SPSS Basics

➢ Tutorial 1: SPSS Windows

There are six different windows that can be opened when using SPSS. The following will give a description of each of them.

The Data Editor

The Data Editor is a spreadsheet in which you define your variables and enter data. Each row corresponds to a case while each column represents a variable. The title bar displays the name of the open data file or "Untitled" if the file has not yet been saved. This window opens automatically when SPSS is started.
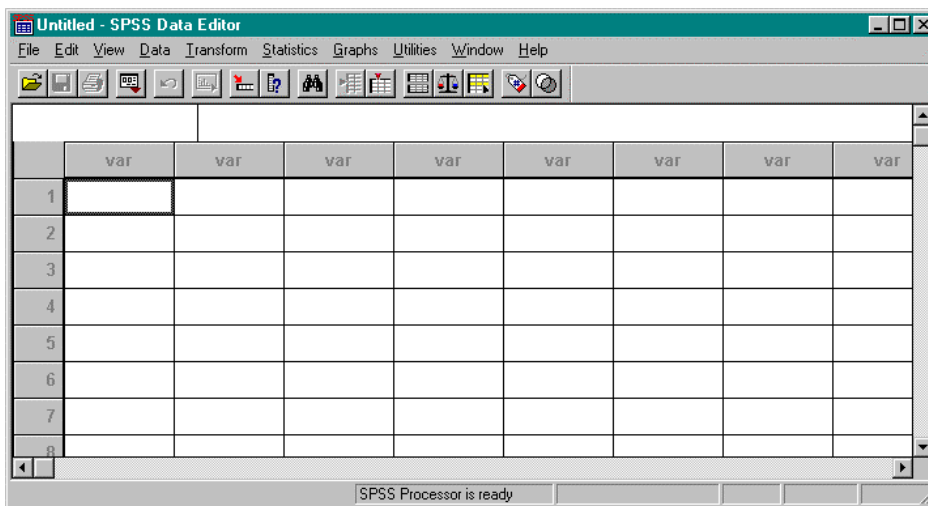
The Output Navigator

The Output Navigator window displays the statistical results, tables, and charts from the analysis you performed. An Output Navigator window opens automatically when you run a procedure that generates output. In the Output Navigator windows, you can edit, move, delete and copy your results in a Microsoft Explorer-like environment.
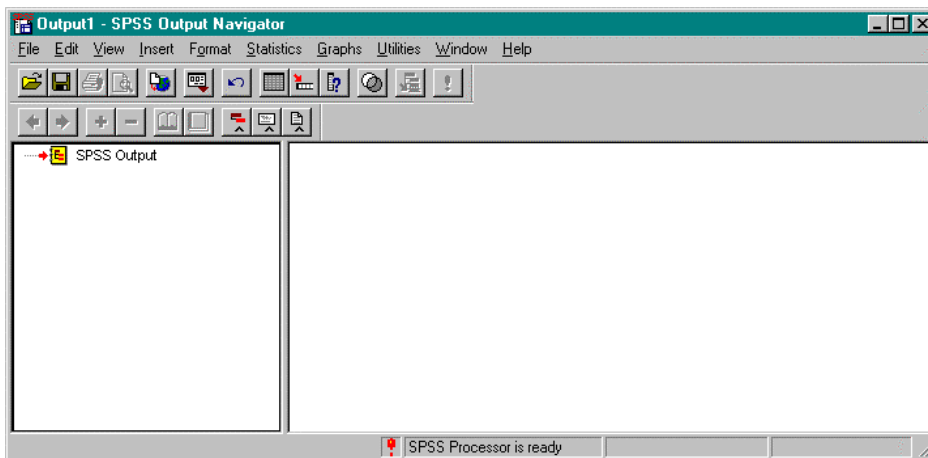
<u>The Pivot Table Editor</u>

Output displayed in pivot tables can be modified in many ways with the Pivot Table Editor. You can edit text, swap data in rows and columns, add color, create multidimensional tables, and selectively hide and show results.

<u>The Chart Editor</u>

You can modify and save high-resolution charts and plots by invoking the Chart Editor for a certain chart (by double-clicking the chart) in an Output Navigator window. You can change the colors, select different type fonts or sizes, switch the horizontal and vertical axes, rotate 3-D scatterplots, and change the chart type.

<u>The Text Output Editor</u>

Text output not displayed in pivot tables can be modified with the Text Output Editor. You can edit the output and change font characteristics (type, style, color, size).

<u>The Syntax Editor</u>

You can paste your dialog box selections into a Syntax Editor window, where your selections appear in the form of command syntax.

➢ Tutorial 2: Starting A SPSS Session

1. Logon to your Polaris account.

2. Select Programs from the Start menu.

3. Select Scientific from the Programs drop down menu.

4. Select SPSS 7.5 from the Scientific drop down menu.

➢ Tutorial 3: Getting Help on SPSS

❑ **Locating Topics in the Help Menu**

1. Select Topics from the Help Menu on the Data Editor.

2. Select the `Contents` tab. This will give a set of books to look under for the required information.

❑ **Searching for Information in the Help Menu**

1. Select Topics from the Help menu.

2. Select the Index tab.

3. Type a word in the text box describing the information to search for. This will give a list of headings on the desired information.

➢ Tutorial 4: Ending A SPSS Session

1. Select Exit SPSS from the File menu on the Data Editor.

# Creating and Manipulating Data in SPSS

When creating or accessing data in SPSS, the Data Editor window is used.

➢ Tutorial 1: Creating a New Data Set

There are three steps that must be followed to create a new data set in SPSS. The following tutorial will list the steps needed and will give an example of creating a new data set.

STEP 1: Defining Variables in a New Data Set

Variables are defined one at a time using the Define Variable dialog box. This box assigns data definition information to variables. To access the Define Variable dialog box, double-click on the top of a column where the word var appears or select Define Variable from the Data menu.
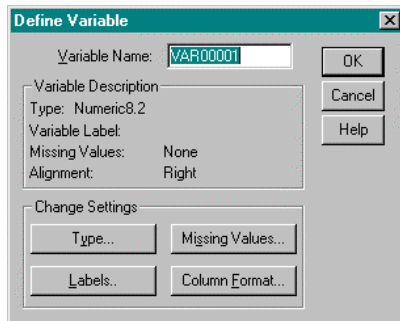
Variable Name: This field describes the name of the variable being defined. To change the name, place the cursor in this field and type the name. The variable name must begin with a letter of the alphabet and cannot exceed 8 characters. Spaces are not allowed within the variable name. Each variable name must be unique.

Type: This field describes the type of variable that is being defined.
To change this field, click on the Type… button. This will open the Define Variable Type: dialog box. Select the appropriate type of data. When done, click on the Continue button.
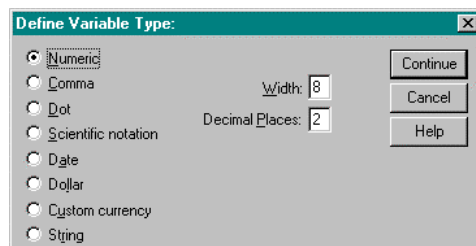

Variable Label: There are two types of variable labels:

   1.   Variable Label: A name for the variable that can be up to 120 characters long and can include spaces (which variable names cannot). If a variable label is entered, the label will be printed on charts and reports instead of the name, making them easier to understand.

2. Value Label: Provides a key for translating numeric data.

To change the variable label, click on the Labels… button. This will open the Define Labels: dialog box. Enter the appropriate information into the fields. When done, click on the Continue button.
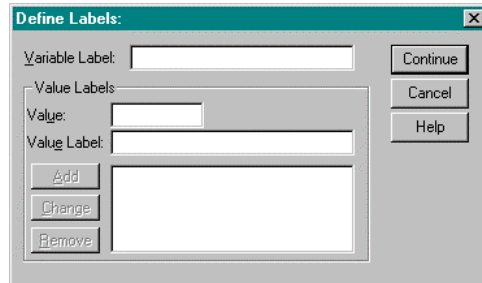
Missing Values: This field indicates which subset of the data will not be included in the data set. To change this field, click on the Missing Values… button. This will open the Define Missing Values: dialog box. Enter the appropriate information into the fields. When done, click on the Continue button.
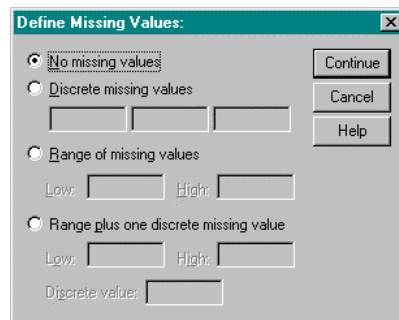
Alignment: This field indicates column alignment and width. To change this field, click on the Column Format… button. This will open the Define Column Format: dialog box. Enter the appropriate information into the fields. When done, click on the Continue button.
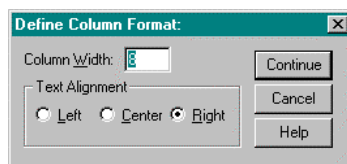
STEP 2: Entering Data in a New Data Set

Once all of the variables are defined, enter the data manually (assuming that the data is not already in an external file). The data is typed into the spreadsheet one cell at a time. Each cell represents an observation.

When information is typed into a cell, it appears in the edit area at the top of the window. The information is entered into the cell when the active cell is changed. The mouse and the tab, enter, and cursor keys can be used to enter data.

To indicate a cell that does not have a data value, a period is entered. A period represents the system-missing value.

4

STEP 3: Saving a New Data Set

Work performed on a data set only lasts during the current session. To retain the current data set, it must be saved to a file.

1. Select Save from the File menu. The Save Data As dialog box opens.

2. From the Save as Type drop-down list, select SPSS (*.sav).

3. From the Save in drop-down list, select the path where the file will be saved.

4. In the File name box, enter a name for the file. SPSS automatically adds the extension .sav.

5. Click Save.

Problem

The following data regarding a person's name, age and weight must be entered into a data set using SPSS.

Name	Age	Weight
Mark	39	250
Allison	43	125
Tom	27	180
Cindy	24	130

Solution

1. Double click on the top of the first column in the Data Editor window. This will open the Define Variable dialog box. Type Name in the Variable Name box.

2. Select Type… in the Change Settings area. This will open the Define Variable Type dialog box. Left click on String.

3. Select Continue. This will close the Define Variable Type dialog box and will re-open the Define Variable dialog box.

4. Click OK. This will define the first column as a string variable called Name.

5. Double click on the top of the second column. This will open the Define Variable dialog box. Type Age in the Variable Name box.

6. Select Type… in the Change Settings area. This will open the Define Variable Type dialog box. Left click on Numeric. In the Width box, set it to 3. In the Decimal Places box, set it to 0.

7. Select Continue. This will close the Define Variable Type dialog box and will re-open the Define Variable dialog box.

8. Click OK. This will define the second column as a numeric variable called Age.

9. Double click on the top of the third column. This will open the Define Variable dialog box. Type Weight in the Variable Name box.

10. Select Type… in the Change Settings area. This will open the Define Variable Type dialog box. Left click on Numeric. In the Width box, set it to 3. In the Decimal Places box, set it to 0.

11. Select Continue. This will close the Define Variable Type dialog box and will re-open the Define Variable dialog box.

12. Click OK. This will define the third column as a numeric variable called Weight.

13. Enter the above information into the cells of the spreadsheet. The Data Editor should look like the following.

14. Select Save from the File menu.

15. Choose the path where the file will be saved.

16. Type temp in the File name box and click Save. SPSS will save this file as temp.sav in the specified directory.

➢ Tutorial 2: Creating a New Data Set From Other File Formats

SPSS is designed to handle a wide variety of formats including:

- Spreadsheet files created with Lotus 1-2-3 and Excel
- Database files created with dBASE
- Tab-deliminated and other types of ASCII text files
- SPSS data files create on other operating systems
- SYSTAT data files

The following tutorial will indicate how to read in a spreadsheet or text file into a data set in SPSS. Examples will be given of each method.

❑ **Reading Spreadsheet Files (Lotus 1-2-3 and Excel)**

Problem

Read the following file, ~/SPSS/nba.xls, into a SPSS data set.

Solution

1. From the File menu, select Open. This will open the Open File dialog box.

2. Change the path name to your home directory and open the SPSS folder. This is where the file to be opened should be.

3. Select Excel(*.xls) (or Lotus(*.w*) for Lotus files) from the Files of type box.
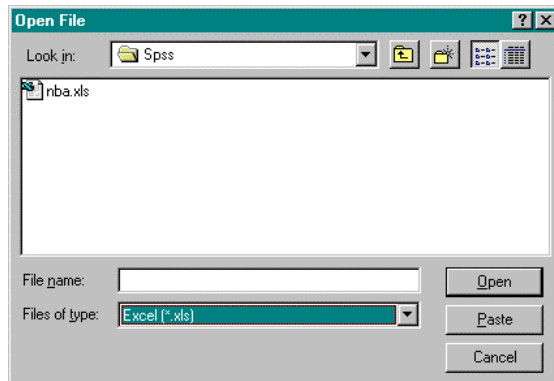
4. Select nba.xls.

5. Click Open. This will open the Opening File Options dialog box. Click on the Read variable names dialog box. Click OK. This will close the Opening File Options dialog box and will open nba.xls in the Data Editor. The Output Navigator will also be opened.
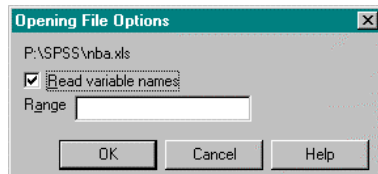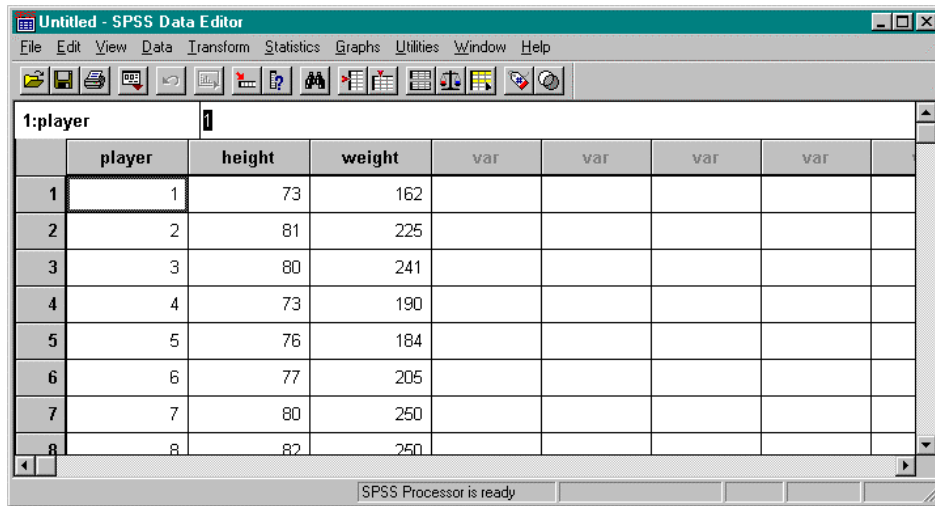
NOTE:

If only a partial file is to be read into SPSS, the following steps are taken.

•   For Lotus files, in the Range box, specify the beginning column letter and row number followed by two periods followed by the ending column letter and row number. Ie. A1..C12
•   For Excel files, in the Range box, specify the beginning column letter and row number followed by a colon followed by the ending column letter and row number. Ie. A1:C12

Window Output

```
Untitled - SPSS Data Editor
File  Edit  View  Data  Transform  Statistics  Graphs  Utilities  Window  Help
```
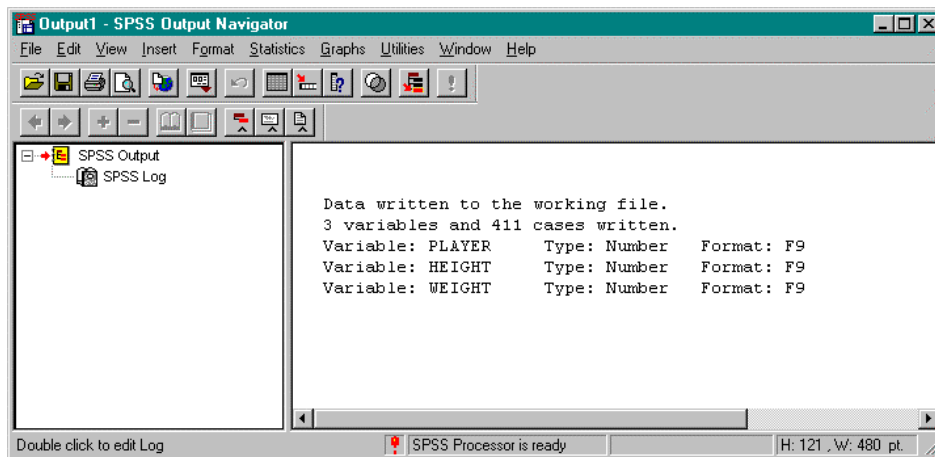
1:player	1

	player	height	weight	var	var	var	var
1	1	73	162				
2	2	81	225				
3	3	80	241				
4	4	73	190				
5	5	76	184				
6	6	77	205				
7	7	80	250				
8	8	82	250				

SPSS Processor is ready

```
Output1 - SPSS Output Navigator
File  Edit  View  Insert  Format  Statistics  Graphs  Utilities  Window  Help
```

SPSS Output
  SPSS Log

```
Data written to the working file.
3 variables and 411 cases written.
Variable: PLAYER    Type: Number   Format: F9
Variable: HEIGHT    Type: Number   Format: F9
Variable: WEIGHT    Type: Number   Format: F9
```

Double click to edit Log   SPSS Processor is ready   H: 121 , W: 480 pt.

❑   **Reading Text Files**

Two ways to read a text file are by using freefield or fixed columns.

Freefield

This method is used if the variables are recorded in the same order for each case but not necessarily in the same column locations.

Problem

Read the following file, ~/SPSS/citydata.txt, into an SPSS data set.

Solution

1. Select Read ASCII Data from the File Menu. From the Read ASCII Data drop down menu, choose Freefield. This will open the Define Freefield Variables dialog box.
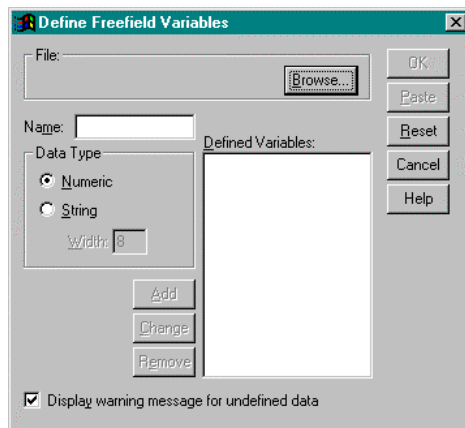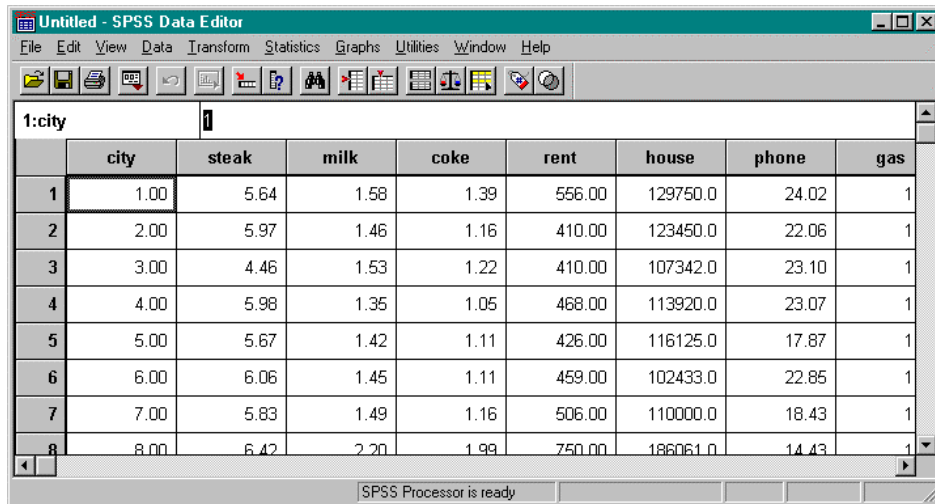
2. Specify the variable name and data type. The following gives a description of each of these fields.

    Name: Variable names must begin with a letter and cannot exceed eight characters. Each variable name must be unique.

    Data Type: Select a data type.

3. Click Add for each separate variable. This will enter the variable name and data type onto the Defined Variables list.

4. Once all variables are defined, click Browse to specify the name of the file to be read. This will open the Define Freefield Variables: Browse dialog box. Change the path name to your home directory and open the SPSS folder. This is where the file to be opened should be.

5. Select citydata.txt and click Open. The Define Freefield Variables dialog box will be returned.

6. Click OK. This will close the Define Freefield Variables dialog box and will open citydata.txt in the Data Editor.

Fixed Columns

This method is used if each variable is recorded in the same column location for each case in the data file.

Problem

Read the following file, ~/SPSS/nba.txt, into an SPSS data set.

Solution

1. Select Read ASCII Data from the File Menu. From the Read ASCII Data drop down menu, choose Fixed Columns. This will open the Define Fixed Variables dialog box which will be used to define each variable.
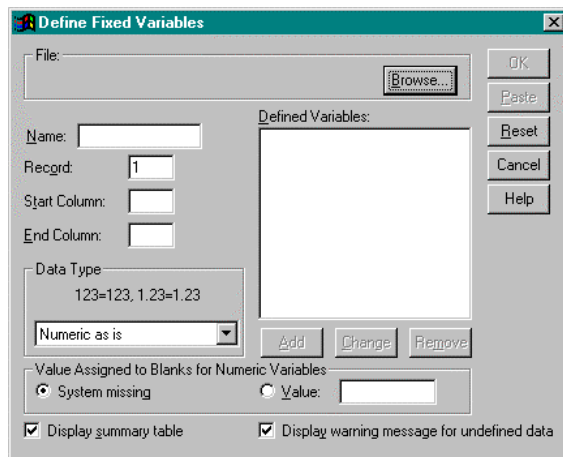
2. Specify the variable name, record, column locations, and data type. The following gives a description of each of these fields.

   Name: Variable names must begin with a letter and cannot exceed eight characters. Each variable name must be unique.

9

Record: A case can have data on more than one line. The record number indicates the line within the case where the variable is located.

Start Column/End Column: These specifications indicate the location of the variable within the record. The value for the variable can appear anywhere within the range of columns.

Data Type: Select a data type.

For this problem, the following is a list of the required information.

Name	Record	Column Locations	Data Type
Player	1	1-3	Numeric as is
Height	1	4-7	Numeric as is
Weight	1	8-12	Numeric as is

3. When all information is added for a variable, click Add. This will enter the record number, start and end columns, variable name, and data type onto the Defined Variables list.

4. Once all variables are defined, click Browse to specify the name of the file to be read. This will open the Define Fixed Variables: Browse dialog box. Change the path name to your home directory and open the SPSS folder. This is where the file to be opened should be.

5. Select nba.txt and click Open. The Define Fixed Variables dialog box will be returned.

6. Click OK. This will close the Define Fixed Variables dialog box and will open nba.txt in the Data Editor.

Window Output
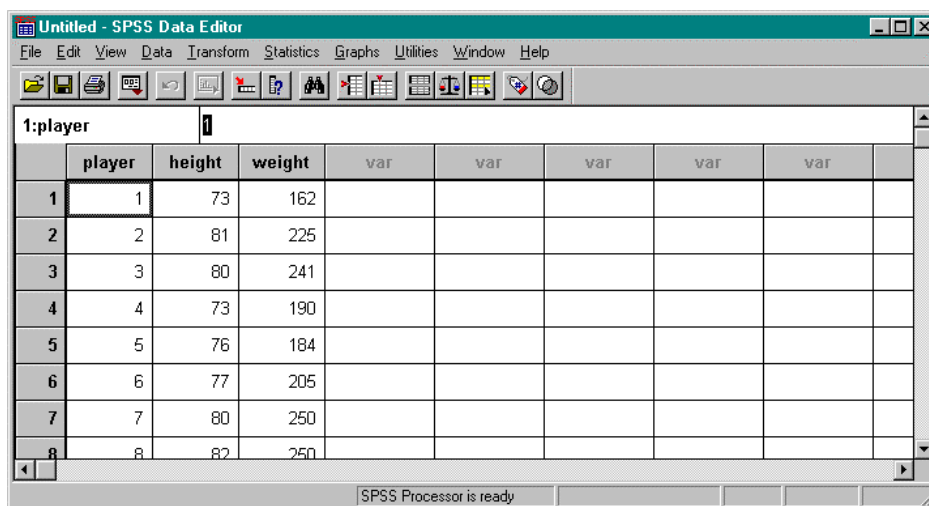
➢ Tutorial 3: Opening an Existing SPSS Data Set

1. Select Open from the File menu. This will open the Open File dialog box.
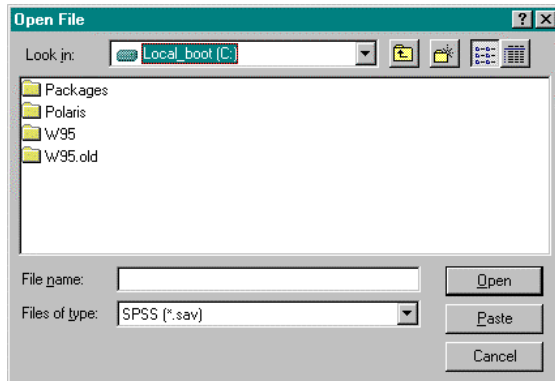
2. From the Files of type drop-down list, select .sav.

3. From the Look in drop-down list, select the appropriate drive where the file is located.

4. In the File name box, type in the name of the file to be opened.

5. Click Open.

➢ Tutorial 4: Printing a Data Set

1. Highlight the data that will be printed. To print all of the data, ignore this step and continue to step 2.

2. Select Print from the File menu. The Print dialog box opens. Change the options where appropriate.
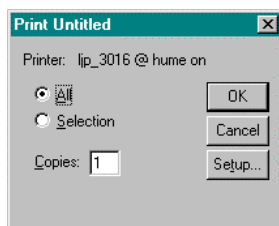
3. Click OK.

Generating Descriptive Statistics in SPSS

The following tutorials will demonstrate how to generate descriptive statistics in SPSS.

> ## Tutorial 1: Mean, Sum, Standard Deviation, Variance, Minimum Value, Maximum Value, and Range

When generating these statistics, the Data Editor must be open with the appropriate data set before continuing.

Problem

Using the data in the file nba.txt that is located in ~/SPSS/, determine the mean, sum, standard deviation, variance, minimum value, maximum value, and range for height only.

Solution

1. From the Statistics menu, select Summarize. From the Summarize drop down menu, select Descriptives. This will open the Descriptives dialog box.
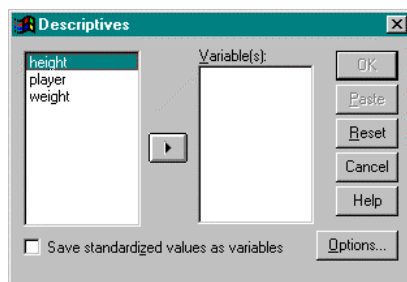
2. In the variable list, select the variable height. Left click on the right arrow button between the boxes to move this variable over to the Variable(s) box. To calculate statistics for many variables, simultaneously add variables to the Variable(s) box.

3. Click on the Options button. This will open the Descriptives: Options dialog box.
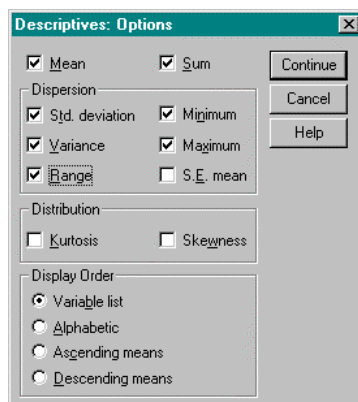
Click on mean, sum, standard deviation, variance, minimum value, maximum value, and range.

Click on the Continue button when done.

4. Click OK. The Descriptives dialog box closes and SPSS activates the Output Navigator to illustrate the statistics.

Window Output

➢ **Tutorial 2: Correlation**

Two or more variables may be included in a correlation matrix. When generating the correlation matrix, the Data Editor must be open with the appropriate data set before continuing.

Problem

Using the data in the file nba.txt that is located in ~/SPSS/, determine the correlation between a player's height and weight.

Solution

1. From the Statistics menu, select Correlate. From the Correlate drop down menu, select Bivariate. This will open the Bivariate Correlations dialog box.
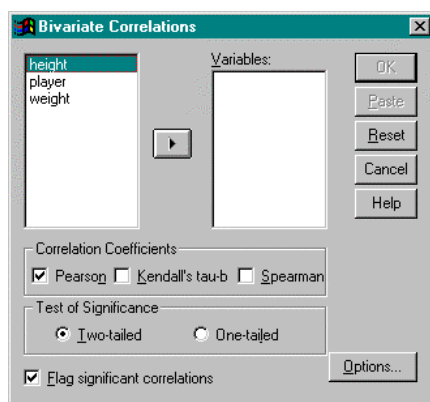
2. In the variable list, select height and weight. Left click on the right arrow button between the boxes to move a variable over to the Variable(s) box.

3. Select the type of correlation coefficients that will be generated. In this case, use Pearson.

4. Select the test of significance to be used. In this case, use two-tailed.

5. Check mark the Flag significant correlations box.

6. Click on the Options…button. This will open the Bivariate Correlations: Options dialog box.
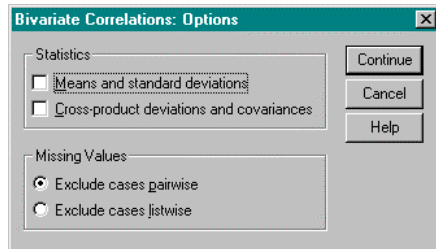
> To display the mean and standard deviation for each variable, select Means and standard deviations. In this case, this option is not used.
>
> To display cross product deviations and covariances for each pair of variables, select Cross-product devations and covariances. In this case, this option will not be used.
>
> When done, click the Continue button.

7. Click OK. The Bivariate Correlations dialog box closes and SPSS activates the Output Navigator. The correlation coefficient for each pair of variables is displayed. The number of cases appears at the bottom.

Window Output

Generating Graphical Statistics in SPSS

The following tutorials introduce how to create scatter plots, histograms, stem and leaf plots, and box plots using the SPSS Graphs menu located on the Data Editor menu bar.

➤ Tutorial 1: How to Generate Scatter Plots

Problem

Using the data in ~/SPSS/nba.txt, create an x-y plot of a player's weight versus height.

Solution

1. From the Graphs menu, select Scatter… This will open the Scatterplot dialog box.
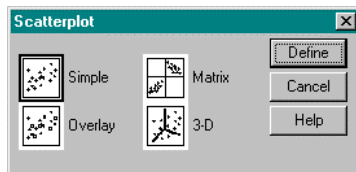
2. Select the Simple icon and click Define. This will open the Simple Scatterplot dialog box.
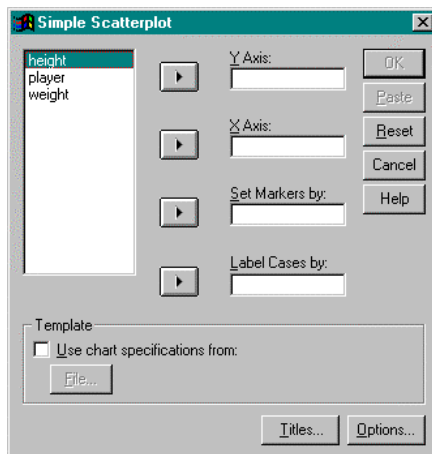
3. From the variable list, select weight. Left click on the right arrow button between the variable list and the Y Axis box to move the variable, weight, to this box.

4. From the variable list, select height. Left click on the right arrow button between the variable list and the X Axis box to move the variable, height, to this box.

5. Click on the Options… button. This will open the Options dialog box.
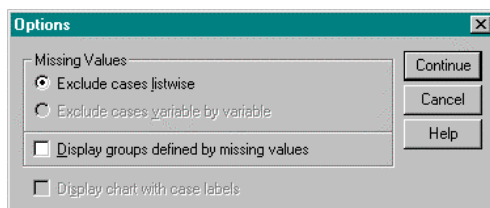
To display a report of missing values, select Display groups defined by missing values. In this case, this option will not be used.

When done, click the Continue button.

6. To display titles, subtitles, or footnotes on the histogram, click on the Titles… button. This will open the Titles dialog box.
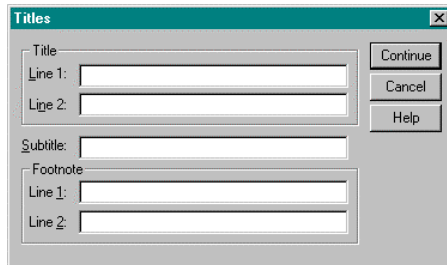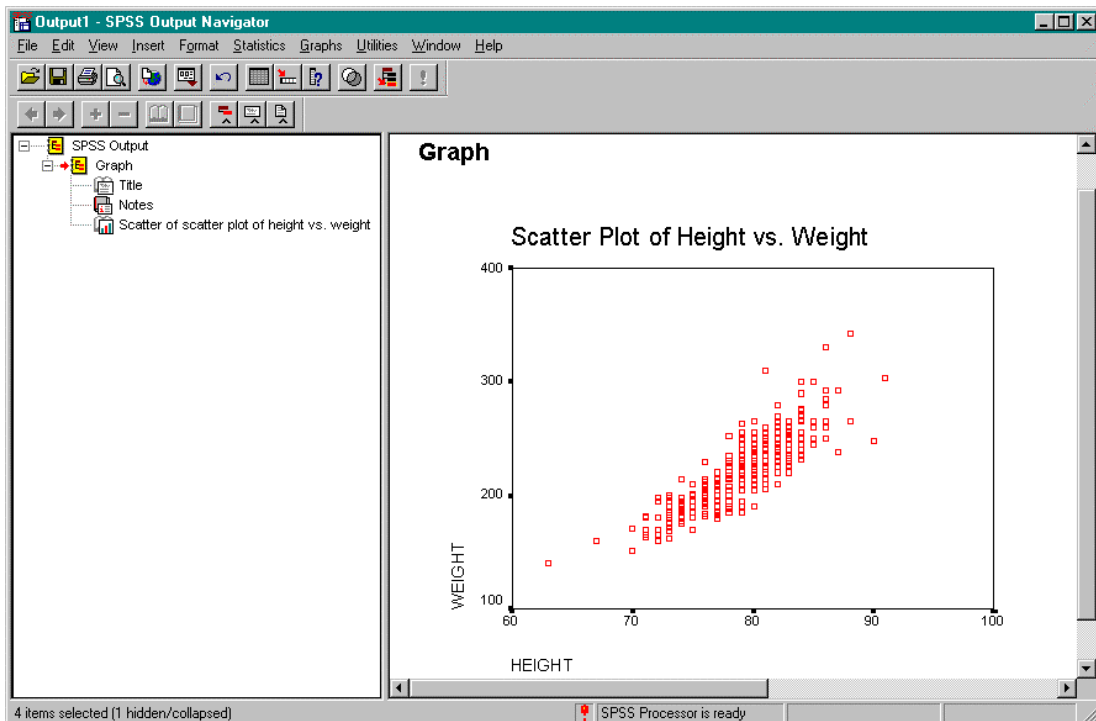

In the Line 1 box, type "Scatter Plot Height vs. Weight".

When done, click the Continue button.

7. Click OK. The Simple Scatterplot dialog box closes and SPSS activates the Output Navigator.

Window Output

➢ Tutorial 2: How to Generate a Histogram

<u>Problem</u>

Using the data in ~ /SPSS/statdata.txt, create histogram of per capita income.

<u>Solution</u>

1. From the Graphs menu, select Histogram… This will open the Histogram dialog box.
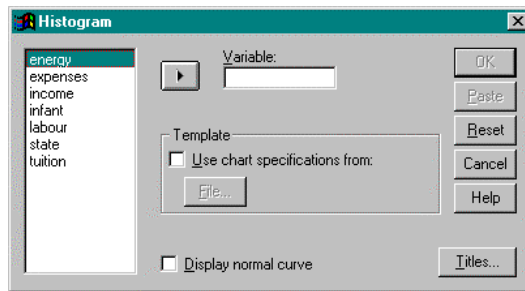
2. From the variable list, select income. Left click on the right arrow button between the variable list and the Variable box to move the variable, income, to this box.

3. Select Display normal curve box to show a normal curve on the histogram.

4. To display titles, subtitles, or footnotes on the histogram, click on the Titles… button. This will open the Titles dialog box.
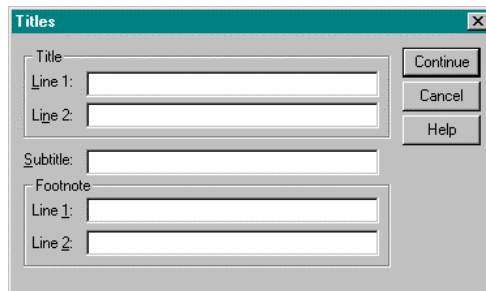
   In the Line 1 box, type "Histogram of Per Capita Income".

   Click on the Continue button when done.

5. Click OK. The Histogram dialog box will close and SPSS activates the Output Navigator to display the histogram.
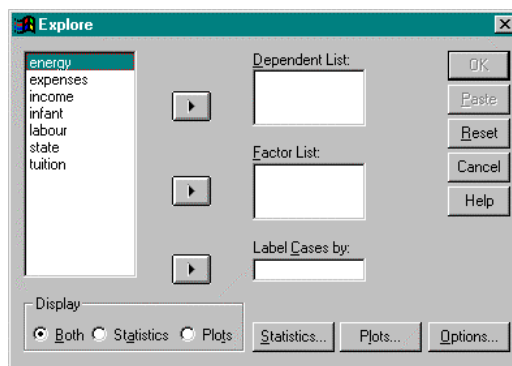
17

Window Output
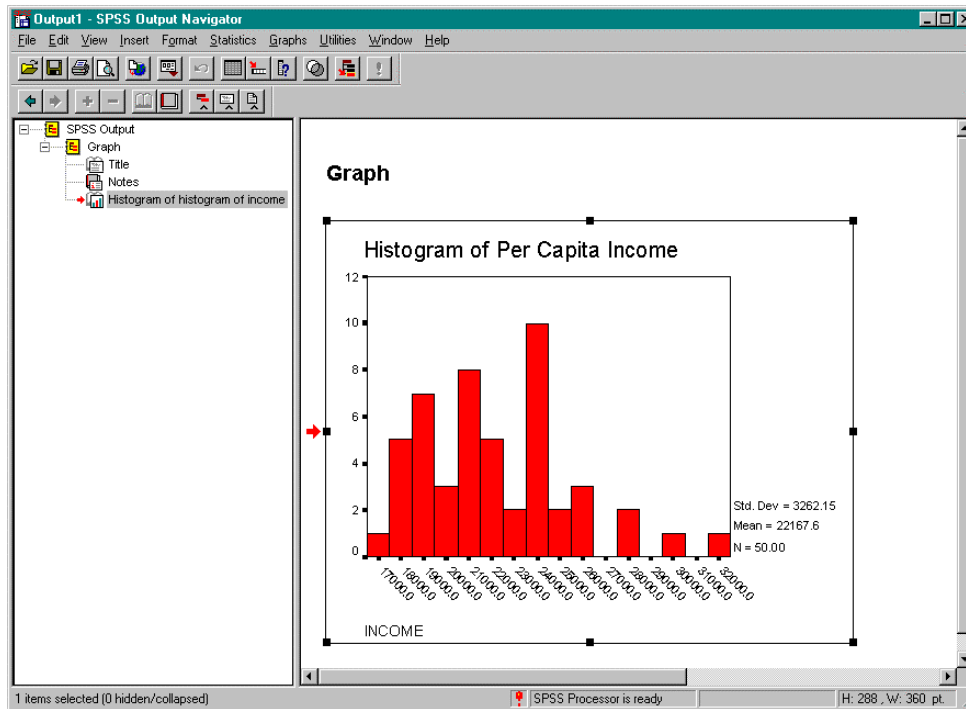
> ## Tutorial 3: How to Generate a Stem and Leaf Plot

Problem

Using the data in ~ /SPSS/statdata.txt, create a stem and leaf plot of per capita income.

Solution

1. From the Statistics menu, select Summarize. From the Summarize drop-down menu, select Explore… This will open the Explore dialog box.

2. From the variable list, select income. Left click on the right arrow button between the variable list and the Dependent List box to move the variable, income, to this box.

3. Click on the Statistics… button. This will open the Explore: Statistics dialog box.

18
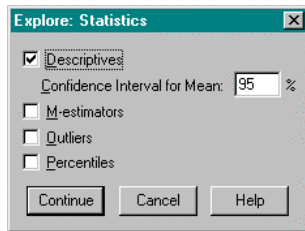
To display descriptive statistics, select Descriptives.

To display maximum likelihood estimators, select M-estimators.

To display cases with the five largest and smallest values, select Outliers.

To display percentiles, select Percentiles.

In this case, none of these options are used.

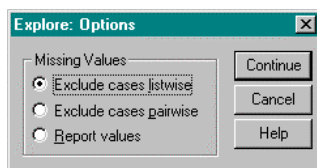When done, click on the Continue button.

4. In the Display area, select Plots. This will display the specified plot only (i.e. no statistics are given).

5. Click on the Plots… button. This opens the Explore: Plots dialog box.
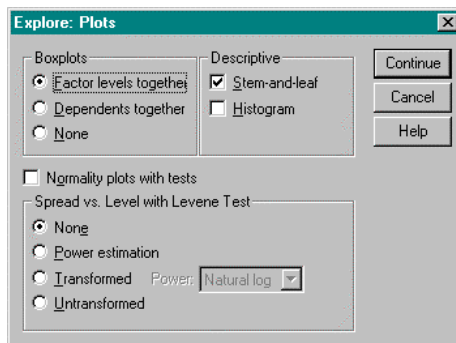
Ensure that the Stem-and-leaf box is selected.

Click on the Continue button.

6. Click on the Options button. This will open the Explore: Options button.

To exclude cases that have missing values for any of the variables used in any of the analyses, select Exclude cases listwise. In this case, this option is used.
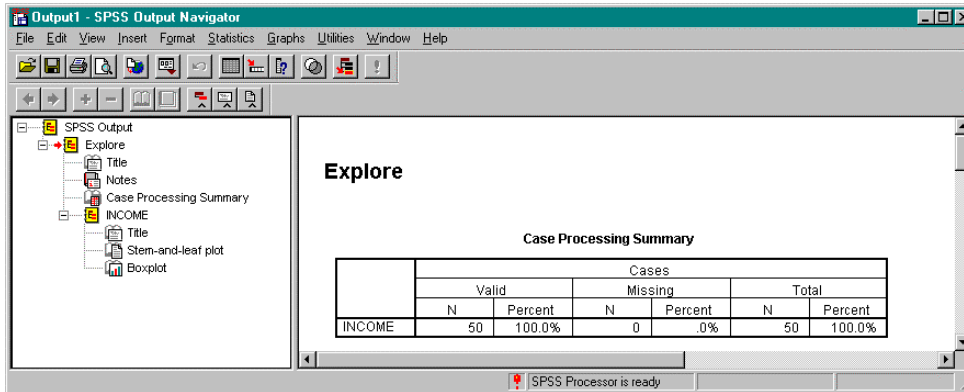
To exclude cases that have missing values for either or both of the pair of variables in a specific correlation coefficient, select Exclude cases pairwise.

However, to treat missing values as a separate category, select Report values.
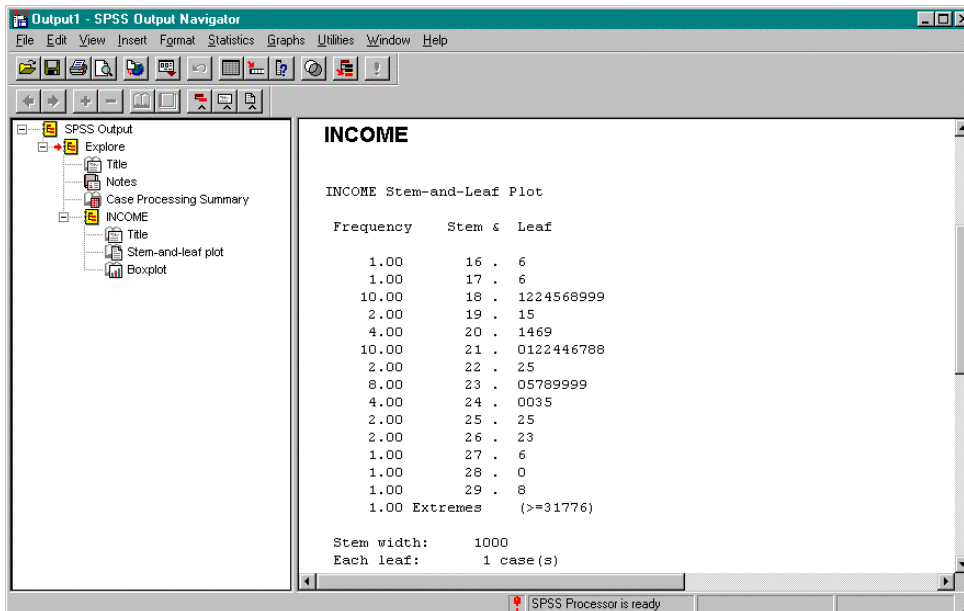
Click the Continue button when done.

7. Click OK. This will close the Explore dialog box and SPSS activates the Output Navigator to display the stem and leaf plot.
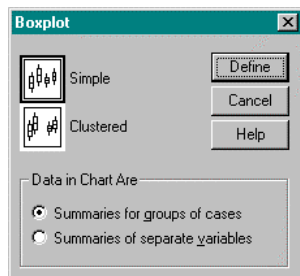
Window Output

➢ **Tutorial 4: How to Generate a Box Plot**

<u>Problem</u>

Using the data in the file, ~ /SPLUS/statdata.dat, produce a boxplot of per capita income

<u>Solution</u>

1. From the Graphs menu, select Boxplot… This will open the Boxplot dialog box.

2. Select the Simple button.

3. Select Summaries of separate variables in the Data in Chart Are area.

4. Click on the Define button. This will open the Define Simple Boxplot: Summaries of Separate Variables dialog box.
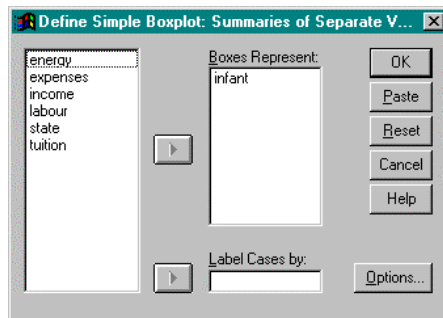
5. From the variable list, select income. Left click on the right arrow button between the variable list and the Boxes Represent box to move the variable, income, to this box.

6. Click on the Options… button. This will open the Options dialog box.
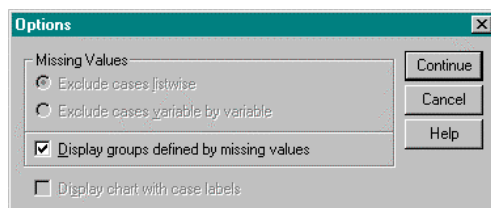
To display a report of missing values, select Display groups defined by missing values. In this case, this option will not be used.

When done, click the Continue button.

7. Click OK. This will close the Define Simple Boxplot: Summaries of Separate Variables dialog box and SPSS activates the Output Navigator to display the box plot.

Window Output
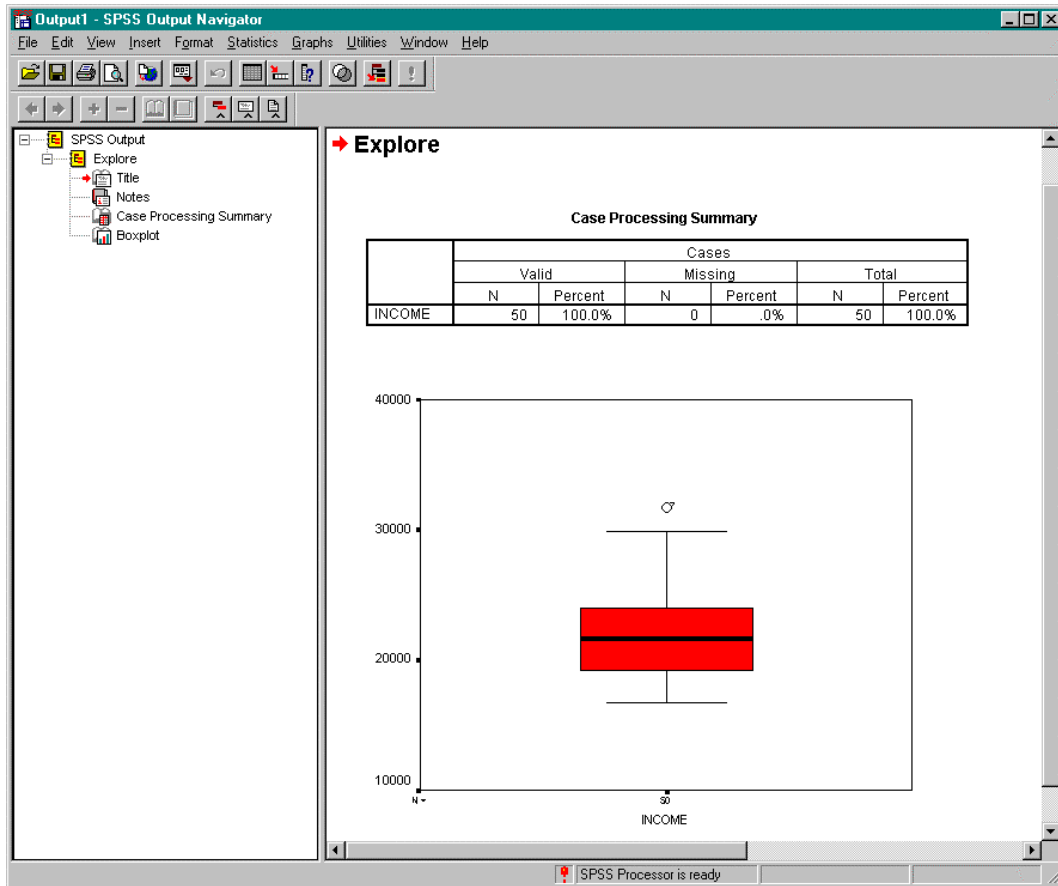
Statistical Models in SPSS

➢ Tutorial 1: Linear Regression

The Regression submenu on the Statistics menu of the Data Editor provides regression techniques. The following tutorial will introduce how to perform linear regression using SPSS. The output contains goodness of fit statistics and the coefficients for the variables.

Problem

Using the data in ~/SPSS/nba.txt, compute a least squares regression line to investigate if a player's height can predict his weight.

Solution

1. From the Statistics menu, select Regression. From the Regression drop down menu, select Linear… This will open the Linear Regression dialog box.
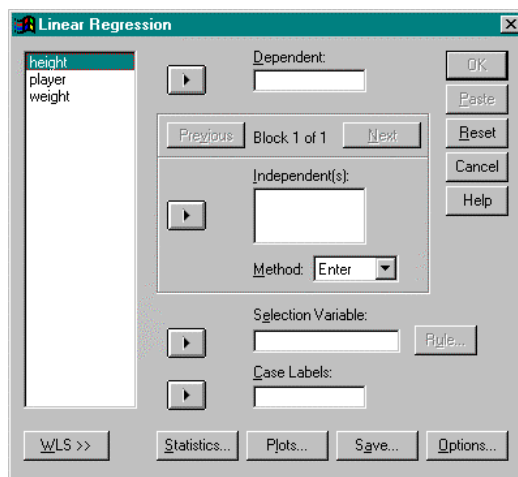
2. From the variable list, select weight. Left click on the right arrow button between the variable list and the Dependent box to move the variable, weight, to this box.

3. From the variable list, select height. Left click on the right arrow button between the variable list and the Independent(s) box to move the variable, height, to this box.

4. Select the method the independent variables are entered into the analysis. From the Method drop-down menu, there is a choice of enter, stepwise, remove, backward, and forward. In this case, we will use the enter method.

5. To limit the analysis to a subset of cases having a particular value for a variable, enter this variable into the Selection Variable box. In this case, this option is not used.

6. Determine the variable that will identify the points on plots. Select the variable and left click on the right arrow between the variable list and the Case Labels box. In this case, this option is not used.
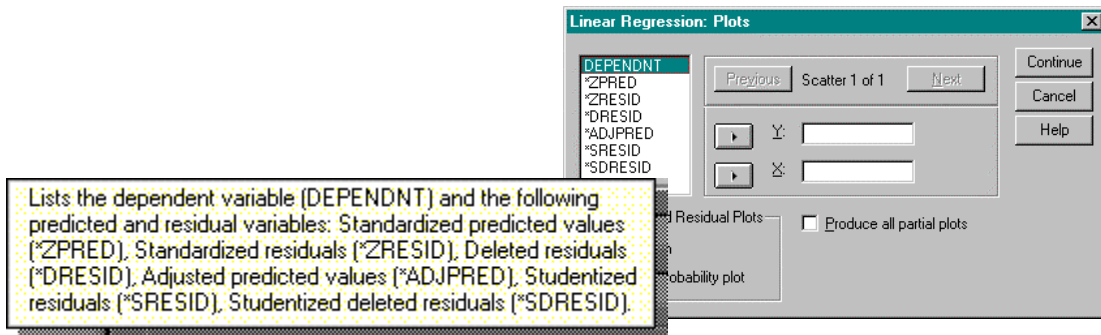
7. To display statistics, click on the Statistics… button. This will open the Linear Regression: Statistics dialog box.
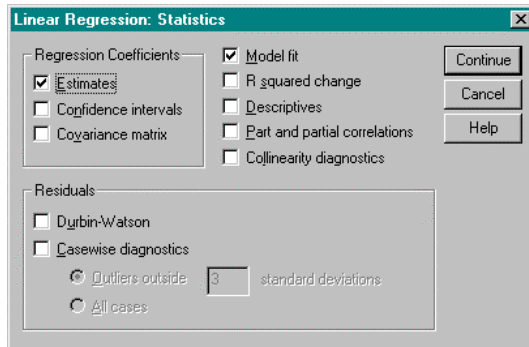
Select the appropriate statistics to be displayed and click on the Continue button when done. In this case, this option is not used.

8. To display specific plots, click on the Plots… button. This will open the Linear Regression: Plots dialog box.

From the variable list, select the variable that will be displayed on the Y axis. Left click on the right arrow button between the variable list and the Y box. Do this also for the X axis. When done, click on the Next button. If more plots are needed, follow the same procedure. In this case, this option is not used.

When done defining the plots, click on the Continue button.

9. To indicate which statistics should be displayed, click on the Save button. This will open the Linear Regression: Save dialog box.
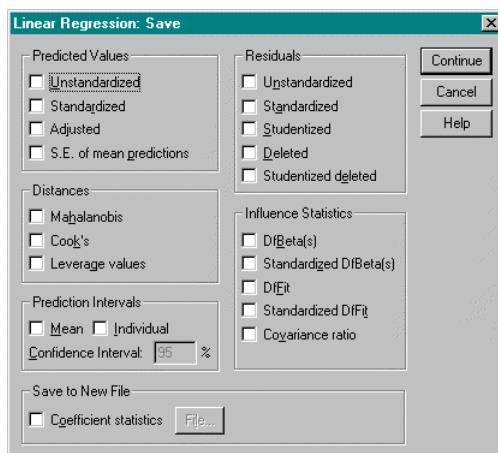
Select the appropriate statistics. To save the coefficient statistics, click on the box and indicate the file to which you want them saved. In this case, this option is not used.

10. To indicate the stepping method criteria, click the Options… button. This will open the Linear Regression: Options dialog box.
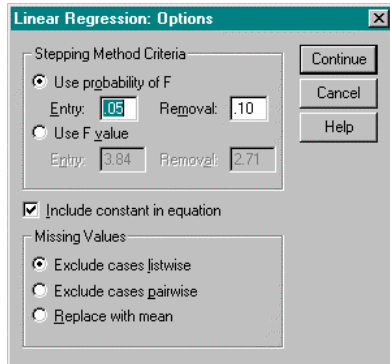
Select the method to be used. When the selection is made, click on the Continue button.

11. Click OK. This will close the Linear Regression dialog box. SPSS activates the Output Navigator to display the results of the analysis.
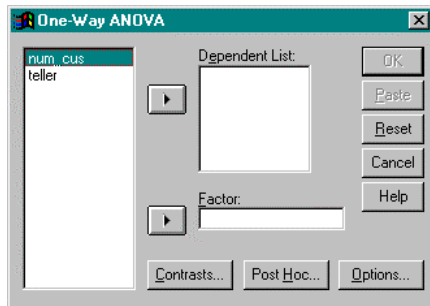
<u>Window Output</u>

## ➢ Tutorial 2: Analysis of Variance

<u>Problem</u>

Using the data in ~/SPSS/teller1.txt, test if the mean number of customers served per hour by each of the four tellers is the same.

<u>Solution</u>

1. From the Statistics menu, select Compare Means. From the Compare Means drop down menu, select One-Way ANOVA… This will open the One-Way ANOVA dialog box.


2. From the variable list, select num_cus. Left click on the right arrow button between the variable list and the Dependent List box to move the variable, num_cus, to this box.

3. From the variable list, select teller. Left click on the right arrow button between the variable list and the Factor box to move the variable, teller, to this box.

4. Click on the Contrasts… button. This will open the One-Way ANOVA: Contrasts dialog box.
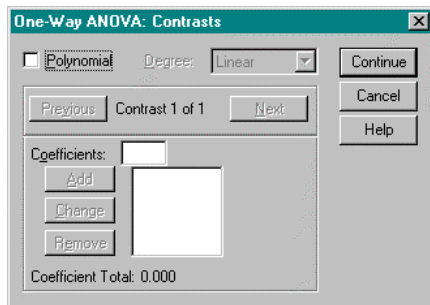
To partition between-groups sum of squares into polynomial trend components, select the Polynomial box and select the highest degree of the polynomial to be modelled. In this case, this option will not be used.

To enter a numeric coefficient value for each level, click Add. However, the number of coefficients must equal the number of groups or the analysis is not performed. Because the levels in this problem are already numeric, this option does not need to be used.

5. Click on the Post Hoc… button. This will open the One-Way ANOVA: Post Hoc Multiple Comparisons dialog box.
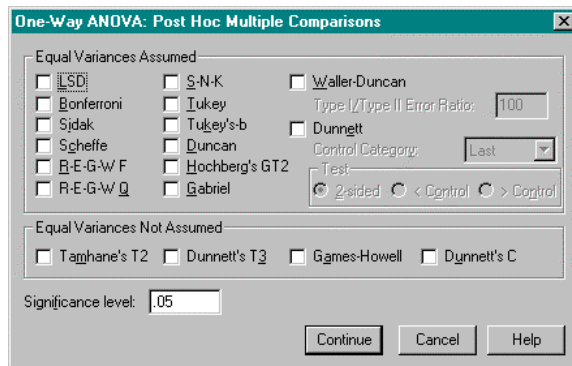

If equal variances are assumed between the different factor levels, select the type of comparison method to be used.

If equal variance are not assumed between the different factor levels, select the type of comparison method to be used.

To get a description on each of the methods listed, right click on the word. A description window will appear.

Click the Continue button when done.

6. Click on the Options… button. This will open the One-Way ANOVA: Options dialog box.
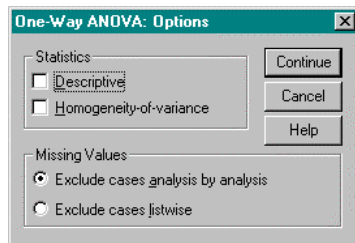
To display descriptive statistics, select Descriptive in the Statistics area. In this case, select this option.

To exclude cases that have missing values for the variable involved in that test, select Exclude cases analysis by analysis. In this case, select this option.

However, to exclude cases that have missing values for any of the variables used in any of the analyses, select Exclude cases listwise.

Click the Continue button when done.

7. Click OK. The One-Way ANOVA dialog box closes and SPSS activates the Output Navigator. The means of the dependent variable for each category of the independent variable can be found under "Descriptives".

Window Output

Output1 - SPSS Output Navigator

File Edit View Insert Format Statistics Graphs Utilities Window Help

SPSS Output
- Oneway
  - Title
  - Notes
  - Descriptives
  - ANOVA

Oneway

Descriptives

			N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
							Lower Bound	Upper Bound		
NUM_CUS	TELLER	1.00	5	21.6000	3.3615	1.5033	17.4262	25.7738	18.00	26.00
		2.00	5	14.8000	1.6432	.7348	12.7598	16.8402	13.00	17.00
		3.00	5	15.0000	3.8079	1.7029	10.2720	19.7280	11.00	21.00
		4.00	5	22.0000	2.9155	1.3038	18.3800	25.6200	19.00	26.00
		Total	20	18.3500	4.5105	1.0086	16.2390	20.4610	11.00	26.00

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
NUM_CUS	Between Groups	238.550	3	79.517	8.596	.001
	Within Groups	148.000	16	9.250		
	Total	386.550	19			

SPSS Processor is ready

H: 121 , W: 480 pt.