SPSS@Oneonta:
Selected Introductory Statistical and Data Manipulation Procedures

Gordon & Johnson
2005

SPSS@Oneonta: Selected Introductory Statistical and Data Manipulation Procedures

This manual was donated by the authors to support the James A. Fraley Scholarship Fund.

Authors:

Lisa A. Gordon, Adjunct Lecturer: Statistics, Department of Mathematics, Computer Science & Statistics, SUNY Oneonta

Steven D. Johnson, Senior Staff Associate, Institutional Research, Division of Finance & Administration & Adjunct Lecturer: Statistics, Department of Mathematics, Computer Science & Statistics, SUNY Oneonta

Comments & Suggestions:

Comments and suggestions may be forwarded to: GORDONLA@ONEONTA.EDU or JOHNSOSD@ONEONTA.EDU.

James A. Fraley Scholarship:

This manual was donated to support the James A. Fraley Scholarship at the SUNY College at Oneonta. All donations aid in the growth of the scholarship's endowment. The Fraley scholarship assists students majoring in the field of statistics at SUNY Oneonta. For information about the Fraley Scholarship contact the Department of Mathematics, Computer Science & Statistics, 238 Fitzelle Hall, SUNY, Oneonta NY 13820 (607-436-3708).

Printing:

First Edition: July 2001; Second printing August 2002. Second Edition: July 2005

Disclaimer:

This document is intended for the use by students in STAT 101: Introductory Statistics and STAT 141: Statistical Software – SPSS; both are courses offered at SUNY Oneonta. It is not intended to be a comprehensive documentation of all that SPSS can provide with regard to a particular statistic or test. The content of the document may change as needed to meet its function as a supplemental reference for STAT 101/141. No living trees were harmed in the printing of this manual.

Copyright:

TABLE OF CONTENTS

INTRODUCTION

This manual is intended as a reference guide to the SPSS statistical software package used by students in the introductory statistics classes at SUNY Oneonta. The manual provides a "how to" approach to each topic presented. It does not discuss every option available for a given topic, but rather provides sufficient instruction to obtain the basic output one might need address data analysis at an introductory level. At the same time, this is not a statistics text. Therefore, the user will need to refer to a statistics text for definitions and further clarification of the procedures presented herein.

DOCUMENT OVERVIEW

Most topics are presented on no more than two pages and include: a brief **OVERVIEW** of the topic; a summary listing of the **STEPS** to complete the procedure; and, for most topics, an **EXAMPLE** of how to use the statistical topic, which includes a more detailed presentations of the steps. Within each procedure's Example, SPSS screen shots show how to access and complete the desired SPSS command. SPSS output is presented so that the user may see the results of the command selected. For some topics the output may be "enhanced" through superimposed notes, which point out various components of the output.

CONVENTIONS USED IN THE MANUAL

There are a number of conventions or means of presenting information that are constant throughout the manual. As one initiates a SPSS command, a **Dialog Box** will open. It is within these Dialog Boxes that one will enter all the information needed to produce the desired output.

STEPS: All bold print, indicates all steps needed to successfully complete the minimum requirements for the SPSS command to provide output.

OPTIONS: With the selected SPSS command, using the Options noted below the **STEPS** will provide additional output or enhance the output.

ADDITIONAL REQUIREMENTS: For some SPSS features it is necessary to go to a second Dialog Box and enter additional information BEFORE selecting the Ok button.

RELATED GRAPHS: For some statistical procedures a reference may be made to a related graph.

RELATED STATISTICAL PROCEDURE: For some Tables and Graphs a related statistical procedure may be noted.

SUMMARY OF STEPS: For some procedures an expanded, more descriptive presentation of the STEPS is presented for the EXAMPLE.

'VARIABLES' used in the examples appear in bold, all capital letters within single quotation marks.

Dialog Boxes, Cells into which one would enter information and some other actions, such as click **OK**, are presented in bold lettering.

ORDERING OF TOPICS

The contents of this manual have been broken into three sections.

Section 1: SPSS Environment - provides information on how to get started using SPSS.

Section 2: Statistical Procedures - provides an overview of many statistics and graphs encountered in an introductory statistics course.

Section 3: Data Manipulation - provides information on many ways to enhance/modify SPSS data files. While these procedures are not necessary to obtain statistical output, they do facilitate data analysis and interpretation.

DATA

The data used for examples throughout this manual were selected from sample data files that are distributed with SPSS. They may be accessed from within SPSS using the following **STEPS: File => Open => Data => set the "Look in:" cell to the folder containing the data file => select the desired data file => Open.** SPSS will enter the data into the Data Editor Window.

The data sets used in this manual are noted below. Coding of variable values may be viewed by selecting the "Variable View" tab in the Data Editor. A printout of variable coding may be obtained by opening the data file and then using the **STEPS: File => Display Data File Information => Working File.** A summary of variables and their formatting will be placed in the Output Viewer. If you are interested in the formatting of a single variable, use the **STEPS: Utilities => Variables.** The File Info printouts for the files noted below are located in **APPENDIX A**.

DATA FILES:

Data: Cars.sav
Data: Employee data.sav
Data: Road Construction Bids.sav
Data: World95.sav

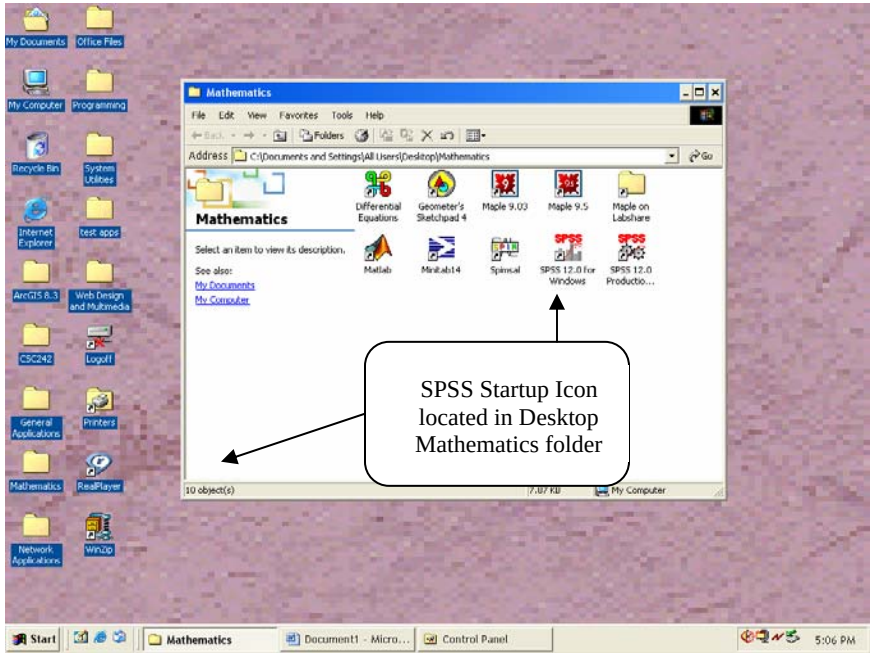
THE SPSS ENVIRONMENT:
HOW TO GET STARTED – SPSS BASICS

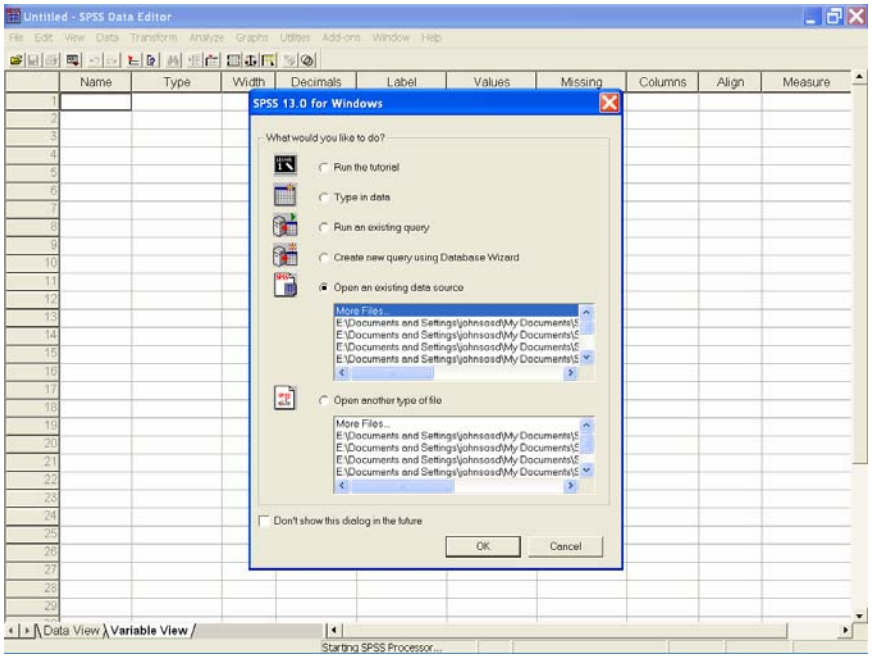
GETTING STARTED WITH SPSS

WHERE TO FIND SPSS: SPSS can be accessed from virtually any PC-based computer lab on the SUNY Oneonta campus. For a listing of the on-campus computer labs and software availability select the following path from the Oneonta Homepage: Technology => Academic Computer Services => Labs.

HOW TO ACCESS SPSS: The location of the SPSS icon may vary from one lab to another. In most cases, it will be located on the desktop in a folder named **Mathematics** If it is not there look in the **General Applications** or **Student Applications** folders.

TO START SPSS: Double-click on the SPSS icon (shown in the left figure) and the Data Editor window below will open.





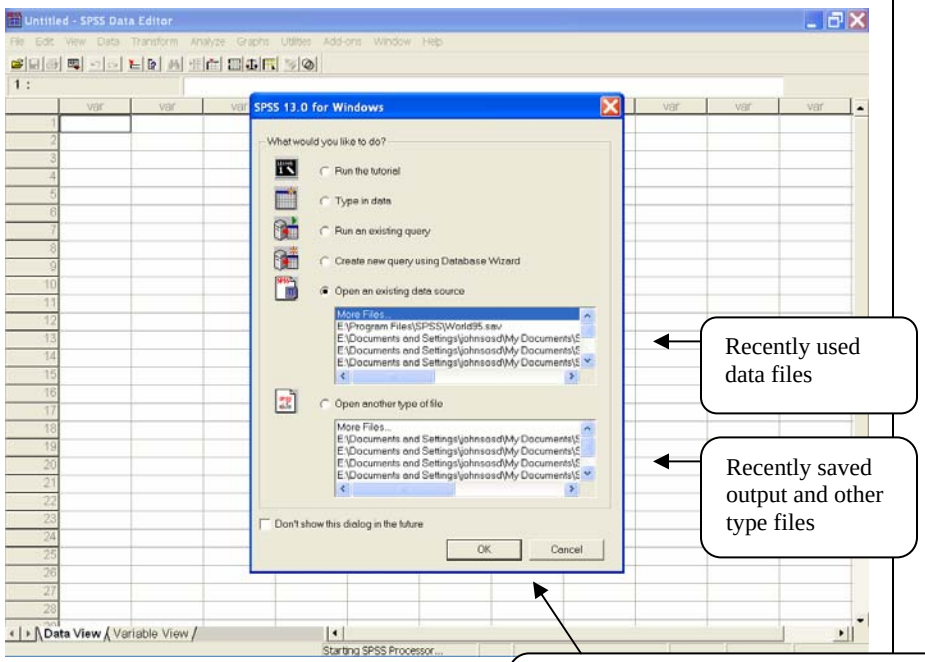
THE INITIAL SPSS WINDOW

When starting SPSS the Data Editor window appears with a dialog box containing start-up options. Recently used data files appear in the top listing and other types of recently used files, such as output and syntax files, will appear in the bottom listing. If the file desired is not listed or you want a blank Data Editor, select cancel. The Data Editor will be ready to receive direct data input or you can use the main menu **File => Open => Data** to locate a specific data file. If you select "OK," SPSS will open the SPSS directory containing over 100 sample data files, including most of the files used in this manual.

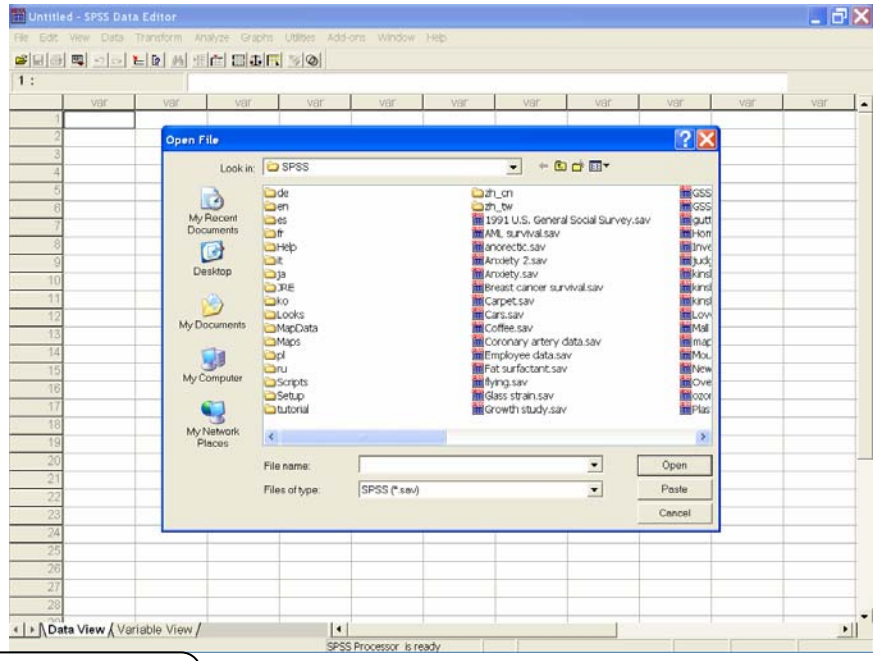
SPSS contains several pre-defined data files for your use. If you wish to select a data file saved in an SPSS subdirectory **select OK** (this is the default directory). If you wish to open a file from another location (such as a disk), select cancel and then use the **STEPS: File** => **Open => Data => and direct SPSS to the file's location** (See topic: RETREIVING A SAVED FILE).

To select a data file: Click on the data file, then select **Open.** Alternately, you may double-click on the file of your choice.

Data Editor: First window to appear when SPSS is opened.



File selection.



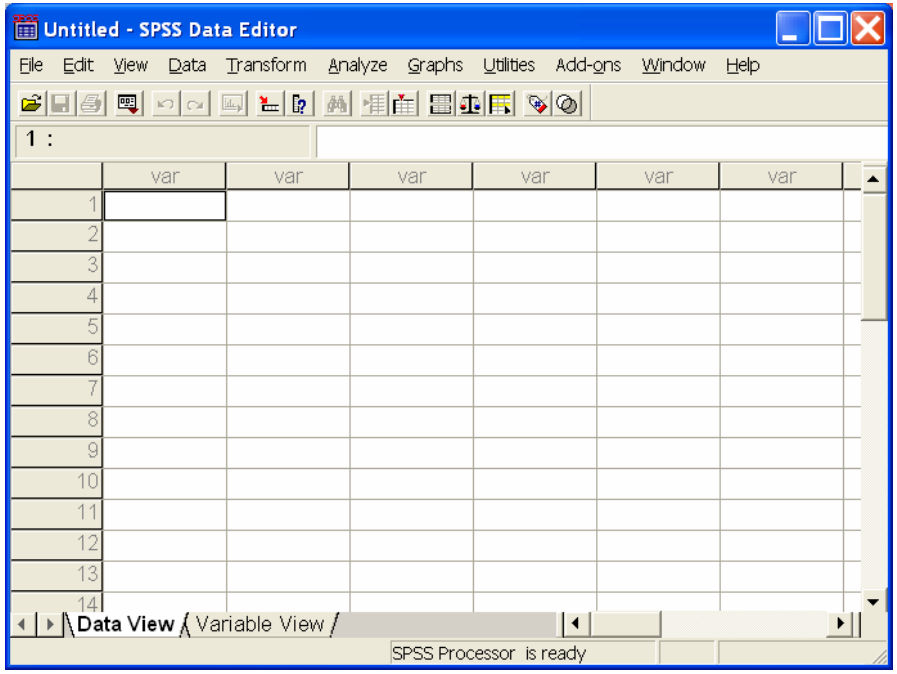
ENTERING DATA INTO A NEW DATA FILE

If you wish to create a new data file, click **Cancel** in the opening dialog box. The opening dialog box will close, and the worksheet shown behind it will be active and ready for you to begin entering data and defining variables. Note that there are two tabs at the bottom of the Data Editor window. When the **Data View** tab is selected one can enter information into the window. When the **Variable View** tab is selected, one can modify information about the file's variables (see the sections on the **DATA EDITOR** and **DEFINE VARIABLES**). In order to change to either view, click on the respective tab in the lower left corner of the worksheet.

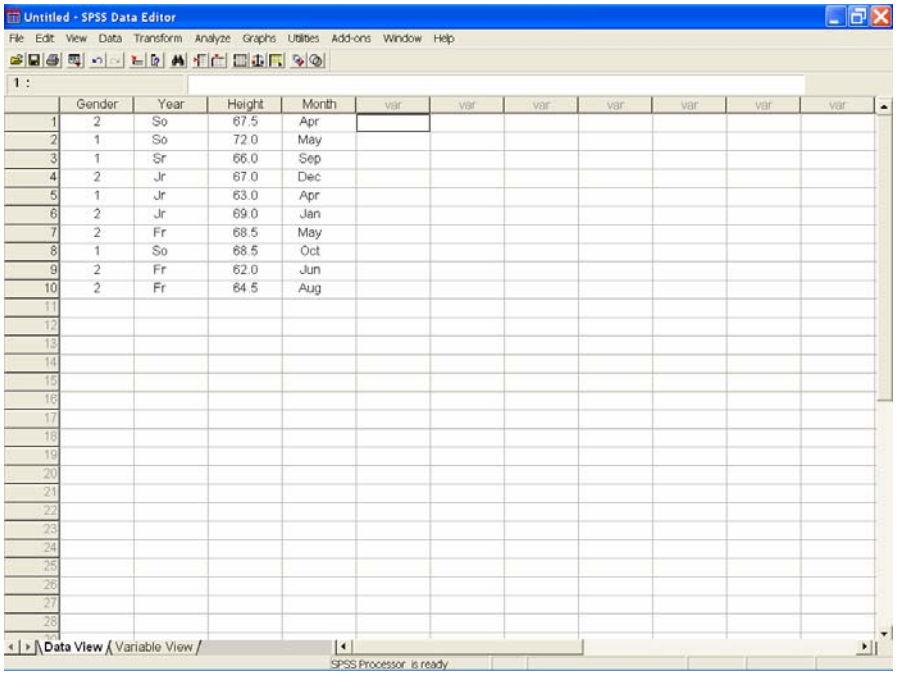
To begin entering your data, click inside the cell where you will enter your first value. Type in the value, and press **Enter** on the keyboard**.** The value will appear in the cell, and you will be taken to the cell immediately under it. If you wish to enter the values of all variables for one case (instead of all cases for one variable), use the right-arrow key on the keyboard instead of pressing **Enter** after entering a value. You will be moved to the next variable for a case. You may move left, right, up or down in **Data View** by using the **arrow keys.** If you make a mistake when entering your data, click once in the cell containing the error, and type in the correct value.

VARIABLES vs. CASES: A **Variable** is a single piece of information collected from every unit in the sample or population of interest. They represent the columns in the data editor. For example, in the data below there are four variables, Gender, Year, Height, and Month. Each member of this population of ten individuals provided information about the variable Gender. In contrast a **Case** contains all of the information (variables) obtained from one source. Cases are the rows in the data editor. So here, the first case is an individual who identified his Gender (2), Year level (So), Height (67.5 in.) and Month (of birth Apr).

Data Editor ready for data entry.



Data manually entered.

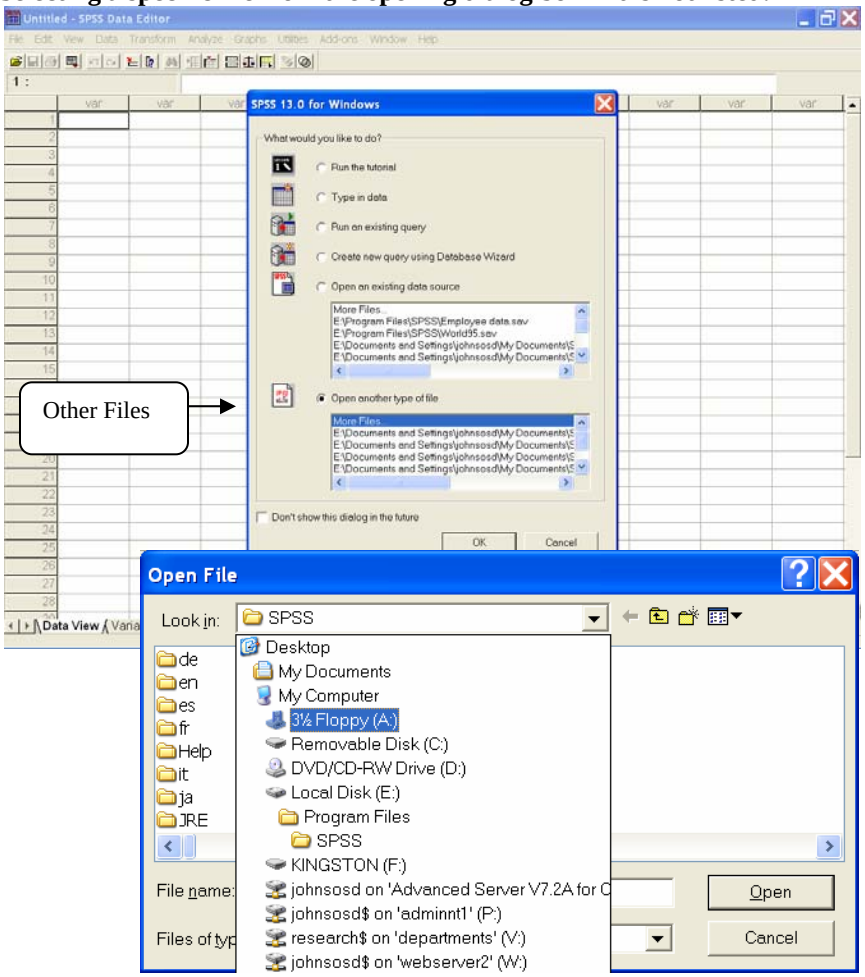


RETRIEVING A SAVED FILE

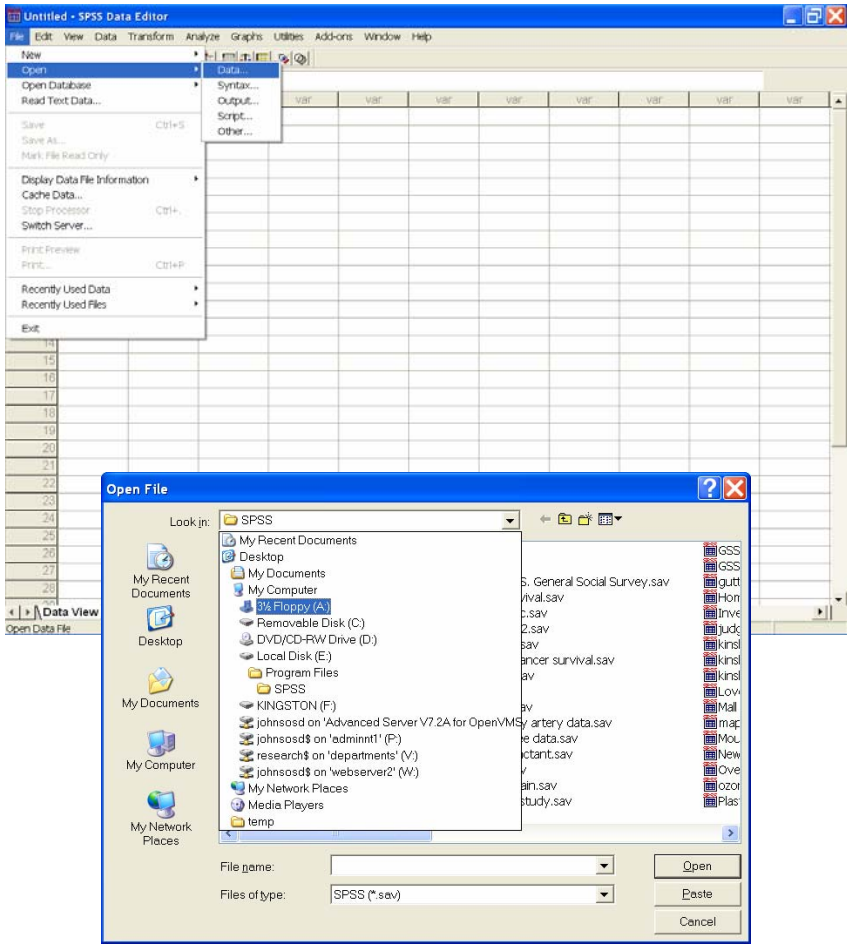
Option 1: To open a file that is not displayed on the recently used listings of the dialog box that appears when SPSS is started, select the **Open another type of file** button, as shown on the left below. In the dialog box that will appear, use the pull-down menu to select where the file is located. For example, in the dialog box shown (below left), the A-drive is highlighted. Selecting that drive will result in the computer listing available files on the A: drive. Note that once you have selected the location, you may need to specify the file type in the **Types of Files** dropdown box.

Option 2: If you have gone beyond the opening screen's dialog box or want to open a file different from the one currently in the Data Editor or one of the other SPSS windows, use the main menu. **STEPS: File => Open => select a file type => select the Location.**

Selecting a specific file from the opening dialog box if it is not listed:



Selecting a file once the opening dialog box has been closed:

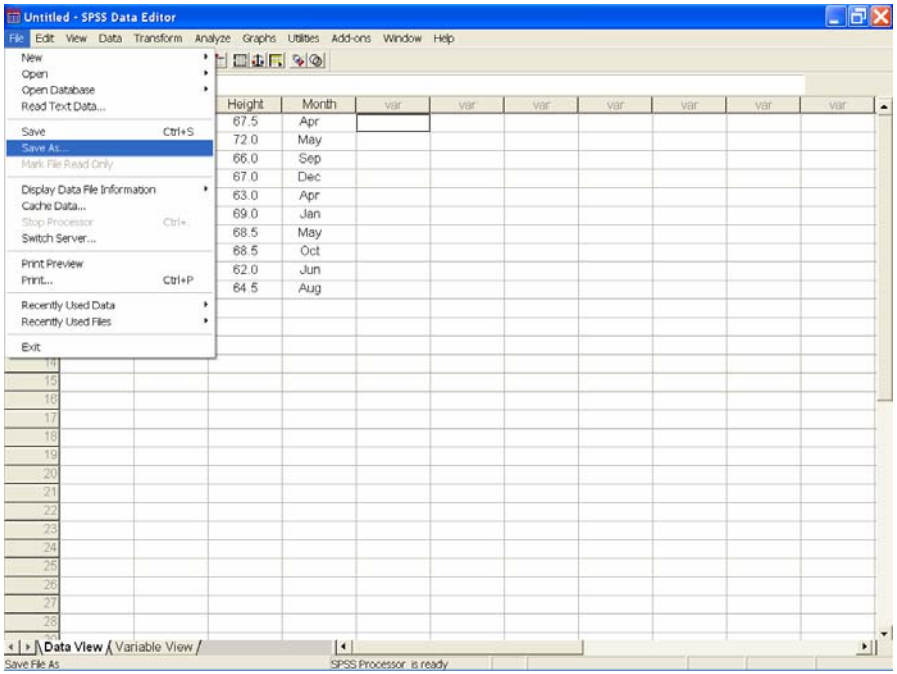


SAVING YOUR WORK

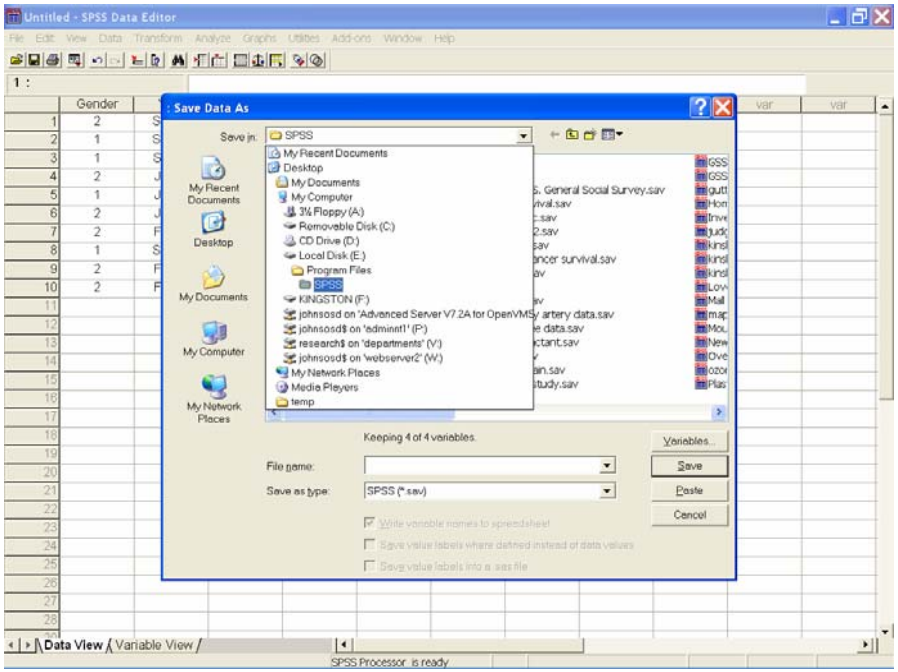
At Oneonta if you wish to save your work, do not save it to the hard drive (C:) as these drives are periodically updated and student files are erased. Instead, save it on a floppy disk, zip disk, flash drive or your **"P:"** drive by selecting **File** from the Main Menu Bar and then using the **Save As** command.

In the dialog box that appears, click once on the down-arrow on the right side of the **Save In** box**.** Select the location to which the file is to be saved. In the **File Name** box, type in a name for your file, then click **OK** and the file will be saved to the location previously selected.

Menu Bar => File => Save As



Select a location via the "Save in" dropdown list



SPSS WINDOWS

SPSS is built around a number of windows, each serving a different function. While several windows are noted below, familiarity with the **Data Editor** and the **Output Viewer** are sufficient to obtain data analysis. The ability to use the **Chart Editor** and the **Pivot Table Editor** will allow one to produce presentation quality tables and charts.

- **CHART EDITOR[1]:** This window allows for the modification of the appearance of graphs and charts.
- **DATA EDITOR[1]:** Usually the entry point for SPSS analysis, this window contains data and editing descriptions for individual variables.
- **DRAFT VIEWER:** Displays output in simple text rather than as pivot tables.
- **OUTPUT VIEWER[1]:** All data analysis output is maintained in this window. Editing of tables and graphs begins here.
- **PIVOT TABLE EDITOR[1]:** This window allows for the modification of the appearance of tables.
- **SYNTAX EDITOR:** Allows for the use of command syntax in the development of procedures.
- **SCRIPT EDITOR:** Allows for the customizing and automation of certain SPSS tasks.

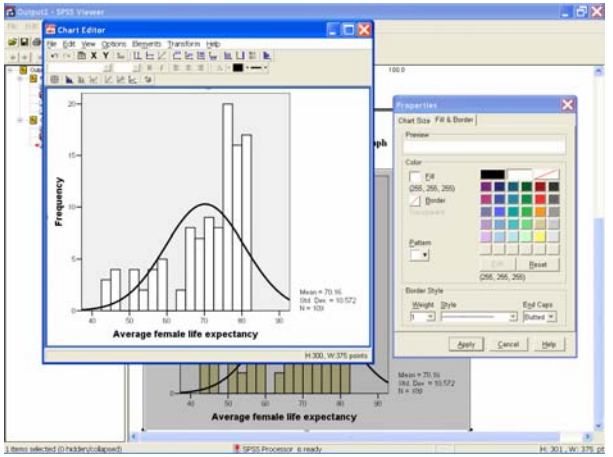
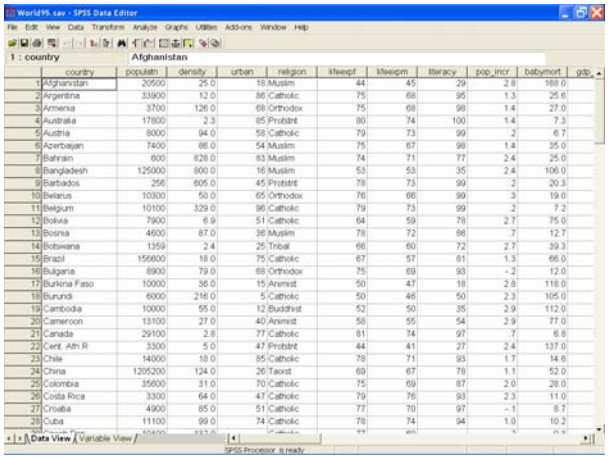
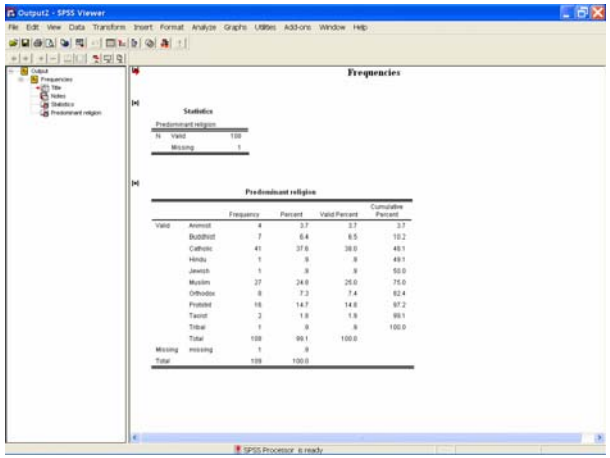


Chart Editor

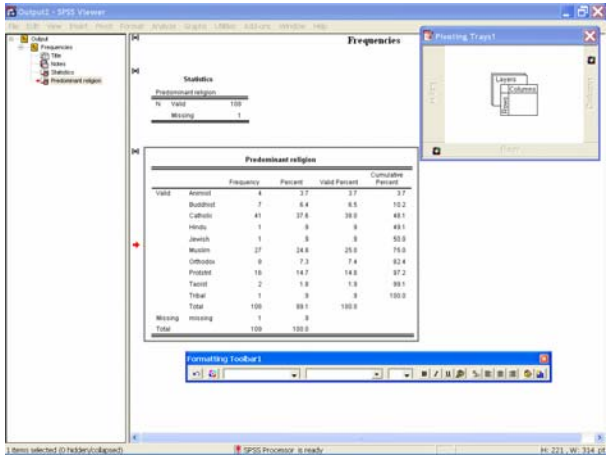
Sample Windows:



Data Editor



Output Viewer



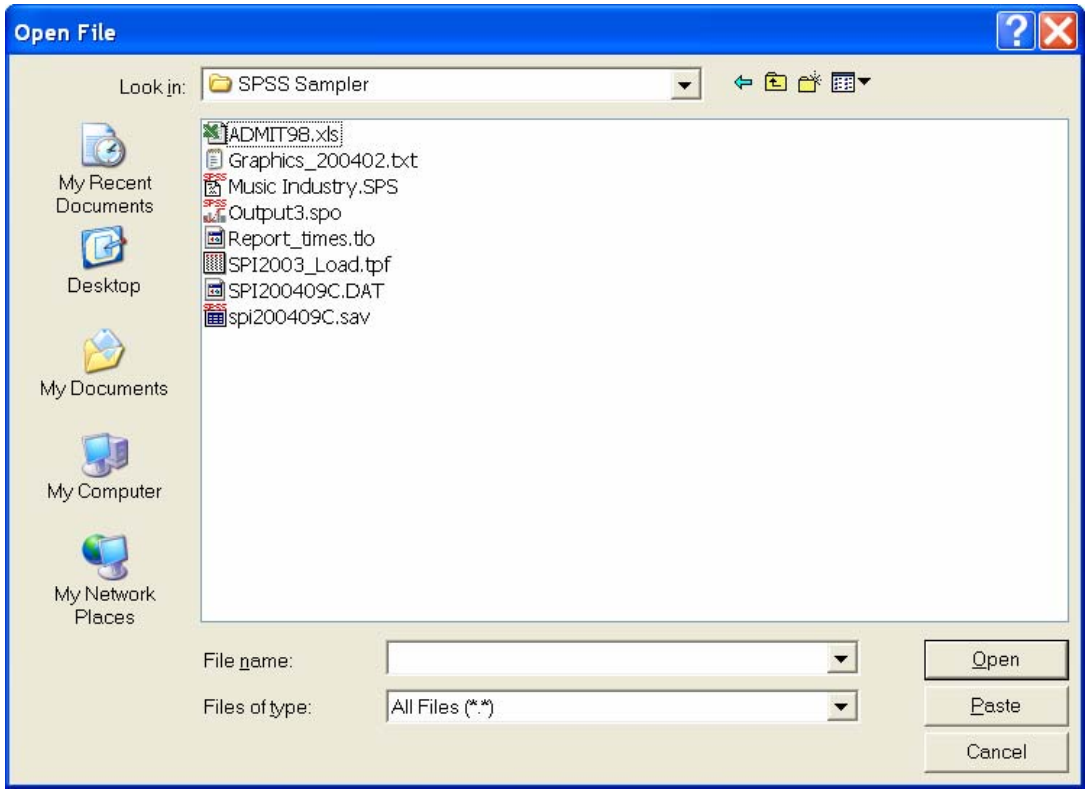
Pivot Table Editor (note box around table has a “shadow” effect around it)

[1]Each of these four windows is more thoroughly presented as separate topics.

SPSS FILE EXTENSIONS

SPSS uses a number of file extensions to identify various types of files. The extensions most frequently encountered at the introductory level identify data and output files

- **DATA FILES** may be identified by the "**.SAV**" extension. In a list of files an icon containing "SPSS" and a small spreadsheet precedes a data file name. Selecting a listing of data files will include Excel files, which can be opened in SPSS without modification or using a data entry wizard.
- **OUTPUT FILES** may be identified by the "**.SPO**" filename extension. In a list of files an icon containing a bar chart precedes an output file.
- **SYNTAX FILES** contain the extension "**.SPS.**" These files contain a listing of steps used to obtain data coding and/or analysis. They are particularly useful where a repetition of tasks might occur either within the context of current data analysis or as a listing of steps for future use.
- **RAW DATA FILES**, those containing data similar to that which one might hand enter, may have the extensions "**.TXT**" or "**.DAT**." To be entered into SPSS these data files would require either a data entry wizard or conversion to some other format accepted by SPSS (e.g. Excel format).
- **DATA WIZARD FILE** has the extension "**.TPF**" and is used to enter .dat and .txt files. Once a set of data have been entered through the data wizard, this type of file may be saved for future use with similarly formatted data.

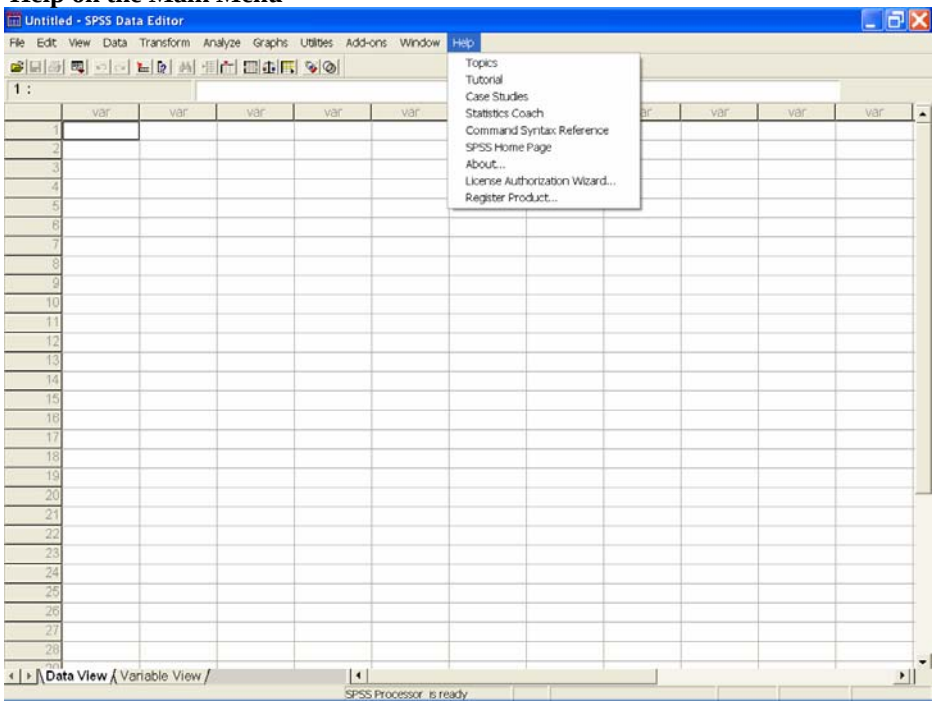


HELP

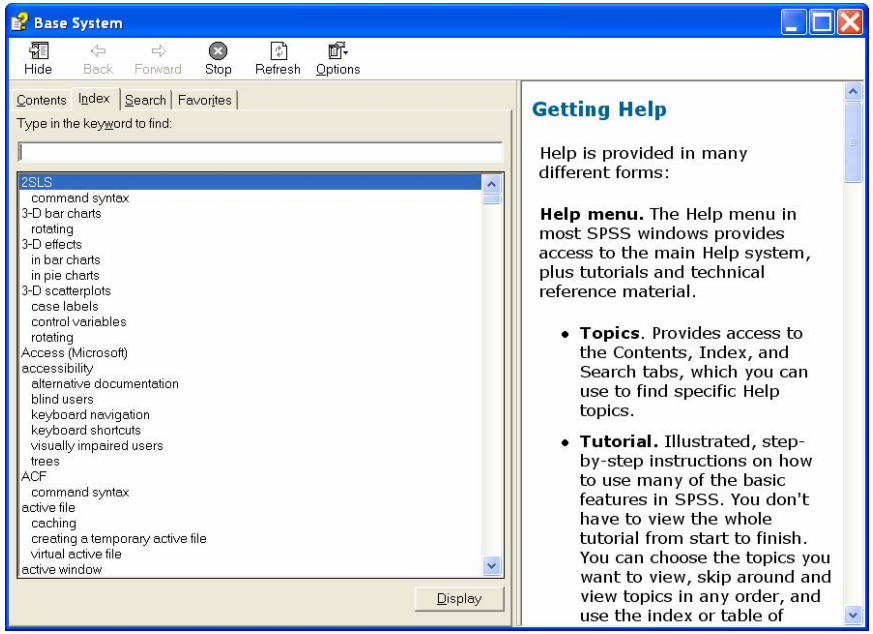
All windows within SPSS contain a link to **Help** on their Main Menus. Additionally, selecting the **F1** key will open the **Help Topics Dialog Box**. Selecting either **F1** or **Topics** from the **Help Menu** initiates the most encompassing version of Help. The Help Topics Dialog Box contains the Contents, Index, and Find sections. From within the Index and Find options one can search for specific topics.

STEPS: Any SPSS Window Main Menu => Help => Make a Selection; Also available from any dialog box.

Help on the Main Menu



Help Topics Dialog Box



Tutorials & Case Studies

A series of brief tutorials and examples (case studies) are available to SPSS users by selecting either the **Tutorials** or **Case Studies** choices from the **Help Menu**.

SPSS Home Page

This is a direct link to SPSS on the Internet.

Command Syntax Reference

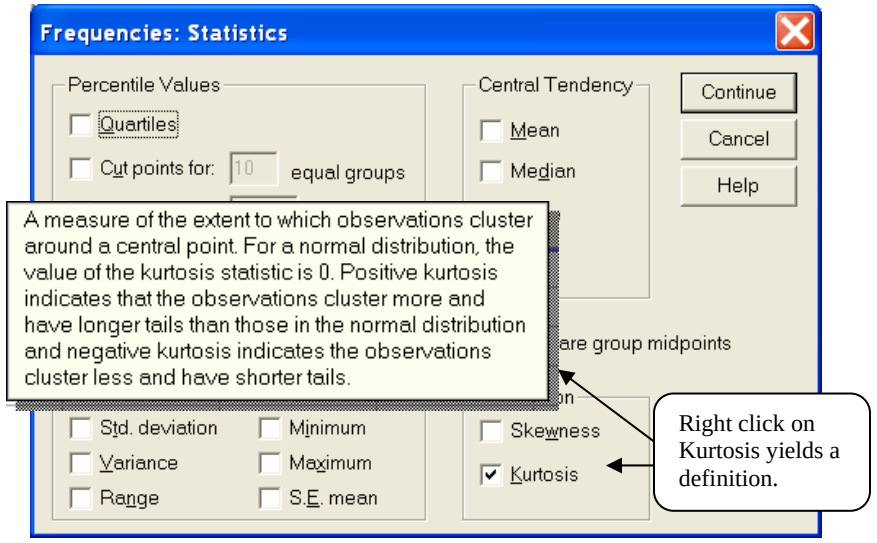
This is a link to .pdf files stored on disk that present how to write syntax.

Dialog Box Help

Within each Dialog Box selecting the Help button will retrieve information about that topic. So, for example, if you has the Frequencies Dialog Box open and selects Help, information about Frequencies appears. From the Frequencies description help window, you can go to the main Help Topics Index, find out how to conduct a frequency calculation, and select related topics.

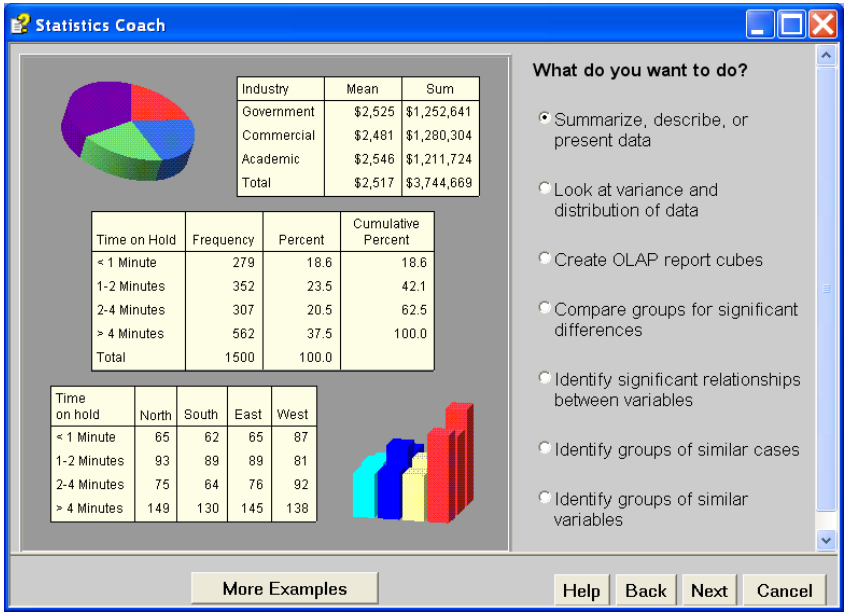
Help with Terminology

Simply placing the mouse cursor on a command or a statistical option and clicking the right button will retrieve pop-up definition for that item. So, for example, right clicking on the word Kurtosis would yield the definition shown below.



Statistics Coach

From within the **Help Dialog Box** one can obtain assistance in determining the statistical analysis approach appropriate for the data one wishes to analyze.



Industry	Mean	Sum
Government	$2,525	$1,252,641
Commercial	$2,481	$1,280,304
Academic	$2,546	$1,211,724
Total	$2,517	$3,744,669

Time on Hold	Frequency	Percent	Cumulative Percent
< 1 Minute	279	18.6	18.6
1-2 Minutes	352	23.5	42.1
2-4 Minutes	307	20.5	62.5
> 4 Minutes	562	37.5	100.0
Total	1500	100.0	

Time on hold	North	South	East	West
< 1 Minute	65	62	65	87
1-2 Minutes	93	89	89	81
2-4 Minutes	75	64	76	92
> 4 Minutes	149	130	145	138

SPSS STATISTICS & GRAPHS:
PRODUCING INTRODUCTORY STATISTICS AND GRAPHS

ANALYSIS OF VARIANCE (ANOVA)

One-Way Analysis of Variance (One-Way ANOVA) is used to examine the hypothesis that the means of several samples are equal. Generally this includes three or more means. Where two samples are involved the two-sample t test is used. Post Hoc tests are used to identify between which means the differences lie.

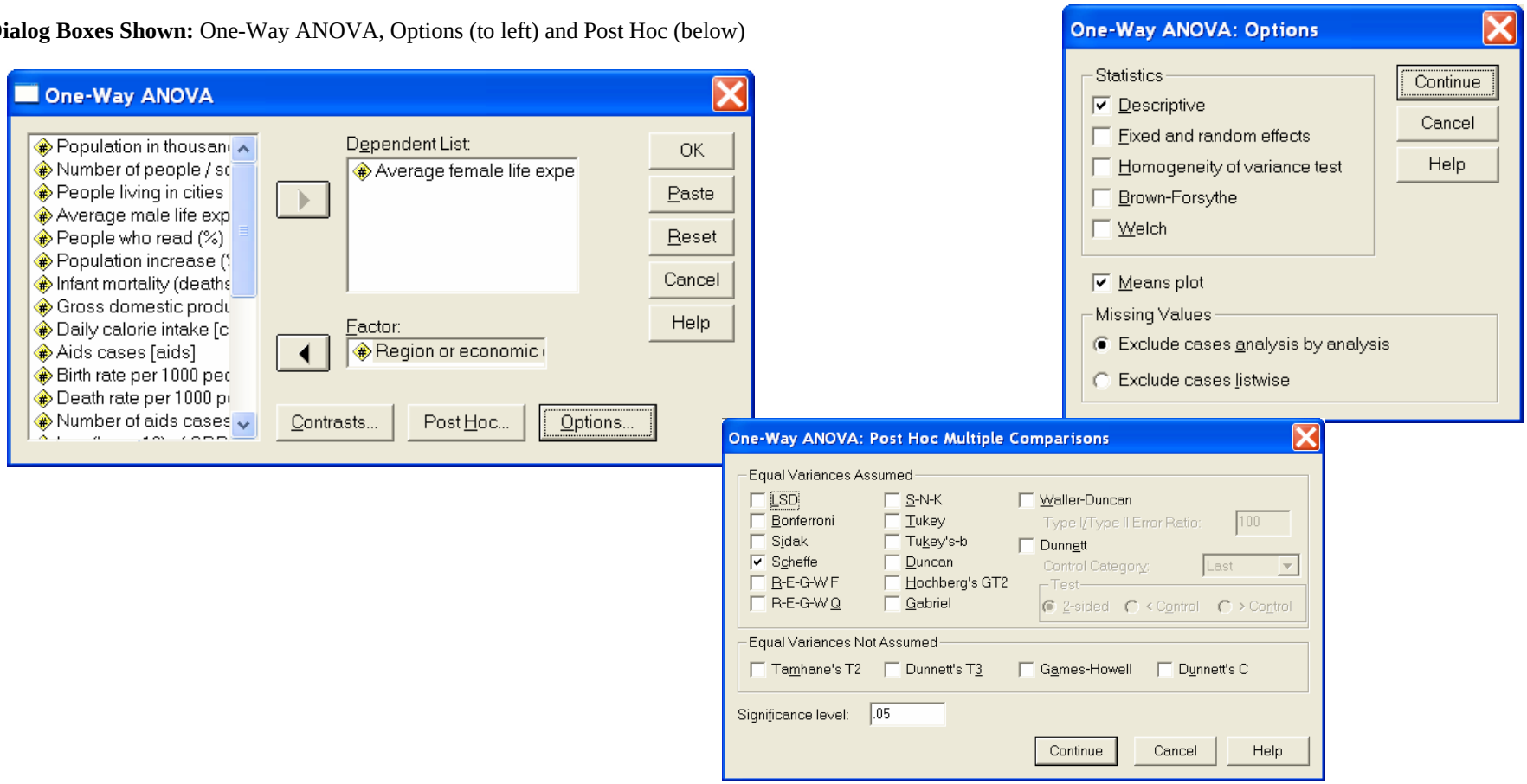
DATA: World95.sav

STEPS: Menu Bar => Analyze => Compare Means => One-Way ANOVA => select Options, Post Hoc data analysis, etc. => OK

EXAMPLE: Does the average life expectancy of females differ across six economic/geographic areas?

SUMMARY OF STEPS: From the Main Menu select **Analyze.** => **Compare Means.** => **One-Way ANOVA.** => Move the variable to be investigated, female life expectancy, into the **Dependent List** and enter grouping variable, geographic region, into the **Factor** box. **Optional output**: Select the Options button to obtain output for descriptive statistics and a means plot. Select the **Post Hoc** button to use one or more of the multiple comparison procedures.

Dialog Boxes Shown: One-Way ANOVA, Options (to left) and Post Hoc (below)



SPSS Output (Selected items): Descriptive statistics for each economic/geographic region; the results of the ANOVA process; a means plot and the results of the Post Hoc analysis using the Scheffe test as an example.

DESCRIPTIVES

Descriptives

Average female life expectancy

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
OECD	21	80.10	1.179	.257	79.56	80.63	78	82
East Europe	14	76.00	1.109	.296	75.36	76.64	74	78
Pacific/Asia	17	67.41	10.886	2.640	61.81	73.01	44	82
Africa	19	54.26	7.978	1.830	50.42	58.11	43	70
Middle East	17	71.59	4.501	1.092	69.27	73.90	63	80
Latn America	21	71.76	7.389	1.612	68.40	75.13	47	79
Total	109	70.16	10.572	1.013	68.15	72.16	43	82

ANOVA

ANOVA

Average female life expectancy

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7568.810	5	1513.762	34.636	.000
Within Groups	4501.539	103	43.704		
Total	12070.349	108			

POST HOC TEST: SCHEFFE

Multiple Comparisons

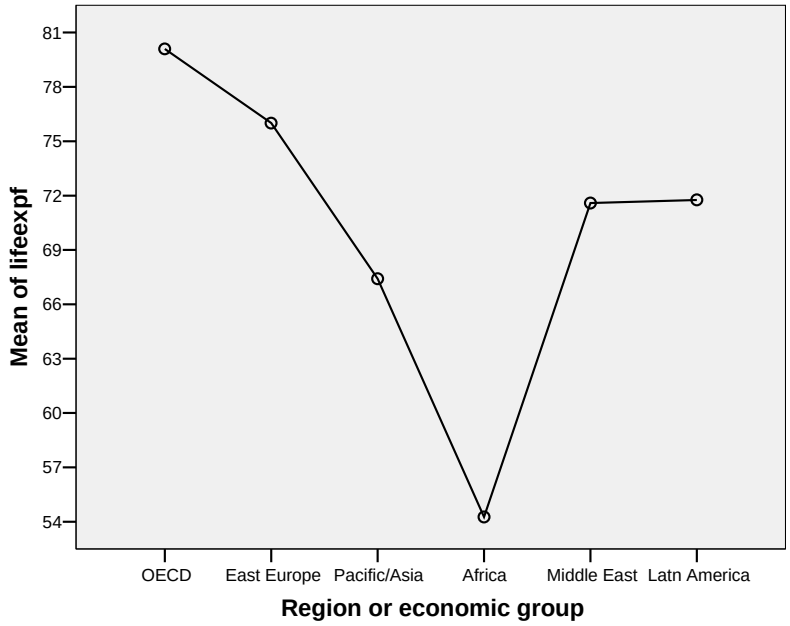
Dependent Variable: Average female life expectancy

Scheffe

(I) Region or economic group	(J) Region or economic group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
OECD	East Europe	4.095	2.281	.666	-3.64	11.83
	Pacific/Asia	12.683*	2.157	.000	5.37	20.00
	Africa	25.832*	2.093	.000	18.73	32.93
	Middle East	8.507*	2.157	.012	1.19	15.83
	Latn America	8.333*	2.040	.008	1.41	15.26
East Europe	OECD	-4.095	2.281	.666	-11.83	3.64
	Pacific/Asia	8.588*	2.386	.030	.49	16.68
	Africa	21.737*	2.329	.000	13.84	29.64
	Middle East	4.412	2.386	.637	-3.68	12.51
	Latn America	4.238	2.281	.632	-3.50	11.98
Pacific/Asia	OECD	-12.683*	2.157	.000	-20.00	-5.37
	East Europe	-8.588*	2.386	.030	-16.68	-.49
	Africa	13.149*	2.207	.000	5.66	20.64
	Middle East	-4.176	2.268	.641	-11.87	3.52
	Latn America	-4.350	2.157	.543	-11.67	2.97
Africa	OECD	-25.832*	2.093	.000	-32.93	-18.73
	East Europe	-21.737*	2.329	.000	-29.64	-13.84
	Pacific/Asia	-13.149*	2.207	.000	-20.64	-5.66
	Middle East	-17.325*	2.207	.000	-24.81	-9.84
	Latn America	-17.499*	2.093	.000	-24.60	-10.40
Middle East	OECD	-8.507*	2.157	.012	-15.83	-1.19
	East Europe	-4.412	2.386	.637	-12.51	3.68
	Pacific/Asia	4.176	2.268	.641	-3.52	11.87
	Africa	17.325*	2.207	.000	9.84	24.81
	Latn America	-.174	2.157	1.000	-7.49	7.14
Latn America	OECD	-8.333*	2.040	.008	-15.26	-1.41
	East Europe	-4.238	2.281	.632	-11.98	3.50
	Pacific/Asia	4.350	2.157	.543	-2.97	11.67
	Africa	17.499*	2.093	.000	10.40	24.60
	Middle East	.174	2.157	1.000	-7.14	7.49

*. The mean difference is significant at the .05 level.

MEANS PLOT



BAR CHART

A **Bar Chart** is used to present categorical data in a graph format. It consists of bars (rectangles) of equal width separated by spaces. Each bar represents a category of the variable. The heights of the bars are proportional to the *frequencies* or *percentages* they represent. **Bar Charts** may be created in conjunction with Frequency Tables or by selection from the **Graphs** Main Menu item. In the former case a simple bar chart will be created for each Frequency Table requested. In the latter case individual **Bar Charts** are created and one may specify certain chart features, such as titles, prior to the chart's creation. Once created graphs and charts may be modified/enhanced through use of the **Chart Editor**. **Bar Charts** assume a number of different formats, including Simple, Clustered and Stacked.

DATA: Employee data.sav

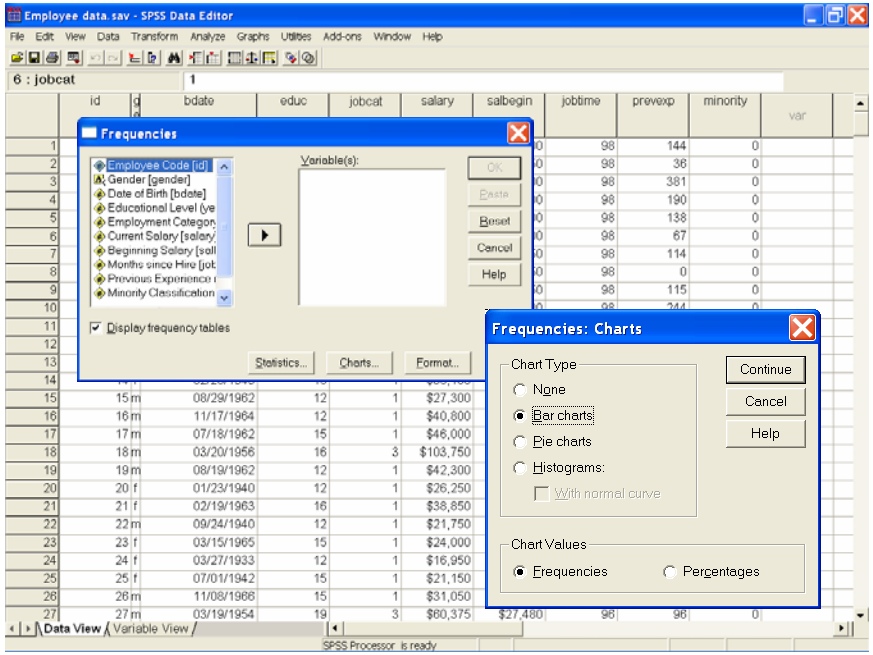
STEPS:

1) METHOD 1: Menu Bar => Analyze => Descriptive Statistics => Frequencies => select variables => Charts button => select Chart Type => Continue => Okay

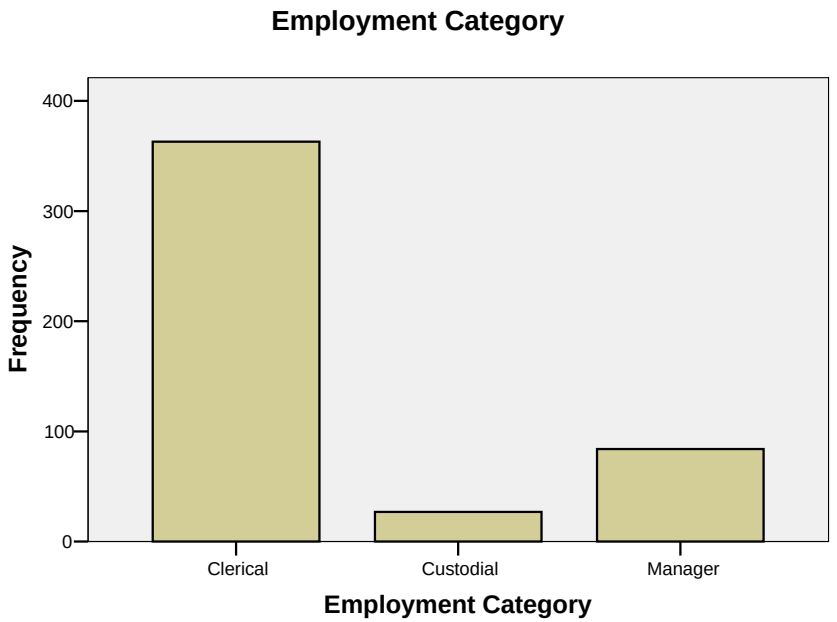
2) METHOD 2: Menu Bar => Graphs => Bar => select options and variables => Okay

BAR CHART EXAMPLE 1: MEHTOD 1 - Obtain a **Simple Bar Chart** of the variable **'JOBCAT'** as part of the output for a **Frequency Table**.

Dialog Box Input: From within the Frequencies dialog box select the Charts button. Select the Bar chart option.

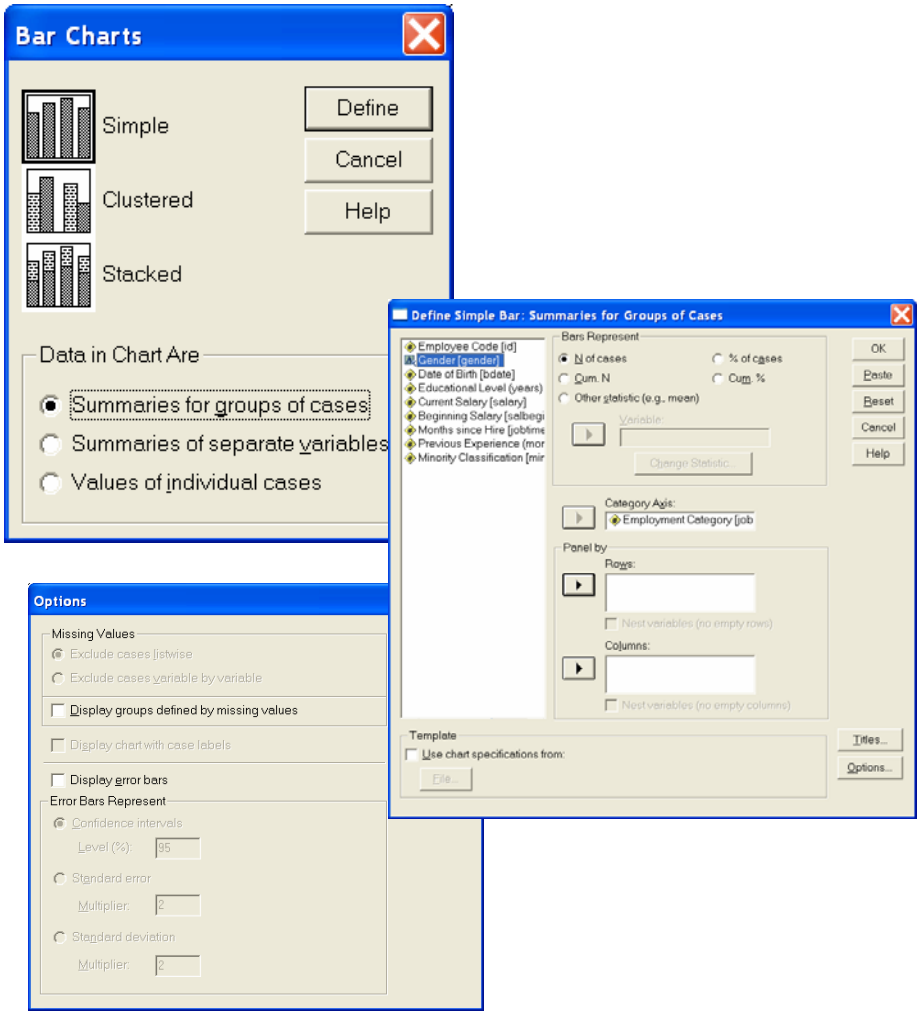


SPSS Output: Bar Chart of 'JOBCAT' (in addition to the frequency table)

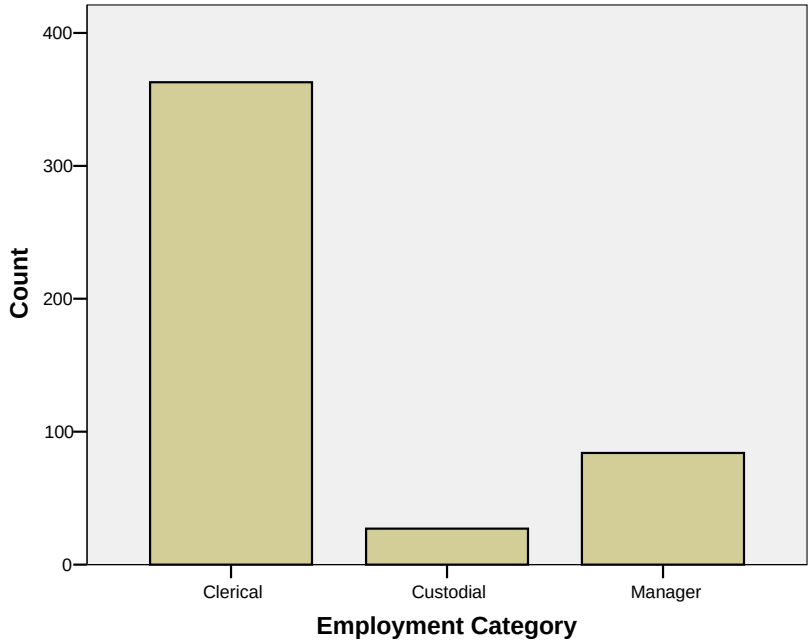


BAR CHART EXAMPLE 2: METHOD 2 - Obtain a **Simple Bar Chart** of the variable **'JOBCAT'** via the **Graphs menu**.

Dialog Box Input: From the Graphs Menu select **Bar** (Chart). Select Simple, then the Define button. Move the **'JOBCAT'** variable into the **Category Axis** cell.
Use the **Options** button to turn off the inclusion of bars for missing values

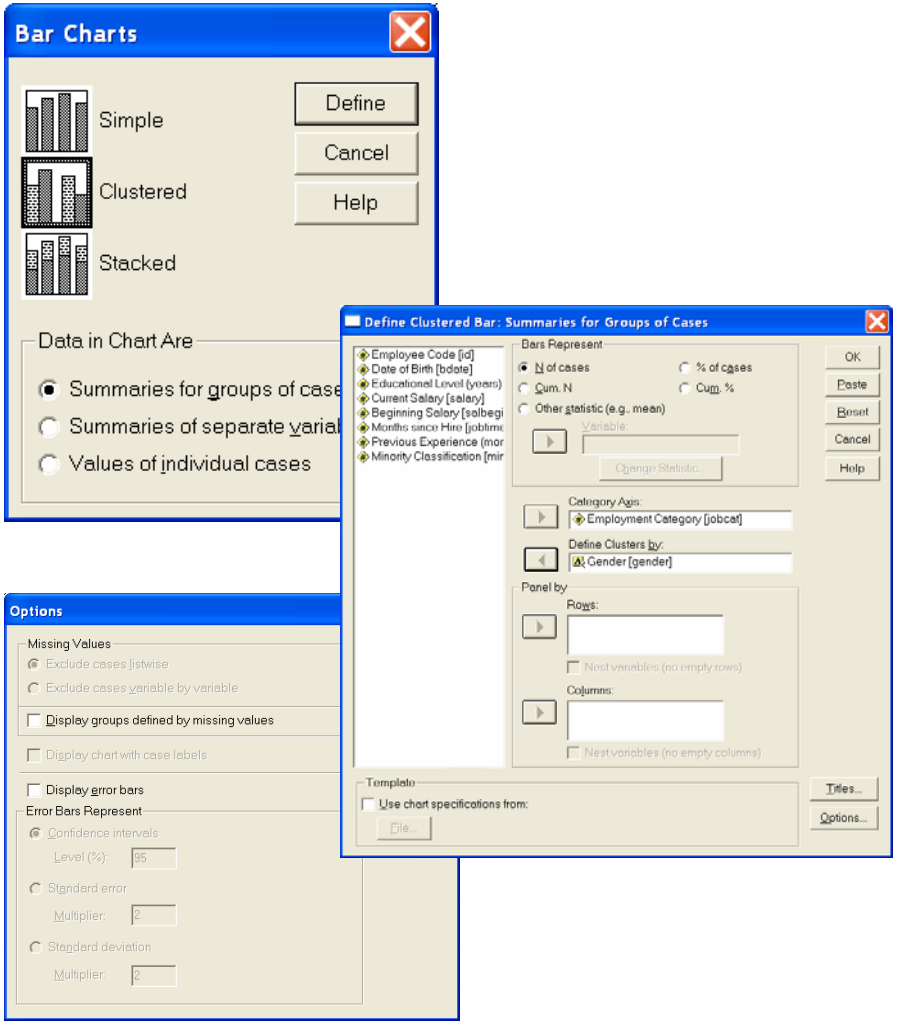


SPSS Output: Simple Bar Chart of 'JOBCAT'

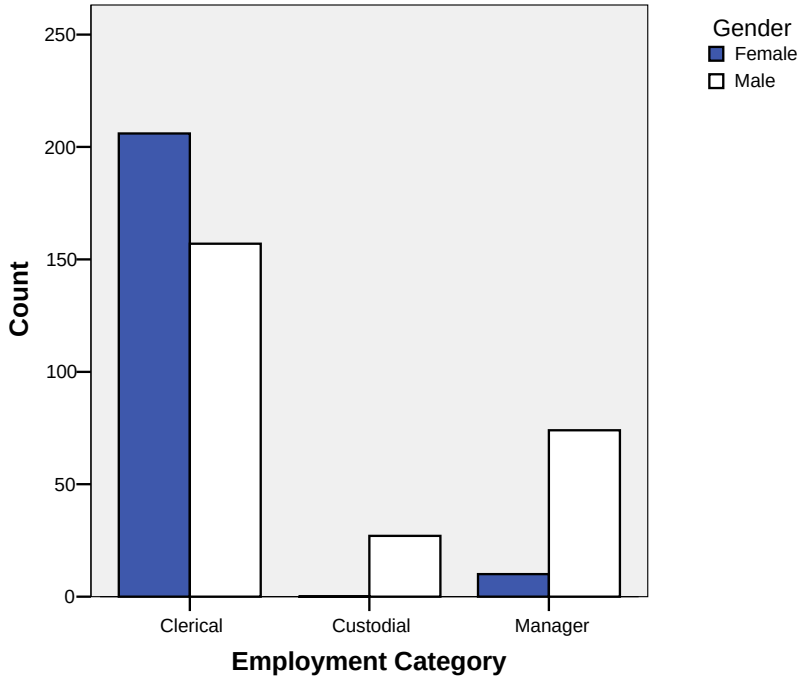


BAR CHART EXAMPLE 3: METHOD 2 -Obtain a **Clustered Bar Chart** of the variable **'JOBCAT'** by the variable **'GENDER'** via the **Graphs menu**.

Dialog Box Input: From the Graphs Menu select **Bar** (Chart). Select Clustered, then the Define button. Move the variable **'JOBCAT'** into the **Category Axis** cell (x-axis) and **'GENDER'** into the **Define Clusters by** cell. Use the **Options** button to turn off the inclusion of bars for missing values

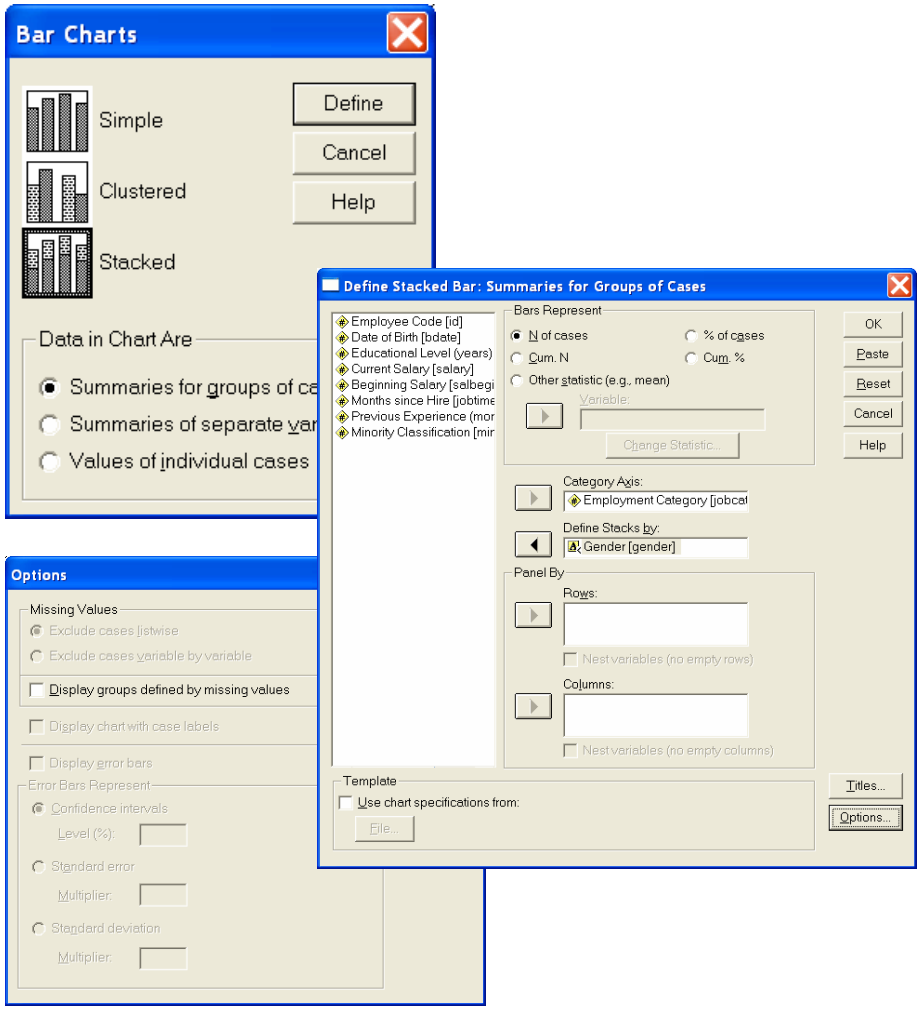


SPSS Output: Clustered Bar Chart of 'JOBCAT' by 'GENDER'

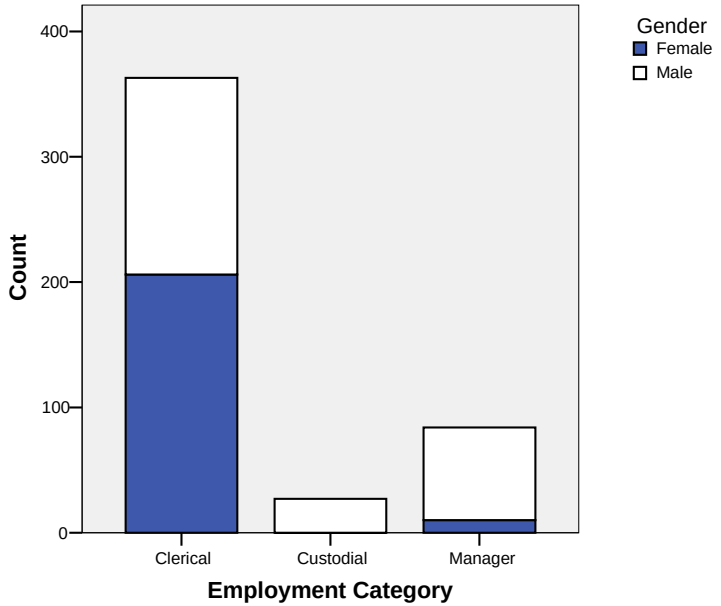


BAR CHART EXAMPLE 4: METHOD 2 - Obtain a **Stacked Bar Chart** of the variable **'JOBCAT'** by the variable **'GENDER'** via the **Graphs menu**.

Dialog Box Input: From the Graphs Menu select **Bar** (Chart). Select Stacked, then the Define button. Move the variable **'JOBCAT'** into the **Category Axis** cell (x-axis) and **'GENDER'** into the **Define Stacks by** cell. Use the **Options** button to turn off the inclusion of bars for missing values



SPSS Output: Stacked Bar Chart of 'JOBCAT' by 'GENDER'



BINOMIAL PROBABILITIES

To obtain probabilities for a variable having a Binomial Distribution, you will need to use the **Compute** function (see **Compute**), as well as some basic arithmetic.

DATA: According to a recent study, 56% (.56) of American mothers with children under the age of 6 years work outside the home.

NOTE: If you are using a blank worksheet, type a number (any number) into at least one cell of the worksheet, otherwise you will get an error message when you use the required **STEPS**.

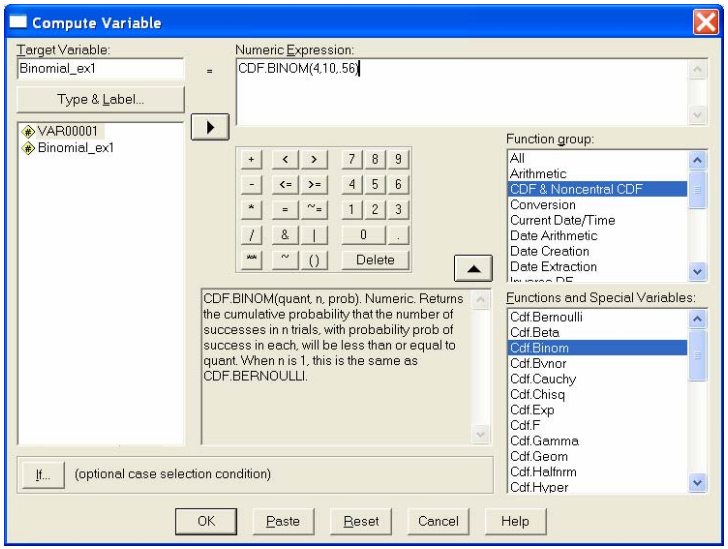
STEPS: Menu Bar => Transform => Compute

NOTE: In each of the examples below, the experiment will involve randomly selecting 10 American mothers with children under the age of 6 years, and asking the question "Do you work outside the home?" Let **x** represent the number of mothers in a sample of 10, who work outside the home.

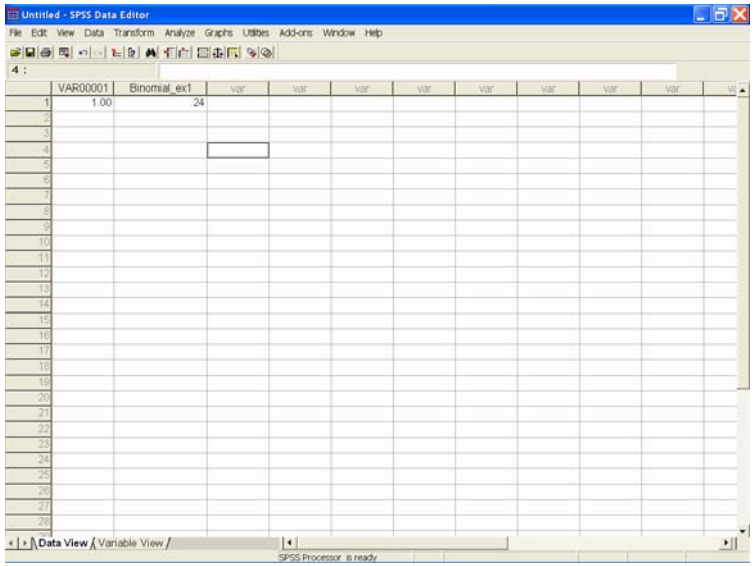
BINOMIAL PROBABILITY EXAMPLE 1: Find **P(x<4)**. Recall from above that n = 10 and p = .56.

SUMMARY OF STEPS: In the **Compute Variable** dialog box, select a name for the column in the worksheet where your results will be stored. Type this name in the **Target Variable** box (below **'BINOMIAL_EX1'**). => From the **Function Group** list select **CDF & Noncentral CDF.** => From the **Functions and Special Variables** list select **CDF.BINOM** and double click on it to bring it into the **Numeric Expression** box. => In the **Numeric Expression** box, the following will appear: **CDF.BINOM (?, ?, ?).** The three question marks are there for you to supply the values of **q, n, and p** respectively, where **q** is the number of successes, **n** is the number of trials, and **p** is the probability of success. To supply these values, **click on the respective question marks,** and type in the values for **q, n, p (4, 10, .56). =>** Click **OK.** => The probability will be displayed in a new column in the Data Editor (in this case a column named **'BINOMIAL_EX1'**).

Dialog Box: Target Variable is named and **Numeric Expression** is selected. Question marks are then replaced with values for **q, n,** and **p**.



SPSS Output: P(x <4) = .24



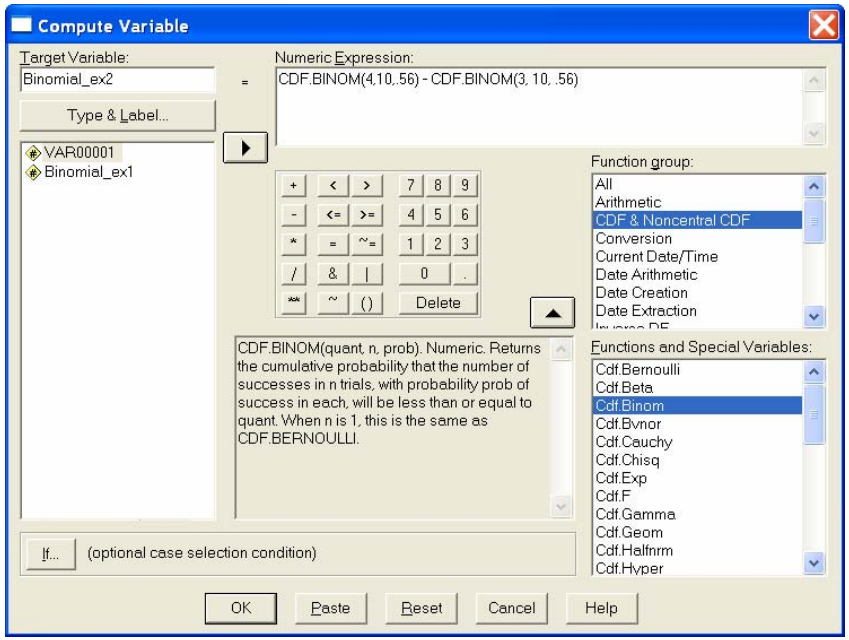
BINOMIAL PROBABILITY EXAMPLE 2: Find **P(x = 4).** Recall from above that n = 10 and p = .56.

NOTE: Since SPSS deals with the cumulative distribution function, **P(x = 4) = P(x < 4) – P(x < 3)**, therefore the formula you use in the **Numeric Expressions** box must reflect this.

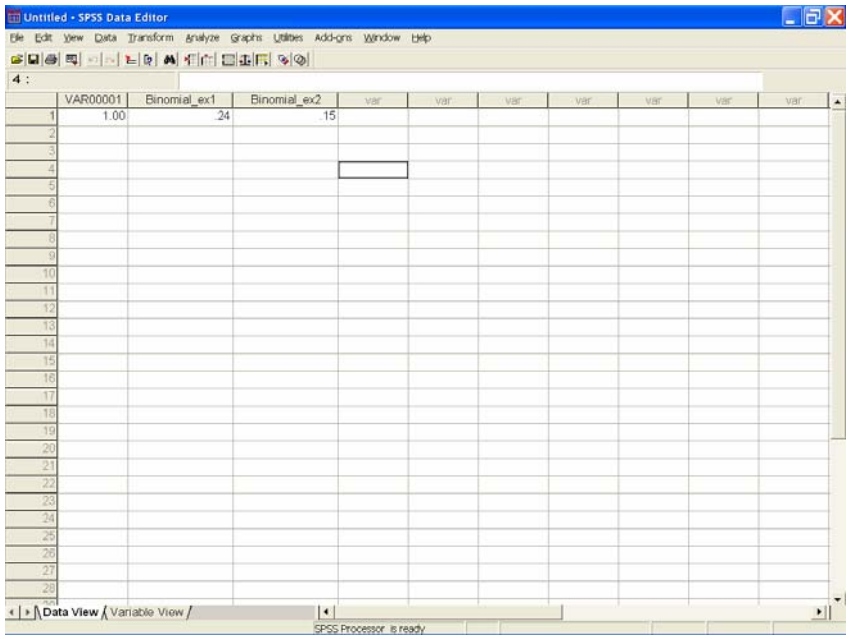
STEPS: Menu Bar => Transform => Compute

SUMMARY OF STEPS: For this example, name your **Target Variable 'BINOMIAL_EX2'.** => From the **Function Group** list select **CDF & Noncentral CDF. =>** From the **Functions and Special Variables** list select **CDF.BINOM** and double click on it to bring it into the **Numeric Expression** box. => Enter a "-"(minus) after the function and then enter the function a second time. At this point the **Numeric Expression** box should include: **CDF.BINOM (?, ?, ?) - CDF.BINOM(?, ?, ?)** . The question marks are there for you to supply the values of **q, n, and p** respectively, where **q** is the number of successes, **n** is the number of trials, and **p** is the probability of success. To supply these values, **click on the respective question marks,** and type in the values for **q, n, p**. Since the first expression will be **P(x<4)**, the values for **q, n, and p** will be **4, 10, and .56** respectively. Since you will be subtracting **P(x<3)**, the values of **q, n, and p** will be **3, 10, and .56** respectively. => Click **OK.** => The probability will be displayed in a new column in the Data Editor (in this case a column named **'BINOMIAL_EX2'**).

Dialog Box: Target Variable is named and **Numeric Expression** is selected. Values for **q, n, and p** are input.



SPSS Output: $P(x = 4) = .15$



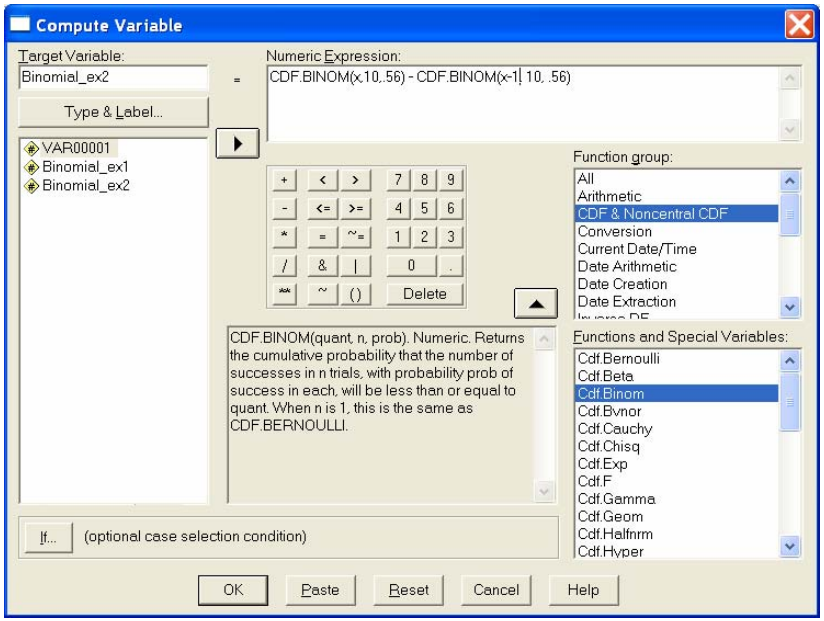
BINOMIAL PROBABILITY EXAMPLE 3: Obtain a table of binomial probabilities for the example used in this section.

NOTE: You will first need to create a variable named **'x'** that has the numbers 0 through 10 in the first 11 cells.

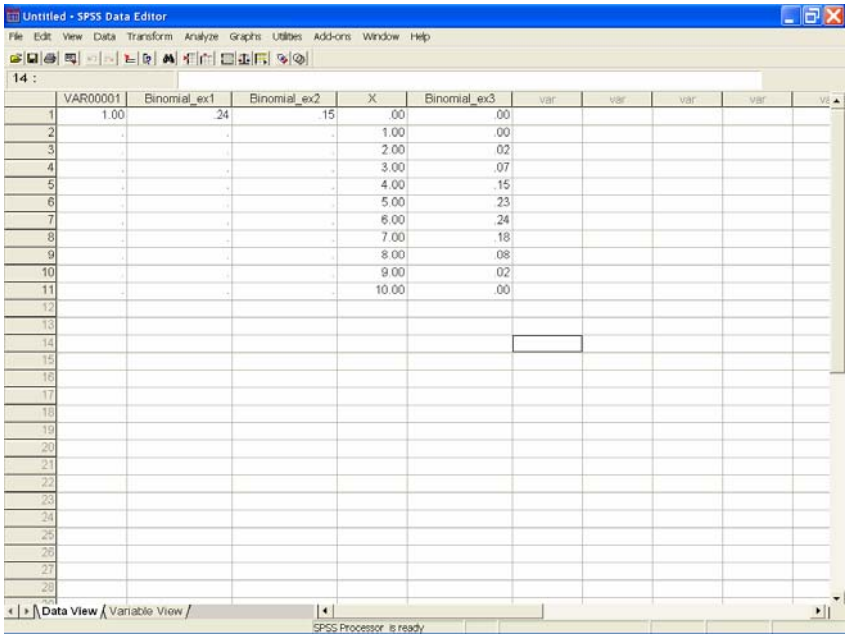
STEPS: Menu Bar => Transform => Compute

SUMMARY OF STEPS: For this example first create the variable **'X'** (see Note). => Name your **Target Variable 'BINOMIAL_EX3'.** => From the **Function Group** list select **CDF & Noncentral CDF. => F**rom the **Functions and Special Variables** list select **CDF.BINOM** and double click on it to bring it into the **Numeric Expression** box. => Enter a "-" (minus) after the function and then enter the function a second time. At this point the **Numeric Expression** box should include: **CDF.BINOM(?, ?, ?) - CDF.BINOM(?, ?, ?)** . The question marks are there for you to supply the values of **q, n, and p** respectively, where **q** is the number of successes, **n** is the number of trials, and **p** is the probability of success. To supply these values, **click on the respective question marks,** and type in the values for **q, n, p**. Since this is a cumulative function and we are seeking the probabilities for every value, **x** and **x – 1** represent the values of **q. (**Example 2 represented an example of finding one value of **'X**.') The first expression will be have the values **x, 10, and .56** for **q, n, and p** and represents the cumulative distribution through the value of '**X.'** Since you will be subtracting the cumulative distribution of the prior value of **'X,'** the resulting probability is that of the specific value. For the second set of question marks the values of **q, n, and p** will be **x-1, 10, and .56** respectively. => Click **OK.** => The probability will be displayed in a new column in the Data Editor (in this case a column named **'BINOMIAL_EX3'**).

Dialog Box: Target Variable is named and **Numeric Expressions** are selected. Values for **q, n, and p** are input.



SPSS Output: Probability distribution is displayed.



BOXPLOT (BOX-AND-WHISKER PLOT)

Based upon the Five-Number Summary, the **Boxplot** is used to graphically display a data set's **Minimum, First Quartile (Q_1), Median (Q_2), Third Quartile (Q_3),** and **Maximum** values. In SPSS, the "whiskers" extend to no more than 1.5 times the value of the Interquartile Range (IQR = $Q_3 - Q_1$) above Q_3 or below Q_1. Any values beyond this are considered outliers, and are labeled by case number.

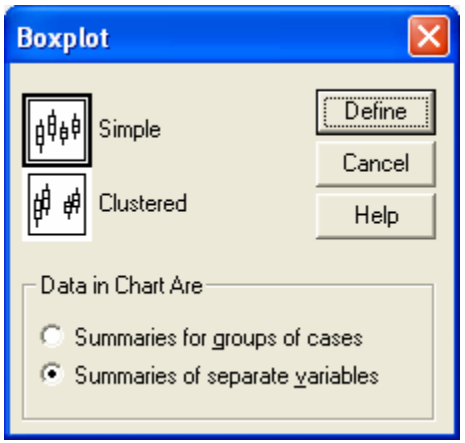
DATA: Cars.sav

STEPS: Menu Bar => Graphs => Boxplot

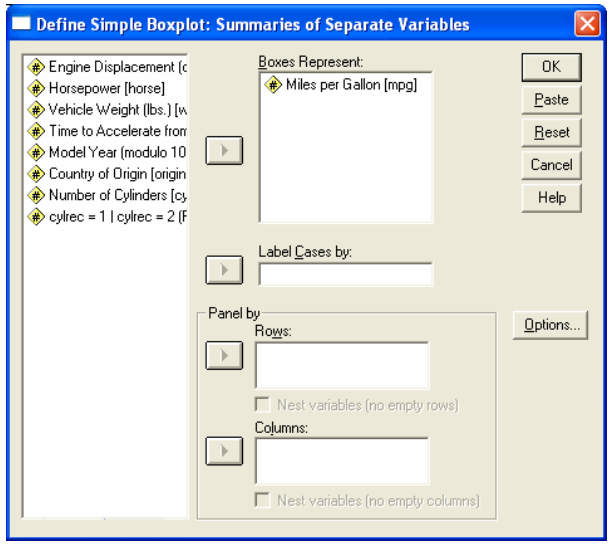
BOXPLOT EXAMPLE 1: Obtain a **Boxplot** for the variable **'MPG.'**

SUMMARY OF STEPS: In the **Boxplot** dialog box, select **Simple** by clicking once on it. => Select **Summaries of Separate Variables** by clicking on the button to the left of it. => Click **Define.** => In the **Define Simple Boxplot** dialog box, select your variable by clicking once on it and then clicking on the right arrow next to the **Boxes Represent**. => Click **OK.** => The **Boxplot** will appear in the Output Viewer.

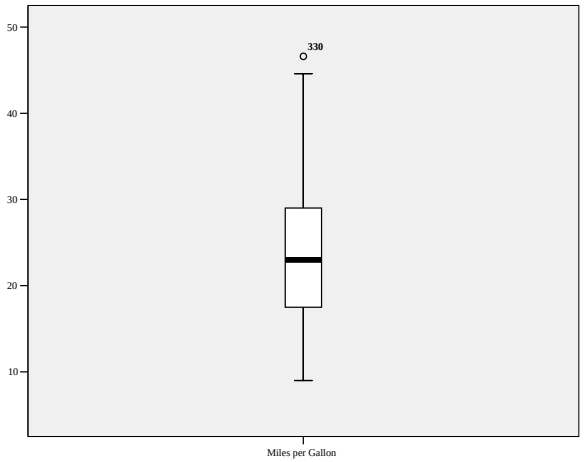
Dialog Box: Boxplot type is selected.



Dialog Box: Variable is selected.



SPSS Output: Boxplot in the output viewer.
(Here 330 represents the case number of an outlier.)

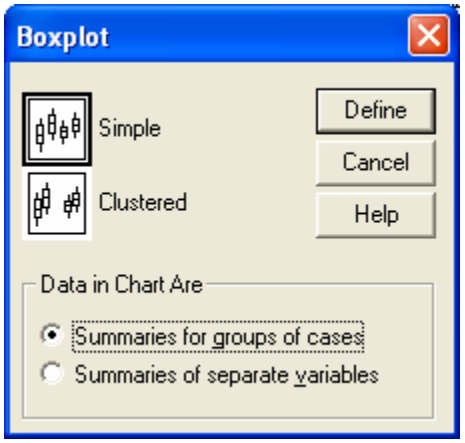


BOXPLOT EXAMPLE 2: Obtain a **Grouped Boxplot** of **'WEIGHT'** by **'ORIGIN.'**

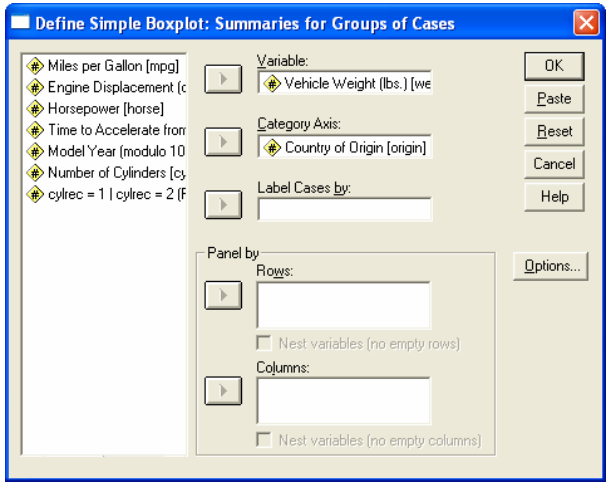
STEPS: Menu Bar => Graphs => Boxplot

SUMMARY OF STEPS: In the **Boxplot** dialog box, select **Simple** by clicking once on it. => Select **Summaries for groups of cases** by clicking on the button to the left of it. Click **Define.** => In the **Define Simple Boxplot** dialog box, select your variable **by clicking once on it and then clicking the right arrow next to the Variable** box . Now select your category variable (that which the boxplots will be grouped by) by clicking once on it and then clicking the right arrow next to the **Category Axis** box. => Click **OK.** => The **Grouped Boxplot** will appear in the output viewer. (**NOTE**: To avoid having a Boxplot for Missing Values, select the Options button and clear the **Display groups defined by missing values** box.)

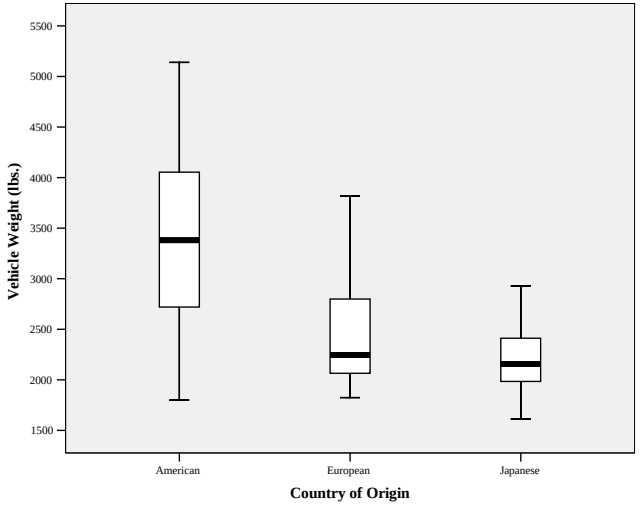
Dialog Box: Boxplot type is selected.



Dialog Box: Variables are selected.



SPSS Output: Grouped Boxplot in output viewer.



CHI-SQUARE (χ^2)

CHI-SQUARE GOODNESS OF FIT TEST: This test uses frequency data to test hypotheses about population proportions. That is, do our observed sample proportions coincide with the proportions stated by the null hypothesis, or are they significantly different?

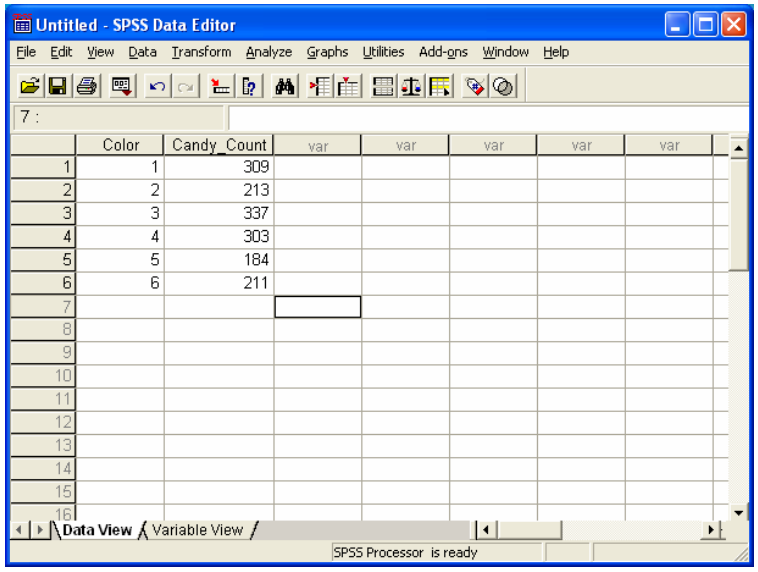
DATA: Students in an Introductory Statistics class were given several bags of M&Ms Plain Candies. They were instructed to sort and count by color. A sum for each color was calculated.

STEPS: (Note that before proceeding here, categories and their corresponding frequencies must be entered into a worksheet.) **1) Enter Data. 2) Data => Weight Cases. 3) Analyze => Nonparametric Tests => Chi-Square**.

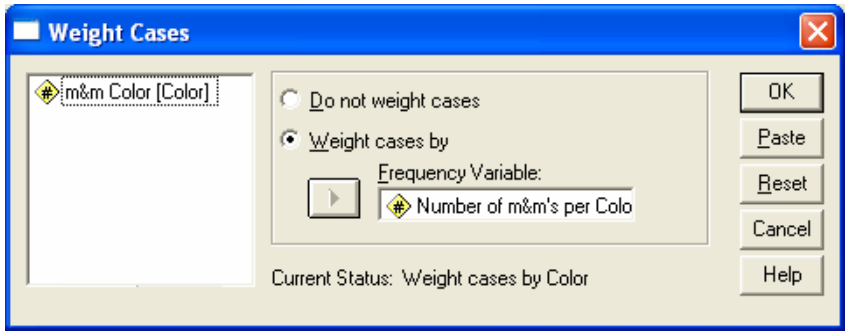
EXAMPLE 1 - $\chi 2$ GOODNESS OF FIT TEST: M&M Mars states that the proportions by color, in bags of Plain m&m's candies is as follows: Blue, .24; Brown, .13; Green, .16; Orange, .20; Red, .13; and Yellow, .14. We will use the **Goodness of Fit Test** to test this hypothesis. The colors are coded as follows in the data sheet: 1 = Blue, 2 = Brown, 3 = Green, 4 = Orange, 5 = Red, 6 = Yellow.

SUMMARY OF STEPS: 1) Enter the data into the Data Editor. **2)** From the Main Menu, select **Data.** => **Weight Cases. Click** on the **Weight cases by** button. Move the variable **containing the frequencies** into the **Frequency Variable** box, by clicking on the variable, and then clicking the right arrow next to the **Frequency Variable** box. => Click **OK.** You will not see anything happen after clicking OK, other than the dialog box disappearing from the screen. (**NOTE**: **Weight Cases** remains in effect until it is deactivated by selecting the **Do not weight cases** option, so remember to come back and turn Weight Cases off after you are done with your analysis.)

1) Data entered into the Data Editor



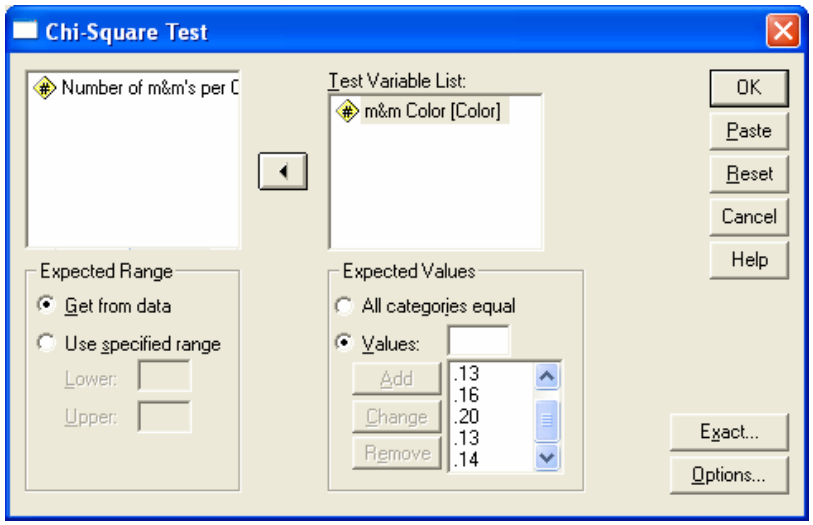
2) Dialog Box: Frequency variable entered into "Frequency Variable" box.



3) Next, select **Analyze => Nonparametric Tests => Chi-Square.** Select **'COLOR'** as the test variable, and move it into the **Test Variable List,** by clicking on the arrow next to the **Test Variables List** box. => Click on the **Values** button, in the **Expected Values** box. You will now enter the **expected proportions**. It is **EXTREMELY** important that you enter the expected proportions in a specific manner. **The first value of the list must correspond to the lowest value of the test variable**. In this example the first value in the list corresponds to the proportion of Blue M&Ms, which is coded with the number 1. Type the proportion of Blue M&Ms, in the **Values** box. Then click **Add.** Continue in this manner for brown, green, orange, red, and yellow, respectively. => Click **OK** and the results of the test will be displayed in the Output Viewer.

NOTE: If all frequencies are equal, values will not need to be entered as described above. Just select the **All categories equal** button, as opposed to the **Value** button.

Dialog Box: Variable is selected, and values are entered.



SPSS Output: Results of Chi-Square.

m&m Color

	Observed N	Expected N	Residual
Blue	309	373.7	-64.7
Brown	213	202.4	10.6
Green	337	249.1	87.9
Orange	303	311.4	-8.4
Red	184	202.4	-18.4
Yellow	211	218.0	-7.0
Total	1557		

Test Statistics

	m&m Color
Chi-Square[a]	44.875
df	5
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 202.4.

CHI SQUARE TEST OF INDEPENDENCE: This test is used to examine whether there exists an association between two qualitative variables.

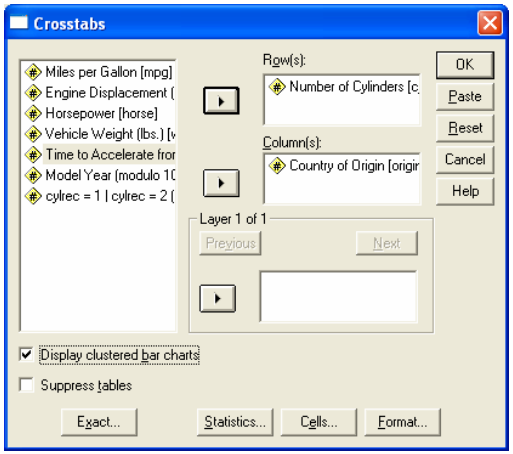
DATA: Cars.sav

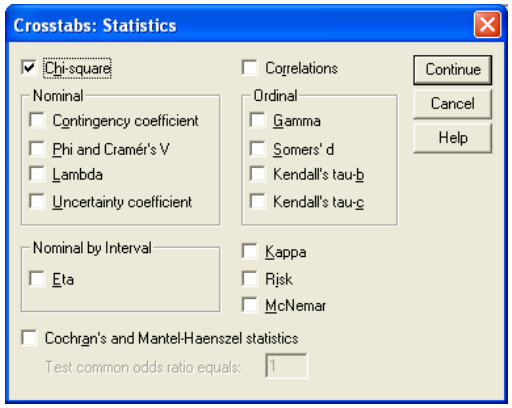
STEPS: Menu Bar => Analyze => Descriptive Statistics => Crosstabs => Enter variables => Statistics button => select Chi-Square => Continue => OK

EXAMPLE 2 - $\chi 2$ TEST OF INDEPENDENCE: Is there a relationship between the number of cylinders a car engine has and its country of origin?

SUMMARY OF STEPS: From the Main Menu select **Analyze** => **Descriptive Statistics** => **Crosstabs** and enter dependent (row) and independent (column) variables. In this case one might consider the **'ORIGIN'** (Country of Origin) the independent variable and **'CYLINDERS'** (Number of Engine Cylinders) as the dependent variable. => Select the **Statistics** button, check **Chi-Square**, and **Continue**. If you do not want a Crosstabs Table check the **Suppress Tables** box. Select other Crosstabs options as desired, then **OK**. THE test results will be output to the Output Viewer window.

Dialog Boxes: Variables entered; Chi-Square selected.





SPSS Output: Crosstabs table (including cell percentages) and results of Chi-Square.

Number of Cylinders * Country of Origin Crosstabulation

Count

		Country of Origin			
		American	European	Japanese	Total
Number of Cylinders	3 Cylinders	0	0	4	4
	4 Cylinders	72	66	69	207
	5 Cylinders	0	3	0	3
	6 Cylinders	74	4	6	84
	8 Cylinders	107	0	0	107
Total		253	73	79	405

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	185.794[a]	8	.000
Likelihood Ratio	217.125	8	.000
Linear-by-Linear Association	129.770	1	.000
N of Valid Cases	405		

a. 6 cells (40.0%) have expected count less than 5. The minimum expected count is .54.

CONFIDENCE INTERVAL FOR A POPULATION MEAN

To construct confidence intervals for a population mean, SPSS uses the One-Sample T-Interval Procedure.

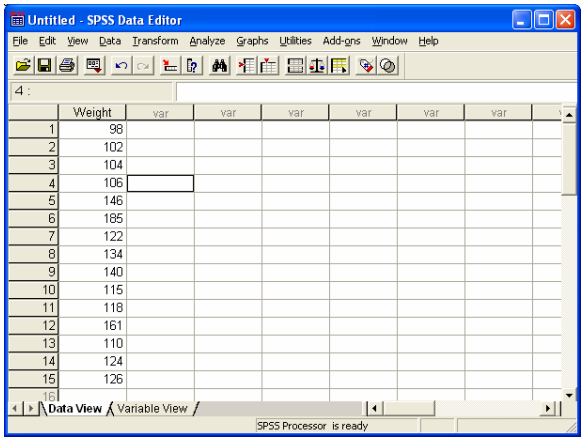
DATA: (Not an SPSS data file.) A random sample of 15 freshman women at a local college yielded a mean weight of 126.07 pounds, with a standard deviation of 24.01 pounds.

STEPS: Menu Bar => Analyze => Descriptive Statistics => Explore

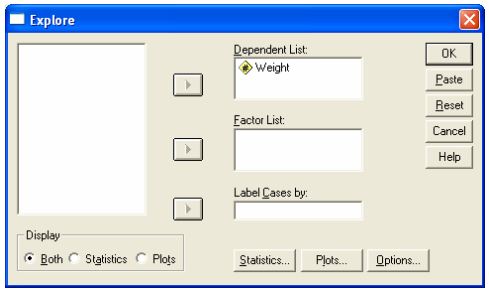
EXAMPLE: Construct a 95% confidence interval for the mean weight, μ, of all freshman women at this college.

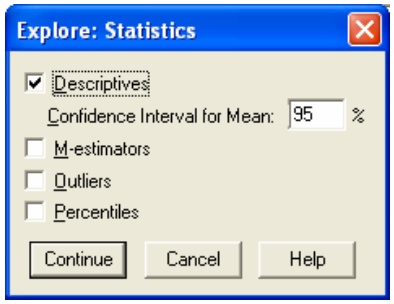
SUMMARY OF STEPS: In the **Explore** dialog box, select your variable by **clicking** once on it and then clicking the right arrow next to the **Dependent List** box. => Click on the **Statistics** option (not the button in the **Display** section of the dialog box). => In the **Explore: Statistics** dialog box, click on the **Descriptives** button. => In the **Confidence Interval for Mean** box, type in the required confidence level. => Click **Continue.** => Click **OK** in the **Explore** dialog box. The confidence interval (along with other descriptive measures of the data) will appear in the Output Viewer.

Data: Weights of 15 women.



Explore Dialog Box: Variable is selected.
Statistics Dialog Box: "Descriptives" and confidence level are selected.





SPSS Output:
95% C.I. = $(112.77 \le \mu \le 139.36)$ lbs.

Descriptives

			Statistic	Std. Error
Weight	Mean		126.07	6.199
	95% Confidence Interval for Mean	Lower Bound	112.77	
		Upper Bound	139.36	
	5% Trimmed Mean		124.35	
	Median		122.00	
	Variance		576.495	
	Std. Deviation		24.010	
	Minimum		98	
	Maximum		185	
	Range		87	
	Interquartile Range		34	
	Skewness		1.176	.580
	Kurtosis		1.254	1.121

CONFIDENCE INTERVAL FOR A POPULATION PROPORTION

Unlike a confidence interval for a population mean, SPSS does not have a specific (straightforward) procedure to find a confidence interval for a proportion. However it can be done using the **Compute** command and the appropriate formula.

DATA: (Not an SPSS file). In a random sample of 1234 Americans, 70% were in favor of the death penalty.

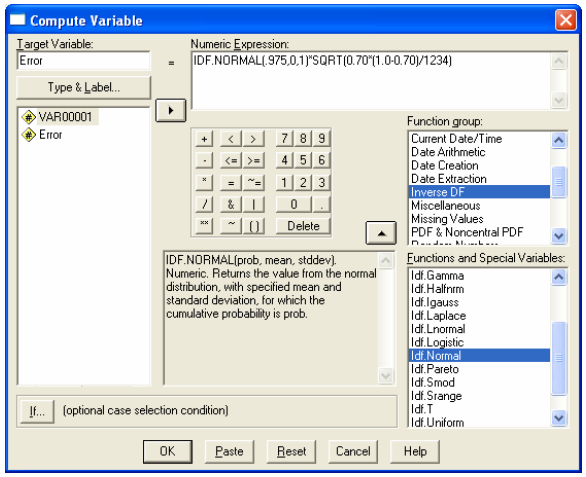
NOTES: 1) If you are using a blank worksheet, type a number (any number) into at least one cell of the worksheet, otherwise you will get an error message when you use the required STEPS. **2)** Consider using a calculator to determine this proportion.

STEPS: Menu Bar => Transform => Compute

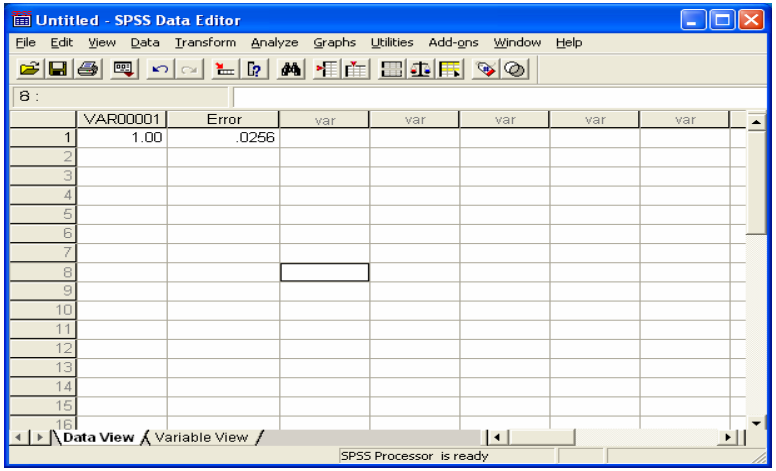
EXAMPLE: Construct a 95% confidence interval for the true proportion **p**, the percentage of all Americans who are in favor of the death penalty.

SUMMARY OF STEPS: In the **Compute Variable** dialog box, you will be calculating the margin of error. => In the **Target Variable** box, type in a name for the variable where the margin of error will appear (here, it is called **'ERROR'**). => Select from the Function Group **"Inverse DF"** and from the **Function and special Variables** select **"Idf.Normal."** => In the Numeric Experssion box the three question marks (IDF.NORMAL(?,?,?)) are replaced by α (here, 1-(0.05/2)=.975), 0 (mean), and 1 (std. dev.). At this point the expression reads: **IDF.NORMAL(.975,0,1)**. Immediately following the preceding, type * **SQRT(0.70 * (1-0.70)/1234)**, where 0.70 is the proportion (p) of Americans in the sample who favor the death penalty, 1-0.70 is the complement of p, and 1234 is the sample size. The equation now reads:
IDF.NORMAL(.975,0,1) * **SQRT(0.70 * (1-0.70)/1234)**. => Select **OK** and the margin of error will appear in a Data Editor column labeled **'ERROR.'**

Compute Dialog Box: Enter the Target variable here named **'ERROR'** and formula for the margin of error.



SPSS Output: Compute statements create a new variable. As such, the output is located in the Data Editor. The margin of error is approximately 0.026, therefore the 95% confidence interval for **p** = .70 +/- .26 or **p** = $(.674 \leq \mathbf{p} \leq .726)$



CROSSTABS (a.k.a. CONTINGENCY TABLES)

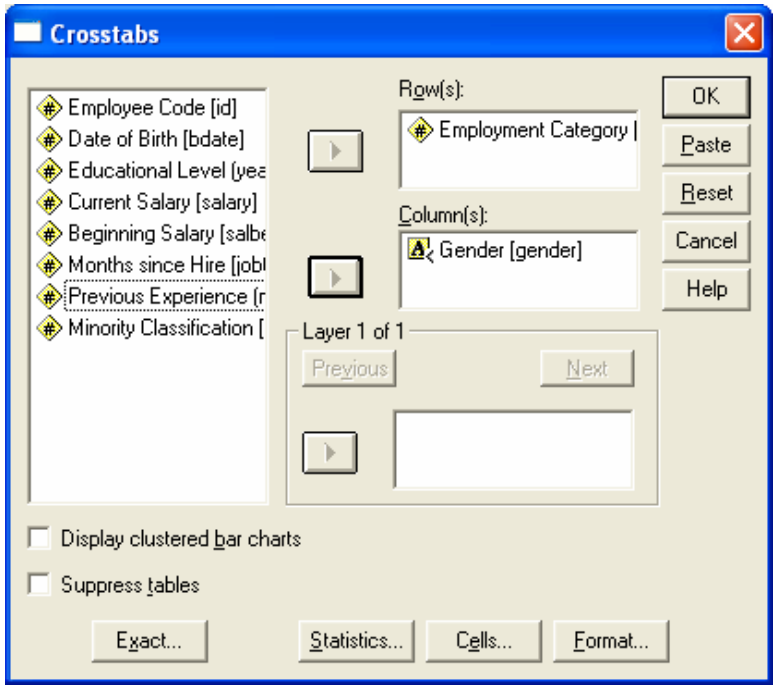
The **Crosstabs** procedure allows you to test measures of association between two categorical variables. Generally, the variable considered to be the independent variable is the columns variable, while the dependent variable is placed in the rows. (However, if placed in the opposite locations, the resulting statistical data is the same.)

DATA: Employee data.sav

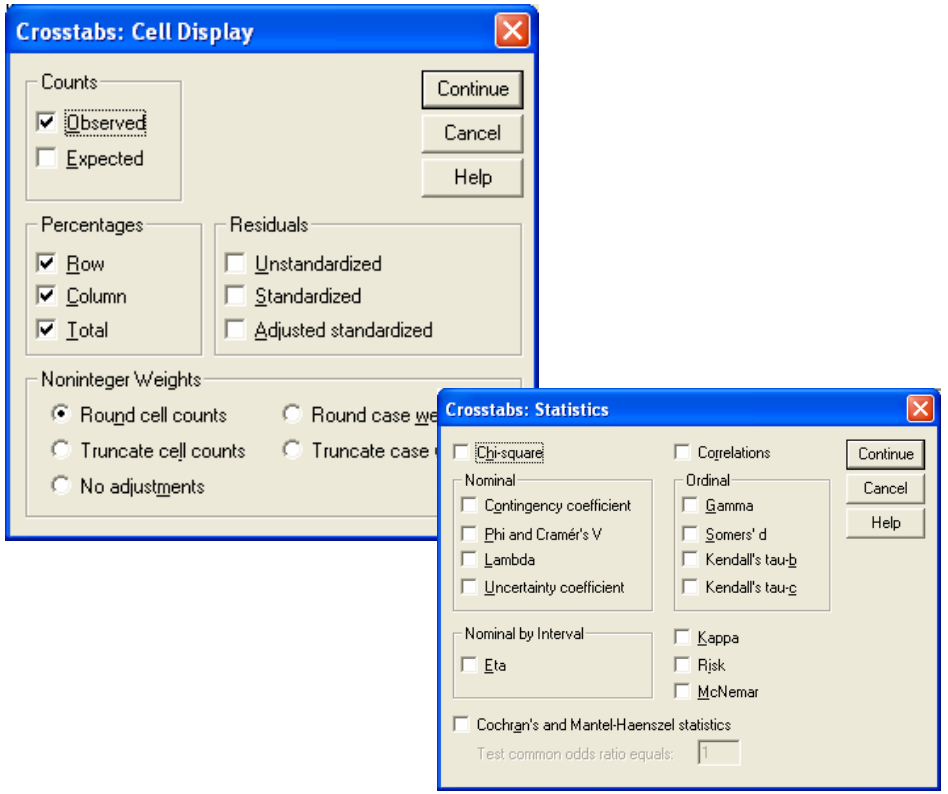
STEPS: Menu Bar => Analyze => Descriptive Statistics => Crosstabs

EXAMPLE: Obtain a Crosstabs of '**JOB CATEGORY**' by '**GENDER**.'

Dialog Box Input: Move the **v**ariable **'JOB CATEGORY'** into the Row(s) Cell and **'GENDER'** is placed into the Column(s) Cell. Percentages for cells and statistics are selected through the buttons at the bottom of the dialog box.



Cell and Crosstabs Statistics options:



Crosstabs Output:

Employment Category * Gender Crosstabulation

			Gender		
			Female	Male	Total
Employment Category	Clerical	Count	206	157	363
		% within Employment Category	56.7%	43.3%	100.0%
		% within Gender	95.4%	60.9%	76.6%
		% of Total	43.5%	33.1%	76.6%
	Custodial	Count	0	27	27
		% within Employment Category	.0%	100.0%	100.0%
		% within Gender	.0%	10.5%	5.7%
		% of Total	.0%	5.7%	5.7%
	Manager	Count	10	74	84
		% within Employment Category	11.9%	88.1%	100.0%
		% within Gender	4.6%	28.7%	17.7%
		% of Total	2.1%	15.6%	17.7%
Total		Count	216	258	474
		% within Employment Category	45.6%	54.4%	100.0%
		% within Gender	100.0%	100.0%	100.0%
		% of Total	45.6%	54.4%	100.0%

If all percentages have been included in the table it can appear somewhat overwhelming, as every data cell in the table will contain a category count and three percentages. You can associate the stacked sets of three percentages located to the right of a row value by remembering **CRCT**.

- **C**: The top number in a data cell represents the **C**ount or number of occurrences for which both variables occurred with the column and row values.
- **R**: The top percentage refers to **R**ow data - what percentage of that value's (row's) respondents are counted in that cell. The total for the row is located in the column furthest right and is referred to as a margin value. Dividing the count for a given cell in the row by the margin value (total) would yield the top percent.
- **C**: The middle percentage refers to **C**olumn data - what percent of that column's respondents are accounted for by the value of the cell. The total for a column value is located at the bottom of the table. Dividing the count for a given cell in a column by the margin value (total) at the bottom would yield the middle percent.
- **T**: The bottom of the three percentages represents a sub-total for the whole table - what percent of all cases a given cell represents. The total number of cases is located in the lower right cell. Dividing the number in any given cell by the total number of cases yields the bottom percent.

Margin Values and Table Totals: The values in the right column represent the sum of the cells across the table. The values across the bottom of the table sum the columns. Both of these sets of values are sometimes referred to as margin values. The lower right cell contains the totals for the entire table.

Reading a Crosstabs Table:

So, how do you make sense of a Crosstabs table? If the headings leave you still uncertain, try the following approach, which uses examples taken from the following table.

What percent of individuals earning over $39,999 are not minority employees? (A **R**ow question)

> Start by going to the Salary value of > $39,999. Go across to the Minority value of "No." In this cell there are 97 employees. Of the 104 employees earning > $39,999, 97 or 93.3% are not minority employees. (Check: 97/104 = .933 =>93.3%)
>
> Is that the correct percentage to select? You can always check by doing the math. It is clear that there are 97 non-minority employees earning >$39,999. So, would those 97 individuals be equal to 97/104 (97/right margin value = total number earning > $39,999), or 97/370 (97/the bottom column total, representing the total number of non-minority employees – 26.2%), or 97/474 (97/the number of non-minority employees earning >$39,999/ all employees = 20.5%)?

What percent of non-minority employees (value = No) earned < $20,000? (A **C**olumn question)

Start by going to the Minority value of "No" column. Go down this column until it crosses with the second variable, Salary, and the value of Salary being referenced (< $20,000). In this cell there is a count of 22. Meaning that 22 non-minority employees earned less than $20,000. They represent 5.9% of all non-minority employees. (Check: 22/370 = .059 => 5.9%)

Minority employees represent what percent of the total workforce? (A **T**otal question)

While this question uses margin values, the same process occurs as for any other cell. Find the requested cell by crossing the minority value of "Yes" with the total row of the table (at the bottom). There are 104 minority employees, representing 21.9% of the workforce. (Check: 104/474 = .219 =>21.9%)

How many Minority employees earn less than $40,000? (A **C**ount question)

Less than $40,000 will include the first two salary ranges in the table. There are 11 Minority employees who earn less than $20,000, and 86 Minority employees who earn $20,000-$39,999. Therefore the number of Minority employees earning less than $39,000 is 97.

Try it out - (answers below)

1) What percent of non-minority employees earn $20,000 - $39,999?
2) Minority employees earning < $20,000 represent what percent of all employees?
3) Minority employees earning < $20,000 represent what percent of all minority employees?
4) Employees earning $20,000 - $39,999 represent what percent of all employees?
5) In the "Yes" minority column, 25.5% represents what measure?
6) Non-minority (No) employees represent what percent of those earning $20,000 - $39,999?
7) Among employees earning > $39,999, what percent are minority employees?
8) How many employees earn < $20,000?
9) What doe the number 104 in the bottom marginal values represent?
10) How many minority employees earn < $20,000?

ANSWERS:
1) Column question – 67.8% (Check: 251/370 = .678 => 67.8%).
2) Totals question – 2.3% (Check: 11/474 = .023 => 21.3%).
3) Column question – 10.6% (Check: 11/104 = .106 => 10.6%).
4) Totals question – 71.1% (Check: 337/474 = .711 => 71.1%).
5) Row question - 25.5% represents the percent of those earning $20,000 - $39,999 who are minority employees (Check: 86/.337 = .255 => 25.5%).
6) Row question – 74.5% (Check: 251/337 = .745 => 74.5%).
7) Row question – 6.7% (Check: 7/104 = .067 => 6.7%).
8) Count question – From the left marginal total there are 33 employees earning < %20,000.
9) Column count – number of minority employees.
10) Column count – 11.

Employees by Salary Range * Minority Classification Crosstabulation

			Minority Classification		
			No	Yes	Total
Employees by Salary Range	< $20,000	Count	22	11	33
		% within Employees by Salary Range	66.7%	33.3%	100.0%
		% within Minority Classification	5.9%	10.6%	7.0%
		% of Total	4.6%	2.3%	7.0%
	$20,000 - $39,999	Count	251	86	337
		% within Employees by Salary Range	74.5%	25.5%	100.0%
		% within Minority Classification	67.8%	82.7%	71.1%
		% of Total	53.0%	18.1%	71.1%
	>$39,999	Count	97	7	104
		% within Employees by Salary Range	93.3%	6.7%	100.0%
		% within Minority Classification	26.2%	6.7%	21.9%
		% of Total	20.5%	1.5%	21.9%
Total		Count	370	104	474
		% within Employees by Salary Range	78.1%	21.9%	100.0%
		% within Minority Classification	100.0%	100.0%	100.0%
		% of Total	78.1%	21.9%	100.0%

DESCRIPTIVE STATISTICS

Descriptive Statistics are used to summarize or describe the important characteristics of a data set. Although there are many types of **Descriptive Statistics**, only the most basic (mean, median, mode, range, variance, standard deviation, quartiles, minimum, and maximum) are shown here. There are two basic ways to obtain descriptive statistics. Each approach provides slightly different options, as well as different output formats.

DATA: Employee data.sav

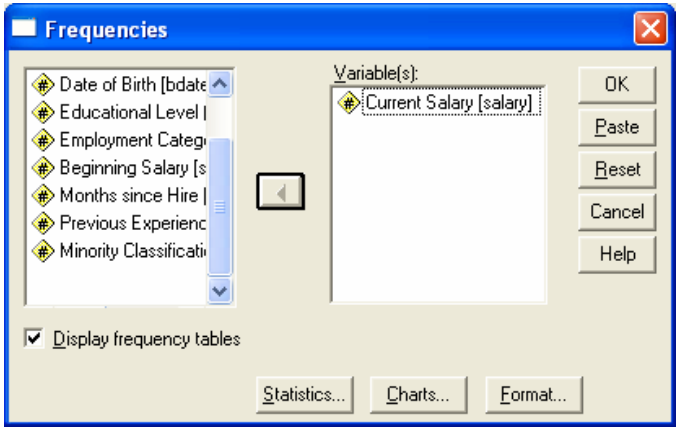
METHOD 1: Descriptives via Frequency command

STEPS: Analyze => Menu Bar => Analyze => Descriptive Statistics => Frequencies

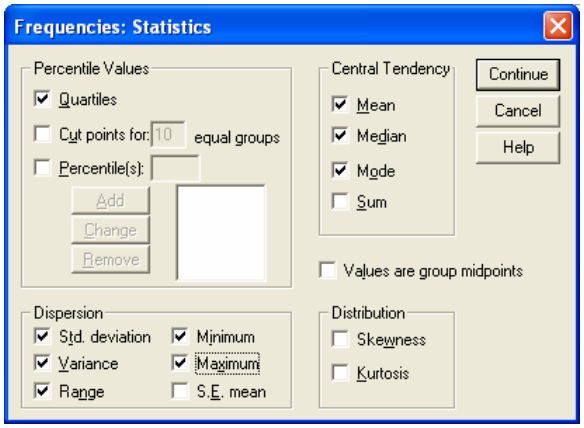
EXAMPLE 1: Obtain the mean, median, mode, range, variance, standard deviation, quartiles, minimum, and maximum for the variable **Salary.**

SUMMARY OF STEPS: In the **Frequencies** dialog box, select your variable by clicking once on it, and then clicking the right-arrow. => If you do not wish to have a frequency table displayed, remove the checkmark from the **Display frequency tables** box. You will get a warning from SPSS when you do this. When this happens, simply click **OK**. => Click the **Statistics** button. In the dialog box that will open, select Quartiles, Standard Deviation, Variance, Range, Minimum, Maximum, Mean, Median, and Mode, by clicking once in the boxes to the left of each. => Selecting **Continue** will return you to the **Frequencies** dialog box.. => Click **OK** and the **Descriptive Statistics** you selected will appear in the output viewer.

Dialog Box: Variable is selected.



Dialog Box: Descriptive Statistics are selected.



SPSS Output: Descriptive Statistics via Frequency command

Statistics

Current Salary

N	Valid	474
	Missing	0
Mean		$34,419.57
Median		$28,875.00
Mode		$30,750
Std. Deviation		$17,075.661
Variance		291578214.5
Range		$119,250
Minimum		$15,750
Maximum		$135,000
Percentiles	25	$24,000.00
	50	$28,875.00
	75	$37,162.50

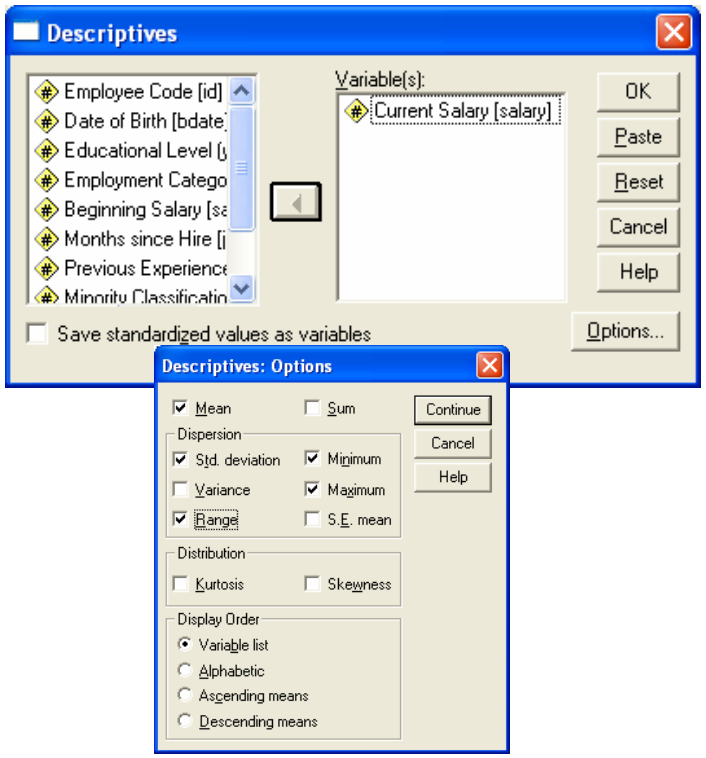
METHOD 2: Descriptives via Descriptive Statistics command

STEPS: Menu Bar => Analyze => Descriptive Statistics => Descriptives

EXAMPLE 2: Obtain **Descriptive Statistics** for the variable '**SALARY.**'

SUMMARY OF STEPS: In the **Descriptives** dialog box, select your variable by clicking once on it and then clicking on the right arrow next to the **Variables** box. => Click on **Options.** => In the **Descriptives Options** dialog box, select the descriptives you wish to obtain, by clicking once on the buttons to the left of each. In this example, mean, standard deviation, range, minimum, and maximum are selected. => Click **Continue.** => Click **OK** in the **Descriptives** dialog box and the statistics you selected will appear in the Output Viewer.

Dialog Box Variable is selected. Statistics are selected using the **Options** button.



SPS Output: Table of Descriptive Statistics.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Current Salary	474	$119,250	$15,750	$135,000	$34,419.57	$17,075.661
Valid N (listwise)	474					

DOTPLOT

A **Dotplot** is used with quantitative data to display the distribution of all values in a data set. For small data sets each dot represents a value of the variable of interest. For larger data sets each dot may represent more than one occurrence of a value.

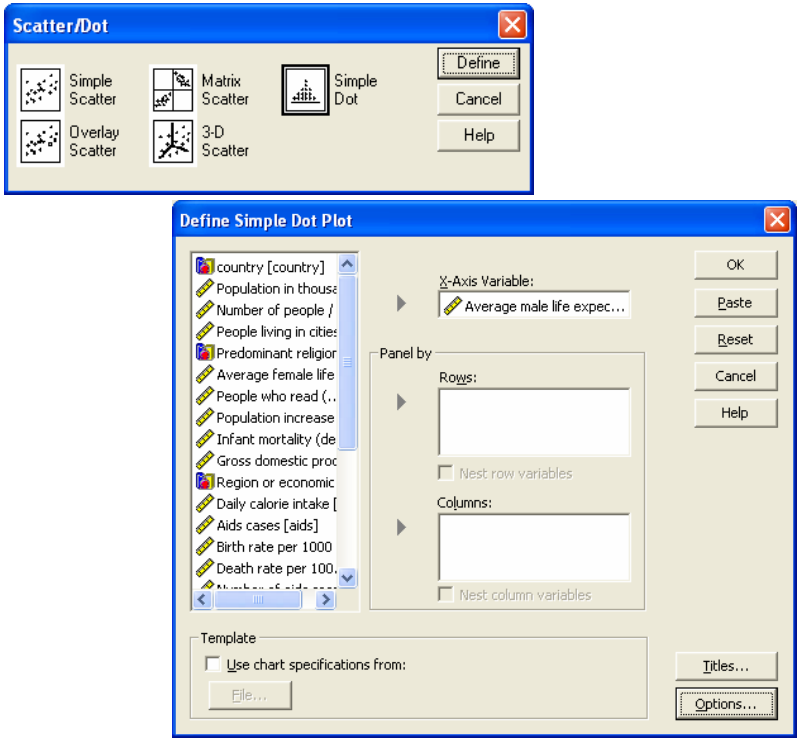
DATA: World95.sav

STEPS: Menu Bar => Graphs => Scatter/Dot

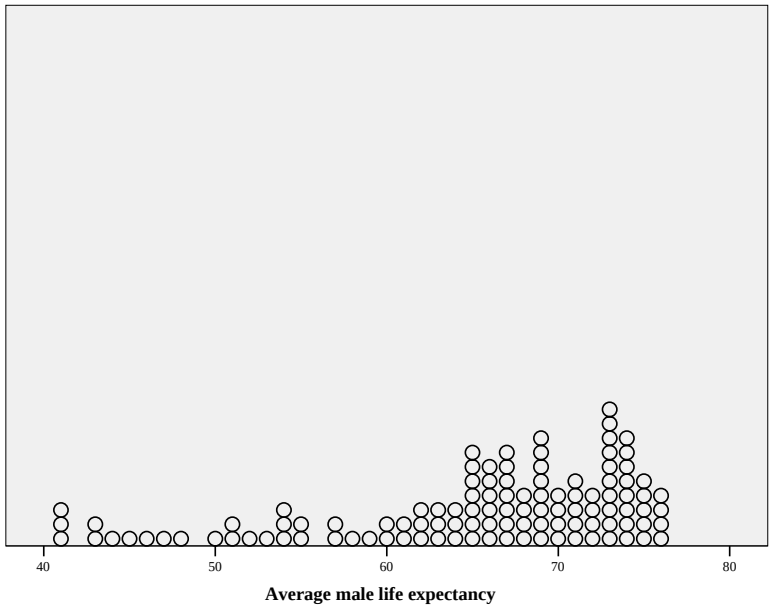
EXAMPLE: Create a **Dotplot** of the variables **'LIFEEXPM'** (Average Life Expectancy for Males).

SUMMARY OF STEPS: In the **Dotplot** dialog box, select **Simple** and then click **Define.** => Select the dependent variable by clicking once on it and then clicking the right arrow next to the **Y-axis.** Select the independent variable by clicking once on it and then clicking the right arrow next to the **X-axis.** => Click **OK** and the **Dotplot** will appear in the output viewer.

Dialog Boxes: Dotplot type is selected => select Define => select Variables.



SPSS Output: Dotplot of Average Life Expectancy for 109 countries.



FIVE-NUMBER SUMMARY

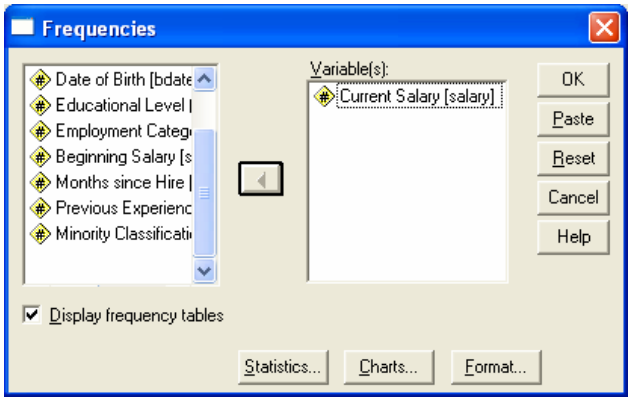
The ***Five-Number Summary*** for a data set consists of the following values: **Minimum, Q_1, Median (Q_2), Q_3, and Maximum**.

DATA: Employee data.sav

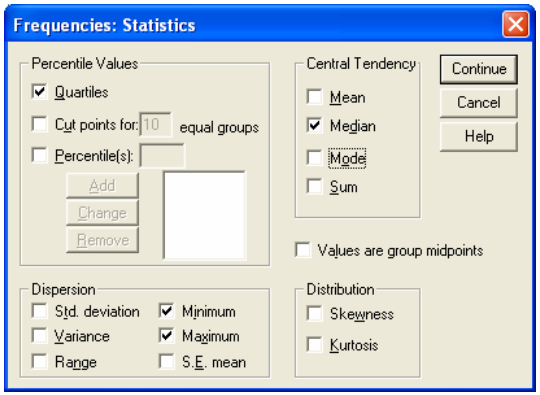
STEPS: Menu Bar => Analyze => Descriptive Statistics => Frequencies

SUMMARY OF STEPS: In the **Frequencies** dialog box, select the variable for which you wish to obtain the **Five-Number Summary** by clicking once on the variable and then click on the right arrow to the left of the **Variables** box. This will move your variable into the **Variables** box. Note that once you do this, the arrow will point to the left. => If in addition to the **Five-Number Summary**, you wish to obtain frequencies for your variable, click once inside the **Display Frequency Tables** box. => Now click once on the **Statistics** button. In the dialog box that opens, select **Quartiles, Minimum, Maximum,** and **Mean** by clicking once inside the box to the left of each. Note that since we have selected **Quartiles**, there is no need to select **Median**. => Click **Continue.** This will return you to the **Frequencies** dialog box. => In the **Frequencies** dialog box click **OK.** The **Five-Number Summary** will appear in the output viewer.

Dialog Box: Variable is selected.



Dialog Box: Select the Five-Number Summary statistics



SPSS Output:

Statistics

Current Salary

N	Valid	474
	Missing	0
Median		$28,875.00
Minimum		$15,750
Maximum		$135,000
Percentiles	25	$24,000.00
	50	$28,875.00
	75	$37,162.50

FREQUENCY POLYGON/AREA GRAPH

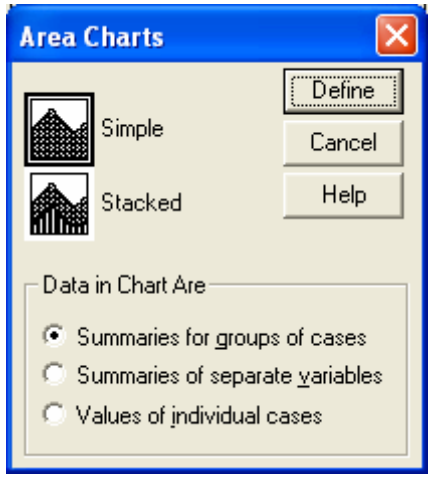
A **Frequency Polygon** graphs data by using lines that connect the midpoints of frequency classes. The height of each midpoint on the graph represents the frequency of occurrence within that class. The line connecting midpoints generally begins and ends touching the x-axis. An **Area Graph** is a Frequency Polygon for which the area under the line has been shaded.

DATA: World95.sav

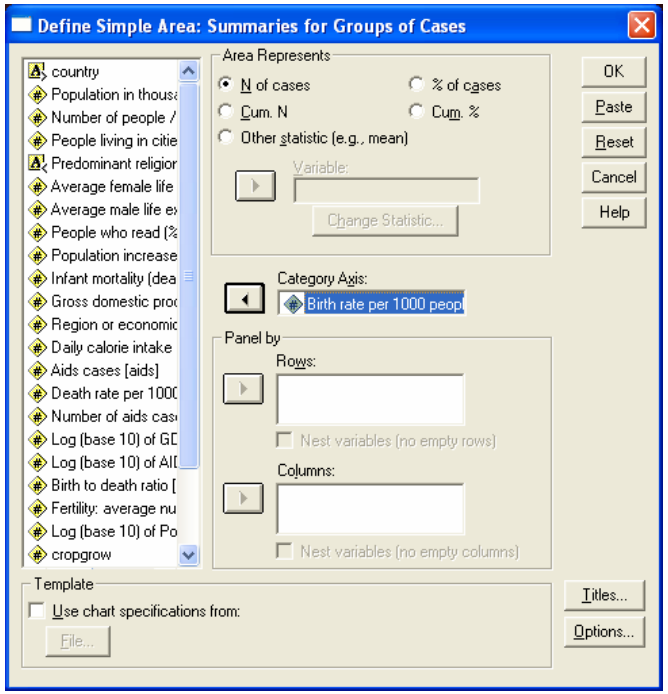
STEPS: Menu Bar => Graphs => Area

SUMMARY OF STEPS: In the initial **Area** dialog box select **Simple** and **Summaries for groups of cases**. => Select **Define**. => Move the variable into the **Category Axis** cell and click **OK**. The **Frequency Polygon** will appear in the output viewer. To convert this graph to an **Area Graph** use the Chart Editor to shade under the line.

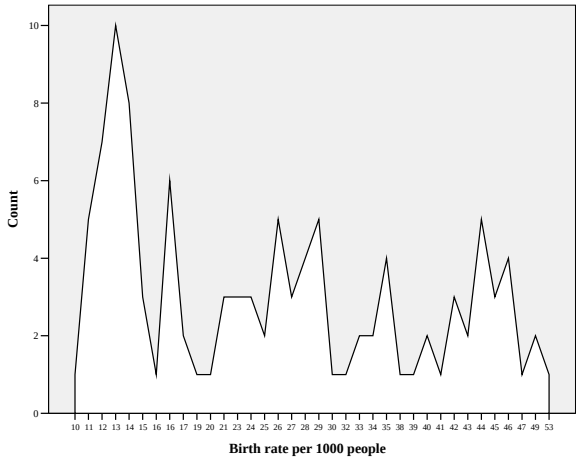
Dialog Box: Select type of graph.



Define the variable and type of data to be graphed.



SPSS Output: Frequency Polygon appears in Output Viewer. (NOTE: compare with Histogram example.)



FREQUENCY TABLES

A **Frequency Table** obtained in SPSS will contain the following for each value of one or more variables: frequency (counts), percents, valid percents[1], and cumulative percents.

DATA: Employee data.sav

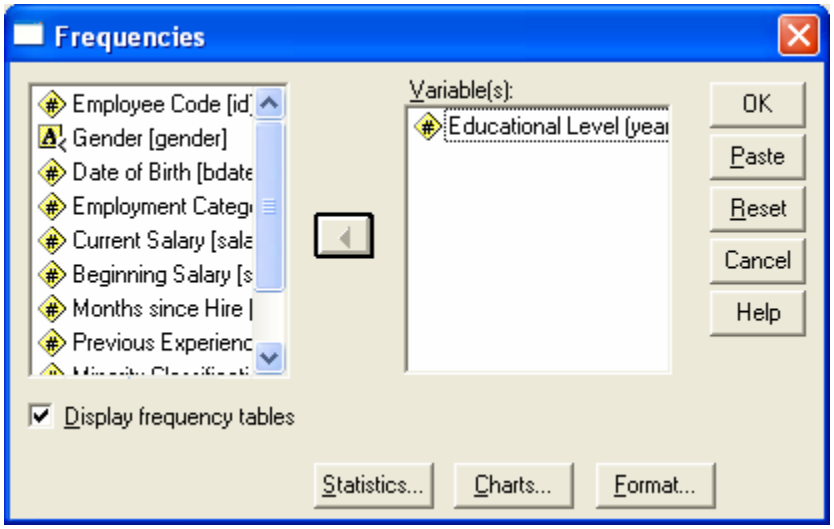
STEPS: Menu Bar => Analyze => Descriptive Statistics => Frequencies

> **OPTIONS:** If in addition to a **Frequency Table,** you wish to obtain any statistics or charts for your variable, click on the **Statistics** or **Charts** boxes. Each of these will take you to other dialog boxes where you can select the type of statistics or charts you wish to obtain.

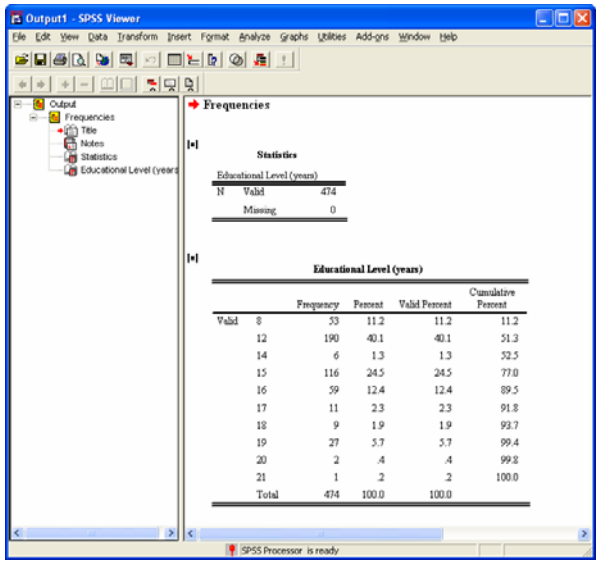
EXAMPLE: Obtain a frequency table of the variable **'EDUC'** (Education Level).

SUMMARY OF STEPS: In the **Frequencies** dialog box, select the variable for which you wish to obtain a **Frequency Table** by clicking once on the variable and then click on the right arrow to the left of the **Variables** box. This will move your variable into the **Variables** box. Note that once you do this, the arrow will point to the left. => Click **OK.** The **Frequency Table** will appear in the Output Viewer. If you want statistics or charts, select those options via the button options.

Dialog Box: A variable is selected.



SPSS Output: Frequency Table appears in output viewer.



NOTES: 1) Frequency Table format**:** A valid percent in SPSS is the percentage, excluding any missing values. **2)** Associated with every frequency table will be a Statistics Table. The statistics table will note the number of valid and missing cases for the frequency table and contain any requested statistics (see Descriptive Statistics).

HISTOGRAM

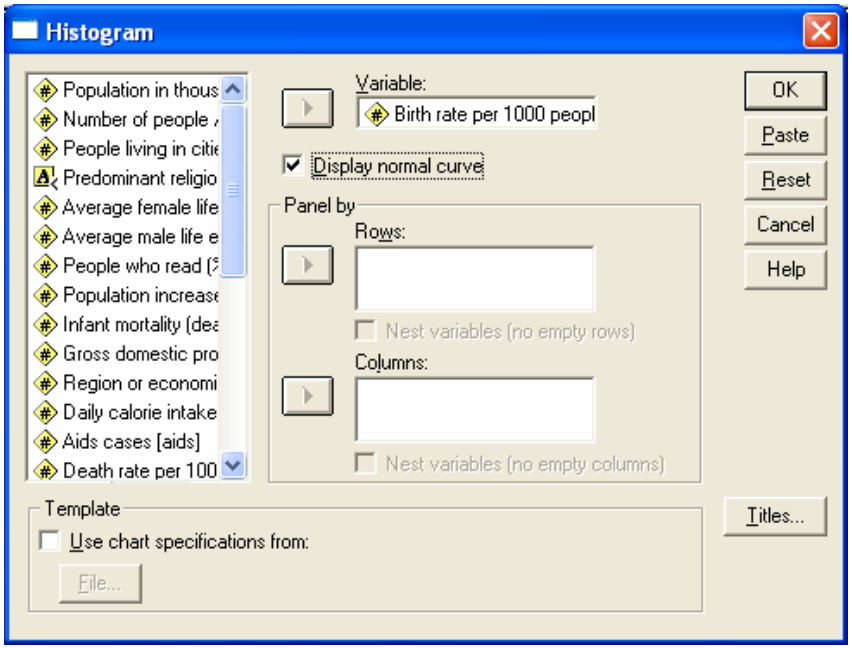
A **Histogram** is a graph of a distribution in which classes of equal width are placed on the horizontal axis, and frequencies, relative frequencies, or percentages are placed on the vertical axis. The frequencies, relative frequencies, or percentages are represented by the heights of the bars. The bars are adjacent to one another.

DATA: World95.sav

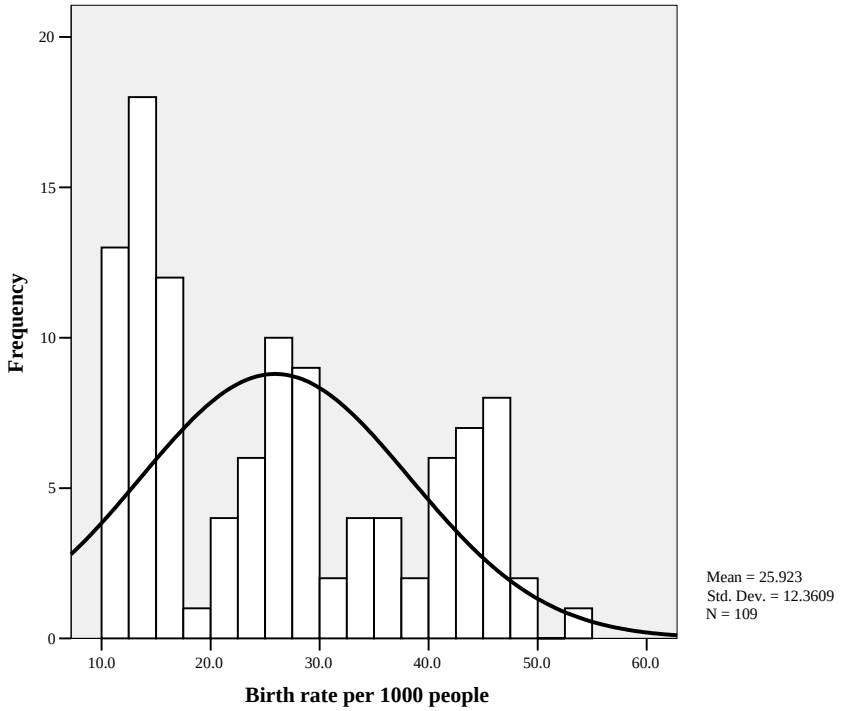
STEPS: Menu Bar => Graphs => Histogram

SUMMARY OF STEPS: In the **Histogram** dialog box select your variable by clicking once on it, and then clicking on the right arrow to the left of the **Variable** box. => Click **OK.** The **Histogram** will appear in the Output Viewer. At times, it is useful to view a normal curve superimposed on the histogram. This is accomplished by selecting the **Display Normal Curve** option.

Dialog Box: Variable is selected.



SPSS Output: Histogram appears in Output Viewer.



HYPOTHESIS TEST FOR A POPULATION MEAN

To test a hypothesis about a population mean, SPSS uses the One-Sample T-Test, and always performs a two-tailed test. Decisions are made based on p-value.

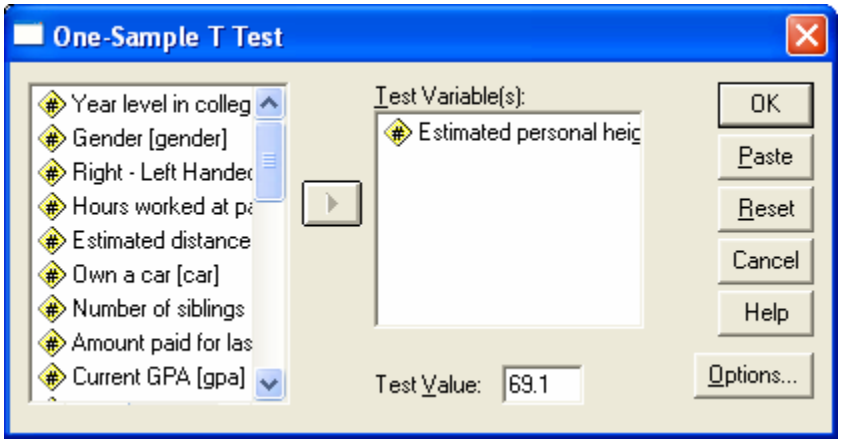
DATA: (Not an SPSS file.) According to the National Center for Health Statistics, the average height for males is 69.1 inches.

STEPS: Menu Bar => Analyze => Compare Means => One-Sample T-Test

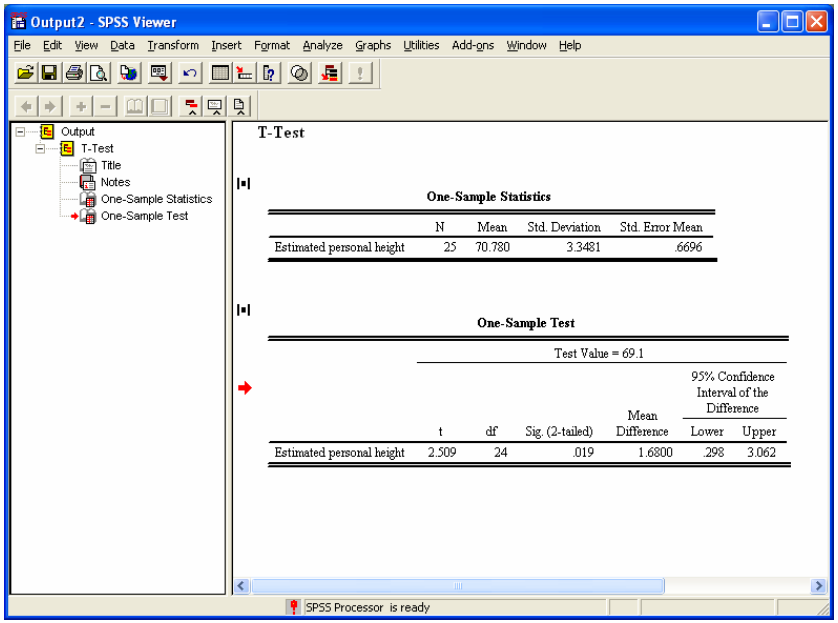
HYPOTHESIS TEST EXAMPLE 1: Using a significance level of 0.05, test the claim that the mean height of males enrolled in an introductory statistics course is different from 69.1 inches ($H_1 : \mu \neq 69.1$).

SUMMARY OF STEPS: In the **One-Sample T-Test** dialog box, select your variable by clicking once on it and then clicking the right arrow next to the **Test Variable(s)** box. => Type your test value in the **Test Value** box. => Click **OK** and the hypothesis test (along with other descriptive measures of the data) will appear in the output viewer.

Dialog Box: Variable is selected and test value is input.



SPSS Output: Hypothesis test and other descriptives appear in output viewer. Note that the p-value (.019) is less than the designated significance level (0.05), therefore we conclude that there is sufficient evidence to reject the null hypothesis.



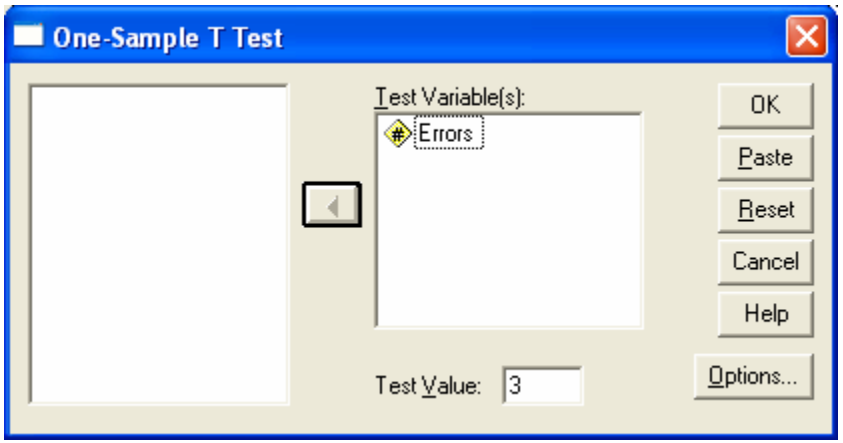
DATA: (Not an SPSS file.) A major catalog company claims that the mean number of shipping errors per 100 shipments is less than 3.

STEPS: Menu Bar => Analyze => Compare Means => One-Sample T-Test

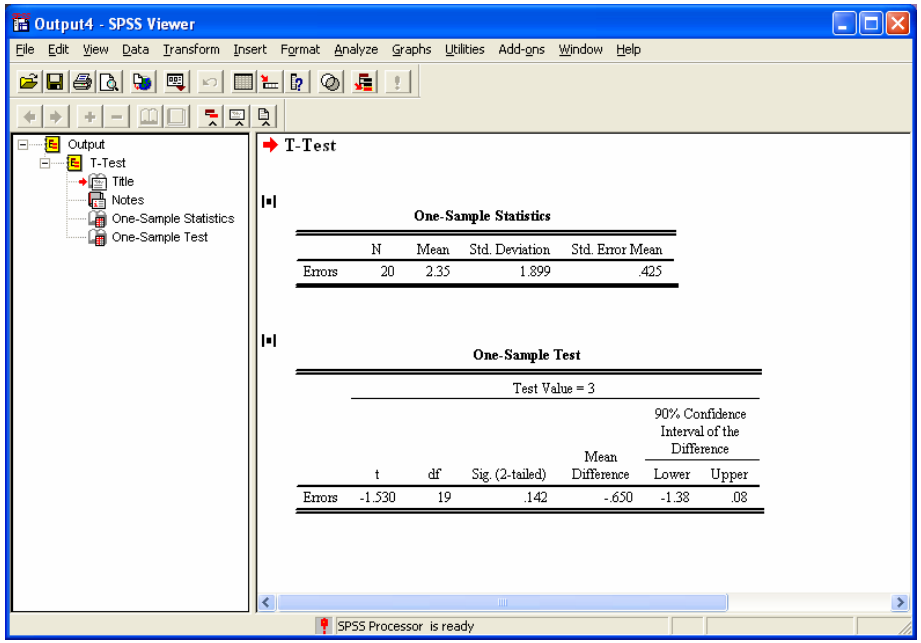
HYPOTHESIS TEST EXAMPLE 2: Test the claim that the mean number of shipping errors per 100 shipments is less than 3 ($H_1 : \mu < 3$), using a 0.10 significance level.[1]

SUMMARY OF STEPS: In the **One-Sample T-Test** dialog box, select your variable by clicking once on it and then clicking the right arrow next to the **Test Variable(s)** box. => Type your test value in the **Test Value** box. => Click **OK** and the hypothesis test (along with other descriptive measures of the data) will appear in the output viewer.

Dialog Box: Variable is selected and test value is input.



SPSS Output: Hypothesis test[1] and other descriptives appear in output viewer.



NOTE: To test a claim that a mean is greater than a given value, you would repeat the process given in Example 2.

[1]Because of the fact that SPSS always performs a two-tailed test, you must divide the p-value by 2. In this example, .142/2 = 0.071 is less than the designated significance level of 0.10 therefore we reject the null hypothesis.

HYPOTHESIS TEST FOR A POPULATION PROPORTION

Unlike a hypothesis test for a population mean, SPSS does not have a specific (straightforward) procedure to conduct a hypothesis test for a population proportion. However, it can be done using the **Compute** command and appropriate formulas.

DATA: (Not an SPSS file). An article in the Daily News reported that 65% of Americans are in favor of the death penalty. In a random sample of 1234 Americans, 70% were in favor of the death penalty.

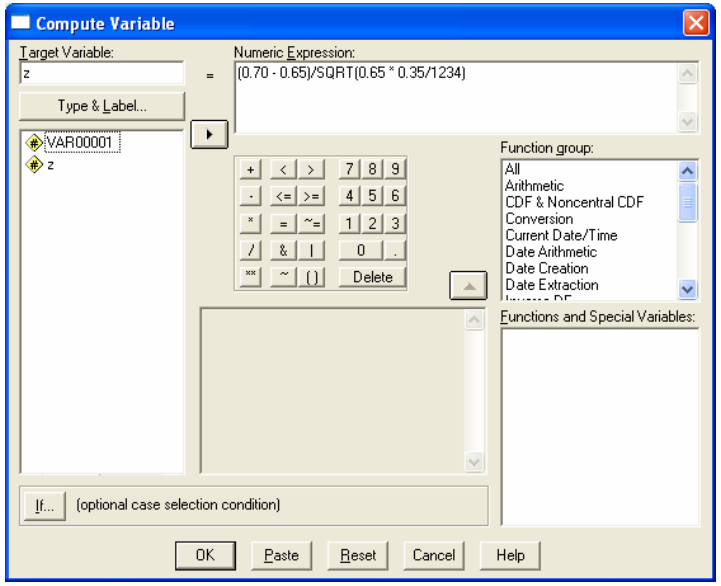
NOTE: If you are using a blank worksheet, type a number (any number) into at least one cell of the worksheet, otherwise you will get an error message when you use the required **STEPS**.

STEPS: (NOTE: This is a two-step process) Menu Bar => Transform => Compute

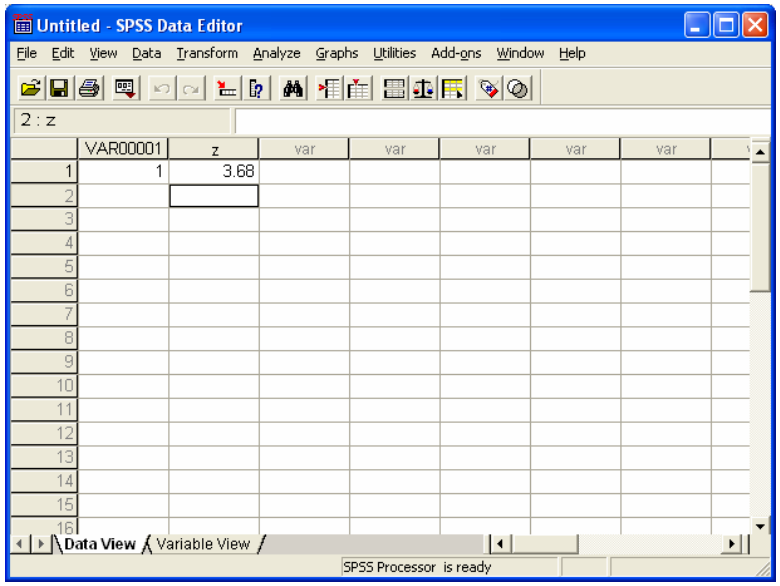
EXAMPLE: Using a significance level of 0.05, test the claim that the actual proportion of Americans who favor the death penalty is greater than 65%.

SUMMARY OF STEPS PART 1 – Computing z: In the **Compute Variables** dialog box, you will first be calculating the test statistic, **z**. In the **Target Variable** box, type in a name for the variable where the test statistic will appear (here, it is called **'z'**). Type the following expression into the **Numeric Expression** box: **(0.70 - 0.65)/SQRT(0.65 * 0.35/1234)**, where 0.70 is p-hat, 0.65 is p**,** 0.35 is 1 minus p-hat, and 1234 is the sample size. => Click **OK**. The value of the test statistic will appear in the **DATA EDITOR** column labeled **'z.'**

Dialog Box: Target variable is named, and formula for the test statistic is input.



SPSS Output: calculated test statistic

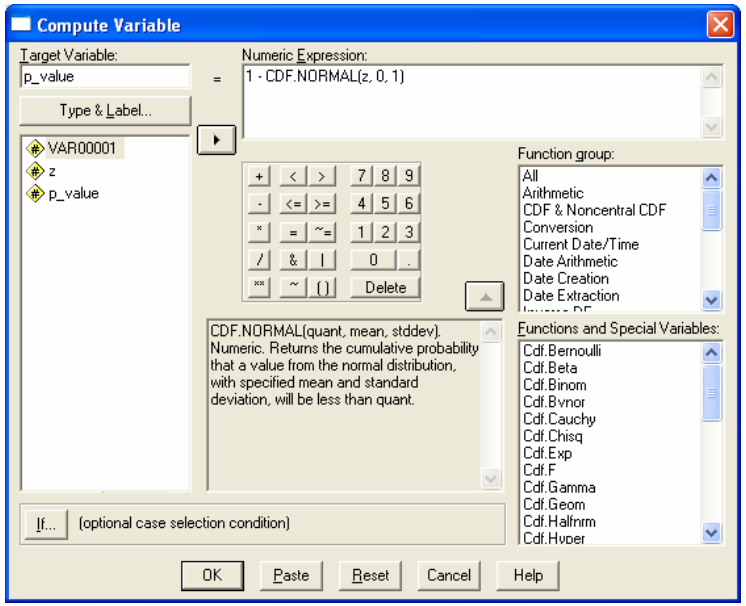


You will now need to calculate the p-value for this test statistic.

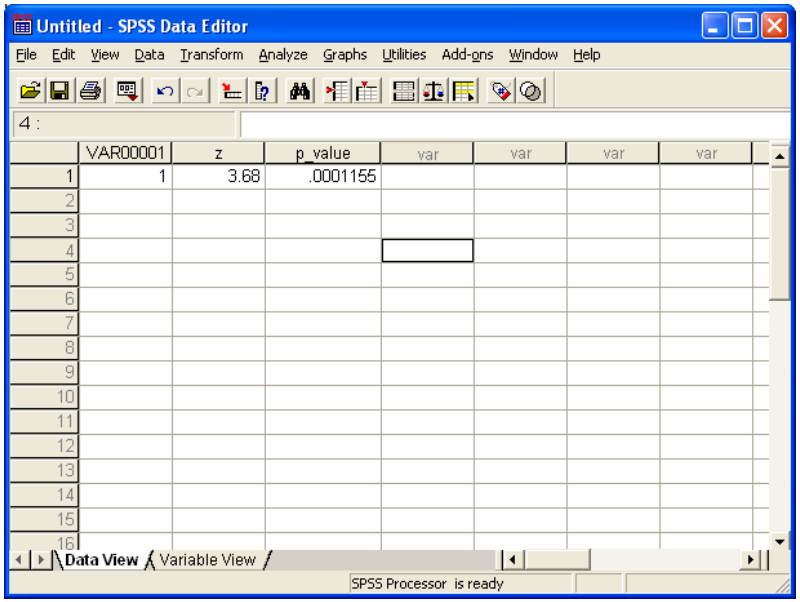
STEPS: Menu Bar => Transform => Compute

SUMMARY OF STEPS PART 2 – Computing p: In the **Compute Variables** dialog box, type in a name for your **Target Variable** (here it is named **'PVALUE'**). In the **Numeric Expressions** box, type in the following expression: **1-CDF.NORMAL(z, 0, 1).** => Click **OK** and the p-value will appear in the column labeled **'PVALUE.'**

Dialog Box: Target variable is named, and formula for the test statistic is input.



SPSS Output: P-value is approximately 0.00012, which is less than the designated significance level. Therefore, the null hypothesis is rejected.



INTERACTIVE GRAPHS

The **Interactive** option on the **Graphs Menu** provides an alternative means for preparing a number of graphs and charts. Charts created through this method open in their own editor as opposed to the **Chart Editor**.

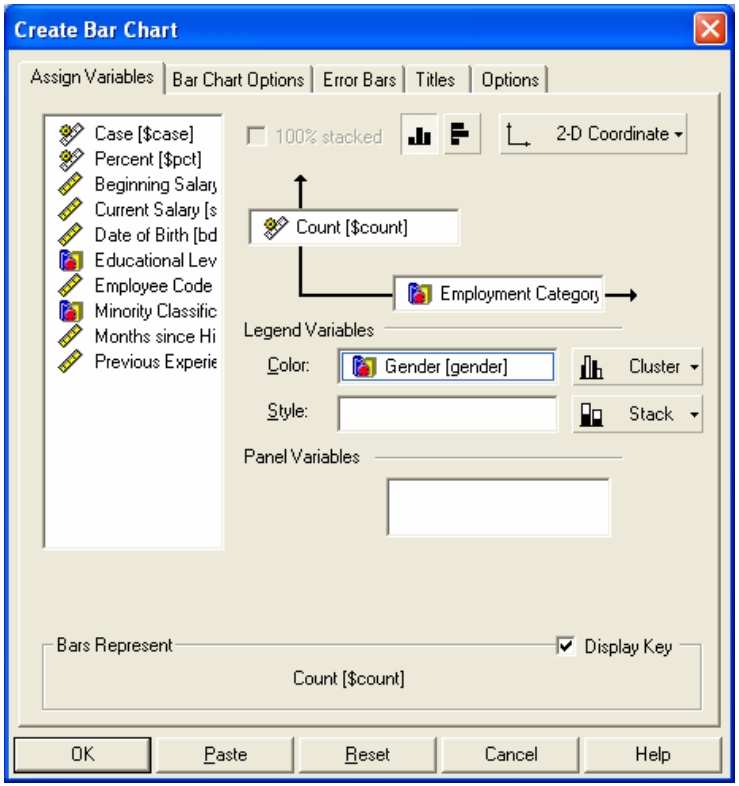
DATA: Employee data.sav

STEPS: Menu Bar => Graphs => Interactive => Select type of chart

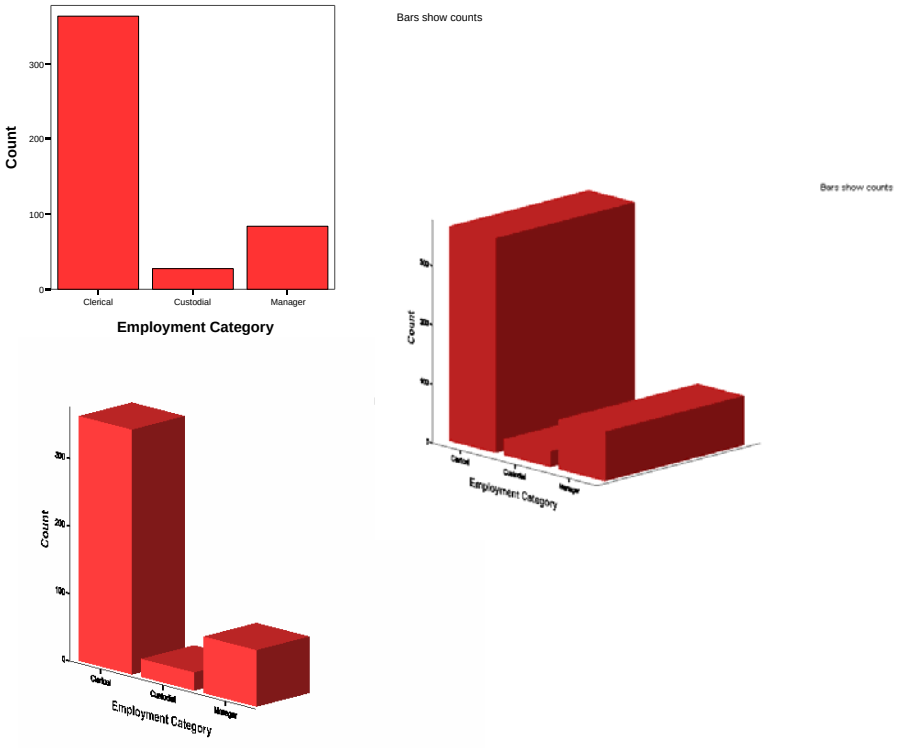
EXAMPLE: Create a **Bar Chart** of the variables **'JOBCAT'** (Job Categories).

SUMMARY OF STEPS: In the **Create Bar Chart** dialog box, select the variable by clicking and holding the mouse cursor and dragging the variable into the x-axis cell. Note that the Count is already in the y-axis cell. Count and Percent represent the type of measures that may be placed on the y-axis. Note that they have similar symbols preceding the measures. => Select **OK** to complete the chart.

Dialog Box Input: Create Bar Chart dialog box with **'JOBCAT'** selected.

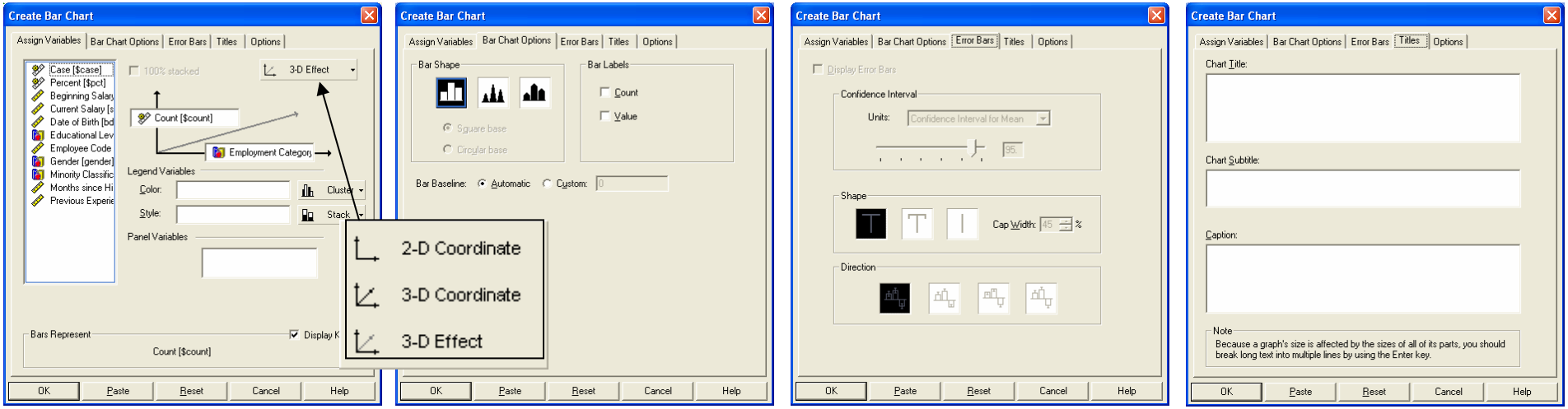


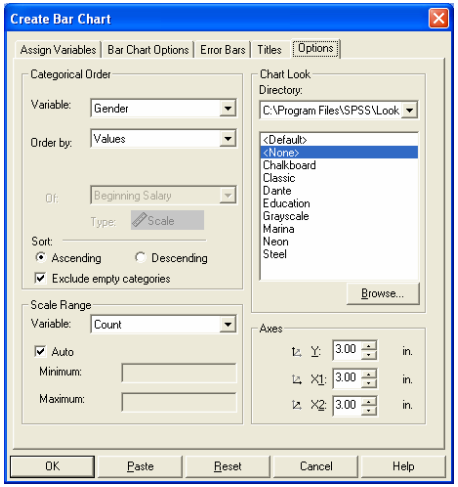
SPSS Output: Interactive Bar Chart (forms shown clockwise from top left: 2-D coordinate; 3-D coordinate; 3-D effect)



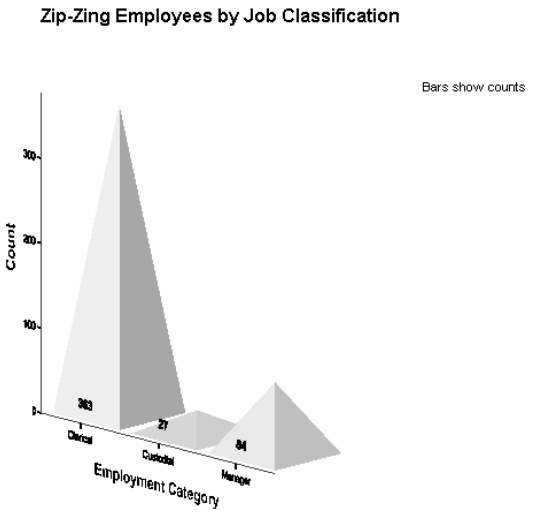
INTERACTIVE GRAPH DIALOG BOX OPTIONS

Options available on the five dialog box tabs are briefly presented here.





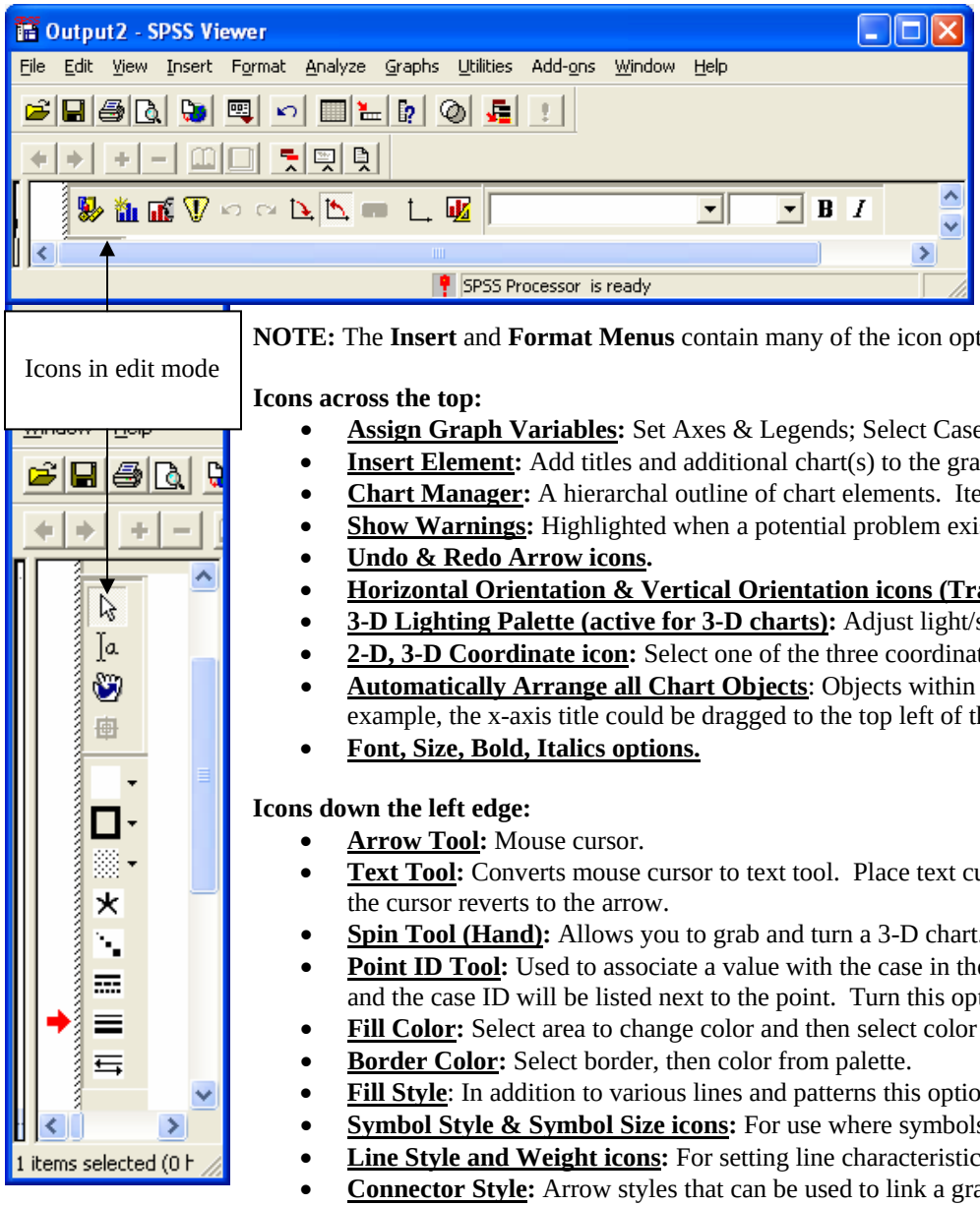
- **Assign Variables:** Identify the x-axis variable and the y-axis measure. The coordinate option (dropdown shown superimposed above) allows for the selection of 2-D or 3-D formats.
- **Bar Chart Options:** Select a **Bar Shape**; **Bar Labels** inserts the actual count on each bar.
- **Error Bars:** Available for scale variables (interval/ratio) where the variable measure is the mean. The length of the bars represents the confidence interval about the mean.
- **Titles:** Assign selected text prior to chart creation.
- **Options:** Set the order of categorical variable values; select pre-designed formats (**Chart Looks**); **Axes** ratios.

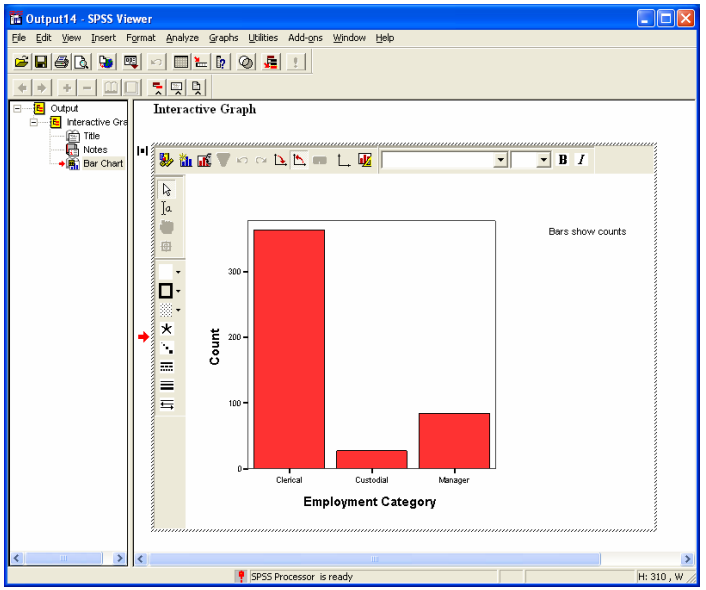


Example: 3-D Effect chart with **Bar Shape** changed to cone with a square base and category **Values** shown on bars. Note the 'Key" in the upper right ("Bars show counts"). This option is on the **Assign Variables** tab, in the lower right. Title and footnote added via the **Titles** tab.

INTERACTIVE GRAPH EDITING

To the right is a 2-D coordinate interactive chart in edit mode. Note the highlighted rectangle about the chart, similar to that used in the **Table Editor**, and the interactive icons above and to the left of the chart.





NOTE: The **Insert** and **Format Menus** contain many of the icon options.

Icons across the top:

- **Assign Graph Variables:** Set Axes & Legends; Select Cases; Pies options.
- **Insert Element:** Add titles and additional chart(s) to the graph as appropriate.
- **Chart Manager:** A hierarchal outline of chart elements. Items may be selected and deleted or hidden via this option.
- **Show Warnings:** Highlighted when a potential problem exists.
- **Undo & Redo Arrow icons.**
- **Horizontal Orientation & Vertical Orientation icons (Transpose Chart Variables)**: Toggle between them.
- **3-D Lighting Palette (active for 3-D charts):** Adjust light/shading, light intensity, and rotation of chart.
- **2-D, 3-D Coordinate icon:** Select one of the three coordinate options after the initial chart has been created.
- **Automatically Arrange all Chart Objects**: Objects within a chart may be moved by clicking on them and dragging them to a new location. For example, the x-axis title could be dragged to the top left of the chart. Clicking on this option, returns objects to their default location.
- **Font, Size, Bold, Italics options.**

Icons down the left edge:

- **Arrow Tool:** Mouse cursor.
- **Text Tool:** Converts mouse cursor to text tool. Place text cursor where you want text, click and enter text. When finished, click the mouse and the cursor reverts to the arrow.
- **Spin Tool (Hand):** Allows you to grab and turn a 3-D chart. Available on the **3-D Lighting Palette**. Turn off by selecting the **Arrow Tool**.
- **Point ID Tool:** Used to associate a value with the case in the data file. For example, in a scatterplot select the **Point ID Tool**, click on a plot point and the case ID will be listed next to the point. Turn this option off by selecting the **Arrow Tool**.
- **Fill Color:** Select area to change color and then select color from palette.
- **Border Color:** Select border, then color from palette.
- **Fill Style**: In addition to various lines and patterns this option contains options changing texture, for example, marble-like styles can be created.
- **Symbol Style & Symbol Size icons:** For use where symbols are needed/used, e.g. scatterplots.
- **Line Style and Weight icons:** For setting line characteristics.
- **Connector Style:** Arrow styles that can be used to link a graph object and a label.

LINE GRAPH

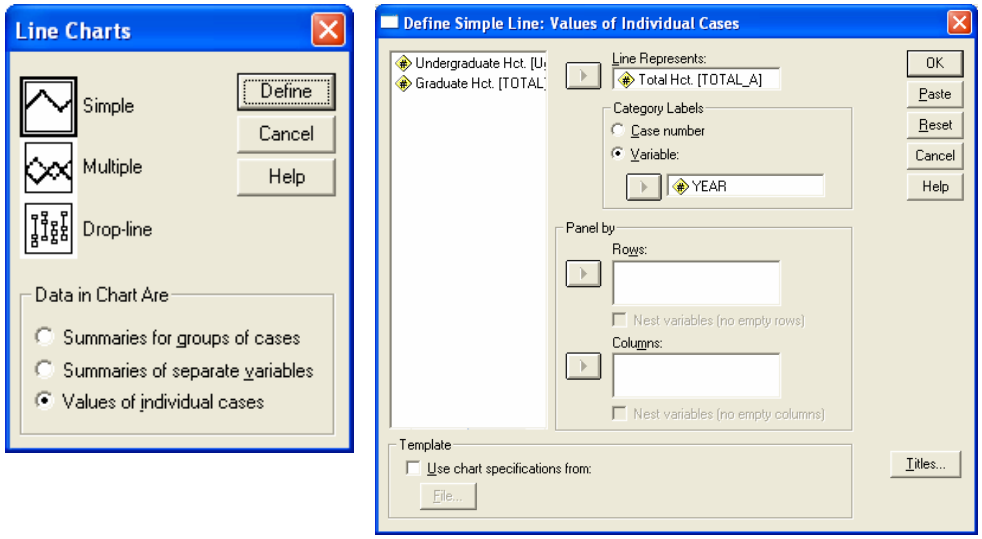
Line Charts are used to present data trends over time. They may present a single variable or be used to compare multiple variables over time.

DATA: World95.sav

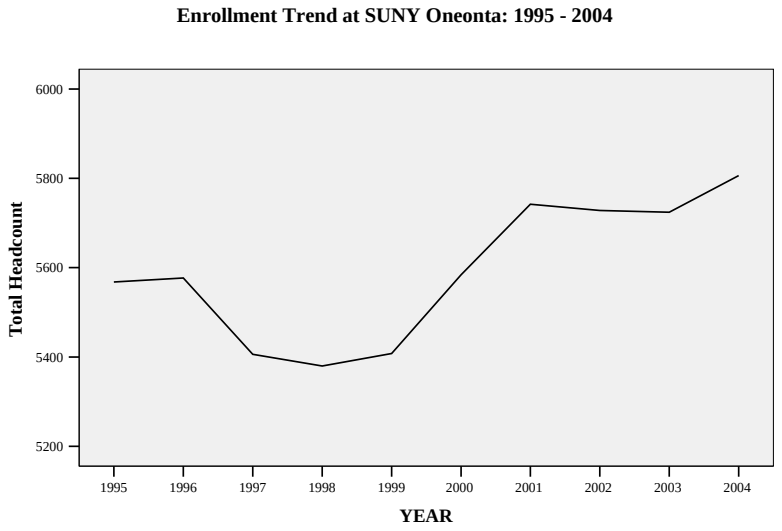
STEPS: Menu Bar => Graphs => Line => Simple; Summary for Groups of Cases => select variable and move it to the category axis box.

EXAMPLE: Create a Line Graph that presents the number of countries in each of the regions within this data set. => In the **Line Charts** dialog box select **Simple** and **Values of individual cases.** => Select **Define.** => Enter the variable to be charted into the **Line Represents** cell. => Enter the category variable, the one used on the x-axis, into the **Category Labels – Variable** cell. => Select **OK** and the chart will open in the Output Viewer.

Dialog Box: Select type chart and data type, then define variables.



SPSS Output: Line Graph as it appears in Output Viewer.



LINEAR CORRELATION

Linear Correlation is used to examine whether or not a relationship exists between two variables. A ***Correlation Coefficient*** will indicate the direction and magnitude of relationship between two variables. Correlation does not imply causation. That is, one can not infer from the correlation that one variable has an affect upon the other. Correlation Coefficients range from –1 to +1, where –1 = perfect negative correlation, 0 = no correlation, and +1 = perfect positive correlation.

DATA: World95.sav

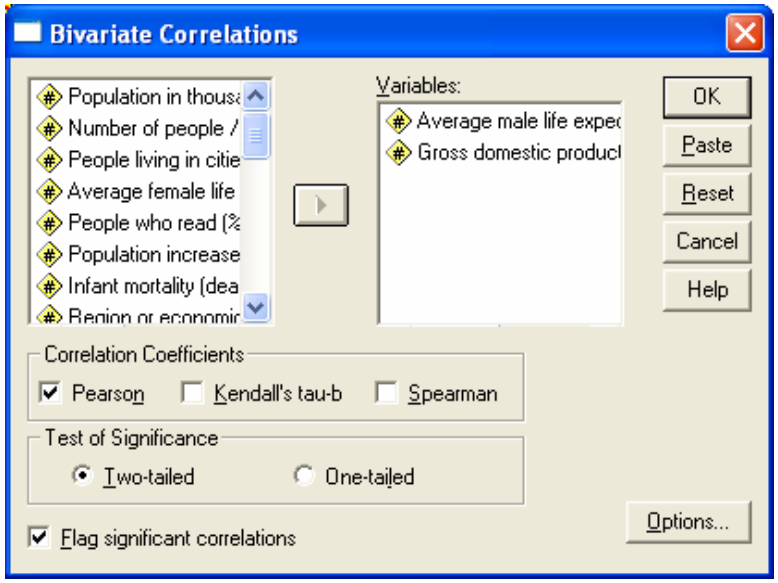
STEPS: Menu Bar => Analyze => Correlate => Bivariate => select at minimum two variables from the list of variables in the left cell and move them into the "Variables" cell. => select the type correlation to be used (here Pearson) => OK.

OPTIONS: Selecting more than two variables will result in each variable being correlated with all other variables individually.

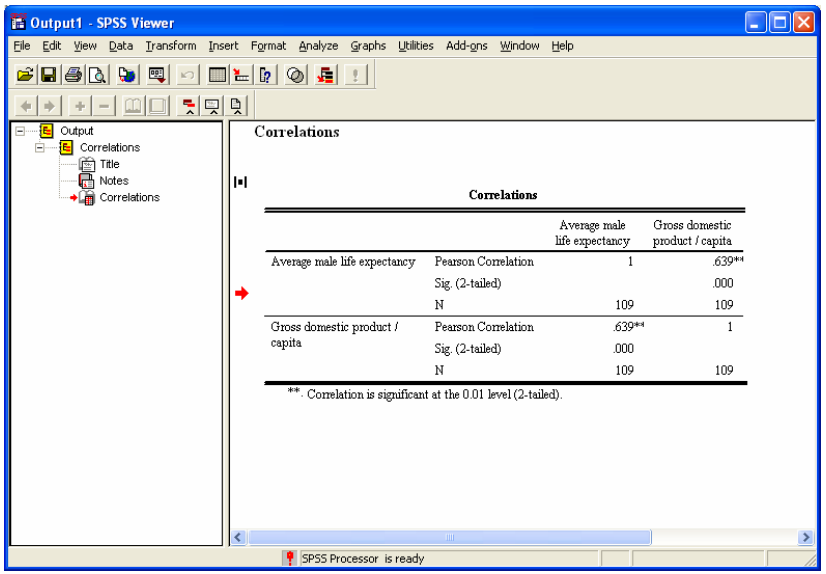
RELATED GRAPH: A ***Scatterplot*** of the two variables of interest is often produced before conducting a correlation. The scatterplot will provide an initial suggestion of whether or not there is a linear relationship between the variables.

EXAMPLE: Is there a relationship between Male Life Expectancy ('**LIFEEXPM**') and a nation's Gross Domestic Product per Capita (**'GDP_CAP'**)? For an initial review create a scatterplot (use **'GDPCAP'** as the **X variable** and **'LIFEEXPM'** as the **Y variable**; see **SCATTERPLOT**). If there is no apparent relationship via the scatterplot, obtaining a correlation may not be warranted. To obtain a correlation enter the two variables into the **Variables** box of the **Bivatiate Correlation** dialog box. => Examine the value of the correlation coefficient and its significance as provided in the output. **NOTE:** The order of entry of the variables has no affect on the correlation coefficient.

Dialog Box Input: Variables **'LIFEEXPM"** and **'GDP_CAP'** entered into the **"Variables:"** cell.



SPSS Output: Correlation and related statistics



LINEAR REGRESSION

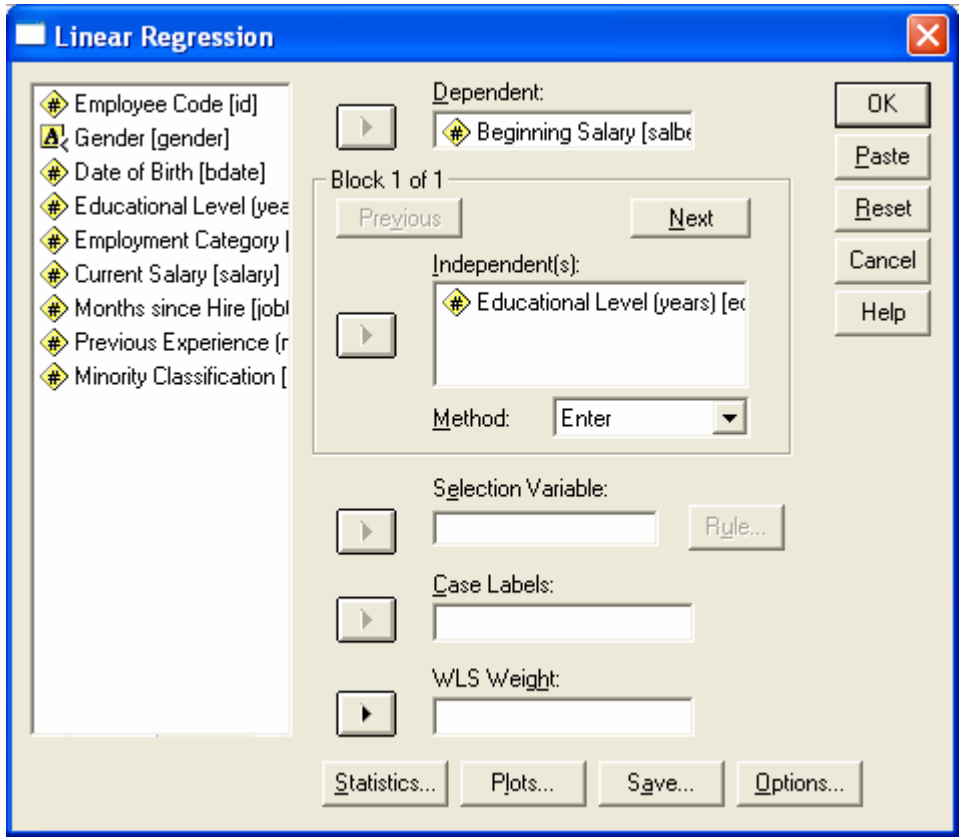
Linear Regression is used to examine the relationship between two or more variables. The method used employs the least-squares criterion to establish a regression line that best fits the data points. The Response variable represents the "Y" variable and the Predictor variable represents the "X" variable in the equation.

DATA: Employee d`ata.sav

STEPS: **Menu Bar => Regression => Linear =>** insert a variable into the **Dependent** cell and a variable into the **Independent** cell.

EXAMPLE: Determine a Regression Equation that predicts Beginning Salary (**'SALBEGIN'**) using Education Level (**'EDUC'**) as the predictor variable.

Dialog Box Input: Variable **'SALBEGIN'** entered into the **Dependent** cell and the variable **'EDUC'** entered into the **Independent** cell.



SPSS Output: Linear Regression - 4 tables (Scatterplot obtained separately)

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Educational Level (years)[a]	.	Enter

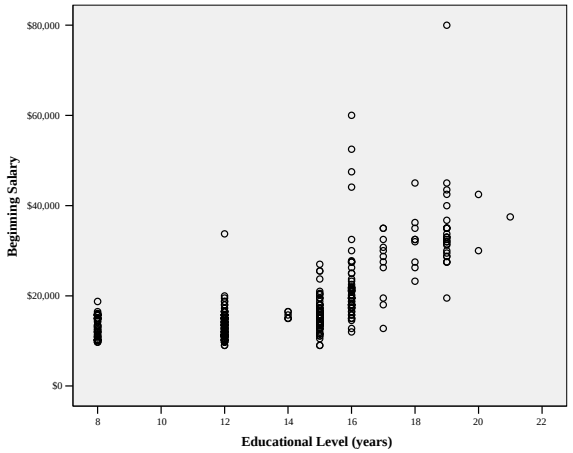
a. All requested variables entered.

b. Dependent Variable: Beginning Salary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.633[a]	.401	.400	$6,098.259

a. Predictors: (Constant), Educational Level (years)



Regression Variables (Above):
Independent Variable (Predictor Variable):Education Level (‘**EDUC’**), which is the only variable entered into the equation. Dependent Variable: Beginning Salary (‘**SALBEGIN’**).

Regression Model Summary (Above**)**:
Correlation Coefficient (r): .663
Coefficient of Determination (r^2): .401 (or 40.1%)

Scatterplot (Above):
Scatterplot showing positive relationship between Education Level and Beginning Salary. (Obtained via Graph => Scatterplot. Shown here to visually present the relationship.)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11747808912.318	1	11747808912.318	315.897	.000[a]
	Residual	17553096053.137	472	37188762.824		
	Total	29300904965.454	473			

a. Predictors: (Constant), Educational Level (years)

b. Dependent Variable: Beginning Salary

ANOVA (Above):
Analysis of Variance output.

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-6290.967	1340.920		-4.692	.000
	Educational Level (years)	1727.528	97.197	.633	17.773	.000

a. Dependent Variable: Beginning Salary

Regression Coefficients (Above):
Direction of Slope and orientation of Correlation Coefficient (+/-): here a “+” meaning a positive Slope and positive Correlation
Slope of Line: 1727.528
Predictor Variable: (EDUC)
y-intercept (Constant): -6290.967
Dependent variable: Noted in the footnote

Regression Equation: SALBEGIN = -6290.967 + 1727.528(EDUC)

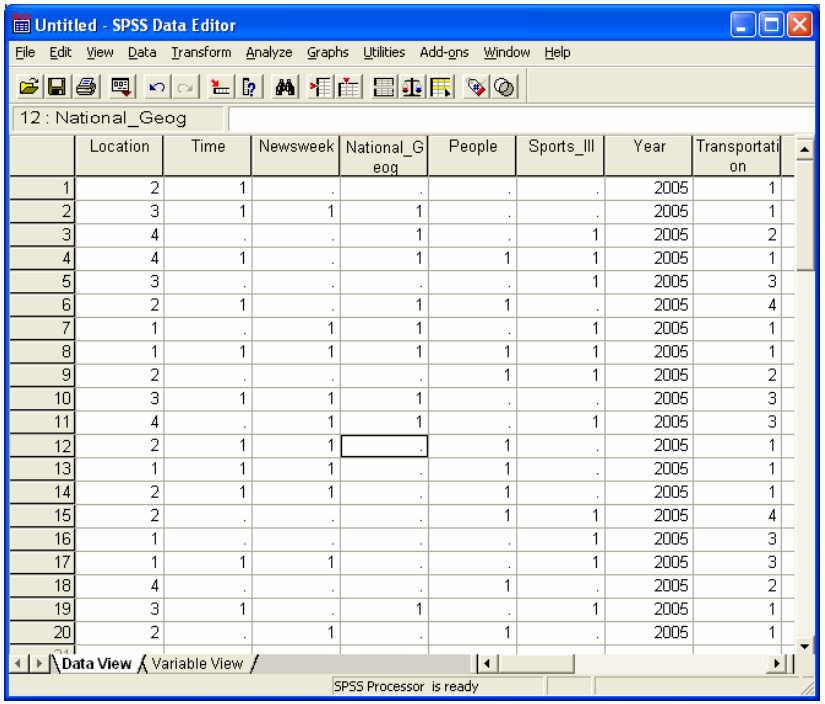
MULTIPLE RESPONSE

Multiple Response analysis is used to summarize data resulting from items for which a respondent could "check more than one response."

DATA: Reading preferences (a non-SPSS data file). Complete file, N = 20, is shown to the right.

STEPS: Menu Bar => Analyze => Multiple Response => Define Sets.

> **OPTIONS:** Once sets are defined, the options of Frequency and Crosstabs become available for analysis



	Location	Time	Newsweek	National_Geog	People	Sports_Ill	Year	Transportation
1	2	1	.	.	.	.	2005	1
2	3	1	1	1	.	.	2005	1
3	4	.	.	1	.	1	2005	2
4	4	1	.	1	1	1	2005	1
5	3	.	.	.	.	1	2005	3
6	2	1	.	1	1	.	2005	4
7	1	.	1	1	.	1	2005	1
8	1	1	1	1	1	1	2005	1
9	2	.	.	.	1	1	2005	2
10	3	1	1	1	.	.	2005	3
11	4	.	1	1	.	1	2005	3
12	2	1	1	.	1	.	2005	1
13	1	1	1	.	1	.	2005	1
14	2	1	1	.	1	.	2005	1
15	2	.	.	.	1	1	2005	4
16	1	.	.	.	.	1	2005	3
17	1	1	1	.	.	1	2005	3
18	4	.	.	.	1	.	2005	2
19	3	1	.	1	.	1	2005	1
20	2	.	1	.	1	.	2005	1

The Multiple Response Question:

While the following question appears to be one, it is actually five related questions.

> Which of the following magazines do you read on a regular basis?
> __ Time
> __ Newsweek
> __ National Geographic
> __ People
> __ Sports Illustrated

The Multiple Response Variable:

Addressing the analysis of this type of survey question requires some planning. Most important, you need to recognize the question as one that an individual could select more than one of the response choices (values) *before* entering the data into a data set. Realizing this, you must treat each response as a separate variable. The values assigned to each response may be the same, such as 1 = "Checked" and 0 = "Not Checked," or a different value may be used for the response to each item within a question, e.g. 1 = *Time*, 2 = *Newsweek*, etc.

Following on this process, multiple frequency tables would be created to present the data from what may initially appear as a single question. Here, a separate table could be created for each magazine. The shortcoming of these frequency tables is that they portray only pieces of a bigger picture. Using the **Multiple Response** statistical option brings all of the pieces into one table for analysis.

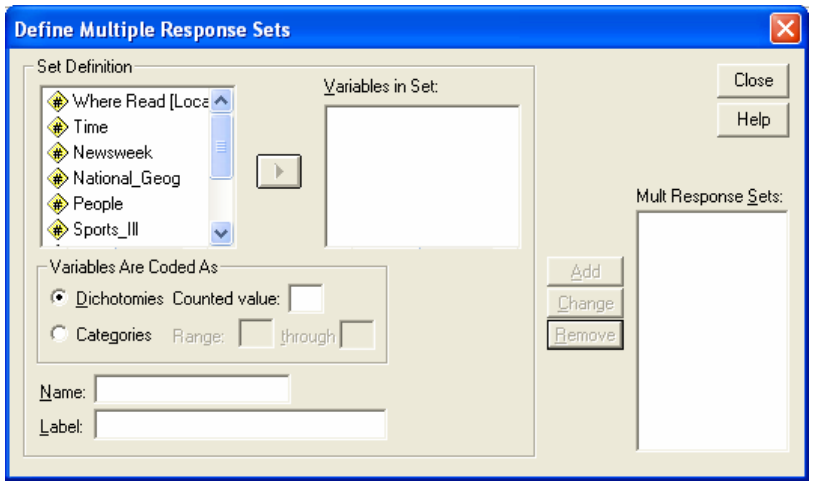
DEFINING A MULTIPLE RESPONSE VARIABLE

Defining Multiple Response Sets: Initially, one must identify the variables to be included in the Multiple Response Set (select and move to the right) via the **Define Sets** option. If the data for each variable have been coded as *Dichotomies*, that is each using the same values(e.g. 1 or 0), the **Dicotomies** option is selected and the value to be reported is identified. If each variable has been coded with different values, such as 1 = *Time* and 2 = *Newsweek*, the **Categories** option is selected.

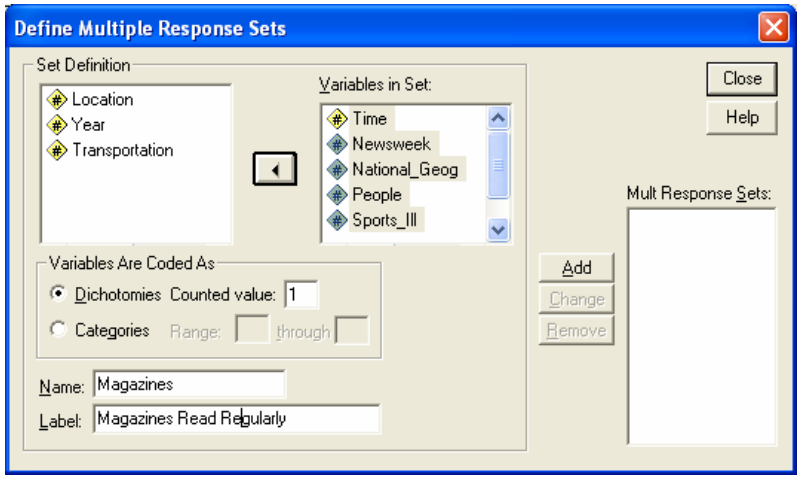
A name must be identified for the **Multiple Response Variable** being created. In the Following example it is **'Magazines.'** The **Variable Label** box is optional. Once the left side of the **Multiple Response Dialog Box** has been completed the new variable name may be "Added" to the **Multiple Response Sets Box**. The new variable will begin with a dollar sign, here '**$Magazines**.' Once all steps have been completed, close the Dialog Box. The multiple response variable has been defined and now the frequency and crosstabs options will be available for analysis of the defined variable.

Creation of a **Multiple Response** variable does not place a new variable into the data set, as occurs when using Recode or Compute. This is a session specific variable that will be lost upon leaving SPSS. Should one wish to maintain this variable definition, it may be pasted into the Syntax Editor and saved for future retrieval. (To save the definition, either a multiple response frequency dialog box or crosstabs dialog box must be pasted.)

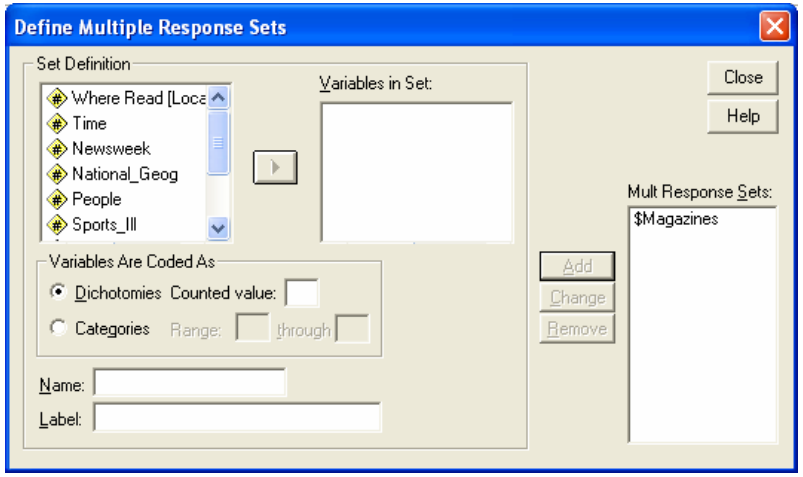
1) Initial Multiple Response dialog box:



2) Variables in the set have been identified, the coding selected (dichotomy) and a variable name assigned.



3) Variable has been moved to the Mult Response Sets box.



MULTIPLE RESPONSE FREQUENCY TABLE

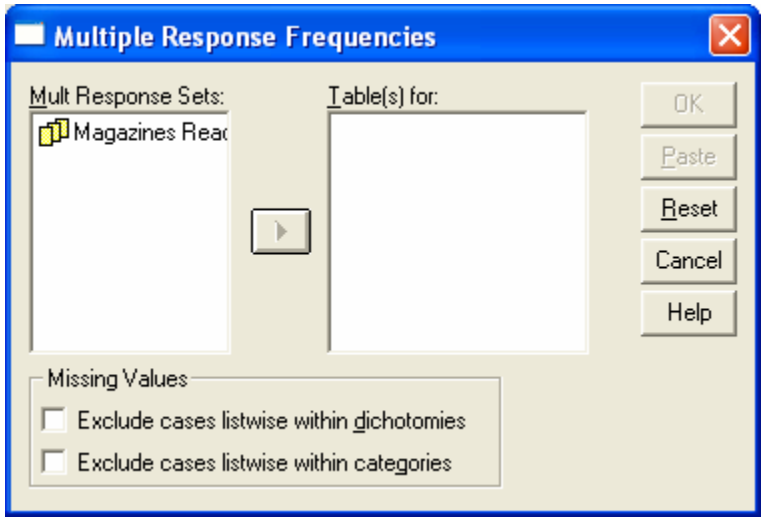
These variables do not become a part of the list of available variables for Frequency tables or Crosstabs. Instead, they must be accessed through the Multiple Response options for Frequencies and Crosstabs. Selecting the **Multiple Response Frequencies** option would yield a table such as the one that follows. Again, this single table combines the data that would be contained in five individual tables.

STEPS: Menu Bar => Analyze => Multiple Response => Frequencies

EXAMPLE: The sample variable asked respondents which magazines they read on a regular basis. Any or all five responses could be selected, yielding five independent frequency tables. By creating a multiple response variable that includes all five responses, a single frequency table summarizing all response options may be obtained.

INTERPRETING THE OUTPUT: The **Case Processing** table indicates that there were 20 respondents. The frequency table indicates that there were 51 responses to this series of questions. Eleven individuals indicated that they read *Time* magazine. These 11 responses represent 21.6% of the 51 responses. Additionally, these 11 individuals represent 55% of all respondents (Percent of Cases, where N = 20).

Multiple Response Define Dialog Box: Move variables to right



SPSS Output: Multiple Response frequency table

Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
$Magazines[a]	20	100.0%	0	.0%	20	100.0%

a. Dichotomy group tabulated at value 1.

$Magazines Frequencies

		Responses		Percent of Cases
		N	Percent	
Magazines Read Regularly[a]	Time	11	21.6%	55.0%
	Newsweek	10	19.6%	50.0%
	National_Geog	9	17.6%	45.0%
	People	10	19.6%	50.0%
	Sports_Ill	11	21.6%	55.0%
Total		51	100.0%	255.0%

a. Dichotomy group tabulated at value 1.

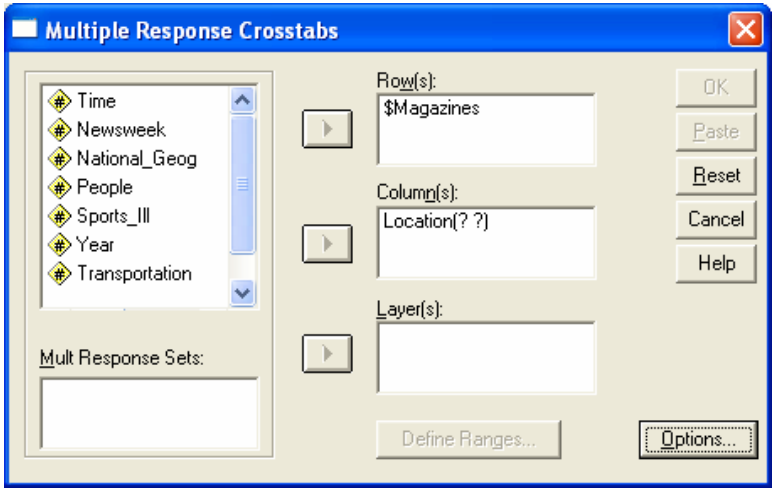
MULTIPLE RESPONSE CROSSTABS TABLE

Using the **Multiple Response Crosstabs** option requires the selection of a second variable. There will be a "(??)" following a variable selected from the top left listing of "traditional" variables if one is selected and moved into either the Rows or Columns boxes. Select the **Define Ranges** button to identify the values of this variable to include in the analysis.

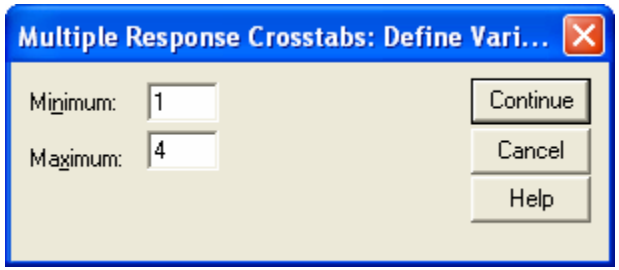
STEPS: Menu Bar => Analyze => Multiple Response => Crosstabs

EXAMPLE: Obtain a Crosstabs of the variable **'$MAGAZINES'** (row variable) with the variable **'LOCATION'** (column variable).

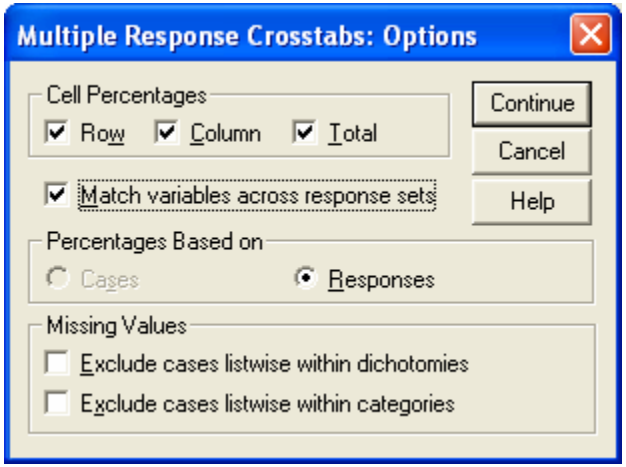
Define the Crosstabs: select the multiple response variable and another variable (can be another multiple response variable); select; select



Define Range to identify the values of the non-multiple response variable, if used.



Options to identify desired percentages.



INTERPRETING MULTIPLE RESPONSE CROSSTABS OUTPUT:

The **Case Processing** table indicates that there were 20 respondents.

While the general appearance of a multiple response crosstabs table looks similar to a crosstabs table generated from traditionally defined variables, it is actually quite different.

Counts: The counts represent the number of responses to each value. Here 3 of the 20 individuals read *Time* at the Library. Those three individuals are also a portion (3) of the 11 individuals who read *Time* (right margin total) and a portion (3) of the individuals who indicated that they read magazines at a library (n = 5).

Row %: The row percent shows that the three individuals who read *Time* at the Library represent 27.3% of all individuals who read *Time* (i.e. sum across the row). So, 27.3% of those who read *Time*, do so at a library (3 of 11).

Column %: The column percent shows that the three individuals who read *Time* at the Library represent 60.0% of all individuals who read magazines at a library (n = 5). So, 60.0% of those who read at a library, read *Time* (3 of 5).

Total %: The total percent indicates that the 3 individuals who read *Time* at the library represent 15% of all respondents (3 of 20).

Margin Values: The total count across the bottom sums to 20 and indicates the primary reading location for the respondents. So, here five respondents (25% of n = 20) indicated they read at a library. The total count and percent in the right margin column indicates the number of the 20 respondents who read a particular magazine (e.g. 11 read *Time*) and what percent of all respondents that count represents (11 of 20 = 55%).

Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
$Magazines*Location	20	100.0%	0	.0%	20	100.0%

$Magazines*Location Crosstabulation

			Where Read				
			Library	Home	Work	On-Line	Total
Magazines Read Regularly	Time	Count	3	4	3	1	11
		% within $Magazines	27.3%	36.4%	27.3%	9.1%	
		% within Location	60.0%	57.1%	75.0%	25.0%	
		% of Total	15.0%	20.0%	15.0%	5.0%	55.0%
	Newsweek	Count	4	3	2	1	10
		% within $Magazines	40.0%	30.0%	20.0%	10.0%	
		% within Location	80.0%	42.9%	50.0%	25.0%	
		% of Total	20.0%	15.0%	10.0%	5.0%	50.0%
	National_Geog	Count	2	1	3	3	9
		% within $Magazines	22.2%	11.1%	33.3%	33.3%	
		% within Location	40.0%	14.3%	75.0%	75.0%	
		% of Total	10.0%	5.0%	15.0%	15.0%	45.0%
	People	Count	2	6	0	2	10
		% within $Magazines	20.0%	60.0%	.0%	20.0%	
		% within Location	40.0%	85.7%	.0%	50.0%	
		% of Total	10.0%	30.0%	.0%	10.0%	50.0%
	Sports_Ill	Count	4	2	2	3	11
		% within $Magazines	36.4%	18.2%	18.2%	27.3%	
		% within Location	80.0%	28.6%	50.0%	75.0%	
		% of Total	20.0%	10.0%	10.0%	15.0%	55.0%
Total		Count	5	7	4	4	20
		% of Total	25.0%	35.0%	20.0%	20.0%	100.0%

Percentages and totals are based on respondents.

a. Dichotomy group tabulated at value 1.

NORMAL PROBABILITIES

To obtain probabilities for a normally distributed variable, you will need to use the **Compute** function (see **Compute**), as well as some basic arithmetic.

DATA: Employee data.sav (assume this to be a population with n = 474, μ = \$34.419.60 and σ = \$17,075.66)

STEPS: Menu Bar => Transform => Compute

EXAMPLE 1: Determine the probability that a randomly selected salary is less than or equal to \$57,000. That is, $\mathbf{P(x \le \$57{,}000)}$.

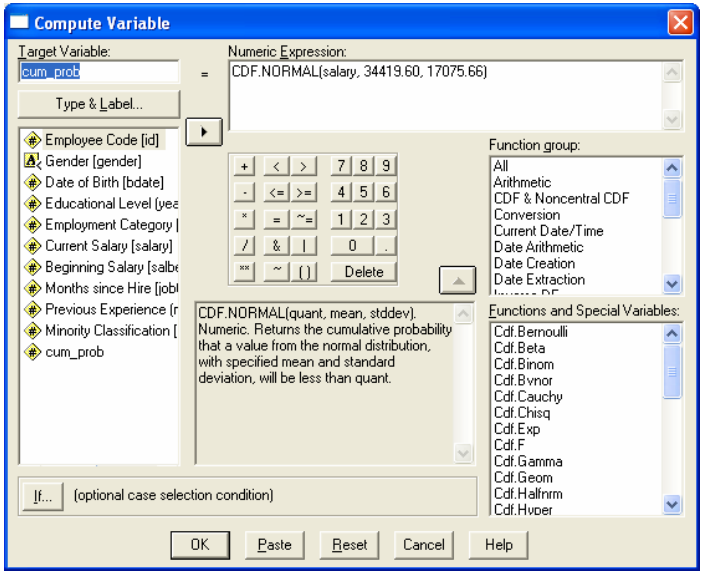
SUMMARY OF STEPS: The variable you are computing is the cumulative probability. It is the probability that a value is less than or equal to a stated value. => Select the **Target Variable** box, and type in a name for the column that will contain the cumulative probabilities. In this example, the column is named '**CUMPROB.**' => Select the **Numeric Expression** box. It is here that you will "build" your formula to calculate the cumulative probabilities. => In the **Function Group** box select **"CDF & Noncentral CD**F." => In the **Functions and Special Variables** box select the function **"CDF.Normal"** and then select the up arrow to move this function to the Numeric Expression box. The expression should read **CDF.Normal(?,?,?)**. Replace the three "?" with the variable '**SALARY**', the population mean, and the population standard deviation. The revised expression should read **CDF.Normal(Salary, 34415.60, 17075.66)**. When you have completed this click **OK.** => A column containing the cumulative probabilities for the variable '**SALARY**' will appear at the end of the worksheet (immediately following the last variable). From this output, you find that $\mathbf{P(x \le \$57{,}000) = .91}$

EXAMPLE 2: Determine the probability that a randomly selected salary is greater than \$57,000. That is, $\mathbf{P(x > \$57{,}000)}$. In order to find this probability, we use the fact that $\mathbf{P(x \le \$57{,}000) = .91}$. As such,

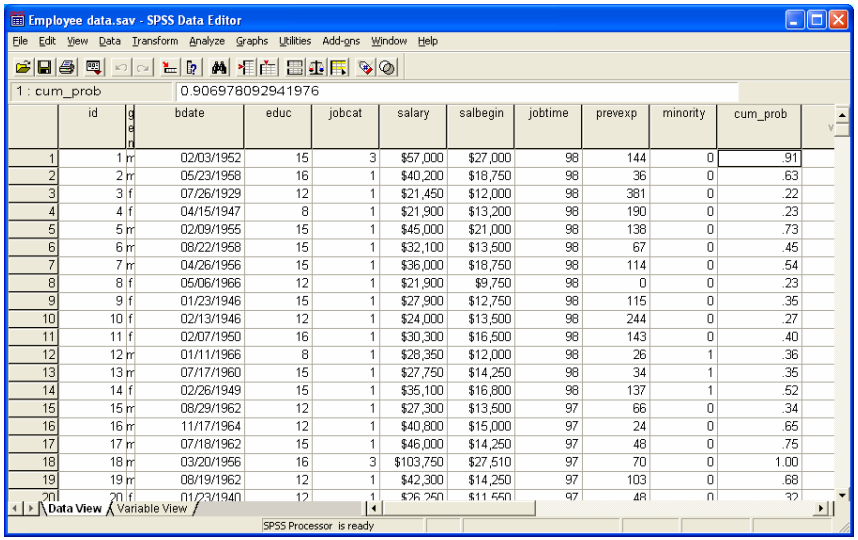
$\mathbf{P(x > \$57{,}000) = 1 - P(x \le \$57{,}000) = 1 - .91 = .09}$.

EXAMPLE 3: Determine the probability that a randomly selected salary is between \$32,100 and \$57,000. That is, $\mathbf{P(\$32{,}100 \le x \le \$57{,}000)}$. In order to find this probability, we again use the fact that $\mathbf{P(x \le \$57{,}000) = .91}$. In addition, we use the fact that $\mathbf{P(x \le \$32{,}100) = .45}$. Then, $\mathbf{P}(\$32{,}100 \le x \le \$57{,}000) = .91 - .45 = .46$.

Dialog Box: Target variable is named, and formula is "built".



SPSS Output: A column containing the cumulative probability is created.



NORMAL PROBABILITY PLOT

Detecting normality from a histogram is sometimes difficult, especially when a data set is small. A **Normal Probability Plot** compares the values of a data set with values we would expect if the data were normally distributed (n scores). The (x,y) pairs (actual data values and n scores) are then plotted. If the plot exhibits a linear pattern, then the data is probably normally distributed. If the plot exhibits significant curvature or more than one outlier is present, then the data is probably not normally distributed.

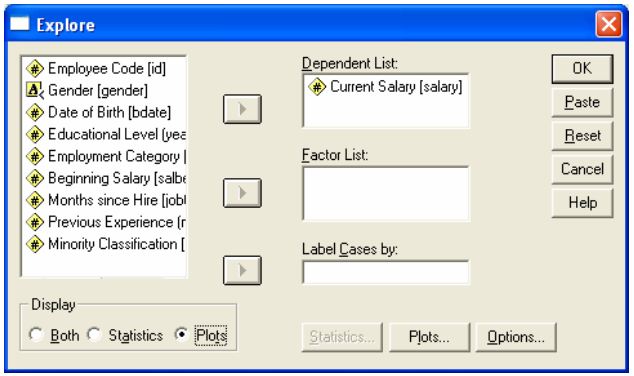
DATA: Employee data.sav

STEPS: Menu Bar => Analyze => Descriptive Statistics => Explore

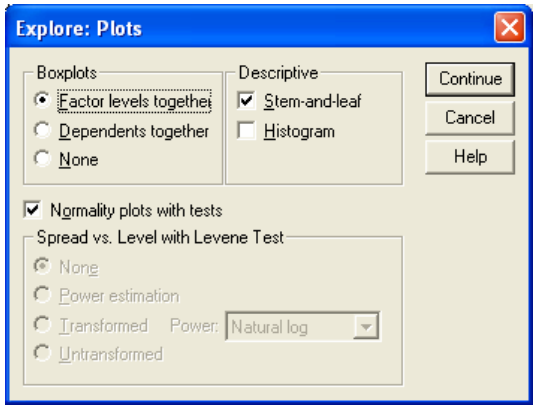
EXAMPLE: Determine if the variable **'SALARY'** is normally distributed.

SUMMARY OF STEPS: In the **Explore** dialog box, select your variable by clicking once on it and then clicking the right arrow next to the **Dependent List** box. => In the **Display** section, select **Plots.** => Click the rectangular box labeled **Plots.** => In the **Explore Plots** dialog box, select **Normality plots with tests** by clicking on the button next to it. => Click **Continue** and you are now back in the **Explore** dialog box . => Click **OK** and the **Normal Probability Plot** will be displayed in the output viewer.

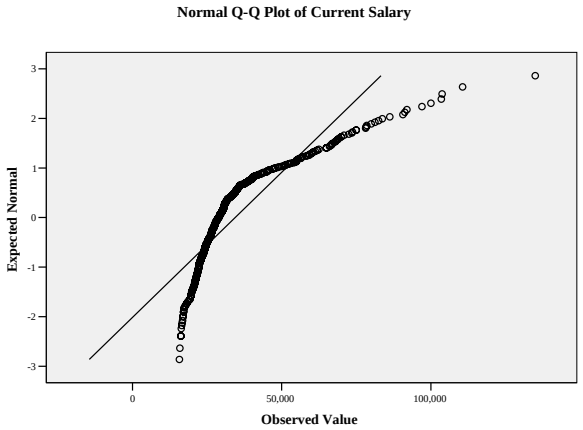
Explore Dialog Box: Variable and Plots display selected.



Plots Dialog Box: Normality Plots is selected.



SPSS Output: Normal Probability Plot



OGIVE

An **Ogive** (oh-jive) is a special type of line graph used to display the cumulative frequency for a given variable.

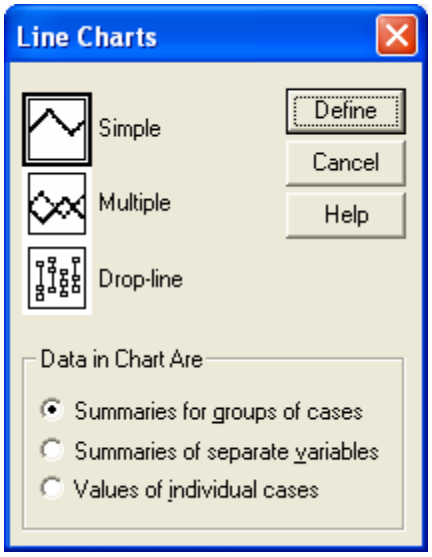
DATA: Employee data.sav

STEPS: Menu Bar => Graphs => Line

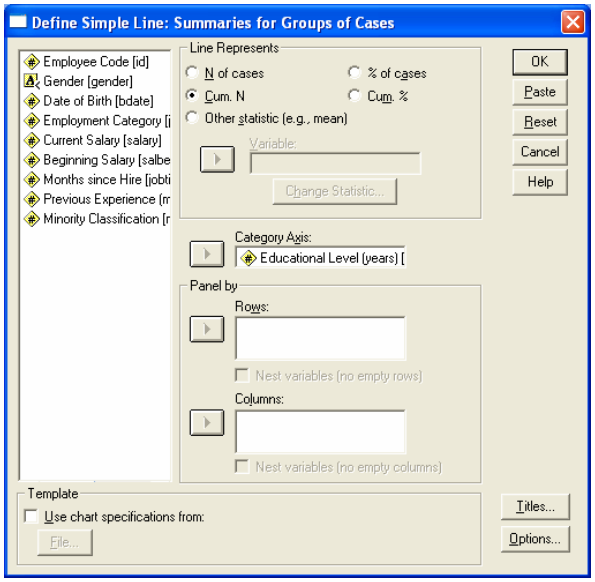
EXAMPLE: Obtain an **Ogive** for the variable **"EDUC."**

SUMMARY OF STEPS: In the **Line Charts** dialog box, select **Simple** and **Summaries for Groups of Cases.** => Click **Define.** => In the **Define Simple Line** dialog box, select **Cum N´** and select your variable by clicking once on it and then clicking on the right arrow next to the **Category Axis** box. => Click **OK** and the **Ogive** will appear in the Output Viewer.

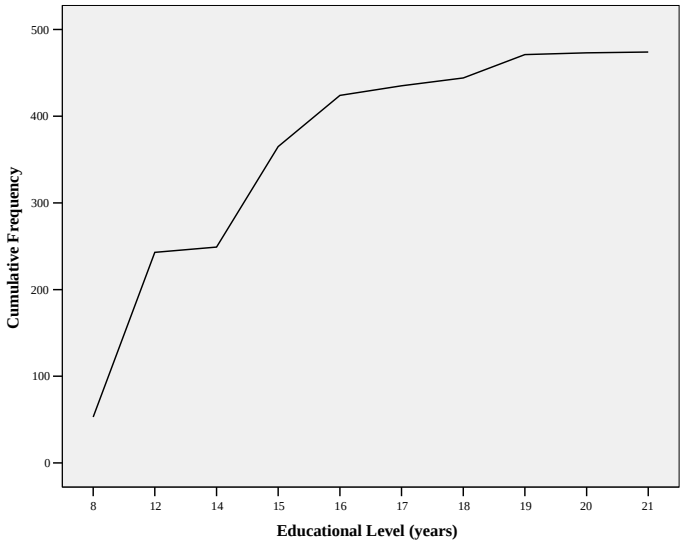
Dialog Box: Line Graph type is selected.



Dialog Box: Variable is selected.



SPSS Output: Ogive for the variable **'EDUC.'**



PARETO CHART

A **Pareto Chart** is a special type of bar chart. Frequencies displayed in a Pareto Chart are ordered from highest to lowest. A cumulative relative frequency curve is also included. The purpose of a Pareto Chart is to call attention to the most frequently occurring values of a nominal variable.

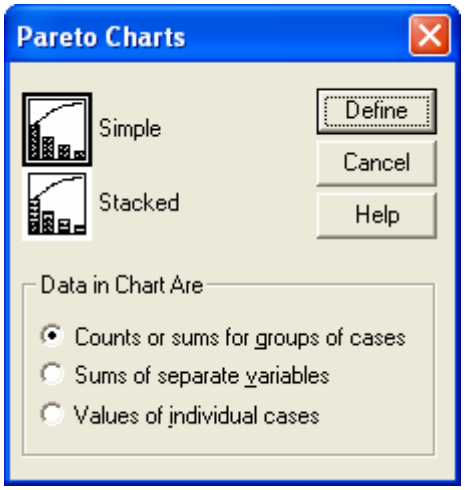
DATA: Employee data.sav

STEPS: Menu Bar => Graphs => Pareto

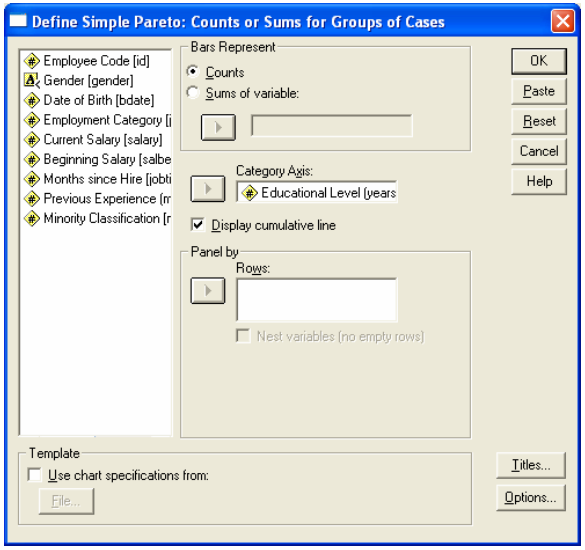
EXAMPLE: Obtain a **Pareto Chart** for the variable **"EDUC."**

SUMMARY OF STEPS: In the **Pareto Charts** dialog box, select **Simple** and **Counts or Sums for Groups of Cases.** Click **Define.** In the **Define Simple Pareto** dialog box, select your variable by clicking once on it and then clicking on the right arrow next to the **Category Axis** box. If you want a cumulative line (**Ogive**) displayed, select the **Display cumulative line** option. Click **OK** and the **Pareto Chart** will appear in the output viewer.

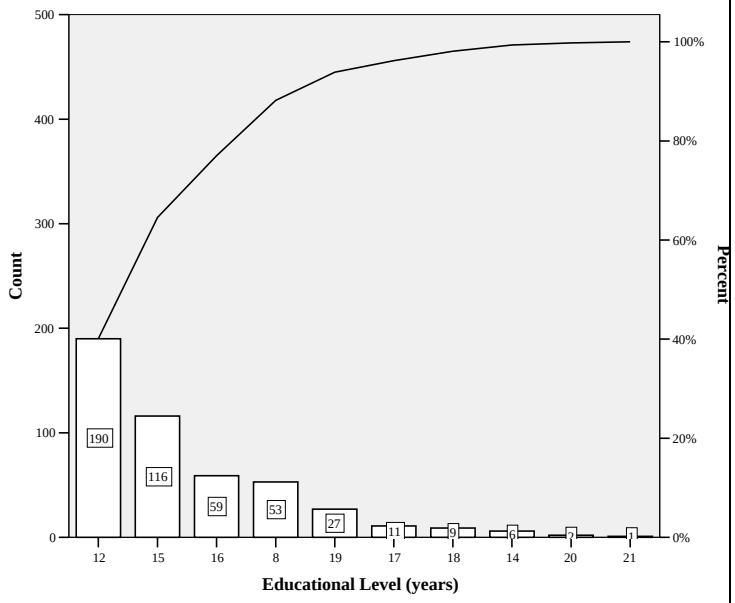
Dialog Box: Pareto Chart type is selected.



Dialog Box: Variable is selected.



SPSS Output: Pareto Chart of **"EDUC"**



PIE CHART

A **Pie Chart** is used for presenting categorical data. The chart consists of a circle subdivided into sections. The size of each section is proportional to the quantity it represents. Although there are many ways to construct a **Pie Chart** in SPSS, only the most basic type will be shown here.

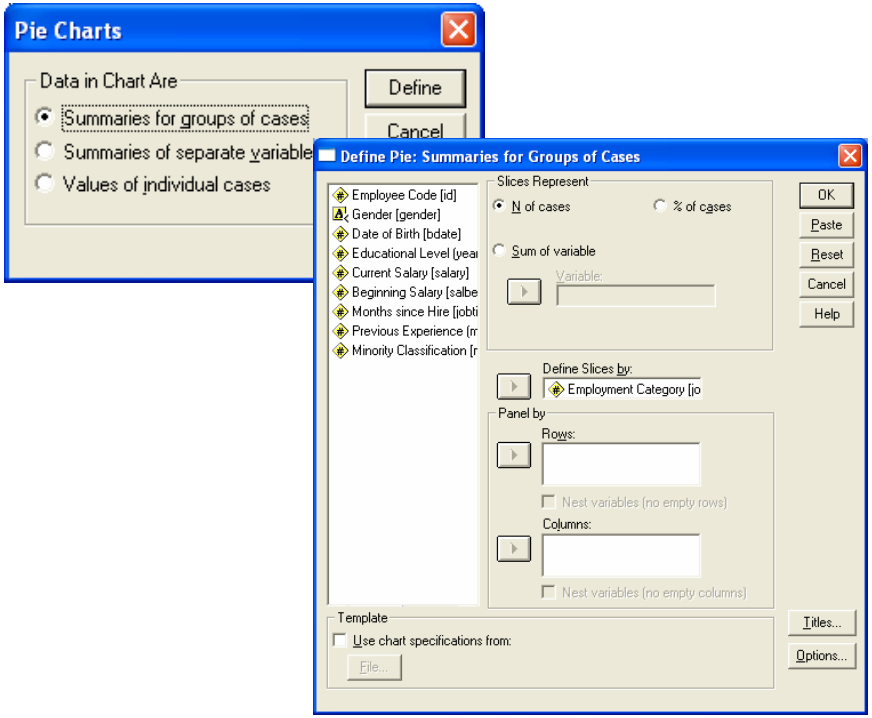
DATA: Employee data.sav

STEPS: Menu Bar => Graphs => Pie

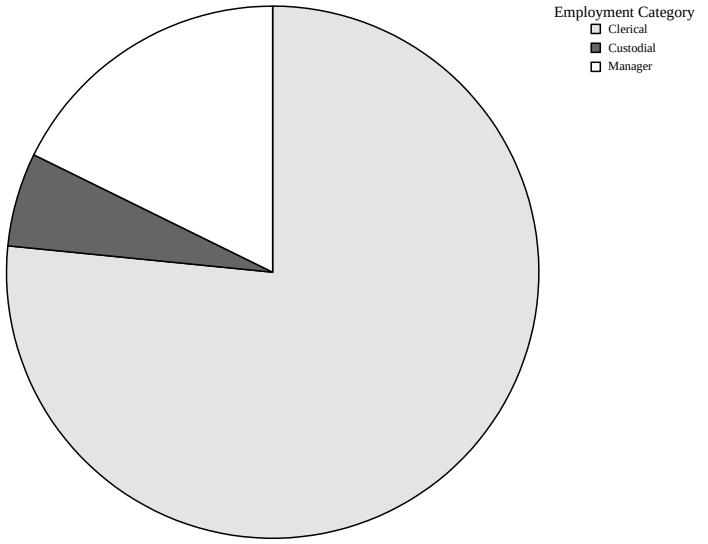
EXAMPLE: Create a Pie Chart of the variable **'JOBCAT'** (Job Category).

SUMMARY OF STEPS: In the **Pie Charts** dialog box, **Summaries for groups of cases** is the default. => Click **Define.** => In the **Define Pie** dialog box **Number of Cases (N of Cases)** is the default. => Select your variable by clicking once on it, and then clicking the right arrow next to the **Define Slices By** box. => Click **OK** and the **Pie Chart** will appear in the Output Viewer.

Dialog Boxes: "Summaries for groups of cases" is the default => Define => Variable is selected.



SPSS Output: Pie Chart of 'JOBCAT"



As is the case for all charts and graphs, the **Chart Editor** (see **Chart Editor** section), contains many useful tools for adding to, and changing the appearance of your chart or graph. Two of these tools (adding counts and percentages, and "exploding" a section) are particularly useful, and will be presented here.

MODIFYING A PIE CHART: Counts, Percentages and Exploding a Slice.

STEPS: Double click on the pie chart in the Output Viewer. This will open the **Chart Editor**. Make the desired modifications and then close the Chart Editor. The edited pie chart will replace the original chart in the Output Viewer.

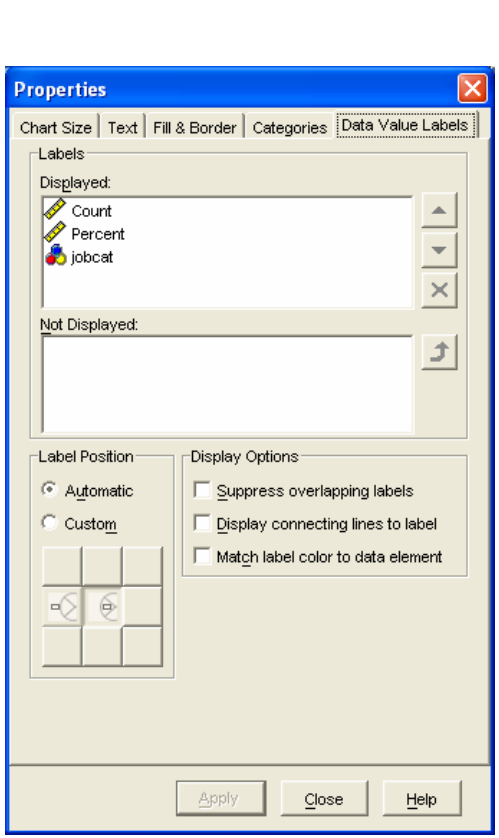
Chart Editor Icons: On the Chart Editor row of icons three assist in applying labels and exploding slices. The **Data Label Mode** icon (looks like crosshairs in a gun sight) is either on or off. When the **Data Label Mode** is "off," the **Data Labels** (looks like a mini bar chart) and the **Explode Slice** (looks like a mini pie chart) are highlighted. In this case either of these actions may be changed. If the Data Labels Mode icon has been selected (toggled on), the Data Labels and Explode Slices options will not be highlighted and will not be operable. All of these options are listed on the Menu Bar => Elements option.

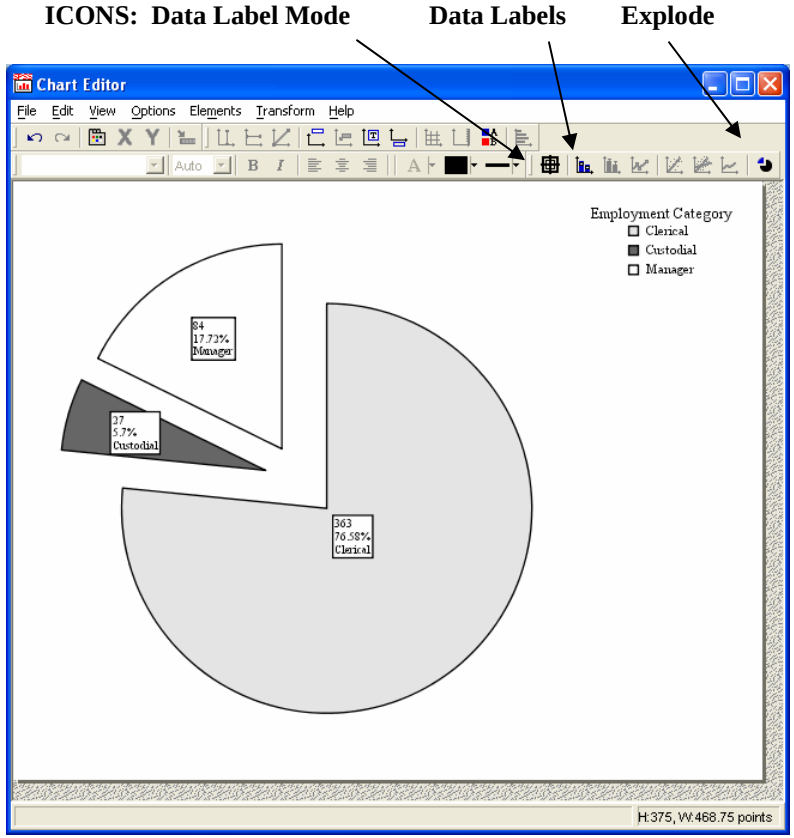
Counts & Percentages: From within the Chart Editor selecting the **'Data Labels'** icon (Main Menu => Elements => Show Labels) toggles labels on/off . If the toggle is selected to show labels, the **Properties** dialog box shown below is displayed. Select the **Data Value Labels** tab, move the desired items into the **Displayed** box and select the **Apply** option.

NOTE: To apply labels to ALL slices just select the icon or click once on the pie and then select the icon. To apply labels to only a selected slice click once on the pie and then a second time on the desired slice. In both cases you will notice a shading line surrounding the portion of the pie that will be affected by subsequent commands.

Clicking on the **Data Label Mode** icon turns the mouse arrow into a replica of the icon. Placing the icon over a slice and clicking the mouse button will display or hide the labels for the selected pie slice. To turn this feature off click on the toolbar icon.

Exploding a Slice: From within the Chart Editor select the **Explode Slice** icon (Main Menu => Elements => Explode Slice). Explode Slices is a toggle on/off option. If you have not selected a specific slice (see preceding NOTE), all slices will be exploded. Selecting a specific slice will result in only that slice being exploded. This process can be repeated with another slice by repeating the preceding process.





SCATTERPLOT

A **Scatterplot** is used to display information about the relationship (or lack of) between two variables. One point is plotted for each (x,y) pair in the data set.

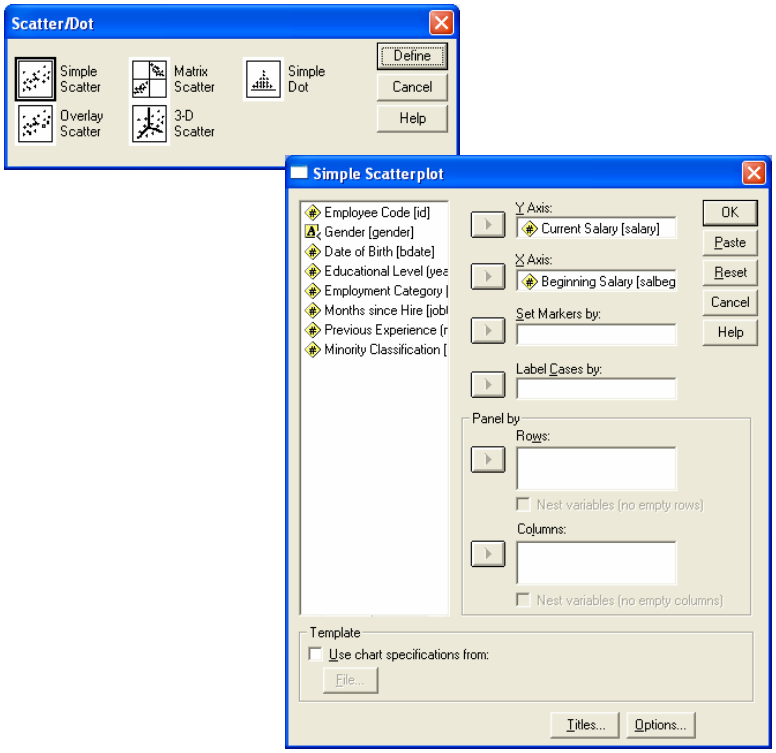
DATA: Employee data.sav

STEPS: Menu Bar => Graphs => Scatter/Dot

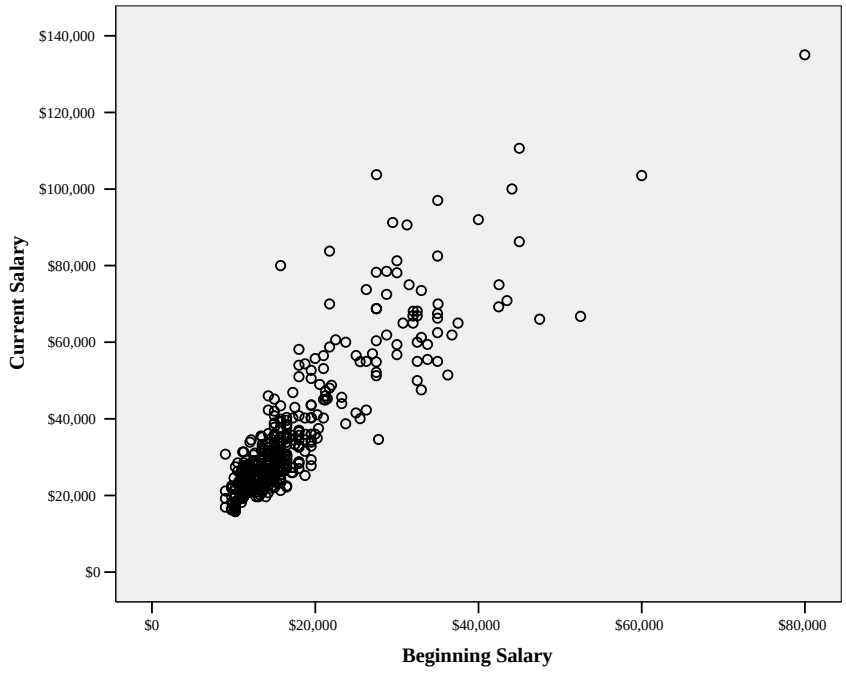
EXAMPLE: Create a **Scatterplot** of the variables **'SALARY'** (Current Salary) and **'SALBEGIN'** (Beginning Salary).

SUMMARY OF STEPS: In the **Scatterplot** dialog box, select **Simple** and then click **Define.** => Select the dependent variable by clicking once on it and then clicking the right arrow next to the **Y-axis.** => Select the independent variable by clicking once on it and then clicking the right arrow next to the **X-axis.** => Click **OK** and the scatterplot will appear in the output viewer.

Dialog Boxes: Scatterplot type is selected => select Define => select Variables.



SPSS Output: Scatterplot of Current Salary by Beginning Salary.



Simple Random Sample (SRS)

A **Simple Random Sample** is the most basic of sampling methods. The sample is chosen in a way such that each sample of a given size is equally likely to be selected. SPSS will allow you to select a **Simple Random Sample** based on a percentage of cases, or an exact number of cases.

DATA: Employee data.sav

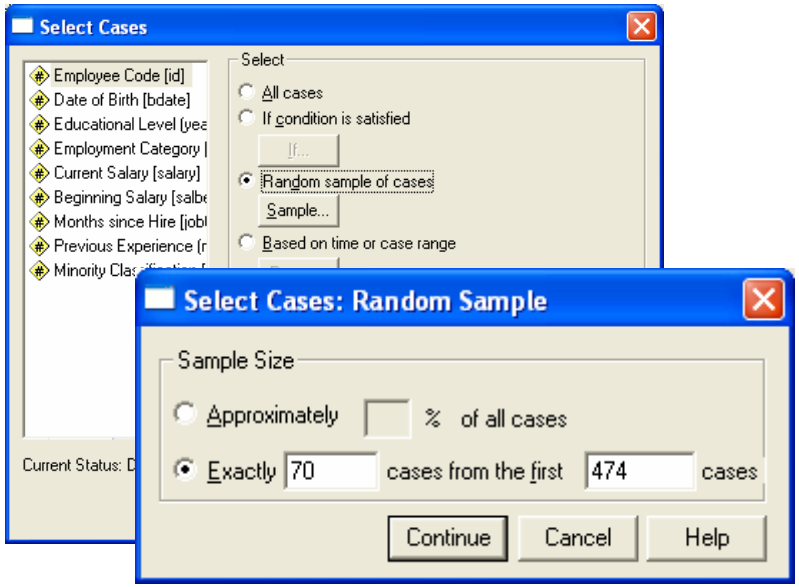
STEPS: Menu Bar => Data => Select Cases.

SUMMARY OF STEPS: In the dialog box that will open, select the **Random Sample of Cases** button and click on the **Sample** box below it. => A second dialog box will open. In this box, you will designate a percentage of cases, or an exact number of cases for your sample. If you wish to select a percentage of cases, click on the **Percentage** button, and type in the desired percentage. If you wish to select a specific number of cases, click on the **Exactly** button and enter the sample size and the population size (or from the first **k** cases, where **k** is dependent on the specific situation). In most cases **k** will be the number of cases in the data file. => Click **Continue.** This will bring you back to the first dialog box. => Select **OK**. The cases in the worksheet that were chosen for the sample will **not** have a line drawn diagonally through their respective case numbers.

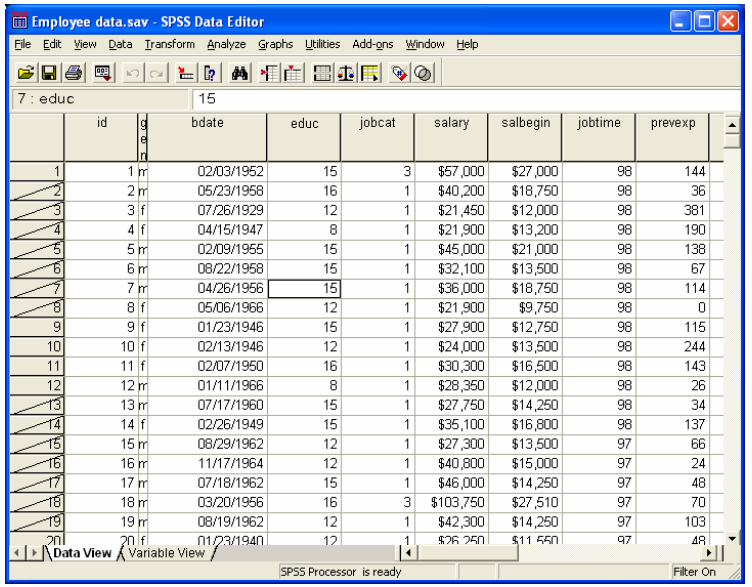
NOTE: When you are done using this sample you must return to the **Select Cases** dialog box and reset the selection to **All Cases.**

EXAMPLE: Select a ***Simple Random Sample*** of size n = 70 from the population of N = 474.

Dialog Box Input: Random sample of cases is selected. The desired number of cases to be chosen from the first **k** cases is designated.



SPSS Output: The sample is denoted by cases for which no line has been drawn through the case number located along the left edge.



STANDARDIZED SCORES (Obtaining z-scores)

To obtain **Standardized Scores**, you will need to use the **Compute** function (see **Compute**). You will also need to know the value of the population mean and standard deviation.

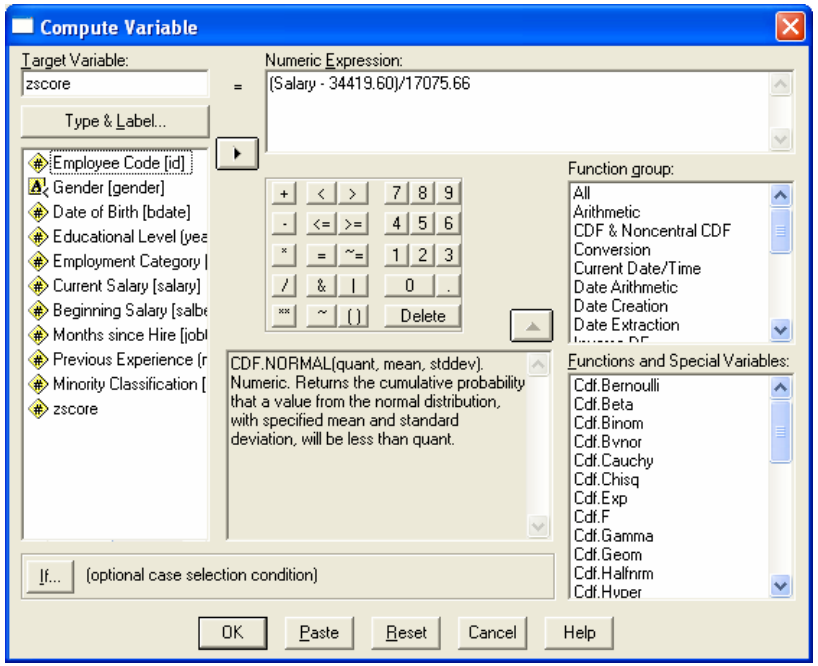
DATA: Employee data.sav (assume this to be a population with n = 474, μ = \$34.419.60 and σ = \$17,075.66)

STEPS: Menu Bar => Transform => Compute

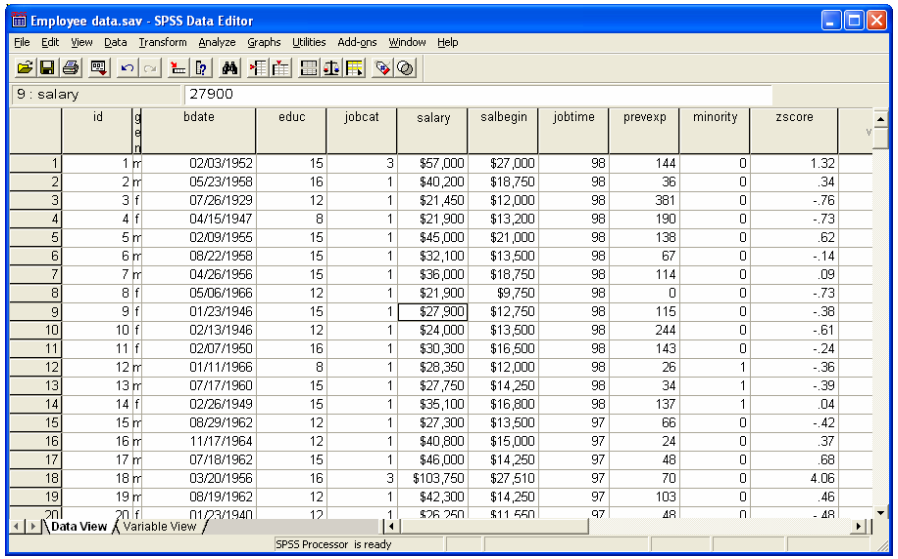
EXAMPLE: Obtain z-scores for the variable **'SALARY'**.

SUMMARY OF STEPS: NOTE: Prior to obtaining z-scores the mean and standard deviation for the variable must be obtained (see Descriptive Statistics). Select the **Target Variable** box, and type in a name for the column that will contain the z-scores. In this example, the column is named **'ZSCORE'.** => Select the **Numeric Expression** box. It is here that you will "build" your formula to calculate the z-scores. Since the population mean and standard deviation are equal to \$34,419.60 and \$17,075.66 respectively, your formula will be **(salary – 34419.60)/17075.66.** => When you have entered the formula, click **OK.** A column containing the z-scores for the variable **'SALARY'** will appear at the end of the Data Editor worksheet (immediately following the last variable).

Dialog Box: Target variable is named, and formula is "built."



SPSS Output: A column containing the z-scores is created in the Data Editor.



STATISTICS & CASE PROCESSING TABLES

Each time an analysis occurs a Statistics Table or a Case Processing Table is included in the analysis output.

Statistics Table: If no specific statistics are requested, the Statistics table simply presents the number of valid and missing cases. If statistics, such as a variable's mean or median is requested as part of the data analysis, these statistics are included in the Statistics Table for that analysis.

Case Processing: When the analysis includes multiple variables, such as in the case of a Crosstabs, a Case Processing table replaces the Statistics table. This table will include information about the valid number of cases for each variable and the number of cases used to process the analysis request.

DATA: Cars.sav

STEPS: None. **These tables are automatically produced in conjunction with other analysis requests.**

SUMMARY OF STEPS: Conduct some form of data analysis.

STATISTICS TABLE: As an example, create a frequency table of the variable **'WEIGHT,'** In its simplest form, a Statistics Table presents the number of valid and missing cases. If descriptive statistics are requested, the Statistics Table will be expanded to include those statistics.

CASE PROCESSING TABLE: As an example create a crosstabs of the variables '**CYLINDERS**' and **'ORIGIN'.** The Case Processing Table will display the number of cases which have data for both variables ("Valid") and the number of cases for which one or both variables is/are missing ("Missing"). Note that the number in the Total cell of the crosstabs table (not shown) reflects the number of valid cases (i.e. those with data for both variables).

Statistics Table accompanying the frequency table of the variable **'WEIGHT'.** (left without stats; right with stats)

Statistics

Vehicle Weight (lbs.)		
N	Valid	406
	Missing	0

Statistics

Vehicle Weight (lbs.)		
N	Valid	406
	Missing	0
Mean		2969.56
Median		2811.00
Std. Deviation		849.827

Case Processing Table accompanying the crosstabs table of **'CYLINDERS'** and **'ORIGIN'.**

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Number of Cylinders * Country of Origin	405	99.8%	1	.2%	406	100.0%

STEM-AND-LEAF PLOT

A **Stem-and-Leaf Plot** is similar to a histogram in that it shows where data is concentrated, the general shape of a distribution, the range of the data, and whether or nor outliers may be present. A **Stem-and-Leaf Plot** places the last digit of a value in the leaf, and all prior digits in the corresponding stem.

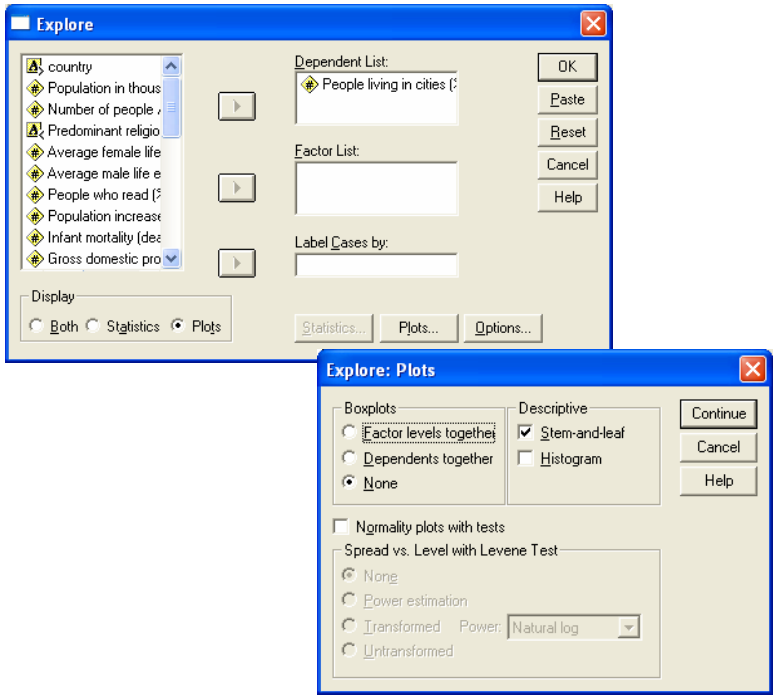
DATA: World95.sav

STEPS: Menu Bar => Analyze => Descriptive Statistics => Explore

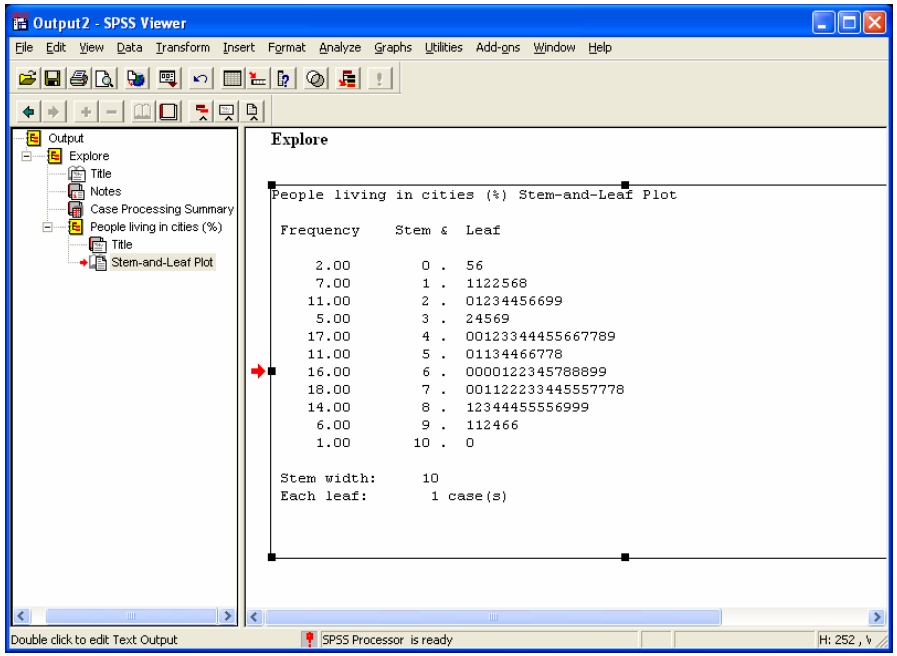
EXAMPLE: Create a Stem-and-Leal Plot of the variable 'URBAN'.

SUMMARY OF STEPS: In the **Explore** dialog box, select your variable by clicking once on it and then clicking the right arrow next to the **Dependent List** box. => In the **Display** section of the dialog box, click on the **Plots** button. => Select the **Plots** button. The **Explore Plots** dialog box will appear. => In the **Explore Plots** dialog box, click **None** under the **Boxplot** section (unless of course, you want a boxplot also). => Click **Stem-and Leaf** under the **Descriptive** section and then the **Continue** button. This will bring you back to the **Explore** dialog box. => Select **OK** and the **Stem-and-Leaf Plot** will appear in the Output Viewer.

Dialog Boxes: Variable is selected => select Plots button => select Stem-and-Leaf => Continue => OK.



SPSS Output: Stem-and-Leaf Plot appears in Output Viewer.



T-TEST

T Tests are used to compare means between two groups or two variables. In the case of the **Independent Samples T Test** (used here in the EXAMPLE), the procedure compares the means for two groups on a single variable, such as the mean current salary for male and female employees. Refer to a statistics text for considerations associated with the use of the test. The **Paired-Samples T Test** (not shown) is used in instances similar to a pre-test post–test situation where one individual has two measures and one is seeking to determine a difference in the two variable means.

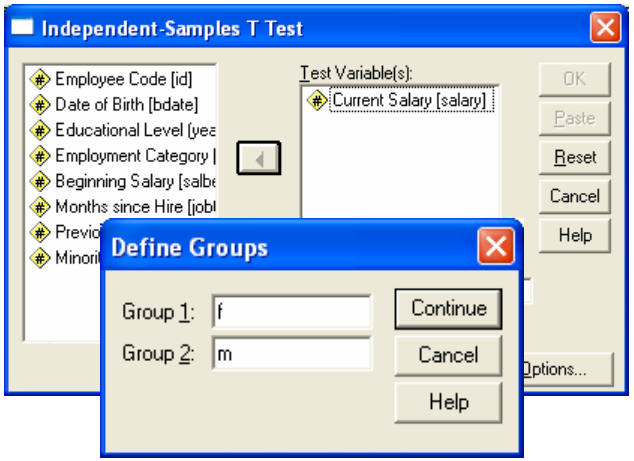
DATA: Employee data.sav

STEPS: Menu Bar => Analyze => Compare Means => Independent Samples T Test => identify variables and define Grouping Variable values => OK

EXAMPLE: Does there exist a difference in the mean salary paid male and female employees?

SUMMARY OF STEPS: Select **Analyze => Compare Means =>Independent Samples T Test**. Enter the variable of interest, **'SALARY,'** into the **Test Variables** cell and the **Grouping Variable**" here **'GENDER,'** into that cell. Note that with the **Grouping Variable** there will be enclosed question marks after the variable. The range of values for this variable must be identified. In the case of **'GENDER'** the values in this data set have been assigned **"m"** for male and **"f"** for female. Other data sets may use numeric values to define variables.

Dialog Boxes: Identify variables of interest and define the Grouping Variable values to be used.



SPSS Output: Descriptive statistics for each group and T Test

Group Statistics

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

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
Current Salary	Equal variances assumed	119.669	.000	-10.945	472	.000	-$15,409.862	$1,407.906	-$18,176.401	-$12,643.322
	Equal variances not assumed			-11.688	344.262	.000	-$15,409.862	$1,318.400	-$18,002.996	-$12,816.728

SPSS DATA EDITING & MANIPULATION:
PROCEDURES TO EDIT, MODIFY, TRANSFORM & REPORT SPSS DATA

CHART EDITOR

As with statistical analyses, charts & graphs are presented in the Output Viewer. The **Chart Editor** is a separate window containing a series editing options. Double clicking on a chart will open the **Chart Editor** and make the selected chart available for modification. Many of the options are presented as icons on a bar below the Main Menu Bar. Editing options are also present on one of the Main Menu categories. Each icon's function may be determined by placing the mouse cursor on the icon and waiting momentarily. When a chart is being edited, its counterpart in the Output Viewer will appear shaded. For this example, a Clustered Bar Chart will be used. Refer also to the **Pie Chart** documentation for additional Chart Editor options.

DATA: Employee data.sav

STEPS: Double Click on any graph or chart.

Initial Bar Chart

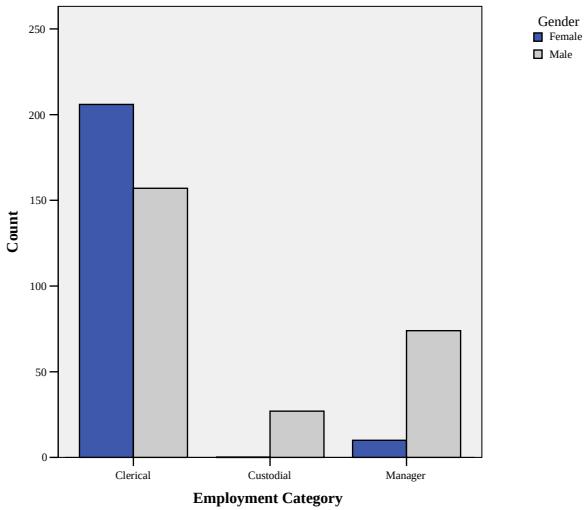
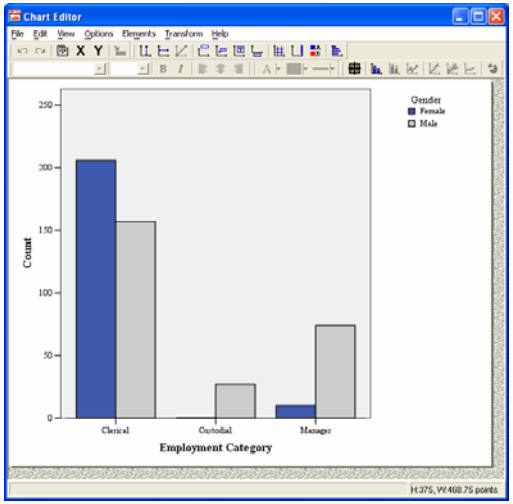


Chart Editor with selected chart opened for editing.



Output Viewer with Chart being edited is shown shaded.

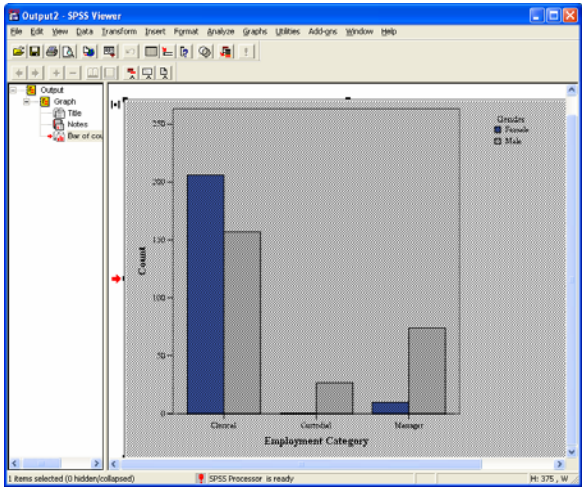
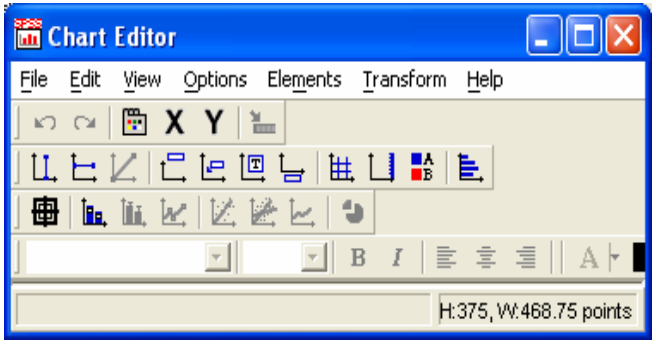
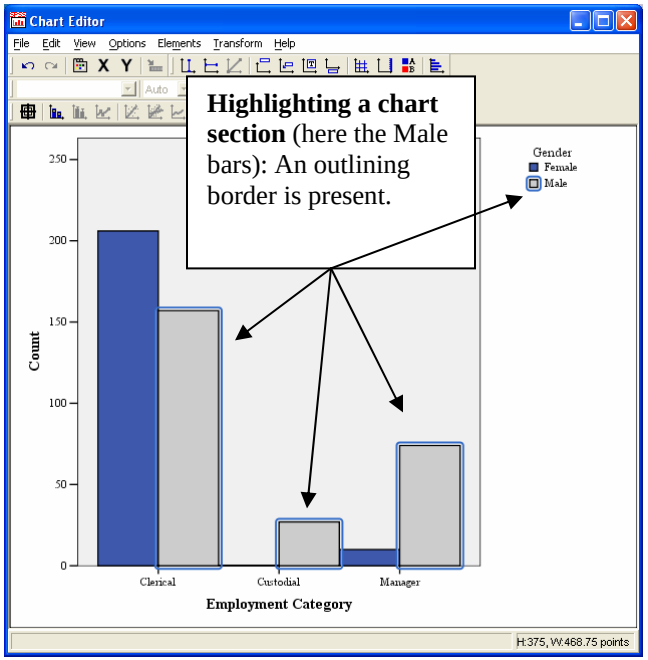


Chart Editor Menu Bar & Toolbar Icons
(compressed to show different toolbars)



- **Edit Menu -Toolbar Icons:** Undo, Redo, Properties, X-Axis, Y-Axis, etc. =>
- **Options Menu – Toolbar Icons:** X & Y Reference Lines, Title, Annotation, Text Box, Footnote, Gridlines, Legend, Transpose, etc. =>
- **Elements Menu – Toolbar Icons**: Data Label Mode, Show Data Labels, Explode Slice, etc. =>
- **Format Toolbar Icons:** Bold, Italics, Centering, Text Color, etc. =>

CHART EDITOR



Moving around in a chart: Although not initially obvious, when a chart is opened in the Chart Editor there are several sections that may be modified. At the outermost perimeter of the chart is an area that may be colored or a border applied. Other sections that may be modified include the area around the x and y axes titles, the bars, the area behind the bars and the legend section. As you click on one of these areas a color border outlining the area will appear.

Highlighting a Bar: One mouse click on any bar in a bar chart will create a border around all the bars. A second click will place the border around the selected bar. In the case of a clustered bar chart, one click selects all bars, two clicks selects all bars of the color where the click occurred, and three clicks narrows the selection down to the specific bar clicked upon.

Editing Text: To edit text, a quick double-click of the mouse will open the Properties dialog box associated with the text. There you can change fonts and size. A SLOW two clicks allows access to the actual text. At that point you can modify the text itself.

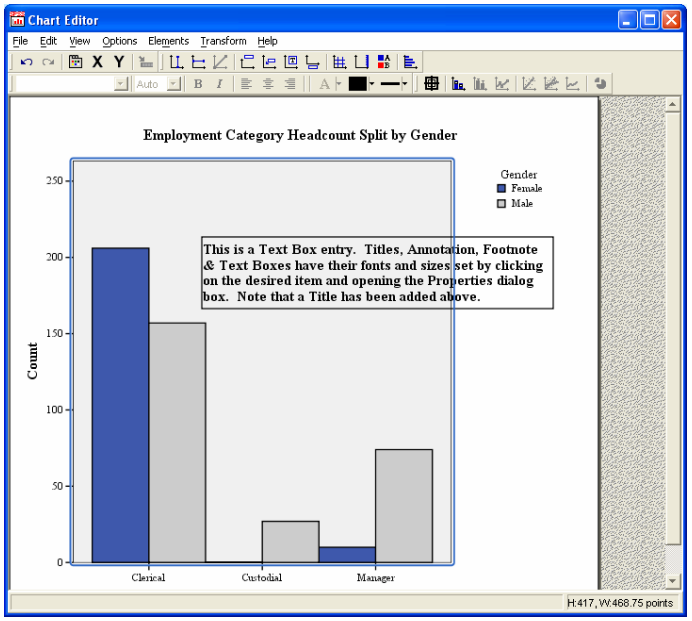
Transpose: On the **Options Menu** and its toolbar, **Transpose** changes the x-y axes orientation of a chart (as shown below).

Transform: The **Transform Menu** displays other charts that the current chart could be transformed into. For example, by selecting one of the available chart choices the clustered bar chart could be converted to a stacked bar chart, a line graph or an area graph.

Titles, Annotation, Footnotes & Textboxes: All of these options are available via the **Options Menu** and its icons.

Exiting the Chart Editor: When all edits have been completed the modified chart is returned to the **Output Viewer** by clicking on the **"X"** in the upper right corner of the Chart Editor or from the **Main Menu Bar => File => Close**.

From the Output Viewer you can save the chart as part of an output file (files with extension .spo) or move a copy into a document.



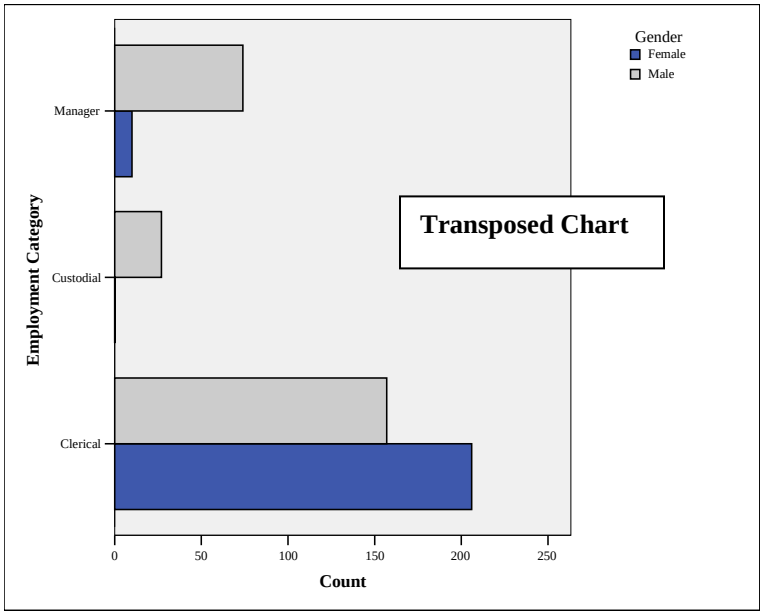


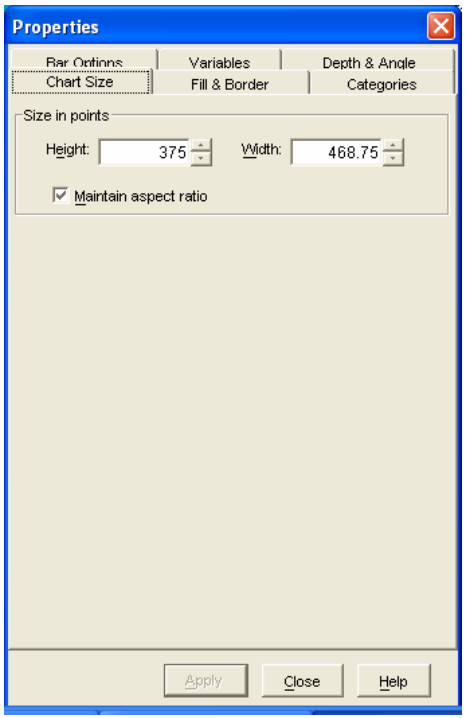
CHART EDITOR – Properties

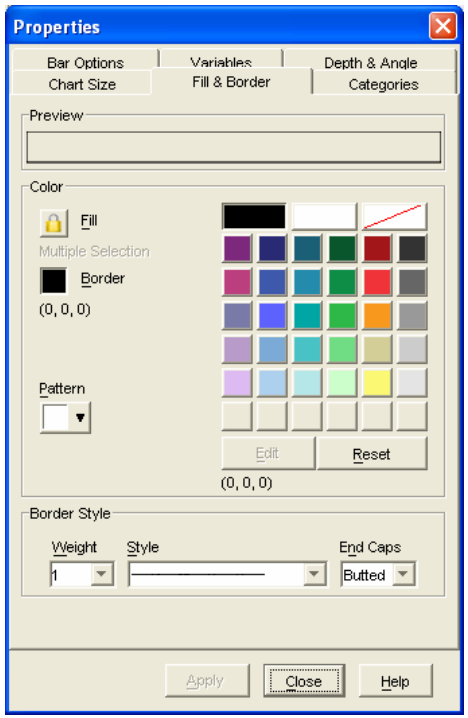
Properties Dialog Box: This dialog box contains many of the editing features that you might want to perform on a chart. The number of options available through the Properties dialog box will vary based upon the type chart being edited and the portion of the chart activated for editing. The following presentation pertains to a clustered bar chart as seen above, but will give some idea of how to move around when using the Properties option.

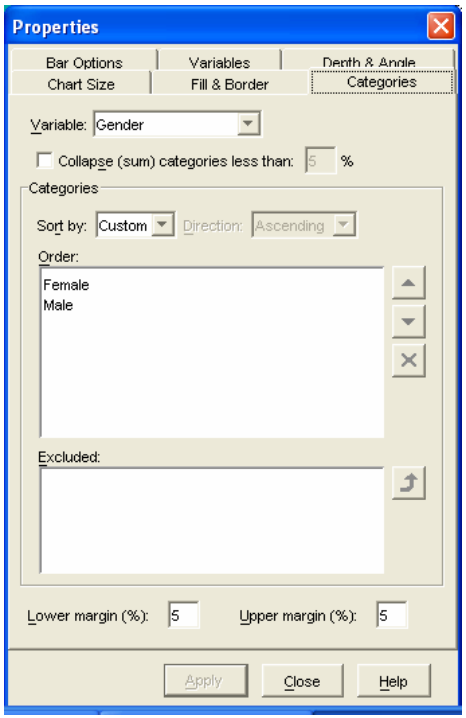
- **Chart Size:** Allows modification of the overall chart size.

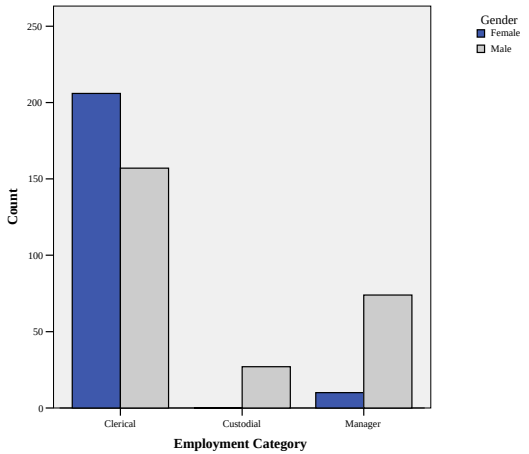
- **Fill & Border:** Border color and thickness as well as bar colors and patterns are set here. Note that to the right there is a lock on the **Fill** option. This is because more than one color of bar has been highlighted. Clicking a second time on one bar will highlight all bars of that color and the lock should be replaced by the bar's current color.

- **Categories:** In the far right dialog box, Female and Male represent the sub-categories of the bar chart. By moving one of the values to the **Exclude** box and selecting the **Apply** button the chart reforms as a bar chart for the remaining value. Categories may also be combined based upon an identified percentage.









<= Initial Bar Chart

Modified Bar Chart =>
1) Chart Size: Not changed.
2) Fill & Borders: Area around chart shaded (light blue), Pattern applied to the dark colored bars.
3) Categories: Female/Male Order switched.

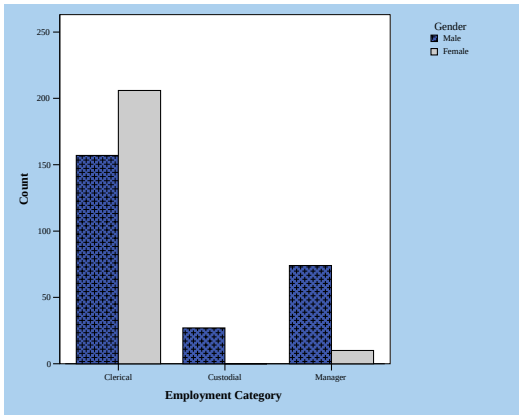
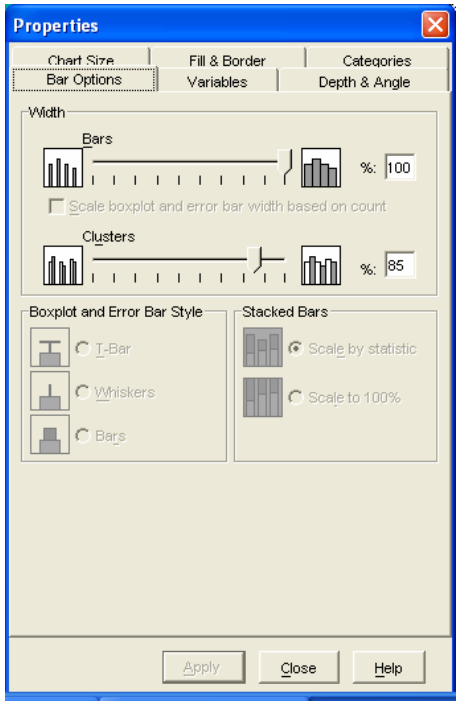
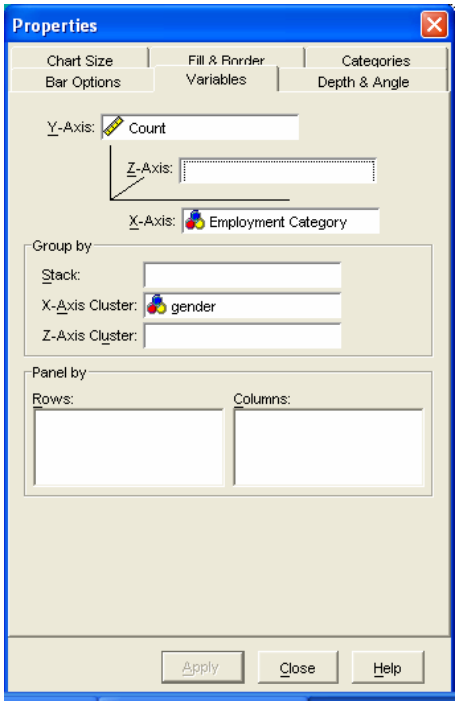


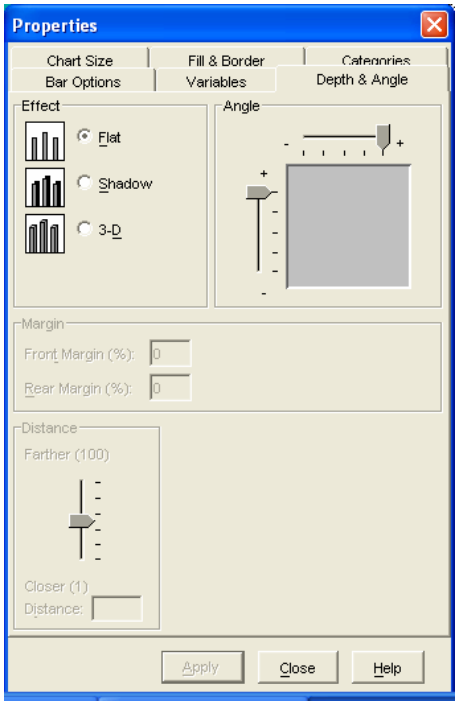
CHART EDITOR –Properties (cont.)

- **Bar Options:** Among the options here are **Bars** which adjusts the width of the bars and **Clusters** which is used to create space between groupings.

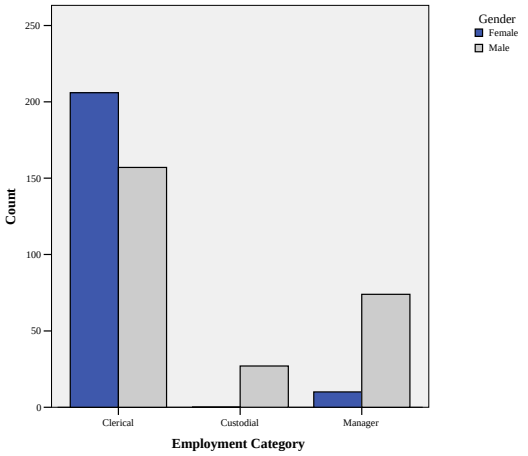
- **Variables:** On this tab you can re-order variables or create different formats for the chart. For example, switching the locations of the two variables would create a clustered bar chart where the x-axis was gender and each gender would have three job categories. By placing the Gender variable in the z-axis a grid-like chart is created where across the x-axis there would be job categories broken down by gender across the z-axis. Think of it as a graphic presentation of a Crosstabs table. An example is presented as the final chart in the Chart Editor section.







- **Depth & Angle:** Three general shapes for bar charts are presented. Selecting either the **Shadow** or **3-D** options activates the **Angle** option. In the **Angle** grid the size of shadows and the tilt of 3-D bars may be adjusted. The **Distance** option changes the size of the chart, making it appear to move forward of backward on the screen.



<= Initial Bar Chart

Modified Bar Chart =>

1) Bar Options: Bars to 75% (moves bars apart), Clusters to 65% (moves clusters apart)

2) Variables: Not changed.

3) Depth & Angle: Changed to shadow style and adjusted the shadow angles relative to the bars.

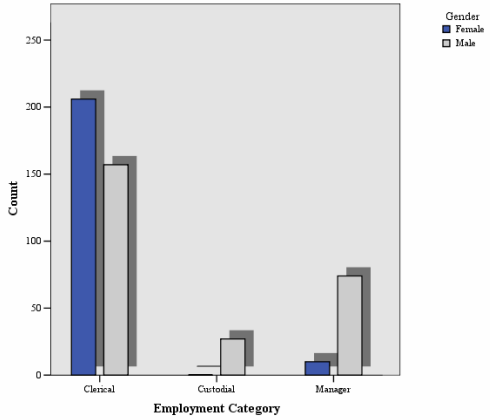
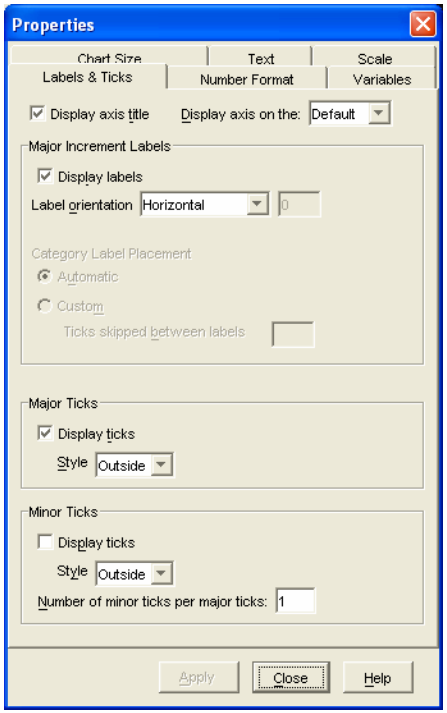
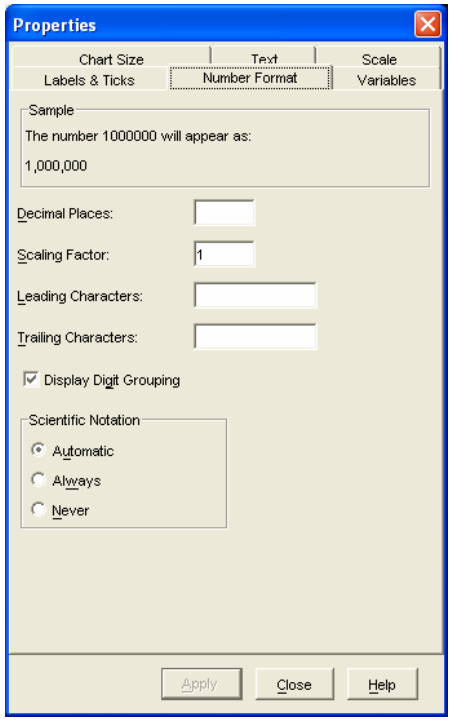
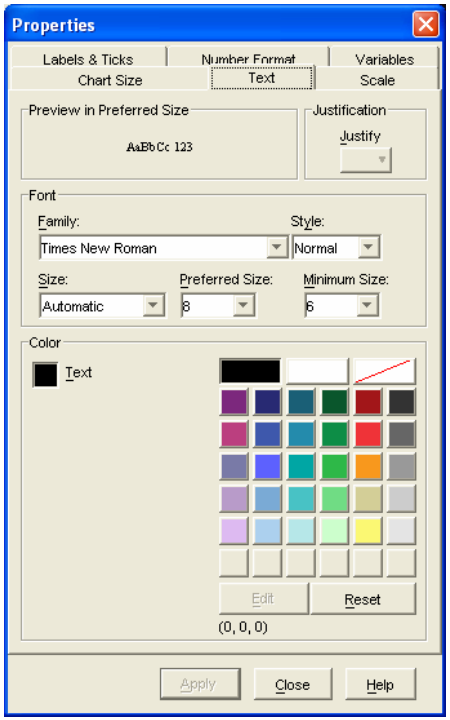


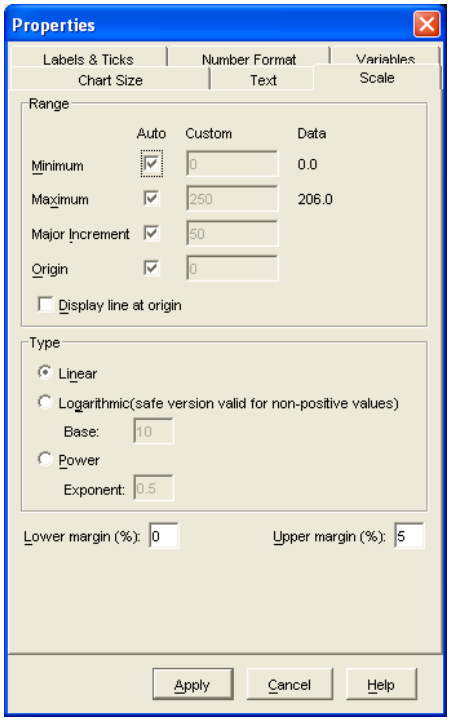
CHART EDITOR –Properties (cont.)

Y-Axis Properties:









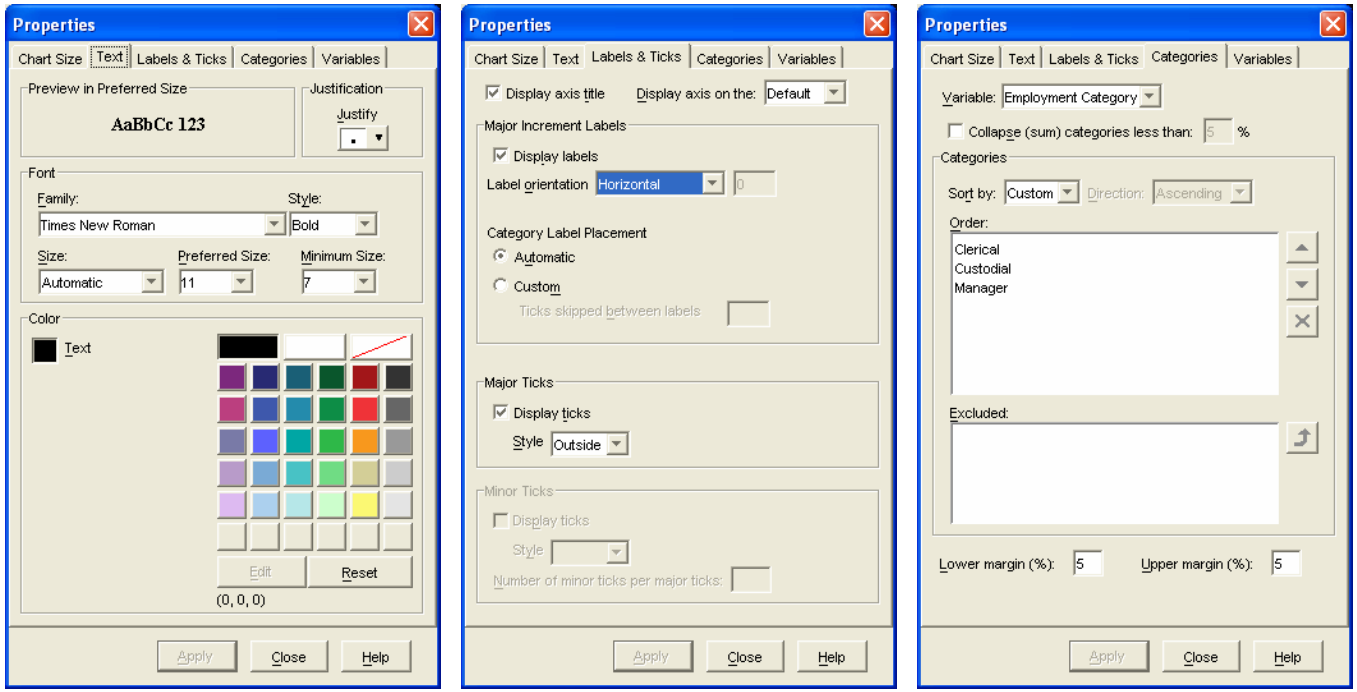
- **Labels & Ticks: Major Increment Label** orientation – horizontal, diagonal, etc.; **Major & Minor Ticks** – Inside, outside or through the y-axis line; Set the number of Minor Ticks between Major Ticks.
- **Number Format:** The **Display Digit Grouping** option places the commas into the presentation as shown in the sample.
- **Text:** Set the color, font and size of y-axis terms here. Note that you have to have selected (clicked on) the text for which theses characteristics are o be set.
- **Scale:** As shown, the structure of the y-axis scale is determined by the program (all boxes checked). By specifying one or more of the **Range** items you can customize the scale to fit your needs.

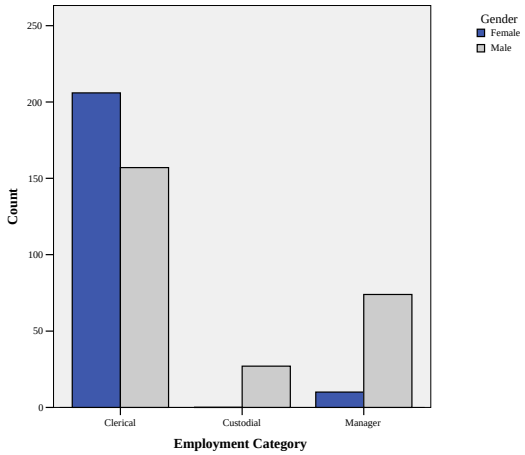
See example of y-axis changes on the next page.

CHART EDITOR –Properties (cont.)

X-Axis Properties:

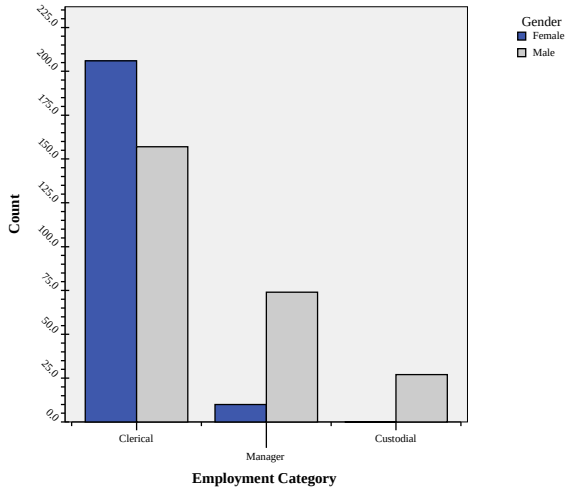
- **Text:** Set the color, font and size of x-axis terms here. Note that you have to have selected (clicked on) the text for which theses characteristics are to be set.
- **Labels & Ticks: Major Increment Label** orientation – horizontal, diagonal, etc.; **Major Ticks** – Inside, outside or through the x-axis line.
- **Categories:** Add or delete and change the order of x-axis variable values.





<= Initial Bar Chart

Modified Bar Chart =>



Y-AXIS:

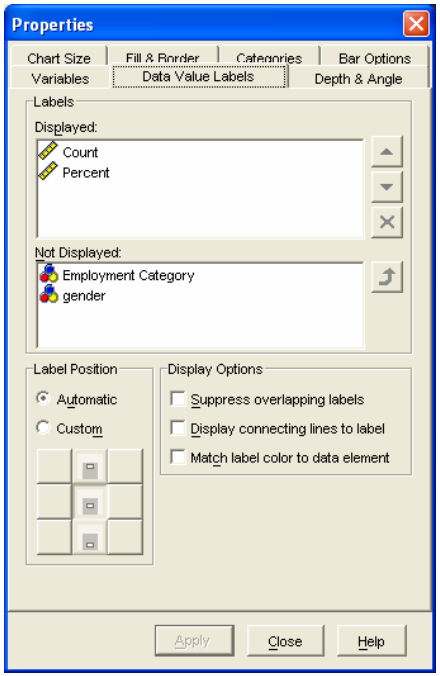
1) Labels & Ticks: Labels Diagonal, Major Tick Style = Through, 4 Minor Ticks between Major Ticks.
2) Number Format: One Decimal place.
3) Text: Not changed.
4) Scale: Maximum set to 225 and Major Increments set to 25 (default was 250 & 50).

X-AXIS:

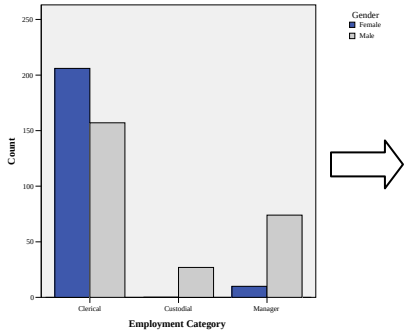
1) Text: Not changed.
2) Labels & Ticks: Labels Staggered.
3) Categories: Reordered from most to least number of employees.

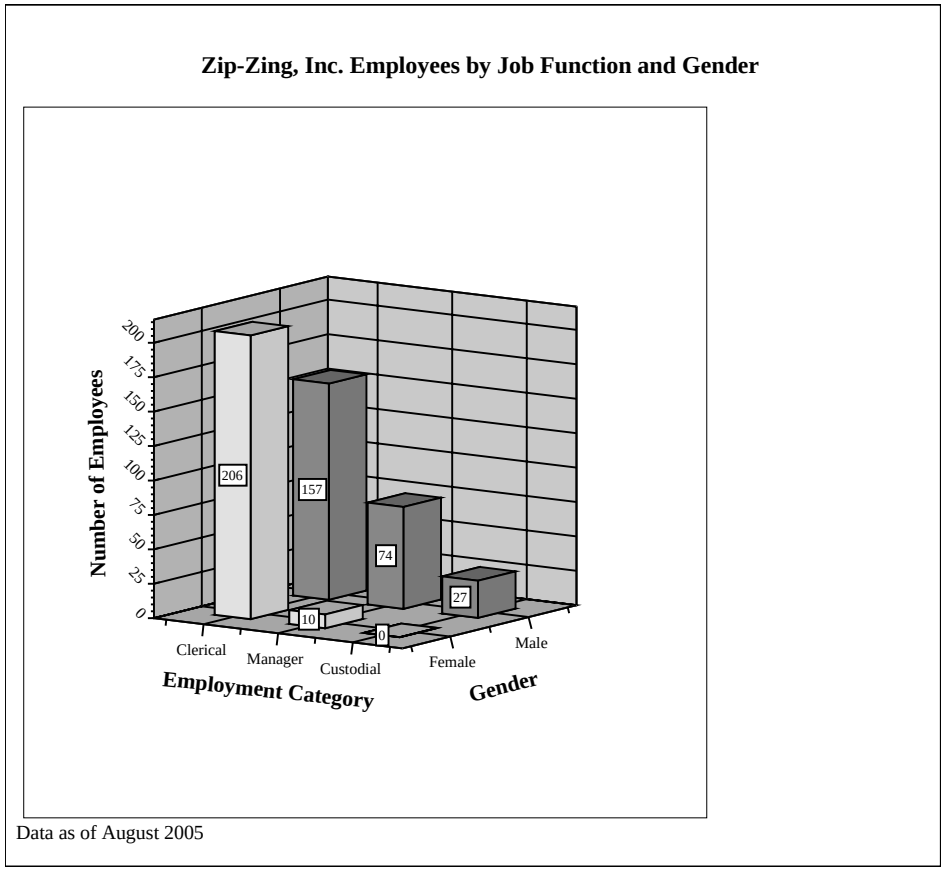
CHART EDITOR –Properties (cont.)

Data Label Mode Properties: (see also **PIE CHART** for additional information on this option.)



Initial Clustered Bar Chart to one of many possible modified versions.





On the **Elements Toolbar** are two icons that identify the contents of value labels and their location. Selecting the **Show Data Labels** icon (looks like a little bar chart) opens the **Data Value Label** section of the **Properties** dialog box. Here the information to be shown as a label for a bar is moved into the top box. Above both a Count and a Percentage will be displayed for selected bars.

The **Data Label Mode** icon (looks like a crosshairs in a gun sight) when clicked changes the mouse arrow to the displayed icon. When placed over a bar and clicked, the defined value labels are included on the chart. Clicking again on the same bar turns off the label.

Question: Suppose that you have several bar charts to edit and move into a report. Now that numerous changes to a bar chart have been made and a considerable amount of time expended, do you have to do this all over again for all the other charts of this type? Of course not… **From within the Chart Editor** save the bar chart's characteristics as a **Template** (File => Save Chart Template). When each bar chart is brought into the Chart Editor for editing, apply the Template (File => Apply Chart Template) and the basic format of the chart will be the same as all others. As charts differ, you may need to do some basic editing, such as titles. This process would apply to other types of charts as well.

COMPUTE

The **Compute** command is used to generate a new variable composed of existing variable values, mathematical operations, and/or functions.

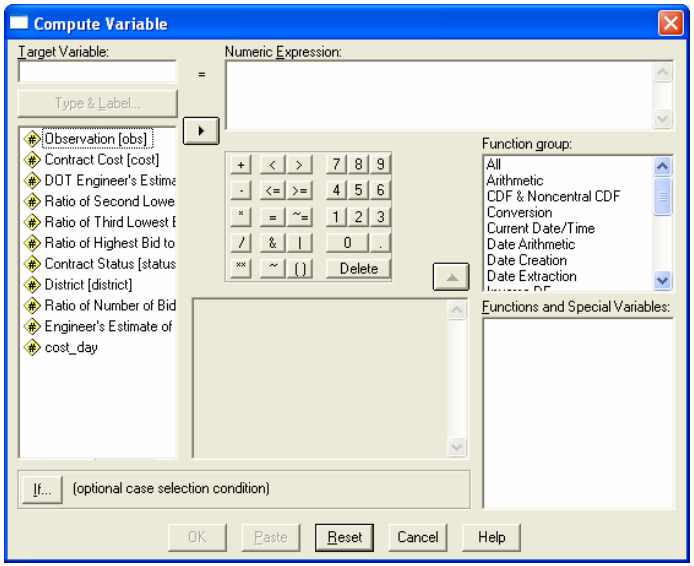
DATA: Road Construction Bids.sav

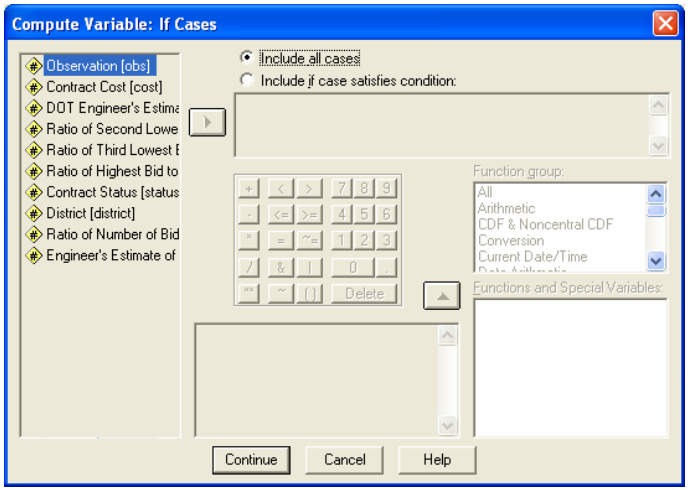
STEPS: Menu Bar => Transform => Compute

SUMMARY OF STEPS: MENU BAR => TRANSFORM => COMPUTE => Name Target Variable => Highlight variable as component of new variable. => Press arrow button. => Enter mathematical operation or function. => Repeat prior two steps until expression is complete. => If all is correct the OK button will be activated. => Exit with New variable by pressing the OK button. => Save the data set.

NOTE: Computed variables will appear **at the end** of the Data Editor variable listing.

STEPS IN DETAIL:

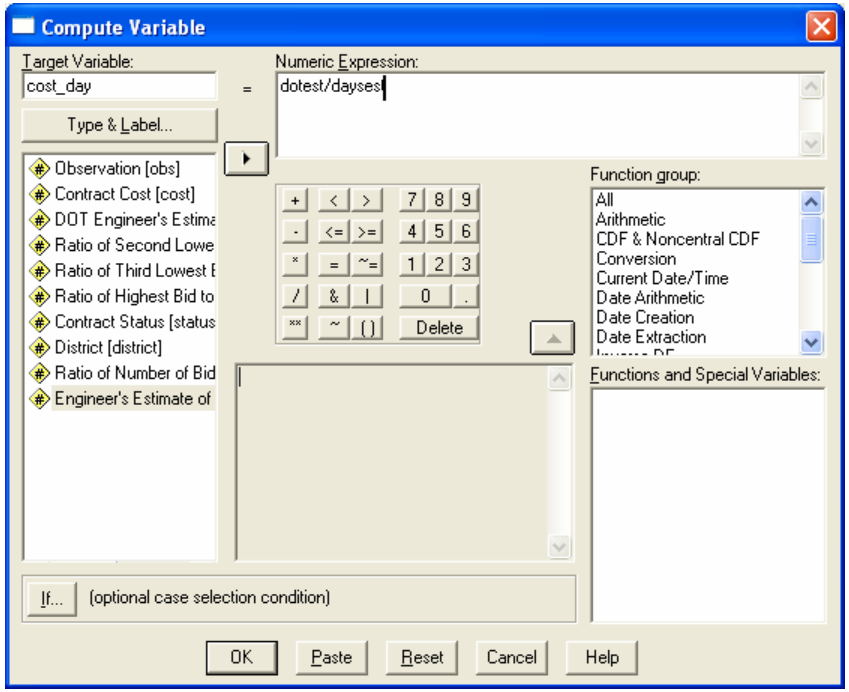




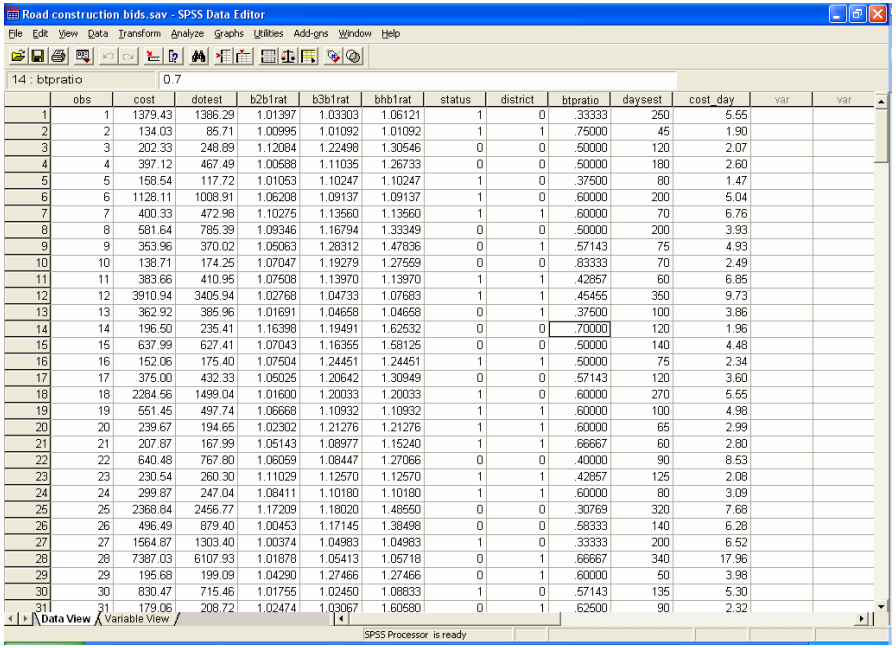
- From the SPSS Main Menu Bar select **Transform**.
- Select **Compute**. A Compute dialog box will open (shown to right).
- Enter a **Target Variable** (the new variable being created) in the box at the upper left.
- Select a variable from the list of existing variables (highlight it by clicking on it) and then click on the **Arrow button**. This variable is added to the **Numeric Expression** box and becomes the first component in the creation of the computed variable. Select an operation from the "keypad" by clicking on it or when appropriate a function by highlighting a **Function Group** and the specific **Function/Special Variable** and then clicking on the **upward pointing Arrow Button**. The operation or function will be added to the formula in the Numeric Expression box. Continue with this process until all variables, operations, and functions comprising the computed variable have been selected. Note that any variable, operation or function may be used more than once in creation of the expression. (Alternatively, you could just type in the equation in the Numeric Expression box.)
- Selecting the **If** box brings up the **Compute Variable: If Cases** dialog box (shown to right). This option would be selected if a new variable is being computed for a selected sub-set of the data set (for example, only males).
- Only when a correctly formatted expression has been completed and a Target Variable has been named, will the **OK** button be highlighted (darken). This could occur before you are finished creating your intended expression, as each addition to it could be sufficient to complete an expression. Click on **OK** when you are finished building an expression. Again, the newly computed variable will appear at the end of the data file.

COMPUTE EXAMPLE: Create a variable that represents an estimate of cost per day for the projects in the data set.

Dialog Box Input: New variable **'COST_DAY',** which is **DOTEST/DAYEST**.



Data Editor with new variable 'COST_DAY'



Other Notes:

The **Reset** button will clear both the Target Variable and Numeric Expression boxes so that you can start over. Commands entered into the Compute dialog box will remain until it is cleared or the SPSS session is ended.

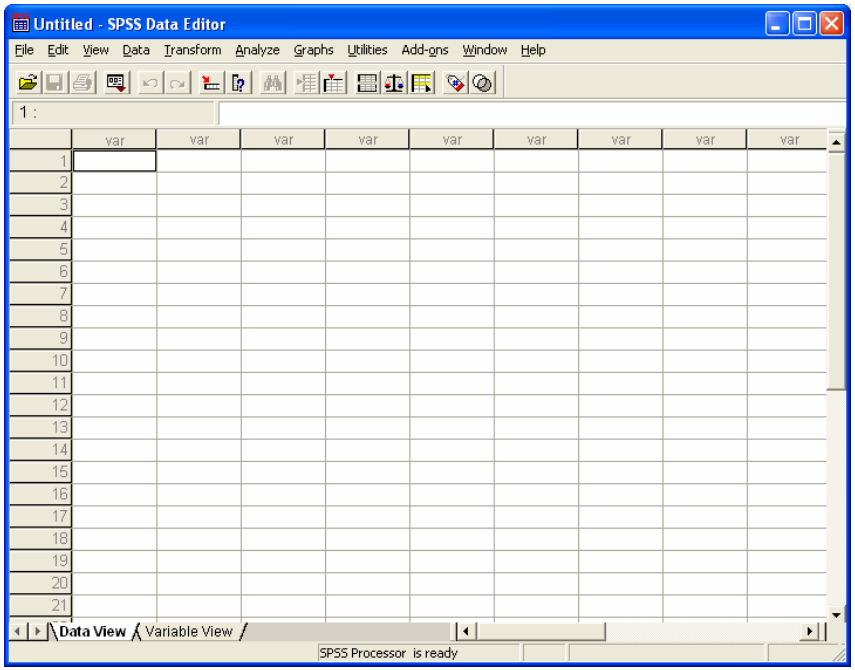
Saving the data set will include the newly computed variable in the data set.

DATA EDITOR

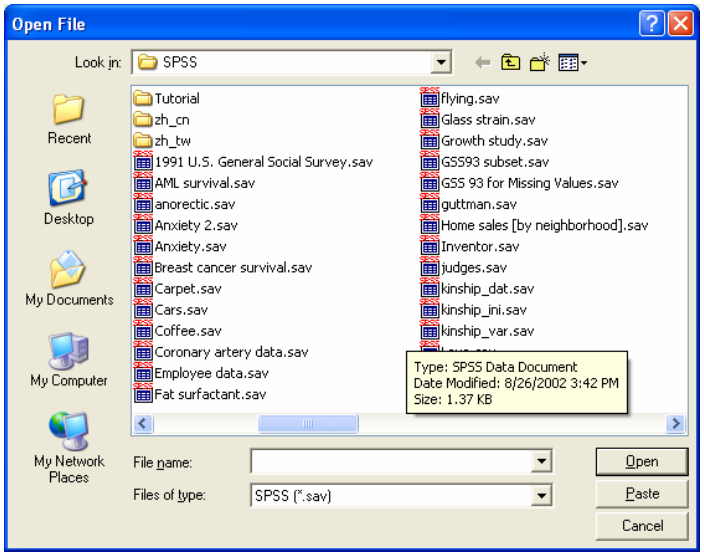
In most instances, SPSS is opened via the **Data Editor**. This editor is where the raw data resides and where modification of the data may be accomplished. The Data Editor is similar in structure to a spreadsheet. The **Columns** represent **Variables**. The **Rows** represent **Cases**[1]. The **Menu Bar** provides a series of options, each of which contains functions related to the menu heading. (See also: **ENTERING DATA INTO A NEW DATA FILE; THE INITIAL SPSS WINDOW;** and **RETREIVING A SAVED FILE**)

At the beginning of a session the Data Editor will be empty, as in the snapshot below.
Data may be entered directly into the Blank Data Editor. Data may be retrieved into the Data Editor via the Menu Bar **File** option.

- **New** => Clears the Data Editor making it ready for direct input;
- **Open** => **Data** will retrieve an existing data file;
- **Read Text Data** => Initiates a Text Wizard through which text data may be imported.



SPSS data files are saved with the extension **".sav"** and are preceded in a file listing by an icon that looks like the Data Editor.



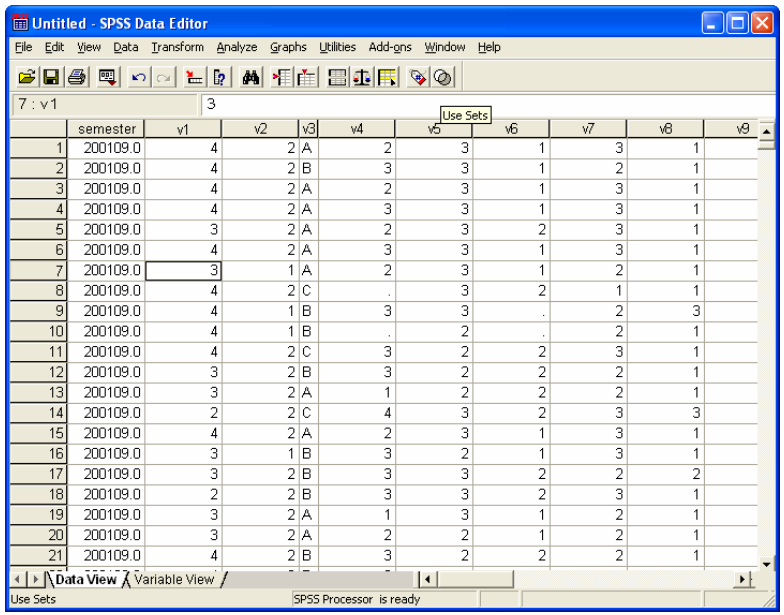
Data View and Variable View:
Note that the Data Editor has two tabs at the bottom left of the window. The **Data View** is where the data reside and where changes in the data occur. The **Variable View** is where you define the characteristics of the variables.

[1] SPSS data files contain only one record per case. During the entry of text data, multiple record cases are converted to a single record format.

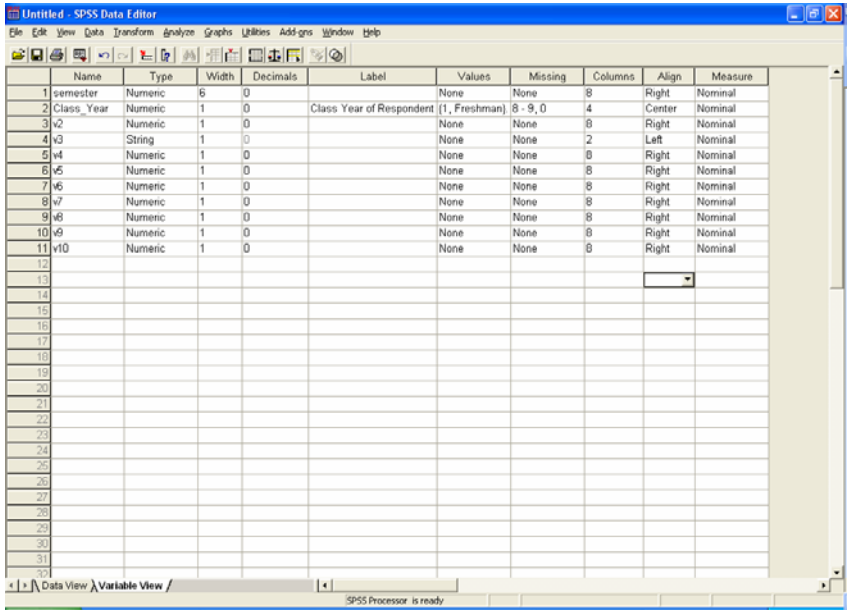
DEFINE VARIABLES

Defining variables allows one the opportunity to modify the general characteristics of a variable – whether it is numeric, string, or of another format. It also facilitates the understanding of output (tables, charts) by allowing for labels to be associated with a variable. The Data Editor has two tabs at the bottom left of the window. The **Data View** is where the data reside and where changes in the data occur. The **Variable View** is where one defines the characteristics of the variables. The snapshots below present the initial Data appearance (Data View, left) and variable definitions (Variable View, right) for a series of variables entered from a text file. The variable **'CLASS_YEAR'** has been edited to show the changes that may be made through variable editing. Frequency Tables of the variable before and after definition are also presented. Steps to modify or define the variables are noted below.

Data prior to defining variables and associated frequency table of variable **v1**.



Defining variable **v1**, renamed **'CLASS_YEAR',** and associated frequency table.



v1

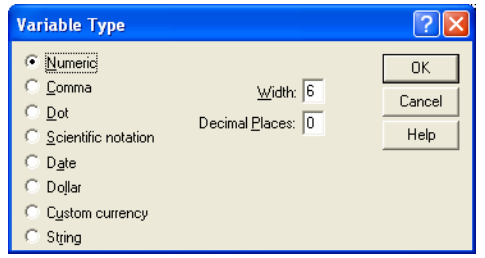
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	3118	22.4	22.7	22.7
	2	3255	23.4	23.6	46.3
	3	3729	26.8	27.1	73.4
	4	3140	22.5	22.8	96.2
	5	53	.4	.4	96.6
	6	412	3.0	3.0	99.6
	7	59	.4	.4	100.0
	Total	13766	98.8	100.0	
Missing	System	170	1.2		
Total		13936	100.0		

Class Year of Respondent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Freshman	3118	22.4	22.7	22.7
	Sophomore	3255	23.4	23.6	46.3
	Junior	3729	26.8	27.1	73.4
	Senior	3140	22.5	22.8	96.2
	Non-Degree	53	.4	.4	96.6
	Graduate	412	3.0	3.0	99.6
	Other	59	.4	.4	100.0
	Total	13766	98.8	100.0	
Missing	System	170	1.2		
Total		13936	100.0		

Defining Variables: There are ten characteristics of a variable and its presentation in the Data Editor that may be changed through the **Variable View**.

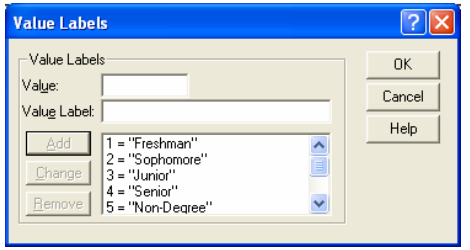
- **Name**: **Variable Names** may be up to 64 characters in length. They must begin with an alpha character. They may contain any digit, a period, @, #, _, or $. They may not contain a space, +, -, *, /, ‘, !, ? or other special characters. They may not end in a period nor should they end in an underscore (though this is legal to do). Each name must be unique. Should you violate any of these rules, SPSS will reject the variable name and request a correction.



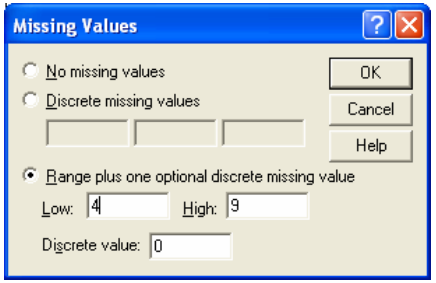
- **Type**: By clicking upon a variable’s type a small button will appear after the listed type. Select the button and a dialog box will open. To change the type of the variable, select an option. (Dialog box shown to the right.)

- **Width & Decimals**: Width and decimals are adjusted by clicking on the current value and using the increase/decrease toggle button that appears. **Width and Decimals** reflect the values set in the **Type** option (above & to right). The data entered will be used by SPSS for calculations, regardless of the settings for the length. So, for example, if the length is set at four characters, two decimals and the number 4.33734 is entered, that number will be used in calculations, but 4.34 will appear in the Data Editor.

- **Variable Label**: A **Variable Label** may be associated with a **Variable Name** simply by typing it into this cell. Assigning a Variable Label will allow for a more descriptive label to be substituted for the Variable Name in tables and charts. For example, above the phrase “Class Year of Respondents” replaces the Variable Name **‘CLASS_YEAR’**. Once created, anywhere the variable name would normally appear in data output, the phrase will appear in its place. This label can be up to 256 characters in length, but shorter is better for table appearance.



- **Value Labels**: This option allows **Value Labels** to be associated with data values (generally numeric data). These labels will then appear in place of data values in output. As an example, if the numeric value of “1” has been recorded to represent “male” respondents, the Value Label “Male” could be assigned to represent the “1” and then **ADD**ed to the list. Similarly, the string value “m” could be assigned the Value Label “Male.” This process is repeated until all possible response values have been assigned Value Labels. If an error of some sort occurs in a Value Label or a different label is desired, highlight the value “label,” select the **Change** button and the value is retrieved for editing. A value-label can also be **Removed**. This label can be up to 60 characters in length, but much shorter is better for table appearance.



- **Missing Values**: Cases for which a data point is truly missing will have a **period** (.) located in the data cell. The period represents the **System Missing** value. These cases are not included in any analysis of that variable. A frequency run on a variable will result in a table listing all recorded values for the variable, plus the number of cases **defaulted to System Missing. If there appear values that are outside the bounds for legitimate responses, perhaps a result of a** data entry error, or values that should be excluded for a specific type of statistical analysis, they may be removed from analysis by assigning their values to missing. The Missing Values option allows discrete values, ranges or a combination of both to be set to missing. These missing values are, however, not added to the System Missing count as they represent a different type of missing data - those that were selected out vs. non-existent. Once coded as missing, these values will appear as separate categories under the Missing Values portion of a frequency table.

- **Columns**: This option changes the width of the columns as seen in the **Data View** and thereby the number of variables visible on the screen. An increase/decrease toggle button appears when the cell is selected. Placing the mouse arrow on the line between variable names in the Data View, pressing the left button and dragging the column’s right border may also be used to adjust a column’s width. It does not change the defined width of the variable.

- **Align**: The data may be aligned left, right or center within the data cells by choosing this cell and selecting the preferred option from the drop down listing.

- **Measure**: Each variable may be identified as a Scale (ratio/interval), Nominal, or Ordinal measurement level variable through a drop down listing. In most cases it is not necessary to change this option. Its primary importance is for data analysis using certain advanced statistical procedures.

EDIT MENU

The options available on the **Edit Menu** vary depending upon which window is active (Data Editor, Output Viewer, Chart Editor, etc.), but generally include those noted below.

STEPS: Menu Bar => Edit => Undo, Redo, Cut, Copy, Paste, Clear, Find, Options (selection varies by window).

- **UNDO:** Undoes the **last** action taken.
- **REDO:** Returns last action.
- **CUT:** Removes selected data and places it on Clipboard.
- **COPY:** Copies selected data and places it on Clipboard.
- **CLEAR/DELETE:** Selected items are removed. **Clear** may be used to delete information from individual cells, or selected (highlighted) rows, columns, or groups of cells in the Data Editor.
- **FIND:** Used to find data values from within a selected variable. Select the variable by clicking once on the variable name or any data cell for the variable. Select **Find** from the **Edit Menu** and a dialog box will appear. In the dialog box, type the value for which you are searching.
- **PASTE:** Places the contents of Clipboard into another area of the current Data Editor, Output Viewer or an external document.
- **OPTIONS:** Used to change the default settings within SPSS.

EXPORTING TABLES AND CHARTS

Tables and charts created in SPSS can be copied from the **Output Viewer** in SPSS to a document created in Microsoft Word or another program. Steps shown here move tables and charts into Microsoft Word. Within SPSS the options are Copy and Copy Objects. From within Word the options are Paste and Paste Special

DATA: Employee data.sav

STEPS: In SPSS: Output Viewer select the table or chart to copy and follow one of the options below.

CHART TRANSFER:

- **SINGLE CHART: From the Output Viewer => Copy => In Word Paste Special => Bitmap.**
- **SINGLE CHART: From the Chart Editor => Edit => Copy Chart => In Word Paste.**
- **MULTIPLE CHARTS: Move individually, see Single Chart**

FREQUENCY TABLES:

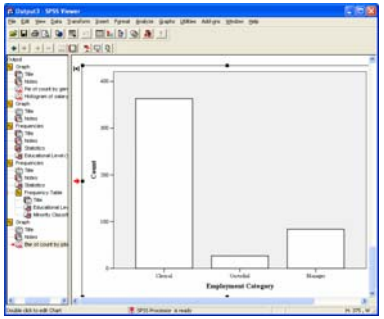
- **SINGLE FREQUENCY TABLE: From the Output Viewer => Copy => In Word Paste Special => Picture (Windows Metafile).**
- **SINGLE FREQUENCY TABLE: From the Output Viewer => Copy Objects=> In Word Paste.**
- **MULTIPLE FREQUENCY TABLES: From the Output Viewer => Copy Objects=> In Word Paste.**

SUMMARY OF STEPS: From the **Output Viewer** select the table or chart to be copied into the Word document. Selection is done by clicking once anywhere within the table or chart, or by clicking on the appropriate item under the SPSS **Output** list on the left side in the **Output Viewer**. =>Open the new or existing Word document into which you wish to copy the table/chart. In order to paste the table/chart into the Word Document, the Word document must be the active window. => In Word, select the location where the table/chart is to be pasted. => Select a Word paste method from among the above options. => Click on the table/chart and boxes will appear around its perimeter. Drag a square to adjust the size of the table/chart. => Double click on the object and the **Word Format Picture** dialog box will open. Select the **Layout Tab** and then select a location format. => The object's position may be adjusted by using the mouse or the arrow keys. **NOTE:** The default layout is In Line With Text - no text in line with the object .

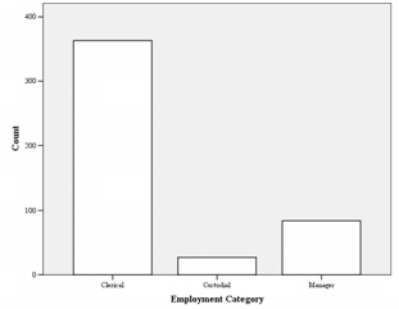
NOTE: If the above process does not provide the table/chart appearance desired try other combinations of SPSS Copy/Copy Object and Word Paste/Paste Special.

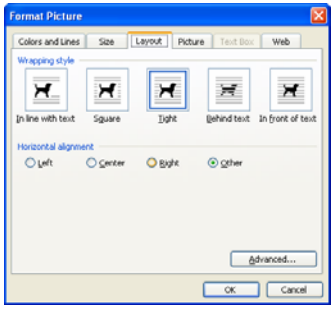
EXAMPLE: Insert a bar chart into Word.

SPSS: Select table/chart (Note the arrow pointing to the selected chart.) => Copy.



Word: Select Edit, Paste Special, Bitmap. Reduce the chart size by clicking on the Chart and dragging a corner marker toward the center. To adjust the location double click on the chart/table and set the Layout option.





FILE & VARIABLE INFORMATION

Quite often you may want to examine the characteristics of one or more variables contained within a data set. In addition to reviewing variable information on the Variable View of the Data Editor, it may be obtained through two utilities: **Variables** and **Data File Information.** The latter provides a printable summary in the Output Viewer and incoudes separate tables of Variable Information and Variable Values.

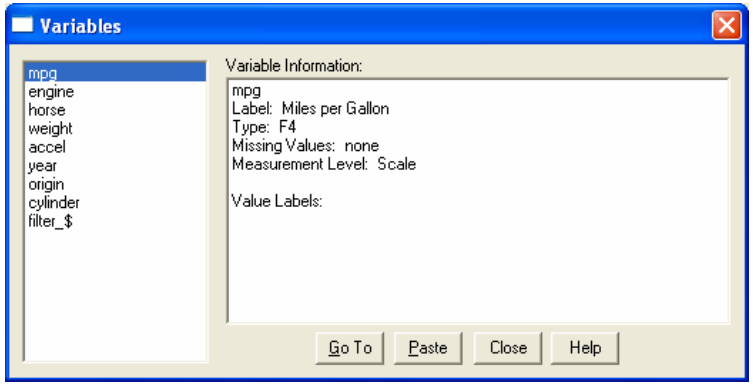
DATA: Cars.sav

VARIABLES:

STEPS: Menu Bar => Utilities => Variables.

EXAMPLE: Use **Variables** to examine the formatting of the variable **'MPG'.**

SUMMARY OF STEPS: Go to the **Utilities Menu.** => Select **Variables.** => In the dialog box click on the variable of interest. => The data characteristics that have been defined will be listed in the right box.



DATA FILE INFORMATION:

STEPS: Menu Bar => File => Display Data File Information => Working File.

EXAMPLE: Use **Data File Information** to obtain a listing of all of the data set's variables, their formatting and values.

SUMMARY OF STEPS: Go to the **File Menu.** => Select **Display Data File Information.** => Select **Working File.** A complete listing of all variables and their formatting will be output to the Output Viewer. A second table contains the response values assigned to variables.

Partial Listing obtained from Data File Information:

Variable Information

Variable	Position	Label	Measurement Level	Column Width	Alignment	Print Format	Write Format
mpg	1	Miles per Gallon	Scale	8	Right	F4	F4
engine	2	Engine Displacement (cu. inches)	Scale	8	Right	F5	F5
weight	3	Vehicle Weight (lbs.)	Scale	8	Right	F4	F4
origin	4	Country of Origin	Nominal	8	Right	F1	F1
cylinder	5	Number of Cylinders	Ordinal	8	Right	F1	F1

Variables in the working file

Variable Values

Value		Label
origin	1	American
	2	European
	3	Japanese
cylinder	3	3 Cylinders
	4	4 Cylinders
	5	5 Cylinders
	6	6 Cylinders
	8	8 Cylinders

INSERT VARIABLE OR CASE

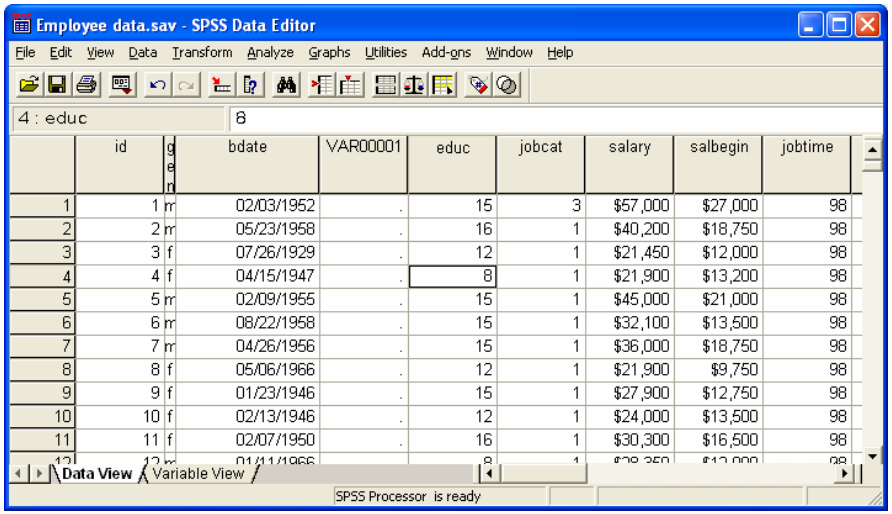
The insert options allow the insertion of a Variable between two existing variables in the Data Editor worksheet or a Case between existing cases. (See also **SORTING VARIABLES AND CASES.)**

DATA: Employee data.sav

INSERT VARIABLE:

STEPS: Menu Bar => Data => Insert Variable

To **insert a variable** locate the area where you want to insert the variable and click on the variable name that is ***to the right*** of where you want the new variable to be. From the **Data Menu**, select **Insert Variable**. In the example (right) the new variable was inserted to the right of the variable **'EDUC'**. Define the variable by selecting the **Variable View** tab and entering appropriate information. (see also **SORTING VARIABLES AND CASES**.)

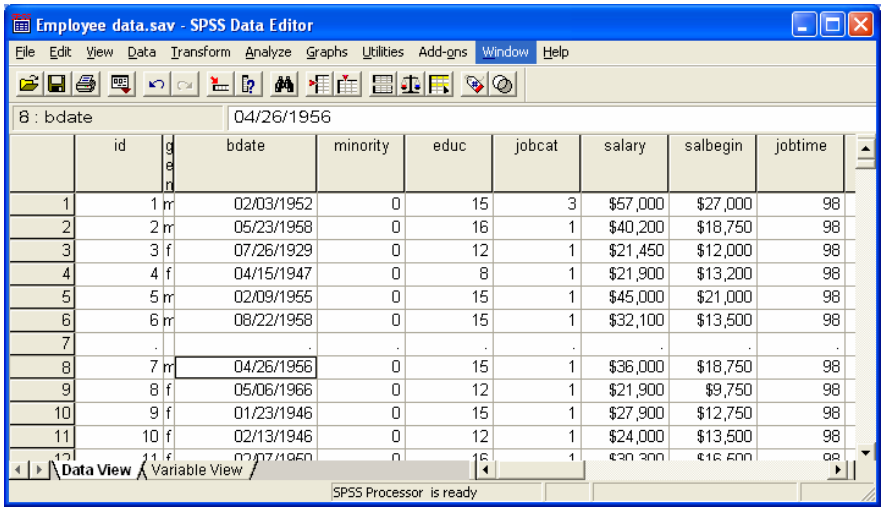


INSERT CASE:

STEPS: Menu Bar => Data => Insert Case

To **insert a case** locate the area where you want to insert the case and click on the case that is ***below*** where you want the new case to be. From the **Data Menu**, select **Insert Case**. In the example (right) the cursor was set on one of the cells in case 7. As the new case was entered case 7 moved down (to case 8). (see also **SORTING VARIABLES AND CASES**.)

.



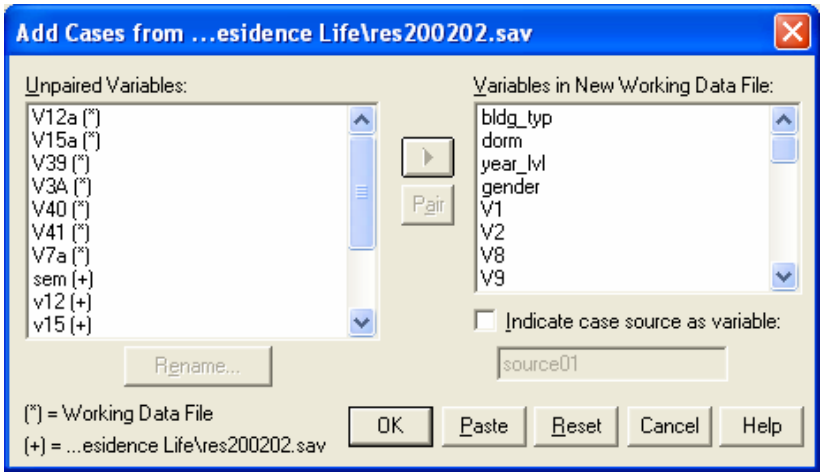
MERGE FILES

The **Merge Files** command allows one to add additional cases or variables to an SPSS data file.

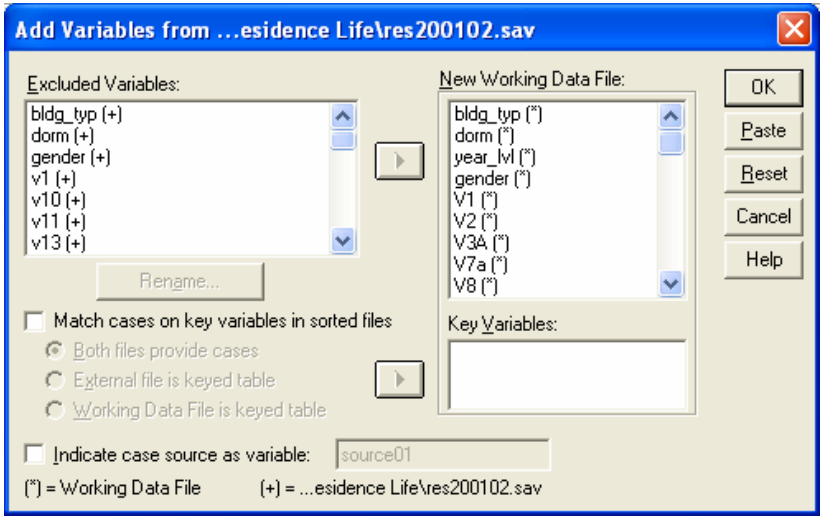
STEPS: Menu Bar => Data => Merge Files => Add Cases or Add Variables

Open the first data file in the Data Editor. From the Main Menu select Data, then Merge Files. The next choice depends upon the type data to be merged.

Add Cases: This option adds more cases to the file. Initially it assumes that the cases to be added have variables in common with the working file. Where a variable match occurs, the variable will appear in the "Variables in the New Working File" cell of the Add Cases Dialog Box. Where a match does not occur, the variables are left in the "Unpaired Variables" cell. They can be added by highlighting them and moving them with the arrow. The default missing value will be assigned to cases for which there is no data relative to a variable which was moved by this latter process.



Add Variables: This option adds more variables to the file. It assumes that the cases to be added have an identifying variable so that the data for cases in the working data set can be linked to data for that case in the data set being added (e.g. a common ID number). The merged data set will contain as many cases as the larger of the two initial data sets. Cases for which there was no information in one of the data sets will have the system missing value inserted into the missing variables.



Notes:

1) Variables followed by an (*) are from the working data set (first one entered into the Data Editor).
2) Variables followed by an (+) are from the data set being added.
3) On occasion a variable may be named differently in files to be merged. They can be merged under the name of the initial file by holding the CTRL key and highlighting the "common" variables. Once selected, the Pair option below the transfer arrow will highlight. Selecting this button will transfer the paired variables into the new data set.
4) The new data set will assume the variable definitions present in the working data set. So, for example, if there are no value labels in the initial data set the added cases, which may have value labels, will default to no value labels.
5) String variables that are to be merged need to have the same length defined.

OUTPUT VIEWER

All statistics, graphs, and tables are output to the Output Viewer. From this window charts and tables may be edited and output may be saved and/or printed.

STEPS: The **Output Viewer** Window opens automatically when a statistical analysis is performed or a graph is created. It can also be accessed by opening a saved output file (i.e. output saved from the Output Viewer).

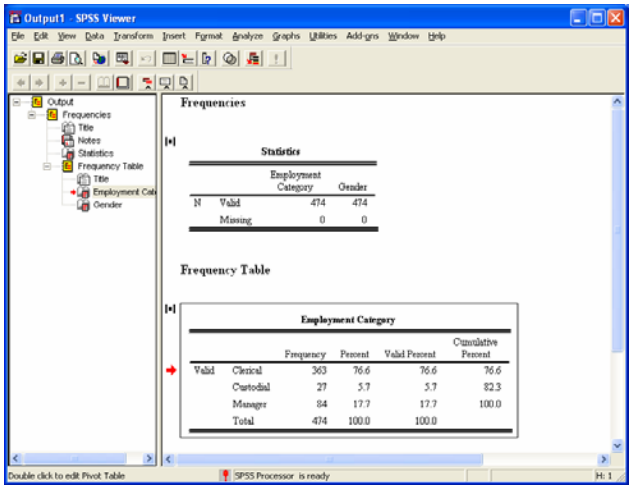
The Output Viewer is partitioned into two sections or windows. The right section contains the results of the statistical procedures or graphing. It is from this portion of the Output Viewer that the Pivot Table Editor and Chart Editor are opened. The left portion of the Output Viewer works as an index to the data analyses conducted. Individual items or groups of items of can be deleted, printed or moved in the listing. They may be "opened" and "closed" by double clicking on the "notebook" icons. The **Red Arrow**, appearing in both windows, indicates what section of the output is being shown. A table/chart will neither appear in the right window nor can it be printed if its index "notebook" is closed

To print or delete a single item, select it either via the index or by clicking on the table/chart in the right window and perform the desired action. To print or delete an entire statistical run (e.g. Title, Notes, Statistics, Table, Chart, etc.) click on the "Yellow notebook" and perform the desired action. If you want to perform an action on all runs present within the Output Viewer, click on the topmost notebook, entitled "Output."

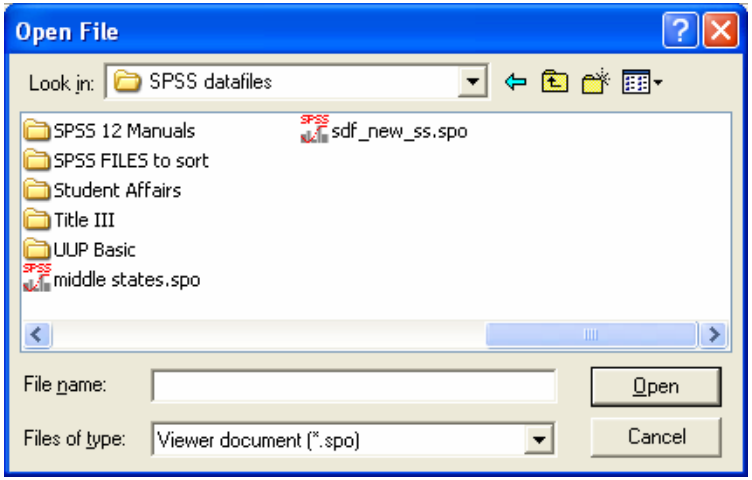
In general a statistical run will include a Title, Notes, Statistics, Table, and/or Chart. Depending upon the default values set for the system some of these items may be closed upon creation. As an example, the Notes table is generally closed as it contains information about the data file - date created, etc. In contrast, the Statistics table, while not something normally printed, would be open as it contains information about the number of valid cases used in the analysis. It would also include descriptive statistics, such as the data's mean and standard deviation, if these statistics had been requested.

Data output is saved in files appended with the extension **".spo"** and are preceded in a file listing with an icon that looks like a bar chart. An existing output file may be opened at any time by selecting **File => Open => Output**.

Output Viewer:



SPSS output is saved with the extension "**.spo**"

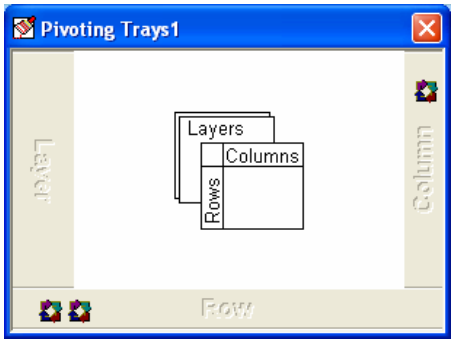


PIVOT TABLES

Frequency tables and crosstabs tables may have the orientation of variables and data values changed through the use of the **Pivot** command.

DATA: Employee data.sav

STEPS: Create a table: Analyze => Descriptive Statistics => Crosstabs. Double click on the table in the Output Viewer (thus putting the table in the Pivot Table Editor) => Select PIVOT from the Main Menu => Pivoting Tray



PivotingTray: Within a Pivoting Tray dialog box will appear a series of multicolored icons (arrows forming a square). Placing the mouse cursor on one of these and pressing the left button will highlight the portion of the table referenced. The icon may be dragged to the Layers, Rows, or Columns section of the Pivoting Tray. Once moved, release of the mouse button will cause the table to be reformatted in the new orientation. A variable icon placed in the Layer section will create individual tables for each category (value) within the variable. A pair of horizontal arrows will appear next to the multicolored icon and are used to move through the Layers of the table. Within the table there will be a dropdown box, containing a listing of the values, through which a layer may be selected.

EXAMPLE: Obtain a Crosstabs of **'JOBCAT'** by **'GENDER.'** Use the Pivot Tray to re-orient the table so that related percentages are together (reverse the order of the two icons in the "Row" section of the Pivoting Tray).

Initial format of a crosstabs table:

Employment Category * Gender Crosstabulation

			Gender		
			Female	Male	Total
Employment Category	Clerical	Count	206	157	363
		% within Employment Category	56.7%	43.3%	100.0%
		% within Gender	95.4%	60.9%	76.6%
		% of Total	43.5%	33.1%	76.6%
	Custodial	Count	0	27	27
		% within Employment Category	.0%	100.0%	100.0%
		% within Gender	.0%	10.5%	5.7%
		% of Total	.0%	5.7%	5.7%
	Manager	Count	10	74	84
		% within Employment Category	11.9%	88.1%	100.0%
		% within Gender	4.6%	28.7%	17.7%
		% of Total	2.1%	15.6%	17.7%
Total		Count	216	258	474
		% within Employment Category	45.6%	54.4%	100.0%
		% within Gender	100.0%	100.0%	100.0%
		% of Total	45.6%	54.4%	100.0%

Format of crosstabs table after using the Pivot feature:

Employment Category * Gender Crosstabulation

			Gender		
			Female	Male	Total
Count	Employment Category	Clerical	206	157	363
		Custodial	0	27	27
		Manager	10	74	84
	Total		216	258	474
% within Employment Category	Employment Category	Clerical	56.7%	43.3%	100.0%
		Custodial	.0%	100.0%	100.0%
		Manager	11.9%	88.1%	100.0%
	Total		45.6%	54.4%	100.0%
% within Gender	Employment Category	Clerical	95.4%	60.9%	76.6%
		Custodial	.0%	10.5%	5.7%
		Manager	4.6%	28.7%	17.7%
	Total		100.0%	100.0%	100.0%
% of Total	Employment Category	Clerical	43.5%	33.1%	76.6%
		Custodial	.0%	5.7%	5.7%
		Manager	2.1%	15.6%	17.7%
	Total		45.6%	54.4%	100.0%

RECODE

The **Recode** function is used to change the raw data values associated with a variable. Most notably, recode is used to reduce the number of value options, thereby consolidating the data into fewer categories. Recode is also used to assign new values in place of existing ones.

DATA: Employee data.sav

STEPS: Menu Bar => Transform => Recode => select a Recode option

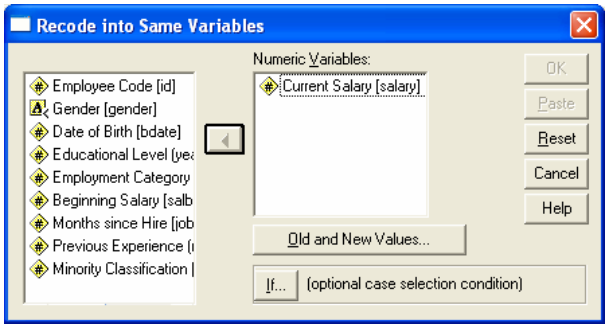
OPTIONS:

- **Into Same Variables**: This option replaces the raw data values of a variable with new values. Once changed, the original data values are lost from the active session. Once the active session is saved, the original data values cannot be retrieved unless the session is saved with a new name or the new values have a one-to-one relationship with the original variable values. Take great care when selecting this option. Given the virtually unlimited number of variables that a SPSS data file may have, there is little need to use this option (see example in final table).

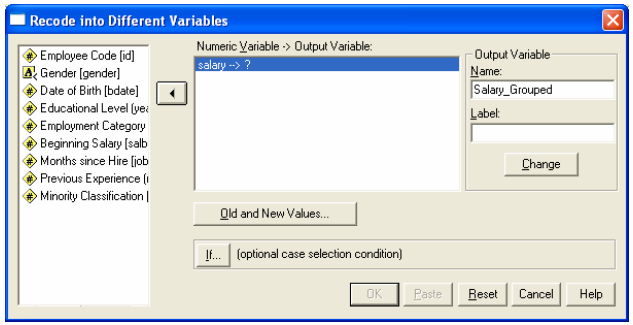
- **Into Different Variables:** This option creates a new variable in the data set, thereby retaining both the original variable and the recoded variable (see example in final table).

SUMMARY OF STEPS: Menu Bar => Transform => Recode => Into Different Variables (or Into Same Variables). => Highlight variable. => Press arrow button. => Enter Output Variable name (for a new variable). => Press **Change** button. => Select **Old and New Values**. => Enter Old and New Values. => Press **Add** button. => Press **Continue** button. => If all is correct the **OK** button will be activated. => **OK**. **NOTE**: new variables will appear **at the end** of the variable listing.

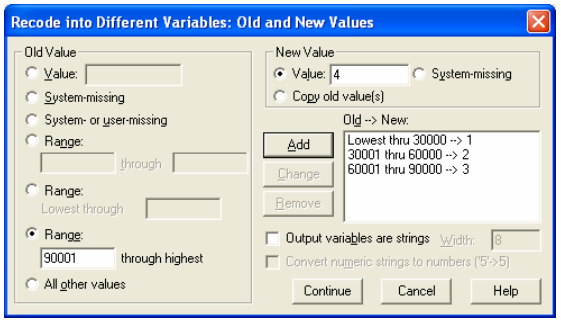
Recode Into Same Variables Dialog Box



Recode Into Different Variables Dialog Box

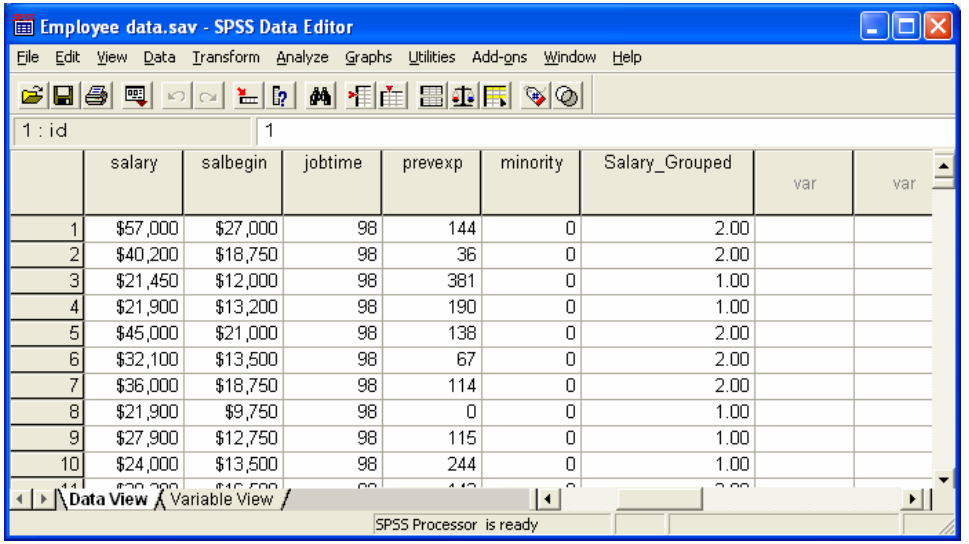


Old and New Values Dialog Box



RECODE EXAMPLE: Recode into a different variable. There are three steps to creating a new variable; all must be completed before the **OK** button will activate. First, after identifying the variable to recode and moving it into the **Numeric Variable -> Output Variable** cell, an **Output Variable** must be identified. In the case of recoding with the Into Different Variable option, the initial variable will look like "S**alary->?**" in the Numeric Variable -> Output Variable cell. Second, once the new variable name, for example "Salary2," is entered in the **Output Variable** cell, the **Change** button must be pressed. The variable statement will be changed to "**Salary->Salary2.** " Third, the recoding plan must be recorded in the **Old and New Values** dialog box (discussed below). Once all of these steps have been completed the **OK** button will be active and he recode may take place. Selecting **OK** and will return SPSS to the **Data Editor**. **The new variable will appear at the end of the variable listing if you selected the Into New Variables option.**

SPSS Output (Right) : New variable in the Data Editor. (Input dialog boxes shown on preceding page.)



	salary	salbegin	jobtime	prevexp	minority	Salary_Grouped	var	var
1	$57,000	$27,000	98	144	0	2.00		
2	$40,200	$18,750	98	36	0	2.00		
3	$21,450	$12,000	98	381	0	1.00		
4	$21,900	$13,200	98	190	0	1.00		
5	$45,000	$21,000	98	138	0	2.00		
6	$32,100	$13,500	98	67	0	2.00		
7	$36,000	$18,750	98	114	0	2.00		
8	$21,900	$9,750	98	0	0	1.00		
9	$27,900	$12,750	98	115	0	1.00		
10	$24,000	$13,500	98	244	0	1.00		

OTHER RECODE CONSIDERATIONS:

- **IF:** The **If** option allows cases containing specific values to be selected. For example, you could use an **If** statement to recode only the values of male employees.

- **OLD AND NEW VALUES:** The **Old and New Values** process will allow you to associate the original data value with a new value. This process can be achieved in a number of ways, which range from creating new values on a one at a time basis to using various range combinations. As an example, the value of "1" could be recoded to a value of "4" by using the following steps. Enter 1 in the **Old Value** cell, then go to the **New Value** cell and enter 4. Once the new value is entered, the **Add** button will darken. Click on it and the recode will appear in the **Old->New** cell. If you wish to edit or delete a recode, click on the recode to be changed in the **Old->New** box and the **Change** and **Remove** buttons will be activated (darken). Once all recodes have been identified, press the **Continue** button and system will go back to the initial **Recode** dialog box.

- **SAVE:** Saving the data file will replace the original data set with a data set containing: a) the new values replacing the original values if **Into Same Variable** is selected; b) the original variable with original values and new variables with the new values if **Into Different Variable** is selected.

- **Recode into the Same Variable vs. a Different Variable:** What is the result?

Five Cases Recoded into Same Variable (cannot reverse recode)

Recode: 1 = 1; 3 = 1; 5 = 5; 7 = 5

(Before) CLASS	*(After)* CLASS
1	1
1	1
3	1
5	5
7	5

(many to one relationship)

Five Cases Recoded into Same Variable (can reverse recode)

Recode: 1 = 1; 3 = 2; 5 = 3; 7 = 4

(Before) CLASS	*(After)* CLASS
1	1
1	1
3	2
5	3
7	4

(one-to-one value relationship)

Five Cases Recoded into Different Variable (saves both value layouts)

Recode: 1 = 1; 3 = 1; 5 = 5; 7 = 5

(Before) CLASS	*(After)* CLASS	CLASS2
1	1	1
1	1	1
3	3	1
5	5	5
7	7	5

SELECTING CASES

Cases in a worksheet may be selected according to specified criteria.

DATA: Employee data.sav

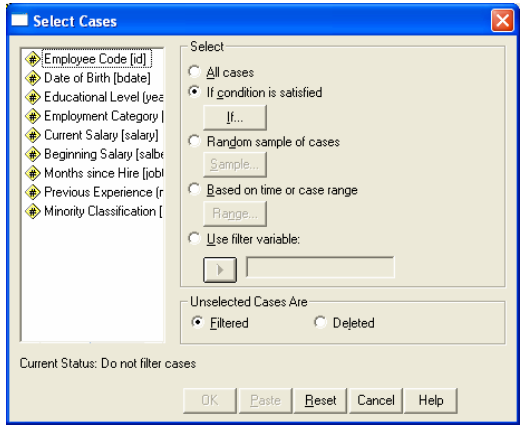
STEPS: Data Editor => Data => Select Cases

EXAMPLE: Select all cases in the worksheet with salaries less than $50,000.

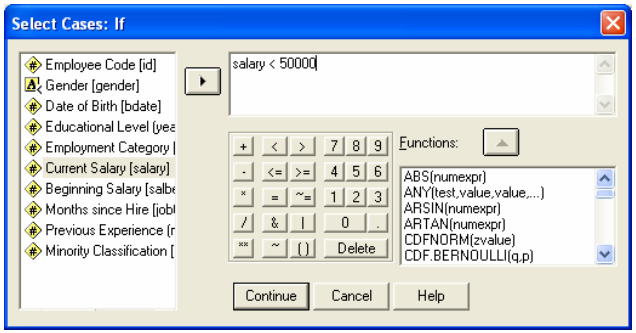
SUMMARY OF STEPS: In the **Select Cases** dialog box click on the button to the left of **If condition is satisfied** and then click on the **If** box. => In the **Select Cases: If** dialog box, you will specify the condition for selection of cases. What is used in this dialog box will vary depending on the situation. In this example the variable '**Salary**' was selected by clicking once on the variable, and then clicking on the right arrow. => The '<' (less than) symbol was then selected by clicking once on it in the keypad section of the dialog box and 50,000 was then typed in. => Click **Continue.** => Click **Ok**. => In the Data Editor cases that **do not meet** the stated criteria, here salaries greater than $50,000, will have a diagonal line drawn through the case number.

NOTES: 1) The defined **Select Cases** statement will remain in effect until the dialog box is re-opened and the **All Cases** option is selected. **2)** Leave the **Unselected Cases Are** option on **Filtered** (default). Doing so will help prevent saving a data file that has had cases removed.

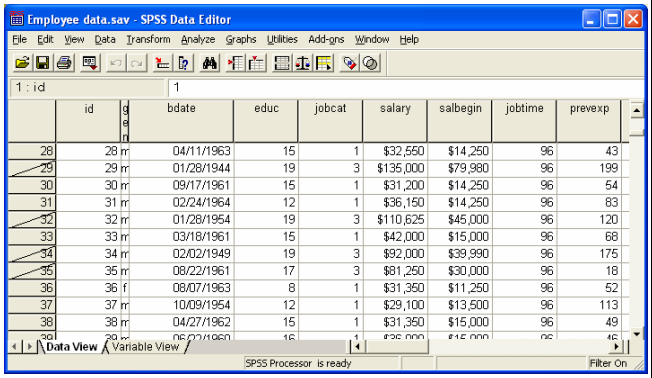
Dialog Box: "Filtered" and **"If"** are selected.



Dialog Box: Condition is specified.



Data Editor showing selected and non-selected cases.



SORTING CASES & VARIABLES

The **"Sort Cases"** command can sort cases in the worksheet, based on one or more sorting variables.

DATA: Employee data.sav

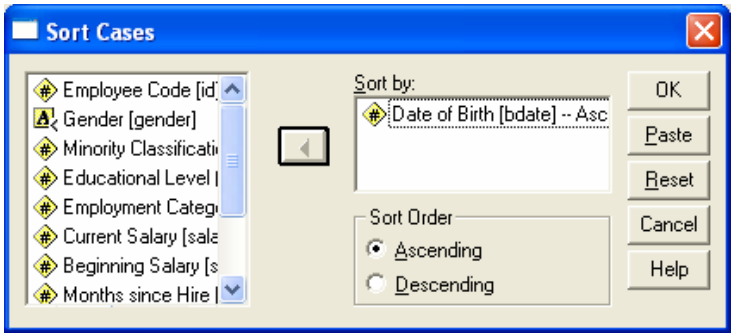
SORT CASES:

STEPS: Menu Bar => Data => Sort Cases

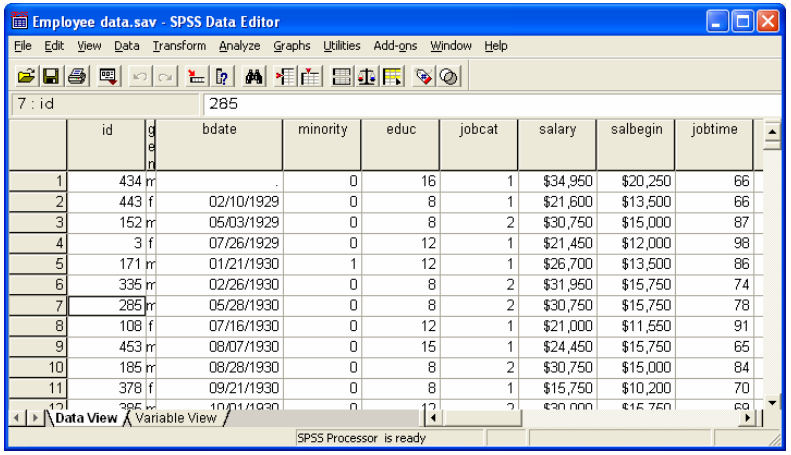
EXAMPLE: Sort cases in ascending order, by the variable **'BDATE'.**

SUMMARY OF STEPS: In the **"Sort Cases"** dialog box, select the variable **'BDATE'** by clicking once on it and then clicking on the right arrow next to the **"Sort By"** box. => Select **Áscending** or **Descending** in the **Sort Order** section of the dialog box. => Click **OK** => Cases are now sorted in the order selected according to birth date.

Dialog Box: Sorting variable and sort order are selected.



Dialog Box: Cases are sorted accordingly.



SORT VARIABLES:

Rearranging the order of variables may be useful after a new variable has been created or added to the data file. **Moving variables is accomplished one variable at a time.** To change the order of variables, use the **Insert Variable** option to create a new variable where you want to relocate a variable (see **INSERT VARIABLE OR CASE**). Select the variable to be moved by clicking once on the **variable name at the top of the column**. From the E**dit Menu**, select **cut** (or right mouse button). This will remove the variable from the Data Editor and place it on the Clipboard. Now click on the new variable name (probably something like **var00001**). From the **Edit Menu**, select **Paste** (or right mouse button) and the cut variable will be placed in this new variable location.

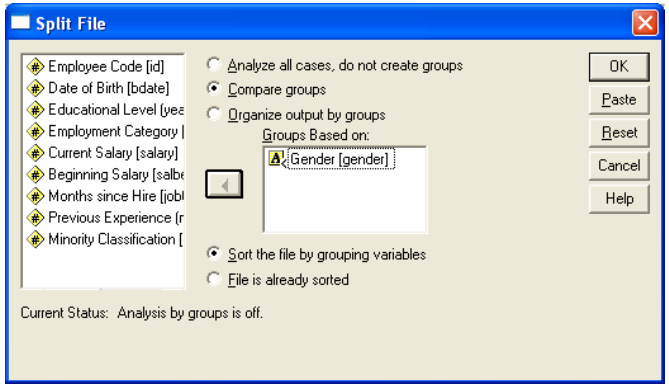
SPLIT A DATA FILE

The **Split File** command splits a worksheet into specified groups based on the values of a selected variable. Data analysis output is based upon the specified grouping.

DATA: Employee.sav

STEPS: Menu Bar => Data => Split File

EXAMPLE: Split the Employee.sav file by the variable **'GENDER'.**

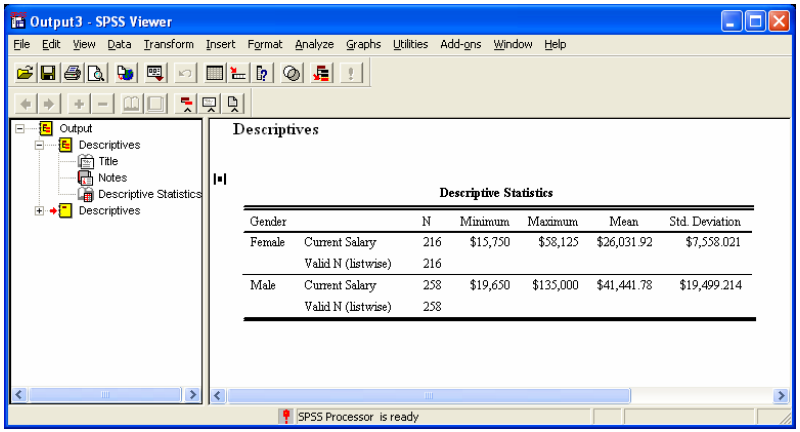


SUMMARY OF STEPS: In the **Split File** dialog box, you may either select **Compare groups** or **Organize output by groups**, by clicking on the button to the left of your choice. => Select your grouping variable, by clicking once on it, and then clicking on the right arrow next to the **Groups Based On:** box. => If the file has not already been sorted, select **Sort the file by grouping variables** by clicking on the button to the left of it. => Click **OK.**

NOTES: 1) The effect of this command is not apparent until data analysis occurs. **2)** This command remains in effect until it is "turned off" by selecting the **Analyze all cases** option. **3)** When **Compare groups** is selected, any analyses will be presented together for comparison. When **Organize output by groups** is selected, analyses will be presented separately for each group.

EXAMPLES: Split the file by the variable **'GENDER'** and obtain descriptive statistics for the variable **'SALARY'**. Output obtained via **Descriptives**.

SPSS Output resulting from selecting **Compare groups**:



SPSS Output resulting from selecting **Organize output by groups**:

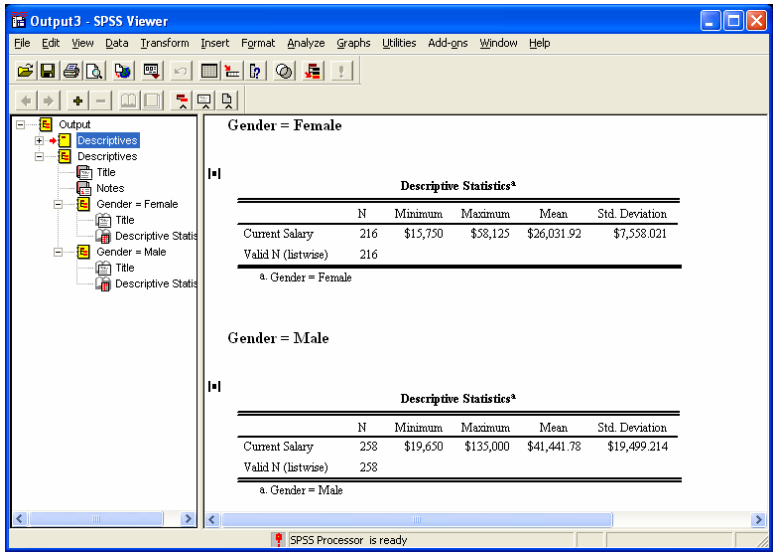
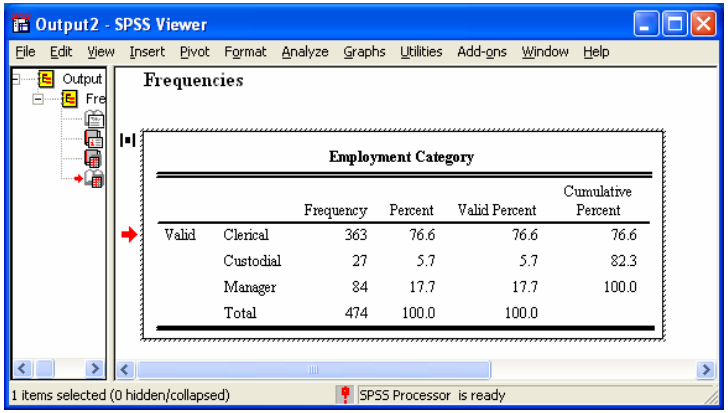


TABLE EDITOR

Tables may be modified through use of the **Table Editor**. When in the **Output Viewer**, Place the mouse cursor on the table to be modified and double click. An outline border will appear around the table. Once in the **Table Editor** any cell may be edited by double clicking on it. (see also **PIVOT TABLES**.) To leave the editor click outside the border.

DATA: Employee data.sav

STEPS: **Output Viewer => Double click on the table to be edited => The table being edited will be surrounded by a highlighted rectangle (shown to the right).**



OPTIONS:

- **TableLooks:** When a table is created its appearance in the Output Viewer is pre-determined by a default setting (Edit => Options =>Table). Once created, a table's overall appearance may be changed using the **TableLooks** command. Choose **Format** from the Menu Bar, then **TableLooks**. Selected from among the variations available or create a format to your liking, based upon one of the existing formats by using the Edit Looks button.
- **Table Properties:** From the **Format** option of the Menu Bar, select **Table Properties**. This editing feature allows one to modify the appearance of cells, borders and footnotes by choosing colors, fonts, line thickness, etc.
- **Cell Properties:** Allows for the adjustment of a selected cell's (or selected cells') alignment, shading, margins, etc. From the Menu Bar select **Format** then **Cell Properties. Note: Changing a cell's Font is done with the Format => Font option**
- **Caption:** A Caption for a table may be added at its end by selecting **Insert => Caption**. To enter information into this cell, double click upon it. To delete a caption, highlight the Caption cell and press Delete.
- **Footnote:** Any cell in the table may be footnoted. Highlight the cell to be referenced and then select **Insert => Footnote**. The Cell will contain a footnote marker and a new cell will be created at the bottom of the table. Double click on the Footnote cell and enter the desired information. To delete a footnote, highlight the Footnote cell and press Delete.

EXAMPLE: Modify the Job Category table by changing the following: TableLook to Modern; total valid values to Regular, bold font style with size 14; shading the Valid Totals row by 10%; adding a footnote to the table title; and adding a caption. When all changes have been made, click anywhere outside of the border.

Initial Table Format (system default):

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	

Table after above modifications:

Employment Category[a]

		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**	

Data Collected Spring 1995

a. Excludes employees at Boise Plant

TEXT FILES

Data stored in a text format may be entered into SPSS through a "**Text Import Wizard**."

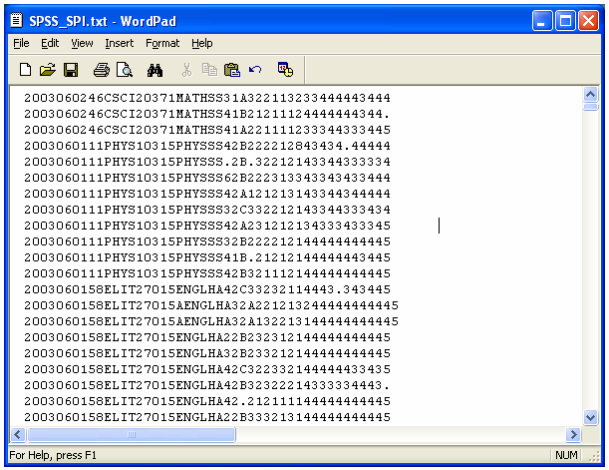
STEPS: File => Read Text Data => identify the text file and select "Open" => follow through the 6 text import steps.

TEXT IMPORT:

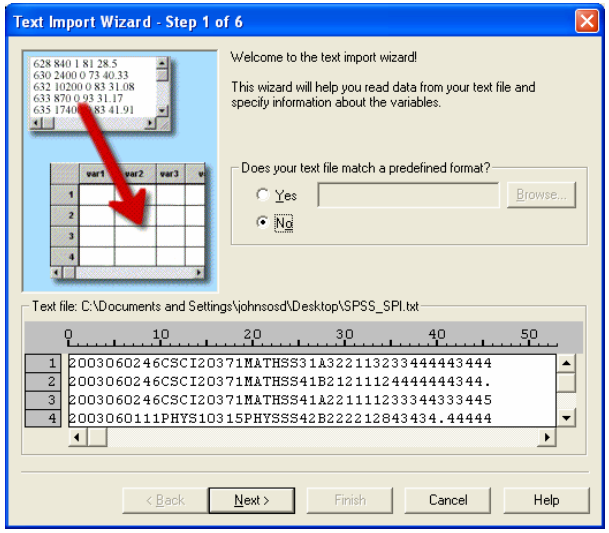
- **Step 1: Select "NO."** The yes option is for instances when a similar data file has been opened and the format established in the Text Import Wizard has been saved.
- **Step 2:** Delimited vs. Fixed Format and Variable Name present – both can be determined by looking at the raw data file.
- **Step 3:** Identify where the data begin in the file and how many lines (Records) there are per Case.
- **Step 4:** Insert breaks between the variables.
- **Step 5:** Identify variable names and variable types (both optional).
- **Step 6:** Finish. (Note: Here is where a text data format can be saved for future use.)

EXAMPLE: Open the text data file **SPSS_SPI.txt**.

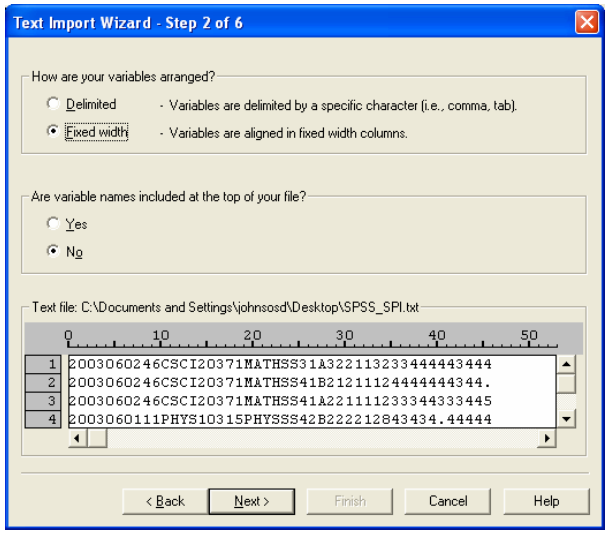
- **Text Data:** (Partial listing)



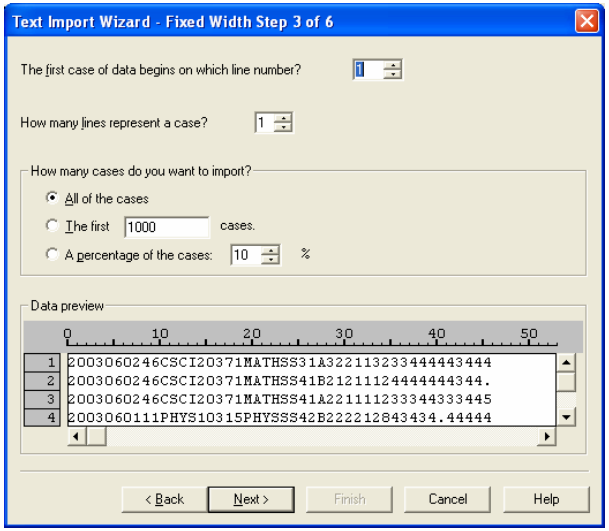
- **Step1:** The Text Wizard shows a small sample or the text data. Generally select "No" for the predefined format question.



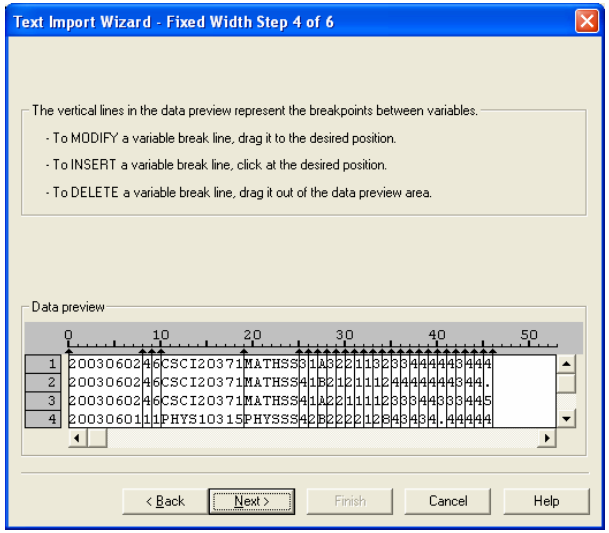
- **Step 2:** Select Delimited vs. Fixed Format and if variable names are included in the file. (Steps 3 & 4 will be slightly different if Delimited Format is selected.)



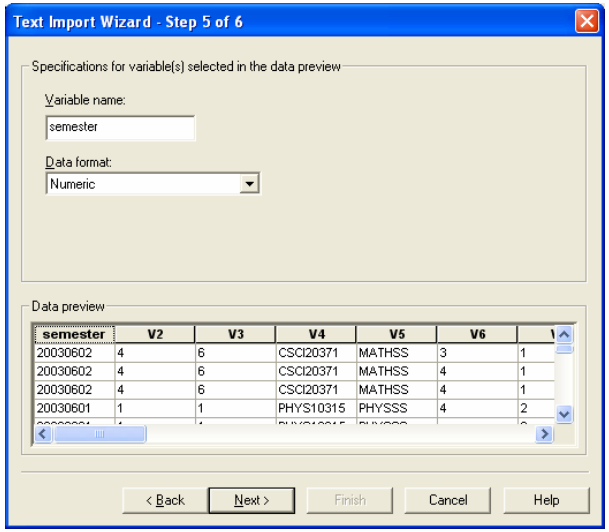
- **Step3:** Identify the number of data lines (Records) per case and how much of the data to be imported.



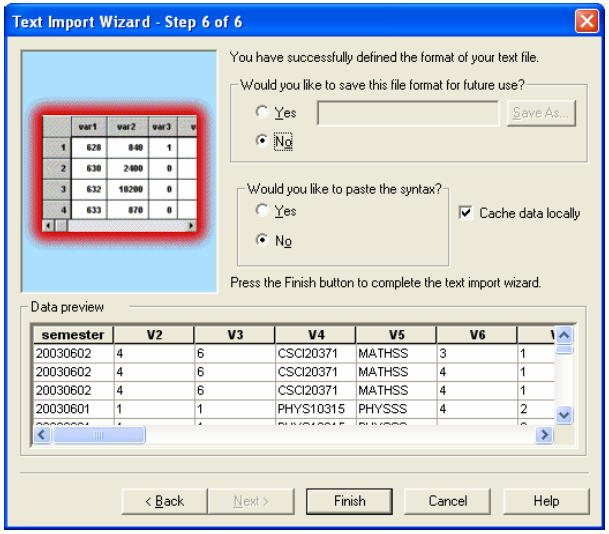
- **Step 4:** For Fixed Format data, as here, insert dividers between variables.



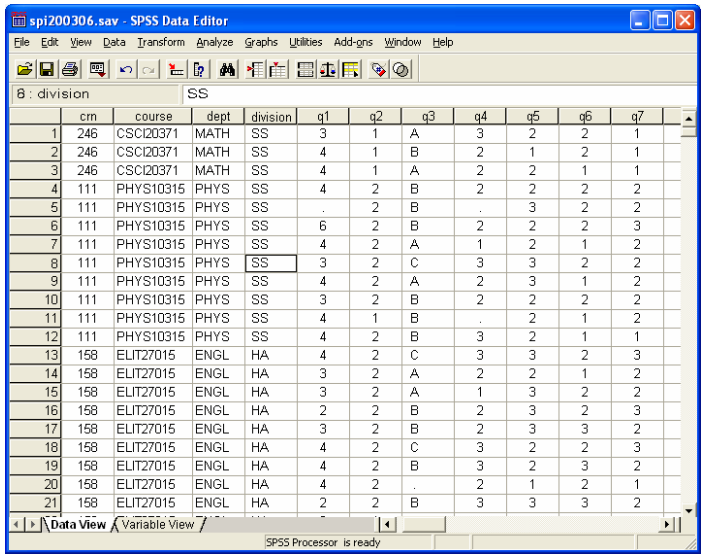
- **Step 5:** Assign variable names and types – both optional as they may be added in the Data Editor.



- **Step 6:** Finish. (Here the format created via the Text Wizard may be saved for future use.)



- **SPSS Data Editor after data import:**



APPENDIX A: Data File Information

In addition to the **Variable View** of the **Data Editor**, information about the coding of variables may be obtained in two other formats.

STEPS For All variables (output to Output Viewer; presented here): Menu Bar => File => Display Data File Information => Working File

STEPS for a Single Variable (output to monitor via dialog box): Menu Bar => Utilities => Variables

CARS.SAV

Variable Information: Cars.sav

Variable	Position	Label	Measurement Level	Column Width	Alignment	Print Format	Write Format	Missing Values
mpg	1	Miles per Gallon	Scale	8	Right	F4	F4	
engine	2	Engine Displacement (cu. inches)	Scale	8	Right	F5	F5	
horse	3	Horsepower	Scale	8	Right	F5	F5	
weight	4	Vehicle Weight (lbs.)	Scale	8	Right	F4	F4	
accel	5	Time to Accelerate from 0 to 60 mph (sec)	Scale	8	Right	F4	F4	
year	6	Model Year (modulo 100)	Ordinal	8	Right	F2	F2	0
origin	7	Country of Origin	Nominal	8	Right	F1	F1	
cylinder	8	Number of Cylinders	Ordinal	8	Right	F1	F1	
filter_$	9	cylrec = 1 | cylrec = 2 (FILTER)	Ordinal	8	Right	F1	F1	

Variables in the working file

Variable Values: Cars.sav

Value		Label
year	0[a]	0 (Missing)
	70	70
	71	71
	72	72
	73	73
	74	74
	75	75
	76	76
	77	77
	78	78
	79	79
	80	80
	81	81
	82	82
origin	1	American
	2	European
	3	Japanese
cylinder	3	3 Cylinders
	4	4 Cylinders
	5	5 Cylinders
	6	6 Cylinders
	8	8 Cylinders
filter_$	0	Not Selected
	1	Selected

a. Missing value

EMPLOYEE DATA.SAV

Variable Information: Employee Data.sav

Variable	Position	Label	Measurement Level	Column Width	Alignment	Print Format	Write Format	Missing Values
id	1	Employee Code	Scale	8	Right	F4	F4	
gender	2	Gender	Nominal	1	Left	A1	A1	
bdate	3	Date of Birth	Scale	13	Right	ADATE10	ADATE10	
educ	4	Educational Level (years)	Ordinal	8	Right	F2	F2	0
jobcat	5	Employment Category	Ordinal	8	Right	F1	F1	0
salary	6	Current Salary	Scale	8	Right	DOLLAR8	DOLLAR8	$0
salbegin	7	Beginning Salary	Scale	8	Right	DOLLAR8	DOLLAR8	$0
jobtime	8	Months since Hire	Scale	8	Right	F2	F2	0
prevexp	9	Previous Experience (months)	Scale	8	Right	F6	F6	
minority	10	Minority Classification	Ordinal	8	Right	F1	F1	9

Variables in the working file

Variable Values: Employee Data.sav

Value		Label
gender	f	Female
	m	Male
educ	0[a]	0 (Missing)
	8	8
	12	12
	14	14
	15	15
	16	16
	17	17
	18	18
	19	19
	20	20
	21	21
jobcat	0[a]	0 (Missing)
	1	Clerical
	2	Custodial
	3	Manager
salary	$0[a]	missing
salbegin	$0[a]	missing
jobtime	0[a]	missing
prevexp	0	missing
minority	0	No
	1	Yes
	9[a]	9 (Missing)

a. Missing value

ROAD CONSTRUCTION BIDS.SAV

Variable Information: Road Construction Bids.sav

Variable	Position	Label	Measurement Level	Column Width	Alignment	Print Format	Write Format
obs	1	Observation	Scale	8	Right	F3	F3
cost	2	Contract Cost	Scale	8	Right	F8.2	F8.2
dotest	3	DOT Engineer's Estimate of Construction Cost	Scale	8	Right	F8.2	F8.2
b2b1rat	4	Ratio of Second Lowest Bid to Lowest Bid	Scale	8	Right	F7.5	F7.5
b3b1rat	5	Ratio of Third Lowest Bid to Lowest Bid	Scale	8	Right	F7.5	F7.5
bhb1rat	6	Ratio of Highest Bid to Lowest Bid	Scale	8	Right	F7.5	F7.5
status	7	Contract Status	Ordinal	8	Right	F1	F1
district	8	District	Ordinal	8	Right	F1	F1
btpratio	9	Ratio of Number of Bidders to Number of Planholders	Scale	8	Right	F7.5	F7.5
daysest	10	Engineer's Estimate of Number of Work Days Required	Scale	8	Right	F3	F3

Variables in the working file

Variable ValuesRoad Construction Bids.sav

Value		Label
status	0	Competitive Contract
	1	Fixed Contract
district	0	Other District
	1	South Florida District

WORLD95.SAV

Variable Information: World95.sav

Variable	Position	Label	Measurement Level	Column Width	Alignment	Print Format	Write Format	Missing Values
country	1	<none>	Nominal	12	Left	A12	A12	
populatn	2	Population in thousands	Scale	8	Right	F8	F8	
density	3	Number of people / sq. kilometer	Scale	8	Right	F8.1	F8.1	
urban	4	People living in cities (%)	Scale	8	Right	F5	F5	
religion	5	Predominant religion	Nominal	8	Left	A8	A8	
lifeexpf	6	Average female life expectancy	Scale	8	Right	F4	F4	
lifeexpm	7	Average male life expectancy	Scale	8	Right	F5	F5	
literacy	8	People who read (%)	Scale	8	Right	F4	F4	
pop_incr	9	Population increase (% per year))	Scale	8	Right	F5.1	F5.1	
babymort	10	Infant mortality (deaths per 1000 live births)	Scale	8	Right	F6.1	F6.1	
gdp_cap	11	Gross domestic product / capita	Scale	8	Right	F6	F6	

Variables in the working file

Variable Information: World95.sav (cont.)

Variable	Position	Label	Measurement Level	Column Width	Alignment	Print Format	Write Format
calories	13	Daily calorie intake	Scale	8	Right	F6	F6
aids	14	Aids cases	Scale	8	Right	F8	F8
birth_rt	15	Birth rate per 1000 people	Scale	8	Right	F5.1	F5.1
death_rt	16	Death rate per 1000 people	Scale	8	Right	F6	F6
aids_rt	17	Number of aids cases / 100000 people	Scale	8	Right	F8.2	F8.2
log_gdp	18	Log (base 10) of GDP_	Scale	8	Right	F8.2	F8.2

Variable Values: World95.sav

Value		Label
religion	[a]	missing
region	1	OECD
	2	East Europe
	3	Pacific/Asia
	4	Africa
	5	Middle East
	6	Latn America
climate	1	desert
	2	arid / desert
	3	arid
	5	tropical
	6	mediterranean
	7	maritime
	8	temperate
	9	arctic / temp
	10	arctic

a. Missing value

INDEX