Psychology 311 – Introduction to Statistics

Lab Manual

Diane Lane

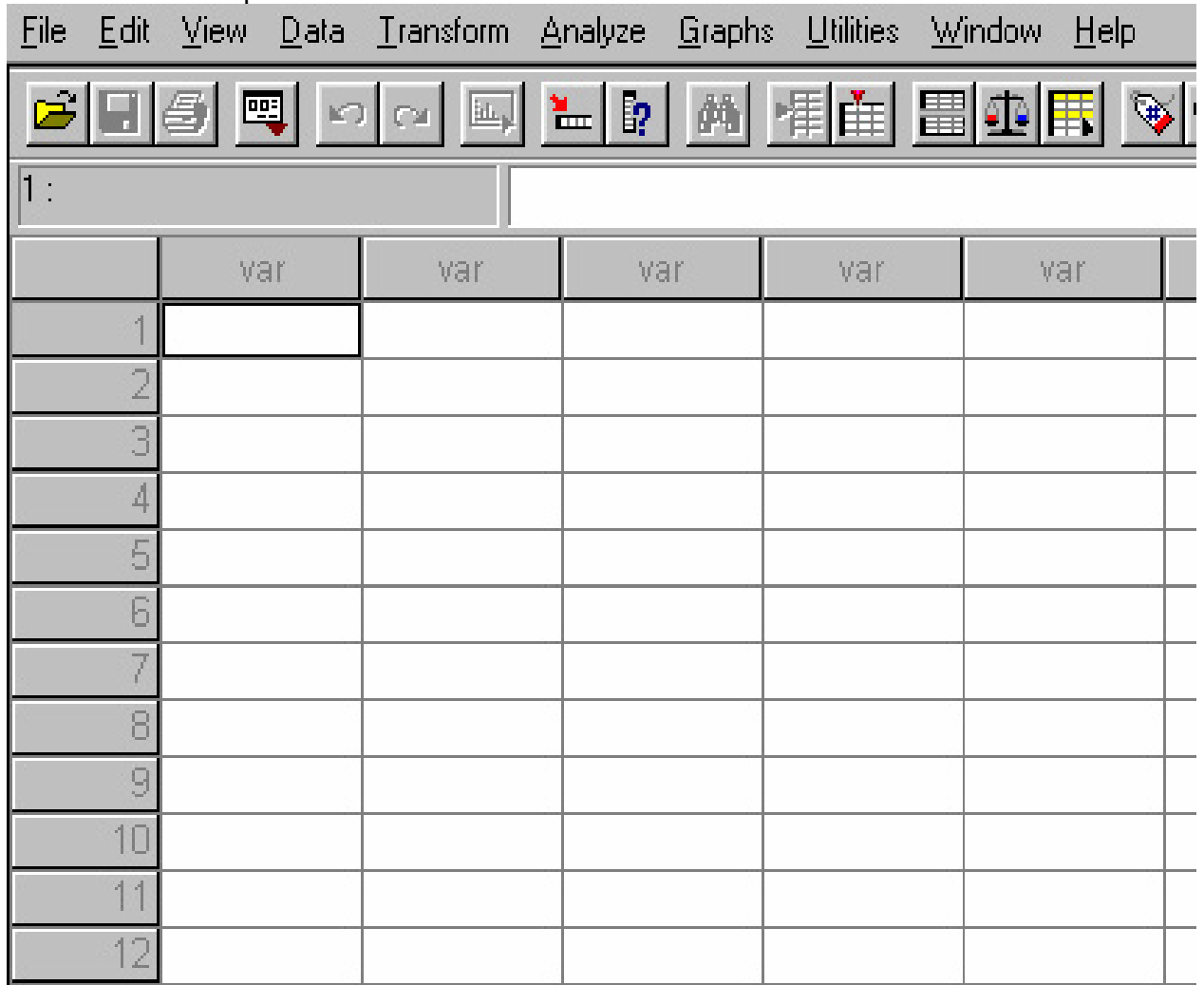### Washington State University Vancouver

Table of Contents

Section 1 - Getting Started

This manual will provide directions for the 10.1 version of SPSS, a statistical program. However, you may use the 7.5 student version to complete lab assignments. For our purposes, there are only minor differences between the two programs.

Opening the Program

Once your computer is on and you are in Windows, click one time on the "Start" button at the left-hand bottom corner of the screen. Go to "Programs" and click on it once. Select SPSS 10.1 from the list of programs by double clicking on it.

SPSS will open to a blank datasheet:

File	Edit	View	Data	Transform	Analyze	Graphs	Utilities	Window	Help

1 :					
	var	var	var	var	var
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					

You can immediately start putting data into the datasheet, or open a previously saved file.

Opening a Saved Data File

        Go to "File" at the top right-hand corner of your screen. Select "Open", then select "Data". From there, select the correct drive (A: from floppy; C: from hard drive), and either select the file from the list provided, or type in the file name. The saved data will open to the datasheet.
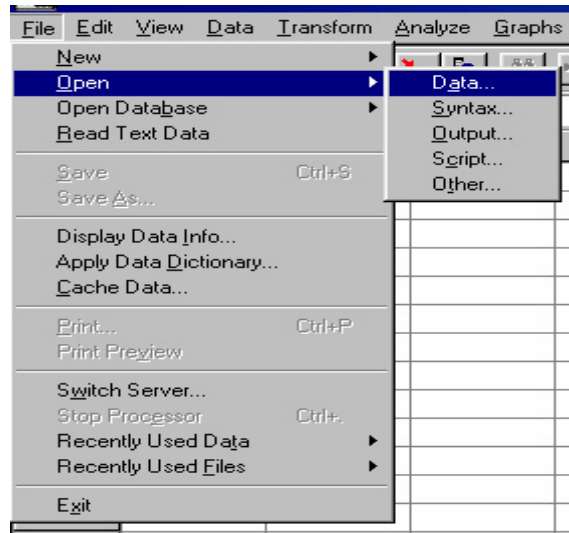

Entering Data to Datasheet

        Variables are listed in columns, not rows. There are two types of variables you can place into the datasheet:
                A. Raw scores - i.e., actual scores such as 145 for a person's weight.
                B. Categorical values - coded numerical values for categorical data.
                        Ex: Freshman = 1
                                  Sophomore = 2
                                  Junior = 3
                                  Senior = 4

Labeling Variables

Variables are labeled by clicking one time on the "Variable View" tab at the bottom left-hand corner of the screen. This will bring you into a screen that summarizes information about each variable. Clicking on the variable of interest will highlight that variable (which should be listed as var00001, var00002, etc.). Once highlighted, change the variable name by typing in the desired name.
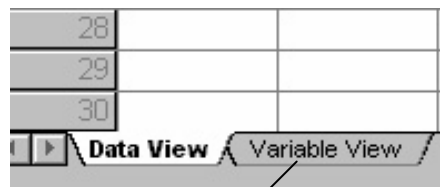
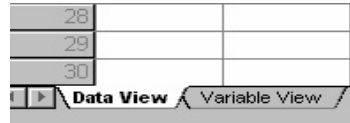|	Name	Type	Width	Decimals	Label	Values	
1	gender	Numeric	8	0		None	No
2	year	Numeric	8	0		None	No
3	weight	Numeric	8	0		None	No
4	iq	Numeric	8	0		None	No
5	pet	Numeric	8	0		None	No
6	stress	Numeric	8	0		None	No
7							

Label Variables   # of Decimals   Categorical Values

Coding Categorical Values

       Categorical values can be entered into the dataset. To identify the categorical values click one time on the "Variable View" tab at the bottom left-hand corner of the screen.

Once in the variable view screen click on the "Values" box that corresponds to the variable of interest.

This will open a "Value Labels" menu box, which allows a values to be assigned to numbers in each variable.
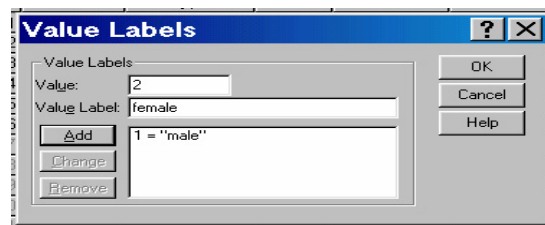
       For example, in the "Gender" variable, 1 signifies a male subject, whereas 2 signifies a female subject. Therefore, type "1" for value, and "male" for value label. Then click add. After clicking add, the program immediately brings you back to the value field to allow labeling of another variable. Continue labeling for each value. When complete click on "OK" one time.

       You will then be brought back to the "Variable View" menu. To return to the datasheet click on the "Data View" tab one time.

Saving Data

To save data, click one time on "File" in the top left-hand corner of the screen. Select "Save". Select drive data will be saved to and type in a file name. **The suffix for any SPSS dataset must be .sav** - SPSS will not recognize a dataset without the suffix .sav.

So for our data, if we wanted to call it stats, our file name would be: stats.sav

Saving Output

To save output, while in output screen, click one time on "File" and select "Save". **The suffix for any SPSS output must be .spo** – SPSS will not recognize an output without the suffix .spo.

So for our output, if we wanted to call it stats, our file name would be: stats.spo

Example Dataset

This example dataset will be used throughout the manual to demonstrate how to use the SPSS program to complete various types of statistics.

Student Name	M/F	Year	Weight	I.Q.	Pet Owner	Stress Level
Jimmy	M	Freshman	220	80	Yes	6
Chris	M	Senior	185	110	No	5
Susan	F	Freshman	104	100	Yes	2
Kay	F	Senior	137	85	Yes	1
Peter	M	Freshman	156	70	No	6
Cindy	F	Junior	130	110	Yes	3
Mary	F	Senior	145	85	Yes	1
Patty	F	Junior	115	90	No	7
Joanne	F	Sophomore	170	115	No	4
Liz	F	Sophomore	102	100	No	5
Renee	F	Junior	149	120	No	3
Rob	M	Sophomore	192	130	No	7
Mike	M	Freshman	180	85	Yes	3
Roger	M	Senior	230	90	No	5
Sandy	F	Freshman	165	115	Yes	1
Steve	M	Junior	199	100	Yes	2
John	M	Sophomore	240	110	Yes	1
Bill	M	Junior	170	100	No	5
Scott	M	Sophomore	175	100	No	4
Nick	M	Freshman	160	90	No	6

Dataset in SPSS:

	gender	year	weight	iq	pet	stress
1	1	1	220	80	1	6
2	1	4	185	110	2	5
3	2	1	104	100	1	2
4	2	4	137	85	1	1
5	1	1	156	70	2	6
6	2	3	130	110	1	3
7	2	4	145	85	1	1
8	2	3	115	90	2	7
9	2	2	170	115	2	4
10	2	2	102	100	2	5
11	2	3	149	120	2	3
12	1	2	192	130	2	7
13	1	1	180	85	1	3
14	1	4	230	90	2	5
15	2	1	165	115	1	1
16	1	3	199	100	1	2
17	1	2	240	110	1	1
18	1	3	170	100	2	5
19	1	2	175	100	2	4
20	1	1	160	90	2	6
21						

You may view the dataset or output file at anytime by clicking on either the SPSS data button (after you name your dataset, the name will be displayed on this button - in this example, the dataset button was named "statslab".) or the SPSS output button at the bottom of the screen.
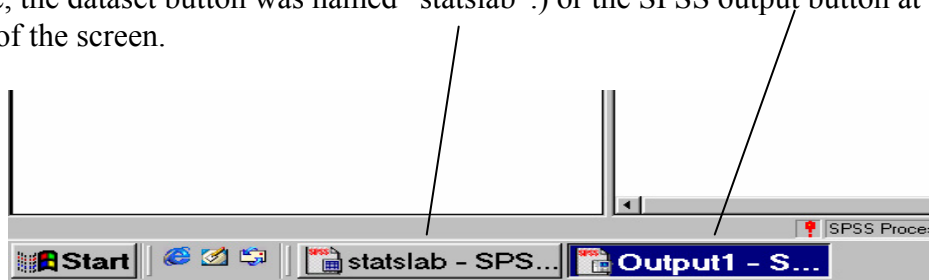
<u>Section 2 - Frequency Distributions and Descriptive Statistics</u>

<u>Frequency Distribution</u>

   To obtain a frequency distribution click one time on "Analyze", go to "Descriptive Statistics", then "Frequencies".
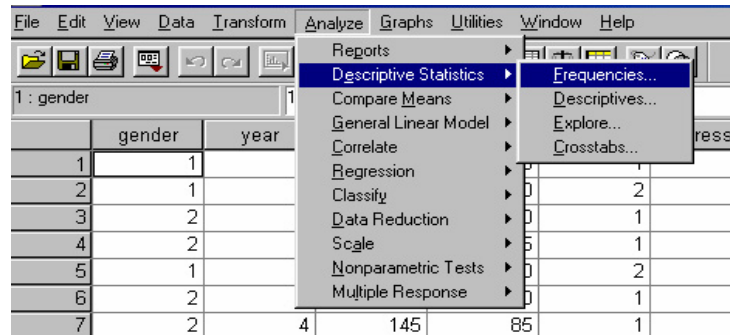
   From the Frequencies menu, place the variable of interest into the "Variables" box by clicking on the variable one time then clicking on the arrow key. (You may use one variable at a time, several variables, or all the variables. You will receive the statistics you select for each variable in the "Variables" box.)
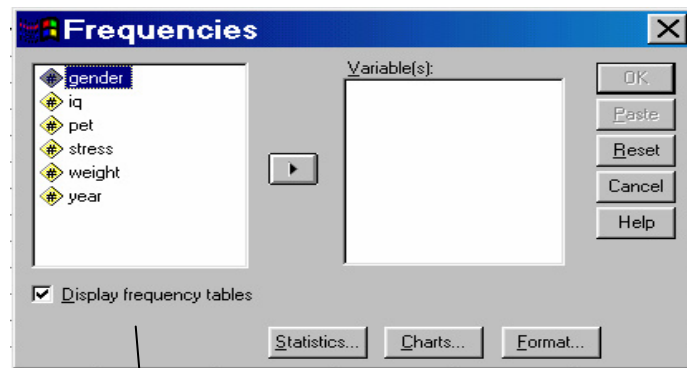

   Click on the "Display Frequency Tables" box to get a frequency table. This table will give you the variable values, its frequency, percentage, valid percentage, and cumulative percentage for each variable placed in the "Variables" box.

For example, if we placed "weight" into the "Variables" box, we would have an output like this:

WEIGHT

N	Valid	20
	Missing	0

WEIGHT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	102	1	5.0	5.0	5.0
	104	1	5.0	5.0	10.0
	115	1	5.0	5.0	15.0
	130	1	5.0	5.0	20.0
	137	1	5.0	5.0	25.0
	145	1	5.0	5.0	30.0
	149	1	5.0	5.0	35.0
	156	1	5.0	5.0	40.0
	160	1	5.0	5.0	45.0
	165	1	5.0	5.0	50.0
	170	2	10.0	10.0	60.0
	175	1	5.0	5.0	65.0
	180	1	5.0	5.0	70.0
	185	1	5.0	5.0	75.0
	192	1	5.0	5.0	80.0
	199	1	5.0	5.0	85.0
	220	1	5.0	5.0	90.0
	230	1	5.0	5.0	95.0
	240	1	5.0	5.0	100.0
	Total	20	100.0	100.0	

Each weight entered into the dataset is listed in the far left column, with the frequency, percent and cumulative percent for each weight.

Measures of Central Tendency and Variance
Click one time on "Analyze", go to "Descriptive Statistics", then "Frequencies".
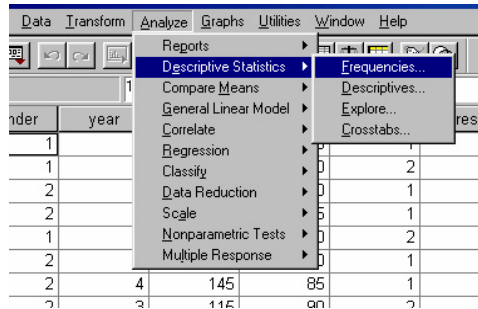
From the Frequencies menu, place the variables of interest into the "Variables" box.
Click one time on the "Statistics" box at the bottom of the screen.
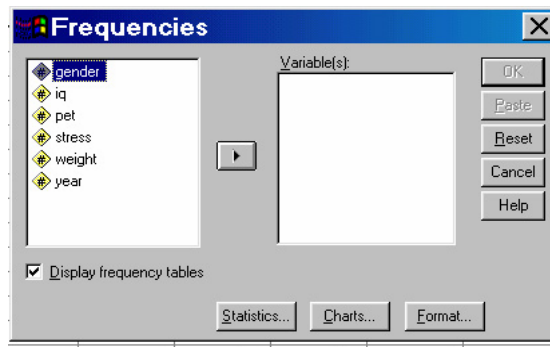
Once in the "Statistics" menu, select desired statistic by clicking on the box to the left of
the desired statistic. Then click "Continue", which will bring you back to the
"Frequencies" menu. If done click "O.K." Selected statistics will be generated in an
output file.
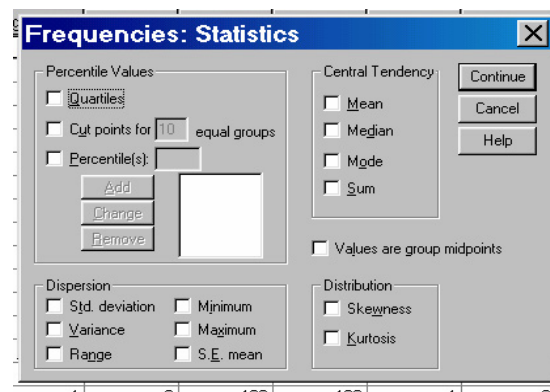
OR

Click one time on "Analyze", go to "Descriptive Statistics", then "Descriptives.
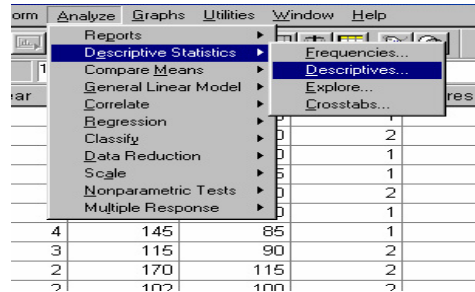
From the Descriptives menu, place the variables of interest into the "Variables" box.
Click on the "Options" box.
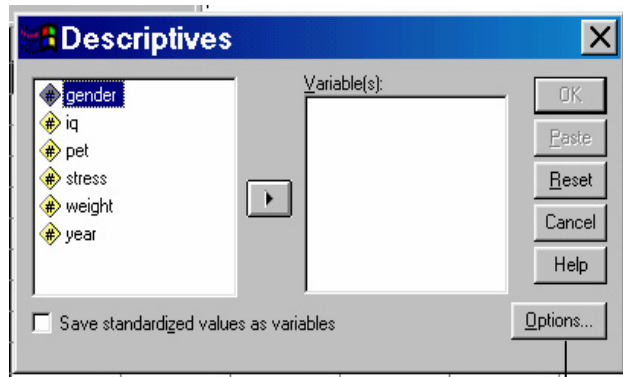
From the options menu select the wanted statistics, then click continue. Once back into
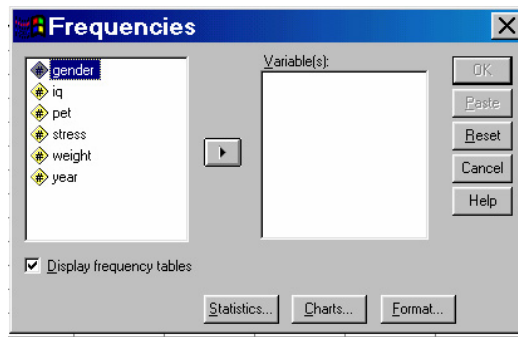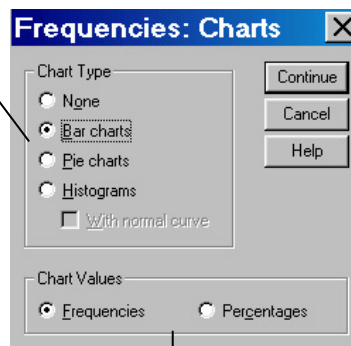the Descriptives menu, click "OK" to complete statistics

Bar Charts and Histograms
Click one time on "Analyze", go to "Descriptive Statistics", then "Frequencies".
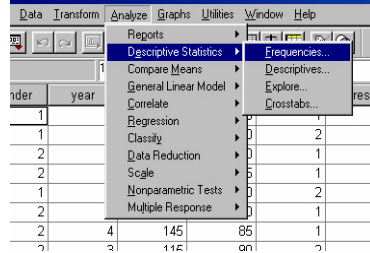
Place the variables of interest into the "Variables" box. Click one time on the "Charts" button at the bottom of the screen.


Select "Bar Chart" or "Histogram". Clicking on "Continue" will return to the Frequencies menu.


Selecting Chart Values: charts can be completed by frequencies (# of each value) or by percentages (percent of total for each value). Make sure to select the correct value depending upon the question asked.
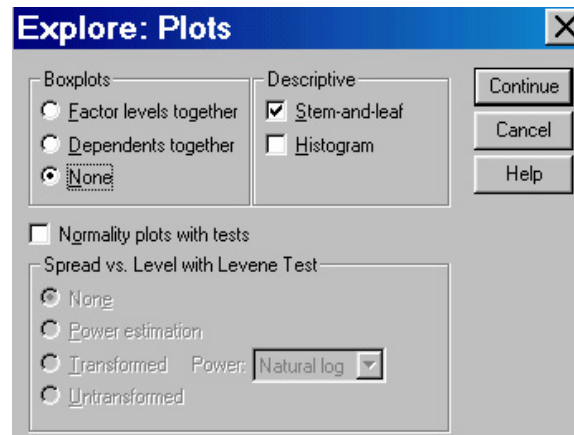
Stem and Leaf Plots

Click one time on "Analyze", go to "Descriptive Statistics", then "Explore".
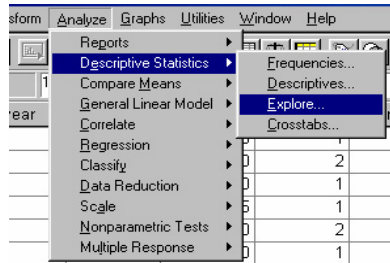
Place the variable(s) of interest into the "Dependent List". Under "Display" box in left lower-hand corner select "Plots". Then click one time on "Plots" button.
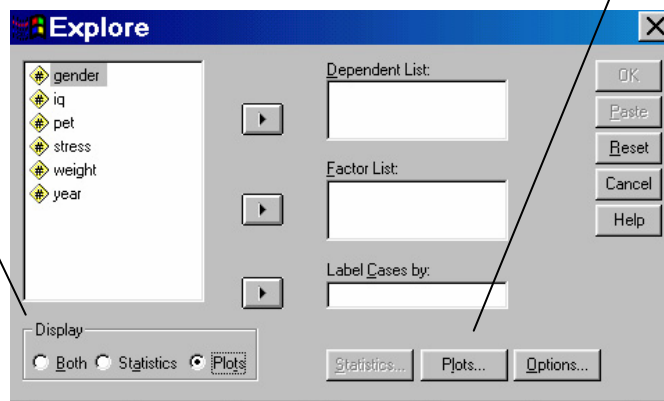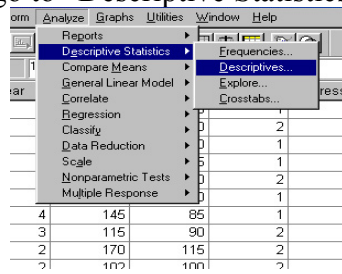

When Plots menu opens, select "Stem-and-Leaf" plot from choices. Then click "Continue", followed by "OK".
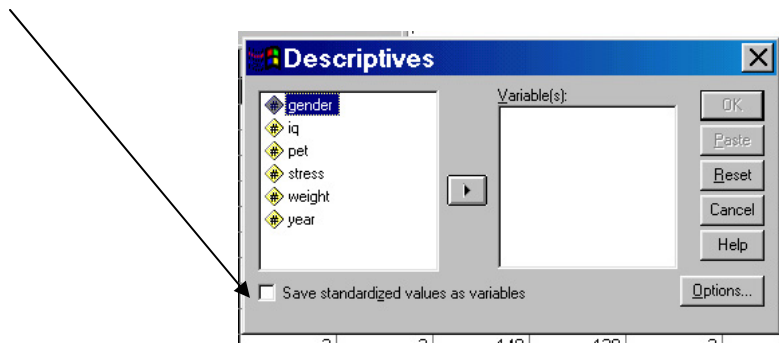
<u>Section 4 – Z-Scores</u>

Click one time on "Analyze", go to "Descriptive Statistics", then "Descriptives".

Place variable(s) of interest into the "Variables" box. Then click on "Save Standardized Values as Variables".


This will create a new variable that will be called zvariable (z before the name of the variable you choose). For example, if we chose weight as our variable, we will now have a new variable, "zweight", **IN OUR DATASET**. Unlike most SPSS calculations, this one will be placed in the dataset, rather than an output file:

	gender	year	weight	iq	pet	stress	zweight
1	1	1	220	80	1	6	1.38991
2	1	4	185	110	2	5	.48569
3	2	1	104	100	1	2	-1.60692
4	2	4	137	85	1	1	-.75437
5	1	1	156	70	2	6	-.26351
6	2	3	130	110	1	3	-.93522
7	2	4	145	85	1	1	-.54770
8	2	3	115	90	2	7	-1.32274
9	2	2	170	115	2	4	.09817
10	2	2	102	100	2	5	-1.65859
11	2	3	149	120	2	3	-.44436
12	1	2	192	130	2	7	.66654
13	1	1	180	85	1	3	.35652
14	1	4	230	90	2	5	1.64826
15	2	1	165	115	1	1	-.03100
16	1	3	199	100	1	2	.84738
17	1	2	240	110	1	1	1.90661
18	1	3	170	100	2	5	.09817
19	1	2	175	100	2	4	.22735
20	1	1	160	90	2	6	-.16018

The reason that z-scores are placed into the dataset rather than an output file is that you may now use the z-scores, instead of the raw scores, for statistical calculations. This can come in handy if you need a standardized number to compare groups tested using different tests/measurement scales.

To demonstrate that you have created this new variable, click one time on "Analyze", go to "Reports", then to "Case Summaries".
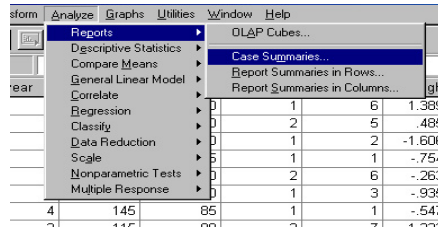
When in the "Case Summaries" menu, place the variables of interest into the "Variables" box, then click "OK" (in this case zweight).
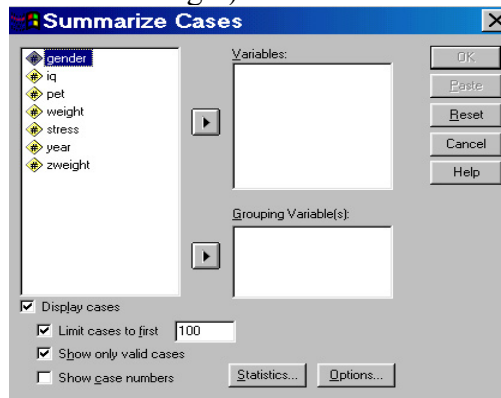
This will give you an output list of your variable in ascending order (NOT the order in your dataset). Here is an example of the output from "zweight":
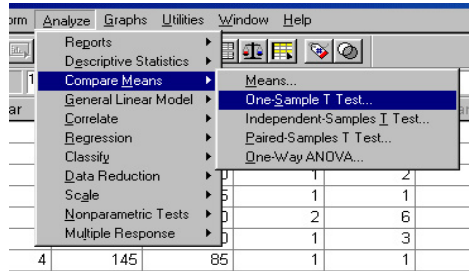
Case Summaries[a]

|	Zscore(W EIGHT)
1	1.38991
2	.48569
3	-1.60692
4	-.75437
5	-.26351
6	-.93522
7	-.54770
8	-1.32274
9	.09817
10	-1.65859
11	-.44436
12	.66654
13	.35652
14	1.64826
15	-.03100
16	.84738
17	1.90661
18	.09817
19	.22735
20	-.16018

Single Sample T-Tests

To obtain a single sample t-test, click one time on "Analyze", go to "Compare Means", then "One-Sample T-Test".

From the one-sample t-test menu, place the variable(s) of interest into the "Test Variable(s)" box. Place the population mean to be tested into the "Test Value" box.
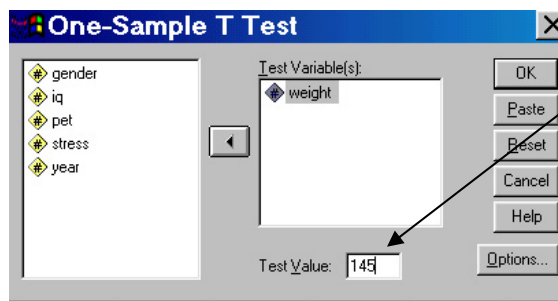
By clicking on the "Options" button, you may set your confidence interval level. The default level set by SPSS is 95%, and unless otherwise stated, this will also be the default for our class.
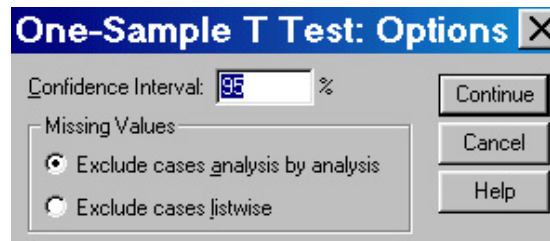
The output generated by the one-sample t-test will look like this:

T-Test

One-Sample Test

|	Test Value = 145				95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
WEIGHT	2.449	19	.024	21.20	3.08	39.32
	1	2	3	4		5

1. Name of variable selected
2. The actual t value calculated by SPSS
3. The degrees of freedom for our t-test
4. The alpha level of our t-test. This is the actual alpha level for our test. If our critical alpha is .05, then any value .05 or less in this box indicates a significant t value. If the alpha level is greater than .05, our t was not significant.
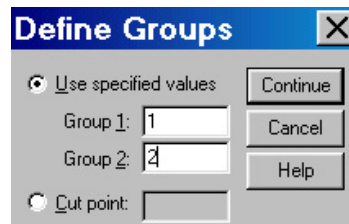5. The confidence intervals for our test.

<u>Independent Samples T-Test</u>

To obtain an independent samples t-test, click one time on "Analyze", go to "Compare Means", then "Independent-Samples T-Test".
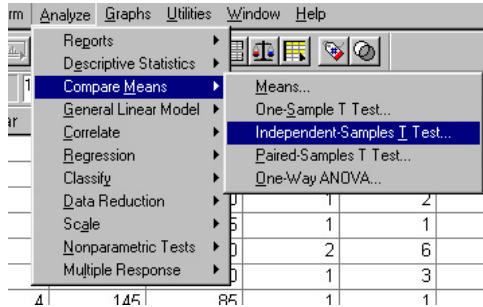
From the independent-samples t-test menu, place the dependent variable into the "Test Variable(s)" box. Place the independent variable into "Grouping Variable" box.
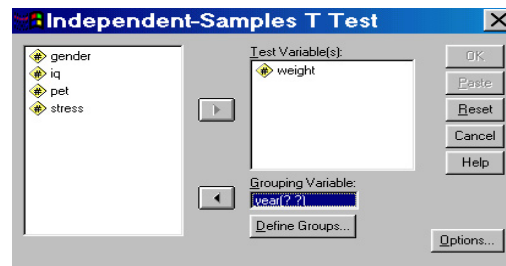

The independent variable has to be defined. For example, in our variable year, 1 = freshmen, 2 = sophomores, 3 = juniors, and 4 = seniors. Since this is a t-test, we can only compare two groups to one another. Therefore, if we were interested examining if there is a significant difference in weight between freshmen and sophomores we would click on the "Define Groups" button and input 1 for "Group 1" and 2 for "Group 2". However, if we were interested in freshmen vs. seniors, we would input 1 and 4.

Once groups are defined, click on the continue button and you will be returned to the independent-samples t-test menu.

At the independent-samples t-test menu, by clicking on the "Options" button, you may set your confidence interval level. The default level set by SPSS is 95%, and unless otherwise stated, this will also be the default for our class.
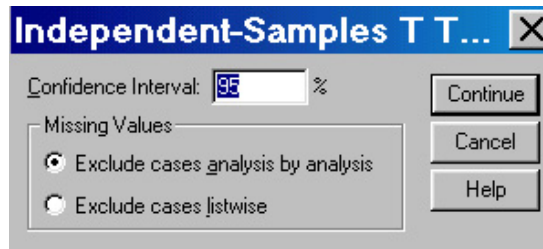
The output generated by the one-sample t-test will look like this:

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
WEIGHT	Equal variances assumed	.190	.673	-.443	9	.668	-11.63	26.269	-71.058	47.791
	Equal variances not assumed			-.431	7.388	.679	-11.63	27.000	-74.806	51.539

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

1. Selected Variable.
2. Test for equal variances (assumption for t-test). Top line equal variance, bottom unequal variance.
3. If the alpha level in the "Sig" box is less than .05, it indicates that there is a significant difference in variances and therefore we should use the lower t value. However, if the "Sig" level is greater that .05 than the variances are equal and we can use the top t value (as in this example – our t value is -.433 with a df = 9).
4. The t value calculated by SPSS.
5. The degrees of freedom for our t value.
6. Our alpha level (see above).
7. Confidence levels (see above).

<u>Section 6 – Analysis of Variance (ANOVA)</u>

<u>One Factor ANOVA</u>
Click one time on "Analyze", go to "Compare Means" and select "One-Way ANOVA".
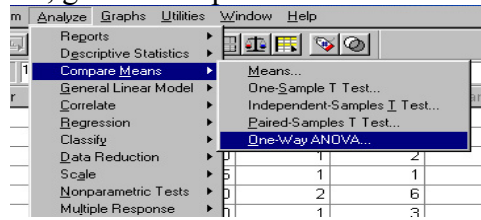
From the ANOVA menu, place the dependent variable into the "Dependent List" box. Place the independent variable into the "Factor" box (in our example we are examining if there are significant differences in weight between freshmen, sophomores, juniors and senior; therefore, weight is placed into the "dependent list" box and year is placed in the "factor: box).
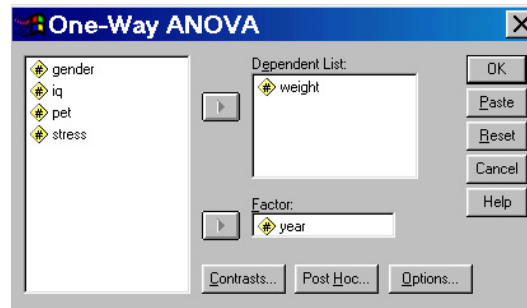
There are three boxes at the bottom of the ANOVA menu. Analysis of Variance only tells you that there is a significant difference between groups, but not which groups specifically. Contrasts allow you to look at differences between 2 groups and should be used if a priori contrasts were designed in the experiment. Post hoc comparisons do the same thing contrast do, however, are used when no specific contrasts were intended in the design of the experiment. The most common post hoc test is the Tukey Test.
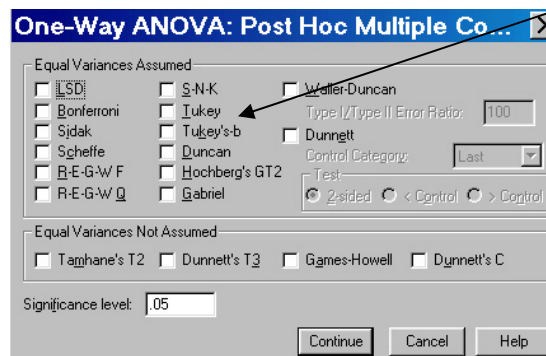
The options button will allow for descriptive statistics for each of the groups and test for homogeneity of variance.
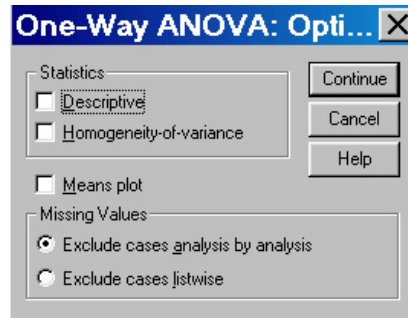
Select the options wanted, then click "OK" from the ANOVA menu. The output generated will look like:

Oneway
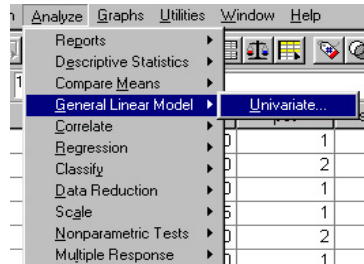
ANOVA

WEIGHT

|	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1669.617	3	556.539	.332	.802
Within Groups	26797.583	16	1674.849		
Total	28467.200	19			

This output should look familiar. It is essentially the same source table that is done by hand in class. The only exception is the last column. The is the actual alpha level for our F value. In this example, the alpha is = .802. Because it is greater than .05, our F value is not significant and we would fail to reject the null hypothesis.

More Complex Analyses

More complex analyses will not be required in this class, however, you may need them in the future. To complete a factorial ANOVA or a repeated factors ANOVA click one time on "Analyze", go to "General Linear Model", then "Univariate".
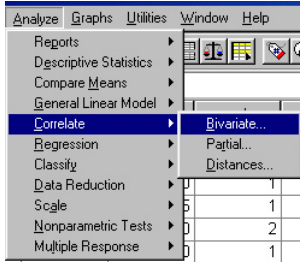

This will get you to the ANOVA menu. Place the dependent variable into the "Dependent Variable" box. The independent variable(s) go into the "fixed factors" box.
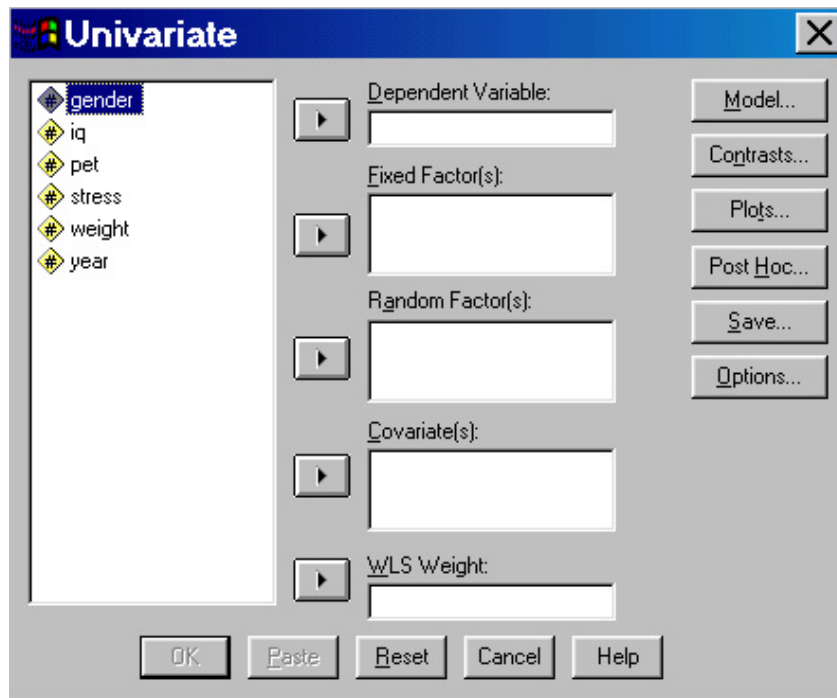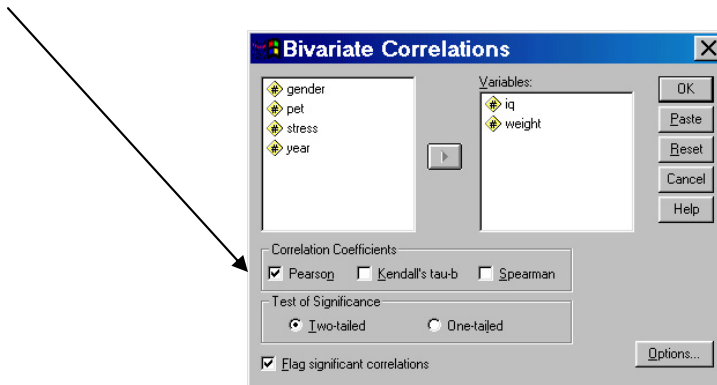
Section 7 – Correlation and Regression

Correlation
Click one time on "Analyze", go to "Correlate", and select "Bivariate".

Once in the Bivariate menu, place the variable of interest into the "Variable(s)" box. Two or more variable may be placed into this box. Make sure to select the Pearson Correlation Coefficient and select whether it is a one- or two-tailed test.

By clicking on the "Options" button, you can get means and standard deviations for each of your variables.
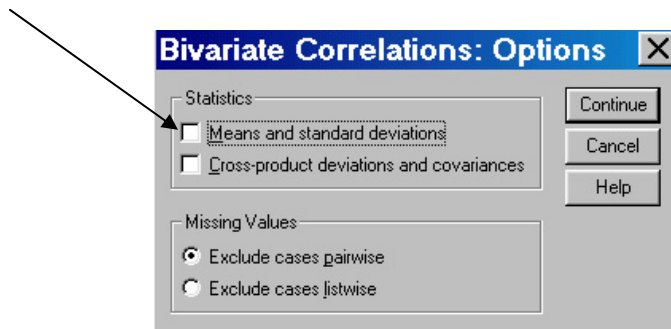
The correlation output will look like:

Correlations

Correlations

|		IQ	WEIGHT
IQ	Pearson Correlation	1	.067
	Sig. (2-tailed)	.	.777
	N	20	20
WEIGHT	Pearson Correlation	.067	1
	Sig. (2-tailed)	.777	.
	N	20	20

The output will contain the Pearson correlation (r value), the alpha value, and the number of subjects for each correlation. In this example, r = .067, with an alpha of .777. Not surprisingly, there is no significant correlation between a person's weight and their IQ.

<u>Regression</u>
Click one time on "Analyze", go to "Regression", and select "Linear".
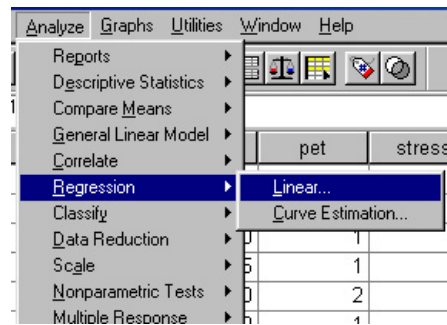
From the linear regression menu, place the dependent variable (Y variable) in the "Dependent" box and the independent variable (X variable) in the "Independent(s)" box. Ignore the rest of the menu options.
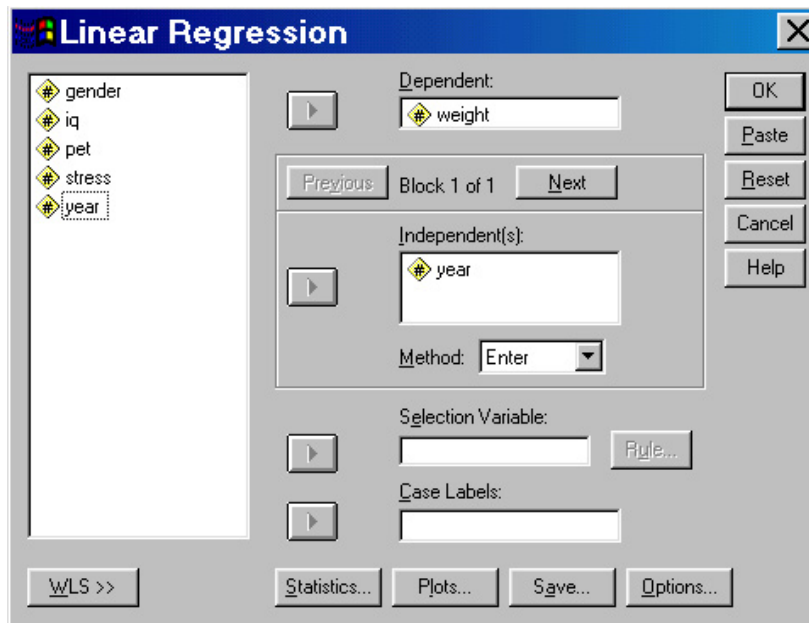
The output will look like:

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.010[a]	.000	-.055	39.766

a. Predictors: (Constant), YEAR

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.013	1	3.013	.002	.966[a]
	Residual	28464.187	18	1581.344		
	Total	28467.200	19			

a. Predictors: (Constant), YEAR

b. Dependent Variable: WEIGHT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	165.377	20.852		7.931	.000
	YEAR	.350	8.026	.010	.044	.966

a. Dependent Variable: WEIGHT

The coefficients box provides the slope and y-intercept necessary to create a regression equation.

Predicted Y = slope * X + Y-intercept

(Constant) B is the Y-intercept (in this example 165.377)
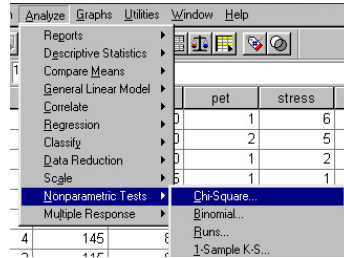The independent variable's B = the slope (in this example Year B = .350)
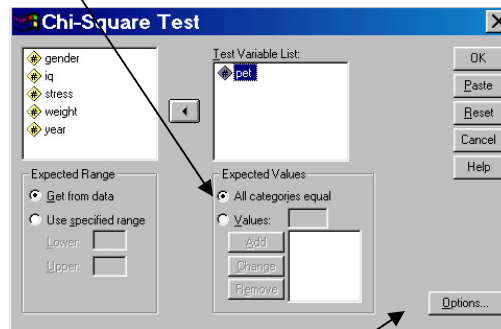
Y = (.350) (X) + 165.377

Section 8 – Chi-Square

No Preference Chi-Square
Click one time on "Analyze", go to "Nonparametric Tests", and select "Chi-Square".

In the Chi-Square menu, place the variable of interest into the "Test Variable List" box. Make sure "All categories equal" is selected.

To obtain descriptive statistics, click on the "Options" button and select "Descriptives".

Chi-Square output from SPSS:

Chi-Square Test

Frequencies

PET

	Observed N	Expected N	Residual
yes	9	10.0	-1.0
no	11	10.0	1.0
Total	20		

Test Statistics

	PET
Chi-Square [a]	.200
df	1
Asymp. Sig.	.655

The chi-square value is .200 with an alpha value of .655, indicating there are no significant frequency differences between groups.

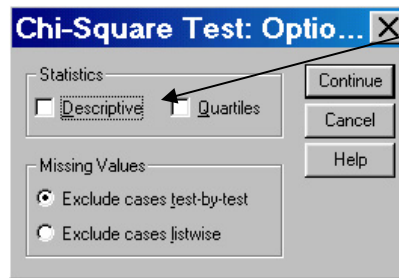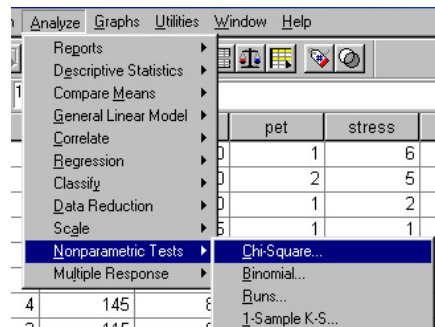No Difference From Population Chi-Square

Click one time on "Analyze", go to "Nonparametric Test", and select "Chi-Square".

From the Chi-Square menu place the variable of interest into the "Test Variable List" box. The range of values possible should be entered in the "Expected Range" box. In this example, the total range possible is from 1 to 4. Next, select "Values" from the "Expected Values" box. Population frequency values must be entered **for each categorical** value. It is important that you enter the values in the same categorical order as is in the "Expected Range" box. For category 1 (freshmen), we have one freshman in our population, 2 sophomores, 8 juniors, and 10 seniors.
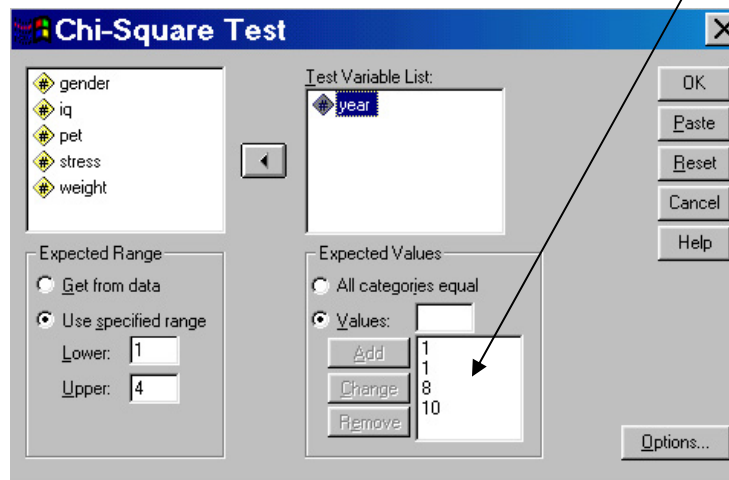
The Chi-Square output from SPSS:

Chi-Square Test

Frequencies

		YEAR		
	Category	Observed N	Expected N	Residual
1	freshman	6	1.0	5.0
2	sophomore	5	1.0	4.0
3	junior	5	8.0	-3.0
4	senior	4	10.0	-6.0
Total		20		

Test Statistics

|	YEAR
Chi-Square a	43.61
df	3
Asymp. Sig.	.000

a. 2 cells (50.0%) have expected frequencies less than 5. The minimum expected cell frequency is 1.0.

The Chi-Square value is equal to 43.61 and the alpha is equal to .000, indicating a significant difference from the population.

Chi-Square Test for Independence
Click one time on "Analyze", go to "Descriptive Statistics", and select "Crosstab".
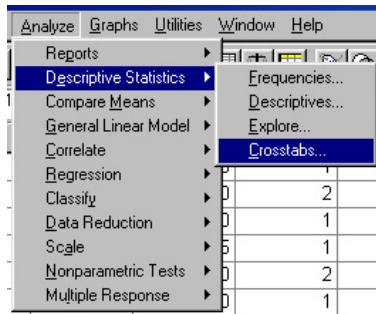
In the crosstabs menu place one variable in the "Row(s)" box and the second variable in the "Columns" box.
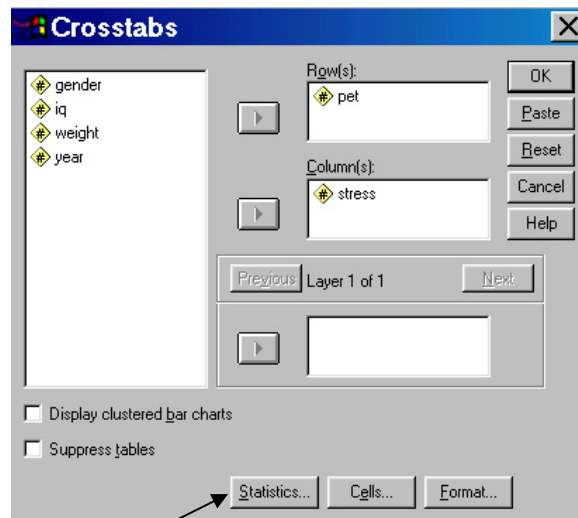
Click one time on the "Statistics" button at the bottom of the menu. In the statistics menu select "Chi-Square"

The Chi-Square output from SPSS:

Crosstabs

Case Processing Summary

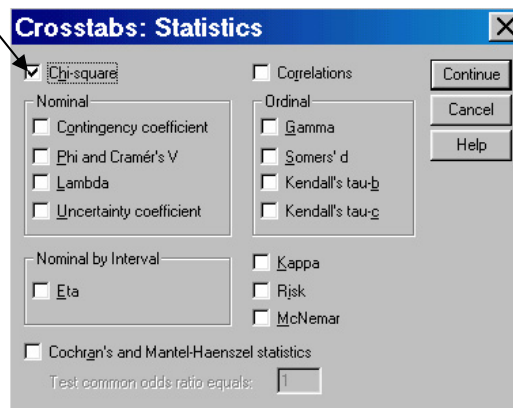|	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
PET * STRESS	20	100.0%	0	.0%	20	100.0%

PET * STRESS Crosstabulation

Count

		STRESS						
		no stress	low stress	med low stress	med stress	med high stress	high stress	extreme stress
PET	yes	4	2	2			1	
	no			1	2	4	2	2
Total		4	2	3	2	4	3	2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14.613ᵃ	6	.023
Likelihood Ratio	19.887	6	.003
Linear-by-Linear Association	10.227	1	.001
N of Valid Cases	20		

The Chi-Square value equals 14.613, with an alpha of .023, indicating a significant difference between groups, or a significant correlation between owning a pet and the amount of stress experienced.

Expectations for Completing Labs

By completing labs, you are conveying that you understand how to use SPSS to compute requested statistics. Thus, not only are you expected to include all necessary output to support your answer, you should also know what is NOT necessary and exclude any extraneous output.

How will I know what is necessary?

Any answer given in a lab assignment, unless otherwise specified, needs output to support the answer. If you have output that is not used in an answer, it is unnecessary and points will be taken off for unnecessary output.

How do I get rid of unnecessary output?

There are three ways to exclude unnecessary output in any assignment.

1. Make sure the only statistics selected are the requested statistics (i.e., the ones you want). Unfortunately, SPSS retains previous settings, and therefore, you may inadvertently get output you do not want. It is best to check that you not only select the statistics you DO want, but also click OFF those statistics you do not want.

2. If you complete the analyses and discover unnecessary calculations in the output, you can delete those calculations while in SPSS (before printing). Simply click once anywhere on the unwanted output. It will be highlighted, then hit the delete button on your keyboard. This will remove the selected output. Be sure to select only the output you want to delete, so you don't lose output you need to answer your lab questions.

3. If you have already printed out your output and notice unnecessary calculations, merely cross them out by hand and write "unnecessary" next to it, indicating that you are aware that this output is not needed to answer the question.

Other Expectations

- All answers on the output must be clearly marked with the number of the question next to the appropriate calculation to indicate you understand where the answer came from.

- You may hand write written responses on the output pages, or double-click on the output and type answers in. Either method is acceptable.

- This manual contains information on using SPSS to calculate statistics. The lab assignments will not only require SPSS calculations, but also information from class about statistics to obtain a complete answer. In other words, you also need to know how to interpret the statistics and output that you receive from SPSS.