HYPOTHESISTESTING

Hypothesis testing is the starting point for learning inferential statistics. We want to know something about the world and we use hypothesis testing to evaluate statements about how the world works.

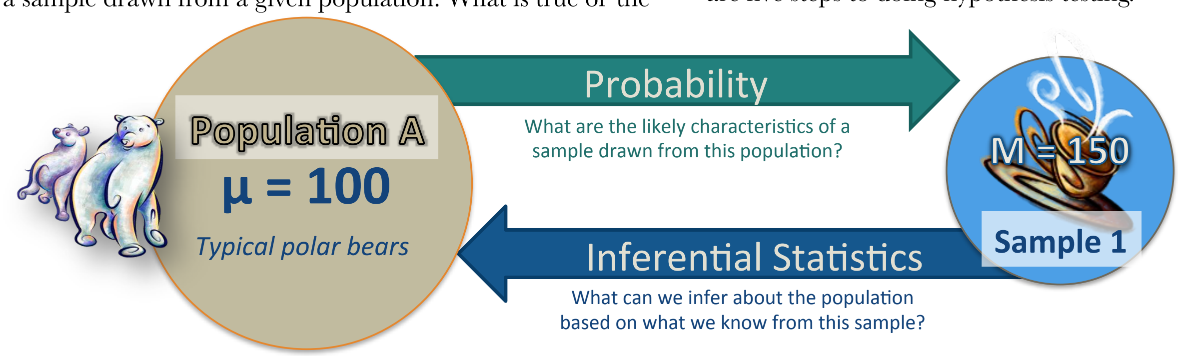
An Introduction to Hypothesis Testing
inferential statistics

By Todd Daniel Ph.D.

Null Hypothesis Statistical Testing (NHST; a.k.a. *significance testing* and *Null Hypothesis Decision Making*) is a statistical method that uses sample data to evaluate assumptions (i.e. test a hypothesis) about a population parameter. Asked more simply, "What can we infer about the population based upon what we have learned from our sample?"

Probability tells us what characteristics we can expect from a sample drawn from a given population. What is true of the population (mean, SD) should also be true of the sample. There will be some random variation, but not too much.

Inferential statistics are *parametric procedures* used to decide whether a given sample mean differs statistically significantly from another mean. We start with a hypothesis that the sample mean is really the same as the population mean. We test whether the differences that we observe are more likely due to random variation or due to an effect from the treatment. There are five steps to doing hypothesis testing.



Hypothesis Testing
A Five-Step Guide

By Todd Daniel Ph.D.

Questions to Ask Before you Begin

What are your Variables?

- Independent variable (IV)
- Dependent variable (DV)

What do you want to know?

- Relationships between groups?
- Differences between groups?

What do the data look like?

- What level are the DV data: NOIR?
- Have the assumptions been met?
- How many IV groups do you have?

What type of test will you use?

- Directional or non-directional H_1?
- Is your alpha level .05 or .01?

Squirrels love this Statistical Hypothesis Inference Testing

Step 1:

Select the Appropriate Statistic

Use the diagrams for Step 1 to choose the right statistical test. Choose the test based on whether you are looking for relationships or differences between groups, the level of the data (NOIR), and whether the assumptions for the test have been met.

***Exploratory data analysis* will help you know what test to choose**

Step 2:

State the Null and Alternative Hypothesis

The null hypothesis states that there is really *no difference* between means, so any apparent differences are due to random variation or *chance*. The alternative hypothesis states that one group mean will be *statistically significantly* different than the other.

Typically the null hypothesis is written as H_0: and the alternative is H_1:

Step 3:

Select a Level of Significance

After choosing a one-tailed vs. two-tailed test, choose a level of significance (a.k.a alpha level). The most common level is α = .05; other common levels are α = .01 for medical research or α = .10 for political polling. The level of significance *sets* your critical value.

For most tests, assume α = .05 and a two-tailed test

Step 4:

Calculate the Statistics

Calculate a test value. When working by hand, compare the test value to the critical value from a table. When using software, the software will calculate a test statistic and compute a probability (p) value. If the p is less than α, the test is *statistically significant*.

In SPSS, the *p* value is in a box labeled "Sig."

Step 5:

Make a Decision!

If your calculated statistic is in the critical region (region of rejection), you reject the null hypothesis. The null says there is no difference between means; if you reject the idea that there is no difference, you are inferring that there there is a difference.

Write up your results in proper APA style

Choosing the Right Statistical Test

Step 1: Looking for Relationships Between Groups

By Todd Daniel Ph.D.

What level is the dependent variable (DV)?

The DV data are at the **Nominal** level

Relationship Between Nominal Variables
Chi-Square test

Are the variables independent or related? Are the data a "good fit" to what was expected? Both IV and DV are nominal/categorical.

The DV data are at the **Ordinal** level

Relationship Between Ordinal Variables
Spearman's Rank Order Correlational Coefficient

Measures how much the variables change together *i.e. strength of linear association*. Both the IV and the DV are ordinal because they are ranked. Spearman is resistant to outliers in scale data.

* Both **nominal** and **ordinal** data are "categorical"

The DV data are at the **Scale** level

Relationship Between Scale Variables
Pearson's Product Correlational Coefficient

Measures how much the variables change together *i.e. strength of linear association*. Both IV and DV are scale level. The *r* value will be between +1.00 and -1.00.

Predictions Using 1 Predictor
Simple Linear Regression

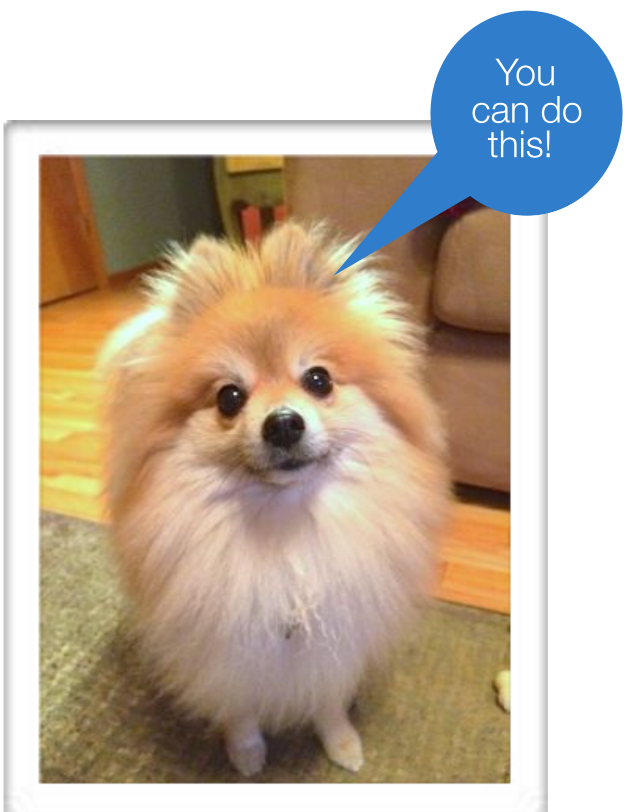
Uses one scale IV to predict one scale DV. In the output, the *a*-value is called **Constant** and the *b*-value is called **b1**.

Equation: $\mathbf{Y = a + b_1X_1}$

Predictions Using 2 or more Predictors
Multiple Regression

Uses multiple scale IVs to predict a single scale DV; this provides a better model to predict the criterion (outcome) variable.

Equation: $\mathbf{Y = a + b_1X_1 + b_2X_2 + \ldots}$



Choosing the Right Statistical Test

Step 1: Looking for Differences Between Groups

By Todd Daniel Ph.D.

What level is the dependent variable (DV)?

DV data are at the **Nominal** level

- **1 Group** → **One-Way Chi-Square test**: Are the data a "good fit" to what was expected? IV is nominal; DV is frequency
- **2 Groups** → **Two-Way Chi-Square test**: Good fit test; IVs nominal with multiple groups, DVs are frequencies; 2x2 crosstabulation

DV data are at the **Ordinal** level

Non-parametric (NP) tests

- **1 Group** → NP One-Sample *t* Test: **Wilcoxin Signed-Rank Test** — Tests whether sample median differs from a hypothesized population median value
- **2 Groups** → NP Independent Samples *t* Test: **Mann-Whitney *U* Rank Sum Test** — Compares two independent samples (IV); preferable with extreme outliers and/or sample sizes less than 20
- **2 Groups** → NP Paired-Samples *t* Test: **Wilcoxin Matched-Pair Signed-Rank Test** — Compares two repeated/paired/dependent samples (IV), DV is ranks with median test
- **3+ Groups** → NP One-Way ANOVA: **Kruskal-Wallis One-Way ANOVA** — Compares a single IV with at least two (typically 3+) independent levels/groups; DV is ranks with median test
- **3+ Groups** → NP Repeated Measures ANOVA: **Friedman Test** — Compares a single IV with at least two (typically 3+) dependent levels; DV is ranks with median test

DV data are at the **Scale** level

Are the DV data **normally distributed** and do the groups have similar **standard deviations**? EDA

NO — Use a non-parametric test

Parametric tests

- **1 Group** → Single Samples: **One-Sample *t* Test** — Compares the mean (DV) of one sample (IV) to a population mean
- **2 Groups** → 2 Independent Samples: **Independent-Samples *t* Test** — Compares the means (DV) of only two independent samples (IV)
- **2 Groups** → 2 Dependent Samples: **Paired-Samples *t* Test** — Compares the means (DV) of only two dependent samples (IV), *i.e. the same people are measured twice or samples have been deliberately matched*
- **3+ Groups** → 3+ Independent Samples: **One-Way ANOVA** — Compares the means (DV) of a single IV with at least two (typically 3+) independent levels (i.e. groups); analyzes variability of groups
- **3+ Groups** → 3+ Dependent Samples: **Repeated Measures ANOVA** — Compares the means (DV) of a single IV with at least two (typically 3+) dependent levels (i.e. repeated measures); analyzes variability

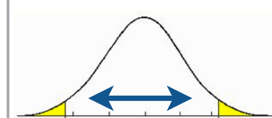
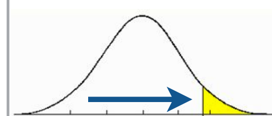
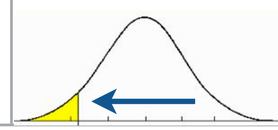
The Null Hypothesis

Step 2: Select a null and alternative hypothesis

By Todd Daniel Ph.D.

Alpha Levels and Directional Tests

Hypotheses predict an outcome that we may or may not find in our experiment. The simplest strategy is to say that there is a relationship between the IV and DV, but we do not know which direction the DV data will change when we manipulate the IV (two-tailed test). A more complicated strategy not only predicts the existence of a relationship but also predicts the direction the DV data will change when we manipulate the IV (one-tailed test).

Type of Test	Example Research Question	Null Hypothesis	Alternative Hypothesis	
Two-Tailed	Will this drug affect stamina?	H_0: $\mu = 100$	H_1: $\mu \neq 100$	
One-Tail Right (Positive)	Will this drug increase concentration?	H_0: $\mu \leq 100$	H_1: $\mu > 100$	
One-Tail Left (Negative)	Will this drug decrease blood pressure?	H_0: $\mu \geq 100$	H_1: $\mu < 100$	

How To Create a Null and Alternative Hypothesis

Questions to Ask Yourself…	**1. Is the research question one-tailed (directional) or two-tailed (non-directional)?** **2. What is the mean of the population?*** *(This number will come from the research question.)* **3. What direction does the dependent variable change? Should it increase or decrease?** If the DV *increases*, you will use > in the alternative hypothesis. If the DV *decreases*, you will use < in the alternative hypothesis. If the DV simply *changes*, you will use ≠ in the alternative hypothesis.
Two-Tailed	***Create the null hypothesis first:*** **H_0: $\mu = 50$ *** ***Then create the alternative:*** **H_1: $\mu \neq 50$** *Example of a two-tailed (non-directional) hypothesis* The height of the average gnome is 28 inches tall. H_0: $\mu = 28$ H_1: $\mu \neq 28$
One-Tailed	***Create the alternative hypothesis first:*** **H_1: $\mu > 50$ *** *Examples of possible null hypotheses*: $\mu = 50$, $\mu \leq 50$, $\mu \geq 50$ ***Then create the null hypothesis with what is left:*** **H_0: $\mu \leq 50$** *Remember that the null hypothesis will always include the equal sign.* *Examples of one-tailed (directional) hypotheses* The average grizzly bear eats more than 12 trout per day. (*one-tail - right*) H_0: $\mu \leq 12$ H_1: $\mu > 12$ Most people wearing diapers are less than 30 months old. (*one-tail - left*) H_0: $\mu \geq 30$ H_1: $\mu < 30$

> greater than
< less than
≥ greater than or equal to
≤ less than or equal to

3 < 5 and 7 > 1
"The alligator always eats more."
The equal sign always goes with the Null Hyp.

**The number 50 is used as a placeholder for the mean of the population. Do not use the number 50 in your hypotheses unless 50 happens to be the mean of the population. And don't include the asterisk, either.*

Calculating a Critical Value

Step 3: Select a level of significance

By Todd Daniel Ph.D.

Alpha Levels and Critical Values

You will compare two scores. Your goal is to set a mathematically-defined limit, “fence,” or “cut-off” beyond which any differences between the two scores are so extreme that you must conclude that they are truly different. The probability of an event occurring at random (by chance) is represented by alpha: α or a

Alpha level: a level of significance, cut-off point, or “fence” that *sets* the critical region (most commonly α = .05)

Critical Region: Area of rejection of the null hypothesis created by the alpha-level “fence”. In the picture below, it is the purple shaded area of the normal curve. The critical region comprises extreme values of a statistic that are unlikely to occur if the Null Hypothesis is true. When the calculated sample value falls **in** the critical region (tail), you reject the null. This means that the difference did not occur by chance.

Calculating the Critical Region (Inclusion and Exclusion)

Setting an alpha level at α = .05 means that 5% of the scores are in the “tail” (*critical region*) and the remaining 95% are in the “body” of the curve. If we set α = .01, then 1% of the scores would be in the tail(s).

In a **one-tail test**, the full 5% of the scores is in the right tail (if you are looking for an increase in scores) or the full 5% of the scores is in the left tail (if you are looking for an decrease in scores)

Two-tail tests assume that the 5% exclusion is from both ends. This means that 2.5% of the scores will be in the right tail and 2.5% of scores will be in the left tail.

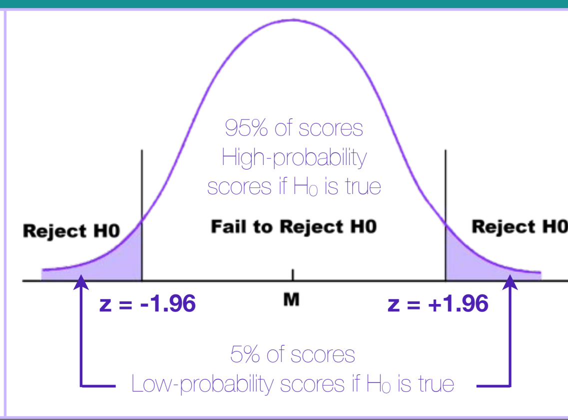
Alpha Levels for a Normal Curve (z-test)
(these values will differ for other distributions)

α = .05 two tail test: ±1.96

α = .05 one-tail test: +1.645 *or* -1.645
(right) (left)

α = .01 two tail test: ±2.575

α = .01 one-tail test: + 2.325 *or* -2.325
(right) (left)



For a two-tail z test, the critical value of z = ±1.96, (*this will be different for a t-test or ANOVA.*)

SPSS will give you a p-value in a box labeled “Sig. 2 Tail”: compare the p value to your α value.

- If p < α you reject the null hypothesis (although you cannot say that it is actually false.)
- If p > α you “accept” * that the null hypothesis is true. [p = .23 > α = .05]

**In truth, we never actually”prove” that the alternative hypothesis is correct and we never actually “accept” the null hypothesis; we only “fail to reject” it.Statistics provide evidence.*

Degrees of Freedom

Step 3: Select a level of significance

By Todd Daniel Ph.D.

Calculating the Critical Region for Other Tests

For other tests (*t-test, ANOVA, Chi-square*), you will look up the critical value on a table. These tables are included at the back of your class notes. Consult the class notes for the type of test you are using to find the appropriate table and instructions on how to use it. Using most tables to calculate the critical region will require knowing the degrees of freedom (df). In most cases, the degrees of freedom will be n-1.

Type of Test	Degrees of Freedom	Example
One-Sample z Test	*none*	The z test does not require degrees of freedom because you are not estimating population parameters
One-Sample T Test	$n-1$	$t(24) = 3.23$, $p = .022$
Independent-Samples T Test	$(n_1 + n_2) - 2$	$t(14) = 1.59$, $p = .135$, *ns*
Paired-Samples T Test	$n-1$ (where n = number of pairs)	$t(7) = 2.76$, $p = .028$
One-Way ANOVA	$df_B = k-1$, $df_W = n-k$, $df_T = n-1$	$F(3, 16) = 10.49$, $p < .001$
Repeated-Measures ANOVA	$df_B = k-1$, $df_W = n-k$, $df_T = n-1$	$F(3, 16) = 10.49$, $p < .001$, *ns*
Correlation	$n-2$ (where n = number of pairs)	$r(48) = +.657$
Linear Regression	*not reported*	Report beta values and *n*
One-Way Chi-Square Test	$k - 1$	$\chi^2(2, N = 70) = 1.15$, $p = .563$, *ns*
Two-Way Chi-Square Test	$(K_{row} - 1) * (K_{column} - 1)$	$\chi^2(3, N = 350) = 46.03$, $p < .05$

n = number of participants
k = number of IV categories

Calculate the Statistic

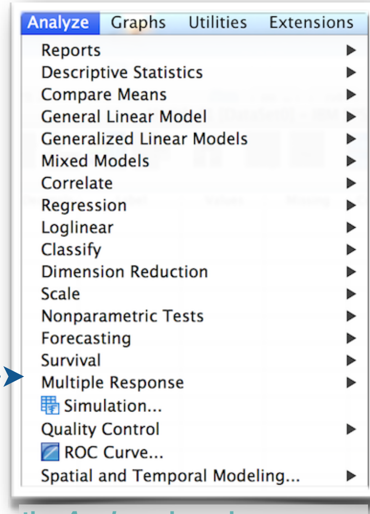
Step 4: Run the appropriate test in SPSS

By Todd Daniel Ph.D.

The Analyze Menu

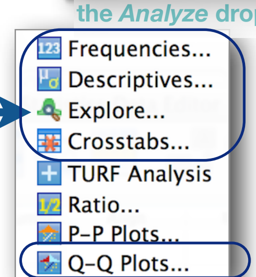
all analysis starts here

When you run a test in SPSS, you begin with the **Analyze** menu. You will not use all of these commands at an introductory level, but you should become familiar with the commands that you will need for this class. When you click on *Analyze*, a drop down menu opens from which you can choose the appropriate test.

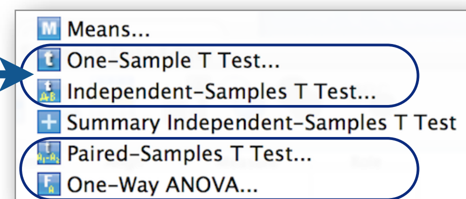


the *Analyze* drop down menu

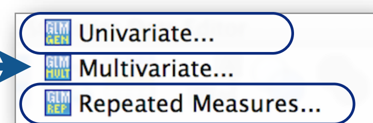
Descriptive Statistics: used for exploring data and calculating descriptive statistics. Both *Frequencies* and *Explore* offer similar options but one may be more suitable to a particular task, so you should know both. *Descriptives* menu is the fast track to descriptive stats, although you might find *Frequencies* more usable. *Crosstabs* is used for the two-way chi-square and cross-tabulation. Q-Q plots are used to assess normality or compare any two distributions.



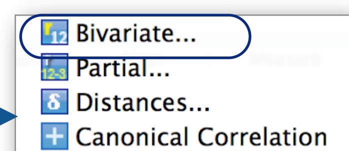
Compare Means: you will use this menu a lot. This is where you find the most commonly used hypothesis testing tools. Here is where you will find three types of *t* tests, as well as simple Analysis of Variance (ANOVA).



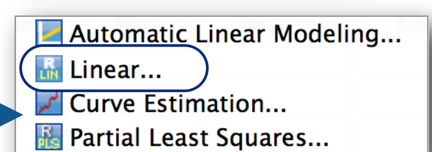
General Linear Model (GLM): You can also do an ANOVA from the GLM menu using the *Univariate* command. GLM allows you to request effect sizes, power, and crate plots. You can also do a two-way ANOVA or a repeated measures ANOVA. *Repeated Measures* is not available in some student versions of SPSS.



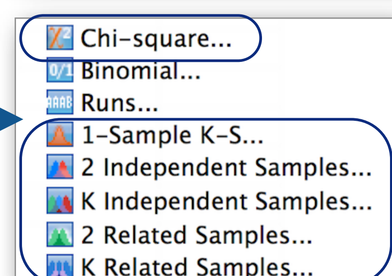
Correlate: Simple correlation, both Pearson and Spearman, is performed from this menu using the *Bivariate* command. *Partial* or *Distances* commands control for a third variable.



Regression: Simple & multiple linear regressions are conducted using the *Linear* command. You can feed multiple IVs into the regression model to create a predictive equation for a single DV.



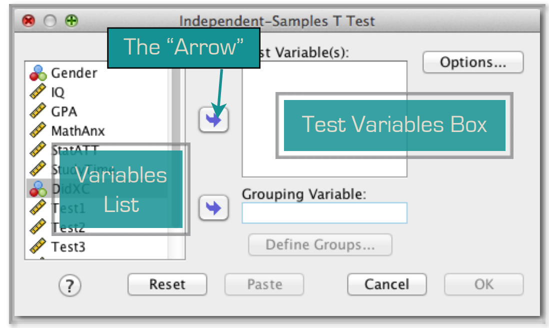
Nonparametric Tests: When you have categorical data (or your scale data violate the assumptions of a parametric test), you can use a nonparametric alternative. This menu is the ***Legacy Dialogs*** menu. Chi-square can be conducted here (or in the *Crosstabs* command in the *Descriptive Statistics* menu). Kolmogorov-Smirnov, Mann-Whitney, Kruskal-Wallis, Wilcoxin, and Friedman are all conducted here.



Calculate the Statistic

Step 4: Run the appropriate test in SPSS

By Todd Daniel Ph.D.



Parametric Analyses in SPSS

In order to run a test, you have to tell SPSS which variables you want to use. You have three options for moving variables from the Variables List on the left to Test Variables boxes.

1. Click on the variable name then click on the arrow between the boxes
2. Click and drag the variable name into the variables box
3. Double-click on the variable name *(works with only one variables box)*

One-Sample T Test

Analyze ➔ Compare Means ➔ One-Sample T-Test

1. Move a scale variable into the Test Variables Box
2. Enter the comparison value (typically the population mean) into the Test Value box. **OK**

Independent-Samples T Test

Analyze ➔ Compare Means ➔ Independent-Sample T-Test

1. Move a scale variable into the Test Variables Box
2. Move a nominal (categorical) variable into the Grouping Variable box.
3. Click on "Define Groups" and enter the number for each group (typically 1 & 2). **Continue** **OK**

Paired-Samples T Test

Analyze ➔ Compare Means ➔ Paired-Samples T-Test

1. Move two scale variables into the first pairs in the Test Variables Box
2. Note: a second pair will appear automatically and can be used or ignored **OK**

One-Way ANOVA

Analyze ➔ Compare Means ➔ One-Way ANOVA

1. Move the DV into the Dependent List box
2. Move the IV into the Factor box
3. Post Hoc: choose a follow up test (Tukey's LSD or Bonferroni are safe bets) **Continue**
4. Options: choose Descriptives. **Continue** **OK**

Repeated-Measures ANOVA

Analyze ➔ General Linear Model ➔ Repeated Measures

1. Type the name of the factor in the Within-Subject Factor Name box (can be any name that fits)
2. Enter the number of levels of that factor. **Add. Define**
3. Move each level of the treatment into the Within Subjects variable box
4. Plots: Move the factor to the Horizontal Axis box. **Add. Continue**
5. Options: Choose Descriptive Statistics, Estimates of Effect Size, and Observed Power. **Continue** **OK**

Correlation

Analyze ➔ Correlate ➔ Bivariate

1. Move all variables to be analyzed into the Test Variables Box
2. Leave the Pearson option on for scale variables. Choose Spearman if your variables are ordinal or highly skewed. **OK**

Simple Linear Regression

Analyze ➔ Regression ➔ Linear

1. Move your criterion (outcome) variable to the Dependent box.
2. Move the predictor variable to the Independent(s) box.
3. Statistics: click R squared change and Descriptives along with the default options. **Continue** **OK**

Multiple Regression

Analyze ➔ Regression ➔ Linear

1. Move your outcome variable to the Dependent box.
2. Move your predictor variables to the Independent box.
3. Method: Enter is safest; Hierarchical for theory-driven; avoid Stepwise (except for some exploratory analysis)
4. Statistics: click R squared change and Descriptives along with the default options. **Continue**
5. Plots: Move *ZPRED to Y and *ZRESID to X; check Histogram and Normal probability Plot **Continue** **OK**

Calculate the Statistic

Step 4: Run the appropriate test in SPSS

By Todd Daniel Ph.D.

Non-Parametric Analyses in SPSS

When running an analysis, it is important to set up your variables properly in SPSS. If you are using correlation, regression, or repeated measures, then your values for each case are listed in separate columns. If you are using independent measures, all scale variables go into the same column with a second column of categorical data to identify the groups. It is also important to know whether your scale data are normally distributed by using exploratory data analysis. If the scale data violate the assumptions of the test you are using (i.e. non-normality), you can use non-parametric alternatives, instead.

One-Way Chi-Square Test

<u>Data Entry</u>

1. In Data View, enter your IV using 1, 2, 3..., for each category (see **Chi Square set-up,** below)
2. Enter the observed frequencies of each category

Data ➜ Weight Cases

3. Select Weight cases and click "weight cases by"
4. Move the Observed/Frequency values into Frequency Variable. **OK**

Analyze ➜ Nonparametric Tests ➜ Legacy Dialogs ➜ Chi-square

5. If testing randomness, leave defaults as-is.
6. If testing against known values, enter them in order under Expected Values (click **Add** after each one)
7. <u>Options</u> click Descriptives. **Continue** **OK**

Two Way Chi-Square Test

<u>Data Entry</u>

1. In Data View, enter your first IV using 1, 2, 3..., for each category (see example below)
2. Enter your second IV in the next column, again using 1, 2, 3..., for the second category
3. Enter the observed frequencies of each category

Data ➜ Weight Cases

4. Select Weight cases and click "weight cases by"
5. Move the Observed/Frequency values into Frequency Variable. **OK**

Analyze ➜ Descriptive Statistics ➜ Crosstabs

6. Move the "predictor"variable into the Rows box
7. Move the "outcome" variable into the Columns box
8. <u>Statistics</u>: Chi-square. **Continue** **OK**

Gender	Outcome	Frequency
1	1	23
1	2	15
1	3	22
2	1	21
2	2	28
2	3	11

Chi Square set-up: in this example Gender (1 = male, 2 = female) and Outcome (1 = completed task on time, 2 = completed after time ran out, 3 = did not complete). The Frequency is recorded for each combination and the data are weighted by the frequency scores.

Mann-Whitney U Rank Sum Test

Analyze ➜ Nonparametric Tests ➜ Legacy Dialogs ➜ 2-Independent Samples

1. Move a scale variable into the Test Variables Box
2. Move a nominal (categorical) variable into the Grouping Variable box.
3. Click on "Define Groups" and enter the number for each group (typically 1 & 2).
4. <u>Options</u>: click on Descriptives. **Continue** **OK**

Wilcoxin Matched-Pair Signed-Rank Test

Analyze ➜ Nonparametric Tests ➜ Legacy Dialogs ➜ 2-Related Samples

1. Move two scale variables into the first pairs in the Test Variables Box (Note: a second pair will appear automatically and can be used or ignored.)
2. <u>Options</u>: click on Descriptives. **Continue** **OK**

Kruskal-Wallis One-Way ANOVA

Analyze ➜ Nonparametric Tests ➜ Legacy Dialogs ➜ K-Independent Samples

1. Move a scale variable into the Test Variables Box
2. Move a categorical variable into the Grouping Variable box.
3. Click on "Define Groups" and enter the range of the groups
4. <u>Options</u>: click on Descriptives. **Continue** **OK**

Friedman Two-Way ANOVA

Analyze ➜ Nonparametric Tests ➜ Legacy Dialogs ➜ K-Related Samples

1. Move a scale variable into the Test Variables Box
2. Move 3 or more categorical variable into the Test Variable box.
3. <u>Statistics</u>: click on Descriptive. **Continue** **OK**

Making the Decision

Step 5: Decide Whether or Not to Reject the Null (H_0)

By Todd Daniel Ph.D.

Compare the Obtained value to the Critical value

Obtained Value = this is the number that you calculated in your statistical test. It is the test result and will be *z* (for *z* tests), *t* (for *t* tests), *F* (for ANOVA), X^2 (for Chi-square tests) or another letter for other, less frequently used tests.

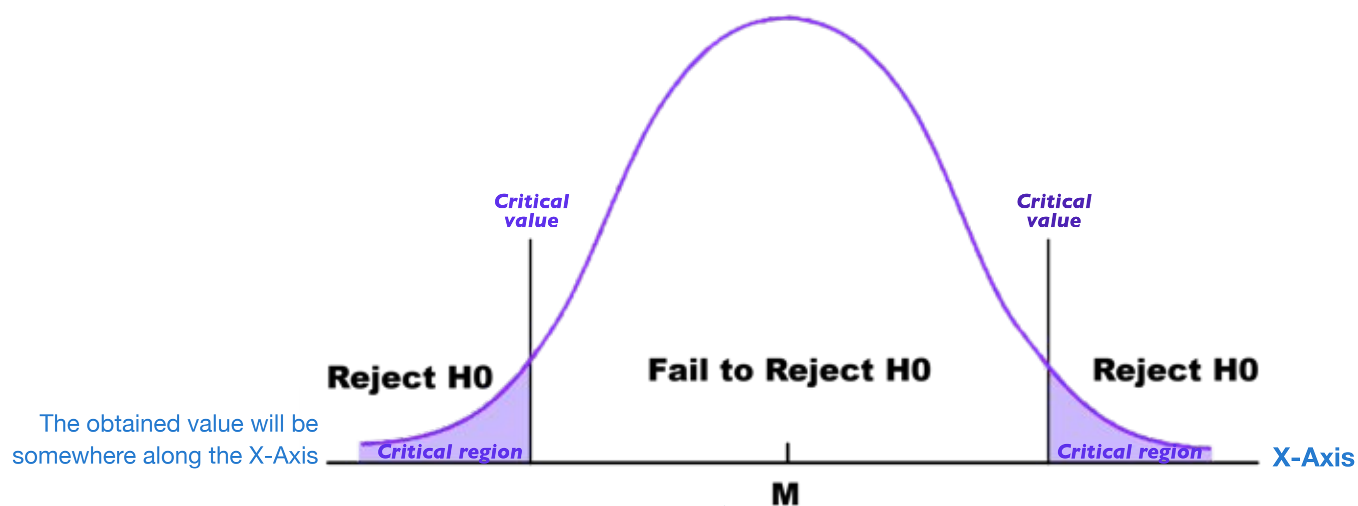
Critical Value = this is the value that you looked up in the table in Step 3. SPSS will not provide a critical value in the output, but will compute a *p* value. In SPSS, the *p* value is in a box labeled "Sig."

Rules for Determining and Reporting Significance

- If your obtained (calculated) value exceeds your critical value, then your test ***is significant***: reject the null, $p < .05$
  - For one-tailed tests, the obtained value must exceed the critical value in the one direction you are testing.
- If your obtained value is not in the critical region, your test is ***not significant***: "accept" the null, $p > .05$, *ns*

  We never really "accept" the null, we just <u>fail to reject</u> it. It is highly unlikely that we would have found these results if the null hypothesis was true.

Sig (2-tailed)	Results are Significant	Results are Not Significant
p value given in SPSS	$p = .042$	$p = .45$, *ns*
p value = .000 in SPSS	$p < .001$	*impossible*
No *p* value (c.f. *doing a test by hand; doing one-tailed test*)	$p < .05$	$p > .05$



Reject
Reject the null hypothesis
Use the alternative hypothesis
Results are <u>in</u> the critical region
Results are significant
There was a significant difference
There was an effect
The treatment worked
$p < .05$

Accept (fail to reject the null)
Fail to reject the null hypothesis
No support for the alt. hypothesis
Results are <u>not in</u> the critical region
Results are not significant
No significant difference
There was not an effect
The treatment made no difference
$p > .05$

Reject
Reject the null hypothesis
Use the alternative hypothesis
Results are <u>in</u> the critical region
Results are significant
There was a significant difference
There was an effect
The treatment worked
$p < .05$

Reporting Your Results

Using proper APA style

By Todd Daniel Ph.D.

Six Steps to Writing the Write-Up in APA Style

When reporting your results, you should begin with a statement about why the test was performed then describe the results. Finally, include the statistics from the test you conducted.

1. **Test**: what is the name of the test you used? (*t, ANOVA, chi-square, correlation, regression*)
2. **Variables**: what variables were you measuring (*IV & DV*) to test what research question?
3. **Measurement**: how were the variables measured/quantified?
4. **Significance**: was the finding significant?
5. **Statistics**: written in APA style, see **General Guidelines for Writing Up Statistics in APA Style** below
6. **Summary**: what do the results indicate and what are the implications of the findings?

- A **(1)** single sample t-test was conducted **(2)** to determine whether students in a statistics class scored higher on the final exam than the class last year. The **(3)** average scores on the final (M = 76.4, SD = 9.34) were were **(4)** significantly higher than last year's class final average (M = 74.5), **(5)** $t(29) = 3.14$, $p = .032$. This indicates that **(6)** the students this year performed significantly better on the final than the class last year.
- A **(1)** one-way ANOVA was conducted **(2)** to test which type of treated wood has superior durability for outdoor decking. Samples of three types of treated wood **(3)**, CCA (M = 1.13), MCP (M = 1.22), and ACQ (M = 1.19) were exposed to artificial weather testing but **(4)** no significant differences were found between their performances **(5)**, $F(2, 298) = 1.42$, *ns*. These result show that **(6)** the three types of wood available for outdoor decking have approximately equal durability.

General Guidelines for Writing Up Statistics in APA Style

Use italics for statistical symbols, but use standard typeface (no bold or italics) when writing (a) Greek letters, (b) subscripts that function as identifiers, and (c) abbreviations that are not variables.

t, F, r, N, n, M, SD, p, ns but $\alpha = .05$, H_0: $\mu = 28$, PTSD

Use parentheses to enclose statistical values.

...was statistically significant (p = .042) for all variables.

Use parentheses to enclose degrees of freedom.

$t(45) = 4.35$

$F(3, 87) = 2.11$

Use brackets to enclose limits of confidence intervals.

95% CIs [3.45, 2.7], [-6.0, 3.89], and [1.89, 7.23]

Use an italicized, uppercase *N* in reference to the total number of subjects or participants in the sample.

N = 328 people participated in the study

Use an italicized, lowercase *n* in reference to only a portion of the sample.

n = 42 of participants were Pacific Islanders

References

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Kahn, J. (2005). Reporting statistics in APA style. Retrieved from http://my.ilstu.edu/~jhkahn/apastats.html Used by permission.

Levene's Test

Testing the assumption of homogeneity of variance

By Todd Daniel Ph.D.

Assumptions matter. If your assumptions are wrong, then the conclusions based on those assumptions will also likely be wrong. Each test we use in statistics is built on certain assumptions about the data. It is important that we make sure that those assumptions have been met before we interpret the hypothesis test. The independent samples *t* test assumes that the groups being compared have approximately equal variance. It would not be valid to compare a very leptokurtic distribution to a very platykurtic distribution. The word *homogeneity* ("of the same nature") describes similarity (approximate equality) between groups.

Levene's Test is a test of homogeneity of variance. It is run automatically by SPSS when you are comparing independent groups. Like any other hypothesis test, Levene's Test begins with the assumption that the groups are equal (no difference). This assumption, in fact, is what you want, since the *t* test assumes (and you would prefer) that the groups have equal (not different) variance.

If Levene's Test **is not** significant (Sig. Value is > .05) then the assumption of homogeneity has been met and you are safe to assume that the groups you are comparing have *equal* variance. Good news. If, however, the Levene's Test **is** significant (Sig. < .05), then the groups have *significantly different* amounts of variance. You cannot assume that the groups have equal variance. (Levene's Test is always a 2-tailed test).

All is not lost, however, when Levene's Test is significant. You can still use the parametric test, but you must interpret the test using the data in the line marked "Equal variances not assumed."

Researchers typically prefer to find significant differences in the *means* between groups, but not in the *variances.* Levene's test measures whether the ***variances*** of the groups are equal; the *t* test tests whether the ***means*** of the groups are equal. Never interpret Levene's Test as a means test; it is not a *t* test.

Independent Samples Test - Levene's Test NOT Significant

	Levene's Test for Equality of Variances		t-test for Equality of Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Equal variances assumed	.503	.490	2.103	50	0.046	10.250
Equal variances not assumed			~~1.998~~	~~47.454~~	~~0.054~~	~~10.250~~

(2) Interpret the t test using top line

(1) Levene's test is not significant (.49 > .05); you can assume that the variances are equal

Independent Samples Test - Levene's Test Significant

	Levene's Test for Equality of Variances		t-test for Equality of Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Equal variances assumed	2.521	0.022	~~2.103~~	~~50~~	~~0.046~~	~~10.250~~
Equal variances not assumed			1.998	47.454	0.054	10.250

(4) Interpret the t test using bottom line

(3) Levene's test is significant (.022 < .05); you cannot assume the variances are equal

Type I and Type II Statistical Errors

Errors in hypothesis testing vs. pregnancy testing

By Todd Daniel Ph.D.

When you make a decision to reject the null hypothesis or not, you could be wrong. The possible errors in hypothesis testing are the same as if you were conducting a pregnancy test.

Pregnancy Test
Assumption is that you are NOT pregnant

		True Condition	
		Not Pregnant	Pregnant
Result of Test	Negative (not pregnant)	CORRECT RESULT AVOID AN ERROR	TYPE II ERROR **False Negative** ***Pregnant But Don't Know It***
	Positive (pregnant)	TYPE I ERROR **False Positive** ***Not Really Pregnant After All***	CORRECT RESULT AVOID AN ERROR

Type I Error: rejecting a null hypothesis that is actually true (false positive)

Concluding that a treatment has an effect when it does not. Type I errors are only a concern when you reject the null hypothesis. The probability of a Type I error is alpha (α) (typically .05 or 5%).

To decrease the chance of a Type I error, you can change the alpha level from .05 to .01; however, this increases the possibility of a Type II error.

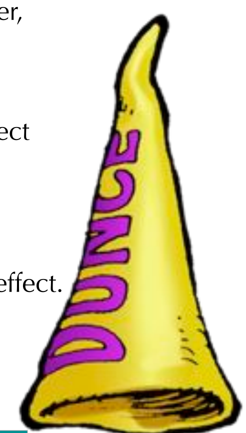
Type II Error: "accepting" a null hypothesis that is actually false (false negative)

Failure to detect an effect that actually exists. Type II errors are only a concern when you fail to reject the null hypothesis. The probability of a Type II error is beta (β).

Type II errors are most likely to happen:

- When **effect size is small** – you are looking for a very tiny effect and it is easy to miss; or,
- When the **n is small** – you do not have a large enough sample size to adequately detect the effect.

To decrease the chance of a Type II error, increase power by increasing sample size (*n*).



Memory Aids: Type **I** is like Pinocchio's nose (*lying*); Type **II** is like a dunce cap (*missed the effect*)

Errors are always written with a Roman numeral (*Type I and Type II, not Type One or Type 2*).

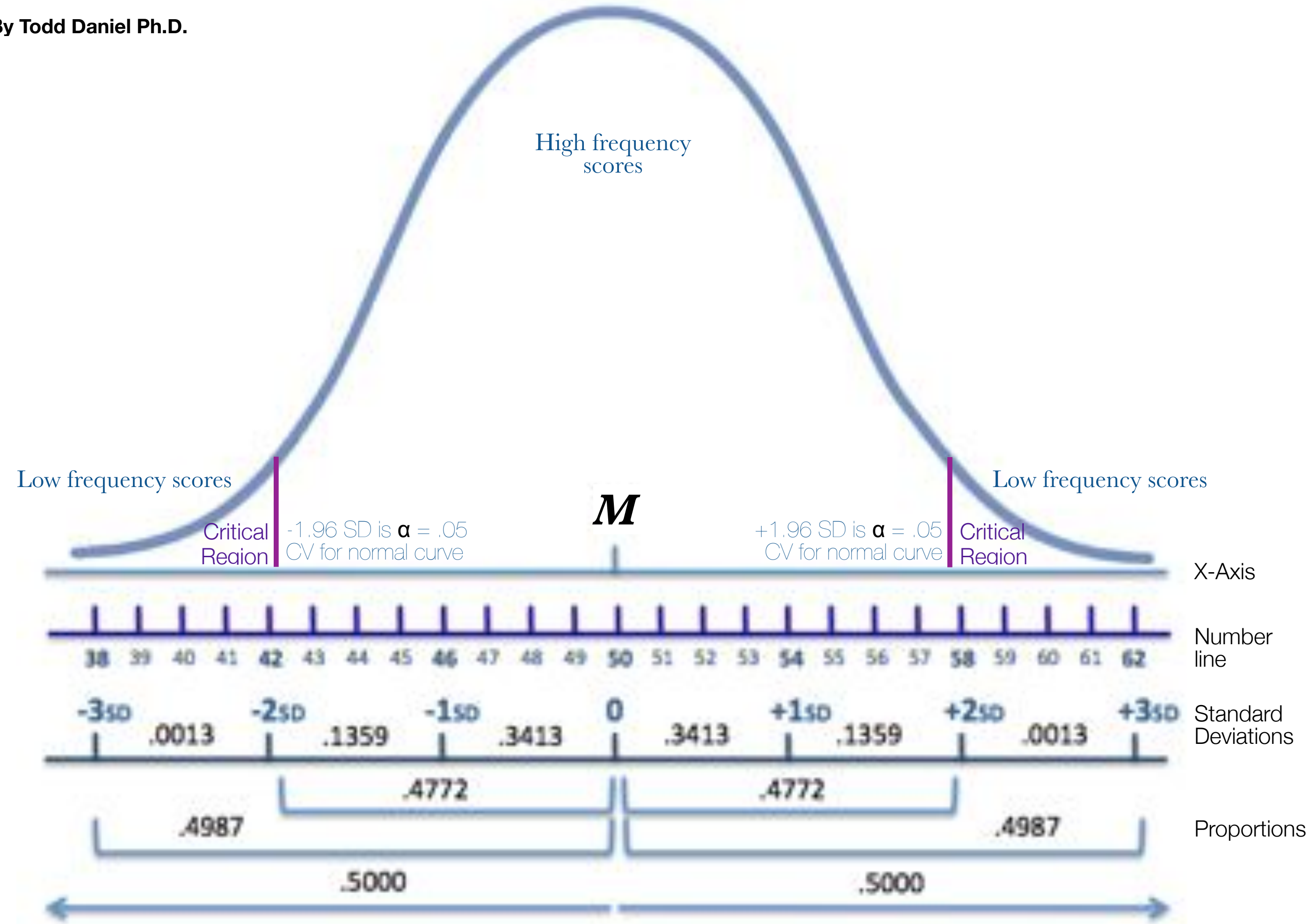
Statistical Errors
Null Hypothesis is that there is NO effect

		True State of the World	
		H0 TRUE (Type I situation)	H0 FALSE (Type II situation)
Sample Outcome	ACCEPT H0	CORRECT DECISION AVOID A TYPE I ☺ ($P = 1 - \alpha$)	TYPE II ERROR ☹ ($P = \beta$)
	REJECT H0	TYPE I ERROR ☹ (EQUALS α)	CORRECT DECISION AVOID A TYPE II ☺ ($P = 1 - \beta$) a.k.a. "power"

The Normal Curve

Mean, Standard Deviation, Raw Data

By Todd Daniel Ph.D.



The Gaussian (Bell) Curve is a theoretical or hypothetical distribution of scores.

Most scores fall in the middle so mean = median = mode

Symmetrical - Each half of the curve is a mirror image of the other. It is perfectly symmetrical.

It has no skew and no kurtosis

Asymptotic - no matter how far tails go out, they never touch the X-axis (symptote)

Frequency - The normal curve is a frequency polygon.

Frequency of x values decreases the farther you move away from the mean (in either direction)

Scores with greater frequency are found in the middle; scores with less frequency are in the tails

Standardization - The curve has a constant, direct relationship with the standard deviation

Each raw score (number) corresponds to a standard deviation (or a fraction of a standard deviation)

A constant percentage of scores will fall under any area of bell curve

When marked in standard deviation units, it becomes a *standard normal curve*

***Z* Scores and the Normal Curve** - A z score distribution is in the shape of a normal curve

A z score distribution has a mean of 0 and a standard deviation of 1.00

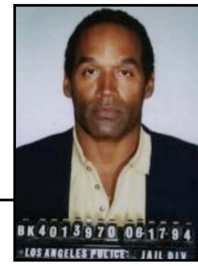
In a normal distribution, the z scores are seldom greater than plus-or-minus 3 standard deviations

Convert a proportion to a decimal: multiply by 100 or move the decimal point right two places: p = .3414 – > 34.13%

Hypothesis Test v. Jury Trial

The theory behind how quantitative research gets done

By Todd Daniel Ph.D.



We want to know if rats who use cocaine run mazes faster than rats who drink alcohol. So we give some rats coke and other rats rum-and-Coke. We measure how quickly each group of rats runs in a maze. But is one group significantly faster? Understanding the logic behind hypothesis testing is easier if you compare it to a jury trial. The data are put "on trial" to see if there is enough evidence to prove the treatment had an effect.

Hypothesis Testing	A Jury Trial
Null Hypothesis: We assume that there is no treatment effect unless there is enough evidence to prove otherwise. H_0: $rats_1 = rats_2$ (*no difference in group means*)	**Presumption of Innocence:** The jury is instructed to assume the defendant is innocent until proven guilty.
The alpha level: We are confident that the treatment does have an effect because it is unlikely the data could occur simply by chance. $\alpha = .05$, two-tail test (*difference of 12 points*)	**Standard of Proof:** The jury must be convinced beyond a reasonable doubt ("*the standard*") before they find a person guilty.
The sample data: The research study is conducted to gather data (*evidence*) to demonstrate that the treatment had an effect. Run the rats through the maze and collect data	**Evidence:** The prosecutor presents evidence to demonstrate that the defendant is guilty.
The critical region: The sample data fall in the critical region (*meet the burden of proof*) or they fall outside the critical region (*there is not enough evidence to reject the null*) A difference of **more than 12** points will be significantly different	**Deliberation:** Either there is enough evidence to meet the burden of proof (*reasonable doubt or preponderance of evidence*) and convince the jury that the defendant is guilty, or there is not.
Conclusion: If the sample data (*evidence*) fall into the critical region (*standard of proof*) we reject the null (*convict*). If the sample data do not fall into the critical region, we fail to reject the null (*acquit*). The means are **15** points apart; they are statistically significantly different	**Verdict:** If the evidence exceeds the burden of proof, the jury votes "guilty". The verdict *declares* the defendant guilty but does not *prove* the defendant is guilty. If there is not enough evidence, the verdict is "not guilty." Being acquitted does not prove the defendant is innocent, either.

10/16