_1	T_1	...	T_1
	T_2	T_2	T_2	T_2	...	T_2
	T_3	T_3	T_3	T_3	...	T_3
	⋮	⋮	⋮	⋮	...	⋮
	T_t	T_t	T_t	T_t	...	T_t

Layout of a randomized block design with b blocks and t treatments

randomized clinical trial– A **clinical trial** where the patients are randomly assigned to different **treatment groups**. See also *randomized controlled clinical trial*.

randomized controlled clinical trial– A **clinical trial** in which subjects are allocated at **random** to the **experimental** and a **concurrent control group**. After the completion of the trial, the results are assessed and compared in terms of the **outcome measure** of interest between the experimental and the control group. Randomized controlled clinical trials are considered as the most scientifically valid method of evaluating the efficacy of a **treatment**.



Schematic diagram of a randomized controlled clinical trial

randomized controlled trial– Same as *randomized controlled clinical trial*.

randomized group design– An **experimental design** that creates one **treatment group** for each **treatment** and assigns each **experimental unit** to one of these groups by a **random** device. See also *randomized block design*.

randomized response model– A technique used in **sample surveys** of human populations to eliminate **response bias** in answering personal and sensitive questions. The procedure introduces an element of chance as to what question a respondent has to answer.

randomized response technique– Same as *randomized response model*.

random model– The term is essentially equivalent to **stochastic model**, and sometimes it is used as a short form for **random effects model**.

randomness– A term used to describe an intuitive concept referring to a condition or property of a phenomenon governed purely by **chance**. See also *random*, *random event*, *random experiment*.

random normal deviates– **Random numbers** generated from a **standard normal distribution**.

random numbers– Random numbers are a collection of digits 0, 1, 2, . . . , 9 arranged as if they had been generated by a **random** device which gives each digit the same **probability** of occurrence. Random numbers are widely used in the selection of a **random sample**. See also *pseudorandom numbers*.

random-numbers table– A listing of numbers generated by a **random** process such that each possible digit is equally likely to precede or follow any other one. Published tables of **random numbers** generated by a computer **algorithm** are widely available to facilitate the selection of **random samples**. L. H. C. Tippett in 1927, M. G. Kendall and B. Smith in

1940, Rand Corporation in 1955, and C. E. Clark in 1966 published the best-known random number tables. Nowadays, computer algorithms for random number generators have largely superseded random-number tables. A short table of random numbers is given below.

Table of random numbers

Row	Column										Row
	1	2	3	4	5	6	7	8	9	10	
1	80083	77093	00960	49851	44218	64603	50045	73159	55805	50067	1
2	22763	43086	98315	90948	77066	47912	58164	50293	32803	55015	2
3	22125	31789	33826	64132	55537	11451	92836	79580	14996	51984	3
4	56241	99012	29886	92789	78115	72669	34419	06357	96818	16337	4
5	49378	85557	71172	30749	54432	92144	22681	49548	18077	30401	5
6	95083	38793	20028	98540	07752	78539	31495	94052	37987	38911	6
7	02803	26490	81174	27904	84943	57181	52137	68864	94549	77710	7
8	42546	61510	57266	84416	54355	74818	65673	98941	24333	45425	8
9	60198	00328	02233	48032	14609	63395	13759	21971	64000	20404	9
10	55536	89600	43238	11102	90620	31173	22357	15252	14569	98341	10
11	16485	41619	57814	18747	28312	93687	03021	20668	45974	63771	11
12	81634	47135	92210	31022	50800	26336	85622	74093	34899	71644	12
13	75281	85184	67672	49786	20730	43161	95372	28160	82440	02757	13
14	45316	21084	13743	48517	01075	42091	93025	92262	42328	51621	14
15	99985	81537	80566	69397	53509	02336	85126	49640	25196	21145	15
16	23050	34065	33474	94498	91298	03595	58587	96149	47680	30561	16
17	71804	028355	46763	86988	19204	27278	16287	85017	68168	61348	17
18	56461	27640	18455	50462	91258	55424	36463	49124	06467	13484	18
19	26409	04456	47172	16686	98951	77734	93342	50827	60020	02820	19
20	34579	53161	29401	14076	19037	83061	46912	16074	68014	71779	20
21	53326	52317	41398	61470	57492	44730	34602	40589	12409	00818	21
22	64454	15627	14444	26788	41024	31498	47423	43207	63501	21043	22
23	12615	35357	85483	83015	79536	5654	94742	38941	36832	70550	23
24	23172	22867	87620	41610	64224	71306	37504	97015	82065	40710	24
25	96310	86555	87851	03749	40471	20834	98170	87168	23027	67084	25
26	07223	92200	83095	54485	68338	48062	22870	11053	93573	83185	26
27	70736	36539	64310	23948	46399	45513	45821	93469	95533	91941	27
28	53140	75281	42302	26586	18095	97262	69518	23908	63082	15251	28
29	33520	99286	75440	29318	80495	92646	03921	60534	06946	75750	29
30	55041	29226	29602	80254	11099	05099	51359	28084	66690	72343	30

Source: Generated by using software.

random outcome— Same as *random event*.

random phenomenon— See *random event*.

random process— See *random event*.

random sample— A **sample** selected in such a manner that every member of the **population** has a fixed and known **probability** of being included in the sample. For a random sample without replacement selected from a **finite population**, every possible sample has equal probability of selection; for a random sample with replacement each item is selected independently of the other item with equal probability. See also *convenience sample, judgment sample, nonprobability sample, probability sample*.

random sampling— A **sampling** scheme wherein each individual or unit is selected entirely by **chance**. Random sampling is one of the best ways of obtaining a **representative sample**. See also *random sample*, *simple random sampling*.

random selection— A method of selecting a **sample** of individuals that uses a truly **random** device. In random selection each individual element in the **population** has an equal chance of being selected. It should not be confused with **haphazard selection**.

random variable— A numerical description of the **outcome** of a **random experiment**. The value of a random variable is determined by a random experiment and thus depends on **chance** and cannot be predicted with certainty. It is also called a **chance variable** or **stochastic variable**.

random variation— Same as *random error*.

random walk— A term used in **stochastic process** to describe the movement of a particle from one point to the other in discrete steps with certain known **probabilities**. Random walks have important applications in many real-life situations such as migration of insects, **sequential sampling**, and diffusion processes. See also *Markov process*.

range— A **measure of variability** or **dispersion** for a **data set** obtained by subtracting the smallest value in a data set from the largest value for **ungrouped data** or between the upper limit of the largest class and the lower limit of the smallest class for **grouped data**. Often used in **quality control** and other works as a quick way to calculate a measure of the dispersion, but is generally not recommended for this purpose because of its sensitivity to **outliers** and the fact that its value increases with **sample size**.

rank— A number indicating the relative position of any one **observation** with respect to the others in a **data set** when the observations are arranged according to their size, from the lowest to the highest. The lowest observation will receive a rank of 1, the second lowest a rank 2; and so forth.

rank correlation— A **nonparametric method** for assessing **association** between two **quantitative variables**. A rank correlation is interpreted the same way as the **Pearson product moment correlation coefficient**. However, a rank correlation measures the association between the ranks rather than the original values. Two of the most commonly used methods of rank correlation are **Kendall's tau** and **Spearman's rho**.

rank correlation coefficient— Same as *rank correlation*.

ranking— The process of assigning **ranks** to a given set of **observations**.

rank of a matrix— The number of rows or columns of a **matrix** that are linearly independent.

rank order— A set of **observations** arranged in order of their rank.

rank-order scale— A scale for **observations** arranged according to their size or magnitude, from the lowest to the highest value or vice versa, in which **ranks** are assigned according to relative position in the scale. The rank-order scale gives the relative position of an observation in a series of measurements. Compare *ordinal scale*.

rank-order statistic— A **statistic** based on the **ranks** of the **sample data**.

Rao–Blackwell–Lehman–Scheffé Theorem– A theorem in mathematical statistics that states that an **unbiased estimator** of a **parameter** based on a complete **sufficient statistic** is the unique minimum variance unbiased estimator of the parameter.

Rao–Cramér inequality– Same as *Cramér–Rao inequality*.

rate– A rate is a measure of the **frequency** of occurrence of a phenomenon. In **vital statistics**, a rate represents the frequency with which a vital event such as birth, death, or disease occurs in a defined population. Although there are some exceptions, the rate is usually calculated by an expression of the form $a/(a + b)$ in which the numerator is also a component of the denominator. It is usually multiplied by a power of 10 to convert the rate from a fraction or decimal to a whole number.

rate of natural increase– Relative change in population size brought about only by the balance between births and deaths; it is obtained as the difference between the crude birth and death rates.

rate of population growth– Relative change in population size brought about as a result of births, deaths, and net migration.

ratio– A ratio is the value obtained by dividing one quantity by another. It is used to show the magnitude of one quantity relative to the magnitude of another. It is calculated by an expression of the form a/b in which the numerator is not a component of the denominator. Thus, in a ratio, the numerator and the denominator usually are separate and distinct quantities. The dimensions of the numerator and denominator may be different so that the ratio has dimensions.

ratio data– **Data** obtained using **ratio scale of measurement**.

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

ratio scale– The process of assigning **measurements** with an **interval scale** that has a true zero point. The ratio scale has four properties: it sorts **observations** into classes, orders them in terms of differences in magnitude, specifies the amount of difference between the observations, and permits the expression of ratios between measurements. Ratio scale yields truly **quantitative data** that can be subjected to all types of mathematical operations. The examples are scales used for measuring height, weight, and cholesterol level. See also *scale of measurement*.

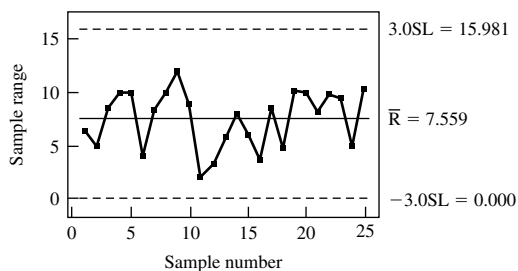
ratio variable– A **continuous variable** measured on a **ratio scale**.

raw data– Same as *raw score*.

raw score– A **score** or **measurement** as originally collected or observed, and has not been modified or transformed in any way.

RBD– Acronym for *randomized block design*.

R chart– A **graphical device** used to control the **variance** of a process by inspecting the **range** of a set of **measurements** taken from various batches or subgroups. The values of the range taken from each subgroup are plotted along the **vertical axis** and can then be used to control within subgroup **spread**. The center line of the *R* chart is the **average** of ranges (\bar{R}) from a pilot set (about 20 subgroups). The **control lines** are based on an **estimate** of within-group **standard deviation** obtained from **variance components** analysis. In practice, the engineer sets the limits at ($D_3 \bar{R}$, $D_4 \bar{R}$) where D_3 and D_4 are obtained from some specially prepared tables.

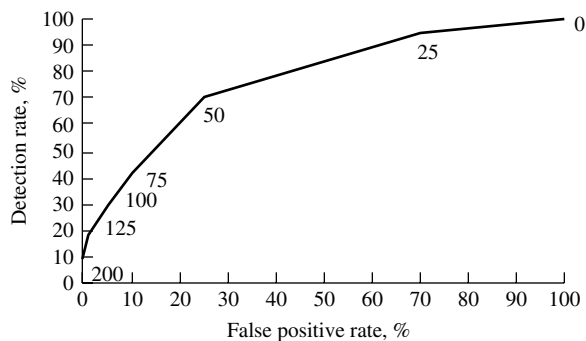
An example of an R chart

RCT– Acronym for *randomized control trial*.

real limits– The lower and upper limits based on the actual values observed before rounding. The real limits of a **class interval** are the boundaries above and below it that include all the values scored as that number. For example, the number 2 has a lower real limit of 1.5 and an upper real limit of 2.5; all values between these two boundaries are scored as 2.

recall bias– A type of **bias** that can occur in a **study design**, particularly in a **retrospective study**, because of different memories of past **exposures** between **cases** and **controls**. See also *information bias*.

receiver operating characteristic curve– In a **diagnostic testing** or **screening test**, a graph showing **sensitivity** or **true positives** on the **y axis** versus the **false positives** on the **x axis**. It is used to assess the property of a **diagnostic test** to discriminate between healthy and diseased individuals. It allows the comparison of performance of different cut points to be made.



A receiver operating characteristic curve

reciprocal transformation– A **transformation** of the form $y = 1/x$ that is particularly useful to stabilize the **variance** of a **data set** when the **standard deviation** of the data set is proportional to the square of the **mean**. If x represents counts, then $y = 1/(x + 1)$ may be used to avoid the possibility of division by zero. The transformation is generally used when $y = 1/x$ has a definite physical meaning and where the possibility of the **random variable** being less than or equal to zero is negligible. For example, **data** on the failures of a machine may be collected as either the interval between failures or the number of failures per unit time. In some cases the transformation can lead to a **linear relationship** between

a pair of **variables**. See also *arc-sine transformation*, *logarithmic transformation*, *power transformation*, *square-root transformation*, *square transformation*.

rectangular array— An **array** of p rows and n columns representing a collection of $p \times n$ data elements comprising n **measurements** on a set of p **variables**.

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pj} & \cdots & x_{pn} \end{bmatrix}$$

Schematic representation of a rectangular array

rectangular distribution— Same as *uniform distribution*.

recursive model— A **causal model** in which there is only one-way causal flow in the system. Thus, in a recursive model, reciprocal causation between **variables** is not permitted. Compare *nonrecursive model*. See also *path analysis*, *structural equation model*.

reference interval— Same as *normal range*.

reference population— The population being chosen as standard for computation of **standardized rates**. In a **sample survey**, the population designated for a particular **sampling design**.

reference range— Same as *normal range*.

region of acceptance— In **hypothesis testing**, the range of possible values of the area in the **sampling distribution** of a **test statistic** that does not lead to rejection of the **null hypothesis**. In other words, it is the region comprising the set of values of a test statistic for which the null hypothesis is accepted. Compare *critical region*.

region of rejection— Same as *critical region*.

regressand— Same as *predicted variable*.

regression— Same as *regression analysis*.

regression analysis— A statistical procedure used to develop a mathematical equation showing how two or more **variables** are related and/or to determine the extent to which one variable changes with changes in another variable or a number of other variables. The procedure allows the unknown value of one variable to be estimated from the known value of one or more other variables. There are a great variety of methods of regression analysis currently being used. See also *multiple regression analysis*, *simple regression analysis*.

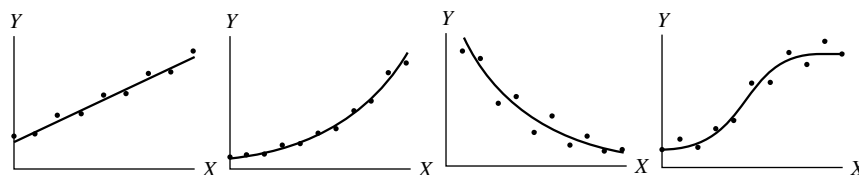
regression artifact— Same as *regression fallacy*.

regression coefficient— The **coefficient** β in the simple **regression equation** $E(Y) = \alpha + \beta X$. It is sometimes called the slope of the **regression line** and is interpreted as the **average** number of units change (increase or decrease) in the **dependent variable** occurring with a unit change in the **independent variable**. In a **multiple regression analysis**, the coefficients are weights applied to the independent variables and are interpreted

as measures of the **effect** of that variable while holding the effects of the other independent variables as **constants**. When the **predictor** is a **categorical variable**, the regression coefficient represents the average of difference between any given level of the variable and the value taken as the baseline or standard. See also *estimated regression coefficient*.

regression constants— The values that determine a **regression line** and locate it in a **cartesian space** are called the regression constants. They are the **slope of the regression line** and its **y** or **x intercept**, depending on whether the prediction is made on the basis of X or Y , respectively.

regression curve— A curve that represents the **regression equation** in a **cartesian space**. For a particular point on the curve, the **abscissa** is the value X and the **ordinate** is $\mu_{Y|X}$, the **mean** of the **distribution** of Y for that specified fixed value X . The word curve is used in contrast to straight line to mean a regression equation of a degree higher than the first. Some examples of regression curves are shown below.



Some examples of regression curves

regression diagnostics— A term used to denote statistical procedures designed to investigate the **assumptions** underlying a **regression analysis**. Regression diagnostics are used to check the assumptions for **normality**, **homoscedasticity**, and/or examine the influence of particular **observations** on the **estimates** of **regression coefficients**. See also *Cook's distance*, *DFBETA*, *DFFITS*, *influence statistics*, *residual analysis*.

regression effect— A term originally used to describe the tendency of certain members of any **population** who, with respect to a given characteristic, are in extreme position (below or above the **average** value) at one time to be in a less extreme position at a later time (either personally or by means of their offspring). Thus, an **observation** that is low or high at the time of first observation will tend to be closer to the **mean** at a later time period. The phenomenon was first noted by Sir Francis Galton who discovered that tall parents do not on the average have as tall offspring and short parents do not on the average have as short offspring.

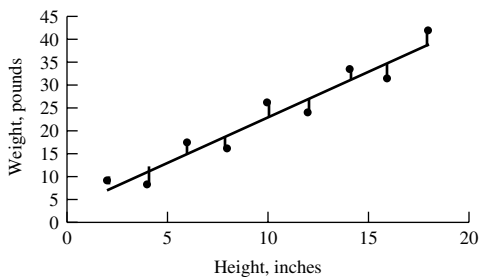
regression equation— An algebraic equation relating the **independent variable(s)** to the **expected value** of the **dependent variable**. A regression equation summarizes the relationship between a **response variable** and one or more **predictor variables**. For a single predictor variable, the regression equation representing a **linear relationship** is written as $E(Y) = \alpha + \beta X$, where Y is a response variable, α is the **intercept**, and β is the **regression coefficient**. It can be used to predict the values of the dependent variable from values of the independent variable(s).

regression fallacy— The incorrect ascription of the **regression effect** to the operation of some important unseen **factor**.

regression forecasting— The use of **regression analysis** for **forecasting a time series**.

regression hyperplane— A **graphical display** of a **regression equation** involving three or more **independent variables**. It is a higher-dimensional equivalent of a **regression line** or **plane**.

regression line– A graphical representation of a regression equation. It is the line drawn on a scatter diagram that best describes the relationship between the **dependent variable** and the **independent variable**. The regression line is usually fitted by using the **method of least squares**.



Regression line of weight on height

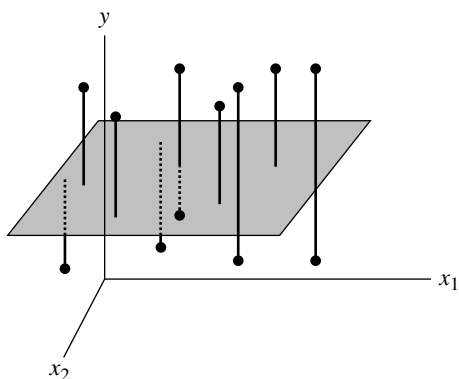
regression mean square– In linear regression analysis, a quantity obtained by dividing the regression sum of squares by its degrees of freedom.

regression method– See *regression analysis*.

regression model– See *regression equation*.

regression modeling– The term is essentially synonymous with **regression analysis**. It is also sometimes used to refer to a number of methods for selecting the “best” possible set of **predictors** when using regression analysis. The three most commonly used procedures for this purpose are **backward elimination**, **forward selection**, and **stepwise regression**. See also *all subsets regression*.

regression plane– The three-dimensional equivalent of a **regression line** that minimizes the sum of the squares of vertical **deviations** between the sample points lying in y versus (x_1, x_2) **cartesian space** and their associated **multiple regression equation** estimates, which all lie on the regression plane. See also *regression hyperplane*.



A best-fitted regression plane

regression sum of squares– See *explained variation*.

regression surface– Same as *regression hyperplane*.

regression through the origin– A term used to denote a **regression analysis** in which the **regression line** passes through the origin. It is used when the true **mean** of the **dependent variable** is known to be zero when the value of the **independent variable** is zero.

regression toward the mean– A term used to describe **regression** when the predicted values on the **dependent variable** show less **dispersion** about the **mean** than the observed values do. This occurs because measures of a **dependent variable** are unreliable and there is a less than perfect relationship between the two **variables**.

regression weight– Same as *regression coefficient*.

regressor– See *independent variable*.

regressor variable– Same as *predictor variable*.

regret– Same as *opportunity loss*.

regret table– In **decision theory**, a table showing the **opportunity-loss** values associated with each possible action/**event** or combination of actions and events.

rejection region– Same as *critical region*.

relative class frequency– The **class frequency** expressed as a **proportion** of the total **frequency**. It is calculated by dividing the class frequency by the total number of **observations**.

relative efficiency– Same as *efficiency*.

relative frequency– Same as *relative class frequency*. The term is also used as a synonym for **empirical probability**.

relative frequency curve– Any curve that represents a **relative frequency distribution**. See also *frequency curve*.

relative frequency distribution– A **frequency distribution** expressed in terms of the **relative frequency**, that is, the fraction or **proportion** of the total number of items in each of several nonoverlapping classes or categories.

relative frequency of an event– **Ratio** of the number of ways an **event** can occur to the total number of possible occurrences. See also *empirical probability*.

relative frequency probability– Same as *empirical probability*.

relative power efficiency– Same as *power efficiency*.

relative risk– The **ratio** of two **risks**. It is also called a **risk ratio**. It is designed to measure the degree of **association** in a **2 × 2 table**. If there is no difference between risks among the two groups, the relative risk will be equal to 1. If the exposed group has higher risk than the unexposed group, the risk ratio will be greater than 1. For example, a relative risk of 3 means that the exposed group is 3 times more likely to have the disease than the unexposed group. In **epidemiology**, the term is used mainly to denote the ratio of risk of disease or death among individuals exposed to a certain health hazard (for example, smokers) to the risk among unexposed (for example, nonsmokers).

relative-value index number– An **index number** constructed by (1) assigning the index number 100 to each item in a list of figures representing a period of time chosen as the base period; (2) finding for each item in each of the other periods under consideration an individual index number or a figure that bears the same relation to 100 that the item in

question bears to its corresponding items in the base period; (3) finally, calculating a **geometric mean** of the individual index numbers for each period.

reliability– The consistency or stability of a measure or test from one occasion to the next. Thus, it is a measure of the reproducibility of a **measurement**. It is measured by the **kappa statistic** for **nominal measures** and by the **correlation coefficient** for numerical measures. In engineering, the reliability of a product is the **probability** that it will perform within specified limits for a specified length of time. See also *Cronbach's alpha*.

REMLE– Acronym for *restricted maximum likelihood estimation*.

repeatability– Same as *reproducibility*.

repeated measurements– A term used to describe **observations** in which the **response variable** for each **experimental unit** is measured on several occasions and possibly under different experimental conditions. Repeated measurements occur frequently in **observational studies** that are longitudinal in nature, and in **experimental studies** involving **repeated measures design**. The repeated measurements are commonly used in a variety of disciplines including health and life sciences, education, psychology, and social sciences. See also *longitudinal study*, *repeated measures analysis*.

repeated measures analysis– Analysis of **repeated measures data** taken on one or more groups of subjects. The main problem with this type of analysis is the lack of **independence** of **observations** taken on a single subject. Repeated measures data are frequently analyzed incorrectly by ignoring the lack of independence of observations. Special statistical methods are often needed for the analysis of this type of **data** that take into account the intercorrelations between the set of **measurements** on the same subject. Analysis as a **split-plot design** is appropriate if the **residuals** from different time periods have equal **correlations**. If the correlation structure is more complex, the appropriate analysis is either a **multivariate analysis of variance**, or one that assures a defined **time-series model**. See also *repeated measures design*.

repeated measures data– Same as *repeated measurements*.

repeated measures design– An **experimental design** that measures the same subjects under two or more experimental conditions or on different occasions on the same **dependent variable**. For example, blood pressure may be measured at successive time periods, say once a week, for a group of patients attending a clinic; or animals are injected with different drugs and **measurements** are made after each injection. The **scores** for each subject are treated as correlated **observations**. In repeated measures design, each subject acts as its own **control**. This helps to control for **variability** between subjects, since the same subject is measured repeatedly. See also *repeated measurements*, *repeated measures analysis*.

replicate– Same as *replication*.

replication– The number of times each **treatment** is repeated in an **experiment**. It is the **sample size** associated with each treatment. The purpose of replication is to obtain more **degrees of freedom** for estimating the **experimental error** and to increase **precision** of **estimates** of **effects**.

representativeness– A term used to describe the extent to which different characteristics of a **sample** accurately represent the characteristics of the **population** from which sample was selected.

representative sample– A **sample** that is similar in terms of characteristics of the **population** to which the findings of a study are being generalized. A representative sample is not biased and therefore does not display any patterns or trends that are different from those displayed by the population from which it is drawn. It is rather difficult and often impossible to obtain a representative sample. **Nonrandom samples** usually tend to have a some kind of **bias**. The use of a **random sample** usually leads to a representative sample.

reproducibility– A term used to refer to the property of **measurements** to reproduce approximately similar results taken under different conditions such as instruments, laboratories, and operations.

resampling– The technique of selecting a **sample** many times and computing the **statistic** of interest with reweighted **sample observations**. Although resampling techniques have been used in **statistical estimation** and **hypothesis testing** for a long time, the computational complexity limited their use to all but the smallest samples. The speed and computing power of modern computers has allowed the statistics with no closed distributional forms or **variance** expressions to be analyzed by resampling techniques. Some commonly used resampling techniques include **bootstrap**, **jackknife**, and their variants.

research design– Same as *study design*.

research hypothesis– Same as *alternative hypothesis*.

residual– In **regression analysis**, the difference between the actual observed value of the **dependent variable** and the value predicted by the **estimated regression model**. It is the portion of the **score** on the dependent variable not explained by independent variables.

residual analysis– The term is used to refer to statistical methods and techniques for checking the **assumptions** of the **regression models** through examination of **residuals**. See also *Cook's distance*, *influence statistics*, *jackknife residuals*, *regression diagnostics*.

residual autocorrelation– The **autocorrelation** calculated from **residuals**. They are useful for checking the **assumptions** of the **regression models**.

residual effect– Same as *residual*.

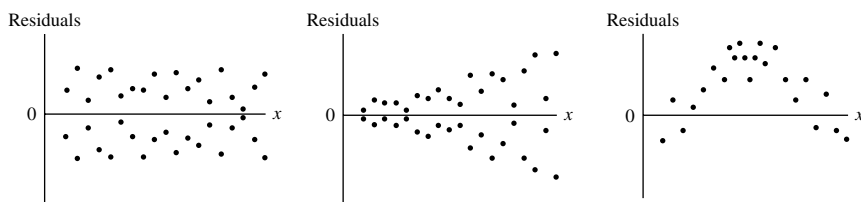
residual error– Same as *error term*.

residual error term– Same as *residual*.

residual maximum likelihood estimation– Same as *restricted maximum likelihood estimation*.

residual mean square– Same as *mean square for error*.

residual plot– The plot of **residuals** against the values of the **independent variable**.



Some examples of residual plots

residual sum of squares– Same as *error sum of squares*.

residual variable– In **path analysis**, an unmeasured **variable** that is posited to cause the **variance** in the **dependent variable** not explained by the **path model**.

residual variation– Same as *unexplained variation*.

response bias– A tendency for individual responses to survey questions to be affected or distorted in some systematic manner.

response measure– Same as *response variable*.

response rate– In a **survey**, the **proportion** of individuals who respond to a particular question or item. Compare *nonresponse rate*.

response surface– A response surface is the geometric representation obtained when a **response variable** is plotted as a function of one or more **quantitative factors** or **variables**.

response value– The particular numerical value assumed by a **response variable**.

response variable– Same as *dependent variable*.

response variate– Same as *dependent variable*.

restricted maximum likelihood estimation– A modification of **maximum likelihood procedure** where **estimators** of **scale parameters** are derived by maximizing the joint likelihood of that part of the **likelihood function** that does not contain any **location parameters**. The term is most commonly used in the context of estimating **variance components** in a **linear model**.

restricted randomization– Same as *block randomization*.

reticulation– The determination of boundaries of census areas, units, and other subdivisions in the country or delineated territory.

retrospective case-control study– Same as *case-control study*.

retrospective cohort study– Same as *historical cohort study*.

retrospective study– A general name for a **research design** in which **data** are collected on life changes of subjects over a specified past period in their lives. Thus, a retrospective study starts with persons already affected by certain condition and looks backward to discover what may have caused the appearance of that effect. Information about possible **exposure factors** is generally obtained by examining past records or interviewing each person and/or the person's relatives. The most common retrospective study is the **case-control study**. See also *cohort study*, *prospective study*.

ridge regression– A type of **regression analysis** designed to address the problem of **multicollinearity** among the **independent variables**.

ridit analysis– A chi-square analysis applied to a $2 \times k$ table to investigate the **independence** or **homogeneity**. For a dose-response analysis, the column variable must be an ordered sequence of numerical values. The analysis is also applicable to situations where the column variable represents an ordinal measure not necessarily numerical.

right-skewed distribution– Same as *positively skewed distribution*.

right-tailed test– Same as *upper-tailed test*.

risk– The **probability** that a person will develop an illness or any other condition over a specified period of time.

risk aversion– In **decision theory**, an attitude according to which a person considers the **utility** of a certain monetary gain to be higher than the **expected utility** of an uncertain prospect of equal **expected monetary gain**.

risk condition– The particular amount or condition of a **risk factor** to which a group or individual was exposed.

risk difference– Same as *absolute risk difference*.

risk factor– In **epidemiology**, a term used to designate a characteristic, such as inheritance, personal behavior, life style, or environmental condition, that is considered to be associated with a given disease or condition not necessarily a casual factor.

risk measure– Same as *measure of risk*.

risk neutrality– In **decision theory**, an attitude according to which a person considers the **utility** of a certain prospect of money to be equal to the **expected utility** of an uncertain prospect of equal **expected monetary gain**.

risk ratio– Same as *relative risk*.

risk seeking– In **decision theory**, an attitude according to which a person considers the **utility** of a certain prospect of money to be lower than the **expected utility** of an uncertain prospect of **equal expected monetary gain**.

robust estimation– A method of **statistical estimation** which is relatively insensitive to failures in the **assumptions** underlying the use of a **statistical model**.

robust estimator– See *robust estimation*.

robustness– A term used to describe the property of a statistical procedure if it is relatively insensitive to violation of certain **assumptions** on which it depends. Such a method remains useful even when one (or more) of its assumptions is (are) violated.

robust procedure– A statistical procedure that is relatively insensitive to violation of **assumptions** underlying its use. For example, **Student's *t* test** is a robust procedure against departures from **normality**. Similarly, **statistical tests** and **confidence intervals** based on **ranks** are robust against the influence of outlying **observations**.

robust regression– A type of **regression** that is relatively insensitive to failures in the **assumptions** of the **regression model**.

robust statistics– See *robust procedure*.

ROC curve– Same as *receiver operating characteristic curve*.

root mean square error– See *mean square error*.

Rosenbaum test– A **nonparametric procedure** for testing the equality of two **scale parameters** having a common **median**. The **test statistic** is based on the total number of **observations** in one **sample** that are either smaller than the smallest value or larger than the largest value in the second sample. The test was proposed by S. Rosenbaum in 1953,

who also gave tables of **critical values** of the test statistic. See also *Ansari–Bradley test*, *Barton–David test*, *Conover test*, *F test for two population variances*, *Klotz test*, *Mood test*, *Siegel–Tukey test*.

rounding– The reporting of numerical information to fewer decimal places by discarding extra digits after a certain number of places and increasing the last of the remaining digits to the next higher digit if the nearest digit being discarded is greater than or equal to 5.

rounding errors– Computing errors caused by rounding values of a quantity to fewer decimal places. Rounding errors can usually be reduced by calculating quantities to more significant figures.

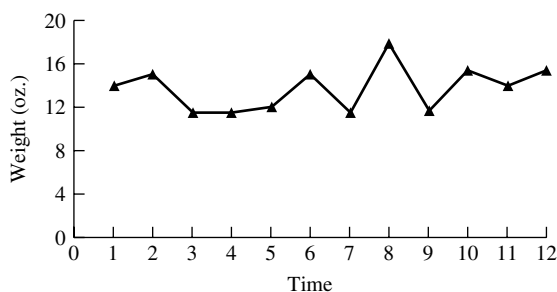
row marginals– In a **cross-tabulation**, the **frequencies** of the **variable** appearing across the rows. Compare *column marginals*.

row sum of squares– Same as *sum of squares for rows*.

Roy's largest root criterion– See *multivariate analysis of variance*.

run– A succession of identical letters or symbols that is followed and preceded by a different letter or no letter at all. The theory of runs allows us to test for **randomness**.

run chart– A simple **graphical device** used to record and display **trends in data** over time. In a run chart, the observed values are plotted on the **vertical axis** and the time they were observed on the **horizontal axis**. The main purpose of the run chart is to monitor a system or a process in order to detect any meaningful changes in the process that may take place over time. The figure below is a run chart that shows the observed weights plotted over time. The graph clearly shows an upward drift in the weights of the product, and it indicates the need for a corrective action on the process.



A run chart

run test– A statistical procedure used to test **randomness** of a sequence of **observations**. The procedure consists of counting the number of **runs** and comparing it with the expected number of runs under the **null hypothesis** of **independence**.