# 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: <u>GORDONLA@ONEONTA.EDU</u> or <u>JOHNSOSD@ONEONTA.EDU</u>.

#### 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:**

Copyright © 2001, 2005 by the authors

Few rights reserved. Any part of this publication may be reproduced for educational purposes provided that appropriate credit is given to its authors and SPSS, Inc. Use on the SUNY Oneonta campus must be coordinated through the authors and the Department of Mathematics, Computer Science & Statistics.

SPSS<sup>TM</sup> is a trademark of SPSS, Inc. The inclusion of SPSS Statistical Software input and output contained in this manual was accomplished using screen shots taken from SPSS version 13.0. For additional documentation, FAQ's, etc. go to the SPSS web site at http://www.spss.com.

### TABLE OF CONTENTS

TABLE OF CONTENTS	2
INTRODUCTION	5
SPSS ENVIRONMENT:	
GETTING STARTED WITH SPSS	3
THE INITIAL SPSS WINDOW	)
ENTERING DATA INTO A NEW DATA FILE	)
RETRIEVING A SAVED FILE	l
SAVING YOUR WORK	2
SPSS WINDOWS	3
SPSS FILE EXTENSIONS	1
HELP	1
SPSS PROCEDURES:	
ANALYSIS OF VARIANCE (ANOVA)	3
BAR CHART	)
BINOMIAL PROBABILITIES	1
BOXPLOT (BOX-AND-WHISKER PLOT)	7
CHI-SQUARE $(\chi^2)$	9
CONFIDENCE INTERVAL FOR A POPULATION MEAN	2
CONFIDENCE INTERVAL FOR A POPULATION PROPORTION	3
CROSSTABS (A.K.A. CONTINGENCY TABLES)	1

DESCRIPTIVE STATISTICS	
DOTPLOT	
FIVE-NUMBER SUMMARY	
FREQUENCY POLYGON/AREA GRAPH	
FREQUENCY TABLES	
HISTOGRAM	
HYPOTHESIS TEST FOR A POPULATION MEAN	
HYPOTHESIS TEST FOR A POPULATION PROPORTION	
INTERACTIVE GRAPHS	
LINE GRAPH	
LINEAR CORRELATION	
LINEAR REGRESSION	
MULTIPLE RESPONSE	
NORMAL PROBABILITIES	
NORMAL PROBABILITY PLOT	
OGIVE	
PARETO CHART	
PIE CHART	
SCATTERPLOT	
SIMPLE RANDOM SAMPLE (SRS)	
STANDARDIZED SCORES (OBTAINING Z-SCORES)	
STATISTICS & CASE PROCESSING TABLES	69

STEM-AND-LEAF PLOT	70
Γ-TEST	71

#### **SPSS DATA MANIPLUATION:**

CHART EDITOR	
COMPUTE	
DATA EDITOR	
DEFINE VARIABLES	
EDIT MENU	
EXPORTING TABLES AND CHARTS	
FILE & VARIABLE INFORMATION	
INSERT VARIABLE OR CASE	
MERGE FILES	
OUTPUT VIEWER	
PIVOT TABLES	
RECODE	
SELECTING CASES	
SORTING CASES & VARIABLES	
SPLIT A DATA FILE	
TABLE EDITOR	
TEXT FILES	
APPENDIX A: DATA FILE INFORMATION	
INDEX	

### **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.



### 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.
Intitled - SPSS Data Editor	🖽 Untilled - SPSS Data Editor
	File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Elle Edit View Data Transform <u>A</u> nalyze Graphs Utilities Add- <u>o</u> ns <u>W</u> indow <u>H</u> elp	eele R oo lo h fil III. Ioo
	1:
	Gender Year Height Month var var var var var var var
1:	1 2 S0 6/.5 Apr
	2 1 50 12.0 may 3 1 Sr 66.0 Sen
var var var var var 🗛	4 2 Jr 670 Dec
	5 1 Jr 63.0 Apr
	6 2 Jr 69.0 Jan
	7 2 Fr 68.5 May
3	8 1 So 685 Oct
	9 2 FF 620 Jun
4	
5	12
	13
	14
7	15
8	18
	12
9	19
10	20
	21
	22
12	23
12	25
	26
14	27
▲ ► Data View / Variable View /	28
SPSS Processor is ready	A State View (Variable View / )     A State View (Variable
	SPSS Processor is ready

### **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.** 



### 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.

Untitled - SPSS Data Editor										i - SPSS Data Editor	
Edit View Data Transform A	nalyze Graphs	Utilities Add-	ons Window	Help						vew Data Transform Analyze Graphs Utilities Add-ons Window Help	
ew pen pen Database pad Text Data	Height	Month	var	var	var	var	Var	var	var I.	Sender La formation to Park t	var var
ve Ctrl+S	67.5 72.0	Apr May								2 S 1 S Save in SPSS + D P ST	101
ve As rk File Read Only	66.0	Sep								1 S My Recent Documents	
olay Data File Information he Data o Processor Ctri+ tch Server	• 63.0 69.0 68.5	Apr Jan May								2         3         My Pecent         My Documents         5. General Social Survey.sav         BCSS           1	
nt Preview nt Ctrl+P	68.5 62.0 64.5	Oct Jun Aug								1     S     Desitiop     CO Drive (c)     pay     Mixed       2     F     Program Files     av     Mixed       2     F     Mixed (c)     av     Mixed	
cently Used Data cently Used Files it 15 16										All Sol Truty (1)	
17 18										Places Keeping 4 of 4 veriables. Variables.	
19 20							-	-		File game: Save	
21							-			Save as type: SPSS (* sav)	
23 24 25										Cencer      Mone venerate remeats to spreadulent      Sprevenue tables where dathed instand of data values      Sprevenue tables into a sea file	
26 27 28											
Data View / Variable View /		14			1	1				a View / Variable View /	1

### **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**<sup>1</sup>: This window allows for the modification of the appearance of graphs and charts. ٠
- DATA EDITOR<sup>1</sup>: 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**<sup>1</sup>: All data analysis output is maintained in this window. Editing of tables and graphs begins here.
- **PIVOT TABLE EDITOR<sup>1</sup>:** 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.

25.6

14.6

52.0 28.0 11.0 8.7 10.2

## Sample Windows:

LINA THE BAR NO

890

6000

1000

3300 4900

ent Am

+ [ + ] Data View / Van

605.0

86 Cathol

SB Orthodo 8 Cathol

15 Proteit 15 Ortho 1 Cathol 5 Cath

18 Orthod

2 Buddhy

Afoh

Data Editor <sup>1</sup>Each of these four windows is more thoroughly presented as separate topics.

1 terms selected (0 hidder/collapsed)	SPSS Processor is ready	The second s	H: 221 , V



· #/## 5 # ##



Chart Editor

The States

1					Fre	quencies	
14							
•		Statistics					
	Predicto	eart intigion					
	N Y#	10	100				
	Mit	ising					
E International							
			Preden	decant religion			
			Frequency	Parcent	Valid Percent	Cumulative Parcent	
	Valid	Avenue	4	3.7	17	3.7	
		Butthirt	1	6.4	8.5	10.2	
		Catholic	41	31.6	38.0	48.1	
		Hindu				49.1	
		Jewith				55.0	
		Muslim	27	24.0	25.0	75.0	
		Orthodox.		7.2	7.4	82.4	
		Protect	18	14.7	14.8	97.2	
		Taolot	2	1.8	1.0	99.1	
		1984				100.0	
		Tutal	108	99.1	100.0		
	Maxing	record					
	104		528	100.8			

### 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.

Open File					? 🛛
Look in: My Recent Documents Desktop My Documents My Computer My Network Places	SPSS Sampler	2.txt PS Of F	•		
	File <u>n</u> ame:			<b>•</b>	<u>O</u> pen
	Files of type:	All Files (*.*)		•	Paste
					Cancel

### 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.

t View Data Transform A 3 🔍 이 🖿 🕼 4 년	nalyze Graphs Utilities Add-ons	Window Help Topics Tutorial		Hide Back Forward Stop Refresh Ontions	
Var         Var           1         2           3         -           4         -           5         -           8         -           9         -           0         -           1         -           2         -           3         -           9         -           1         -           2         -           3         -           4         -           5         -           6         -           7         -           8         -           9         -           1         -           2         -           3         -           4         -           5         -           6         -           7         -           8         -           9         -           1         -           2         -           3         -           4         -           5         -           6         - <th>Var         Var           Image: Im</th> <th>Var Case Studies Statistics Coach Command Syntax Reference SPSS Home Page About Register Product</th> <th>ar. var. var. var ar. var. var.</th> <th>Qontents       Index       Search       Favorites         Type in the keyword to find:         2SLS         command syntax         3-D bar charts         rotating         3-D effects         in bar charts         3-D scatterplots         cases (Microsoft)         access (Microsoft)</th> <th>Getting Help Help is provided in many different forms: Help menu. The Help menu in most SPSS windows provides access to the main Help system, plus tutorials and technical reference material. • Topics. Provides access to the Contents, Index, and Search tabs, which you can use to find specific Help topics. • Tutorial. Illustrated, step- by-step instructions on how to use many of the basic features in SPSS. You don't have to view the whole tutorial from start to finish. You can choose the topics y want to view, skip around a view topics in any order, an use the index or table of</th>	Var         Var           Image: Im	Var Case Studies Statistics Coach Command Syntax Reference SPSS Home Page About Register Product	ar. var. var. var ar. var. var.	Qontents       Index       Search       Favorites         Type in the keyword to find:         2SLS         command syntax         3-D bar charts         rotating         3-D effects         in bar charts         3-D scatterplots         cases (Microsoft)         access (Microsoft)	Getting Help Help is provided in many different forms: Help menu. The Help menu in most SPSS windows provides access to the main Help system, plus tutorials and technical reference material. • Topics. Provides access to the Contents, Index, and Search tabs, which you can use to find specific Help topics. • Tutorial. Illustrated, step- by-step instructions on how to use many of the basic features in SPSS. You don't have to view the whole tutorial from start to finish. You can choose the topics y want to view, skip around a view topics in any order, an use the index or table of

#### **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.

Percentile Values  Quartiles  Cut points for: 10 equal groups	Central Tendency <u>M</u> ean Me <u>d</u> ian	Continue Cancel Help
A measure of the extent to which observati around a central point. For a normal distrib value of the kurtosis statistic is 0. Positive I indicates that the observations cluster mo	ons cluster pution, the kurtosis	
have longer tails than those in the normal ( and negative kurtosis indicates the observ cluster less and have shorter tails.	distribution /ations are group mid	lpoints

#### **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.



# **SPSS STATISTICS & GRAPHS:** PRODUCING INTRODUCTORY STATISTICS AND GRAPHS

### ANALYSIS OF VARIANCE (ANOVA)

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

**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.

**One-Way ANOVA: Options** 

One-Way ANOVA     Population in thousant     Number of people / sc     People living in cities     Average male life exp     People who read (%)     Population increase (     Aids cases [aids]     Bith rate per 1000 pec     Death rate per 1000 pic     Number of aids cases ▼     Contrasts     Post Hoc     Options	OK Paste Reset ancel Help	Statistics       Continue <ul> <li>Descriptive</li> <li>Eixed and random effects</li> <li>Homogeneity of variance test</li> <li>Brown-Forsythe</li> <li>Welch</li> </ul> Help <ul> <li>Means plot</li> <li>Missing Values</li> <li>Exclude cases <u>a</u>nalysis by analysis</li> <li>Exclude cases listwise</li> </ul> <ul> <li>Exclude cases listwise</li> </ul> <ul> <li>Means plot</li> <li>Missing Values</li> <li>Exclude cases listwise</li> </ul> <ul> <li>Means plot</li> <li>Missing Values</li> <li>Exclude cases listwise</li> </ul>
	Equal Variances Assumed Equal Variances Assumed Bonferroni Iukey Ty Sjdak Tukey's-b Du Scheffe Duncan Cr R-E-G-W F Hochberg's GT2 R-E-G-W Q Gabriel Equal Variances Not Assumed Tamhane's T2 Dunnett's T3 Ga Significance level: 05	aller-Duncan pe I/Type II Error Ratio: 100 nngtt introl Category: Last 2-sided C < Control C > Control mes-Howell Dunnett's C ontinue Cancel Help

**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**

#### <u>ANOVA</u>

Average female life expectancy

Between Groups

Within Groups

Total

Sum of Squares

7568.810

4501.539

12070.349

	Descriptives							
Average female lif	e expectancy							
					95% Confider Me	ce Interval for ean		
	Ν	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
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

#### POST HOC TEST: SCHEFFE

Multiple Comparisons

Dependent Variable: Average female life expectancy Scheffe

		Mean Difference			95% Confid	ence Interval
(I) Region or economic group	(J) Region or economic group	(I-J)	Std. Error	Sig.	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

#### MEANS PLOT



ANOVA

5

103

108

Mean Square

1513.762

43.704

F

34.636

Sig.

.000

df

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

### **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

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

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.



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



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



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



### **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'**).



#### **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'**).



#### 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

selected. Values for **q**, **n**, and **p** are input.

**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**. => 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 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'**).

Target Variable: Binomial_ex2	Numeric Expression:           =         CDF.BINOM(x:10,.56) - CDF.BINOM(x=1, 10, .56)	
Type & <u>L</u> abel		
<ul> <li>✔AR00001</li> <li>Binomial_ex1</li> <li>Binomial_ex2</li> </ul>	+     >     7     8     9       +     <	oup: ncentral CDF n te/Time netic tion ction
	CDF.BINOM(quant n, prob). Numeric: Returns the cumulative probability that the number of successes in n trials, with probability prob of success in each, will be less than or equal to quant. When n is 1, this is the same as CDF.BERNOULLI.	nd Special Variables Illi y
(optional case sele	ection condition) Cdf. Haffarr Cdf. Hyper	n

Dialog Box: Target Variable is named and Numeric Expressions are

**SPSS Output:** Probability distribution is displayed.

1:										
	VAR00001	Binomial_ex1	Binomial_ex2	Х	Binomial_ex3	var	Var-	var	VBr	Vá
1	1.00	.24	_15	.00	.00					
2				1.00	.00					
3				2.00	.02					
4				3.00	.07					
5	+			4.00	.15					
6		11		5.00	.23					
7				6.00	.24					
8				7.00	.18					
9				8.00	.08					
10	Q.			9.00	02					
11	+	2.5		10.00	.00					-
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24							-			
25										
26										
-										

### **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 ( $\mathbf{Q}_1$ ), Median ( $\mathbf{Q}_2$ ), Third Quartile ( $\mathbf{Q}_3$ ), 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.
Boxplot       Define         Image: Boxplot       Image: Define         I	Define Simple Boxplot: Summaries of Separate Variables       Image: Comparison of Compar	(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.)



# **CHI-SQUARE** $(\chi^2)$

CHI-SOUARE 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.)

	The Many Data Taranée Anima Casha Uliking Addang Window Hab	
		weight cases
<pre>/ . 1 2 3 4 5 6 7 7 8 9 10 11 12 13 14 15 16 16 </pre>	Color       Candy_Count       var       var	cases by Erequency V (* Number us: Weight ca



Wumber of m&m's per Colo

OK.

Paste

Reset

Cancel

Help

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. <u>The first value of the list must correspond to the lowest value of the test variable</u>. 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.

				0		
				m&n	n Color	
hi-Square Test				Observed N	Expected N	Residual
Number of m&m's per C	Test Variable List:	ОК	Blue	309	373.7	-64.7
	🚸 m&m Color [Color]	Paste	Brown	213	202.4	10.6
	त 👘 👘	Reset	Green	337	249.1	87.9
		Cancel	Orange	303	311.4	-8.4
		Help	Red	184	202.4	-18.4
Dected Range	Expected Values		Yellow	211	218.0	-7.0
Lise specified range	Ali categolijes equal     Values:		Total	1557		
Lower:	Add .13 A .16 .20	Evact	Test	t Statistics		
	Remove 14	Dations		m&m Col	or	
			Chi-Squar	å 44.87	'5	
			df		5	
			Asymp. Si	g00	00	
			a. 0 cell The r	ls (.0%) have experimentation	xpected frequenc	ies less than 5. w 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.

Crosstabs	
<ul> <li>Miles per Gallon (mpg)</li> <li>Engine Displacement (</li> <li>Horsepower [horse]</li> <li>Vehicle Weight [hos.] [v</li> <li>Time to Accelerate fror</li> <li>Model Year (modulo 12)</li> <li>Cyltec = 1   cyltec = 2 (</li> </ul>	Bgw(s):       □K         Paste       Paste         Column(s):       Cancel         Help       Help         Layer 1 of 1       Next.         Preyrous       Next.
Suppress <u>t</u> ables	
Egact	Statistics Cells Format
Egact	Statistics Cglls Format

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

Number of Cylinders \* Country of Origin Crosstabulation

Count

		Co	ountry of Orig	gin	
		American	European	Japanese	Total
Number of	3 Cylinders	0	0	4	4
Cylinders	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 <sup>a</sup>	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,  $\mu$ , 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.	<b>SPSS Output:</b> 95% C.I. = $(112.77 \le \mu \le 139.36)$ lbs	s.
He Edit Yew Data Iransform Analyze Graphs Utilities Add-ons Window Help		Descriptives	
			Statistic Std. Error
₩eiαht var var var var var var	Dependent List: OK	Weight Mean	126.07 6.199
98	Paste	95% Confidence Interval I ower Bound	112 77
2 102 3 104 4 106	Eactor List: Eactor	for Mean Upper Bound	139.36
5 146 6 185	Нер	5% Trimmed Mean	124.35
7 122	Label <u>C</u> ases by:	Median	122.00
9 140		Variance	576 495
	Uisplay		24.010
12 161		Std. Deviation	24.010
		Minimum	98
14 124 15 126		Maximum	185
16 A Data View (Variable View (	Explore: Statistics	Range	87
SPSS Processor is ready		Interquartile Range	34
	✓ Descriptives	Skewness	1,176 .580
	Confidence Interval for Mean: 95 %	Kurtosis	1 254 1 121
	M-estimators	Kuitosis	1.254 1.121
	I <u>P</u> ercentiles		
	Continue Cancel Help		

### **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  $\alpha$  (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.'



**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  $\mathbf{p} = .70 + .26$  or  $\mathbf{p} = (.674 \le \mathbf{p} \le .726)$ 

ile <u>E</u> dit	⊻iew <u>D</u> ata	Transform Analy	/ze <u>G</u> raphs	Utilities Add-	ons <u>W</u> indow	Help		
2 8	a 🔍 🗠	a 🔚 🕼	#4 <u>*1 (</u> ***		<b>S</b>			
8 :								
	VAR00001	Error	var	var	var	var	var	
1	1.00	.0256						
2								
3								
4								
5								_
6								
7								+-
8								+-
9								-
10								+-
11								+-
12								+-
14								+
15								+
16								+.
	ata View 🖌 🗸	riable View 🖊			4			I

### **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.'

<b>Dialog Box Input:</b> Move the variable ' <b>JOB CATEGORY</b> ' into the Row	1 Cell and Crosstabs Statistics options:				
Dialog Box Input: Move the variable 'JOB CATEGORY' into the Row and 'GENDER' is placed into the Column(s) Cell. Percentages for cells a statistics are selected through the buttons at the bottom of the dialog box.	r(s) Cell Cell and Crosstabs Statistics options: and Crosstabs: Cell Display Counts C				
<ul> <li>Display clustered bar charts</li> <li>Suppress tables</li> <li>Exact</li> <li>Statistics</li> <li>Cells</li> <li>Format</li> </ul>	Image:				

#### **Crosstabs Output:**

			Gei	_	
			Female	Male	Total
Employment	Clerical	Count	206	157	363
Category		% 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%

**Employment Category \* Gender Crosstabulation** 

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 Count 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 Column 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 Row 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 Column 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 Total 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 Count 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.

Employees by Selery Penge \* Minerity Classification Crosstabulation

Try it out - (answers below)

1) What percent of non-minority employees earn \$20,000 - \$39,999?		noyees by building ite	ige wintority clussifie	uuon eros	5000000	<u> </u>
<ul> <li>2) Minority employees earning &lt; \$20,000 represent what percent of all employees?</li> <li>2) Winority employees earning &lt; \$20,000 represent what percent of all employees?</li> </ul>				Mino Classifi	ority ication	
3) Minority employees earning < \$20,000 represent what percent of all minority employees?				No	Yes	Total
4) Employees earning \$20,000 - \$39,999 represent what percent of all employees?	Employees	< \$20,000	Count	22	11	33
<ul> <li>5) In the "Yes" minority column, 25.5% represents what measure?</li> <li>6) Non-minority (No) employees represent what percent of those earning \$20,000 - \$39,999?</li> </ul>	by Salary Range		% within Employees by Salary Range	66.7%	33.3%	100.0%
<ul> <li>7) Among employees earning &gt; \$39,999, what percent are minority employees?</li> <li>8) How many employees earn &lt; \$20,000?</li> </ul>			% within Minority Classification	5.9%	10.6%	7.0%
9) What doe the number 104 in the bottom marginal values represent?			% of Total	4.6%	2.3%	7.0%
10) How many minority employees earn < \$20.000?		\$20,000 - \$39,999	Count	251	86	337
			% within Employees by Salary Range	<sup>s</sup> 74.5%	25.5%	100.0%
ANSWERS:			% within Minority Classification	67.8%	82.7%	71.1%
1) Column question $-07.8\%$ (Check: $251/5/0 = .078 \Rightarrow 07.8\%$ ). 2) Totals question $-2.3\%$ (Check: $11/474 = .023 \Rightarrow 21.3\%$ )			% of Total	53.0%	18.1%	71.1%
3) Column question $-10.6\%$ (Check: $11/1/4 = .023 = 21.5\%$ ).		>\$39,999	Count	97	7	104
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			% within Employees by Salary Range	93.3%	6.7%	100.0%
employees (Check: 86/.337 = .255 => 25.5%). 6) Row question – 74.5% (Check: 251/337 = .745 => 74.5%).			% within Minority Classification	26.2%	6.7%	21.9%
7) Row question $-6.7\%$ (Check: $7/104 = .067 => 6.7\%$ ).			% of Total	20.5%	1.5%	21.9%
8) Count question – From the left marginal total there are 55 employees earning < %20,000. 9) Column count – number of minority employees	Total		Count	370	104	474
10) Column count – 11.			% 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.	<b>SPSS Output: Descriptive Statistics</b> via Frequency command
► Frequencies          Image: Prequencies       Variable(s):       OK         Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience         Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience         Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous Experience       Image: Previous	Percentile Values       Central Tendency            Quartiles           Median             Quartile(s):           Median             Percentile(s):           Mgde             Add           Sum             Dispersion           Minimum             Maximum           Skewness             Maximum           Mutosis	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         \$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.

Descriptiv	es									
Employee (     Date of Birt     Educationa     Employmer	Code [id] <u> </u>	able(s): Current Salary [sa	lary] OK Paste Reset			De	escriptive Sta	tistics		
🛞 Beginning 🤅	Salary [sa		Cancel		Ν	Range	Minimum	Maximum	Mean	Std. Deviation
Months sin     Previous Fi	ce Hire []		Lista	Current Salary	474	\$119,250	\$15,750	\$135,000	\$34,419.57	\$17,075.661
Minoritu Ela				Valid N (listwise)	474					
	Descriptives: Options	Continue Cancel m Help ean								

# 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.



# 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:
Dialog Box: Variable is selected.	Dialog Box: Select the Five-Number Summary statistics	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.



# **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<sup>1</sup>, 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.
Current Salary [sala Beginning Salary [s Months since Hire [ Previous Experienc Months characterized Months characterized Experienc Experien	Image: Program         Educational Level (years)           Valid         Process         Valid Processi         Processi           Valid         0         311.2         11.2         11.2           12         190         40.1         40.1         51.3           14         6         1.3         1.3         32.5           15         116         64.5         24.5         77.0           16         5.9         1.2         1.2         4.3           19         9.19         1.9         93.7           19         2.7         5.7         5.7         99.4           20         2         .4         .4         99.5           21         1         2         .2         100.0           Total         47.4         100.0         100.0

**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.
Histogram       Image: Constraint of the con	<figure></figure>

# 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.

Year level in colleg     Gender [gender]     Bight - Left Handed     Hours worked at pr     Dwn a car [car]     Mumber of siblings     Amount paid for las     Current GPA [gpa]         Test ⊻alue: 69.1	Image: Section of the section of t	ert Format Analyze Graph	s Utilities Add-gns Window Help One-Sample Statistics N Mean Std. Deviation Std. Error Mean ight 25 70.780 3.3481 6696 One-Sample Test Test Value = 69.1 95% Confidence Interval of the Difference t df Sig. (2-tailed) Difference Lower Upper ight 2.509 24 .019 1.6800 .298 3.062
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

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.<sup>1</sup>

**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.



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

<sup>1</sup>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.'

Compute Variable			×
Target Variable: z Type & Label ♥ VAR00001 ♥ z	Numeric Expression: = (0.70 - 0.65)/SQRT(0.65 * 0.35/1234) + < > 7 8 9 - <= > = 4 5 6 × = ~= 1 2 3 / & 1 0 . * ~ (1) Delete	Function group: All Arithmetic CDF & Noncentral CDF Conversion Current Date/Time Date Arithmetic Date Arithmetic Date Creation Date Extraction Euroctions and Special Varia	New York Contraction of the second se
	OK Paste Reset Cancel H	Help	

**SPSS Output:** calculated test statistic

🗑 Untit	led - SPSS D	ata Editor						
<u>File E</u> dit	⊻iew <u>D</u> ata	<u>T</u> ransform <u>A</u>	<u>I</u> nalyze <u>G</u> rapł	ns <u>U</u> tilities A	dd- <u>o</u> ns <u>W</u> ind	ow <u>H</u> elp		
<b>2</b>	<b>ð 🖳</b> 🖂	a 🔚 🕻	2 🙀 📲	i∎ 🖬 🏚	<u>r</u> 🔊 🖉	1		
2 : z								
	VAR00001	z	var	var	var	var	var	1 A
1	1	3.68						
2								
3								
4								—— I
5								
7								
9								
10								
11								
12								
13								
14								
15								
16  ⊂ [ } ] A Da	ata View 🖉 🗸	ariable View	/					   ▶   ▼
	A TON A TO	andoro viow ,	, 	PSS Processor	is ready			

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.'

Compute Variable				🛅 Unti	tled - SPSS D	ata Editor				
Target Variable:	Numeric Expression:			<u>File E</u> di	t <u>V</u> iew <u>D</u> ata	<u>T</u> ransform <u>A</u>	<u>I</u> nalyze <u>G</u> raphs	Utilities Add	- <u>o</u> ns <u>W</u> indow	He
p_value	= 1 - CDF.NURMAL(z, U, 1)			🖻 🖬	🞒 🖳 🗠	o 🗠 🔚 🛚	2 M M T		S 🛛 🖉	
Type & Label				4 :						
🚸 VAR00001	· • · ·	Function group:			VAR00001	z	p_value	var	var	Т
() ₹	+ < > 7 8 9	All	~	1	1	3.68	.0001155			
₩ p_value	- <= >= 4 5 6	CDF & Noncentral CDF		2	2					
	× = ~= 1 2 3	Conversion		3						+
	/ &   0 .	Date Arithmetic		4						+
	** ~ () Delete	Date Creation								+
		Functions and Special) (a)	rishlas:	7						+
	Numeric. Returns the cumulative probability	Cdf.Bernoulli		8	1					t
	that a value from the normal distribution, with specified mean and standard	Cdf.Beta Cdf.Binom		9						T
	deviation, will be less than quant.	Cdf.Bvnor		10	1					+
		Cdf.Cauchy Cdf.Chisg		11						+
		Cdf.Exp		12						+
		Cdf.Gamma		14						+
If foptional case set	ection condition)	Cdf.Lieom Cdf.Halfnrm		15	i					t
2	,	Cdf.Hvper			[ 	- vielete Vierre	,			T
Γ	OK Paste Reset Cancel	Help			ata view A v	anapie view ,		S Processor is	ready	-

nately 0.00012, which is less than the designated ull hypothesis is rejected.

var

.

<u>اا</u>

var

# **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.

Create Bar Chart						
Assign Variables Bar Ch	art Options Error Bars Titles Options					
<ul> <li>Case [\$case]</li> <li>Percent [\$pct]</li> <li>Beginning Salary</li> <li>Current Salary [s</li> </ul>	100% stacked     100% stacked     100% stacked					
Date of Birth [bd	Count [\$count]           Image: Second					
Months since Hi	Legend Variables <u>Color:</u> <u>Col</u>					
	≥gyle: Stack ▼ Panel Variables					
Bars Represent ────────────────────────────────────						
OK <u>P</u> a	ste <u>R</u> eset Cancel Help					

**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.

Create Bar Chart 🛛 🔀	Create Bar Chart 🔀	Create Bar Chart 🛛	Create Bar Chart 🛛 🔀
Create Bar Chart  Assign Variables Bar Chart Options   Error Bars   Titles   Options    Create Gare [Socae]  Perevice (Socae]  Current Salavy [s Date of Birth [bd Garder (gender] Minority Classific Minor	Create Bar Chart  Assign Variables Bar Chart Options Error Bars Tales Options Bar Labels  C Grouplar base C Crouplar base Bar Baseline: C Automatic C Quetom: C Copylor base C Copylor base C Copylor base C C C Copylor base C C C C C C C C C C C C C C C C C C C	Create Bar Chart          Assign Variables       Bar Chart Options       Error Bars         Display Error Bars       Confidence Interval         Units:       Confidence Interval for Mean         Units:       Confidence Interval for Mean         Shape       Cap Width: 45 🖃 %	Create Bar Chart           Assign Variables       Bar Chart Options       Error Bars       Titles         Chart Title:
Bars Represent Count (\$count)	3-D Coordinate 3-D Effect	Direction	Note           Because a graph's size is affected by the sizes of all of its parts, you should break long text into multiple lines by using the Enter key.           OK         Paste         Reset         Cancel         Help



- <u>Assign Variables</u>: 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.
- <u>Titles</u>: Assign selected text prior to chart creation.
- <u>Options</u>: Set the order of categorical variable values; select predesigned 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.

#### Zip-Zing Employees by Job Classification



## **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.



TOUTPUT14 - SPSS Viewer

E Output

Ele Edit View Insert Format Analyze Graphs Ublittles Add-ons Window Help

Interactive Graph

1 🗧 🗖 🧕

# 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.



# 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.

Bivariate Correlations		$\mathbf{\overline{X}}$	Bile Edit View Data Transform Insert	Format Analyze Graphs Utilit	ies Add- <u>o</u> ns <u>W</u> indow	Help	
Population in thouse     Number of people (	Variables: Average male life experience	ОК					
Average female life	Gross domestic product	Paste <u>R</u> eset	Cutput     Gotput     Gotput     Gotput     Gotput     Gotput     Gotput     Gotput     Gotput	Correlations	Correlations		
People who read (%		Cancel				Average male life expectancy	Gross domestic product / capita
Population increase				Average male life expectancy	Pearson Correlation	1	.639**
🗰 Infant mortality (dea		Help		•	Sig. (2-tailed)	100	.000
🛞 Beaion or economic 💟				Gross domestic product /	Pearson Correlation	639*	× 1
Correlation Coefficients				capita	Sig. (2-tailed)	.000	-
Conciación Coemcientis					N	109	109
✓ Pearson	<u>S</u> pearman			**. Correlation is significant	nt at the 0.01 level (2-ta	iled).	
🖲 <u>I</u> wo-tailed 🔷 One-ta	ailed						

# 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.

Minority Classification [   Method: Enter     Selection Variable:     Case Labels:     WLS Weight:
----------------------------------------------------------------------------------------------------

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

## Variables Entered/Removed<sup>®</sup>

Model	Variables Entered	Variables Removed	Method
1	Educational Level <sub>a</sub> (years)		Enter

a. All requested variables entered.

b. Dependent Variable: Beginning Salary

## Regression Variables (Above):

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

# Model SummaryModelRR SquareAdjusted RStd. Error of<br/>the Estimate1.633a.401.400\$6,098.259

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

#### **Regression Model Summary** (Above): Correlation Coefficient (r): .663

Coefficient of Determination  $(r^2)$ : .401 (or 40.1%)



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

## Coefficientsa

		Unstandardized Coefficients		Standardized Coefficients	_	
Model		В	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)

#### ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11747808912.318	1	11747808912.318	315.897	.000 <sup>a</sup>
	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.

## **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

## 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.

🛅 Untit	led - SPSS D	ata Editor							×
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	<u>T</u> ransform <u>A</u>	<u>I</u> nalyze <u>G</u> rapł	ns <u>U</u> tilities A	dd- <u>o</u> ns <u>W</u> indo	ow <u>H</u> elp			
<b>2</b>	a 🔍 🗠	a 🔚	2 M M		5 0	1			
12 : Na	ational_Ge	og							
	Location	Time	Newsweek	National_G eog	People	Sports_III	Year	Transportati on	-
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	2005	3	_
17		1	1			1	2005	3	_
10	4	. 1				. 1	2005	2	_
19			. 1	1	. 1		2005	1	_
20	zta View 1 V	richle Mow	1				2005		- 11 •
	ata view X V	anable view j	/	PSS Processor	is ready			<u> </u>	Ш_

## 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. <u>This is a session specific variable that will be lost upon leaving SPSS</u>. 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.)

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

Define Multiple Respons	e Sets	
Set Definition Contraction Contraction Contraction Set Definition	Variables in Set:         Image: Set the set	Close Help Mult Response <u>S</u> ets:
Variables Are Coded As <u>D</u> ichotomies Counted Categories Range: <u>Name:</u> Magazines Label: Magazines Read R	value: 1 phrough	<u>Add</u> Change <u>R</u> emove

1) Initial Multiple Response dialog box:

Define Multiple Response Sets	X
Set Definition Variables in Set:	Close Help Mult Response <u>S</u> ets:
Variables Are Coded As	Add Change Bemove



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).

The response Denne Daning Done move variables to right	SPSS Output: Mu	altiple Response	frequend	cy table	
		Case Sum	mary		
Multiple Response Frequencies			Cases		
Mult Response Sets: Table(s) for:		Valid	Missing	5	Total
	Ν	Percent N	N Pe	rcent I	N Percent
Paste	\$Magazines <sup>a</sup> 20	100.0%	0	.0%	20 100.0%
<u>R</u> eset	a. Dichotomy grou	p tabulated at val	ue 1.		
Cancel		\$Magazine	s Freque	ncies	
Help					
			Re	sponses	Dercent
Missing Values			N	Percent	of Cases
Exclude cases listwise within <u>d</u> ichotomies	Magazines	Time	11	21.6%	55.0%
Exclude cases listwise within categories	Read Regularly	lead a Newsweek 10	19.6%	50.0%	
	Regularly	National_Geog	g 9	17.6%	45.0%
		People	10	19.6%	50.0%
		Sports_Ill	11	21.6%	55.0%
	Total		51	100.0%	255.0%

## 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

Multiple Response Cro	sstabs		
Time     Newsweek     National Geog	$\rightarrow$	Ro <u>w(s):</u> \$Magazines	OK <u>P</u> aste
Area and a coope     Area and a coope     Area and a coope     Area and a coope and a	$\rightarrow$	Colum <u>n(s):</u> Location(? ?)	<u>R</u> eset Cancel Help
Mult Response Sets:	$\rightarrow$	Layer(s):	
		Define Ranges	Deptions

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

Multiple F	Response Cros	stabs: Define Vari 🔀
Mi <u>n</u> imum:	1	Continue
Ma <u>x</u> imum:	4	Cancel
		Help

**Options** to identify desired percentages.

Multiple Response Crosstabs: Options	: 🔀
Cell Percentages	Continue
🔽 Ro <u>w</u> 🔽 <u>C</u> olumn 🔽 <u>T</u> otal	Cancel
Match variables across response sets	Help
Percentages Based on	
C Cages C Besponses	
Missing Values	
📃 🔲 Exclude cases listwise within dichotomies	
Exclude cases listwise within categories	

## 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													
	V	alid	Total											
	N Percent N Percent													
\$Magazines*Location	20	100.0%	0	.0%	20	100.0%								

### **\$Magazines\*Location Crosstabulation**

				Wher	e Read		
			Library	Home	Work	On-Line	Total
Magazines	Time	Count	3	4	3	1	11
Read	% within \$Magazine		27.3%	36.4%	27.3%	9.1%	
Regularly		% 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
	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,  $\mu = $34.419.60$  and  $\sigma = $17,075.66$ )

## **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,  $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 CDF**." => 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 **P**(**x** ≤ \$57,000) = .91

**EXAMPLE 2:** Determine the probability that a randomly selected salary is greater than \$57,000. That is, P(x > \$57,000). In order to find this probability, we use the fact that  $P(x \le $57,000) = .91$ . As such,  $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,  $P(\$32,100 \le x \le \$57,000)$ . In order to find this probability, we again use the fact that  $P(x \le \$57,000) = .91$ . In addition, we use the fact that  $P(x \le \$57,000) = .91 - .45 = .46$ .

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

🔲 Compute Variable		X
Target Variable: cum_prob	Numeric <u>E</u> xpression: = CDF.NORMAL(salary, 34419.60, 17075.66)	
Type & Label		~
Employee Code [id]     A; Gender [gender]     Date of Birth [bdate]     Educational Level (yea     Employment Category       Current Salary [salary]     Beginning Salary [salary     Months since Hire [job]	+         <         7         8         9         All           +         <         >         4         5         6         CDF & Noncentral CDF            <         >         4         5         6         CDF & Noncentral CDF            <         >         1         2.3         Current Date/Time         Date Creation               Delete         Date Creation         Date Extraction	
Previous Experience (r Minority Classification [ cum_prob	CDF.NORMAL(quant, mean, stddey). Numeric. Returns the cumulative probability that a value from the normal distribution, with specified mean and standard deviation, will be less than quant. Cdf. Bernoulli Cdf. Bernoulli Cdf. Broom Cdf. Bro	
(optional case sele	ction condition)	~
	OK Paste Reset Cancel Help	

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

🛗 Employ	🖩 Employee data.sav - SPSS Data Editor														
Eile Edit	⊻iew <u>D</u> ata <u>T</u> rans	sform <u>A</u> nalyze <u>G</u> r	aphs <u>U</u> tilities	Add- <u>o</u> ns <u>₩</u>	indow <u>H</u> elp										
<b>6</b>															
1 : cum_	1: cum_prob 0.906978092941976														
	id g e	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority	cum_prob	v				
1	1 m	02/03/1952	15	3	\$57,000	\$27,000	98	144	0	.91					
2	2 m	05/23/1958	16	1	\$40,200	\$18,750	98	36	0	.63					
3	3 f	07/26/1929	12	1	\$21,450	\$12,000	98	381	0	.22					
4	4 f	04/15/1947	8	1	\$21,900	\$13,200	98	190	0	.23					
5	5 m	02/09/1955	15	1	\$45,000	\$21,000	98	138	0	.73					
6	6 m	08/22/1958	15	1	\$32,100	\$13,500	98	67	0	.45					
7	7 m	04/26/1956	15	1	\$36,000	\$18,750	98	114	0	.54					
8	8 f	05/06/1966	12	1	\$21,900	\$9,750	98	0	0	.23					
9	9 f	01/23/1946	15	1	\$27,900	\$12,750	98	115	0	.35					
10	10 f	02/13/1946	12	1	\$24,000	\$13,500	98	244	0	.27					
11	11 f	02/07/1950	16	1	\$30,300	\$16,500	98	143	0	.40					
12	12 m	01/11/1966	8	1	\$28,350	\$12,000	98	26	1	.36					
13	13 m	07/17/1960	15	1	\$27,750	\$14,250	98	34	1	.35					
14	14 f	02/26/1949	15	1	\$35,100	\$16,800	98	137	1	.52					
15	15 m	08/29/1962	12	1	\$27,300	\$13,500	97	66	0	.34					
16	16 m	11/17/1964	12	1	\$40,800	\$15,000	97	24	0	.65					
17	17 m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0	.75					
18	18 m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0	1.00					
19	19 m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0	.68					
201	20 f	01/23/1940	12	1	\$26.250	\$11.550	97	48	Π	32					
<u>∢  </u> ⊁]∖Dat	a view A variable	e view /	SPSS Proce	ssor is ready	<u> </u>						<u> </u>				

# 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 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.



# 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 <u>Line Charts</u> 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.



# 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"
Pareto Charts         Simple         Simple         Stacked         Define         Cancel         Help         Data in Chart Are         Counts or sums for groups of cases         Sums of separate variables         Values of individual cases	Define Simple Pareto: Counts or Sums for Groups of Cases            Employee Code [id]         Gender [gende]         Out as of brink [bdate]         Finployment Category [is]         Sums of variable:             Current Salary [salary]         Menority Salary [salary]         Minority Classification [iii]         Previous Experience (m         Minority Classification from:             Paret by             Regresent             Viset variables (no empty rows)               Category Agis:                  Paret by                  Regresent                  Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                 Regresent                    Regresent                 Regresent	500 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 100% 1

# **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.



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.

<u>Chart Editor Icons</u>: 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.

<u>Counts & Percentages</u>: 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.



## 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

s to be choss  Select Cas  Case  Case  Select Case  Selec	sen from the first k cases is designated.         ses       Select         Code [id]       All cases         al Level (yea       If gondition is satisfied         in Category         If gondition is satisfied         in Category         Findom sample of cases         Salary (salary)       Sample         © Based on time or case range       Sample Size         Sample Size       Approximately       % of all cases         • Exactly 70       cases from the first       474
Current Status:	Exactly 70 cases from the first 474 cases

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

🛅 Emplo	🖬 Employee data.sav - SPSS Data Editor													
<u>Eile E</u> dit	Elle Edit View Data Iransform Analyze Graphs Utilities Add-gns Window Help													
	ele por le m tit de sol													
7 : edu	с	15												
	id g e n	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	-					
1	1 m	02/03/1952	15	3	\$57,000	\$27,000	98	144						
2	2 m	05/23/1958	16	1	\$40,200	\$18,750	98	36						
3	3 f	07/26/1929	12	1	\$21,450	\$12,000	98	381						
4	4 f	04/15/1947	8	1	\$21,900	\$13,200	98	190						
_5	5 m	02/09/1955	15	1	\$45,000	\$21,000	98	138						
6	6 m	08/22/1958	15	1	\$32,100	\$13,500	98	67						
-7	7 m	04/26/1956	15	1	\$36,000	\$18,750	98	114						
8	8 f	05/06/1966	12	1	\$21,900	\$9,750	98	0						
9	9 f	01/23/1946	15	1	\$27,900	\$12,750	98	115						
10	10 f	02/13/1946	12	1	\$24,000	\$13,500	98	244						
11	11 f	02/07/1950	16	1	\$30,300	\$16,500	98	143						
12	12 m	01/11/1966	8	1	\$28,350	\$12,000	98	26						
13	13 m	07/17/1960	15	1	\$27,750	\$14,250	98	34						
14	14 f	02/26/1949	15	1	\$35,100	\$16,800	98	137						
	15 m	08/29/1962	12	1	\$27,300	\$13,500	97	66						
16	16 m	11/17/1964	12	1	\$40,800	\$15,000	97	24						
	17 m	07/18/1962	15	1	\$46,000	\$14,250	97	48						
	18 m	03/20/1956	16	3	\$103,750	\$27,510	97	70						
	19 m	08/19/1962	12	1	\$42,300	\$14,250	97	103						
20  ▲   ▶  \Da	ata View <u>↓</u> Variabl	01/23/1940	12	1	\$26.250	\$11.550	97	48	<b>П</b> .					
	~		SPSS Proces	sor is ready	J			Filter O	n //					

## 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,  $\mu = $34.419.60$  and  $\sigma = $17,075.66$ )

**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).

Compute Variable			Employ	ee data.sav - S	PSS Data Editor	L under	• • • • • • • • • • • • • • • • • • • •	1						
			Elle Edit y	new Data Irans	rorm <u>A</u> nalyze <u>G</u> r	apris <u>U</u> tilities	Add-ons wi	naow <u>H</u> eip						
arget Variable:	Numeric <u>Expression:</u>			) 🖳 🖂 🖾				0						
zscore	= (Salary - 34419.60)/17075.66	<u>^</u>	9 : salary	/	27900									
Type & <u>L</u> abel		~		id g e	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority	zscore	
Employee Code (id)	l ▶ l'		1	1 m	02/03/1952	15	3	\$57,000	\$27,000	98	144	0	1.32	
Candas [aandas]		Function group:	2	2 m	05/23/1958	16	1	\$40,200	\$18,750	98	36	0	.34	
az Gender (gender)	+ < > 7 8 9	All 🔼 🔼	3	3 f	07/26/1929	12	1	\$21,450	\$12,000	98	381	0	76	<u> </u>
Date of Birth [bdate]	- <= >= 4 5 6	Arithmetic	4	4 1	04/15/194/	8	1	\$21,900	\$13,200	98	190	U	/3	-
🏶 Educational Level (yea		Conversion		517	02/09/1955	15	1	\$45,000	\$21,000	96	130	0	.62	-
🚸 Employment Category	* = ~= 1 2 3	Current Date/Time		7 m	04/26/1956	15	1	\$36,000	\$18,750	98	114	0		
Current Salary [salary]		Date Arithmetic	8	8 f	05/06/1966	12	1	\$21,900	\$9,750	98	0	0	73	_
🔶 Beginning Salaru (salbr		Date Creation	9	9 f	01/23/1946	15	1	\$27,900	\$12,750	98	115	0	38	
🔶 Monthe since Hire fisht		Date Extraction 🔍 🗸	10	10 f	02/13/1946	12	1	\$24,000	\$13,500	98	244	0	61	
			11	11 f	02/07/1950	16	1	\$30,300	\$16,500	98	143	0	24	<u> </u>
Previous Experience (r	CDF.NORMAL(quant, mean, stddev).	Eunctions and Special Variables:	12	12 m	01/11/1966	8	1	\$28,360	\$12,000	98	26	1	36	-
Minority Classification [	Numeric. Returns the cumulative probability	Cdf.Bernoulli	13	13 m	07/17/1960	15	1	\$27,750	\$14,250	98	34 197	1	39	-
🚸 zscore	that a value from the normal distribution,	Cdf.Beta	14	14 I	08/29/1962	13	1	\$27,300	\$13,500	97	66	0	- 42	<u> </u>
	with specified mean and standard	Cdf.Binom	16	16 m	11/17/1964	12	1	\$40,800	\$15,000	97	24	0	.37	
	deviation, will be less than quart.	Cdf Cauchu	17	17 m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0	.68	
		Cdf Chisa	18	18 m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0	4.06	
		Cdf Evo	19	19 m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0	.46	
		Cdf.F	201 ∢ [ ► ]\ Data	View Variable	01/23/1940	12	1	\$26.250	\$11.550	97	48	n	- 48	•
	▲	Cdf.Gamma		A random	, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	SPSS Proce	ssor is ready							
		Cdf Goom												-

# 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).

Stat 'WI	<b>Statistics Table</b> accompanying the frequency table of the variable <b>'WEIGHT'.</b> (left without stats; right with stats)						Case Processing Table 'ORIGIN'.	accomp	oanying the	crosstał	os table of <b>'(</b>	CYLINI	DERS' and
Statistics Statistics							G		G				
	Vehicle Weight (lbs.)			Vehicle Wei	ght (lbs.)			Cas	e Processing	g Summ	ary		
_	N Valid 406 Missing 0		N	Valid	406	_			С	ases			
				Missing	0		V	alid	Mi	ssing	Т	otal	
=			Mean		2969.56		Ν	Percent	Ν	Percent	Ν	Percent	
			Median		2811.00	Number of Cylinders	405	99.8%	1	2%	406	100.0%	
				Std. Deviation	on	849.827	* Country of Origin	105	JJ.070	1	.270	-100	100.070

# **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.



# **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.


# **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.
- <u>Options Menu Toolbar Icons:</u> X & Y Reference Lines, Title, Annotation, Text Box, Footnote, Gridlines, Legend, Transpose, etc.
- <u>Elements Menu Toolbar Icons</u>: 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.

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

**<u>Transform</u>:** 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**

<u>Properties Dialog Box</u>: 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.

- <u>Chart Size</u>: Allows modification of the overall chart size.
- **<u>Fill & Border</u>**: 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.
- <u>Categories</u>: 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.



Properties	Properties 🛛 🗙	Properties 🛛 🗙
Bar Oxtinns     Variables     Depth & Angle       Chart Size     Fill & Border     Categories       Size in points     Height:     375 ÷     Width:     468.75 ÷       Image: Maintain aspect ratio     Image: Maintain aspect ratio     Image: Maintain aspect ratio	Bar Options Variables Depth & Angle Chart Size Fill & Border Categories Preview Color Ell Multiple Selection Border (0, 0, 0) Pattern Edit Reset (0, 0, 0) Border Style Weight Style The style End Caps Butted The Style	Bar Options       Variables       Derth & Annie         Chart Size       Fill & Border       Categories         Variable:       Gender       Image: Collapse (sum) categories less than:       Image: Solution:         Collapse (sum) categories less than:       Image: Solution:       Image: Solution:       Ascending         Order:       Image: Solution:       Ascending       Image: Solution:       Image: Solution:         Order:       Image: Solution:       Ascending       Image: Solution:       Image: Solution:         Female       Image: Solution:       Image: Solution:       Image: Solution:       Image: Solution:         Excluded:       Image: Solution:       Image: Solution:       Image: Solution:       Image: Solution:         Lower margin (%):       Image: Solution:       Image: Solution:       Image: Solution:       Image: Solution:
Apply Close Help	Apply Close Help	Apply Close Help

<= 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.



- **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 ٠ 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.

Properties	Properties	Properties 🛛 🗙
Chart Size       Fill & Border       Categories         Bar Options       Variables       Depth & Angle         Width       Image: Second structure       %: 100         Scale boxplot and error bar width based on count       %: 85         Boxplot and Error Bar Style       %: 85         Image: Second structure       %: 85         Boxplot and Error Bar Style       Stacked Bars         Image: Second structure       %: Scale by statistic         Image: Second structure       %: Scale to 100%	Chart Size Fill & Ronder Categories Bar Options Variables Depth & Angle Y-Axis: Count Z-Axis: Employment Category Group by Stack: X-Axis Cluster: gender Z-Axis Cluster: Columns:	Chart Size Fill & Border Celenonies Bar Options Variables Depth & Angle Effect Depth & Angle Shadow Shadow Margin Front Margin (%): 0 Rear Margin (%): 0 Distance Farther (100) Closer (1) Distance:
Apply <u>Close</u> <u>Help</u>	Apply Close Help	Apply Close Help

• 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.





### **Y-Axis Properties:**

PropertiesX	PropertiesX	Properties 🛛	Properties 🔀
Properties       Image: Chart Size       Text       Scale         Labels & Ticks       Number Format       Variables         Image: Chart Size       Display axis title       Display axis on the:       Default Image: Chart Size         Image: Chart Size       Display axis title       Display axis on the:       Default Image: Chart Size         Image: Chart Size       Display labels       Image: Chart Size       Image: Chart Size       Image: Chart Size         Image: Category Label Placement       Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size         Image: Category Label Placement       Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size         Image: Category Label Placement       Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size         Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size         Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size         Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: Chart Size         Image: Chart Size       Image: Chart Size       Image: Chart Size       Image: C	Properties       Image: Chart Size       Text       Scale         Chart Size       Text       Scale         Labels & Ticks       Number Format       Variables         Sample       The number 1000000 will appear as:       1,000,000         Decimal Places:       Image: Characters:       Image: Characters:         Irailing Characters:       Image: Characters:       Image: Characters:	Properties       Labels & Ticks     Number Format     Variables       Chart Size     Text     Scale       Preview in Preferred Size     Justification       AuBbCc 123     Justify       Font	Properties       Image: Chart Size       Number Format       Variables         Chart Size       Text       Scale         Range       Auto       Custom       Data         Minimum       Image: Chart Size       0.0       0.0         Maximum       Image: Chart Size       206.0         Major Increment       Image: Solution       Image: Chart Size         Origin       Image: Chart Size       Image: Chart Size         Type       Image: Chart Size       Image: Chart Size
Major Ticks ✓ Display ticks Style Outside ▼ Minor Ticks ✓ Display ticks Style Outside ▼ Number of minor ticks per major ticks: 1 Apply	✓ Display Digit Grouping         Scientific Notation         ④ Automatic         ④ Always         ⑥ Never	Iext	Image: Construction control in the construction of the

- 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.

### **X-Axis Properties:**

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



Chart Size Text Labels & Ticks Categories Variables          Chart Size Text Labels & Ticks Categories Variables         Display axis title       Display axis on the:         Default Y         Major Increment Labels         Display labels         Label grientation         Horizontal         Category Label Placement         Automatic         Custom         Display ticks         Style         Style         Minor Ticks         Display ticks         Style         Minor Ticks         Display ticks         Style         Style         Number of minor ticks per major ticks:	Properties 🔀	Properties
V Display axis title Display axis on the: Default     Major increment Labels   V Display labels   Label grientation   Category Label Placement   Category Label Placement   Category Label Placement   Category Label Placement   Custom   Ticks skipped between labels     Major Ticks   Variable:   Enployment Category     Order:   Category Label Placement   Custom   Ticks skipped between labels     Major Ticks   Variable:   Style   Numor Ticks   Display ticks   Style   Number of minor ticks per major ticks:     Lower margin (%):     Style     Category     Major Ticks     Image: Style     Version: Style     Lower margin (%):     Support     Category: Style     Image: Style <	Chart Size   Text   Labels & Ticks   Categories   Variables	Chart Size   Text   Labels & Ticks Categories   Variables
Major Increment Labels   Image: Display labels   Label grientation   Horizontal   Image: Categories   Category Label Placement   Image: Categories   Category Label Placement   Image: Categories   Categories   Categories   Sort by: Custom   Image: Categories   Categories   Sort by: Custom   Image: Categories   Categories   Sort by: Custom   Image: Categories   Image: Categories   Sort by: Custom   Image: Categories   Ima	☑ Display axis <u>ti</u> tle <u>D</u> isplay axis on the: Default <u>▼</u>	⊻ariable: Employment Category ▼
Category Label Placement   Category Label Placement   Category Label Placement   Custom   Custom   Ticks skipped between labels     Major Ticks   V Display ticks   Style   Outside     Minor Ticks   Display ticks   Style   Minor Ticks     Lower margin (%):   5     Upper margin (%):     5	Major Increment Labels	Collap <u>s</u> e (sum) categories less than: 5 %
Label grientation   Horizontal   Category Label Placement            • Automatic             • Custom             • Custom             Ticks skipped between labels      Major Ticks             Major Ticks             Style            Outside              Minor Ticks             Style            Style            Minor Ticks               Style            Minor Ticks               Lower margin (%):            Style              Lower margin (%):              Lower margin (%):              Lower margin (%):	I Display labels	Categories
Category Label Placement Automatic Custom Ticks skipped between labels  Major Ticks Style Outside   Minor Ticks Lower margin (%): 5 Upper margin (%): 5	Label orientation Horizontal	Sort by: Custom 💌 Direction: Ascending 💌
Clerical   Custom   Custom   Ticks skipped between labels     Major Ticks   V Display ticks   Style   Outside     Minor Ticks     Display ticks   Style   Vumber of minor ticks per major ticks:     Lower margin (%):   5     Upper margin (%):     5	Category Label Placement	Order:
Custom Ticks skipped between labels Major Ticks ✓ Display ticks Style Outside ✓ Minor Ticks Display ticks Style ✓ Number of minor ticks per major ticks: Display ticks Style ✓ Number of minor ticks per major ticks: Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ Customal Manager ✓ ✓ Customal Manager ✓ Customal Manager ✓ Manager ✓ Customal Manager ✓ ✓ Customal ✓ ✓ Customal Manager ✓ ✓ Customal Manager ✓ ✓ Customal ✓ Customal Customal ✓ ✓ Customal Customal ✓ Customal ✓ ✓ Customal ✓ Customal Customal ✓ Customal ✓ ✓ Customal ✓ Customal ✓ Customal ✓ Customal Customal ✓ Customal Customal Customal Customal ✓ Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal Customal	Automatic	Clerical
Ticks skipped between labels     Major Ticks     Image: Style Outside Image: Style	C Custom	Custodial Manager
Major Ticks   ✓ Display ticks   Style Outside    Minor Ticks   ✓ Display ticks   Style    ✓ Display ticks   ✓ D	Ticks skipped between labels	
Major Ticks Style Outside Minor Ticks Style Mumber of minor ticks per major ticks: Umber of minor ticks per major ticks: Mumber of minor ticks: Mumbe		
Image: Style Outside Image: Style Outside Image: Style Image: Sty	Maior Ticks	
Style     Outside       Minor Ticks       Display ticks       Style       Number of minor ticks per major ticks:	✓ Display ticks	Excluded:
Minor Ticks Display ticks Style Lower margin (%): 5 Upper margin (%): 5	Style Outside V	t
Minor Ticks       Display ticks       Style       Number of minor ticks per major ticks:         Image: Style         Image: Style <td></td> <td></td>		
Display ticks       Style       Number of minor ticks per major ticks:         Image: Style         Ima	-Minor Ticks-	
Style	Display ticks	
Number of minor ticks per major ticks:	Style	Lower margin (%): 5 Upper margin (%): 5
	Number of minor ticks per major ticks:	- , - , -
Opply Class Hala		
Show Close Leib Show Close Leib	Apply Close Help	Apply Close Help



#### <= Initial Bar Chart

#### Modified Bar Chart =>

#### Y-AXIS:

- 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.



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.

Employment Categor

# 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.

Compute Variable			🗰 Road constr	uction b	oids.sav - Sl	PSS Data Ec	litor									ſ
get Variable:	Numeric Expression:		Eile Edit View	Data ∏r	ansform Ana	alyze <u>G</u> raphs	Utilities Ad	d-ons <u>W</u> indo∙	v <u>H</u> elp							
st_day =	dotest/dayses	~			> 🏪 📴	<u># 1</u>		<u>s so</u>								
1			14 : btpratio		0.7											
Type & Label			ob	s	cost	dotest	b2b1rat	b3b1rat	bhb1rat	status	district	btpratio	daysest	cost_day	var	
		~	1	1	1379.43	1386.29	1.01397	1.03303	1.06121	1	0	.33333	250	5.55		-
Observation [obs]	ľ	Eurotion group:	2	- 2	202.33	248.89	1.12084	1.22498	1.30546		0	.7 5000	120	2.07		+
Contract Cost [cost]		r unction group.	4	4	397.12	467.49	1.00588	1.11035	1.26733	0	0	.50000	180	2.60		t
Contract Cost [Cost]	+ < > 7 8 9	All 🔨	5	5	158.54	117.72	1.01053	1.10247	1.10247	1	0	.37500	80	1.47		
DUT Engineer's Estima	· <= >= 4 5 6	Arithmetic	6	6	1128.11	1008.91	1.06208	1.09137	1.09137	1	0	.60000	200	5.04		1
Ratio of Second Lowe		CDF & Noncentral CDF	/	/	400.33	4/2.98	1.10275	1.13560	1.13560	1	1	.60000	200	<u> </u>		-
Batio of Third Lowest F	× = ~= 1 2 3	L'onversion	9	9	353.96	370.02	1.05063	1.28312	1.47836	0	1	.57143	200	4.93		+
		Current Date/Time	10	10	138.71	174.25	1.07047	1.19279	1.27559	0	0	.83333	70	2.49		+
Hatio of Highest Bid to		Date Arithmetic	11	11	383.66	410.95	1.07508	1.13970	1.13970	1	1	.42857	60	6.85		
Contract Status [status	** ~ () Delete	Date Creation	12	12	3910.94	3405.94	1.02768	1.04733	1.07683	1	1	.45455	350	9.73		_
District [district]		Date Extraction	13	13	362.92 196.60	385.96	1.01691	1.04658	1.04668	0	1	.37500	100	3.86		+
Datio of Number of Did		Functions and Constal Mariables	14	15	637.99	627.41	1.16356	1.19491	1.62552	0	0	50000	120	4.48		+
	<u>^</u>	Functions and Special Variables:	16	16	152.06	175.40	1.07504	1.24451	1.24451	1	1	.50000	75	2.34		t
Engineer's Estimate of			17	17	375.00	432.33	1.05025	1.20642	1.30949	0	0	.57143	120	3.60		
			18	18	2284.56	1499.04	1.01600	1.20033	1.20033	1	0	.60000	270	5.55		1
			19	19	551.45 120.67	497.74	1.05568	1.10932	1.10932	1	1	.60000	100	4.98		+
			23	20	207.87	167.99	1.05143	1.08977	1.15240	1	1	.66667	60	2.80		H
			22	22	640.48	767.80	1.06059	1.08447	1.27066	0	0	.40000	90	8.53		t
			23	23	230.54	260.30	1.11029	1.12570	1.12570	1	1	.42857	125	2.08		
			24	24	299.87	247.04	1.08411	1.10180	1.10180	1	1	.60000	80	3.09		1
			25	25	2368.84	2456.77 879.40	1.17209	1.18020	1.48550	0	0	.30769	320	6.28		+
			20	27	1564.87	1303.40	1.00374	1.04983	1.04983	1	0	.33333	200	6.52		+
1		1	28	28	7387.03	6107.93	1.01878	1.05413	1.05718	0	1	.66667	340	17.96		t
(optional case selection o	condition)		29	29	195.68	199.09	1.04290	1.27466	1.27466	0	1	.60000	50	3.98		
		1	30	30	830.47	715.46	1.01755	1.02450	1.08833	1	0	.57143	135	5.30		1
			31	311	179.06	208.72	1.02474	1.03067	1.60580	0	i 11	.62500	90	2.32		

# **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**<sup>1</sup>. 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.

🛅 Untit	🖩 Untitled - SPSS Data Editor 🔹 🔲 🛛 🔀										
<u>File E</u> dit	⊻iew <u>D</u> ata	<u>T</u> ransform &	<u>A</u> nalyze <u>G</u> raph	ns <u>U</u> tilities A	idd- <u>o</u> ns <u>W</u> indo	ow <u>H</u> elp					
2	ð 🖳 🗠	o 🖂 🔚 🛙	? 颅 唱!	t 📰 🏚	<b>F</b> 🏹						
1:											
	var	var	var	var	var	var	var	var	var 🔺		
1											
2											
3											
4											
5											
7											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21			,						<b>•</b>		
<u> </u>	ata View <u>{</u> ∀	ariable View	/								
	SPSS Processor is ready										

# 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.

Open File		? 🛛
Look jn:	C SPSS	· ← 🗈 📸
Recent Desktop My Documents My Computer	Tutorial 2h_cn 2h_tw 1991 U.S. General Social Survey.sav AMI survival.sav anorectic.sav Anxiety 2.sav Anxiety 2.sav Breast cancer survival.sav Carpet.sav Coronary artery data.sav Employee data.sav Employee data.sav Fat surfactant.sav	flying.sav Growth study.sav Growth study.sav Marking.dat.sav Kinship_dat.sav Kinship_dat.sav Kinship_var.sav Type: SP5S Data Document Date Modified: 8/26/2002 3:42 PM Size: 1.37 KB
My Network Places	File name: Files of type: SPSS (".sav)	✓     ✓     ✓     ✓     Paste     Cancel

#### 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.

<sup>1</sup> 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.

🛗 Untit	led - SPSS D	ata Editor								(	
<u>Eile E</u> dit	<u>V</u> iew <u>D</u> ata	Iransform @	analyze <u>G</u> raph	15	Utilities Add-ç	ons <u>W</u> indow	Help				
<b>2</b>	a 🔍 🗠	o 🔚	2 M M	r <b>i</b> tte		<b>\$</b> 0					
7 : v1		3				Lice St	-+-				
	semester	v1	v2	vЗ	٧4	1050 0	v6	Ι	v7	v8	v9 🔺
1	200109.0	4	2	А	2	3		1	3	1	-
2	200109.0	4	2	В	3	3		1	2	1	
3	200109.0	4	2	А	2	3		1	3	1	
4	200109.0	4	2	А	3	3		1	3	1	
5	200109.0	3	2	А	2	3		2	3	1	
6	200109.0	4	2	А	3	3		1	3	1	
7	200109.0	3	1	А	2	3		1	2	1	
8	200109.0	4	2	С		3		2	1	1	
9	200109.0	4	1	В	3	3			2	3	
10	200109.0	4	1	В		2			2	1	
11	200109.0	4	2	С	3	2		2	3	1	
12	200109.0	3	2	В	3	2		2	2	1	
13	200109.0	3	2	А	1	2		2	2	1	
14	200109.0	2	2	С	4	3		2	3	3	
15	200109.0	4	2	А	2	3		1	3	1	
16	200109.0	3	1	В	3	2		1	3	1	
17	200109.0	3	2	в	3	3		2	2	2	
18	200109.0	2	2	В	3	3		2	3	1	
19	200109.0	3	2	А	1	3		1	2	1	
20	200109.0	3	2	А	2	2		1	2	1	
21	200109.0	4	2	в	3	2		2	2	1	-
<   ►  \D	ata View 🖌 🗸	ariable View	/		-	•					
Use Sets			5	PSS	Processor is re	eady					

Defining variable v1, renamed 'CLASS\_YEAR', and associated frequency table.

ĺ	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
I	semester	Numeric	6	0		None	None	8	Right	Nominal
2	Class Year	Numeric	1	0	Class Year of Respondent	(1, Freshman).	8-9.0	4	Center	Nominal
3	v2	Numeric	1	0		None	None	8	Right	Nominal
4	v3	String	1	0		None	None	2	Left	Nominal
5	v4	Numeric	1	0		None	None	8	Right	Nominal
6	v5	Numeric	1	0		None	None	8	Right	Nominal
7	v6	Numeric	1	0		None	None	8	Right	Nominal
8	v7	Numeric	1	0		None	None	8	Right	Nominal
9	v6	Numeric	1	0		None	None	8	Right	Nominal
10	v9	Numeric	1	0		None	None	8	Right	Nominal
11	v10	Numeric	1	0		None	None	8	Right	Nominal
12										
13										
14										
16										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										

		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 Respon	dent	
		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		

83

**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.
- <u>**Type**</u>: 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.)
- <u>Width & Decimals</u>: 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.
- <u>Variable Label</u>: 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.
- <u>Value Labels</u>: 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 ADDed 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.
- <u>Missing Values</u>: 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.

- <u>Columns</u>: 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.
- <u>Measure</u>: 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.

Variable Type		? 🛛
Numeric     Domma     Dot     Scientific notation     Date     Dollar     Custom currency     String	<u>W</u> idth: <mark>6</mark> Decimal <u>P</u> laces: 0	OK Cancel Help



Missing Values	? 🗙
C No missing values	ОК
C Discrete missing values	Cancel
	Help
<u>Bange plus one optional discrete missing v</u>	alue
Low: 4 <u>H</u> igh: 9	
Di <u>s</u> crete value: 0	

# **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 <u>Chart Editor</u> => 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.



# 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							Va
			Measurement						Value		
Variable	Position	Label	Level	Column Width	Alignment	Print Format	Write Format		origin	1	Α
mpg	1	Miles per Gallon	Scale	8	Right	F4	F4			2	E
engine		Engine								3	Ja
engine	2	Displacement	Scale	8	Right	F5	F5		cylinder	3	3
		(cu. inches)								4	4
weight	3	Vehicle	Scale	8	Right	F4	F4			5	5
		Weight (lbs.)			8					6	6
origin	4	Country of Origin	Nominal	8	Right	F1	F1	-		8	8
cylinder	5	Number of Cylinders	Ordinal	8	Right	F1	F1	_			

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**.)

🛅 Empl	🖩 Employee data.sav - SPSS Data Editor 🔹 🔲 🗖 🔀													
<u>File E</u> dit	<u>File Edit View D</u> ata <u>T</u> ransform <u>A</u> nalyze <u>G</u> raphs <u>U</u> tilities Add- <u>o</u> ns <u>W</u> indow <u>H</u> elp													
<b>2</b>	<b>619</b> <u>1</u>													
4 : edu	4:educ 8													
	id	g e n	bdate	VAR00001	educ	jobcat	salary	salbegin	jobtime					
1	1	m	02/03/1952		15	3	\$57,000	\$27,000	98					
2	2	m	05/23/1958		16	1	\$40,200	\$18,750	98					
3	3	f	07/26/1929		12	1	\$21,450	\$12,000	98					
4	4	f	04/15/1947		8	1	\$21,900	\$13,200	98					
5	5	m	02/09/1955		15	1	\$45,000	\$21,000	98					
6	6	m	08/22/1958		15	1	\$32,100	\$13,500	98					
7	7	m	04/26/1956		15	1	\$36,000	\$18,750	98					
8	8	f	05/06/1966		12	1	\$21,900	\$9,750	98					
9	9	f	01/23/1946		15	1	\$27,900	\$12,750	98					
10	10	f	02/13/1946		12	1	\$24,000	\$13,500	98					
11	11	f	02/07/1950		16	1	\$30,300	\$16,500	98					
<b>▲   ▶  </b> \0	ata View 🔏 🗸	ariable	01/11/1966		l <sup>e</sup>	1	00 0E0	£10 000	90 09	<b>_</b>				
				SPSS Proces	sor is ready									

### **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**.)

🛅 Emple	oyee data.sa	w - Si	PSS Data Editor							X				
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	<u>T</u> rans	sform <u>A</u> nalyze <u>G</u> r	aphs <u>U</u> tilities	Add- <u>o</u> ns 🛛	/indow <u>H</u> elp								
8:bda	6 : bdate 04/26/1956													
	id	g e n	bdate	minority	educ	jobcat	salary	salbegin	jobtime					
1	1	m	02/03/1952	0	15	3	\$57,000	\$27,000	98	Π				
2	2	m	05/23/1958	0	16	1	\$40,200	\$18,750	98	Π				
3	3	f	07/26/1929	0	12	1	\$21,450	\$12,000	98	Γ				
4	4	f	04/15/1947	0	8	1	\$21,900	\$13,200	98	Г				
5	5	m	02/09/1955	0	15	1	\$45,000	\$21,000	98	Π				
6	6	m	08/22/1958	0	15	1	\$32,100	\$13,500	98	Π				
7														
8	7	m	04/26/1956	0	15	1	\$36,000	\$18,750	98	Π				
9	8	f	05/06/1966	0	12	1	\$21,900	\$9,750	98	Π				
10	9	f	01/23/1946	0	15	1	\$27,900	\$12,750	98	Π				
11	10	f	02/13/1946	0	12	1	\$24,000	\$13,500	98	Τ_				
	ata View 🔏 🗸	f ariable	02/07/1950	0	16	1	¢20 200	\$16 ANN	09 •	Ī				
				SPSS Proces	sor is ready					1				

# **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 <u>Pair</u> 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.



Add Variables fromesidence Life\res2	00102.sav	
Excluded Variables: bldg_typ (+) dorm (+) gender (+) v1 (+) v10 (+) v11 (+) v13 (+) Rename Match cases on key variables in sorted files Both files provide cases External file is keyed table External file is keyed table Working Data File is keyed table [Indicate case source as variable: source01 (") = Working Data File (+) =esidence Life keyed	New Working Data File:           bldg_typ (*)           dorm (*)           year_M(*)           gender (*)           V1 (*)           V2 (*)           V3A (*)           V7a (*)           V8 (*)           V8 (*)           vaiables:	OK <u>P</u> aste <u>B</u> eset Cancel Help

# **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"
Cutiput] - SPSS Viewer  Constant in the interviewer i	Open File   Look in:   SPSS datafiles   SPSS 12 Manuals   SPSS FILES to sort   Student Affairs   Title III   UUP Basic   middle states.spo
Image: specific display="2">Preventioned Percentioned Percentide Percentide Percentioned Percentioned Percentioned Percention	File name:

# **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).

	Employme	ent Category * Gender C	rosstabula	tion	
			Gei	nder	_
			Female	Male	Total
Employment	Clerical	Count	206	157	363
Category		% 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
Total		% 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	Clerical	206	157	363
	Category	Custodial	0	27	27
		Manager	10	74	84
	Total		216	258	474
% within Employment	Employment	Clerical	56.7%	43.3%	100.0%
Category	Category	Custodial	.0%	100.0%	100.0%
		Manager	11.9%	88.1%	100.0%
	Total		45.6%	54.4%	100.0%
% within Gender	Employment	Clerical	95.4%	60.9%	76.6%
	Category	Custodial	.0%	10.5%	5.7%
		Manager	4.6%	28.7%	17.7%
	Total		100.0%	100.0%	100.0%
% of Total	Employment	Clerical	43.5%	33.1%	76.6%
	Category	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:**

• <u>Into Same Variables</u>: 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).

**Old and New Values Dialog Box** 

• <u>Into Different Variables</u>: 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 Same Variables	Recode into Different Variables	Recode into Different Variables: Old and New Values
Numeric <u>Variables:</u> ©K     © Employee Code [id]     © Date of Birth [bdate]     © Educational Level (yes     © Employment Category     © Beginning Salary (salar)     © Months since Hire [iob     Previous Experience (i     Minority Classification          Id. (optional case selection condition)	Employee Code [id]     Gender [gender]     Date of Bith [bdate]     Educational Level (per     Engloyment Category     Beginning Salay [sab     Months since Hire [ide     Previous Experience (     Minority Classification     [i (optional case selection condition)     [i (optional case selection condition)	Old Value       New Value         ✓ Yalue:       ✓ Yalue:         ✓ System-missing       ✓ Yalue:         ✓ System-or yser-missing       C Copy old value(s)         ✓ Barge:       Uld → New:         Øld → New:       Øld → New:         Øld → N

**Recode Into Different Variables 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 "Salary-?" 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.)

## **OTHER RECODE CONSIDERATIONS:**

🛅 Emple	🛗 Employee data.sav - SPSS Data Editor											
<u>Eile E</u> dit	<u>File E</u> dit <u>V</u> iew <u>D</u> ata Iransform <u>A</u> nalyze <u>G</u> raphs <u>U</u> tilities Add- <u>o</u> ns <u>W</u> indow <u>H</u> elp											
<b>2</b>	e e e e e e e e e e e e e e e e e e e											
1:id	1:id 1											
	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						
	nta View 🔏 🗸	ariable View /	/ 00	4.40								
			S	PSS Processor	is ready							

- 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 (cannot rev Recode: 1 = 1; 3	into Same Variable erse recode) = 1; 5 = 5; 7 = 5	Five Cases Recoded (can rever Recode: 1 = 1; 3	ses Recoded into Same Variable (can reverse recode)Five Cases Recoded into D (saves both value)ode: 1 = 1; 3 = 2; 5 = 3; 7 = 4Recode: 1 = 1; 3 = 1;			Variable = 5
(Before)	(After)	(Before)	(After)	(Before)	(A)	fter)
CLASS	CLASS	CLASS	CLASS	CLASS	CLASS	CLASS2
1	1	1	1	1	1	1
1	1	1	1	1	1	1
3	1	3	2	3	3	1
5	5	5	3	5	5	5
7	5	7	4	7	7	5
(many to one	relationship)	(one-to-one va	lue relationship)			

# 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.	<b>Dialog Box:</b> Condition is specified.	Data Editor showing selected and non-selected cases.
Select Cases  Select	Select Cases: If            Employee Code [id]	Imployee data.sav - SPSS Data Editor       Imployee data.sav - SPSS Data Editor         Ele Edit yiew Data Transform Analyze graphs Utilities Addrons Window Help         I id       I         I id       I id         I

# **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.

Fort Cases		×	Empl	o <mark>yee data.sav - S</mark> <u>V</u> iew <u>D</u> ata <u>T</u> rans	P <mark>SS Data Editor</mark> sform <u>A</u> nalyze <u>G</u> ra	aphs <u>U</u> tilities /	Add- <u>o</u> ns <u>W</u> indo	low <u>H</u> elp			
Employee Code [id]	v: ate of Birth [bdate] Asc	ок	<b>≩</b> .∎ 7:id	🞒 💷 🖂	▶ 😥 🛤 🔸 285	minority	educ	iohcat	ealary	salhenin	inhtime
<ul> <li>Minority Classification</li> <li>Educational Level I</li> <li>Employment Catego</li> <li>Current Salary [salation]</li> <li>Beginning Salary [salation]</li> <li>Months since Hire I</li> </ul>	Order scending escending	Paste Reset Cancel Help	1 2 3 4 4 5 6 7 8	434 m 443 f 152 m 3 f 171 m 335 m 285 m 108 f	02/10/1929 05/03/1929 07/26/1929 01/21/1930 02/26/1930 05/28/1930 07/16/1930		16 8 12 12 12 8 8 8 8 12	1 1 2 1 1 2 2 2 1	\$34,950 \$21,600 \$30,750 \$21,450 \$26,700 \$31,950 \$30,750 \$21,000	\$20,250 \$13,500 \$15,000 \$12,000 \$13,500 \$15,750 \$15,750 \$11,550	66         66           66         87           98         86           74         78           91         91
			9 10 11	453 m 185 m 378 f	08/07/1930 08/28/1930 09/21/1930	0	15 8 8	1 2 1	\$24,450 \$30,750 \$15,750	\$15,750 \$15,000 \$10,200	65 84 70

### **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 Edit 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.

🛗 Outpu	t3 - SPSS V	iewer									
<u>File E</u> dit	⊻iew <u>D</u> ata	Transform	Insert	Format	<u>A</u> nalyze <u>G</u> raphs <u>U</u>	<u>J</u> tilities Ad	dd- <u>o</u> ns <u>W</u> ind	low <u>H</u> elp			
<b>2</b>   <b> </b>  {	3 🖪 词			- [? ] (	ا یے او						
+ +	+ - 🕮		ŇĎ	1							
	utput Deservinting		I	)escrip	ives						
	- 📄 Title	5									
	- 🔂 Notes - 👍 Descrip	tive Statistics	1-1			D	escriptive St	atistics			_
	Descriptives	s		Gender		N	Minimum	Maximum	Mean	Std. Deviation	-
				Female	Current Salary	216	\$15,750	\$58,125	\$26,031.92	\$7,558.021	
					Valid N (listwise)	216					
				Male	Current Salary	258	\$19,650	\$135,000	\$41,441.78	\$19,499.214	-
					Valid N (listwise)	258					
											-
<		>	<								
				9	SPSS Processor is re	eadv					
						,			1		1

SPSS Output resulting from selecting Compare groups:

SPSS Output resulting from selecting Organize output by groups:

Output3 - SPSS Viewer							
Eile Edit Yiew Data Iransform Insert	Format <u>A</u> nalyze (	Graphs	Utilities Add-	ons <u>W</u> indow	Help		
	- 🕞 🔊 🚛	0					
		<u>·</u>					
E Output (	jender = Female	•					
Descriptives							
Notes			Descriptiv	e Statistics <sup>a</sup>			
Gender = Female		N	Minimum	Maximum	Mean	Std. Deviation	
Descriptive Statis	Current Salary	216	\$15,750	\$58,125	\$26,031.92	\$7,558.021	
E Gender = Male	Valid N (listwise)	216					
Descriptive Statis	a. Gender = Fem	ale					
-w .							
	<sup>1</sup> ondor – Molo						
	Sender – Ivraie						
1-1							
···			Descriptiv	e Statistics <sup>a</sup>			
		N	Minimum	Maximum	Mean	Std. Deviation	
	Current Salary	258	\$19,650	\$135,000	\$41,441.78	\$19,499.214	
	Valid N (listwise)	258					
	a. Gender = Mal	e					
< > > <							>
	📍 SPSS Proc	essor is r	ready				

🔜 Split File		
<ul> <li>Employee Code [id]</li> <li>Date of Birth [bdate]</li> <li>Educational Level (yea</li> <li>Employment Category  </li> <li>Current Salary [salary</li> <li>Beginning Salary [salary</li> <li>Months since Hire [job]</li> <li>Previous Experience (r</li> <li>Minority Classification [</li> </ul>	Analyze all cases, do not create groups     Compare groups     Diganize output by groups     Groups Based on:     Area and a strength of the file by grouping variables     File is already sorted	OK Paste Beset Cancel Help
Current Status: Analysis by	groups is off.	

# 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:**

- <u>TableLooks</u>: 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.
- <u>**Table Properties:</u>** 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.</u>
- <u>Cell Properties</u>: 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
- <u>Caption</u>: 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.
- <u>Footnote</u>: 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.

In	itial Ta	ble Format	t (system de	fault):		
			Employ	yment Categ	Jory	
			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	

#### 🖬 Output2 - SPSS Viewer File Edit View Insert Pivot Format Analyze Graphs Utilities Add-ons Window Help E Output Frequencies E Fre Employment Category A state Cumulative Frequency Percent Valid Percent Percent Valid Clerical 363 76.6 76.6 76.6 Custodial 27 5.7 5.7 82.3 17.7 17.7 100.0 Manager 84 Total 474 100.0 100.0 Litems selected (0 hidden/collapsed) 👎 SPSS Processor is ready

# Table after above modifications:

### Employment Category<sup>a</sup>

		Frequency	Percent	Valid Percent	Cumulative Percent
	Clerical	363	76.6	76.6	76.6
\$7.11.1	Custodial	27	5.7	5.7	82.3
valid	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. ٠
- Step 2: Delimited vs. Fixed Format and Variable Name present both can be determined by looking at the raw data file. ٠
- Step 3: Identify where the data begin in the file and how many lines (Records) there are per Case. ٠
- Step 4: Insert breaks between the variables. ٠
- Step 5: Identify variable names and variable types (both optional). ٠
- Step 6: Finish. (Note: Here is where a text data format can be saved for future use.) ٠

### **EXAMPLE:** Open the text data file **SPSS\_SPLtxt**.

#### **Text Data:** (Partial listing) .

E SPSS_SPL.txt - WordPad		
	• Step1: The Text Wizard shows a small sample or the	he text data. Generally select "No" for the predefined format
	question.	
2UU3U6U246CSC12U371MATHSS31A322113233444443444 2003060246CSC12U371MATHSS41B21211124444444344	Taut Impact Wirard Step 1 of 6	
2003060246CSCI20371MATHSS41A221111233344333445	Text import wizard - step 1 of 6	
2003060111PHYS10315PHYSS542B222212843434.44444	Welcome to the text import wizard	• Step 2: Select Delimited vs. Fixed Format and if variable
2003060111PHYS10315PHYSSS.2B.322121433443333334	628 840 1 81 28.5 630 2400 0 73 40.33	names are included in the file (Stops 3 & 4 will be
2003060111PHYS10315PHYSSS62B222313343343433444	632 10200 0 83 31.08 This wizard will help you read data from your text file and	names are included in the me. (Steps 5 & 4 win be
2003060111PHT510315PHT55542&1212131433443434444	635 1740 83 41.91 specify information about the variables.	slightly different if Delimited Format is selected.)
2003060111PHYS10315PHYSSS42A231212134333433345		
2003060111PHYS10315PHYSSS32B222212144444444445		Text Import Wizard - Step 2 of 6
2003060111PHYS10315PHYSSS41B.21212144444443445	Does your text file match a predefined format?	
2003060111PHYS10315PHYS5S42B321112144444444445	Proven	
2003060158ELIT27015AENGLHA32A221213244444444445		How are your variables arranged?
2003060158ELIT27015AENGLHA32A132213144444444445	e Ng	C Delimited
2003060158ELIT27015ENGLHA22B232312144444444445		
2003060158ELIT27015ENGLHA32B233212144444444445		<ul> <li>Eixed widthiger - Variables are aligned in fixed width columns.</li> </ul>
2003060158ELTT27015ENGLHA42C322332144444433435		
2003060158ELT127015ENGLH&42.212111144444444445	Text file: C\Documents and Settings\iobnsosd\Deskton\SPSS_SPLtxt	
2003060158ELIT27015ENGLHA22B333213144444444445		Are variable names included at the top of your file?
	0 10 20 30 40 50	
For Help, press F1 NUM	1 2003060246CSCI20371MATHSS31A322113233444443444	· Tez
	2 2003060246CSCI20371MATHSS41B212111244444444444444	
	3 2003060246CSCI20371MATHS541A221111233344333445	
	4 2003060111PHYS10315PHYSSS42B222212843434.44444 🗸	
		Text file: C:\Documents and Settings\johnsosd\Desktop\SPSS_SPI.txt

Cancel

Help

<u>N</u>ext>

Help

Cancel

2003060246CSCI20371MATHSS31A322113233444443444

2003060246CSCI20371MATHSS41B21211124444444344. 3 2003060246CSCI20371MATHS541A221111233344333445 4 2003060111PHYS10315PHYSS542B222212843434.44444

<u>N</u>ext >

1

2

•

< <u>B</u>ack

Step3: Identify the number of data lines (Records) per case and how much of the data to be imported. ٠

The first case of data begins on which line number?   Text Import Wizard - Fixed Width Stap 4 of 6 For the vertical lines in the data preview represent the basedpoints between variables. To Motion Cases: The vertical lines in the data preview represent the basedpoints. To NODERT a variable treak line, dag to to the detered position. To NODERT a variable treak line, dag to to the date preview area. The vertical lines in the data preview area. To NODERT a variable treak line, dag to to the date preview area. To NODERT a variable treak line, dag to to the date preview area. Data preview Data preview Quest 10000 Cases: Quest 1	Text Import Wizard - Fixed Width Step 3 of 6	• Step 4: For Fixed Format data, as here, insert divid	ers between variables.
How many cases do you want to mport?	The first case of data begins on which line number?	Text Import Wizard - Fixed Width Step 4 of 6	• Step 5: Assign variable names and types – both option as they may be added in the Data Editor.
<ul> <li>C MODIFY a variable break line, deg it to the desired position.</li> <li>To MODIFY a variable break line, deg it to the desired position.</li> <li>To INSERT a variable break line, deg it out of the data preview area.</li> <li>Data preview</li> <li>0 a preview</li></ul>	How many cases do you want to import?	The vertical lines in the data preview represent the breakpoints between variables.	Text Import Wizard - Step 5 of 6
<ul> <li>     The first 1000 cases.       <ul> <li>A generatage of the cases: 10</li></ul></li></ul>	All of the cases	<ul> <li>To MODIFY a variable break line, drag it to the desired position.</li> </ul>	
A gencentage of the case:       0 → %         Oda preview       - To DELETE a variable break line, drag it out of the data preview area.         0 da preview       - To DELETE a variable break line, drag it out of the data preview area.         0 da preview       - To DELETE a variable break line, drag it out of the data preview area.         0 da preview       - To DELETE a variable break line, drag it out of the data preview area.         0 da preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview       - To DELETE a variable break line, drag it out of the data preview area.         0 data preview	C Ihe first 1000 cases.	<ul> <li>To INSERT a variable break line, click at the desired position.</li> </ul>	Specifications for variable(s) selected in the data preview
Data preview         10       20       30       40       50         2       2003 6024 46 CS C120371 INATHSS 31 JA3221 J1323 94444493444       1       1         2       2003 6024 46 CS C120371 INATHSS 91 JA221 J11233 9443 934 44       1       1       2003 6024 46 CS C120371 INATHSS 91 JA221 J1123 93 443 93 94 95       1       2003 602 46 CS C120371 INATHSS 94 JA221 J1123 93 443 93 44 95       1       2003 602 46 CS C120371 INATHSS 94 JA221 J112 23 94 493 94 94 4       1       1       2003 602 24 6CS C120371 INATHSS 94 JA221 J112 23 94 494 94 94 4       1       1       2003 602 24 6CS C120371 INATHSS 94 JA221 J112 23 94 494 94 94 4       1       1       2003 602 24 6CS C120371 INATHSS 94 JA221 J112 23 94 494 94 94 4       1       1       2003 602 24 6CS C120371 INATHSS 94 JA221 J112 23 94 494 94 94 4       1       1       2003 602 24 6CS C120371 INATHSS 94 JA221 J112 23 94 494 94 94 4       1       1       2003 602 24 6CS C120371 INATHSS 94 JA221 J112 23 94 494 94 94 4       1       1       2003 602 46 SC C120371 INATHSS 94 JA221 J112 23 94 94 94 94 94 4       1       1       2003 602 46 SC C120371 INATHSS 94 JA221 J112 J23 94 94 94 94 94 4       1       1       2003 602 46 SC C120371 INATHSS 94 JA221 J112 J23 94 94 94 94 94 4       1       1       2003 602 46 SC C120371 INATHSS 94 JA22 JA2 3       1       2003 602 4 6 SC CS C20371 INATHSS 94 JA22 JA2 3       1       2003 96 02 4 6 SC CS C20371 INATHSS 94 JA2 3       1       2003 96 02 4 6 SC CS C20371 INA	C A percentage of the cases: 10 📫 %	- To DELETE a variable break line, drag it out of the data preview area.	<u>V</u> ariable name:
Data preview         1       2003060246CSC120371MATHSS31A322113233444449344.         2003060246CSC120371MATHSS41A22111234444449344.         2003060246CSC120371MATHSS41A22111234444449344.         2003060216CSC120371MATHSS41A22111234444449344.         2003060216CSC120371MATHSS41A2211123444444944.         2003060211PHYS1031SPHYSSS42B222212843434.444444         4         20030602146CSC120371MATHSS41A22111223944943443444         2003060246CSC120371MATHSS91AB2D2111D2054949549445         2003060246CSC120371MATHSS91AB2D211D2054949589498         2003060246CSC120371MATHSS91AB2D211D2058949589498         2003060214CSC120371MATHSS91AB2D211D2058949589498         2003060211PHYS1031SPHYSSS42B22212843444.14444         2003060211PHYS1031SPHYSSS42B2221212919494944.44444         2003060214CSC120371MATHS91AB2D211D2058940589485         2003060211PHYS1031SPHYSS842B222112919494944.44444         2003060211PHYS1031SPHYSS42B222112919494944.44444         4       2003060214CSC120371MATHS941A221111205894059445         2003060211PHYS1031SPHYSS42B222112919494944         4       2003060211PHYS1031SPHYSS42B222112920494544444444444444444444444444444444			semester
< Back       Next>       Finish       Cancel       Help         < Back       Next>       Finish       Cancel       Help         < Back       Next>       Finish       Cancel       Help	Data preview         0         10         20         30         40         50           1         D003060246CSC120371MATHSS31A32211323344443444         4         4         4           2         D003060246CSC120371MATHSS41B212111244444444444         4         4         4           3         D003060246CSC120371MATHSS41B222111233344333445         4         4         4         4	Data preview           0         10         20         30         40         50           1         2003060246CSC120371MATHSSb1Ab22b113b2354A94A94A94A9         ▲         2003060246CSC120371MATHSSb1Ab22b113b2594A94A94A94A94A94A94A94A94A94A94A94A94A9	Data format: Numeric
<u>Back Next&gt;</u> <u>Finish Cancel Help</u> <u>Concel Help</u>	C Back Nexts Finish Cancel Help		semester V2 V3 V4 V5 V6 VA
K         Rack         Next>         Finish         Cancel         Help			20030602 4 6 CSCI20371 MATHSS 3 1 -
Kext>         Finish         Cancel         Help           20030601         1         1         PHYS10315         PHYSSS         4         2		· · · · · · · · · · · · · · · · · · ·	20030602 4 6 CSCI20371 MATHSS 4 1
		< <u>B</u> ack <u>Next</u> Finish Cancel Help	20030601 1 1 PHYS10315 PHYSSS 4 2

Step 6: Finish. (Here the format created via the Text Wizard may be saved for future use.) ٠



#### SPSS Data Editor after data import: •

🛅 spi20	0306.sa	v - SPSS Data	Editor									
<u>File E</u> dit	⊻iew D	ata <u>T</u> ransform	Analyze	Graphs U	tilities Ad	d- <u>o</u> ns <u>W</u> in	idow <u>H</u> elp					
<b>6</b>	<b>s</b> 🔍	<b>n</b> 🗠 🔚	i? #	「」	⊞  <b>⊕ </b> F	<b>,</b> 📢 🤇	<b>)</b>					
8 : divi	sion	:	SS									
	cm	course	dept	division	q1	q2	q3	q4	q5	q6	q7	<b>_</b>
1	246	CSCI20371	MATH	SS	3	1	A	3	2	2	1	
2	246	CSCI20371	MATH	SS	4	1	В	2	1	2	1	
3	246	CSCI20371	MATH	SS	4	1	A	2	2	1	1	
4	111	PHYS10315	PHYS	SS	4	2	В	2	2	2	2	
5	111	PHYS10315	PHYS	SS		2	В		3	2	2	
6	111	PHYS10315	PHYS	SS	6	2	В	2	2	2	3	$\square$
7	111	PHYS10315	PHYS	SS	4	2	A	1	2	1	2	
8	111	PHYS10315	PHYS	SS	3	2	С	3	3	2	2	
9	111	PHYS10315	PHYS	SS	4	2	A	2	3	1	2	
10	111	PHYS10315	PHYS	SS	3	2	В	2	2	2	2	
11	111	PHYS10315	PHYS	SS	4	1	В		2	1	2	
12	111	PHYS10315	PHYS	SS	4	2	В	3	2	1	1	
13	158	ELIT27015	ENGL	HA	4	2	С	3	3	2	3	
14	158	ELIT27015	ENGL	HA	3	2	A	2	2	1	2	
15	158	ELIT27015	ENGL	HA	3	2	А	1	3	2	2	
16	158	ELIT27015	ENGL	HA	2	2	в	2	3	2	3	$\square$
17	158	ELIT27015	ENGL	HA	3	2	В	2	3	3	2	$\square$
18	158	ELIT27015	ENGL	HA	4	2	С	3	2	2	3	$\square$
19	158	ELIT27015	ENGL	HA	4	2	в	3	2	3	2	$\vdash$
20	158	ELIT27015	ENGL	HA	4	2		2	1	2	1	$\square$
21	158	ELIT27015	ENGL	HA	2	2	в	3	3	3	2	Τ.
< D  \ NDa	nta View	Variable View	v 7		-	1		1			1	Πľ
		A	. /	SDSS D	processor is	ready						

Numeric		•					
) ata preview -							
) ata preview semester	V2	V3	V4	V5	V6		1^
) ata preview semester 20030602	<b>V2</b>	<b>V3</b>	V4 CSCI20371	V5 MATHSS	<b>V6</b>	1	-
) ata preview semester 20030602 20030602	<b>V2</b> 4 4	6 6	V4 CSCI20371 CSCI20371	V5 MATHSS MATHSS	3 4	1	1^
ata preview semester 20030602 20030602 20030602	<b>V2</b> 4 4 4	<b>V3</b> 6 6 6	V4 CSCI20371 CSCI20371 CSCI20371	V5 MATHSS MATHSS MATHSS	3 4 4	1 1 1	<u>`</u>
ata preview           semester           20030602           20030602           20030602           20030602           20030602           20030602	<b>V2</b> 4 4 4 1	6 6 6 1	V4 CSCI20371 CSCI20371 CSCI20371 CSCI20371 PHYS10315	V5 MATHSS MATHSS MATHSS PHYSSS	3 3 4 4 4	1 1 1 2	1

# **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

									Value		Label
			Vari	able Information:	Cars.sav				year	0 <sup>a</sup>	0 (Missing)
			Measurement						·	70	70
Variable	Position	Label	Level	Column Width	Alignment	Print Format	Write Format	Missing Values		71	71
mpg	1	Miles per Gallon	Scale	8	Right	F4	F4			72	72
engine		Engine								73	73
-	2	Displacement	Scale	8	Right	F5	F5			74	74
horse	2	(cu. menes)	Saala	o	Diaht	E5	E5			75	75
norse	3	Horsepower	Scale	8	Right	F5	F3			76	76
weight	4	Vehicle Weight (lbs.)	Scale	8	Right	F4	F4			77	77
accel		Time to								78	78
	5	Accelerate	Saala	Q	Dight	E4	E4			79	79
	5	from 0 to 60 mph (sec)	Scale	0	Rigin	Г4	Г4			80	80
vear		Model Year								81	81
yeur	6	(modulo 100)	Ordinal	8	Right	F2	F2	0		82	82
origin	7	Country of	Nominal	8	Pight	F1	F1		origin	1	American
	/	Origin	Nominai	0	Right	1.1	1.1			2	European
cylinder	8	Number of	Ordinal	8	Right	F1	F1			3	Japanese
	0	Cylinders	orunnur	0	Right				cylinder	3	3 Cylinders
filter_\$	0	$cylrec = 1 \mid$		0	D' 14	<b>F1</b>	F1			4	4 Cylinders
	9	cyrec = 2 (FILTER)	Ordinal	8	Right	FI	FI			5	5 Cylinders
Variables i	n the working	filo								6	6 Cylinders
v arrables r	ii uie workilig									8	8 Cylinders

filter\_\$ 0 Not Selected

Variable Values: Cars.sav

1 Selected

a. Missing value

# EMPLOYEE DATA.SAV

variable filler individual Elipitoyet Data.sav											
			Measurement						Value		Label
Variable	Position	Label	Level	Column Width	Alignment	Print Format	Write Format	Missing Values	gender	f	Female
id	1	Employee Code	Scale	8	Right	F4	F4			m	Male
gender	2	Gender	Nominal	1	Left	A1	A1		educ	$0^{a}$	0 (Missing)
bdate	3	Date of Birth	Scale	13	Right	ADATE10	ADATE10			8	8
educ	4	Educational	Ordinal	8	Right	F2	F2	0		12	12
		Level (years)								14	14
jobcat	5	Employment Category	Ordinal	8	Right	F1	F1	0		15	15
salary		Current		0	D: 1.	DOLLADO	DOLLADO	<b>*</b> 0		16	16
•	6	Salary	Scale	8	Right	DOLLAR8	DOLLAR8	\$0		17	17
salbegin	7	Beginning Salary	Scale	8	Right	DOLLAR8	DOLLAR8	\$0		18	18
iobtime		Months since								19	19
Jootinie	8	Hire	Scale	8	Right	F2	F2	0		20	20
prevexp		Previous								21	21
	9	Experience (months)	Scale	8	Right	F6	F6		jobcat	$0^{a}$	0 (Missing)
minority	10	Minority		0	D' 1/	<b>F1</b>	<b>F1</b>	0		1	Clerical
	10	Classification	Ordinal	8	Kignt	FI	F1	9		2	Custodial
Variables in the working file									3	Manager	

Variable Information: Employee Data.sav

#### Variable Values: Employee Data.sav

9<sup>a</sup> a. Missing value

\$0<sup>a</sup>

\$0<sup>a</sup>

 $0^{a}$ 

0

1

missing

missing

missing missing

9 (Missing)

No Yes

salary

salbegin

jobtime

prevexp

minority 0

# **ROAD CONSTRUCTION BIDS.SAV**

			Measurement				
Variable	Position	Label	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

#### Variable Information: Road Construction Bids.sav

### Variable ValuesRoad Construction Bids.sav

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

Variables in the working file

## WORLD95.SAV

			Measurement					
Variable	Position	Label	Level	Column Width	Alignment	Print Format	Write Format	Missing Values
country	1	<none></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	

Variable Information: World95.sav

Variables in the working file

### Variable Values: World95.sav

							Value		Label
							religion	а	missing
							region	1	OECD
	Va	wahla Tufawa atia						2	East Europe
	va	riable informatio	on: world95.sav (	cont.)				3	Pacific/Asia
		Measurement		A 11				4	Africa
	Label	Level	Column Width	Alignment	Print Format	Write Format		5	Middle East
	intake	Scale	8	Right	F6	F6		6	Latn America
	Aids cases	Scale	8	Right	F8	F8	climate	1	desert
	Birth rate per							2	arid / desert
	1000 people	Scale	8	Right	F5.1	F5.1		3	arid
	Death rate							5	tropical
	per 1000	Scale	8	Right	F6	F6		6	mediterranean
	people							7	maritime
	Number of							8	temperate
	aids cases / 100000	Scale	8	Right	F8.2	F8.2		9	arctic / temp
	people							10	arctic
	Log (base 10) of GDP_	Scale	8	Right	F8.2	F8.2	a. Mis	ssing val	ue

Ξ

Variable

calories

death\_rt

aids\_rt

log\_gdp

aids birth\_rt Position

13

14

15

16

17

18

# **INDEX**

#### Α

ACCESS SPSS, 8 ACCESSING SPSS, 8 Add Cases Merge Files, 89 Add Variables Merge Files, 89 ANALYSIS OF VARIANCE, 18 ANOVA Analysis of Variance, 18 Area Graph. *See* Frequency Polygon

#### В

Bar Chart Clustered, 21, 22, 23 Simple, 20 BAR CHART, 20 BINOMIAL PROBABILITIES, 24 BOX-AND-WHISKER PLOT, 27 Boxplot Grouped, 28 BOXPLOT, 27

#### С

Caption, 97 Cell Properties, 97 Chart Editor Bar Options – Bars & Clusters, 76 Bar Y-axis Scale, 77 Chart Categories, 75 Chart Depth & Angle, 76 Chart Fill & Borders, 75 Chart Size, 75 Chart Size, 75 Chart Titles, Footnotes, etc., 74 Chart Variables, 76 Data Label Mode, 79 Editing Text, 74 Exiting the Chart Editor, 74 Highlighting a Bar, 74

Moving around in a chart, 74 Properties Dialog Boxes, 75 Template, 79 Transform a chart, 74 Transpose, 74 X-axis Categories, 78 X-axis Labels & Ticks, 78 X-axis Properties, 78 Y-axis Labels & Ticks, 77 Y-axis Number Format, 77 Y-axis Properties, 77 CHART EDITOR, 73 Chi Square **TEST OF GOODNESS OF FIT, 29 TEST OF INDEPENDENCE, 31** CHI-SQUARE, 29 CLEAR Edit Menu (Data Editor), 85 COMPUTE, 80 CONFIDENCE INTERVAL FOR A **POPULATION MEAN. 32** CONFIDENCE INTERVAL FOR A **POPULATION PROPORTION, 33** CONTINGENCY TABLES see Crosstabs, 34 CONVENTIONS USED IN THE MANUAL. 5 Conventions used in this Manual **ADDITIONAL REQUIREMENTS, 5 OPTIONS**, 5 PATH. 5 **RELATED GRAPHS, 5** RELATED STATISITCAL PROCEDURES, 5 SUMMARY OF STEPS, 5 COPY Edit Menu (Data Editor), 85 Correlation. See Linear Correlation Correlation Coefficient. See Linear Correlation CROSSTABS. 34 CUT Edit Menu (Data Editor), 85

### D

DATA EDITOR. 82 Data File Descriptions finding, 6 Data File Information File - data descriptions, 87 DATA FILES, 6 **DEFINE VARIABLES, 83 Defining Variables** Data Editor, 84 **Descriptive Statistics** Mean, median, mode, range, variance, std. dev., quartiles, minimum, maximum. See via Descriptives, 38 via Frequencies. 37 **DESCRIPTIVE STATISTICS, 37 Display Data File Information**, 87 **DOCUMENT OVERVIEW, 5** DOTPLOT, 39

EDIT MENU, 85 ENTERING DATA INTO A NEW DATA FILE, 10 EXPORTING TABLES AND CHARTS, 86

F

Ε

FILE & VARIABLE INFORMATION, 87 File Information, 87 FIND Edit Menu (Data Editor), 85 Five-Number Summary, 27 FIVE-NUMBER SUMMARY, 40 Footnote, 97 frequencies, 20 frequency distributions, 20 FREQUENCY POLYGON, 41 FREQUENCY TABLES, 42

### G

#### **GETTING STARTED WITH SPSS, 8**

### н

Help Dialog Box Help, 16 Help with Terminology, 16 SPSS Home Page, 16 Statistics Coach, 16 Syntax Guide, 16 Tutorials, 15 HELP, 15 HISTOGRAM, 43 HYPOTHESIS TEST FOR A POPULATION MEAN, 44 HYPOTHESIS TEST FOR A POPULATION PROPORTION, 46

## \_\_\_\_\_I

Insert Case, 88 Insert Variable, 88 INSERT VARIABLE OR CASE, 88 Interactive Graph Dialog Box Options, 49 Editing, 50 INTERACTIVE GRAPHS, 48 Interquartile Range, 27 INTRODUCTION, 5

#### L

LINE GRAPH, 51 LINEAR CORRELATION, 52 Linear Regression Least-Squares Criterion, 53 LINEAR REGRESSION, 53

#### Μ

MERGE FILES, 89 Missing Values Data Editor, 84 Multiple Response, 57, 58 Crosstabs, 58 Frequencies, 57 MULTIPLE RESPPONSE, 55

#### Ν

Normal curve, 43 NORMAL PROBABILITIES, 60 NORMAL PROBABILITY PLOT, 61

### 0

Ogive, 62 ORDERING OF TOPICS, 6 OUTPUT VIEWER, 90

# Р

Pareto Chart Cumulative frequency curve, 63 PARETO CHART, 63 PASTE Edit Menu (Data Editor), 85 Pie Chart Editing, 65 PIE CHART, 64 PIVOT TABLES, 91 Post Hoc Tests Analysis of Variance, 18

# R

RECODE, 92 RETREIVING A SAVED FILE, 11

# S

SAVING YOUR WORK, 12 SCATTERPLOT, 66 SELECTING CASES, 94 Simple Random Sample, 67 Sort Cases, 95 Sort Variables, 95 SORTING CASES & VARIABLES, 95 SPLITTING A FILE, 96 SPSS ENVIRONMENT Entering Data INTO A New Data File, 10 Getting Started with SPSS, 8 HELP, 15 Saving Your Work, 12 SPSS FILE EXTENSIONS, 14 SPSS Windows, 13 The Initial SPS Window, 9 SPSS FILE EXTENSIONS, 14 SPSS Home Page, 16 SPSS WINDOWS, 13 STANDARDIZED SCORES, 68 STATISTICS & CASE PROCESSING TABLES, 69 STEM-AND-LEAF PLOT, 70 Syntax Guide, 16 System Missing Data Editor, 84

### Т

TABLE EDITOR, 97
Caption, Cell Properties, Footnote, Table Properties, TableLooks, 97
TABLE OF CONTENTS, 2
Table Properties, 97
TableLooks, 97
Test of Goodness of Fit. See Chi-Square, See Chi-Square
Test of Independence. See Chi-Square
TEXT FILES
Importing, 98
THE INITIAL SPSS WINDOW, 9
T-Test
Independent Samples T Test, 71 Paires Samples T Test, 71 T-TEST, 71

#### U

UNDO Edit Menu (Data Editor), 85

### V

Variable information, 87 Variable View Data Editor, 83 Variables, 87 Utilities - data descriptions, 87

### W

WHERE TO FIND SPSS, 8

#### Ζ

z-scores, 68 Obtaining, 68