Reading SPSS Output

Variable: **Waistline**

1. **<u>Obtaining summary measures:</u>** 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

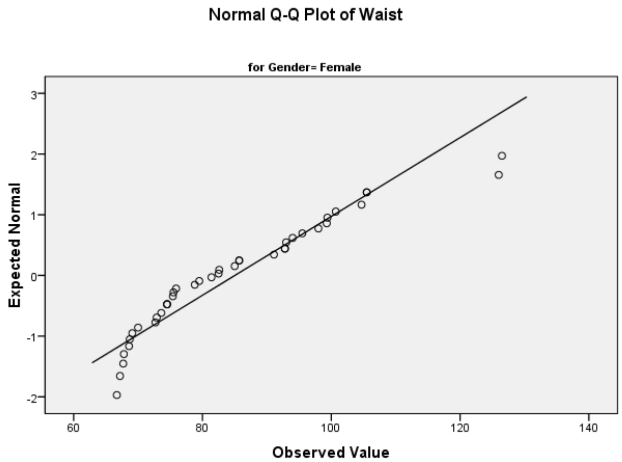
	Gender			Statistic	Std. Error
Waist	Female	Mean		85.0325	2.43517
		99% Confidence Interval for Mean	Lower Bound	78.4383	
			Upper Bound	91.6267	
		5% Trimmed Mean		83.7472	
		Median		81.9500	
		Variance		237.202	
		Std. Deviation		15.40136	
		Minimum		66.70	
		Maximum		126.50	
		Range		59.80	
		Interquartile 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

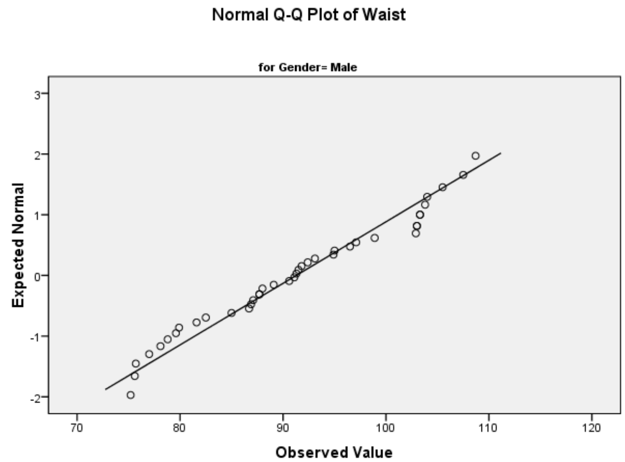
2. **<u>Checking Normality.</u>** Together with the above table, you will also get results of the Shapiro-Wilk test to determine if it is reasonable to assume that both data sets come from normal population. The result for the "**waist"** variable is given below

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	6.25250	2.89162	.49573	12.00927
	Equal variances not assumed			2.162	66.378	.034	6.25250	2.89162	.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.

5. SPSS output for the chi-square test using Occupation data.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	65.524(a)	3	.000
Likelihood Ratio	74.295	3	.000
N of Valid Cases	490		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 34.75.

a. The X^2**_obs** that we compute in class is the Pearson Chi-Square. In this case, X^2**_obs=65.524** and its corresponding p-value is given in the last column of that row under "Asymp. Sig (2-sided)".
b. One condition that we need to satisfy for the Chi-square test is the minimum expected count for each cell. For the test to be appropriate for the data, we want the expected counts to be at least 5. This is checked by SPSS and a note is included at the bottom of the table of results.

6. SPSS output for ANOVA using Price Promo data.

Descriptives

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	40	4.2240	.27341	.04323	4.1366	4.3114	3.62	4.79
3	40	4.0628	.17424	.02755	4.0070	4.1185	3.60	4.42
5	40	3.7590	.25265	.03995	3.6782	3.8398	3.25	4.41
7	40	3.5487	.27503	.04349	3.4608	3.6367	2.97	4.01
Total	160	3.8986	.35931	.02841	3.8425	3.9547	2.97	4.79

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.989	3	3.663	59.903	.000
Within Groups	9.539	156	.061		
Total	20.527	159			

7. SPSS output for linear regression using Health Exam data.

In this example, weight is the response variable (*y*) and waistline is the explanatory variable (*x*).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	908[a]	. 825	.522	14.68086

a Predictors: (Constant), SysBP

- About 82.5% of the variability of **weight** can be explained by **waistline**.

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-51.728	11.129		-4.648	.000
	SysBP	2.395	.125	.908	19.180	.000

a Dependent Variable: DiasBP

- Regression Line: **weight = -51.728 + 2.395*waistline.**