

## Hypothesis Testing

Below is a sample data set that we will be using for today's exercise. It lists the heights for 10 men and 12 women collected at Truman State University. The data will be entered in the same way that it was in the **SPSS introduction** handout. In addition to the "Height" and "Gender" columns, we will add another column "Hfemale". Click on the variable view tab and name the new column. Then return to the data view spreadsheet and enter the 12 heights for the females in this new column.

Male	Female
69	65
70	61
65	67
72	65
76	70
70	62
70	63
66	60
68	66
73	66
	65
	64

### Analyzing data

We commonly are interested in comparing the mean of the population to some known standard. We do this by creating a hypothesis, sampling from the population, calculating a test statistic from the sample, and then determining if the sample could reasonable come from the population if the null hypothesis was correct. Below are several examples illustrating various hypothesis tests.

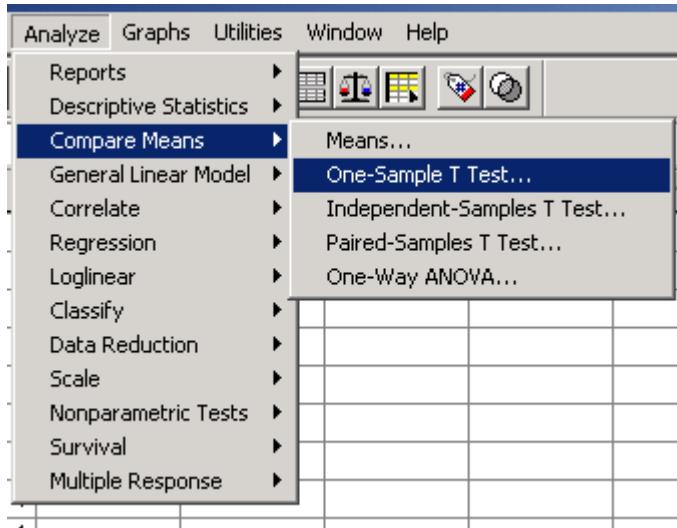
#### 2-tailed Hypothesis test

A researcher is interested in knowing if the average height for women at Truman State University is equal to 62.5 inches. The appropriate hypothesis are

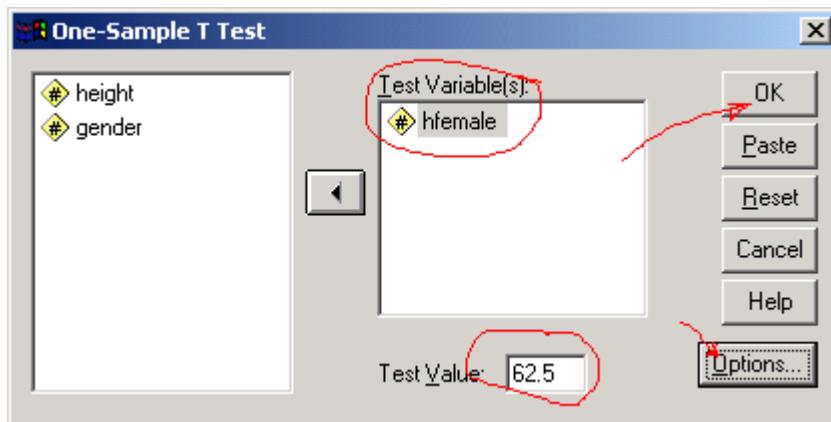
$$H_o : \mu_F = 62.5 \text{ inches}$$

$$H_A : \mu_F \neq 62.5 \text{ inches}$$

The appropriate test for this hypothesis is the one sample t-test. To run this test, click on **Analyze-Compare Means-One Sample T-test**.



This will bring up the One-Sample T Test window. To run the test we need to move “hfemale” into the test variable and change the test value to 62.5 [the value of the null hypothesis]. The standard hypothesis test will be run at a significance level of 5%, but if you did wish to change this you can click the options button and change this setting.



SPSS will not indicate whether the null hypothesis has been rejected. That is up to the researcher to determine from the presented results. Below is the output received from SPSS for the female heights. We will use the P-value method to determine if the null hypothesis should be rejected. The output contains the t value for the data and its degrees of freedom. Additionally, it displays the test value [under  $H_0$ ] and a column labeled Sig.(2-tailed).

One-Sample Test						
	Test Value = 62.5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
HFEMALE	2.522	11	.028	2.00	.25	3.75

This “Sig” value is what we refer to as the P-value. It is the probability of seeing a test score this extreme or more [in the direction of the alternative hypothesis] if the null hypothesis is true. The smaller the number, the more rare our test score is under  $H_0$  and the more likely that the null hypothesis isn’t true. Using the .05 significance level as our cutoff, we find the P-value of .028 to be in our rejection region. We reject the null hypothesis and conclude that there is sufficient evidence that the heights of women at Truman are not equal to 62.5 inches.

#### One tailed hypothesis test (example 1)

A researcher believes that the female population at Truman State is taller than the average height of the general population of the United States [62.5 inches].

$$H_o : \mu_F = 62.5 \text{ inches} \quad H_A : \mu_F > 62.5 \text{ inches}$$

We conduct the analysis on SPSS exactly as it was done before. The only difference will be in the interpretation of the computer output. The P-value for the test statistics is .028, but we need to remember that our test is a one sided test and the P-value output by SPSS is for a 2 sided test. As we are not interested in scores occurring in the lower tail, we must divide the outputted P-value by 2. So the P-value for the test is .014, resulting in our rejection of the null hypothesis. We would conclude that there is sufficient evidence that Truman State females are taller than the general population.

### **IMPORTANT MISINTERPRETATIONS ARE POSSIBLE!!!**

Unfortunately, because SPSS does not automatically conduct a one sided test, researchers can make incorrect conclusions by naively dividing the P-value by 2 and making a decision. A researcher must utilize common sense to realize which side of the hypothesis their data lies upon. We may get a “rare” P-value but it might be in the opposite direction of the alternative hypothesis.

#### One tailed hypothesis test (example 2)

Assume that the researcher believes that the female population at Truman is actually **shorter** than the general population [62.5 inches].

$$H_o : \mu_F = 62.5 \text{ inches} \quad H_A : \mu_F < 62.5 \text{ inches}$$

The hypothesis are setup as before, but now the alternative hypothesis states that women are shorter than the general population. Running our SPSS analysis exactly as before we once again obtain a P-value of .014, indicating that we have a rare sample if the average height of women is truly 62.5 inches. But this time we **CANNOT** reject  $H_o$  and rule in favor of  $H_A$ . The P-value tells us that getting an average this far away from 62.5 is a rare event, but looking at the data tells us its an even rarer event if the average height was actually shorter [as the alternative hypothesis assumes!]. So we conclude that there is not sufficient evidence that Truman females are shorter than the general population.

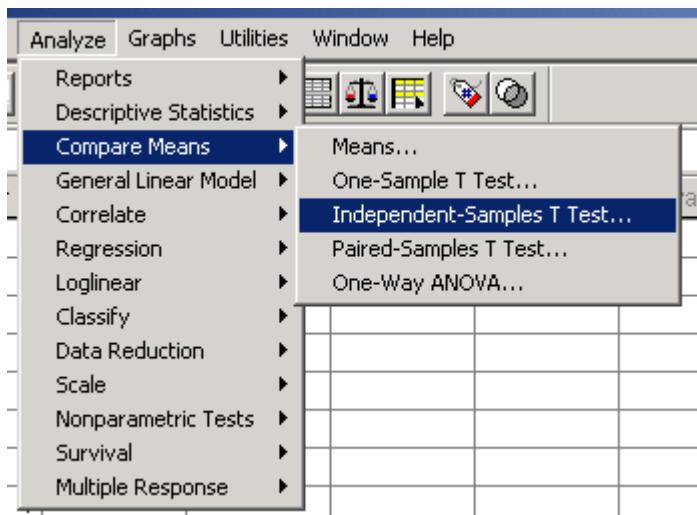
The true P-value for this test is actually obtained by dividing the Sig value by 2 and subtracting that from 1. So P-value is .986. This indicates that we would almost always expect test scores to be more extreme [in the direction of the alternative] if the null hypothesis were true [this score is not rare at all!]. Basically the data shows almost no evidence to support the alternative hypothesis.

### Comparing 2 independent samples

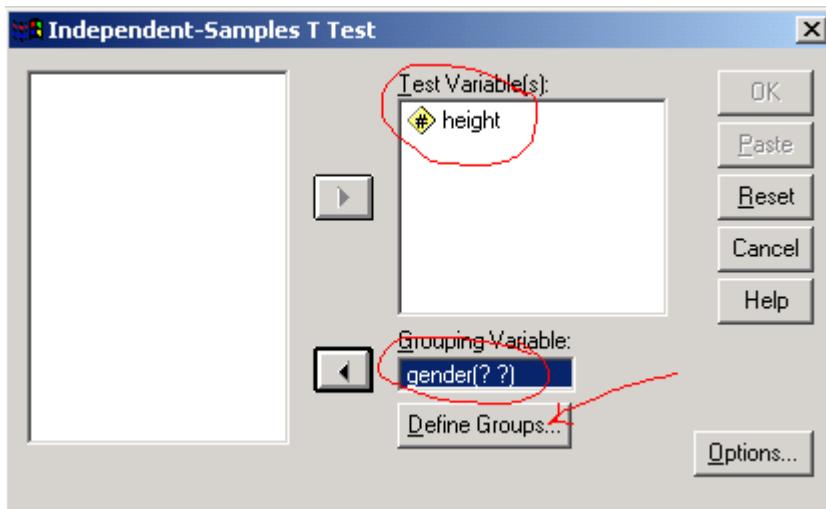
One question of interest in this data set might be to determine if there is a statistical difference in the heights of men and women. The appropriate hypotheses for this question is

$$H_o : \mu_M - \mu_F = 0 \quad H_A : \mu_M - \mu_F \neq 0$$

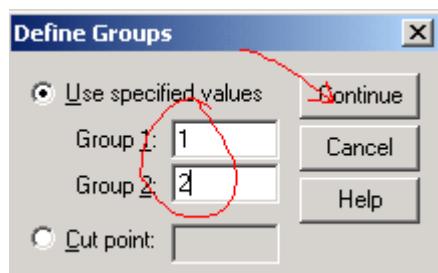
The appropriate test for this hypothesis is the independent samples t-test. To run this test, click on **Analyze-Compare Means-Independent Samples t-test**



We then move the height and gender to their appropriate fields and click on the **Define groups** button.



We choose which groups we wish to compare [we can only choose 1 & 2 in this example] and hit continue.



The output window will contain the results of the hypothesis test. The top row of the table assumes that the variances of the two groups are equal. The bottom row assumes that the variances are unequal and SPSS will calculate the correct degrees of freedom  $\Delta$ , which we did by hand. As before the table contains t-observed, the df, and the P-value for the test. We can see that we have a P-value of .000 [which means its zero to 3 decimal places so it could be .0004] so we reject the null hypothesis and conclude that there is sufficient evidence that the average height of men and women are not the same.

If we check the calculation for df assuming unequal variances

$$\Delta = \frac{\left[ \frac{3.25^2}{10} + \frac{2.75^2}{12} \right]^2}{\frac{(3.25^2)^2}{10} + \frac{(2.75^2)^2}{12}} = 17.77$$

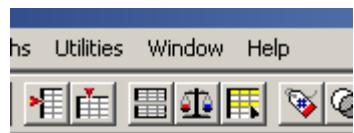
We see that the hand calculations match the SPSS output. Note also that we receive the same conclusion regardless of which test we use, though in some cases they may lead to different conclusions.

### Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
HEIGHT	Equal variances assumed	.085	.774	4.229	20	.000	5.40	1.28
	Equal variances not assumed			4.162	17.765	.001	5.40	1.30

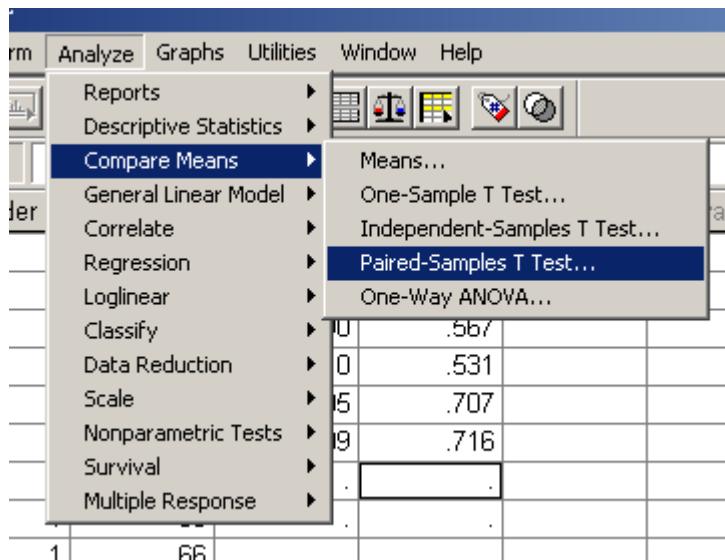
### Comparing 2 dependent samples (Paired t-test)

If the test subjects are matched and paired by certain traits when assigned to the treatment groups, we no longer have independent samples. Now each measurement in a treatment group can be reasonably paired with a measurement in the other treatment group [and this was done before the experiment was run!]. The data will be placed into 2 columns with each row corresponding to a “paired” data set. For the river water quality example done in class we have 6 locations with a measurement at the top and bottom at each location.

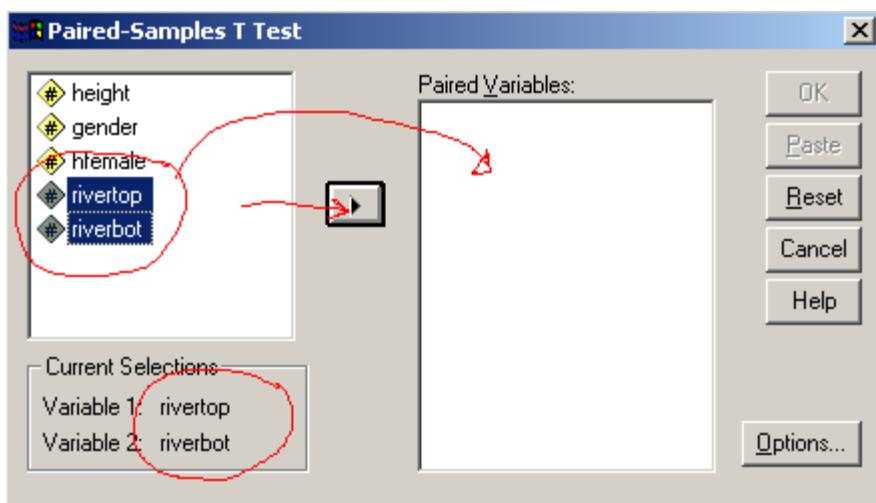


rivertop	riverbot	
.415	.430	
.238	.266	
.390	.567	
.410	.531	
.605	.707	
.609	.716	
.	.	
.	.	
.	.	

To test if there is a significant difference between top and bottom we run the paired Samples t test in SPSS.



This will bring up the data window. We highlight the two variables representing our paired data and then use the arrow key to select them.



Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair	RIVERTOP	.44450	6	.14177	5.79E-02
1	RIVERBOT	.53617	6	.17133	6.99E-02

**Paired Samples Test**

	Paired Differences						t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference								
				Lower	Upper							
Pair 1 RIVERTOP - RIVERBOT	-0.0917	.0607	2.48E-02	-.15536	-2.8E-02	-3.700	5		.014			

We can see that the p value for the test is well below our cutoff of 0.05 and so we would reject the null hypothesis and conclude there is significant evidence that the concentration of pollutants differs from the top and bottom of the river.