

Statistics Outreach Center Short Course

**Introduction to SPSS**

Wednesday, July 1, 2009  
3:00 – 5:00 pm  
N106 LC

---

**Topics Covered:**

- SPSS Windows and Files
  - Inputting Data
  - Descriptive Statistics
  - Statistical Graphics
  - Advanced Data Techniques
  - Inferential Statistics, including:
    - Chi-square and T-tests
    - One-way ANOVA
    - Correlations
    - Regression
  - Exporting Output to Other Software
  - Some Further SPSS Resources
- 

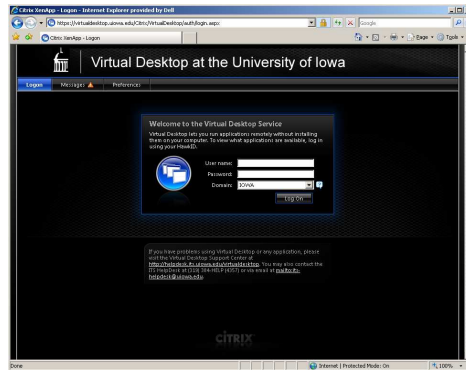
**Overview**

This course is designed for beginning SPSS users, providing a basic introduction to SPSS through the topics listed above. The first sections introduce users to the SPSS for Windows environment and discuss how to create or import a dataset, transform variables, and calculate descriptive statistics. The remaining sections describe some commonly used inferential statistics, graphical display of output, and other related topics. During this tutorial, a sample dataset, *employee data.sav*, is used for all examples. This example dataset is provided with recent versions of SPSS. It can be found in the root SPSS directory.

Although we will be using copies of SPSS that are installed on the lab computers, the University also allows you to access SPSS through the Virtual Desktop. Virtual Desktop lets you run applications remotely without installing them on your computer. Virtual Desktop is available at:

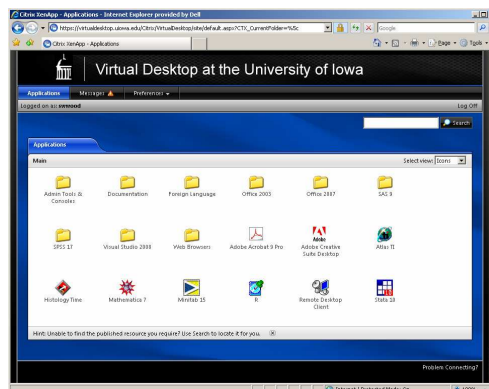
<https://virtualdesktop.uiowa.edu>

1. Log in using your HawkID at <https://virtualdesktop.uiowa.edu>.

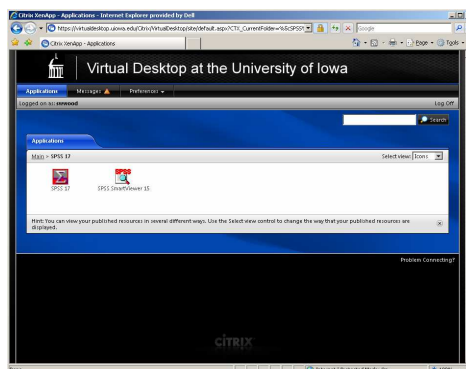


2. Click *End detection process*.

3. Choose the SPSS 17 folder from the “Main” list of applications.



4. Click the SPSS 17 icon.

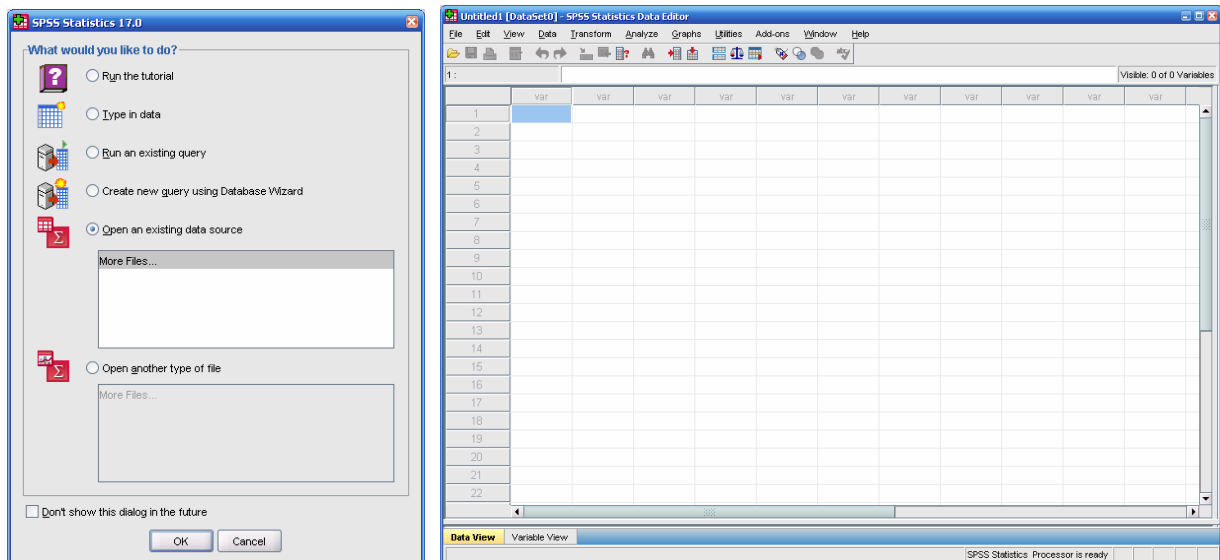


**Notes:**

- On some University computers, SPSS is available without going through the Virtual Desktop. To start SPSS, go to the *Start* icon on your Windows computer. You should find an SPSS icon under the *Programs* menu item.
- When using the Virtual Desktop to access SPSS, you can only open and save files from your University of Iowa personal drive (the H: drive).
- When using the Virtual Desktop, a dialog box may appear asking for read/write access. To my knowledge, you can simply click OK with the defaults.

When the program opens, it will present you with a “What would you like to do?” dialog box.

For now, click the **Cancel** button.



## Section 1: SPSS Windows and Files

### 1.1. SPSS Windows and Files

The *Data Editor*, *Output Viewer*, and *Syntax Editor* windows are the most frequently used in analyzing data in SPSS. Basic information on these files is summarized below.

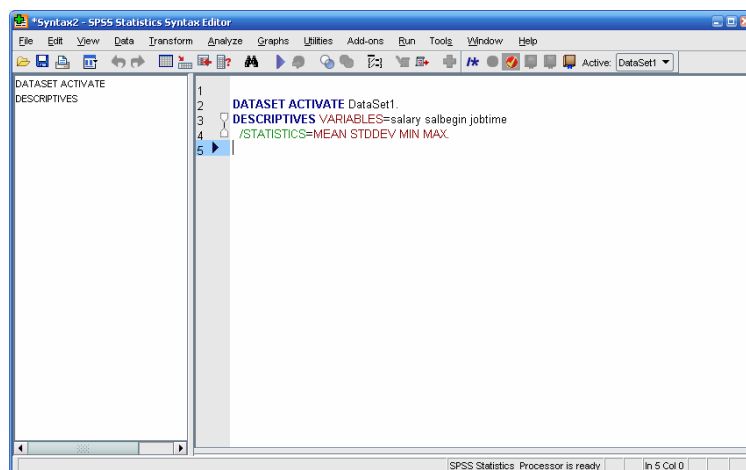
Windows	File Suffix	Function
Data Editor	.sav	Define, enter, and edit data and run statistical analyses
Output Viewer	.spv	Contain the results of all statistical analyses and graphical displays of data.
Syntax Editor	.sps	Compose SPSS commands and submit them to the SPSS processor. This window is activated when you click on the <i>Paste</i> function.

Note: Output Viewer files that were created with previous versions of SPSS (such as SPSS 15) can not be opened in SPSS 17. The program SPSS SmartViewer 15 (available on the Virtual Desktop) can be used to open these older output files (old Output Viewer files end with .spo). Likewise, SPSS 17 output files can not be opened in older versions of SPSS.

The *Syntax Editor* allows you to keep a record of what analyses you conduct during an SPSS session. It is highly recommended that you work with the Syntax features (using the Paste button) as you work and saving your Syntax file before exiting SPSS. This way, you have a log of the analyses conducted during an SPSS session

To run all the syntax from the *Syntax Editor*, use the menu option **Run > All**.

To run a portion of the syntax, highlight the selection and use **Run > Selection**.



## Section 2: Creating and Importing Datasets

### 2.1. Entering Data

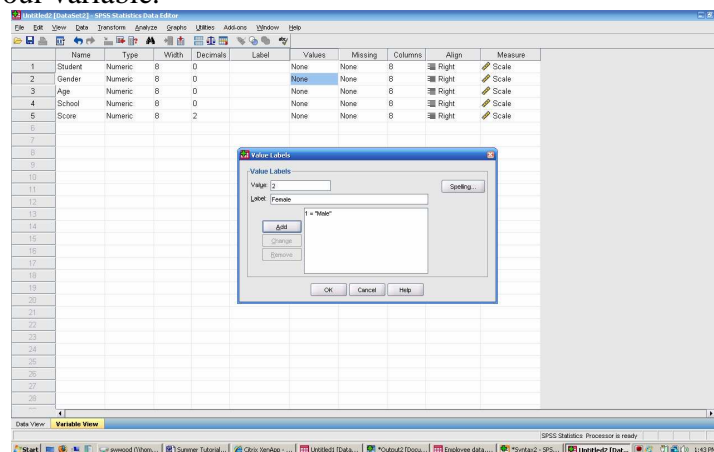
Assume that you collected the following information from a random sample of 100 students.

Student	Gender	Age	School	Test Score
1	Male	12	School A	56.64
2	Female	11	School B	60.83
3	Male	10	School C	44.25
4	Male	12	School A	49.13
5	Female	11	School C	30.67
...	...	...	...	...
98	Male	10	School B	32.45
99	Female	11	School C	45.67
100	Female	13	School B	55.12

As you type in your data, you will need to keep in mind each of the following.

- Create a separate line for each student.
- Create a column for each variable of interest. In this example, we will use five columns, one for each of the following variables (student, gender, age, school, test score).
- Develop a numerical code for the gender and school variables. For the gender variable, we will assign the value of 1 to Male and the value of 2 to Female. For the school variable, we will assign the value of 1 to School A, the value of 2 to School B, and 3 to School C.

Step 1. Define Variables and their characteristics in the **Variable View** window. To create the label of “Male” for a value of 1 and “Female” for a value of 2, click on **Label** to provide a label for your variable.



Step 2. After defining all the variables, click on the **Data View** window (lower left hand corner). You should see all of the variable names that you have entered.

## 2.2. Variable View Information

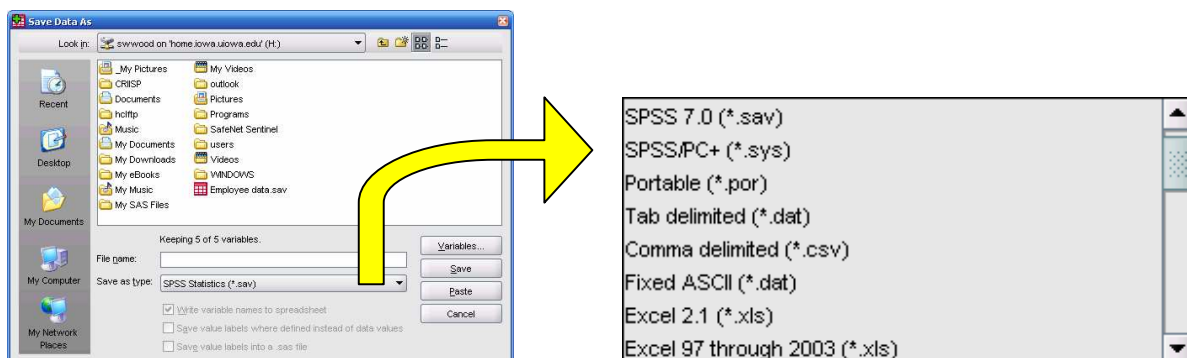
Most of the information in the Variable View window can be left alone. However, to get the most out of your data set (and to make things easier when running analyses), it is a good idea to fill out the Variable View “spreadsheet”. The following is a list of what each column of Variable View means.

- **Name:** The name of your variable. It can be up to 64 characters long. Variable names cannot contain spaces or most symbols (punctuation, etc). Names must start with a letter.
- **Variable Type:** You can format the way that the numbers look using the variable type option (click on the ... button). Variable types include numeric, comma, dot, scientific notation, date, dollar, custom currency, and string.
- **Width:** The length of the string, in characters, or the number of places before the decimal point in a number.
- **Decimals:** The number of places after the decimal point to include.
- **Label:** Use this option to write a description of what the variable means. Good for surveys, where the original question can be placed here.
- **Values:** If a variable is discrete (contains a finite number of possible values), then you can define the values using this option. Great for gender and yes/no variables, since it is easy to forget how you defined these variables numerically.
- **Missing:** If a particular value (such as 999 or -999) is defined as a missing value, you can inform SPSS through this option.
- **Columns:** Total length of the variable value to be displayed (in characters).
- **Align:** How you would like the values to be aligned in the Data View.
- **Measure:** What kind of variable do you have? Is it a “Nominal” variable (where values are simply un-order-able labels), “Ordinal” (where the values are discrete, yet order-able), or “Scale” (where the values are continuous, or can be interpreted as continuous)?

## 2.3. Saving an SPSS file

To save the data that you created, choose the following menu options from the menu in the Data Editor window: **File > Save As...**

You can save a copy of the dataset in another format (e.g., Excel, ASCII, etc) by selecting the format from the list of the *Save as type* menu:

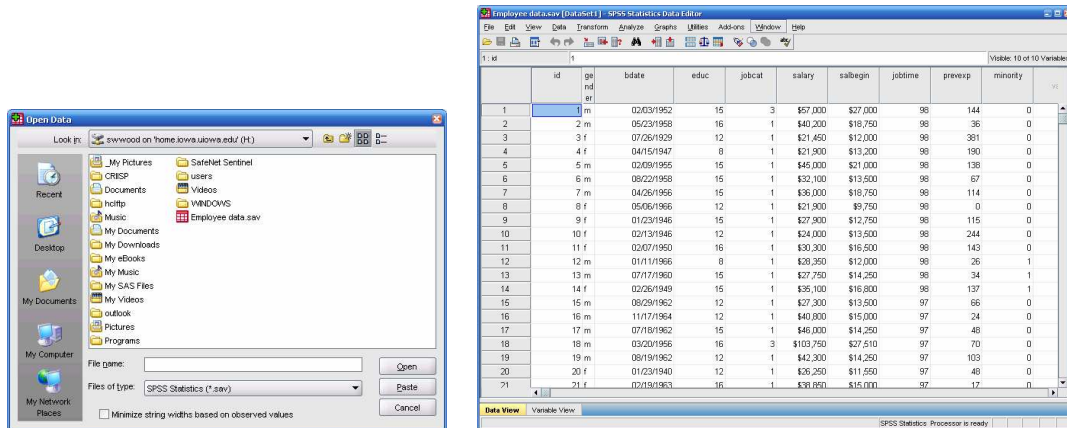


## 2.4. Opening an Existing Dataset

An existing data file can be opened in the Data Editor. From the Data Editor window, choose the following menu options:

**File > Open > Data...**

The *Open File* dialog box should automatically open to the SPSS directory of example files. Choose *employee data.sav* from the list and click **Open**.

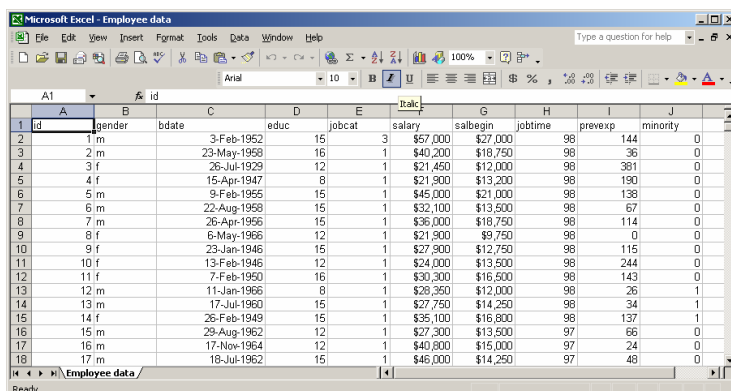


## 2.5. Opening an Excel Spreadsheet

Sometimes, data are set up in the form of an Excel spreadsheet. SPSS can easily read in data from Excel and reformat it as an SPSS data set. From the menu in the Data Editor window, choose the following menu options:

**File > Open > Data...**

The *Open File* dialog box defaults to SPSS data file types. To open Excel files, change the *Files of Type* option to “*Excel (\*.xls)*”. After selecting your Excel file, click **Open**. A dialog box similar to the one below will appear to confirm your selection.

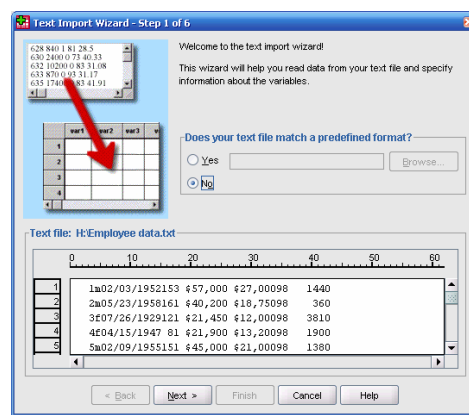


## 2.6. Opening Data from a Text File

Very large text files will often be saved as a text (\*.txt) document (or sometimes a \*.dat or a \*.csv document). You can import these data sets into SPSS using the Text Import Wizard. From the menu in the Data Editor window, choose the following menu options:

**File > Open > Data...**

The *Open File* dialog box defaults to SPSS data file types. To open text files, change the *Files of Type* option to “Text Documents (\*.txt)”. After selecting your text file, click **Open**. The Text Import Wizard will guide you through the process.



The 6 steps of the Text Import Wizard are:

1. Does your text file match a predefined format? Typically, the answer is **No**.
2. How are your variables arranged? Are variable names included on the top of your file?
3. The first case of data begins on which line number? How are your cases represented (one per line)? How many cases do you want to import?
4. Which delimiters appear between variables? (Delimited format) Where are the breakpoints between variables? (Fixed-Width format)
5. What are the variable names and formats?
6. Would you like to paste the syntax?

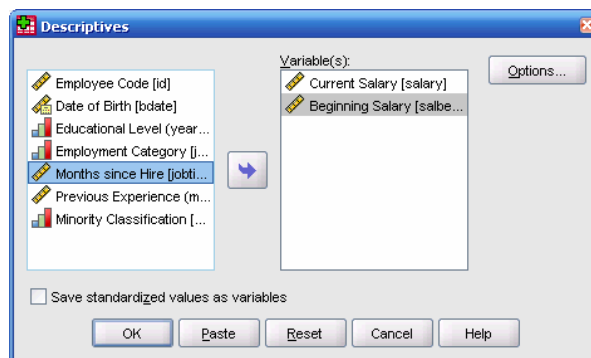


### Section 3: Describing Your Data (Numerically)

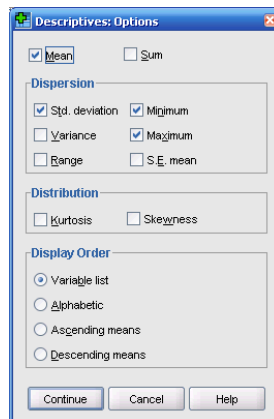
#### 3.1. Descriptive Statistics (Means, Standard Deviations, etc. for Continuous Variables)

Summary (or descriptive) statistics are available under the *Descriptives* option available from the *Analyze* and *Descriptive Statistics* menus:

**Analyze > Descriptive Statistics > Descriptives...**



Select the variables you are interested in and click the “triangle” button to move the variables place. To view the available descriptive statistics, click on the button labeled **Options**. This will show the following dialog box:



After selecting the statistics you would like, click **Continue**. Output can be generated by clicking on the **OK** button in the *Descriptives* dialog box. Statistics will be in the Output Viewer.

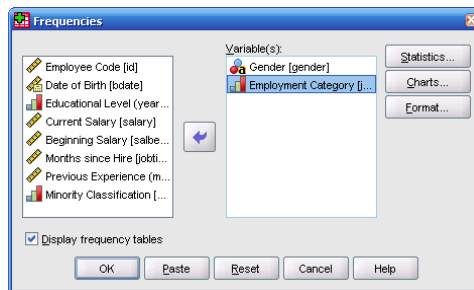
**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
Current Salary	474	\$15,750	\$135,000	\$34,419.57	\$17,075.661
Beginning Salary	474	\$9,000	\$79,980	\$17,016.09	\$7,870.638
Valid N (listwise)	474				

**3.2. Frequencies (For Discrete Variables)**

The **Frequencies** option allows you to obtain the number of people within each employment category in the dataset. The **Frequencies** procedure is found under the **Analyze** menu:

**Analyze > Descriptive Statistics > Frequencies...**



Click on the “**Statistics...**” button to see what descriptive statistics are available. Note that percentiles and descriptive statistics can be calculated in the **Frequencies** menu.

The example in the above dialog box would produce the following output:

**Gender**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	216	45.6	45.6	45.6
	Male	258	54.4	54.4	100.0
	Total	474	100.0	100.0	

**Employment Category**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Clerical	363	76.6	76.6	76.6
	Custodial	27	5.7	5.7	82.3
	Manager	84	17.7	17.7	100.0
	Total	474	100.0	100.0	

To create contingency tables, you can use the **Crosstabs** menu.

**Analyze > Descriptive Statistics > Frequencies...**

Contingency tables will be used later in the tutorial.

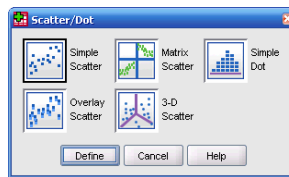
## Section 4: Describing Your Data Graphically

The easiest way to produce graphs in SPSS is to use the **Legacy Dialogs**. For additional formatting options, consider using the **Graphs > Interactive** menu.

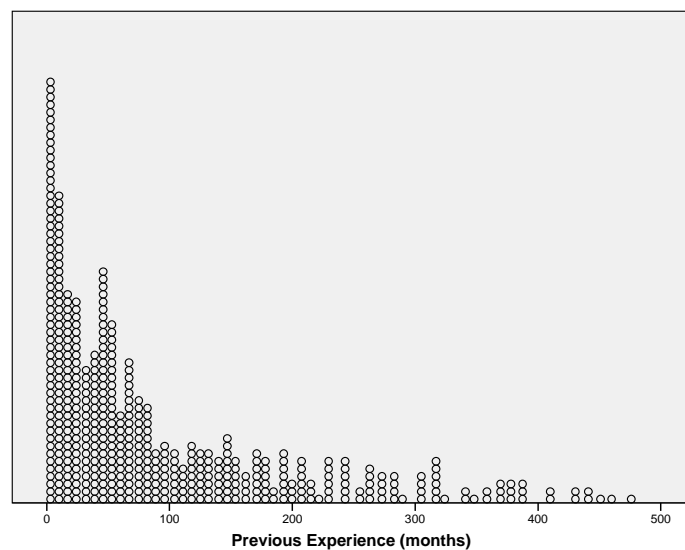
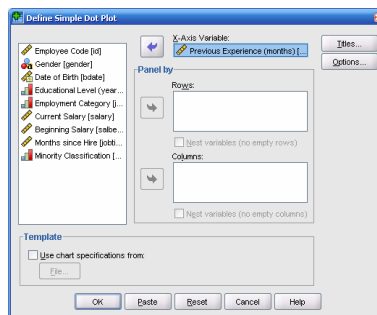
### 4.1. Dot Plot

A dot plot provides a simple way of visualizing your data. Let's look at "Previous Experience" using a dot plot. **Graphs > Legacy Dialogs > Scatter/Dot...**

We will use the "Simple Dot" option.



Add the previous experience variable to the "X-Axis Variable" box using the triangle button.



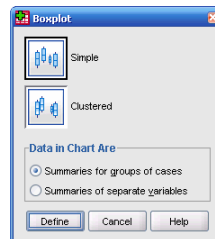
## 4.2. Box Plot

A box plot uses quantiles to help visualize the data. Although some information is not considered, box plots are convenient, compact ways of summarizing data. In SPSS, a categorical variable is required, along with a continuous variable.

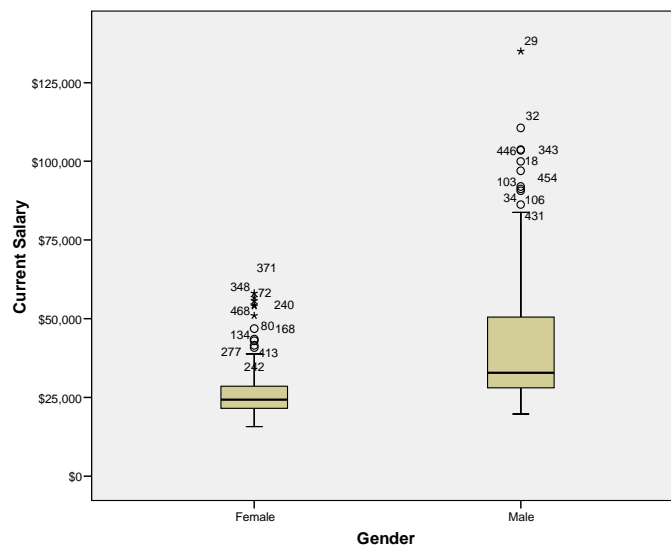
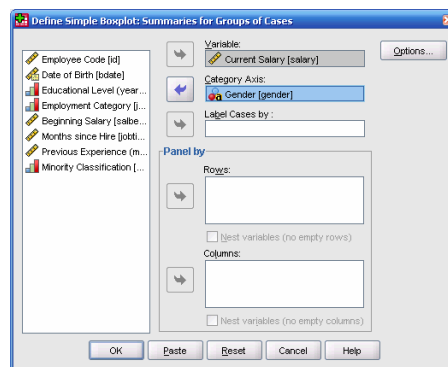
Let's look at the distribution of salary for males and females. A separate box plot will be created for males and females.

### Graphs > Legacy Dialogs > Boxplot...

A simple box plot will suffice.



Add the "Salary" variable to the *Variable* box; add the "Gender" variable to the *Category Axis* box. Click **OK** to continue.

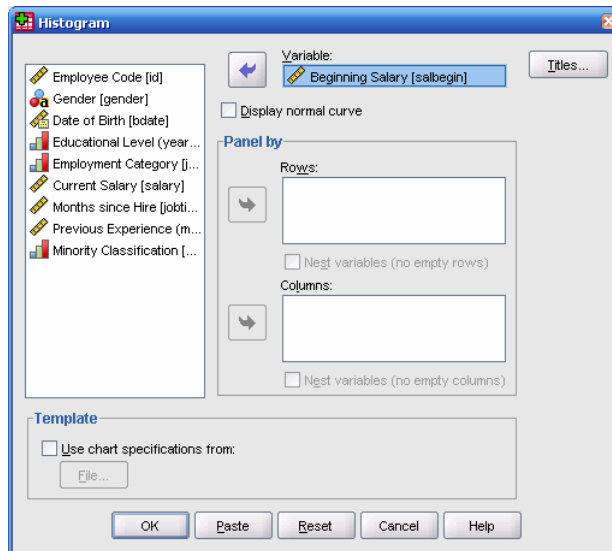


### 4.3. Histogram

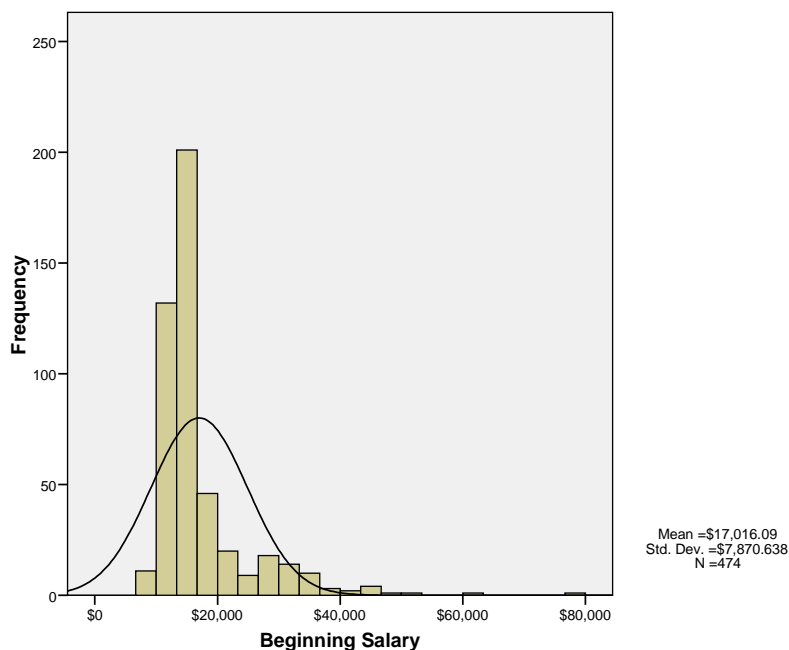
Histograms are commonly used with larger data sets. Histograms can help an analyst quickly determine if data is normally distributed or if the data are skewed.

Let's create a histogram of the Beginning Salary of the employees.

#### Graphs > Legacy Dialogs > Histogram



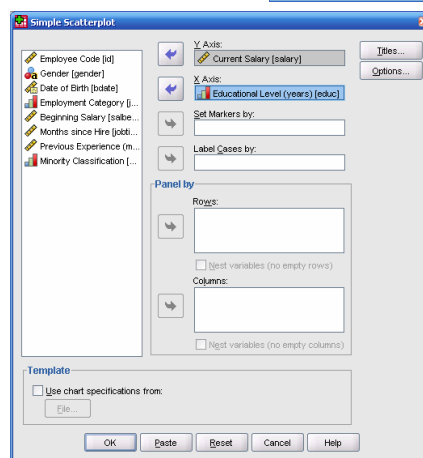
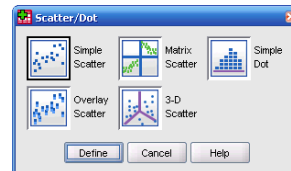
Input “*Beginning Salary*” into the *Variable* box using the “Triangle” button. To see if the data are normal, let's also check the box next to “*Display Normal Curve*”. Click **OK**.



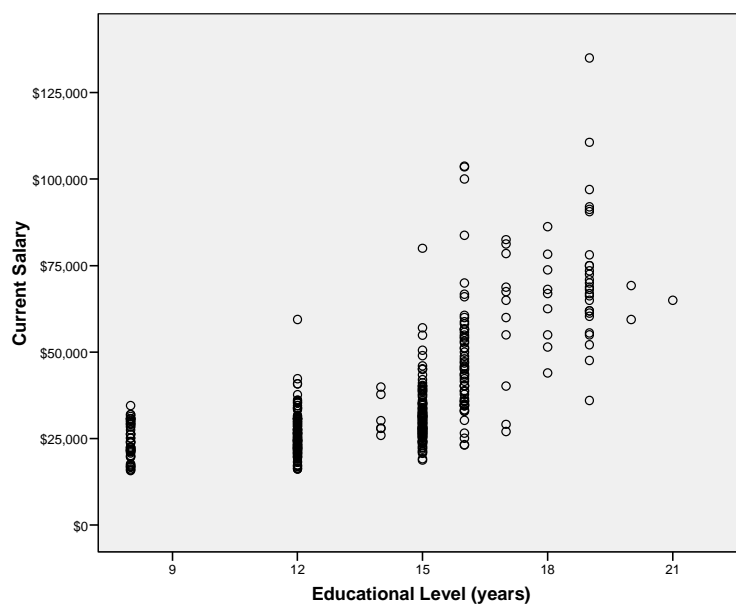
### 4.4. Scatter Plot

Prior to conducting a correlation analysis, it is advisable to plot the two variables to visually inspect the relationship between them. To produce scatter plots, select the following menu option:

**Graphs > Legacy Dialogs > Scatter/Dot...**



This plot indicates that there is a positive, fairly linear relationship between current salary and education level. In order to test whether this apparent relationship is statistically significant, we could run a correlation.



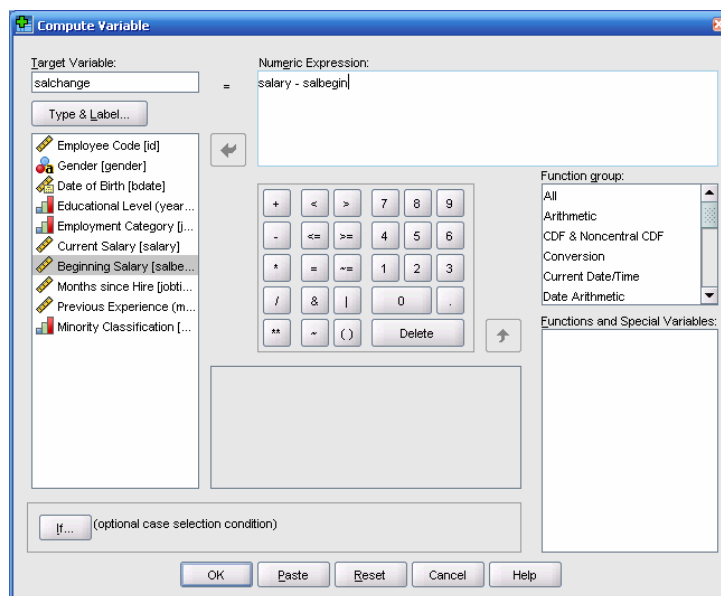
## Section 5: Advanced Data Tools

### 5.1. Compute Variables

New variables can be created using the *Compute* option available from the menu in the Data Editor:

#### Transform > Compute Variables...

To create a new variable, type its name in the box labeled *Target Variable*. The expression defining the variable being computed will appear in the box labeled *Numeric Expression*.



This new variable, *salchange*, will appear in the rightmost column of the working dataset.

### 5.2. Recode Variables

Often, it may be necessary to take a continuous variable and turn it into a categorical variable. For example, let's say that we want to take the salary variable and define a new variable as follows:

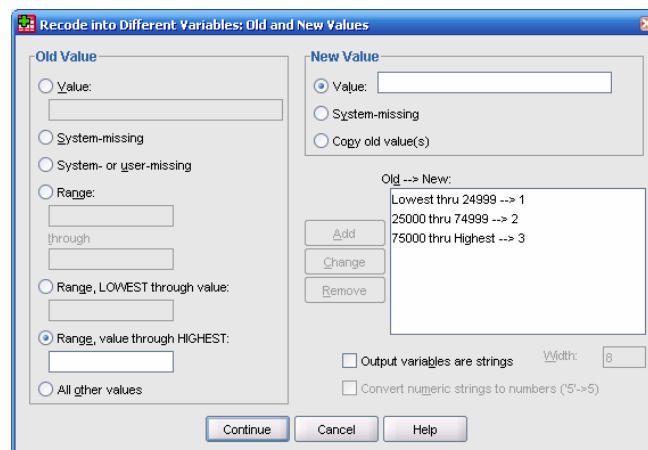
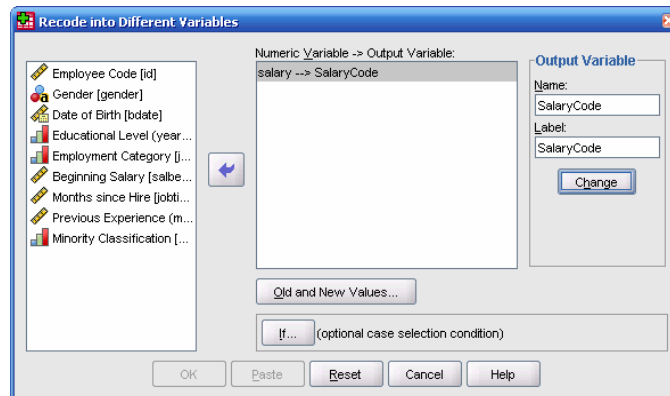
- 1 = Employees whose salaries are less than \$25,000
- 2 = Employees whose salaries are between \$25,000 and \$75,000
- 3 = Employees whose salaries are greater than \$75,000

To do this, we need to go to the *Recode into Different Variables* menu.

## Transform > Recode into Different Variables...

Here are a rundown of steps necessary to do the recoding defined above.

1. Select the “*Salary*” variable and click the “Triangle” button.
2. Type in a name and label for the new variable. I chose “*SalaryCode*” for both. Click the **Change** button.
3. Click **Old and New Values...**
4. Click “*Range, LOWEST through value*” and type in 24999. Under the New Value listing, click “*Value*” and type in 1. Click the **Add** button to add to the list.
5. Repeat, using the “*Range*” option and type in 25000 and 74999. Under the New Value listing, click “*Value*” and type in 2. Click the **Add** button to add to the list.
6. Click “*Range, value through HIGHEST*” and type in 75000. Under the New Value listing, click “*Value*” and type in 3. Click the **Add** button.
7. Click **Continue** and **OK** to create the new variable.



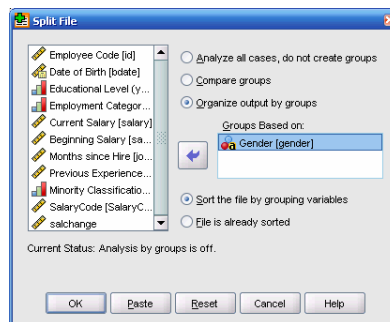


### 5.3. Split File

Often, an analyst wants to run separate tests or graphs for different subgroups. Example of this would include running descriptive statistics by gender or by level of education. This section will show how you can split a file so that separate analyses are run. For this example, we will split the file by gender.

#### Data > Split File

Click on the “*Organize output by groups*” option and select “*gender*” using the triangle button. Click **OK**.



To test this out, let’s run descriptive statistics on the variables “*Beginning Salary*” and “*Months since Hire*”.

#### Gender = Female

Descriptive Statistics<sup>a</sup>

	N	Minimum	Maximum	Mean	Std. Deviation
Beginning Salary	216	\$9,000	\$30,000	\$13,091.97	\$2,935.599
Months since Hire	216	63	98	80.38	9.676
Valid N (listwise)	216				

a. Gender = Female

#### Gender = Male

Descriptive Statistics<sup>a</sup>

	N	Minimum	Maximum	Mean	Std. Deviation
Beginning Salary	258	\$9,000	\$79,980	\$20,301.40	\$9,111.781
Months since Hire	258	63	98	81.72	10.351
Valid N (listwise)	258				

a. Gender = Male

Remember to change the option back to “**Analyze all cases, do not create groups**” when you are done!!

## Section 6: An Introduction to Inferential Statistics in SPSS

### 6.1. T test

There are three types of *t* tests; the options are all located under the *Analyze* menu item:

- Analyze > Compare Means > One-Sample T test...**
  - > **Independent-Samples T test...**
  - > **Paired-Samples T test...**

One-Sample T Test: Used to test if the mean of a single variable equals a pre-defined number.

$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$$

Independent-Samples T Test: Used to test if the mean of two variables are equal.

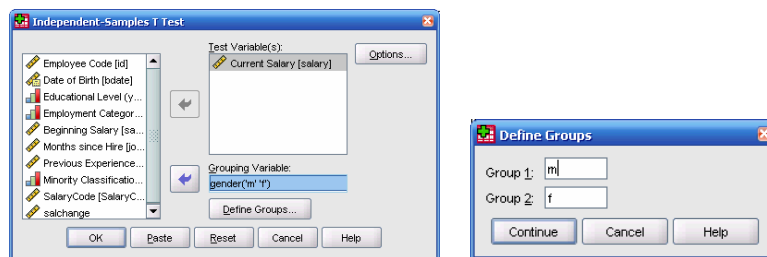
$$H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2$$

Paired-Samples T Test: Used to test if the mean difference between two dependent samples is 0.

$$H_0: d = 0 \quad H_1: d \neq 0$$

Example: Does the population mean for salary differ for males and females?

Here, we will use an **Independent-Samples T Test**. When current salaries for male and female groups are being compared in this analysis, the *salary* and *gender* variables are selected for *Test Variable* and *Grouping Variable* boxes, respectively.



Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Current Salary	Male	258	\$41,441.78	\$19,499.214	\$1,213.968
	Female	216	\$26,031.92	\$7,558.021	\$514.258

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	95% Confidence Interval of the Difference	
Current Salary	Equal variances assumed	119.669	.000	10.945	472	.000	\$15409.86	\$1,407.906	\$12,643.322	\$18,176.401
	→	Equal variances not assumed			11.688	344.262	.000	\$15409.86	\$1,318.400	\$12,816.728

In the above example, in which the null hypothesis is that males and females do not differ in their salaries, the  $t$  statistic under the assumption of unequal variances has a value of 11.688, and the degrees of freedom has a value of 344.262 with an associated significance level of .000. The significance level tells us that the hypothesis of no difference is rejected under the .05 significance level. Accordingly, we conclude that males and females differ in their salaries.

## 6.2. One-Way ANOVA

There are two ways of comparing group differences for one or more independent and dependent variables in SPSS: the *One-Way ANOVA* and *General Linear Model (GLM)* procedures in the *Analyze* menu. For this section, we will use the *One-Way ANOVA* procedure.

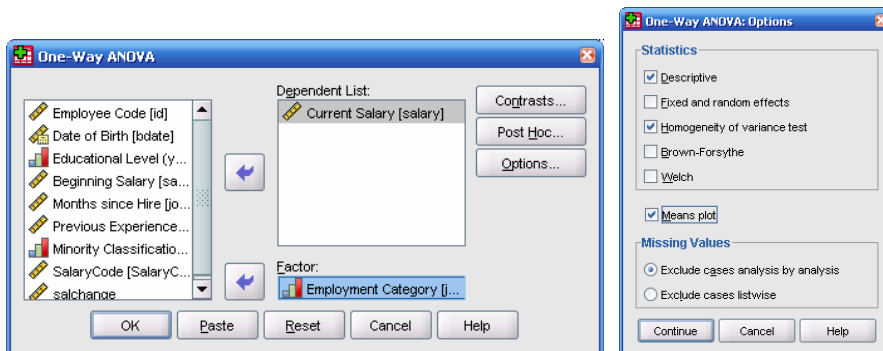
### Analyze > Compare Means > One-Way ANOVA...

One-Way ANOVA: Used to test if the population means of two or more variables are equal.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k \quad H_1: \text{At least one } \mu_i \neq \mu_j$$

Example: Does the population mean for current salary differ by employment category?

To conduct the one-way ANOVA, first select the independent and dependent variables to produce the following dialog box:



The above options will produce the following output (some output is omitted):

Test of Homogeneity of Variances

Current Salary

Levene Statistic	df1	df2	Sig.
59.733	2	471	.000

ANOVA

Current Salary

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.9E+010	2	4.472E+010	434.481	.000
Within Groups	4.8E+010	471	102925714.5		
Total	1.4E+011	473			

In the above example in which the hypothesis is that three categories of employment do not differ in their salaries, the  $F$  statistic has a value of 434.481 with the associated significance level of .000 (Technically, the  $p$ -value is less than 0.001). The significance level tells us that the hypothesis of no difference among three groups is rejected under the .05 significance level. Accordingly, we conclude that the three groups of employment (*Clerical, Custodial, and Manager*) differ in their salaries.

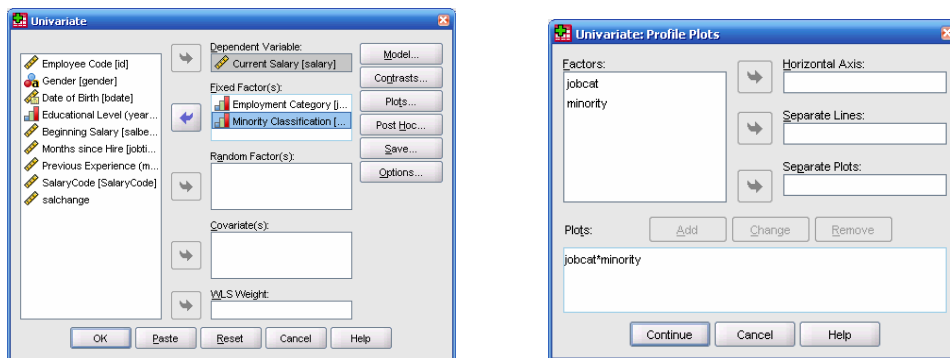
### 6.3. Two-Way ANOVA

When there are more than two independent variables, the analysis is done by selecting *General Linear Model (GLM)* procedures in the *Analyze* menu.

Analyze > **General Linear Model** > **Univariate...**

Example: Is current salary dependent on minority and employment category?

For this example, the Dependent Variable is **Current Salary** (*salary*) and the Fixed Factors are **Minority** (*minority*) and **Employment Category** (*jobcat*). The screen will appear as follows.



Between-Subjects Factors

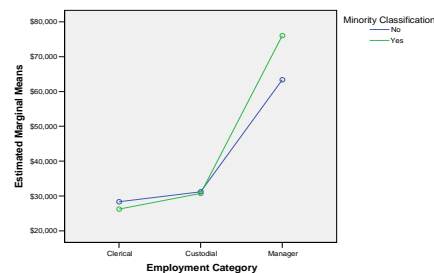
		Value Label	N
Employment Category	1	Clerical	363
	2	Custodial	27
	3	Manager	84
Minority Classification	0	No	370
	1	Yes	104

Tests of Between-Subjects Effects

Dependent Variable: Current Salary					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9.034E+010 <sup>a</sup>	5	1.807E+010	177.742	.000
Intercept	1.537E+011	1	1.537E+011	1511.773	.000
jobcat	2.596E+010	2	1.298E+010	127.699	.000
minority	237964814	1	237964814.4	2.341	.127
jobcat * minority	788578413	2	394289206.5	3.879	.021
Error	4.757E+010	468	101655279.9		
Total	6.995E+011	474			
Corrected Total	1.379E+011	473			

a. R Squared = .655 (Adjusted R Squared = .651)

Estimated Marginal Means of Current Salary



Taking a significance level of .05, the *jobcat* main effect and the *jobcat by minority interaction* are significant. The change in the simple main effect of one variable over levels of the other is most easily seen in the graph of the interaction. If the lines describing the simple main effects are not parallel, then a possibility of an interaction exists. The presence of an interaction was confirmed by the significant interaction in the summary table.

### 6.4. Correlations

To obtain a bivariate correlation, choose the following menu option:

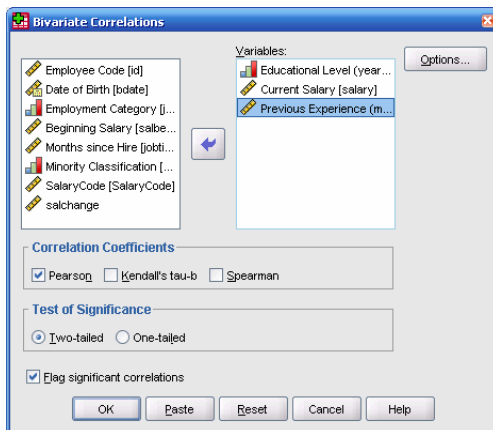
**Analyze > Correlate > Bivariate...**

Test for a Correlation: Used to test if the correlation between two variables is 0.

$$H_0: \rho = 0 \quad H_1: \rho \neq 0$$

Example: What are the correlations between education level, current salary, and previous experience? Do any of these correlations differ from zero?

With the selections shown below, we would obtain correlations between *Education Level* and *Current Salary*, between *Education Level* and *Previous Experience*, and between *Current Salary* and *Previous Experience*. We will maintain the default options shown in the above dialog box in this example. The selections in the above dialog box will produce the following output:



**Correlations**

		Educational Level (years)	Current Salary	Previous Experience (months)
Educational Level (years)	Pearson Correlation	1.000	.661**	-.252*
	Sig. (2-tailed)	.	.000	.000
	N	474	474	474
Current Salary	Pearson Correlation	.661**	1.000	-.097*
	Sig. (2-tailed)	.000	.	.034
	N	474	474	474
Previous Experience (months)	Pearson Correlation	-.252*	-.097*	1.000
	Sig. (2-tailed)	.000	.034	.
	N	474	474	474

\*\* . Correlation is significant at the 0.01 level (2-tailed).  
\* . Correlation is significant at the 0.05 level (2-tailed).

The above output shows there is a positive correlation between employees' number of years of education and their current salary. This positive correlation coefficient (.661) indicates that there is a statistically significant ( $p < .001$ ) linear relationship between these two variables such that the more education a person has, the larger that person's salary is. Also observe that there is a statistically significant ( $p < .001$ ) negative correlation coefficient for the association between education level and previous experience (-.252) and for the association between employee's current salaries and their previous work experience.

## 6.5. Regression

Regression models can be used to help estimate a (dependent) variable based on information from other (independent) variables.

Overall Model Fit (F-Test): Used to test if the regression model is “better” than using only the mean of the dependent variable.

$$H_0: Y = \beta_0 \qquad H_1: Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Test for a Single  $\beta_k$ : Used to test if  $\beta_k$  differs from zero.

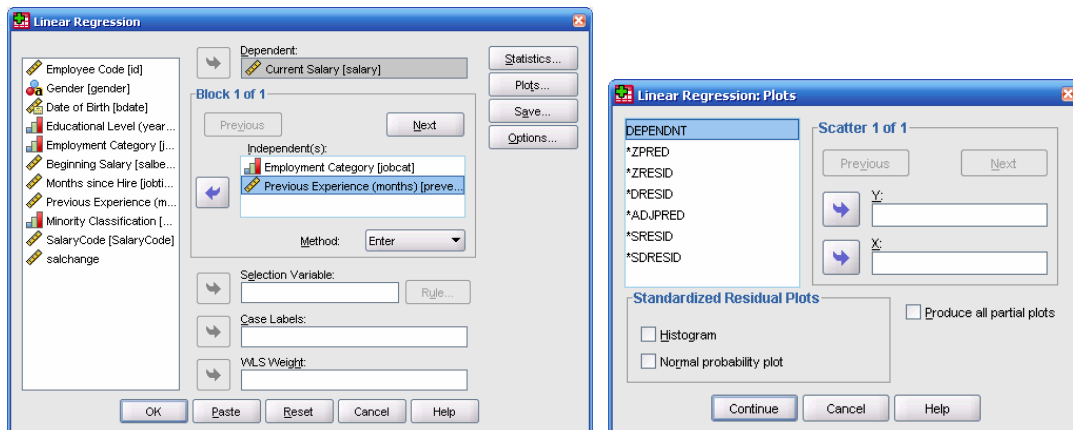
$$H_0: \beta_k = 0 \qquad H_1: \beta_k \neq 0$$

Example: What is the regression model for using “educational level” and “years experience” to predict salary?

Everything we need to create a linear regression model is located in the following menu:

**Analyze > Regression > Linear...**

The variable we are trying to “predict” is the *Current Salary* variable, which goes in the *Dependent* box. *Employment Category* and *Previous Experience* go in the *Independent(s)* box. Although there are many options available in the linear regression dialog box (such as the Plots box), we will only consider the defaults.



Results appear on the next page.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.794 <sup>a</sup>	.630	.629	\$10,407.649

a. Predictors: (Constant), Previous Experience (months), Employment Category

ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.7E+010	2	4.345E+010	401.121	.000 <sup>a</sup>
	Residual	5.1E+010	471	108319164.3		
	Total	1.4E+011	473			

a. Predictors: (Constant), Previous Experience (months), Employment Category

b. Dependent Variable: Current Salary

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12116.105	1067.489		11.350	.000
	Employment Category	17431.593	620.131	.789	28.110	.000
	Previous Experience (months)	-23.986	4.585	-.147	-5.232	.000

a. Dependent Variable: Current Salary

So, what does the output tell us?

- $R^2 = 0.630$ , which suggests 63% of the variance in salary can be “accounted for” by information about employment category and previous experience.
- An  $F$ -statistic of 401.121 ( $p$ -value  $< 0.001$ ) suggests that a regression model containing employment category and previous experience is “better” than a model without any predictor variables (using the mean salary as an estimate for anyone).
- The regression equation is:

$$\text{Salary} = 12116 + 17431 * \text{Employment Category} + (-23.986) * \text{Previous Experience}$$

- The  $t$ -statistics for each  $\beta_i$  is large enough in magnitude to reject the null hypothesis that each  $\beta_i = 0$ .

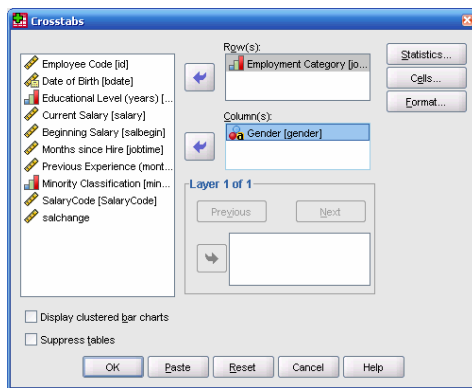
### 6.6. Chi-Square Tests

Chi-square tests are frequently used in contingency tables to determine what relationships exist between two discrete variables.

To create a contingency table in SPSS, we used the *Crosstabs* menu.

**Analysis > Descriptive Statistics > Crosstabs...**

Let's create a contingency table tallying the employees by *gender* and by *employment category*.



**Employment Category \* Gender Crosstabulation**

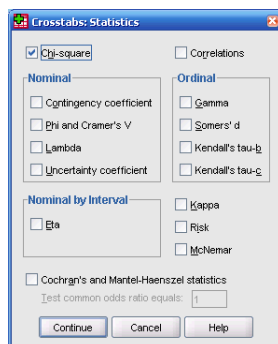
Count		Gender		Total
		Female	Male	
Employment Category	Clerical	206	157	363
	Custodial	0	27	27
	Manager	10	74	84
Total		216	258	474

Using the *Crosstabs* menu, we can also compute a chi square test for this data. Click on the **Statistics...** button and click on *Chi-square*.

Chi Square Test for 2 x 2 Contingency Table: Used to test if relationships occur across groups.

$H_0$ : There is NOT a relationship between variable A and variable B.

$H_1$ : There is a relationship between variable A and variable B.



**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	79.277 <sup>a</sup>	2	.000
Likelihood Ratio	95.463	2	.000
N of Valid Cases	474		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.30.

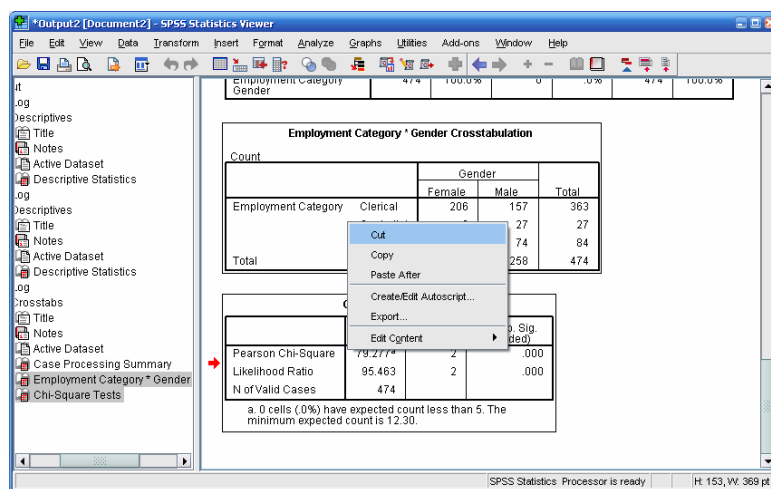
Since  $\chi^2 = 79.277$  (and  $p < 0.001$ ), we can conclude that there is a relationship between gender and employment category. Namely, there are more custodians and managers are male than female, and there are more clerical workers that are female.



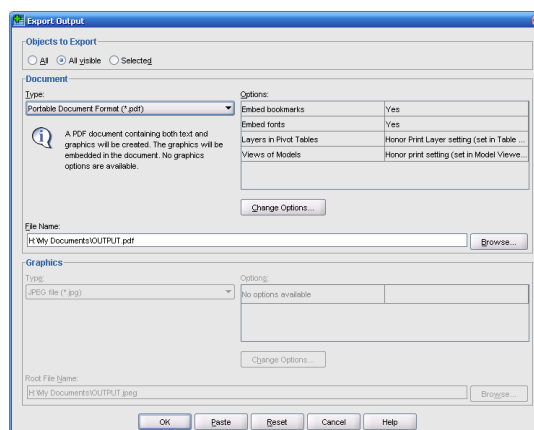
## Section 7: Exporting Your Results

SPSS output can be exported into other applications including Word documents, spreadsheets, and databases. There are at least two ways for doing this: **Copy** and **Export**. Each will produce a different appearance in displaying outputs.

One easy way for doing this is to copy the selected output and paste them on the software that you want to use. For example, we can export our results from the Chi-Square Test to Word and edit the exported tables in Word. From the *Output Viewer* window, highlight the results and select the “Copy” option by clicking the right button on your mouse. After copying the selected tables, paste and edit them in Word. You may need to resize or crop images.



To export the entire output window into a file, use the **Export** option. Consultants frequently do this, since SPSS output windows can only be opened in SPSS. The results can be exported into a Word document (\*.doc), an HTML file (\*.html), a PDF file (\*.pdf), a text file, an Excel file, or a Powerpoint file.



Remember that SPSS data sets can be exported into other formats by using the **File > Save As...** dialog box and changing the format under **Save as type**.

## Section 8: Some Further SPSS Resources

For more information on SPSS, try the following resources:

- Students and faculty of the College of Education are eligible for free statistical consulting at the Statistics Outreach Center. To make a free statistical appointment with the Statistical Outreach Center, visit [www.education.uiowa.edu/statoutreach](http://www.education.uiowa.edu/statoutreach), 200 Lindquist Center South, or email at [sheila-barron@uiowa.edu](mailto:sheila-barron@uiowa.edu).
- Visit the SPSS homepage, [www.spss.com](http://www.spss.com). The site includes a variety of helpful resources, including the SPSS Answer Net for answers to frequently asked technical questions, and an FTP site where you can find SPSS macros and an online version of the SPSS Algorithms manual. The support website is available at [www.support.spss.com](http://www.support.spss.com).
- A good book that is used frequently at the Statistics Outreach Center is *Discovering Statistics Using SPSS (2<sup>nd</sup> Edition)* by Andy Field (Sage, 2005). This entertaining text (seriously!) contains a wealth of information about running SPSS. It also provides some statistical background for many tests and analyses. This book is available at the University's Engineering Library. The book is also available in paperback on Amazon.com for \$55-\$70.
- For other books on SPSS, head for the Engineering Library (Seamans Center) and look for books with call numbers that start with "HA 32". A number of books are available (including, apparently, *SPSS for Dummies*), but make sure you get current books. Try to stick with books published after 2004, which covers SPSS version 13 and beyond.