SPSS Assignment #2 (C.I. and H.T.)

The main goal of this assignment is to give you some experience on how to use SPSS to construct confidence intervals and to test hypothesis. After you do this assignment, you should be able to read SPSS output and make the proper inference based on the values found in the SPSS output. This assignment is worth 35 points. You can work in groups of 2 – 3 people.

I. Open the SPSS data file **HealthExam.sav**. You can download it from our course webpage at http://websites.uwlax.edu/storibio/ (then click STAT **245**, then **Health Exam Data**). Clicking on it should start the SPSS program automatically and your data will show up in an SPSS window.

II. This data are from the U.S. Department of Health and Human Services, National Center for Health Statistics, Third National Health and Nutrition Examination Survey. It has a total of 80 cases (40 males and 40 females) with each case having values for 14 variables. These variables are listed below. For this assignment, you will only work with 3 variables (***Gender, Height, and Weight***).

1. ***Gender***
2. ***Age*** (in years)
3. ***Height*** (in inches)
4. ***Weight*** (in pounds)
5. ***Waist*** (circumference in cm.)
6. ***Pulse*** (pulse rate in beats per minute)
7. ***SysBP*** (systolic blood pressure in mmHg)
8. ***DiasBP*** (diastolic blood pressure in mmHg)
9. ***Cholesterol*** (in mg)
10. ***BodyMass*** (body mass index)
11. ***Leg*** (upper leg length in cm)
12. ***Elbow*** (elbow breadth in cm)
13. ***Wrist*** (wrist breadth in cm)
14. ***Arm*** (arm circumference in cm)

Note that in the lower left corner of your SPSS spreadsheet you will find two options: ***Data View*** and ***Variable View***. Clicking on the "Variable View" will show you the list of variables in the data with their characteristics. This is where you can modify the characteristics of variables. Clicking on the "Data View" will take you back to the actual values of the variables. This is where you can change the values of the variables.

III. First, use SPSS to fill the table below for variables ***Height*** and ***Weight***. You can use the following commands to study a particular variable: "**Analyze**"→ "**Descriptive Statistics**" → "**Explore**". Then choose the variable of interest. To get summary measure for males and females separately, put the variable ***gender*** in the "**factor list**" box. Later in part IV, when you want to construct qqplots and perform the Shapiro-Wilk test of normality, click on "**plots**", then check the "**normality plots with tests**".

Parameter	Sample Mean	Sample Std. Dev. (S_X)	Standard Error of the Sample Mean	99% C.I. Lower Bound	99% C.I. Upper Bound
μ (all)	$\overline{X}$	SD (all)	SD ($\overline{X}$)		
μ_F (female)	$\overline{X}_F$	SD (female)	SD ($\overline{X}_F$)		
μ_M (male)	$\overline{X}_M$	SD (male)	SD ($\overline{X}_M$)		
$\mu_M - \mu_F$	$\overline{X}_M - \overline{X}_F$	XXXXXX	SD ($\overline{X}_M - \overline{X}_F$)		

To fill the last row, use the following commands: "**Analyze**"→ "**Compare Means**"→ "**Independent-Sample T-Test**". Put ***Gender*** in the "**Grouping Variable**" box and click "**Define Groups**" and type M for group 1 and F for

group 2, then hit **continue.** Then, hit **OK**. Construct this table for both the ***Height*** and ***Weight*** variables. [Note: ***the default confidence level is 95%, so make sure that you change it to 99%***]

IV. Use SPSS to conduct significance tests to test the null hypothesis that the mean value of the males is equal to the mean value of the females for both ***Height*** and ***Weight*** variables, using the two-sided alternative with α=0.01. The two commonly used statistical procedures to do this are: (1) the Independent samples t-test and (2) the Mann-Whitney nonparametric (or distribution-free) test.

You can access the t-test procedure by choosing "**Analyze**"→ "**Compare Means**" → "**Independent Samples T Test**". Use ***Gender*** for "**grouping variable**" (Just like you did earlier). But before using the t-test procedure, show that the normality assumption is reasonable using QQ-plots and the Shapiro-Wilk normality test (see instructions of part III). Once you have shown that the normality assumption is reasonable, go ahead and perform the t-test. Use the result of the Levene's test for equality of variance to choose the appropriate test (whether equal variance or not) and then summarize your results in tables that will include the observed t-statistic value, the degrees of freedom, and the correct p-value for your alternative. Write a short discussion and conclusion about your results.

If one of the samples violates the normality assumption (based on the qq-plot and Shapiro-Wilk test), you will have to use a nonparametric test (Mann-Whitney test) to compare these two samples. You can do this by choosing, "**Analyze**"→ "**Nonparametric Test**" → "2 **Independent Samples**". Click on "**Fields**" tab and then choose the variable you wish to test and move it to the "**Test Field"** box and use ***Gender*** for "**groups**". Then hit "**run**". Copy and paste the resulting table. Write a short discussion and conclusion about your results.

V. Here are the things that I want you to include in your final work.

1. Using the variable ***Height***, complete the table of summary statistics of part III.

2. Copy and paste the 2 qq-plots and the test of normality table into your word document. Comment on the normality of the data. Did you see strong evidence against the normality assumption?

   i. If you decide to use the t-test.
      1. Give the p-value associated to the Levene's test for Equality of Variances. Based on this p-value, what can you say about the variances of male values and female values? Is it safe to assume that they are equal or did you find enough evidence to say that they are not equal?
      2. Construct a table that includes the observed value of the t-statistic, its degrees of freedom, and the p-value for the t-test to test H_o: $\mu_1=\mu_2$ vs. H_1: $\mu_1\neq\mu_2$. Write a conclusion based on the result of the t-test.
      3. Comment on how you can obtain the same result using the confidence interval for $\mu_1 - \mu_2$.

   ii. If you decide to use the nonparametric Mann-Whitney test.
      1. Copy and paste the resulting table.
      2. Using the appropriate (asymptotic) p-value, test H_o: $\mu_1=\mu_2$ vs. H_1: $\mu_1\neq\mu_2$. Write a conclusion based on the result of the Mann-Whitney test.

3. Repeat steps 1 and 2 for the variable ***Weight***.

Reading SPSS Output for SPSS Assignment #2

Variable: **Waistline**

1. **Obtaining summary measures:** Click on "Analyze" → "Descriptive Statistics" → "Explore". Move the variable "**waist**" into the "dependent list" (putting "**gender**" in the "Factor list" will give you summary measures for males and females separately). To get the qq-plots and the Shapiro-Wilk test, make sure you click on "plots" then check the box for "Normality plots with tests". Below is the SPSS output that you will get:

Descriptives

	Gender			Statistic	Std. Error
Waist	Female	Mean		85.0325	2.43517
		99% Confidence Interval for Mean	Lower Bound	78.4383	
			Upper Bound	91.6267	
		5% Trimmed Mean		83.7472	
		Median		81.9500	
		Variance		237.202	
		Std. Deviation		15.40136	
		Minimum		66.70	
		Maximum		126.50	
		Range		59.80	
		Interquartile Range		22.05	
		Skewness		.962	.374
		Kurtosis		.611	.733
	Male	Mean		91.2850	1.55930
		99% Confidence Interval for Mean	Lower Bound	87.0626	
			Upper Bound	95.5074	
		5% Trimmed Mean		91.2333	
		Median		91.2000	
		Variance		97.256	
		Std. Deviation		9.86185	
		Minimum		75.20	
		Maximum		108.70	
		Range		33.50	
		Interquartile Range		18.78	
		Skewness		.037	.374
		Kurtosis		-1.058	.733

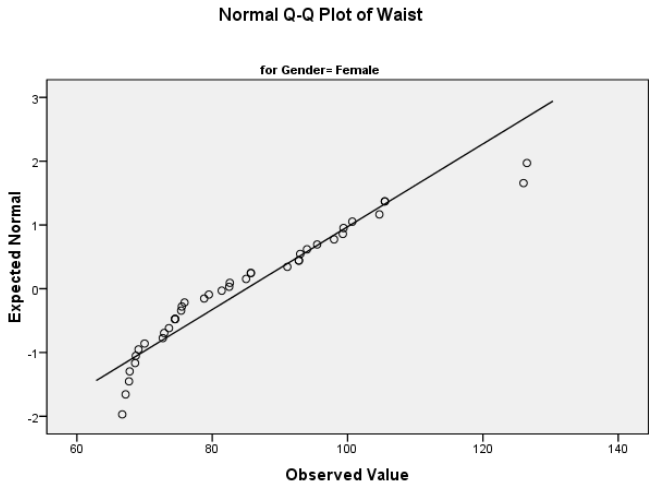
I don't want you to copy and paste this whole table. Just pick out the correct values to put in your table.

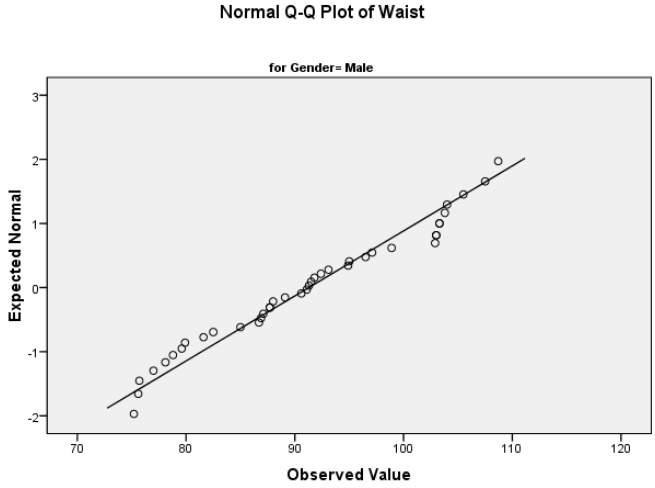
2. **<u>Checking Normality.</u>** Together with the above table, you will also get results of the Shapiro-Wilk test to determine if it is reasonable to assume that both data sets come from normal population. The result for the "**waist"** variable is given below

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Gender	Statistic	df	Sig.	Statistic	df	Sig.
Waist	Female	.148	40	.027	.905	40	.003
	Male	.131	40	.084	.952	40	.090

Note that the p-value for the Shapiro-Wilk test are 0.003 and 0.090 (in the last column under "Sig."). This implies that the female data set is not normal because the p-value was smaller than alpha=.05. You can also see a curve pattern in the corresponding qq-plots (see left figure below), suggesting that the female data is not normal.





3. **Nonparameteric Test:** Because the female data is not normal, we cannot use the ordinary 2-sample independent t-test. Instead, we are going to use a non-parametric test, called the Mann-Whitney test. You can do this by choosing, "**Analyze**"→ "**Nonparametric Test**" → "2 **Independent Samples**". Choose the variable you wish to test and use ***Gender*** for "**grouping variable**". Click on "**Fields**" tab and then choose the variable you wish to test and move it to the "**Test Field"** box and use ***Gender*** for "**groups**". Then hit "**run**". You will then get something like the table below:

   The p-value that you want to look for is the value in the third column, labeled as " Sig.". In this particular example, the p-value is 0.010. Because this p-value is smaller than alpha=0.05, we reject the null hypothesis and conclude that the mean waistline for males is not equal to the mean waistline for females.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Waist is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	.010	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

4. **T-test:** Let's suppose we that we can assume that both data sets are reasonably normal, just so that I can illustrate how to perform the t-test. You can access the t-test procedure by choosing "**Analyze**"→ "**Compare Means**" → "**Independent Samples T Test**". Use ***Gender*** for "**grouping variable**" (Just like you did earlier), then click on "Define Groups", and type "M" for group 1 and "F" for group 2., then hit "continue" and then click "ok". You will then get the table below

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
Waist	Equal variances assumed	7.430	.008	2.162	78	.034	6.2525	2.8916	.49573	12.00927
	Equal variances not assumed			2.162	66.38	.034	6.2525	2.8916	.47982	12.02518

When using the t-test, you need to decide if you can assume equal variances. From the table above, we see that the p-value for the Levene's test for equality of variance is 0.008 (under "sig"). Since this value is less than alpha=0.05, this implies that the variances cannot be assumed to be equal. Therefore, you should use the t-test result given in the second row "Equal variances not assumed". The corresponding t_obs is 2.162, df=66.378, and the p-value is 0.034. Since this p-value is smaller than alpha=0.05, we reject the null hypothesis.