SPSS Instructions
Data Analysis Assignment #3

To Open the Data File

Begin the assignment by locating the data file WAGES2003 posted on the Data Analysis page of the course website. To open this file in SPSS, you must first save it to the hard drive of your computer. Click on the WAGES2003 link and select the save option. You can save the data file to any location on the hard-drive; however, I recommend that you save it to the desktop (Be sure to erase the file upon completion of the assignment if you are working from a university lab).

Once you have the file saved, open the SPSS program by navigating through the start menu. When SPSS opens, make sure the option to open an existing data file has been selected and click on more files. Then navigate to the location where you stored the WAGES2003 file, click open. Check the data view window to see if the data were imported correctly. Your data set should have 100 observations and 12 variables. Now you are ready to begin the assignment.

Hypothesis Testing for Means in SPSS

SPSS is capable of conducting three different types of hypothesis test for means including a one-sample t-test, a two-sample t-test (i.e. Independent Samples t-test), and a paired sample t-test. In the third data analysis assignment you will work with only the first two methods. Hypothesis test is a fairly straightforward procedure in SPSS. Given sample data that meet the conditions required for inferences about means, the program easily calculates test statistics and P-values for test of equality. However, it is your responsibility to identify the appropriate test and hypothesis and to interpret the output correctly. You must select a variable for which you would like to conduct a test and provide an estimate of the population mean (i.e. the value from your null hypothesis). SPSS will then generate a test statistic and P-value (for a two tailed test) which you must then interpret to make a decision about the null hypothesis.

The examples that follow illustrate how one sample t-test and two sample t-test are preformed in SPSS.

One Sample T-Test

The SATScores05 data set records the verbal and math SAT test scores for 150 randomly selected students at a selective liberal arts college. Use this data to determine whether the average verbal test scores at this college are significantly higher than the national average of 520. Use a significance of 5 percent.

Identify the appropriate null and alternative hypotheses to address this research question?

Ho: $\mu = 520$
Ha: $\mu > 520$

Identify the mean and standard deviation from this sample:

Beginning in the data view window of SPSS select the Analysis tab from the menu at the top of the screen. Then select the *Descriptive Statistics* option followed by *Descriptives*. Select your variable of interest (In this example our variable of interest is verbal) from the list on the left side of the dialog box and use the arrow to move it the *Variables* box on the right. Then click OK. You should be able to identify the sample mean, standard deviation and sample size from the output provided.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Verbal	150	310	800	530.03	94.497
Valid N (listwise)	150				

Notice our sample mean in this example (530.03) exceeds the national average of 520 points. However, we must conduct a formal hypothesis test to determine whether this evidence suggests that we can confidently reject the null hypothesis.

The test statistic for this one sample t-test is: $t = \dfrac{\bar{y} - \mu}{\dfrac{s}{\sqrt{n}}}$

While you could calculate this value by hand given the descriptive statistics and null hypothesis, SPSS can calculate this value for you much more quickly.

SPSS will calculate the value of our test statistic if you select *Analyze/CompareMeans/ One-Sample t-Test* from the menu options. Once you have navigated to the test menu, you should move the variable of interest (Verbal) to the test variable dialog box. Finally insert the hypothesized value for the mean of this variable (the value provided by the null hypothesis - 520) into the test value box. SPSS generates the following output:

One-Sample Test

	Test Value = 520					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Verbal	1.300	149	.195	10.033	-5.21	25.28

This output includes the following:

1. A calculation of the test statistic (t) according to the equation above: 1.3 in this example

2. The degrees of freedom (df): 149 in this example

3. A precise estimate of the P-value for a two tailed test which is labeled as Sig. (2-tailed) in the output above. In this example the reported P-value is 0.195; however, our alternative hypothesis is that the mean verbal test score is larger than the national average. Thus we are conducting a 1-tailed test and must divide the P-value reported in this table by 2 before comparing it to our significance level $\alpha = 0.5$.

4. The mean difference column reports the difference between our sample mean (530.03) and the hypothesized value of the true population mean (520).

5. The final two columns of this output give the upper and lower limits of the confidence interval for the mean difference $(\bar{y} - \mu)$. Note that this is not the confidence interval for the true population mean. SPSS will not automatically generate the types of confidence intervals that we have studied in chapter 23; however, you can easily construct these intervals given the descriptive statistics of the data.

Now that SPSS has given us the value of the test statistic and the P-value we can easily make a decision about the null hypothesis using either the critical t-value method or the p-value approach. If you select the critical t-value approach you must identify the critical t value in the t distribution table. *See class notes for details on how to make this decision. Do not forget to summarize your conclusions in context!!!*

Note you should also be able to interpret the P-value in this context as a conditional probability: Example – The probability that we would observe an average verbal SAT score of 530.03 from a sample of 150 students is 0.0975 if the true average score at this college is equal to the national average of 520.

Two Sample T-Test

Nationally females tend to have higher Verbal scores than males. Is that true at this college? Use a significance level of 5 percent.

Identify the mean and standard deviation for the two groups:

Descriptives

	Gender			Statistic	Std. Error
Verbal	1	Mean		530.47	10.864
		95% Confidence Interval for Mean	Lower Bound	508.82	
			Upper Bound	552.11	
		5% Trimmed Mean		529.93	
		Median		530.00	
		Variance		8851.468	
		Std. Deviation		94.082	
		Minimum		330	
		Maximum		740	
		Range		410	
		Interquartile Range		140	
		Skewness		.057	.277
		Kurtosis		-.599	.548
	2	Mean		528.51	11.128
		95% Confidence Interval for Mean	Lower Bound	506.34	
			Upper Bound	550.69	
		5% Trimmed Mean		526.68	
		Median		520.00	
		Variance		9163.514	
		Std. Deviation		95.726	
		Minimum		310	
		Maximum		800	
		Range		490	
		Interquartile Range		150	
		Skewness		.312	.279
		Kurtosis		-.305	.552

In this data the gender variable takes a value of 1 for females and 2 for males (Sorry, I forgot to put the definitions in the data!). Notice that in this sample of students the females have a slightly higher average verbal score. However, we want to know whether the difference in the test scores across the two groups is sufficiently large enough for us to confidently reject the null hypothesis.

Hypothesis: The research question is do females have higher average verbal test scores at this college? In other words, is the difference between the average verbal test scores between females and males greater than zero?

Ho: $\mu_f - \mu_m = 0$
Ha: $\mu_f - \mu_m > 0$

The test statistic for this two sample t-test is:

$$t = \frac{(\bar{y}_f - \bar{y}_m) - (\mu_f - \mu_m)}{\sqrt{\frac{s_f^2}{n_f} + \frac{s_m^2}{n_m}}}$$

SPSS will conduct this test if you select *Analyze/CompareMeans/Independent-Samples t test* from the menu options. Once you have navigated to the test menu, you should move the variable of interest (Verbal) to the test variable dialog box. You must then split the sample into two groups by moving the variable that defines the groups (gender in this example) to the *Grouping Variable* window. Next, click on the *define groups* option to define how the groups are labeled for this variable. In our example the first group is females (always follow the order of the null hypothesis or your results will be backwards). Females are labeled as a 1 in the data set, so in the *Group 1* box insert a 1. In the *Group 2* box insert a 2 for males. Now click continue and then OK. SPSS generates the following output:

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Verbal	1	75	530.47	94.082	10.864
	2	74	528.51	95.726	11.128

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Verbal	Equal variances assumed	.054	.817	.126	147	.900	1.953	15.550	-28.777	32.683
	Equal variances not assumed			.126	146.860	.900	1.953	15.552	-28.781	32.687

For the purpose of this class we will always assume that the variances of the two groups are not equal therefore you can ignore the first row of information.

Here the value of test statistic calculated using the above formula is 0.126, while the P-value for a 1-tailed test is 0.45 (=0.90/2).

Caution: To identify the degrees of freedom necessary to find the critical t-value use the small of the two groups (n=74 for males, df = 73). In the output above, SPSS has apparently used the more complicated formula for finding the degrees of freedom found on page 547 of your textbook. While there is certainly nothing wrong with this method, let's use the smaller of the two samples to be consistent with the procedure we use when conducting hypothesis test by hand.

If you have followed all of these instructions and are still experiencing trouble, send me an email (kbarbou@ilstu.edu) or plan on attending one of the SPSS tutorial sessions.