Statistics Glossary

Symbols:

α	alpha	significance level; probability of a type I error
β	beta	probability of a type II error
μ	mu	population mean
ν	nu	degrees of freedom
π	pi	ratio of a circle's circumference to its diameter, ≈ 3.1416
ρ	rho	Pearson product-moment population correlation coefficient
σ	sigma	population standard deviation; standard error
$\sum$	sigma	summation
df	degrees of freedom	
E(X)	expected value of X	
H_0	null hypothesis	
H_A	alternative hypothesis	
i	interval size	
n	number of observations in a sample	
PR	percentile rank	
p	probability of a success	
p(X)	probability of event X	
Q	semi-interquartile range	
q	probability of a failure	
r	Pearson product-moment sample correlation coefficient	
r^2	proportion of variance in y accounted for by x	
S	sample standard deviation (S^2 is the sample variance)	
SS	sum of squares	
X	sample score	
$\overline{X}$	sample mean	
Z'	Fisher's transformation of r	
z	standard score	

Definitions:

Analysis of variance (ANOVA)

A procedure for determining how much of the total variability among scores to attribute to a range of sources of variation and for testing hypotheses concerning some of the sources

Completely randomized design (CRD)

A study in which the assignment of participants to treatment levels is completely random; each participant is in only one treatment condition

Confidence interval

A range of values computed from data so that a specified percentage (often 95%) of all possible random samples from the same population will give intervals that contain the true population value

Correlation coefficient

A number that represents the degree of association or *strength of relationship* between two variables

Critical region

The *region for rejecting* the null hypothesis; determined by H_A and α

Cumulative frequency distribution

A distribution that shows the number, proportion, or percentage of scores that occur below the real upper limit of each interval *(including all intervals below)*

Dependent samples

The selection of participants in one sample is *affected* by the selection of participants in the other sample; keywords "matched" or "repeated"

Matched sample: matching each participant in the experimental condition with a participant in the control condition on some variable that is correlated with the dependent variable

Repeated measures: observing the same participants under both the experimental and control conditions

Histogram

>Similar to a bar graph, but used for quantitative variables; constructed by placing vertical bars over the real limits of each interval, with the height of each bar corresponding to the frequency of the interval

Independent samples

>The selection of participants in one sample is *not affected* by the selection of participants in the other sample; keyword "random"

Level of significance

>The probability that is the largest risk a researcher is willing to take of rejecting a true null hypothesis

Mean

>*Average*; sum of the scores divided by the number of scores

Median

>The *middle value* that divides the data into two equal groups

Mode

>The score or qualitative category that occurs with *greatest frequency*

Normal distribution

>A probability distribution that is *unimodal and symmetrical*; the mean, median, and mode are all the same value (the highest point on the curve)

Outliers

>Scores that *differ so markedly* from the main body of data that their accuracy is questioned

***p*-value**

>The probability of obtaining a value of the test statistic equal to or more extreme than that observed, given that the null hypothesis is true

Parameter

>Descriptive measure for a *population*; usually represented by Greek letters

Percentile (point)

>A point on the measurement scale below which a specified percentage of scores falls

Percentile rank

> The percentage of the scores of the distribution that fall below that score

Population

> The collection of *all* people, objects, or events having one or more specified characteristics

Power

> The probability of correctly rejecting the null hypothesis; $1 - \beta$

Random assignment

> The method of placing participants into the treatment groups in which each participant has an equal chance of being placed in any of the groups

Random sampling

> The method of drawing samples from a population such that every possible sample of a particular size has an *equal chance of being selected*

Relative frequency distribution

> A distribution that shows the *proportion or percent frequency* for each interval

Residual (prediction error)

> The difference between a person's actual score and predicted score

Sample

> A *subset* of a population

Sampling distribution

> A probability distribution in which the random variable is a statistic based on the results of more than one trial

Semi-interquartile range

> Half the distance between the first quartile point and the third quartile point

Standard deviation

> Measure of the spread of data that is based on every score in a distribution

Standard score

> A number that expresses the value of a score relative to the mean and standard deviation of its distribution

Skewed distributions

Distributions that are *asymmetrical*; there are two types

Negatively skewed: longer tail extends to the *left*

Positively skewed: longer tail extends to the *right*

Statistic

Descriptive measure for a *sample*; usually represented by English letters

Type I error

Rejecting a true null hypothesis

Type II error

Retaining a false null hypothesis