

A Review of Basic Statistical Tests

First Edition, Revised | Summer 2006

Test Name	When to Use It	How to Report It	Degrees of Freedom	What the Results Mean
Tests that compare only two groups				
One-Sample z-Test	To compare a single sample mean to the population mean if the population standard error is known.	$Z = \text{____}, p < .05$	$(N - 1)$	Compare the Z score to the normal distribution table. For a two-tailed test, your results are significant if Z is greater than or equal to +/- 1.96.
One-Sample t-Test	To compare a single sample mean to the population mean when the population standard error is NOT known.	$t(df) = \text{____}, p < .05$	$(N - 1)$	Compare the t-score to a t-distribution table. A significant t-value means that your sample mean is significantly different from the population mean.
Independent Groups t-Test	When you compare two sample means in which the samples are made up of different groups of individuals (between subjects).	$t(df) = \text{____}, p < .05$	$[(n - 1) + (n - 1)]$ or $(N - 2)$	Compare the t-score to a t-distribution table. A significant t-value means that your group means are significantly different from one another.
Repeated Measures t-Test	When you compare two sample means in which the samples are made up of the same group of individuals tested twice (within subjects).	$t(df) = \text{____}, p < .05$	$(N - 1)$	Compare the t-score to a t-distribution table. A significant t-value means that your group means are significantly different from one another.
Tests that compare three or more groups				
One-Way Between Subjects ANOVA	When you compare three or more sample means in which the samples are made up of different individuals.	$F(df_{bet}, df_{w/in}) = \text{____}, p < .05$	$df_{bet} = (k - 1)$ $df_{w/in} = (N - k)$	A significant F-value means that one of the groups is different from one of the other groups. You must run post-hoc tests to figure out which groups differ.
One-Way Repeated Measures ANOVA	When you compare three or more sample means in which the samples are made up of the same individuals tested multiple times.	$F(df_{IV}, df_{Error}) = \text{____}, p < .05$	$df_{IV} = (k - 1)$ $df_{Error} = [(k - 1) \times (N - 1)]$	A significant F-value means that one of the groups is different from one of the other groups. You must run post-hoc tests to figure out which groups differ.
Tukey's HSD Post-Hoc Test	After you find a significant F-value in ANOVA, you use this test to find out which conditions differ from which other conditions.	Consult a Statistics Text.	N/A	This is a <i>pairwise</i> test, meaning it tests all possible pairs of conditions to see if they differ. Using this test helps guard against Type I Errors.
Tests that examine two or more Independent Variables				
Multi-Factor Between Subjects ANOVA	Use when testing two or more IVs, and each cell represents a different group of participants. Yields multiple F-tests.	$F(df_A, df_{error}) = \text{____}, p < .05$ $F(df_B, df_{error}) = \text{____}, p < .05$ $F(df_{A \times B}, df_{error}) = \text{____}, p < .05$	<i>For a basic two-way design:</i> $df_A = k_A - 1$ $df_B = k_B - 1$ $df_{A \times B} = df_A \times df_B$ $df_{error} = (N - k_A - k_B)$	This test yields multiple F-scores: one for every IV, and one for every possible interaction. You must evaluate all F-scores for significance.
Multi-Factor Within Subjects ANOVA	When testing two or more IVs and all participants receive all conditions of the study. Yields multiple F-tests.	$F(df_A, df_{error}) = \text{____}, p < .05$ $F(df_B, df_{error}) = \text{____}, p < .05$ $F(df_{A \times B}, df_{error}) = \text{____}, p < .05$	Consult a Statistics Text.	This test yields multiple F-scores: one for every IV, and one for every possible interaction. You must evaluate all F-scores for significance.
Mixed-Factor ANOVA	When testing two or more IVs, where one IV is between-subjects (e.g. gender) and the other is within subjects (e.g. treatment vs. control).	$F(df_A, df_{error}) = \text{____}, p < .05$ $F(df_B, df_{error}) = \text{____}, p < .05$ $F(df_{A \times B}, df_{error}) = \text{____}, p < .05$	Consult a Statistics Text.	This test yields multiple F-scores: one for every IV, and one for every possible interaction. You must evaluate all F-scores for significance.
Correlational Tests (Test for relationships)				
Pearson Product-Moment Correlation	Tests the <i>relationship</i> between two variables.	$r(df) = \text{____}, p < .05$	$N - 2$	Pearson's r tells you the strength and direction of a relationship between two variables. It ranges from -1 to +1.
Multiple Regression: Overall Model Test	Tests the ability of a set of variables (predictors) to predict the outcome of a DV (criterion).	$R^2 = \text{____}$ $F(df_{reg}, df_{resid}) = \text{____}, p < .05$	$df_{reg} = p$ $df_{resid} = N - p - 1$	R^2 tells you the proportion of variance in your criterion explained by all your predictors. It ranges from 0 to 1. The F-test tells you if your R^2 is significant.
Multiple Regression: Coefficient Tests	Tests the ability of one predictor in a set of predictors to predict the outcome of the criterion. Used in conjunction with the above test.	$b = \text{____}, t(df) = \text{____}, p < .05$	$df = N - p - 1$	You will compute one t-test for each predictor in your model; these tests tell you if the individual predictor is significant.
Measures of Effect Size				
Eta-Squared (η^2)	Eta-squared is a universal measure of effect size. It can be used on designs with any number of groups.	eta-squared = ____ or $\eta^2 = \text{____}$	<i>No df necessary</i> Eta-squared is calculated: (SS_{IV} / SS_{total})	Like R^2 , eta-squared tells you the proportion of variance in your DV that is explained by your IV. It ranges from zero to one.
Cohen's d	This will give you a measure of effect size, but only when comparing two groups.	Cohen's d = ____	<i>No df necessary</i> Cohen's d is calculated: $[(\text{mean}_1 - \text{mean}_2) / \text{std. dev}]$	Cohen's d tells you the number of standard deviations that separate your two group means.

$k = \#$ of groups | $p = \#$ of predictors | $N =$ total sample size | $n =$ group sample size

Quick Glossary of Common Statistics Terms:

Assumptions

Each test has a set of assumptions that must be met in order to be conducted properly. Check with a statistics text for the assumptions of the above tests.

Alternative Hypothesis

Complements the null hypothesis, and tests a specific prediction about a relationship between two or more variables.

Between-Subjects

Design term referring to the fact that participant groups are composed of separate individuals.

Cell

A "cell" refers to a unique combination of IV levels. A 2x3 design has 6 cells, a 3x2x2 design has 12 cells, etc. (see *Multi-Factor ANOVA Notation*)

Interaction (two-way)

An interaction occurs in a multi-factor design when the effect of one IV differs based on the level of a second IV.

Levels of Measurement

This refers to the properties of measured data: Nominal, Ordinal, Interval and Ratio. All tests on this guide require Interval or Ratio data.

Mean

The arithmetic average of a set of numbers.

Median

The middle number in a set, when the numbers are arranged in ascending order.

Mode

The most frequently occurring number in a set.

Multi-Factor ANOVA Notation

ANOVAS are often described as 2x2, 4x5, 3x2x3, etc. These numbers represent the

number of levels in each of the IVs present. (3x2x3 = 3 IVs; 3 levels in the first IV, 2 in the second IV and 3 in the third IV)

Null Hypothesis

Hypothesis that there is no relationship between two or more variables. The Null Hypothesis can never be proven—just disproven.

Power

The probability that you will correctly reject the null hypothesis. Having many participants, low measurement error and looking for large effects all increase power.

Repeated Measures

Design term referring to the fact that the same group of participants received all experimental conditions.

Standard Deviation

The extent to which scores deviate from the mean—the larger the standard

deviation, the more spread out the scores are.

Statistical Significance

A test results is considered "significant" if there is less than a 5% chance that a positive result is due to a Type I Error.

Standard Error of the Mean

Measure of how much error one may expect between a sample mean and the true population mean.

Type I Error

Rejecting the null hypothesis when you shouldn't have. This is the worse of the two forms of errors.

Type II Error

Failing to reject the null hypothesis when you should have rejected it.