Chapter 9
Testing Hypotheses

- Overview
- The null-hypothese
- One and two-tailed tests
- 5 Steps for testing hypotheses
- Type 1 and Type 2 Errors
- Z tests and t tests

Testing Hypotheses (and Null Hypotheses)

Testing Hypotheses is a procedure that allows us to evaluate hypotheses that are typically drawn from a theory and based on sample statistics.

Example of a theory: UNT sociology majors are brighter than the average UNT student.

Example of an hypothesis: UNT Sociology majors have a higher GPA than the UNT student population.

Example of a theory: Commitment to an organization (Organizational Commitment) is affected by the characteristics of the organization.

Example of an hypothesis: Organizational commitment is affected by the procedures used to do the work.

Example of a null-hypothesis:

Organizational commitment is not affected by the procedures used to do the work.

Research Hypothesis

A research hypothesis is a statement typically reflecting a relationship between two variables that can be statistically tested.

By examining the truth of an hypothesis, we are able to draw conclusions about the broader theory from which the hypothesis was derived.

That is, if the hypothesis is found to be true, this lends support for the theory from which the hypothesis was drawn. If the hypothesis is found to be false, this provides evidence that the theory is false.

Null Hypothesis

A null hypothesis is a statement of "no difference." That is, rather than suggesting a relationship exists between two variables, the null hypothesis suggests there is "no relationship" between the variables.

Example of null hypothesis: Organizational commitment is not related to the procedures used to do the work.

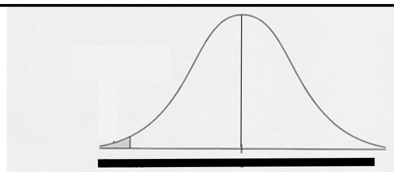
If the null hypothesis is found to be false (i.e., we reject the null hypothesis) then we have support for the research hypothesis and the broader theory. Statisticians typically test the null hypothesis rather than the research hypothesis. The reasons are statistically based.

One-Tailed Tests

One-tailed hypothesis test – A hypothesis to be tested where the sample statistic is believed to be either higher or lower when compared to another group.

Example of one-tailed test: The average family income of Native Americans is less than that of the average U.S. family.

In this example the sample statistic is "average family income of Native Americans" and the comparison group is the average income of a family in the U.S.



Av. Income for Population of Native Am. — Av. U.S. Income

It is one-tailed since we are predicting that the income of Native Americans falls to the left of the income of the U.S. population. Because it falls to the left, it is referred to as a “left-tailed test”.

What would be the null hypothesis?

One-Tailed Tests

- **Right-tailed test** – A one-tailed test in which the sample outcome is hypothesized to be at the right tail of the sampling distribution. (e.g., Native Americans have higher average family incomes than that of the U.S.)
- **Left-tailed test** – A one-tailed test in which the sample outcome is hypothesized to be at the left tail of the sampling distribution. (e.g., Native Americans have lower family incomes than that of the U.S.)

Two-Tailed Tests

Two-tailed hypothesis test – A hypothesis test in which a sample statistic might fall within either tail of the sampling distribution.

We are not sure which tail of the curve the statistic is likely to fall.

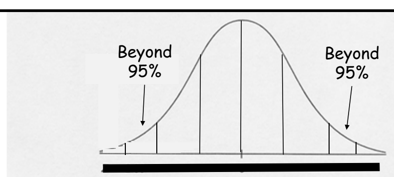
Example of two-tailed test: The average family income of Native Ams. is not the same as that of the average U.S. family (we haven’t specified greater or lesser than)—what would be the null hypothesis?

Conducting a one-tailed test

The null hypothesis is typically tested:
for example, we might test the null hypothesis: there is no difference between the average GPA score of sociology majors and that of all UNT students.

(fictitious data:)
GPA for a sample of sociology majors = 2.9
Sample size (N) = 50
Standard Error = .2
GPA for all UNT students = 2.6

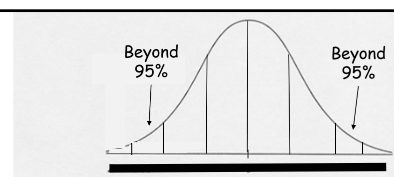
In class: plot the information above on a normal curve. Use the GPA for all UNT students as the mean of the normal curve. Show Z scores and raw scores for three standard errors.



Z scores	-3	-2	-1	0	+1	+2	+3
Raw scores				2.6	2.8	3.0	3.2

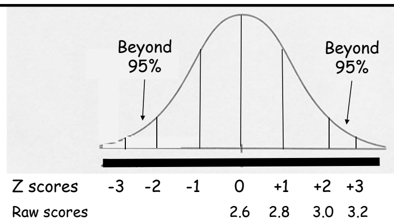
The null hypothesis assumes no difference. If the mean GPA score for sociology majors falls within our 95% confidence interval (plus or minus 2 SEs), then:

we can conclude, with 95% confidence, that our sample statistic is “no different” than the comparison group (that is, the average GPA for soci majors is no different than that of all UNT students). The difference observed (2.9 vs 2.6) is simply due to sampling error.

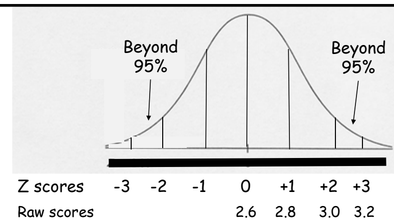


Z scores	-3	-2	-1	0	+1	+2	+3
Raw scores				2.6	2.8	3.0	3.2

In other words, we have used our sample statistics and the average GPA score of all UNT students to create a confidence interval where we can be 95% confident that, if the average sociology GPA falls within the confidence interval, then any difference found between our sample sociology GPA (2.9) and the GPA for all UNT students (2.6) is due to sampling error and not due to a real difference between the two groups.

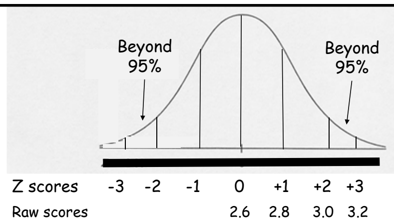


On the other hand, if the sample statistic falls beyond the 95% range, then we conclude that we are 95% confident that the difference we found (2.9 vs 2.6) IS due to a real difference between the two groups.



What statisticians usually say is:

the chance that there is "no difference" between the two groups is 5% or less. If the probability of "no difference" is very small, we reject the null hypothesis of no difference.



<u>In class exercise</u>: In our example, do we reject or accept the null hypothesis? Explain your answer.

If the mean GPA score for sociology majors (2.9) falls within our 95% confidence interval (plus or minus 2 SEs), then we will accept the null hypothesis. And, we will assume the difference we find is due to sampling error.

The Five Steps In Hypothesis Testing: #1

1. Make sure the data meet the assumptions for hypothesis testing
   --a random sample is being used
   --either the population is normally distributed or the sample taken from the population has over 50 cases (this will allow us to apply the Central Limit Theorem)
   --knowing the level of measurement of the data.

The Five Steps In Hypotheses Testing: #2

2. State the (a) research hypothesis, (b) the null hypotheses and select (c) alpha.

   2a. What is a <u>research hypothesis</u> (H_1) –

   A statement typically reflecting a relationship between two variables that can be statistically tested.

   It is typically drawn from a theory. If the research hypothesis is supported by the data then this supports the broader theory.

2b. What is a <u>null hypothesis</u> (Ho) –

A statement of "no difference," which contradicts the research hypothesis.

Example: The average GPA of UNT sociology majors is no different than that for all UNT students.

If we reject the null hypothesis, this provides support for the research hypothesis.

2c. What is an alpha?

Alpha is the probability level we have chosen at which the null hypothesis will be rejected.

In our example above we selected an "alpha" of .05. At .05 alpha, there is only a 5% probability that the null hypothesis is true, i.e., that there is no difference between the sample statistic and comparison group. If the probability is at or smaller than alpha (in this case .05), we will reject the null hypothesis of no difference and assume that the difference found is a true difference (not simply due to sampling error).

In other words, we ask the question: If there is actually no difference between the two groups, what is the probability that we would have randomly selected a sample with a statistic (e.g., average sociology GPA) this much larger or smaller than that of the comparison parameter (e.g.,whole UNT student population)?

We then calculate the probability of getting the difference we found (assuming no difference). If we find that there is very little chance (5% or less) of drawing a sample with this much difference, then we will conclude that the difference is a real difference and not due to sampling error. We will reject the null hypothesis and accept the research hypothesis that the two groups are different.

The Five Steps In Hypotheses Testing: #3

3. Select the sampling distribution and specify the test statistic.

Our sampling distribution will be either the Z distribution or t distribution. We will use either the Z or t distribution to assist us in testing our null hypothesis (so far we have used only the Z distribution—where we have used the normal curve).

Our test statistic will be either the Z statistic or the t statistic. We will use either the Z or t statistic to assist us in testing our null hypothesis (so far we have used only Z statistics such as 1.96, and 2.58)

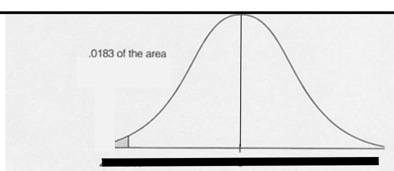
The Five Steps In Hypotheses Testing: #4

4. Calculate the test statistic, e.g., convert the sample mean to a Z statistic or t statistic.

Let's use an example to calculate a Z statistic: If we assume that our sample mean family income for Native Americans is $22,400 and that our mean family income for the population as a whole is $28,985:

then we would ask:

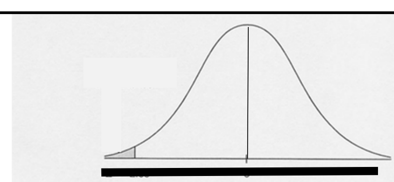
What are the chances that we would have randomly selected a sample of Native Americans with an average family income of $22,400 if there is really no difference between their income and that of the population as a whole ($28,985)?



$22,400 $28,985

We can determine the chances or probability because of what we know about the sampling distribution and subsequently the normal curve.

We begin by assuming the two groups are NOT different (null hypothesis). If we took 100 samples, each of the sample statistics (such as the means) is expected to be located somewhere around the statistic of the comparison group (28K). If our one sample statistic is not located somewhere around the comparison group statistic, then we will assume the difference found is a real difference and not due to sampling error.



$22,400 $28,985

We can take our sample statistic (we'll use the mean), convert it into a Z score, and find the Z score on the normal curve. Typically, if the sample statistic (in this case mean) is two standard errors or further from the mean, we will reject the null hypothesis. That is, if the **"p value" of the sample statistic** is .05 or less, we will reject the null hypothesis (in some cases we may decide to use three standard errors as the level at which to reject the null hypothesis in which case the p value of the sample statistic must be .01 or less).

So, lets suppose:

The mean family income for our sample of Native Americans = $22,400 with a sample size of 100

The mean family income for U.S. population = $28,985
The standard deviation for U.S. population = $7,345

We want to test the null hypothesis that Native American income is no different than the U.S. population income.

$$Z = \frac{\bar{Y} - u_y}{\frac{\sigma_{\bar{y}}}{\sqrt{N}}} \quad \text{or} \quad \frac{\text{Group Statistic} - \text{Population Parameter}}{\frac{\text{Population SD}}{\sqrt{N}}}$$

The group statistic is the mean family income for our sample of Native Americans = $22,400 and the sample size (N) is 100.

The mean family income for U.S. population = $28,985
The standard deviation for U.S. population = $7,345

In-class Exercise: What is the Z?

$$Z = \frac{\bar{Y} - u_y}{\frac{\sigma_{\bar{y}}}{\sqrt{N}}} \quad \text{or} \quad \frac{\text{Group Mean} - \text{Population Mean}}{\frac{\text{Population SD}}{\sqrt{N}}}$$

$$\text{Standard Error} = \frac{\sigma_{\bar{y}}}{\sqrt{N}} = \$7{,}345/10 = \$734.5$$

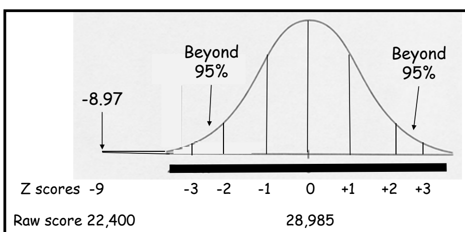
$$Z = \frac{22{,}400 - 28{,}985}{734.5} = \frac{-6{,}585.0}{734.5} = -8.97$$

A Z statistic of -1.96 would tell us that there is a .05 (5%) chance of getting this difference ($22,400 vs $28,985), when there is actually no difference. That is, there is a 5% chance that the difference found is due to sampling error.

A Z statistic of -2.58 would tell us that there is a .01 (1%) chance (or probability) of getting this difference, when there is actually no difference.

A Z statistic of -3.27 would tell us that there is a .001 (.1%) probability of getting this difference, when there is actually no difference.

.05, .01, and .001 are called alpha levels when they are selected as the cut off point for significance.



Our Z statistic is -8.97. This is far beyond the scientifically accepted “cut off” of a Z of -1.96. This tells us that the probability of getting this difference, when there is actually no difference, is extremely small. Because it is beyond 1.96, we will reject the null hypothesis and assume that this is a real difference in income between the average Native American family and the average family in the U.S.

The Five Steps In Hypotheses Testing: #5

5. Making a decision and interpreting the results

5a. Making a decision: Again, if our Z statistic has a p value (probability value) equal to or smaller than the alpha level we selected (e.g., .05 or .01, or .001), then we can reject the null hypothesis.

Since the alpha is typically set at .05 in the social sciences and our P value is much smaller than .05 (as indicated by the Z score), we will reject our null hypothesis.

5b. Interpreting the results:

The "Z statistic" of -8.97 is statistically significant, that is, it is very unlikely that the difference we found occurred by chance or sampling error.

We can say that the difference between the family income of Native Americans and the U.S. population is significant beyond the .001 level.

We can also say that the average family income of Native Americans was substantially less than the average income of families in the U.S. when the survey was taken. We say "substantially" because there is a $6,400 difference and we consider this to be a very big (substantive) difference. Of course, this is only our opinion of the difference.

Summary: Steps in Testing an Hypothesis

1. Verify that assumptions are met
2. State research and null hypotheses and alpha (our example hypothesized that the average income of Native Americans was less than the U.S. population as a whole)
3. Select sampling distribution and test statistic to be used (Z or t statistic)
4. Compute test statistic
5. Make a decision and interpret results

Errors that we try to avoid

Type I error: if the null hypothesis is rejected when it is actually true. (this is the most common error because researchers want to reject the null hypotheses so they can accept their hypotheses).

Type II error: if the null hypothesis is accepted when it is actually false.

Type I and Type II Errors and their relationship to alpha

- When selecting our alpha, we need to be aware that if we set **alpha too large** (e.g. p value of.10) we may create a **Type I** error—that is, we might reject the null hypothesis when it is actually true.
- Or, if we set the **alpha too small** (e.g., .001) we may create a **Type II** error by failing to reject a false null hypothesis.

Let's take another example:

--Our research hypothesis is that the salary for women is less than that for the U.S. population.

--We will test the null hypothesis that the salary for women is no different than that for the U.S. population as a whole. (Is this a one-tailed or two tailed test?)

--We sample at least 50 women so that our theoretical sampling distribution will be normally distributed (a required assumption).

--The test statistic used is either the **Z statistic or t statistic** and since we have the population standard deviation we will use the Z statistic.

In-class: compute the test statistic.
The formula for the Z statistic is:

$$Z = \frac{\overline{Y} - u_y}{\frac{\sigma_y}{\sqrt{N}}} \quad \textbf{or} \quad \frac{\text{Sample Mean} - \text{Population Mean}}{\frac{\text{Population SD}}{\sqrt{N}}}$$

Where the population mean is $28,985
the sample mean for women is $24,100
the population standard deviation of 23,335
and the sample size is 100

Computing the test statistic. The formula for the Z statistic is:

$$Z = \frac{\overline{Y} - u_y}{\frac{\sigma_y}{\sqrt{N}}} \quad \text{or} \quad \frac{24{,}100 - 28{,}985}{\frac{23{,}335}{\sqrt{100}}} = -2.09$$

Where the population mean is $28,985
the sample mean for women is $24,100
the population standard deviation of 23,335
and the sample size is 100

Make a Decision and Interpret the results. In our example:

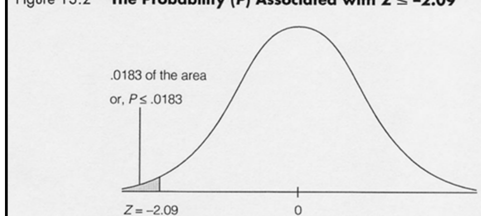
--we confirm that the Z is on the left tail of the distribution (-2.09)

--the Z score is smaller than our selected alpha of .05. We know this because a Z score of -1.96 is at the .05 level and our Z score of -2.09 is further out than -1.96.

--thus, we can reject the null hypothesis of no difference and can conclude that the average income for women is less than that of the general population.

Probability Values

Figure 13.2 **The Probability (P) Associated with $Z \leq -2.09$**



$24,100 — Women

$28,985 — U.S. Population

When to use the t statistic and when to use the Z statistic

1. The Z statistic can only be used if the **population standard deviation** is known. Typically, this is not the case.
2. When the **sample standard deviation** must be used in lieu of the population SD then the t statistic should be used.
3. The formula for the t statistic is identical to the formula for the Z statistic except that **the sample SD is used in place of the population SD**

Computing the t statistic. The formula for the t statistic when <u>comparing a group to the population</u> (there is a different formula for the t statistic when comparing two groups)

$$t = \frac{\overline{Y} - u_y}{\frac{S_y}{\sqrt{N}}} \quad \textbf{or} \quad \frac{\text{Sample Statistic} - \text{Population Parameter}}{\frac{\text{Sample SD}}{\sqrt{N}}}$$

Steps for interpreting the t statistic

Unfortunately, locating where the t statistic falls on the normal curve is not as easy as when using the Z statistic (that is, unlike Z values, you cannot use the normal curve when using t values).

Once the t statistic is calculated it is compared to the t value needed to reject the null hypothesis. The t value needed can be found on a t distribution table (page 484-5 in textbook) and will vary depending on whether the researcher has chosen an alpha of .05, .01, etc.

How to use the t distribution table to determine significance

(1) Determine the **degrees of freedom** your sample provides (this is typically: N-1) and then locate the DF on the t-distribution table (table is on page 484-5) .

(2) Find on the table: the **alpha** which you selected at the start of the statistical analysis (an alpha of .05 and a two-tailed test are typically used by researchers)

(3) **Find the intersecting point** where the DF and the alpha cross. At the intersecting point you will find the t value needed to reject the null hypothesis.

t distribution table

Table 13.2 **Values of the *t* Distribution**

	Level of Significance for One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test					
df	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
10	1.372	1.812	2.228	2.764	3.169	4.587
15	1.341	1.753	2.131	2.602	2.947	4.073
20	1.325	1.725	2.086	2.528	2.845	3.850
25	1.316	1.708	2.060	2.485	2.787	3.725
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Source: Abridged from R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Table 111. Copyright © R. A. Fisher and F. Yates 1963. Reprinted by permission of Pearson Education Limited.

(4) **Compare** the t value calculated from the data to the t value identified on the t distribution table. If the calculated t value is <u>larger</u> than the t value found in the table, then the null hypothesis can be rejected and the difference between the group and the population can be considered statistically significant (but not necessarily "substantively" significant).

In-Class Exercise:

The population mean income is $28,985 and the sample mean income for women is $24,100 with a sample standard deviation of 24,897 and sample size of 100.

$$t = \frac{\overline{Y} - u_y}{\frac{S_y}{\sqrt{N}}} \quad \textbf{or} \quad \frac{\text{Group Mean – Population Mean}}{\frac{\text{Sample SD}}{\sqrt{N}}}$$

The population mean is $28,985 and the sample mean for women is $24,100 with a sample standard deviation of 24,897 and sample size of 100.

$$\frac{24{,}100-28{,}985}{\frac{24{,}897}{\sqrt{100}}} = \frac{-4{,}885}{2{,}489.7} = -1.96$$

Finding the t statistic in the t distribution table

Our **degrees of freedom** for this example is N – 1 or 99 and our t statistic is -1.96 (the larger the t statistic the more likely it will be significant).

On page 484-5 of your book we can find the **t distribution table**. It displays the **degrees of freedom** for 60 and for 120. Since ours is 99 it is less than 120. Therefore, to be conservative we will use 60 DF.

We can assume a **one-tailed test** since existing knowledge indicates that women make less than the population as a whole and certainly not more (the mean will fall on the left side of the curve).

t distribution table

Table 13.2 **Values of the *t* Distribution**

	Level of Significance for One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test					
df	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
10	1.372	1.812	2.228	2.764	3.169	4.587
15	1.341	1.753	2.131	2.602	2.947	4.073
20	1.325	1.725	2.086	2.528	2.845	3.850
25	1.316	1.708	2.060	2.485	2.787	3.725
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Source: Abridged from R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Table 111. Copyright © R. A. Fisher and F. Yates 1963. Reprinted by permission of Pearson Education Limited.

Determining Statistical Significance

Since our t statistic is -1.96 we can conclude statistical significance at the .05 level.

Would our findings be significant if we had chosen an alpha of .01?

Would our findings be significant using a 2-tailed test (women's income is different from the general population)?

Comparing the Sample Statistics of Two Groups

(Presented above is a comparison of a group's sample statistic to a population parameter)

Example for comparing two groups:

Comparing the mean salary of new sociology professors (group 1) to the mean salary for new engineering professors (group 2). Previously we were comparing the group statistic (such as the mean salary of sociology professors) to the population parameter (such as the mean salary of the whole U.S. population).

Steps for comparing the sample statistics of two groups are the same as that for comparing a sample statistic to the population parameter with three exceptions:

(1) the formula for calculating the t statistic is different

(2) calculating the degrees of freedom is different, and

(3) must now determine whether the two groups have equal or unequal variances. (If the Levene's test is significant then their variances are unequal)

Formula for calculating the t statistic for comparing two groups:

(You will not be required to calculate this comparison because the formula for determining the Standard Error of the Differences Between the Means is complex. We will use the computer to do the comparison and then we will determine whether the null hypothesis can be rejected.)

$$t = \frac{\text{Mean of } 1^{st} \text{ Group} - \text{Mean of } 2^{nd} \text{ Group}}{\text{Standard Error of the Differences Between the Means}}$$

Step 4: Calculation of the t Statistic

$$t = \frac{\text{Mean of 1}^{st}\text{ Group} - \text{Mean of 2}^{nd}\text{ Group}}{\textbf{Standard Error}\text{ of the Differences Between the Means}}$$

$$\textbf{SE} = \sqrt{\frac{(N_1-1)\,SD^2y_1 \;-\; (N_2-1)\,SD^2y_2}{(N_1 + N_2) - 2}}\;\sqrt{\frac{N_1+N_2}{N_1N_2}}$$

The sample findings were:

Males: mean = grade 4.39; SD=1.70; N=380

Females: mean = grade 4.61; SD=1.62; N=394

IMPORTANT You will not be asked to calculate the t statistic for comparing two groups. However, you are expected to use the computer to calculate the t statistic.

Example: Comparing Two Sample Means (male and female job burnout)

In SPSS:

1. Analyze
2. compare means
3. independent sample t test
4. move V101 (sex) to "group variable" box
5. click "define group"
6. in group 1 put "1" (female) and in group 2 put "2" (male)
7. click continue
8. move "burnout", V102 (age), V100 (education), V114 (# residents assigned) to "test variables" box and then click Okay

凩訝

(see you later)