Glossary of Symbols

Mathematical Symbols

$b > c$	b greater than c
$b < c$	b less than c
$b = c$	b equal to c
$b \approx c$	b approximately equal to c
$+$	Plus
$-$	Minus
$\pm$	Plus or minus
∞	Infinity
$\sqrt{\ }$	Square root
$\times$	Multiplication or interaction in ANOVA
$\lvert c \rvert$	Absolute value of c; negative signs are ignored
$n!$	Factorial; $n(n-1)(n-2) \ldots (3)(2)(1)$
e	The number 2.718. . .
$\ln(c)$	Natural logarithm; logarithm to base e
$\log(c)$	Common logarithm; logarithm to base 10

Statistical Symbols

a	Intercept
a, b, c, and d	Frequencies in a 2 × 2 table
A, B, and C	Factors in ANOVA
b	Regression coefficient
C	Correction term for the mean
d	Cohen's measure of effect size
D	Difference between ranks or scores
df	Degrees of freedom
f	Frequency
H	Test statistic for Kruskal-Wallis test
H_0	Null hypothesis

k	Number of levels in one-way ANOVA
lsd	Least significant difference
MS	Mean square
n	Sample size
N	Sample size in analysis of variance
p	p value; also proportion or probability
r	Correlation coefficient
R	Sum of the ranks of the group with the smaller n
R_i	Score i's rank
r_S	Rank-order correlation or Spearman's rho
s	Sample standard deviation
s^2	Sample variance
s_p^2	Pooled variance
$s_{y.x}^2$	Error variance
2 × 2 table	Table with two rows and columns
2sd advantage	Two standard deviation advantage
S	Subject or person
S/A	Subjects within levels of A
SS	Sum of squares
T	Sum of scores or total
TOT	Total variability
U	Test statistic for Mann-Whitney test
x	Number of successes in n trials
X and Y	Variables
$\bar{X}$	Sample mean
$\hat{Y}$	Predicted score of Y
z	Fisher's z transformation
Z	Standard normal distribution

Greek Letters

α	Alpha: probability of making a Type I error
β	Beta: probability of making a Type II error
μ	Mu: population mean
ϕ	Phi: correlation between two dummy coded dichotomies
ρ	Rho: population correlation coefficient
σ	Sigma: population standard deviation
σ^2	Sigma squared: population variance
Σ	Summation sign
χ^2	Chi-square distribution
ω^2	Omega squared

Glossary of Terms

Aggregation: Creating a score that is an average or sum of other scores.
Alpha: Probability of making a Type I error.
Alternative hypothesis: Hypothesis that is true if the null hypothesis is false.
Analysis of variance: Procedure for testing the differences between means.
Analysis of variance table: Table with sums of squares, mean squares, degrees of freedom, and F ratios.
ANOVA: Analysis of variance.
Antilog: For $x = \log(y)$, y is the antilog of x; inverse logarithm function.
Arcsin transformation: Two-stretch transformation of proportions that stretches less than probit and logit.
Asymmetric distribution: Distribution whose shape changes when its mirror image is examined.
Bar graph: Graph of the frequencies of a nominal variable.
Bimodal distribution: Distribution with two peaks.
Binomial distribution: Distribution that describes the probability of x successes in n independent trials.
Cell: Particular row and column combination.
Central limit theorem: With increasing sample size, the distribution of the mean approaches a normal distribution, regardless of the shape of the original distribution of the scores.
Central tendency: Typical value of an observation from the sample.
Chi square distribution: Sampling distribution with a lower limit of zero and no upper limit; sum of independent Z^2 values; χ^2.
Chi square test of independence: Test to evaluate whether two nominal variables are associated.
Circle diagram: Representation of the partitioning of sums of squares and degrees of freedom in analysis of variance.
Class interval: Range of possible scores that can be a member of a given class.

Class midpoint: One-half the sum of a class's lower and upper limits.
Class width: Difference between adjacent lower limits.
Coefficient of variation: Standard deviation divided by the mean.
***Cohen's* d:** Measure of effect size in a two-group study; difference between the means divided by the pooled within-groups standard deviation.
Complete model: Model that contains the term that is to be tested.
Concave curvilinearity: Relationship that begins negative and becomes positive; U shape.
Constant in model: Term added to every score; often the population mean of the dependent variable.
Contrast: Set of weights assigned to levels of the independent variable in ANOVA; weights that are chosen for theoretical reasons and must sum to zero.
Convex curvilinearity: Relationship that begins positive and becomes negative; inverted U shape.
Correction term of the mean: Squared sum of all the observations which is divided by the total number of observations; symbolized by C.
Correlated correlations: Two or more correlations computed using the same sample of objects.
Correlation coefficient: Regression coefficient between Z scored variables that varies from -1 to $+1$; r.
Criterion variable: Outcome or dependent variable in a regression equation.
Critical value: Value that the test statistic must meet or exceed to be deemed statistically significant.
Cummulative frequency: Sum of the frequencies of all classes that are less than or equal to the class's upper limit.
Curvilinearity: Nonlinear relationship in which the relationship changes direction.
Data: Numerical values given to objects.
Datum: Single score.
Degrees of freedom for a contrast: One.
Degrees of freedom for χ^2 goodness of fit test: Number of levels of the nominal variable less one.
Degrees of freedom for χ^2 test of independence: $(r - 1)(c - 1)$.
Degrees of freedom for error variance in a regression equation: $n - 2$.
Degrees of freedom for* F *in one-way ANOVA: $k - 1$ in the numerator and $N - k$ in the denominator.
Degrees of freedom for interaction in two-way ANOVA: $(a - 1)(b - 1)$.
Degrees of freedom for pooled variance: $n_1 + n_2 - 2$.
***Degrees of freedom for* t:** For one-sample test, $n - 1$; for two-sample test, $n_1 + n_2 - 2$; for a test of a single correlation or regression coefficient, $n - 2$.
Degrees of freedom of the standard deviation: Sample size minus one, or $n - 1$.

Dependent variable: Outcome or variable caused by the independent variable.
Descriptive statistics: Numerical values that summarize sample data.
Dichotomy: Nominal variable with two levels.
Distribution: Shape of a sample or population; usually represented by a histogram.
Distribution-free test: Procedure for testing a model that makes no distributional assumptions.
Distribution-tied test: Test that assumes a normal distribution that is analogous to a distribution-free test.
Dummy coding: Numbers used to create a dummy variable.
Dummy variable: Numerical variable that is created by assigning arbitrary numbers to the levels of a nominal variable.
Ecological fallacy: Inferring individual relations from aggregate relations.
Effect size: Measure of the strength of effect as opposed to its p value.
Efficient statistic: Statistic with a relatively small standard error.
Error in a regression equation: Observed score minus the predicted score; the vertical distance in the scatterplot from the regression line to the point.
Factor: Nominal independent variable in ANOVA.
Factorial design: In two-way ANOVA, the creation of all possible combinations of two independent variables.
F *distribution:* Sampling distribution that is the ratio of two independent variances.
Fisher's z transformation: Transformation of a correlation that makes its distribution approximately normal.
Flat distribution: Distribution in which all scores are equally likely.
Flat transformation: Transformation that changes the shape of a distribution into flat one; rank order and percentile rank.
Frequency: Number of observations that fall in a cell of a table or the number of observations in a class interval.
Frequency table: Table with the classes and their frequencies.
Friedman two-way ANOVA: Test used to evaluate the medians and other aspects of two or more nonindependent groups.
Goodness of fit χ^2 test: Test to compare the observed distribution of a nominal variable to a predicted distribution.
Histogram: Graph of the frequency table of a distribution with the X axis being the classes and the Y axis being the frequency.
Hotelling test: Test of the equality of two nonindependent correlations in which two of the variables are in common.
Independent groups: Two or more samples that contain different persons who do not influence one another.
Independent sampling: If one object is sampled, every other object in the population has the same probability of being sampled.

Independent variable: Causal variable in a model.
Inferential statistics: Using sample data to draw conclusions about the population; tests of models.
Interaction: Effect of an independent variable changes as a function of a second variable.
Intercept: Predicted value of Y when X is zero in a regression equation in which X is the predictor and Y the criterion.
Interquartile range: Difference between the upper median and the lower median.
Interval level of measurement: Measurement level at which numbers can be used to quantify differences between objects.
Kruskal-Wallis ANOVA: Test used to compare the medians and other aspects of two or more independent groups.
Leaf: In a stem and leaf display, the next digit after the stem.
Least significant difference test: Post hoc test of means in one-way ANOVA; Tukey lsd.
Leptokurtic distribution: Distribution that has a high peak in the center and skinny tails.
Linearity: One-unit change in X produces the same change in Y regardless of where the change in X comes.
Logarithm: If $x^y = b$, y is the logarithm of b to base x.
Logit difference: In a 2×2 table the difference between logits; also the natural logarithm of the odds ratio.
Logit transformation: Natural logarithm of the odds.
Log linear model: Model for multiple nominal independent variables and a nominal dependent variable.
Lower median: Median of scores below the median of the sample.
Lowest lower limit: Lower limit of the lowest class interval.
McNemar's test: Test of the effect of a dichotomous independent variable on a dichotomous dependent variable when groups are nonindependent.
Main effect: In two-way ANOVA the effect of an independent variable averaged across levels of the other independent variable.
Mann-Whitney test: Distribution-free test that compares the medians and other aspects of two independent groups; U.
Margin: In a table, sum of frequencies across a row or a column.
Mean: Sum of the observations divided by the sample size.
Mean square: In ANOVA the sum of squares divided by degrees of freedom.
Measurement: Assignment of numbers to objects by a rule.
Median: Middle observation in a sample.
Mode: Most frequent observation in a sample.
Model: Mathematical equation specified by a theory.
Negative association: As one variable increases, the other decreases.

Negative skew: Distribution with a long, skinny tail on the left side.
Nominal level of measurement: Measurement level at which only differentiation of objects is possible.
Nonindependent groups: Two or more samples that contain the same persons or sampling units.
Nonlinearity: Relationship between two variables that varies in strength as a function of one variable.
Normal distribution: Unimodal, symmetric, bell-shaped distribution with limits of positive and negative infinity.
Normalized ranks transformation: Transformation that alters a variable's distribution to make the distribution more normal.
No-stretch transformation: Constant multiplied or added to each score; basic shape of the distribution not altered.
Null hypothesis: Constraint on the complete model that is present in the restricted model; H_0.
Odds: Proportion divided by the quantity one minus the proportion.
Odds ratio: In a 2×2 table (ad)/(bc).
Omega squared: Measure of variance explained in one-way ANOVA.
One-stretch transformation: Transformation to remove positive skew, which stretches the left side of the distribution; square root, logarithm, and reciprocal.
One-tailed test: Test in which only one alternative hypothesis is considered.
One-way analysis of variance: Method used to test for differences between independent means.
Operational definition: Set of procedures used to measure a construct.
Ordinal level of measurement: Measurement level at which objects can be rank ordered.
Outlier: Extremely large or small score.
Paired t test: Test of the difference between two nonindependent means.
Parameter: Quantity computed using all objects in the population, often symbolized by a Greek letter.
Part-whole problem: Two variables, one of which is derived from the other.
Pearson Filon test: Test of the equality of two nonindependent correlations in which none of the variables are in common.
Percentage difference: In a 2×2 table, the difference between percentages computed across either rows or columns.
Percentile rank: Percentage of scores that the object is greater than.
Phi: Correlation between two dummy-coded dichotomies.
Platykurtic distribution: Distribution with a low peak in the center and fat tails.
Pooled variance: Weighted average of variances used in two-group *t* test, where the weights are sample size minus one for each group.

Population: All possible observations.
Positive association: As one variable increases, the other increases.
Positive skew: Distribution with a long, skinny tail on the right side.
Post hoc test of means: Test in which all possible pairs of means are compared.
Power: Probability of rejecting the restricted model when the restricted model is false; one minus the probability of making a Type II error.
Power efficiency: Ratio of sample size needed for a distribution-tied test to the sample size needed for a distribution-free test in which the same power is achieved and the assumptions of the distribution-tied test hold.
Predicted score: In a regression equation, the intercept plus the predictor score times the regression coefficient.
Probit transformation: Two-stretch transformation of proportions based on the standard normal distribution.
p ***value:*** The probability of obtaining a value of the test statistic at least as large as the one obtained.
Random assignment: Each object having the same probability of being assigned to a level of the independent variable.
Random sample: Each object equally likely to be chosen from the population.
Range: Crude measure of variability; largest score minus the smallest score.
Rank-order correlation: Spearman's rho; correlation between ranks; r_S.
Rank-order transformation: Scores rank ordered from smallest to largest and the smallest score assigned a 1, the next a 2, and so on.
Reciprocal transformation: One-stretch transformation in which one is divided by the score; $1/X$.
Rectangular distribution: Flat distribution.
Regression coefficient: Measure of association of how much a one-unit change in the predictor variable creates in the criterion variable.
Regression equation: Criterion equals the intercept plus the regression coefficient times the predictor.
Regression toward the mean: Predicted scores in a regression equation are less variable than the scores of the criterion.
Relative frequency: One hundred times the frequency divided by sample size.
Reliability: Proportion of true variance in a variable.
Repeated measures design: All subjects measured at each level of the independent variable.
Residual variable: All other sources of variation in the dependent variable besides that due to the independent variables.
Restricted model: Model that is a constrained version of the complete model, the constraint being the null hypothesis.
Restriction in range: Variable with limited variability.

Robust statistic: Statistic not influenced much by outliers.
Sample: Set of scores that refer to different objects.
Sample size: Number of observations in the sample; n.
Sampling distribution: Distribution of a statistic that is created by drawing repeated samples and recomputing the statistic.
Sampling error: The fact that a statistic changes when it is recomputed using a different sample.
Scatterplot: Graph to represent the association between two variables; variables form the axes and points are the data.
Significance level: Alpha or the probability of making a Type I error.
Sign test: Distribution-free test for evaluating the difference between the medians and other aspects of two nonindependent groups.
Skew: Long, skinny tail on just one side of a distribution.
Slope: Regression coefficient; linear measure of association.
Smoothed frequency: One-half the class's frequency plus one-quarter the sum of the adjacent class frequencies.
Smoothing: Procedure to make a frequency table less influenced by choice of lowest lower limit and class width.
Spearman's rho: Correlation coefficient of ranks; r_S.
Standard deviation: Measure of variability that uses all observations; square root of the variance; s.
Standard error of the mean: Standard deviation divided by the square root of the sample size.
Standard normal distribution: Z distribution; normal distribution with a mean of zero and a variance of one.
Standard score: Z score; score minus the mean and the difference divided by the standard deviation.
Statistic: Quantity computed from sample data.
Statistically significant: p value is equal to or less than the significance level.
Stem: Lower limit of a class used to represent a class interval in a stem and leaf display.
Stem and leaf display: Vertical histogram that essentially preserves the raw data.
Sum of squares: In ANOVA the numerator of a mean square.
Summation sign: Symbol that represents the sum of all the scores; Σ.
Symmetric distribution: Distribution that when folded vertically perfectly coincides.
Tail of a distribution: Frequency of very large or very small values.
t distribution: Sampling distribution used to test hypotheses about means and to test correlation coefficients.
Test statistic: Quantity computed from sample data used to evaluate the plausibility of a restricted model.
Tukey least significant difference test: Post hoc test in one-way ANOVA.

Two-standard deviation advantage: Measure of effect size of r that equals how much more likely someone who is one standard deviation above the mean on X will outscore on Y someone who is one standard deviation below the mean on X.
Two-stretch transformation: Transformation that is used to remove lower and upper limits, commonly used on proportions; arcsin, probit, and logit.
Two-tailed test: Test in which the two alternative hypotheses are considered.
Two-way analysis of variance: Procedure to evaluate models with two nominal independent variables and an interval dependent variable.
Type I error: Rejecting the restricted model when it is true; alpha, or α.
Type II error: Retaining the restricted model when it is false.
Unbiased statistic: Statistic whose mean of the sampling distribution equals the population parameter that the statistic is estimating.
Unimodal distribution: Distribution with one peak.
Unit in a distribution: Smallest possible difference between a pair of scores.
Unit of measurement: Term that defines the meaning of a one-point difference between two scores.
Upper median: Median of scores above the median of the sample.
Variability: How much the observations differ from one another.
Variance: Measure of variability that is based on deviations from the mean; s^2.
X *axis:* Horizontal (left to right) axis in a graph.
Y *axis:* Vertical (up and down) axis in a graph.
Z *distribution:* Standard normal distribution; a normal distribution with a mean of zero and a variance of one.
Z *score:* Score in which the mean has been subtracted and this difference is divided by the standard deviation.
z *transformation:* Transformation to make the distribution of r more normal, commonly called Fisher's r to z transformation.
Z *transformation:* Scores in which the mean has been subtracted and this difference is divided by the standard deviation.