AN INTRODUCTION TO SPSS

Abstract

This manuscript is designed for a new user of SPSS, it involves reading a data set into SPSS, data manipulation, simple data visualizing tools, and some common statistical analyses. For additional information or assistance, please contact or visit the Statistical Consulting Software desk.

This document was prepared by members of the Statistical Consulting Service at Purdue University.

Updated August 2016

Contents

Original document was written June 2006, updated in July 2012.

This update was completed in August 2016 –the following graduate students in the Department of Statistics at Purdue University contributed to this update:

Eonyoung Cho
Evidence Matangi
Hui Sun
Yunfan Li

Introduction

About this Document

This manual was written by members of the Statistical Consulting Service as an introduction to SPSS 22. It is designed to assist new users in familiarizing themselves with the use of SPSS. This document only discusses some basic data visualization and statistical analysis procedures. For more in-depth information about any SPSS-related problem, the software consulting desk, located in MATH G175, is open Monday – Friday from 10:00 am to 4:00 pm when classes are in session at Purdue University.

What is SPSS?

SPSS is a powerful statistical software program with a graphical interface designed for ease of use. Almost all commands and options can be accessed using pull down menus at the top of the main SPSS window. This design means that once you learn a few basic steps to access programs, it is very easy to expand your knowledge in using SPSS through the help files. To access the online SPSS help, you click on *Help* in the menu and then click on *Topics* if you want help by topic or on *Tutorials* for step-by-step hands-on guide.

How to get SPSS

SPSS is installed on all ITaP (Information Technology at Purdue) machines in all ITaP labs around campus. To get into the program: click *Start* → *Programs* → *Standard Software* → *Statistical Packages* → *IBM SPSS Statistics 22*

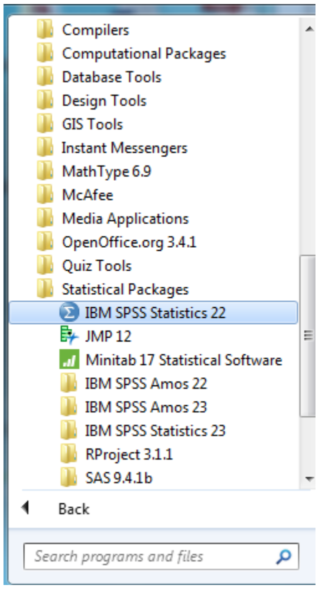


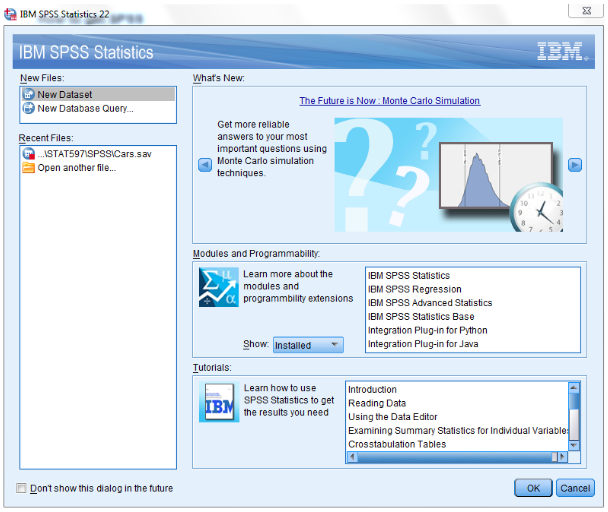
Figure 1: Accessing SPSS on an ITaP PC

Stewart Center Room G31 will loan SPSS CDs overnight *to Purdue employees and graduate students only* to install SPSS on their home computers or laptops. Remember to take your ID to sign out the CDs!

SPSS may also be launched using the *goremote* facility provided through ITaP. The website address is: https://goremote.itap.purdue.edu/Citrix/XenApp/auth/login.aspx . Once logged in, go to Standard Software → Statistical Packages → SPSS 22.

Opening SPSS Data

When SPSS is launched, a pop-up window (**Error! Reference source not found.**) with a few options will appear. Assume the goal is to analyze a data set, one can select *New Dataset* or open a file recently used or another file under *Recent Files* and then click OK. The other windows shows *What's New*, *Modules and Programmability* and *Tutorials*, which help one to navigate SPSS.



Get more reliable answers to your most important questions using Monte Carlo simulation techniques.

Learn more about the modules and programmability extensions

Learn how to use SPSS Statistics to get the results you need

Figure 2: IBM SPSS Welcome Screen

Sometimes you have already entered the SPSS session as described above, worked on a data set for a while, and then want to open and work on another data set. You do not have to quit the current SPSS session to perform this. Simply click on the *File* menu, follow *Open* then *Data*… and find your file (Figure 3).

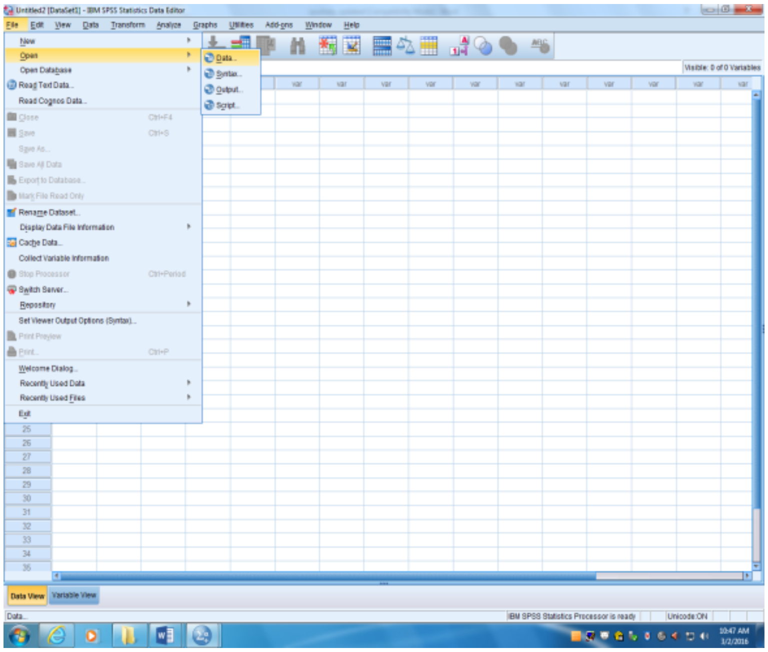


Figure 3: Opening Data From Within SPSS

Opening data from external files

Opening CSV or Excel files

Once in SPSS, in the *SPSS Data Editor* click on *File*, then *Open* and then choose data as shown in Figure 3, and *Enter* and the screen as shown in Figure 4 is given. In *Look in*: specify the location of the datafile, under *File name*: specify its name; and under *Files of type*: specify the file type. The dataset we are working with is a called Cars.csv as shown in Figure 4. (Download a copy of this data set click here)

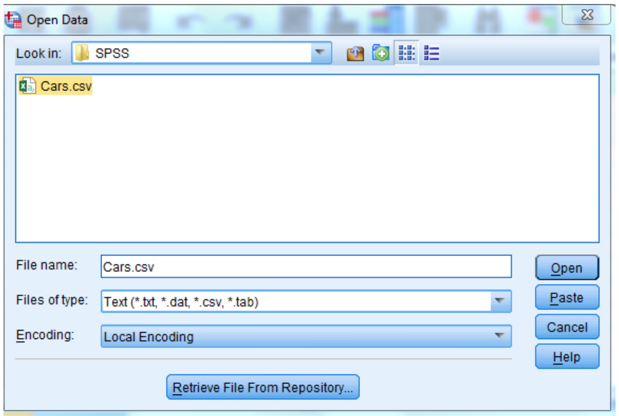
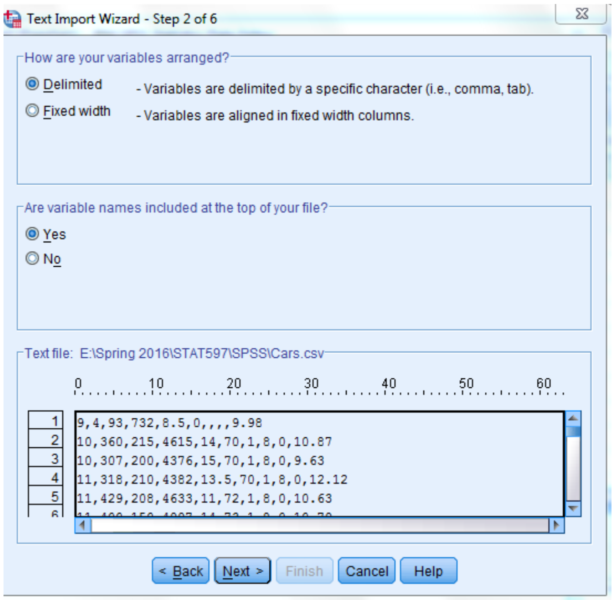


Figure 4: Opening external data files in SPSS

Figure 5 (below) shows how one imports the data from Car.csv. Since this file has variable names at the top of the file, then click the *Yes* button under "*Are variable names included at the top of your file?*" Click on *Next* until the data pops in the Data View. Similar steps can be taken to import data from Excel and many spreadsheets and text files.


Text Import Wizard - Step 2 of 6

How are your variables arranged?

● Delimited - Variables are delimited by a specific character (i.e., comma, tab).

○ Fixed width - Variables are aligned in fixed width columns.

Are variable names included at the top of your file?

● Yes

○ No

Text file: E:\Spring 2016\STAT597\SPSS\Cars.csv

```
1  9,4,93,732,8.5,0,,,,9.98
2  10,360,215,4615,14,70,1,8,0,10.87
3  10,307,200,4376,15,70,1,8,0,9.63
4  11,318,210,4382,13.5,70,1,8,0,12.12
5  11,429,208,4633,11,72,1,8,0,10.63
```

Figure 5: Data importing in SPSS

Cars Data (used in the Examples)

The cars data consists of measurements taken from a sample of 406 cars on 8 variables. Three variable are measured on a nominal scale, the country of origin (America, Japan, and Europe); type of cylinder, with 5 categories, 3-cylinder, 4-cylinder, 5-cylinder, 6-cylinder, 7-cylinder, and 8-cylinder; and the model year from 1970 to 1982. The other five variables are measured on a ratio scale which are the vehicle weight, time to accelerate, horsepower, miles per gallon, and engine displacement.

SPSS Windows

The SPSS program has three main types of windows: the data editor, output window and syntax window. The data editor window is open by default, and contains the data set. It consists of two views, the Data View and the Variable View. This window is described in more detail in the section on Working With Data and Variables. Data files are saved with a file type of .sav.

The output window holds the results of analyses. This window will open automatically once an analysis is requested. The tables of the Output Viewer are saved (click *File*, *Save* or *Save As*) with a file type of .spv, which can only be opened with SPSS software.

The syntax window contains written commands corresponding to each menu command and options. Syntax can be created by hitting Paste instead of OK on main windows for each procedure. Using Paste will not cause the procedure to be performed. To run procedures from the syntax window, click on ▶. The syntax window will only open if a syntax file is opened by the user, or if the paste option is used when executing a command. Output and syntax files can be saved and opened using the File menu. Multiple output and syntax files can be open at the same time. Syntax files are saved as plain text and almost any text editor can open them, but with a file extension of .sps.

Dialogue Boxes

Although each dialog box is unique, they have many common features. A fairly typical example is the dialog box for producing frequency tables (tables with counts and percents). To bring up this dialog box from the menus in the data window, click on *Analyze* → *Descriptive Statistics* → *Frequencies.*

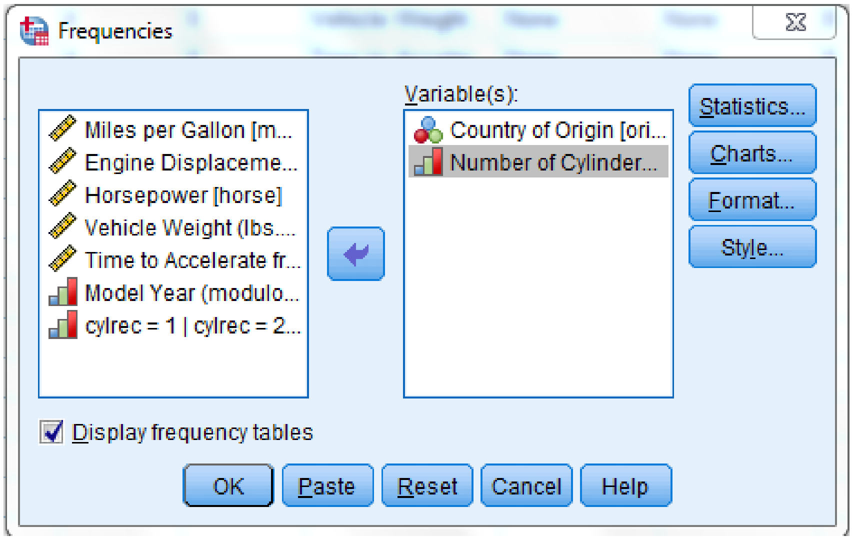
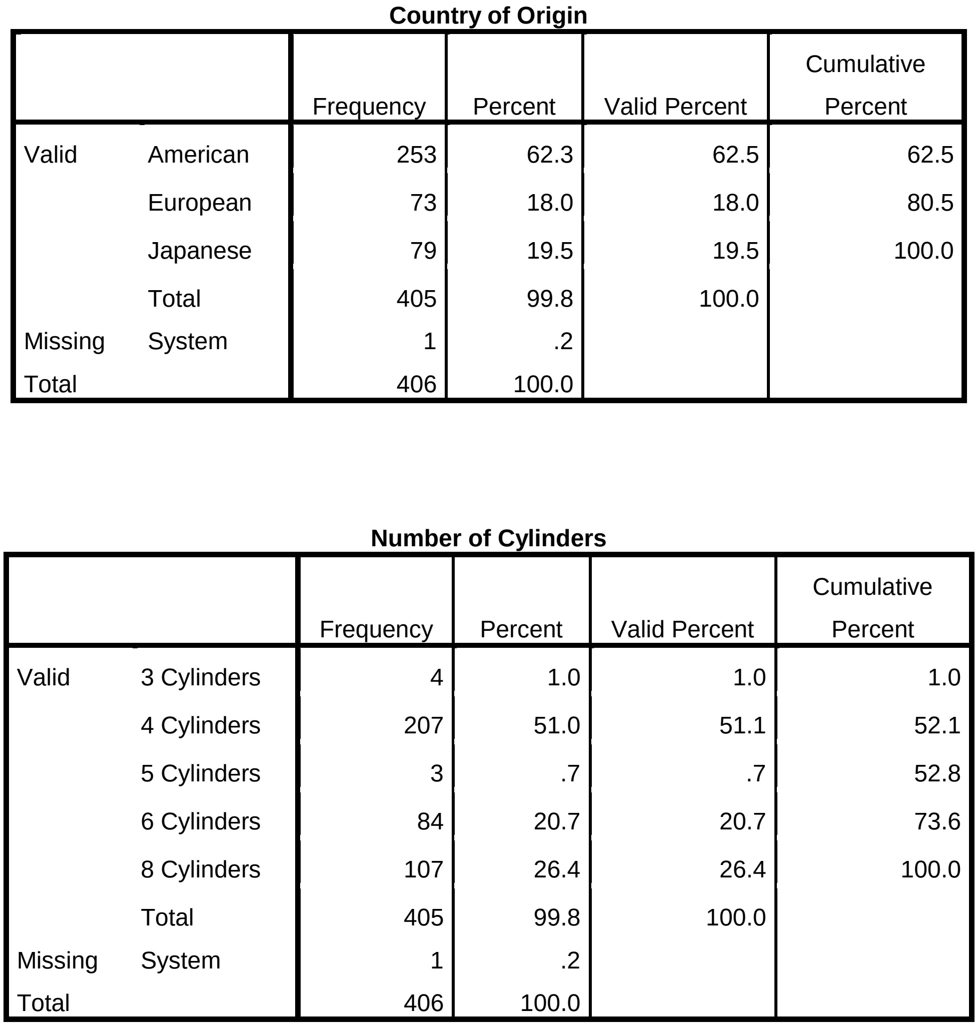


Figure 6: Dialogue Box

On the left in Figure 6, there is a variable selection list with all of the variables in your data set. If your variables have variable labels, what you see is the beginning of the variable label. To see the full label as well as the variable name [in square brackets], hold your cursor over the label beginning. Select the variables you want to analyze by clicking on them (you may have to scroll through the list). Then click the arrow button to the right of the selection list, and the variables are moved to the analysis list on the right. If you change your mind about a variable, you can select it in the list on the right and then click the arrow button to move it back out of the analysis list. On the far right of the dialog are several buttons that lead to further dialog boxes with options for the frequencies command. At the bottom of the dialog box, click *OK* to issue your command to SPSS, or *Paste* to have the command written to a Syntax Editor.

Frequency Table



Country of Origin

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	American	253	62.3	62.5	62.5
	European	73	18.0	18.0	80.5
	Japanese	79	19.5	19.5	100.0
	Total	405	99.8	100.0	
Missing	System	1	.2		
Total		406	100.0		

Number of Cylinders

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3 Cylinders	4	1.0	1.0	1.0
	4 Cylinders	207	51.0	51.1	52.1
	5 Cylinders	3	.7	.7	52.8
	6 Cylinders	84	20.7	20.7	73.6
	8 Cylinders	107	26.4	26.4	100.0
	Total	405	99.8	100.0	
Missing	System	1	.2		
Total		406	100.0		

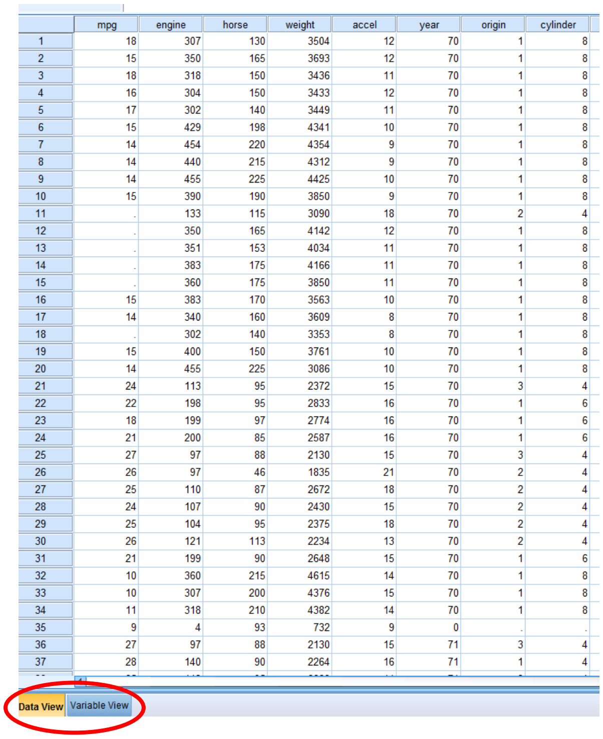
Figure 7: Frequency output for the “Country of Origin” and “Number of Cylinders”

If you return to a dialog box you will find it opens with all the specifications you last used. This can be handy if you are trying a number of variations on your analysis, or if you are debugging something. If you'd prefer to start fresh you can click the *Reset* button.

Working with Data and Variables

Viewing data and variables

Data in SPSS can be viewed in two different ways: data view and variable view. The data view allows the user to look at the entire data set, with each row showing a different observation, and each column representing a different variable. Another way to view the data is to use the variable view. This shows the variable names and general properties for each variable. The user can alternate between these views using the tabs at the bottom left hand side of the SPSS data editor window, Figure 8 below shows the data view.

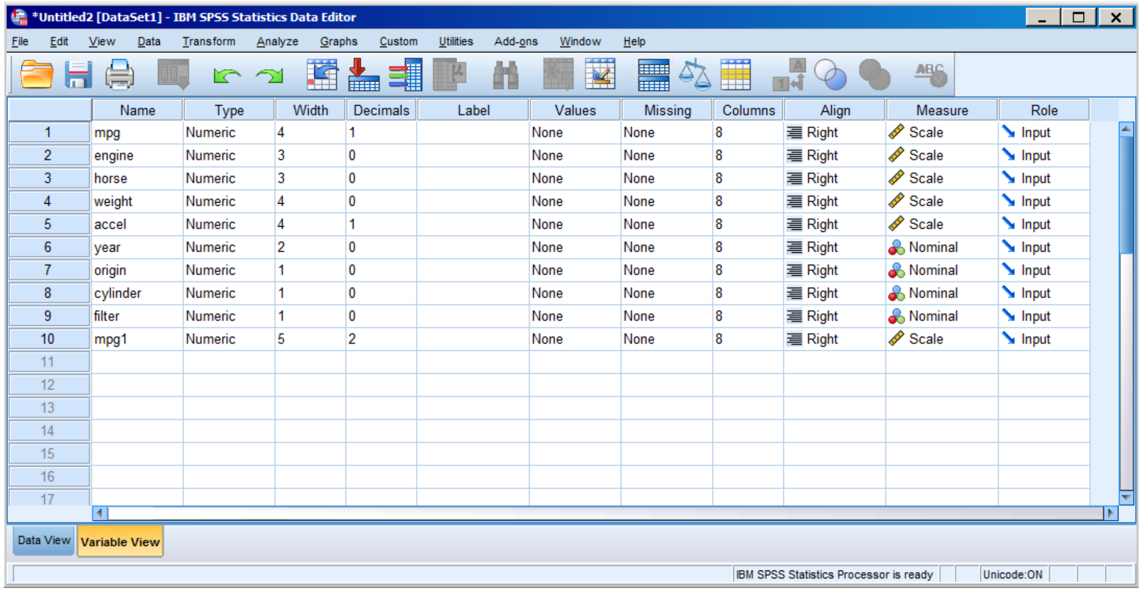


	mpg	engine	horse	weight	accel	year	origin	cylinder
1	18	307	130	3504	12	70	1	8
2	15	350	165	3693	12	70	1	8
3	18	318	150	3436	11	70	1	8
4	16	304	150	3433	12	70	1	8
5	17	302	140	3449	11	70	1	8
6	15	429	198	4341	10	70	1	8
7	14	454	220	4354	9	70	1	8
8	14	440	215	4312	9	70	1	8
9	14	455	225	4425	10	70	1	8
10	15	390	190	3850	9	70	1	8
11	.	133	115	3090	18	70	2	4
12	.	350	165	4142	12	70	1	8
13	.	351	153	4034	11	70	1	8
14	.	383	175	4166	11	70	1	8
15	.	360	175	3850	11	70	1	8
16	15	383	170	3563	10	70	1	8
17	14	340	160	3609	8	70	1	8
18	.	302	140	3353	8	70	1	8
19	15	400	150	3761	10	70	1	8
20	14	455	225	3086	10	70	1	8
21	24	113	95	2372	15	70	3	4
22	22	198	95	2833	16	70	1	6
23	18	199	97	2774	16	70	1	6
24	21	200	85	2587	16	70	1	6
25	27	97	88	2130	15	70	3	4
26	26	97	46	1835	21	70	2	4
27	25	110	87	2672	18	70	2	4
28	24	107	90	2430	15	70	2	4
29	25	104	95	2375	18	70	2	4
30	26	121	113	2234	13	70	2	4
31	21	199	90	2648	15	70	1	6
32	10	360	215	4615	14	70	1	8
33	10	307	200	4376	15	70	1	8
34	11	318	210	4382	14	70	1	8
35	9	4	93	732	9	0	.	.
36	27	97	88	2130	15	71	3	4
37	28	140	90	2264	16	71	1	4

Figure 8. Changing data view/variable view

Define variable properties

To define or change the attributes of variables, change to “Variable View” to see a list of all the variables with their properties from the current data file. Click or double click the variable you would like to specify or change. Descriptions of each attribute are shown below in Figure 9.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	mpg	Numeric	4	1		None	None	8	Right	Scale	Input
2	engine	Numeric	3	0		None	None	8	Right	Scale	Input
3	horse	Numeric	3	0		None	None	8	Right	Scale	Input
4	weight	Numeric	4	0		None	None	8	Right	Scale	Input
5	accel	Numeric	4	1		None	None	8	Right	Scale	Input
6	year	Numeric	2	0		None	None	8	Right	Nominal	Input
7	origin	Numeric	1	0		None	None	8	Right	Nominal	Input
8	cylinder	Numeric	1	0		None	None	8	Right	Nominal	Input
9	filter	Numeric	1	0		None	None	8	Right	Nominal	Input
10	mpg1	Numeric	5	2		None	None	8	Right	Scale	Input
11											
12											
13											
14											
15											
16											
17											

Figure 9. Variable attributes.

Name is the name of the variable. Rules for establishing variable names can be found on **IBM SPSS** *help → Command Syntax Reference → Universals → Variables → Variable Names.*

Type is the type of a variable. Common options are Numeric for numbers, Date for dates, and String for character strings. The string option allows the user to type in any set of characters including punctuation marks and blank spaces. It is ideal for inputting open- ended questions which are not coded.

Width is the maximal number of characters or digits allowed for a variable. Generally a width large enough to accommodate all the possible values of the variable should be chosen; otherwise any values with length greater than the specified value will be truncated.

Decimals are valid for numeric variables only. It specifies the number of decimals to be kept for a variable. All the extra decimals will be rounded up and the rounded numbers will be used in all the analysis, so be careful to specify the number of decimals to fit the required precision.

Label is the descriptive label for a variable. One can assign descriptive variable labels up to 256 characters long, and variable labels can contain spaces and reserved characters not allowed in variable names.

Values is the descriptive value labels for each value of a variable. This is particularly useful if the data file uses numeric codes to represent non-numeric categories (for example, codes of 1 and 2 for male and female).

Missing specifies some data values as user-missing values. Refer to the Missing Values section for more detail.

Columns is the column width for a variable. Column formats affect only the display of values in the Data Editor. Changing the column width does not change the defined width of a variable. If the defined and actual width of a value are wider than the column, asterisks (*) are displayed in the Data view. Column widths can also be changed in the Data view by clicking and dragging the column borders.

Align controls the display of data values and/or value labels in the Data view. The default alignment is right for numeric variables and left for string variables. This setting affects only the display in the Data view.

Measure is the level of measurement as scale (numeric data on an interval or ratio scale), ordinal, or nominal. Nominal and ordinal data can be either string (alphanumeric) or numeric. Nominal and ordinal are both treated as categorical. The variable, origin (Country of Origin) is measured on a nominal scale as the cars are distinguished on the basis of a name or label, i.e. American, European, and Japanese; whilst the variable gallon (miles per gallon) is measured on a scale, specifically, ratio measurement scale because the difference between measurements and ratios are meaningful, and that they have a true zero value.

To download the Cars data file as an SPSS file (i.e. with the .sav extension and all variable attributes edited as in the example) already click here.

Missing Values

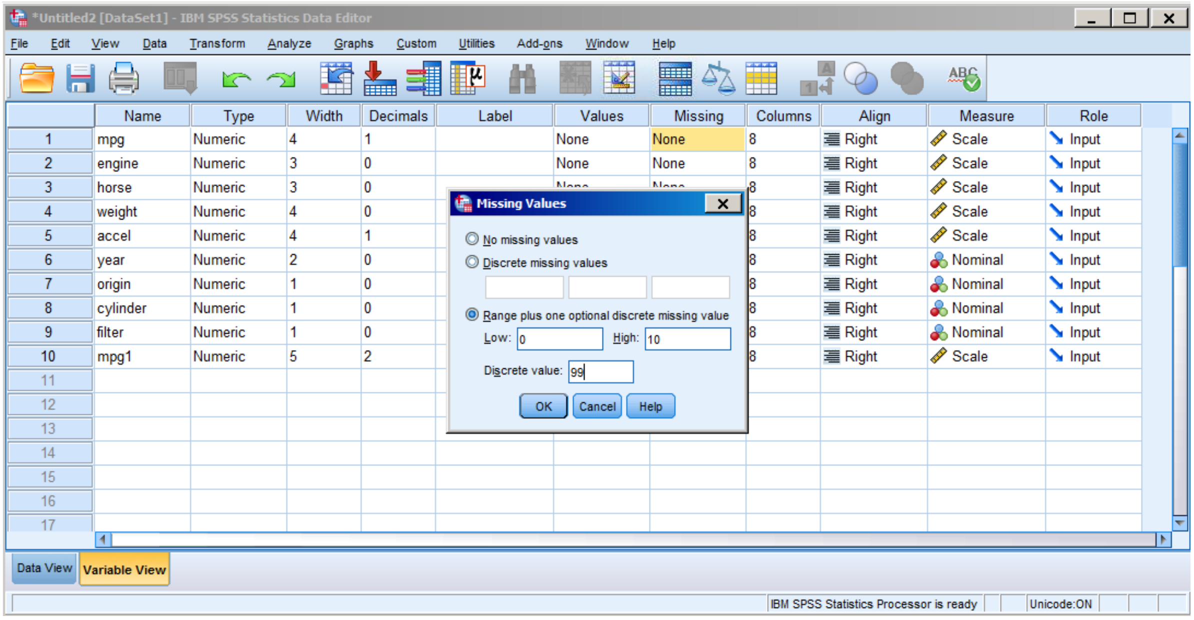
Missing values are a topic that deserves special attention. This section explains why they arise and how to define them. In SPSS there are two types of missing values: user defined missing values and system missing values. By default in SPSS, both types of missing values will be disregarded in all statistical procedures, except for analyses devoted specifically to missing values, for example, replacing missing values. In frequency tables, missing values will be shown, but they will be marked as such and will not be used in computation.

User Defined Missing Values

User defined missing values indicate data values that are either missing, due to reasons like non-response, or are not desired to be used in most analyses (e.g. “Not Applicable”.) By default SPSS uses “.” to represent missing values. In some cases, there might be the need to distinguish between data missing because a respondent refused to answer and data missing because the question did not apply to that respondent, and thus would like more than one expression for missing values. One can achieve this by setting up the “Missing” property of the corresponding variable to specify some data values as missing values. These options allows one to enter up to three discrete missing values, a range of missing values, or a range plus one discrete missing value.

All string values, including null or blank values, are considered valid values unless they are explicitly defined as missing. To define null or blank values as missing for a string variable, enter a single space in one of the fields for discrete missing values.

In Figure 8, you will notice missing values denoted by “.”, for the variable *mpg* observations 11-15. The example in Figure 10 below shows how to specify user defined missing values for variable *mpg* by setting up its *Missing*” property.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	mpg	Numeric	4	1		None	None	8	Right	Scale	Input
2	engine	Numeric	3	0		None	None	8	Right	Scale	Input
3	horse	Numeric	3	0		None	None	8	Right	Scale	Input
4	weight	Numeric	4	0				8	Right	Scale	Input
5	accel	Numeric	4	1				8	Right	Scale	Input
6	year	Numeric	2	0				8	Right	Nominal	Input
7	origin	Numeric	1	0				8	Right	Nominal	Input
8	cylinder	Numeric	1	0				8	Right	Nominal	Input
9	filter	Numeric	1	0				8	Right	Nominal	Input
10	mpg1	Numeric	5	2				8	Right	Scale	Input
11											
12											
13											
14											
15											
16											
17											

Missing Values

○ No missing values

○ Discrete missing values

● Range plus one optional discrete missing value

Low: 0 High: 10

Discrete value: 99

OK Cancel Help

Figure 10. User defined missing values

System Missing Values

System missing values occur when no value can be obtained for a variable during data transformations. For example, if there are two variables, one indicating a person's gender and the other whether she or he is married and you create a new variable that tells whether (a) a person is male and married, (b) female and married, (c) male and not married, all females that are not married will have a system missing value (".") instead of a real value.

Modifying and creating new variables

Insert

The easiest way to manually input a new variable is to scroll through the data view spreadsheet horizontally until the first empty column is encountered, and entering in the data. The new variable can be named appropriately in the variable view spreadsheet. Alternatively, selecting the "Insert Variable" option under the "Data" menu allows insertion of a new variable at other locations in the table. By default, this inserts the new variable in the first column of the spreadsheet, but this can be changed by highlighting the column to the right of the desired location.

Recode

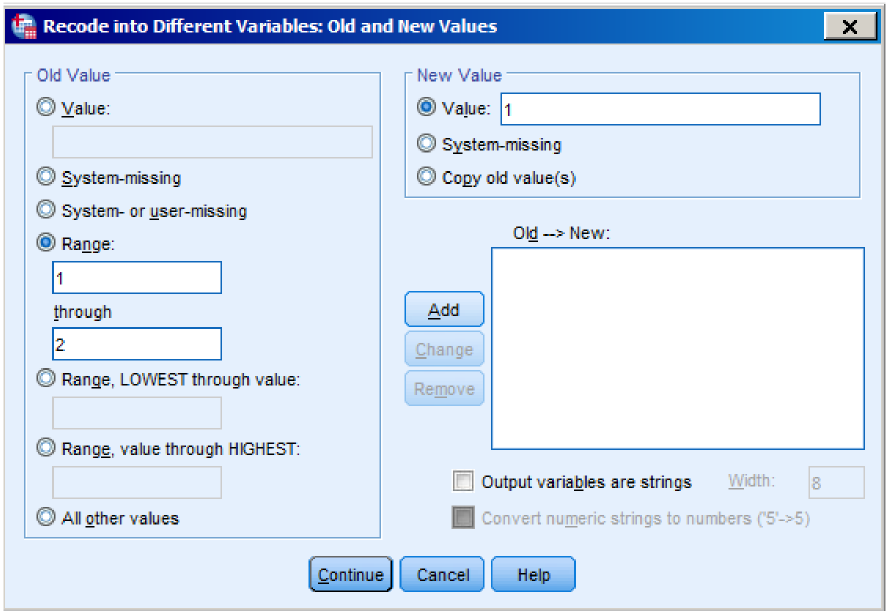
The recode function can be used to collapse ranges of data into categorical variables, and reassigning existing values to other values. To create a new variable as a function of another (log, sin, etc), use "Compute" (described in the next section.)

1. Select *Recode into Different Variables* under the *Transform* menu. *Recoding into Same Variables* is not recommended, since it will change existing variables and you will lose the original values.

2. Select each variable to be transformed, and move it into the section on the right hand side using the ➡ button. Note that the same transformation will be applied to all of these variables. If different types of transformations are required, each transformation needs to be done separately.

3. If new variables are being created, define name for the output variable on the right hand side. If desired, a label can be entered as well, though it is not required. Once the desired name and label are entered, you must click the *Change* button.

4. Select the *Old and New Values* button and the window below (Figure 11) will appear. In the *Old Value* side of the window, select the appropriate original values to be recoded.


Recode into Different Variables: Old and New Values
Old Value
○ Value:
○ System-missing
○ System- or user-missing
○ Range, LOWEST through value:
○ Range, value through HIGHEST:
○ All other values
New Value
○ System-missing
○ Copy old value(s)
Old --> New:
Add
Change
Remove
□ Output variables are strings
Width: 8
□ Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

● Range: 1 through 2

● Value: 1

Figure 11. Recode into Different Variables: Old and New Values window

a) By selecting *Value* one can specify a value to replace (*e.g.* “male” or “1”). It is case sensitive, so “A” and “a” are considered to be different values.

b) *System-Missing* and *System- or user-missing* allows missing values to be replaced by actual values. It is not recommended to recode missing values using this method, unless there is a strong reason to do so. Missing values should be handled with care, using techniques such as multiple imputation.

c) The three range options partition numeric variables into categories. The above figure demonstrates how a range of continuous variables can be condensed into a category. Rather than running any procedures to find out the range of variables, the range options with *LOWEST through value:* and *value through HIGHEST:* can be used to catch every point in the data set.

d) *All other values* can be used to pick up values not specifically referenced elsewhere.

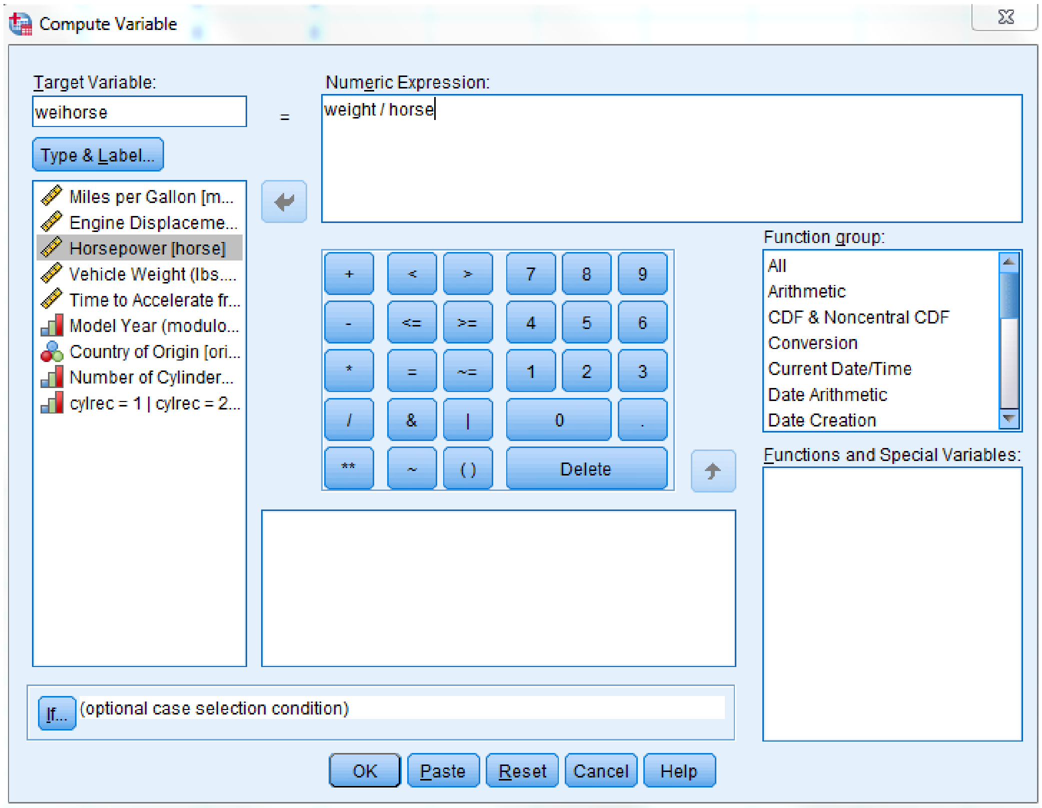
5. On the *New Value* side, type in the new value. Then click the *Add* button to add it to the *Old->New* list. When recoding into different variables, one has the option of changing numbers to strings, or converting numbers saved as strings to numbers. Unless otherwise specified, the new variable will

be saved in the same format as the original variable. Click *Continue* to close the window. On the main screen, click *OK* or *Paste* to finish.

Compute

Suppose you want to create a new variable, measuring the ratio of the vehicle's weight to its horsepower, you define the new variable as *weihorse* for the weight per unit horsepower.

To create a new variable as a function of one or more existing variables, select *Compute* from the *Transform* menu. Enter the name of the new variable, *weihorse*, in the *Target Variable* box. In the *Numeric Expression* box, use the keypad, function list, and the variable list to write out the equation used to compute the new variable, (in this example: weight/horse). Click *OK* or *Paste* to close the window.



Compute Variable

Target Variable: weihorse

Numeric Expression: weight / horse

Figure 12. Compute new variable, weihorse

Creating Graphs

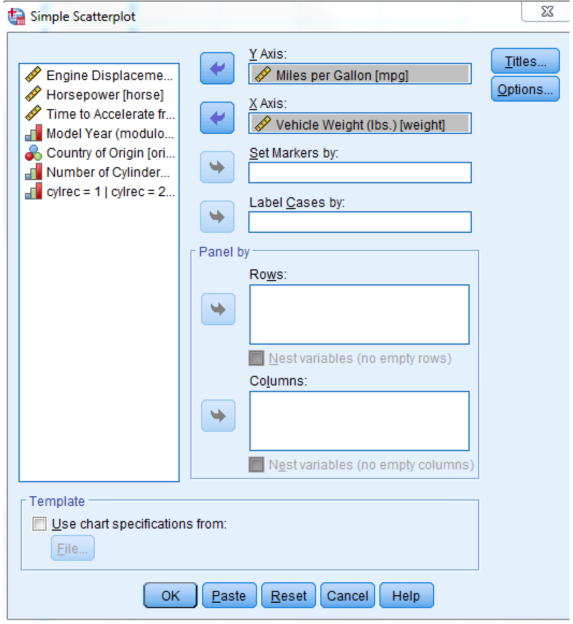
Graphs in SPSS may be generated using one of two options. The first option is the *Legacy Dialogs*, which allows one to create basic charts and graphs. The second option is to use the *Chart Builder* which allows one to generate charts either from a predefined gallery or by specifying individual parts (for example, axes and bars). The steps to create a few common graphs are shown below. However, SPSS has the ability to produce many other graphs such as population pyramid, error bar, and 3-D bar chart. The *Chart Builder* allows more flexibility in creating graphs.

For any graph generated in SPSS, one can double click on the graph to invoke a Chart Editor window, inside which one can double click any part of the graph to edit it. For instance, the title of the graph can be edited by double clicking the title area of the graph. So are labels of axis, type of points, color of lines, size of the box, and other features of a graph.

Scatterplot

Suppose, we seek to investigate the linear relationship between miles per gallon and the vehicle weight, we first plot a scatterplot to see the direction in which they are related.

We will introduce the Simple Scatterplot. In the "Graphs" menu, choose *Legacy Dialogs* → *Scatter/dot*. Select *Simple Scatter*, click on the *Define* button to get the window shown below. Select a variable for the Y-axis and a variable for the X-axis. These variables must be numeric and not in date format. One can also select a categorical variable to define rows of panels and another categorical variable to define columns of panels. Using the "Title" button one may specify the title, subtitles and the footnotes for the plot. In the following example we are plotting mpg against vehicle weight, using model year to define rows of panels, see Figure 13 and 14.



Simple Scatterplot

Engine Displaceme...
Horsepower [horse]
Time to Accelerate fr...
Model Year (modulo...
Country of Origin [ori...
Number of Cylinder...
cylrec = 1 | cylrec = 2...

Y Axis: Miles per Gallon [mpg]

X Axis: Vehicle Weight (lbs.) [weight]

Set Markers by:

Label Cases by:

Panel by

Rows:

□ Nest variables (no empty rows)

Columns:

□ Nest variables (no empty columns)

Template

□ Use chart specifications from:

Figure 13. Generating a simple scatterplot

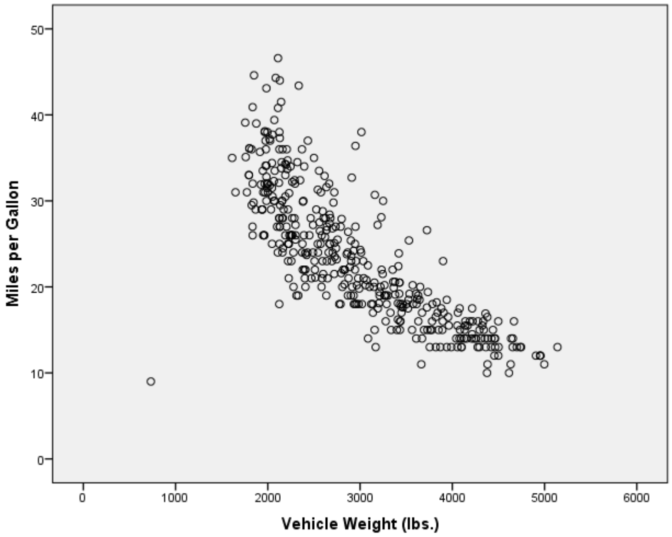
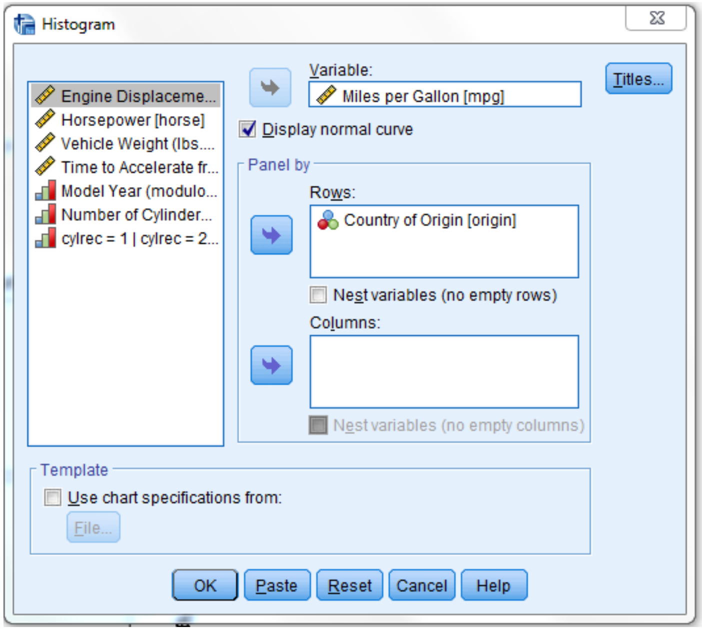


Figure 14: Scatterplot of Miles per Gallon vs. Vehicle weight (lbs.)

Histogram

A histogram shows the distribution of a single numeric variable. By selecting *Legacy Dialogs* → *Histogram* in the *Graphs* menu, one can generate a histogram. One can check the *Display normal curve* option to require an estimated normal curve displayed over the histogram. Suppose, you want to draw histograms of the miles per gallon based on the origin of the vehicle, in the *Panel by* dialogue box, either in the rows or columns, you can put the variable, *origin*, as shown below in Figure 15.



Histogram

Variable: Miles per Gallon [mpg]

☑ Display normal curve

Panel by

Rows: Country of Origin [origin]

☐ Nest variables (no empty rows)

Columns:

☐ Nest variables (no empty columns)

Template

☐ Use chart specifications from:

Figure 15. Generating a histogram

Graph titles can be specified by clicking on the *Titles* button, and these will be shown in the output as in Figure 16, below.

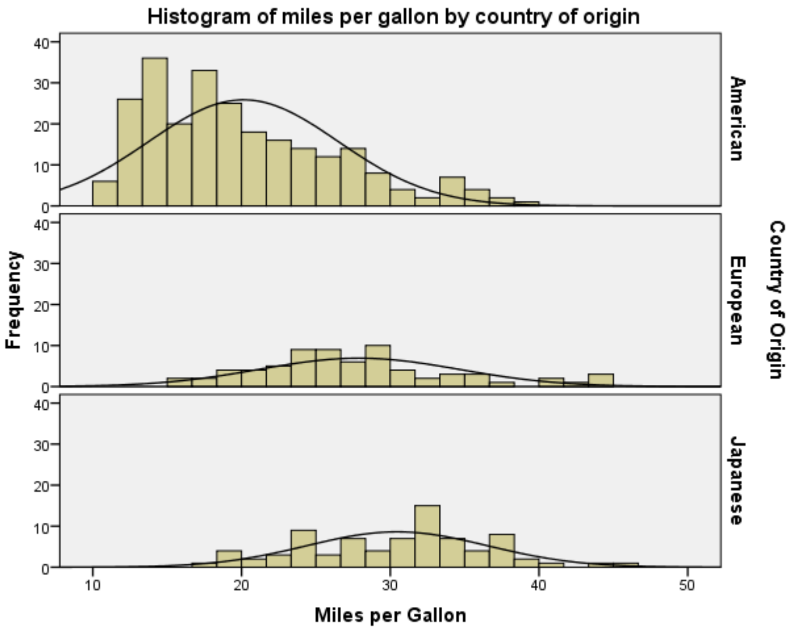
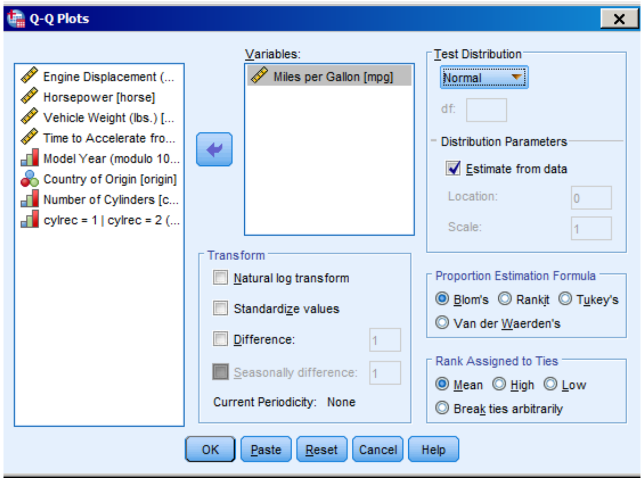


Figure 16: Histogram of Miles per Gallon by country of origin

Q-Q Plot

The Q-Q Plot (quantile-quantile plot) procedure plots the quantiles of a variable's distribution against the quantiles of a variable from a test distribution. Q-Q plots are generally used to investigate whether the distribution of a variable is consistent with a proposed distribution. Specifically, Q-Q plots can be used to investigate whether a variable (e.g. residuals in a regression model) follows a Normal distribution. If the distribution of the variable and the proposed distribution are the same, points in the Q-Q plot follow a straight line. If the distributions are not similar, points in the Q-Q plot deviate from the straight line.

Suppose you want to generate a Q-Q plot with a Normal distribution as the test distribution. Select *Descriptive Statistics* → *Q-Q Plots* in the *Analyze* menu. Enter the variables you want to plot into the *Variables* box, and select *Normal* by clicking *Test Distribution*. Click *OK* to generate the plot.



Engine Displacement (...
Horsepower [horse]
Vehicle Weight (lbs.) [...
Time to Accelerate fro...
Model Year (modulo 10...
Country of Origin [origin]
Number of Cylinders [c...
cylrec = 1 | cylrec = 2 (...

Variables:
Miles per Gallon [mpg]

Test Distribution
Normal
df:

Distribution Parameters
☑ Estimate from data
Location: 0
Scale: 1

Transform
□ Natural log transform
□ Standardize values
□ Difference: 1
□ Seasonally difference: 1
Current Periodicity: None

Proportion Estimation Formula
● Blom's ○ Rankit ○ Tukey's
○ Van der Waerden's

Rank Assigned to Ties
● Mean ○ High ○ Low
○ Break ties arbitrarily

Figure 17. Generating a QQ plot

The Q-Q plot output for the previous example is given below in Figure 18.

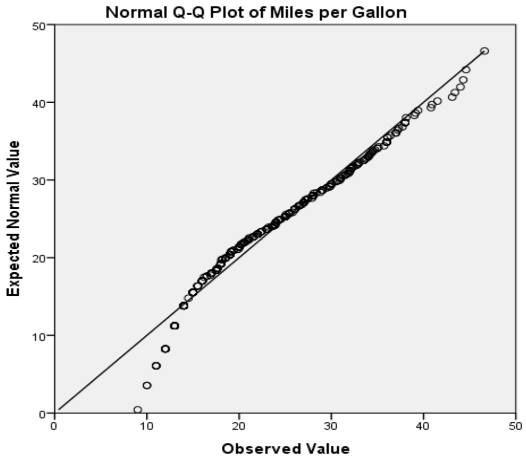
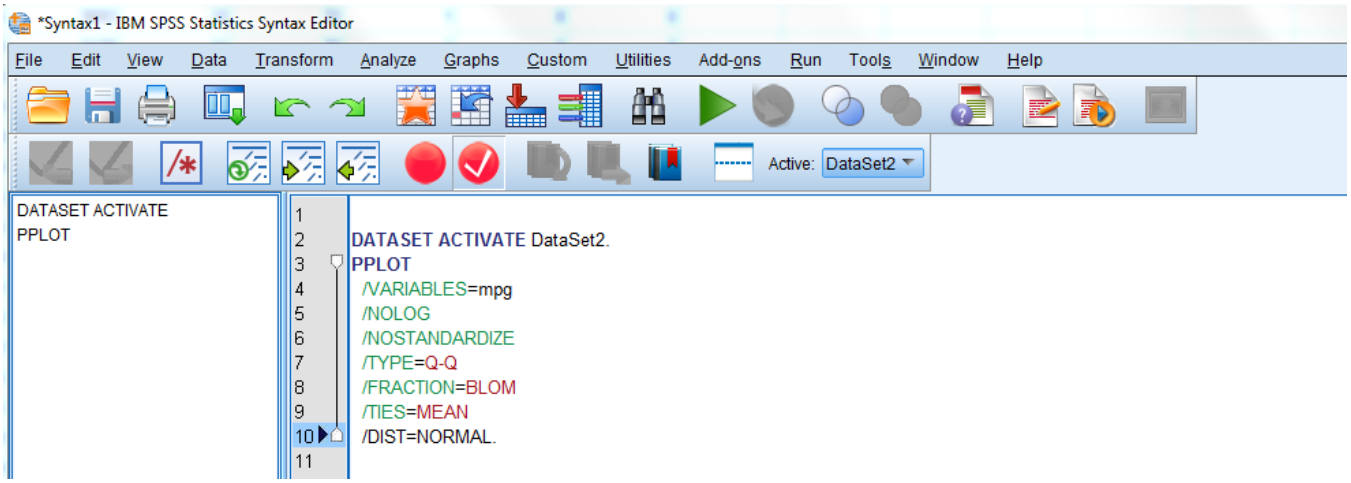


Figure 18: Q-Q plot of Miles per Gallon

Syntax File

Here is an example of the syntax for the Q-Q plot in Figure 18. After selecting *Descriptive Statistics* → *Q-Q Plots* in the *Analyze* menu. You enter the variables, *mpg* (Miles per Gallon), you want to plot into the *Variables* box, and select *Normal* by clicking *Test Distribution*. Then you click *Paste* to generate the syntax, below in Figure 19.



```
DATASET ACTIVATE DataSet2.
PPLOT
  /VARIABLES=mpg
  /NOLOG
  /NOSTANDARDIZE
  /TYPE=Q-Q
  /FRACTION=BLOM
  /TIES=MEAN
  /DIST=NORMAL.
```

Figure 19: Syntax for the QQ plot for the variable "Miles per Gallon", mpg

You can save the syntax as a *.sps* file for later access in running the analysis.

Analyze Data

Descriptive Statistics

In the *Analyze* menu, the option *Descriptive Statistics* produces a submenu with the choices *Frequencies*, *Descriptives*, *Explore*, *Crosstabs*, and *Ratio*. Of these, *Crosstabs* and *Descriptives* have some particularly useful features which this manual will cover. For more information on the other three, more information can be found in the SPSS help menu, which is discussed on section “Help in SPSS” of this manual.

Descriptives

The descriptives procedure calculates univariate statistics for selected variables. In addition, it provides the option of creating a standardized variable for the selected variables. Simply check the box at the bottom of the window to save the standardized variable. The options menu provides a list of univariate statistics available. For more statistics or computing statistics by group, see the *Means* procedure under *Compare Means*.

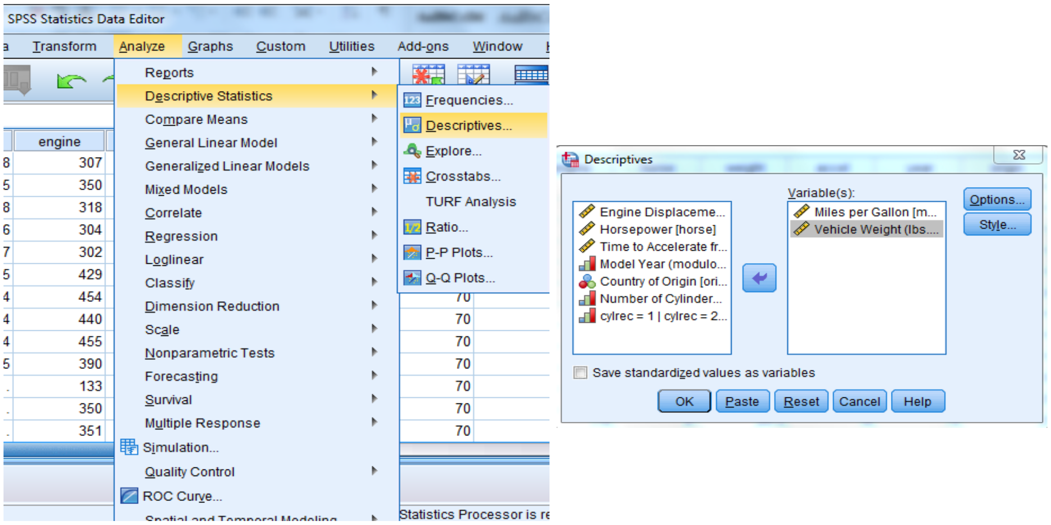
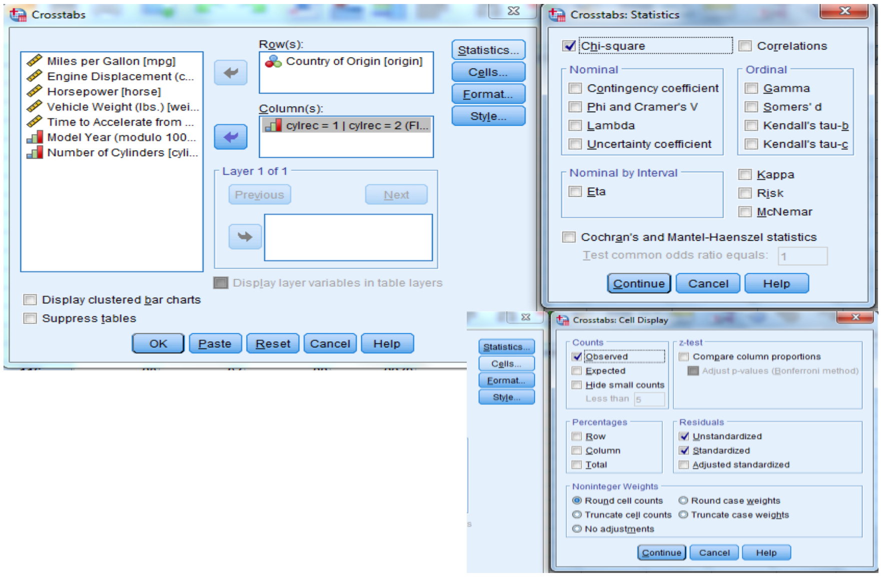


Figure 20: Descriptives

Crosstabs

The Crosstabs procedure forms two-way and multi-way tables and provides a variety of tests and measures of association for two-way tables. Multi-way tables are formed using the ‘Layer’ button. Note that tests are not made across layers. When layers are used, comparisons are made for the row and column variables at each value of the layer variable. The *Statistics* button at the bottom allows various statistics to be computed, including correlations and Chi-square tests. To help uncover patterns in the data that contribute to a significant chi-square test, the *Cells* button provides options for displaying expected frequencies and three types of residuals (deviates) that measure the difference between observed and expected frequencies. Each cell of the table can contain any combination of counts, percentages, and residuals selected, see Figure 21 and 22.



Crosstabs

- Miles per Gallon [mpg]
- Engine Displacement (c...
- Horsepower [horse]
- Vehicle Weight (lbs.) [wei...
- Time to Accelerate from ...
- Model Year (modulo 100...
- Number of Cylinders [cyli...

Row(s): Country of Origin [origin]

Column(s): cylrec = 1 | cylrec = 2 (FI...

Layer 1 of 1

□ Display layer variables in table layers

□ Display clustered bar charts

□ Suppress tables

Crosstabs: Statistics

☑ Chi-square

□ Correlations

Nominal: □ Contingency coefficient □ Phi and Cramer's V □ Lambda □ Uncertainty coefficient

Ordinal: □ Gamma □ Somers' d □ Kendall's tau-b □ Kendall's tau-c

Nominal by Interval: □ Eta

□ Kappa □ Risk □ McNemar

□ Cochran's and Mantel-Haenszel statistics

Test common odds ratio equals: 1

Crosstabs: Cell Display

Counts: ☑ Observed □ Expected □ Hide small counts Less than 5

z-test: □ Compare column proportions □ Adjust p-values (Bonferroni method)

Percentages: □ Row □ Column □ Total

Residuals: ☑ Unstandardized ☑ Standardized □ Adjusted standardized

Noninteger Weights: ● Round cell counts ○ Round case weights ○ Truncate cell counts ○ Truncate case weights ○ No adjustments

Figure 21: Crosstabs

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Country of Origin * cylrec = 1 | cylrec = 2 (FILTER)	398	98.0%	8	2.0%	406	100.0%

Country of Origin * cylrec = 1 | cylrec = 2 (FILTER) Crosstabulation

			cylrec = 1 | cylrec = 2 (FILTER)		Total
			Not Selected	Selected	
Country of Origin	American	Count	107	146	253
		Residual	39.0	-39.0	
		Standardized Residual	4.7	-2.9	
	European	Count	0	70	70
		Residual	-18.8	18.8	
		Standardized Residual	-4.3	2.6	
	Japanese	Count	0	75	75
		Residual	-20.2	20.2	
		Standardized Residual	-4.5	2.7	
Total		Count	107	291	398

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	83.873[a]	2	.000
Likelihood Ratio	118.660	2	.000
Linear-by-Linear Association	71.467	1	.000
N of Valid Cases	398		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 18.82.

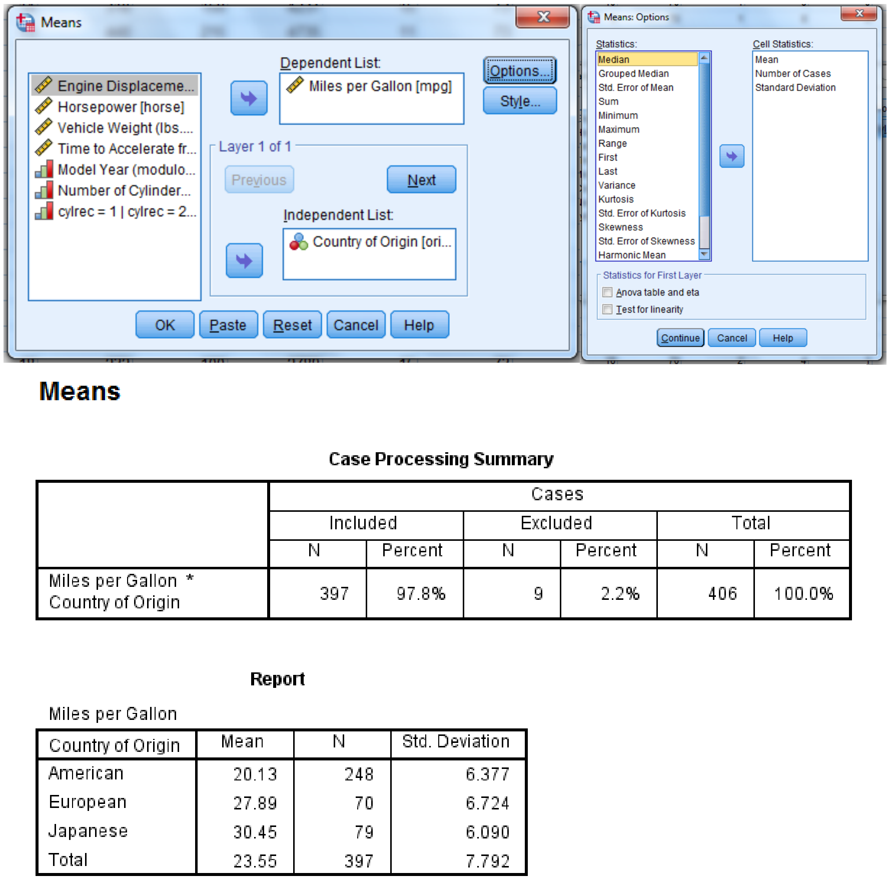
Figure 22: Crosstabs results

Compare Means

From the *Compare Means* option in the *Analyze* menu, one can perform t-tests, and one-way ANOVA, and calculate univariate statistics for variables.

Means

The *Means* window includes two fields: *Dependent List* and *Independent List*. The procedure calculates univariate statistics (e.g. mean, median, and standard deviation) for variables in the *Dependent List* field, grouped using variables in the *Independent List* field. By default, the mean, standard deviation, and sample size are displayed. More statistics can be selected using the *Options* button, see Figure 23.



Dependent List: Miles per Gallon [mpg]

Independent List: Country of Origin [ori...

Cell Statistics: Mean, Number of Cases, Standard Deviation

Means

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Miles per Gallon * Country of Origin	397	97.8%	9	2.2%	406	100.0%

Report

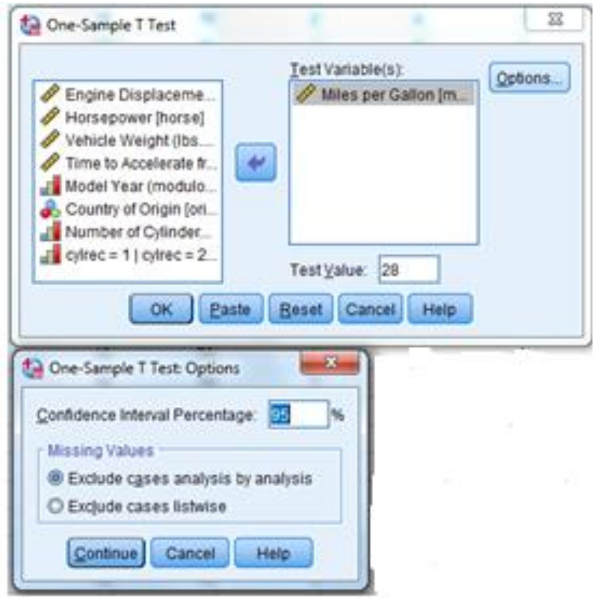
Miles per Gallon

Country of Origin	Mean	N	Std. Deviation
American	20.13	248	6.377
European	27.89	70	6.724
Japanese	30.45	79	6.090
Total	23.55	397	7.792

Figure 13: Means Procedure and Options Window

One Sample T test

The One Sample t-test procedure (Figure 24 and 25) analyzes whether the mean of a single variable differs from a specified constant. The options here allow you to specify the confidence level and how to handle the missing values. The example in Figure 23 tests whether the mean of *mpg* differs from 28 using a confidence level of 95%. The options for Missing Values are explained below and the output is shown in Figure , where the numbers in the second table correspond to the value of the t statistic, the degree of freedom for the t test, the p-value of the test (here, .000 is less than 0.05 which indicates that the mean of *mpg* significantly differs from 28), the difference between the real mean and the test value, the lower and the upper values of the 95% confidence intervals for the difference. From the table, we can see that the 95% confidence interval for *mpg* is (28-5.26, 28-3.72) = (22.74, 24.28).



Test Variable(s): Miles per Gallon [m...

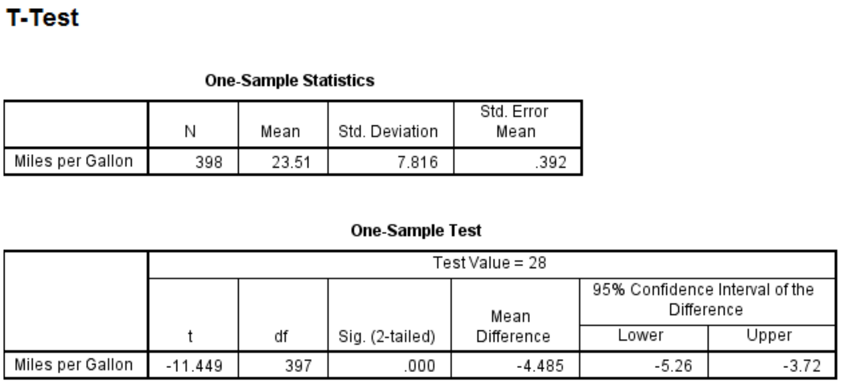
Test Value: 28

Confidence Interval Percentage: 95 %

● Exclude cases analysis by analysis

○ Exclude cases listwise

Figure 24: One-Sample T-test procedure



T-Test

One-Sample Statistics

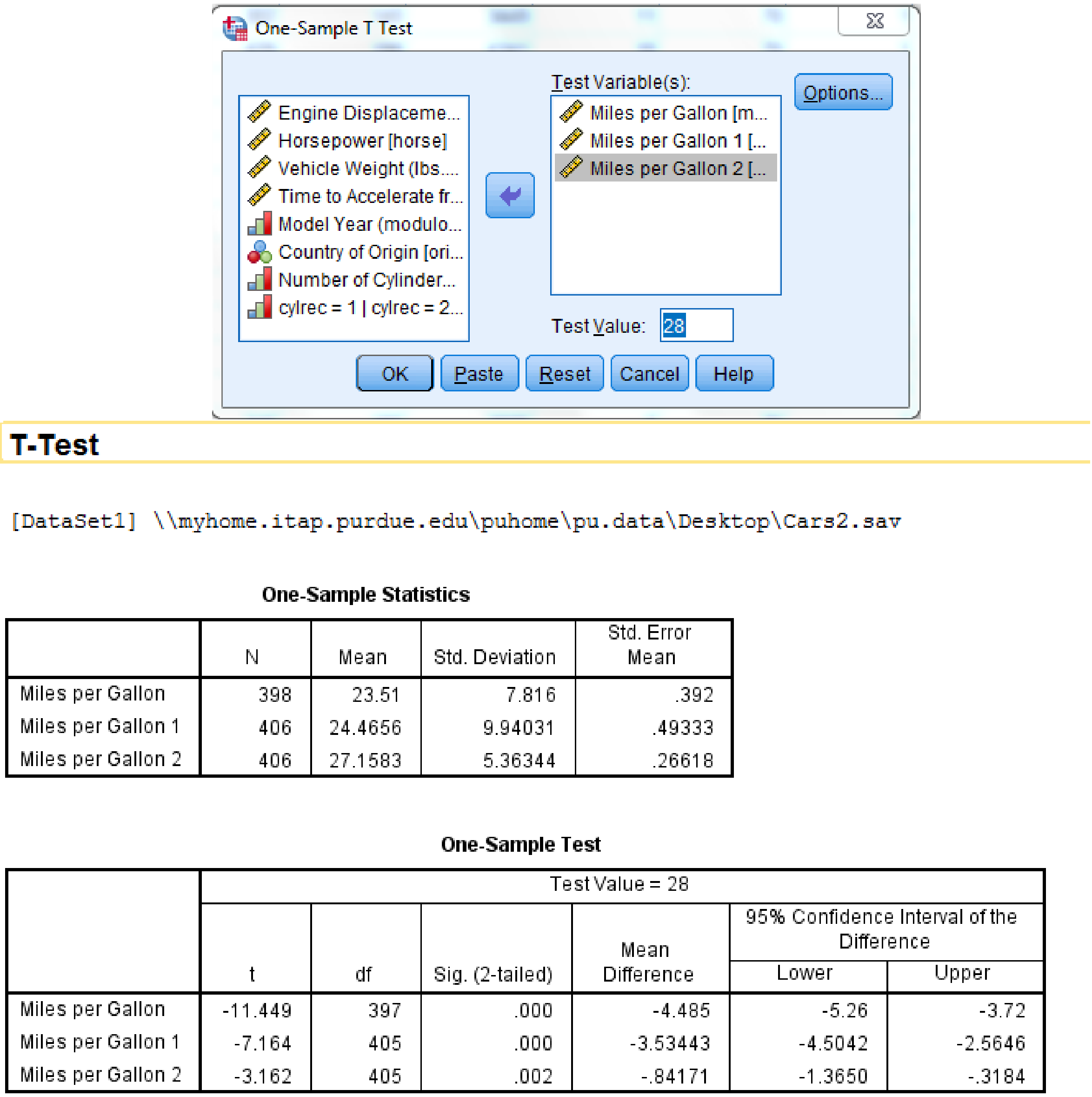
	N	Mean	Std. Deviation	Std. Error Mean
Miles per Gallon	398	23.51	7.816	.392

One-Sample Test

	Test Value = 28					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Miles per Gallon	-11.449	397	.000	-4.485	-5.26	-3.72

Figure 25: One Sample T-test results

If the user has more than one variable to be tested against the same value, one can conduct all the one sample t tests in one step by putting all the variables of interest in the *Test Variable(s)* field. In the case of having missing values, you have two options *Exclude cases analysis by analysis* and *Exclude cases listwise*. If the first option is chosen (as is done most often), each t test will use all cases that have valid data for the variable tested and sample sizes may vary from test to test. If the second option is used, each t test will use only cases that have valid data for all variables used in any of the t tests requested and the sample size is constant across tests. In this CARS data, since we don't have those variables, we create two new variables *mpg1(mean=25, sd=10), mpg2(mean=27, sd=5))* and save this data as CARS2. After then, we can do one sample t test for each 3 variables (*mpg, mpg1, mpg2)* against the value 28, see Figure 26.



Test Variable(s):
Miles per Gallon [m...
Miles per Gallon 1 [...
Miles per Gallon 2 [...

Test Value: 28

T-Test

```
[DataSet1] \\myhome.itap.purdue.edu\puhome\pu.data\Desktop\Cars2.sav
```

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Miles per Gallon	398	23.51	7.816	.392
Miles per Gallon 1	406	24.4656	9.94031	.49333
Miles per Gallon 2	406	27.1583	5.36344	.26618

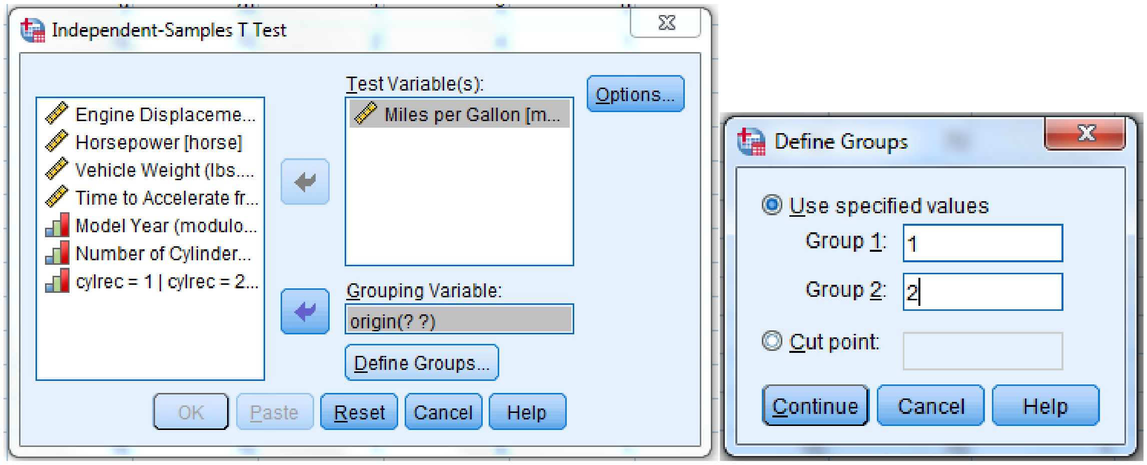
One-Sample Test

	Test Value = 28					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Miles per Gallon	-11.449	397	.000	-4.485	-5.26	-3.72
Miles per Gallon 1	-7.164	405	.000	-3.53443	-4.5042	-2.5646
Miles per Gallon 2	-3.162	405	.002	-.84171	-1.3650	-.3184

Figure 26: One Sample T-test results for multiple tests

Independent Samples T test

The Independent-Samples t-test procedure (Figure 27) compares means for two groups of cases. It has the same options as the One-Sample t test. The example in Figure compares the means of *mpg* for two *origin* groups. You still need to specify which two groups are to be compared as follows: first click on the *Define Groups* tab, then in the pop-up window enter the corresponding values of the *Grouping Variable*, click on *Continue* to return to the main window and the *origin(? ?)* will become *origin(1 2).* SPSS does not allow to type text when you define specified values. Because alphabetical order, the number of specified values is 1: American, 2: European, and 3: Japanese. Finally, click on the activated OK tab to finish the procedure.



Test Variable(s): Miles per Gallon [m...

Grouping Variable: origin(? ?)

● Use specified values

Group 1: 1

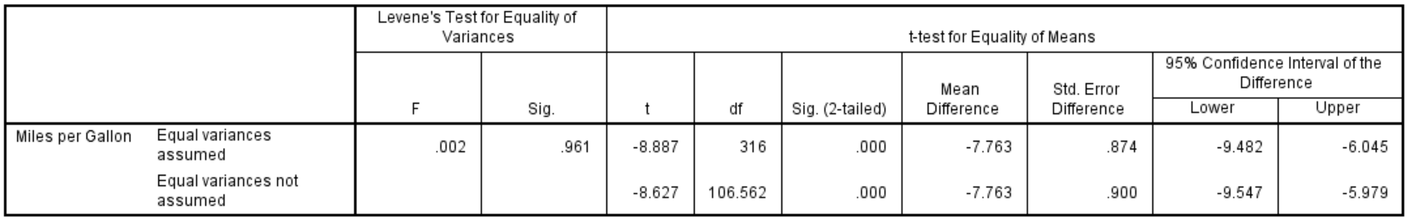
Group 2: 2

Figure 27: Independent Samples *t*-test

In the output (Figure 28), the first table displays statistics for each of the two origin groups, American and European. In the second table, the first two columns are results for testing if the two groups have equal variances (here the big p-value of Levene's Test for Equality of Variance, 0.961, indicates equal variances); the next columns list two testing results according to whether equal variances are assumed or not. They also have meanings similar to those in one sample t test except that the difference now refers to the difference of the two group means. From the table we can see that since p value=0.000 < 0.05, there are significant difference between *mpg* in two different countries. If the first two columns of Figure 28 are results for testing when the two groups do have not equal variances (if the p-value of Levene's test is less then equal to .05), we will read the bottom row of Figure 28. We can see that since p value=0.000 < 0.05, there are significant difference between *mpg* in two different countries.

Group Statistics

	Country of Origin	N	Mean	Std. Deviation	Std. Error Mean
Miles per Gallon	American	248	20.13	6.377	.405
	European	70	27.89	6.724	.804

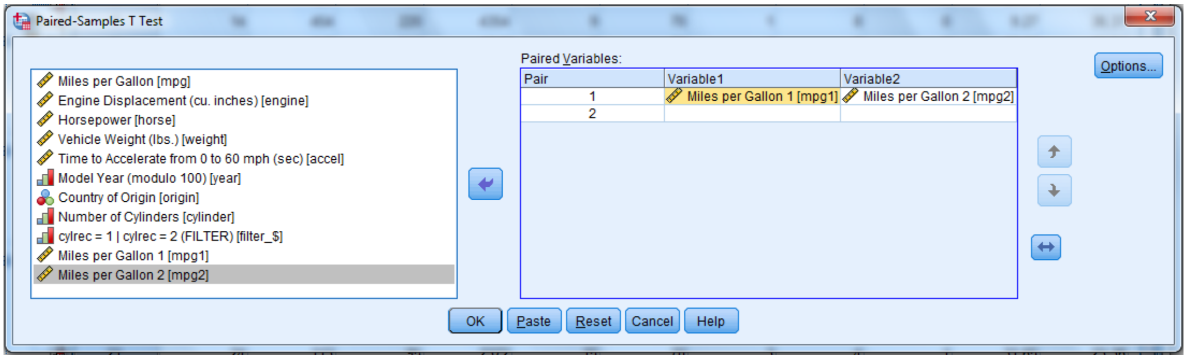


		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Miles per Gallon	Equal variances assumed	.002	.961	-8.887	316	.000	-7.763	.874	-9.482	-6.045
	Equal variances not assumed			-8.627	106.562	.000	-7.763	.900	-9.547	-5.979

Figure 28: Output for Independent Samples *t*-test

Paired Samples T Test (also called Matched Pairs t test)

The Paired-Samples T-Test procedure (Figure 29) compares the means of two variables for a single group. It has the same options as the One-Sample T Test. The example in Cars2 data (Figure 27), compares the means of *mpg1* and *mpg2* (suppose that *mpg1* is mile per gallon from test track1 and *mpg2* is mile per gallon from test track2) for all the observations. Since *mpg1* and *mpg2* are test results from a same car but two different tracks, we need to consider using paired samples t test instead of Independent Samples T Test. The paired variables are selected by highlighting the two variables to be paired in the list and click on the *OK* button.



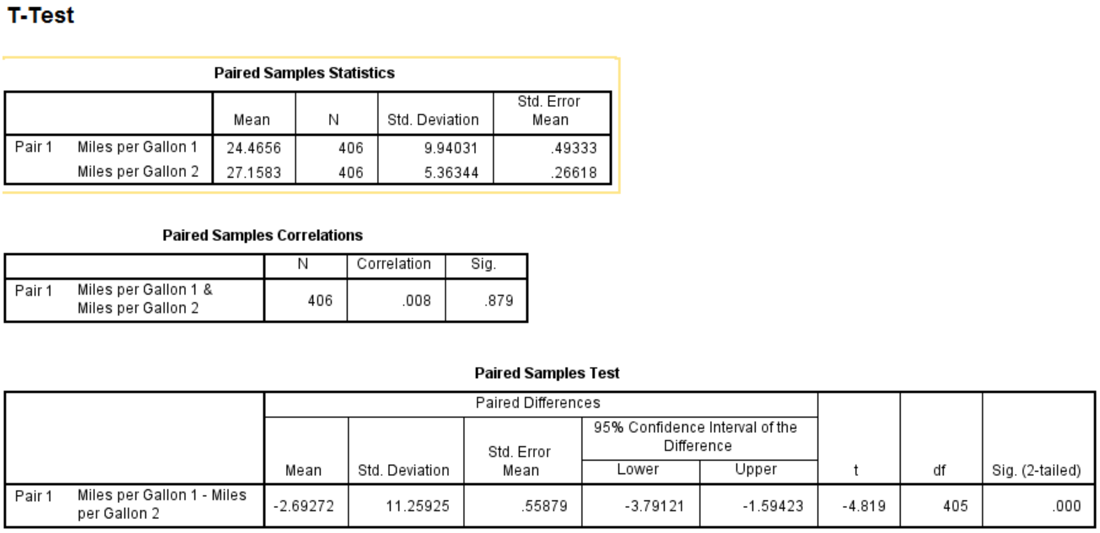
- Miles per Gallon [mpg]
- Engine Displacement (cu. inches) [engine]
- Horsepower [horse]
- Vehicle Weight (lbs.) [weight]
- Time to Accelerate from 0 to 60 mph (sec) [accel]
- Model Year (modulo 100) [year]
- Country of Origin [origin]
- Number of Cylinders [cylinder]
- cylrec = 1 | cylrec = 2 (FILTER) [filter_$]
- Miles per Gallon 1 [mpg1]
- Miles per Gallon 2 [mpg2]

Paired Variables:

Pair	Variable1	Variable2
1	Miles per Gallon 1 [mpg1]	Miles per Gallon 2 [mpg2]
2		

Figure 29: Paired Samples t test

In the output, the first table displays statistics for the two variables; the second table has the correlation between the two variables and the p-value indicating whether the correlation is significant (in Figure 30, p-value = 0.000 which means that there is a significant difference between these two variables); the third table has the format of a one sample t test, testing on whether the paired difference is equal to zero. A difference of zero indicates that the means are the same. Note that the differences will be calculated with the first chosen variable – second chosen variable. In this case *mpg1 – mpg2.*



T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Miles per Gallon 1	24.4656	406	9.94031	.49333
	Miles per Gallon 2	27.1583	406	5.36344	.26618

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Miles per Gallon 1 & Miles per Gallon 2	406	.008	.879

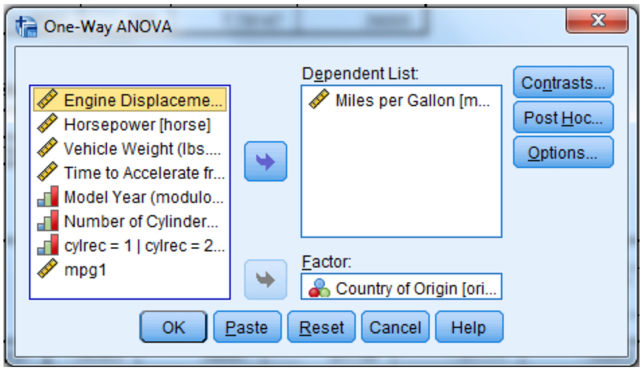
Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Miles per Gallon 1 - Miles per Gallon 2	-2.69272	11.25925	.55879	-3.79121	-1.59423	-4.819	405	.000

Figure 30: Paired t test Output

One-Way ANOVA

The One-Way ANOVA procedure (Figure) produces a one-way analysis of variance for a quantitative dependent variable by a single factor (independent) variable. The example below fits a one-way ANOVA model for *mpg* by factor *origin*. The One-Way ANOVA analysis can also be carried out by following *Analyze* → *General Linear Model* → *Univariate*. The options here are also included in the Univariate option for General Linear Model.

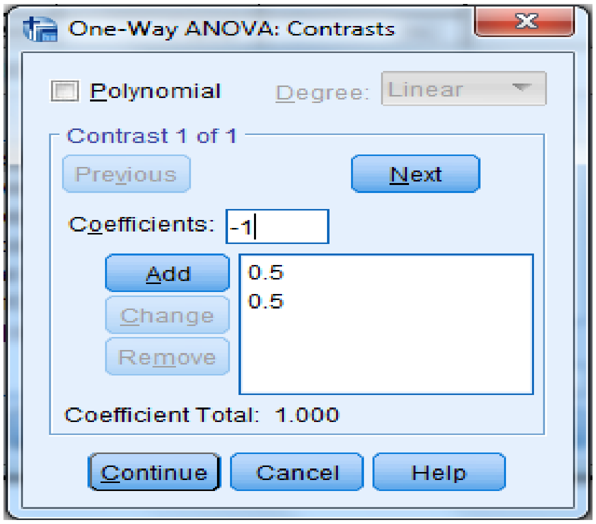


Dependent List: Miles per Gallon [m...

Factor: Country of Origin [ori...

Figure 31: One-Way ANOVA procedure

The Contrasts button allows one to display the tests for contrasts. One example of a contrast is: $\frac{\text{mean of group1+mean of group 2}}{2}$ – mean of group 3 , which compares the average of group 1 mean and group 2 mean with the mean of group 3 and can be used to test whether group 3 is significantly different from the other two groups. This contrast is specified as below. Each coefficient is entered in the *Coefficients* field first, and put into the big field below by clicking on the *Add button*. The order of the coefficients is important because it corresponds to the (ascending) order of the category values of the factor variable. Notice the coefficients for a contrast MUST sum to zero. The Polynomial option is used to test for a polynomial (Linear, Quadratic, Cubic, 4th or 5th, chosen from the pull-down Degree list) trend of the dependent variable across the ordered levels of the factor variable. Then click *Continue.*



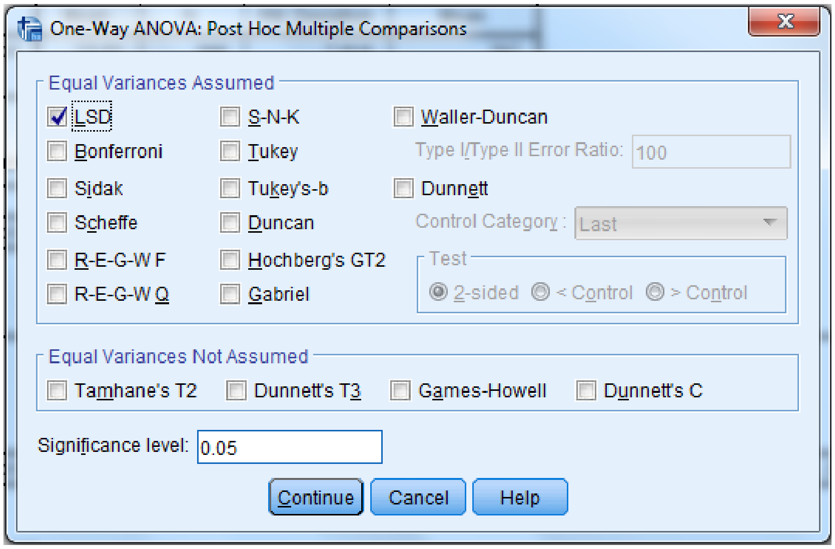
□ Polynomial Degree: Linear

Coefficients: -1

0.5
0.5

Coefficient Total: 1.000

Figure 32: ANOVA Contrasts



One-Way ANOVA: Post Hoc Multiple Comparisons

Equal Variances Assumed

☑ LSD □ S-N-K □ Waller-Duncan

□ Bonferroni □ Tukey Type I/Type II Error Ratio: 100

□ Sidak □ Tukey's-b □ Dunnett

□ Scheffe □ Duncan Control Category: Last

□ R-E-G-W F □ Hochberg's GT2 Test

□ R-E-G-W Q □ Gabriel ● 2-sided ○ < Control ○ > Control

Equal Variances Not Assumed

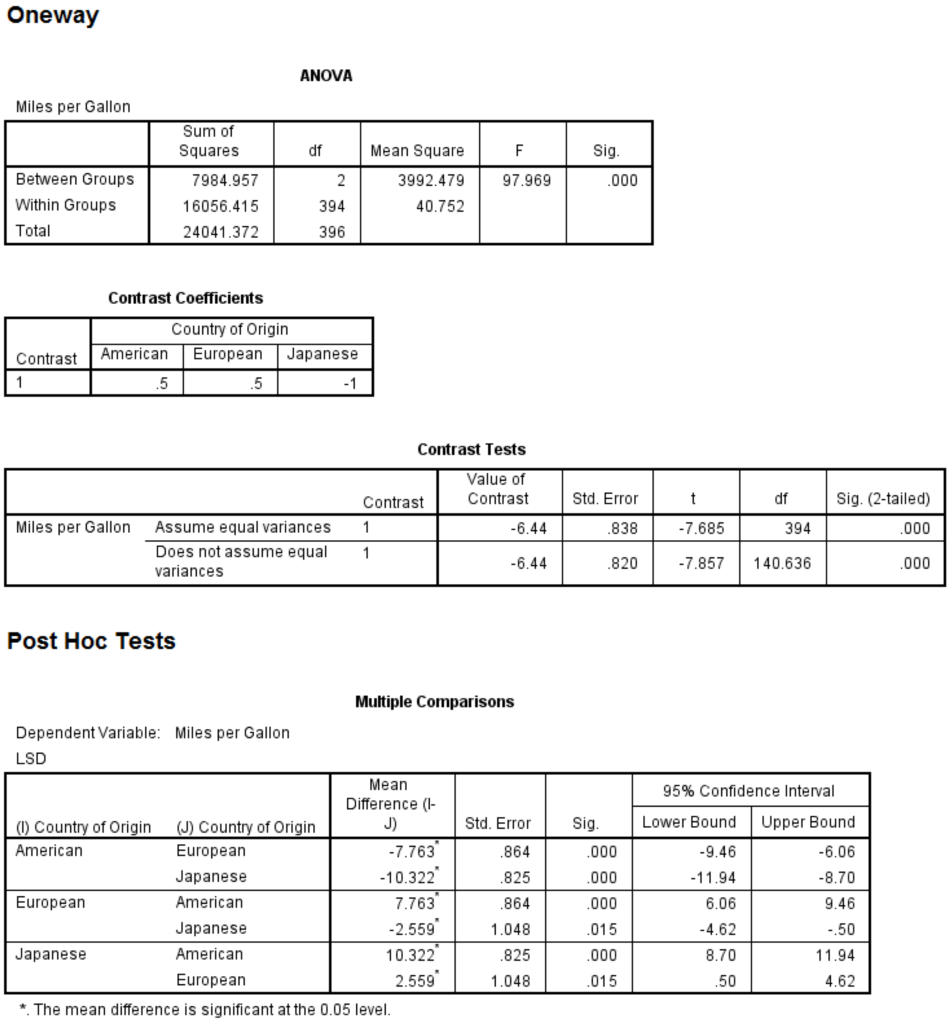
□ Tamhane's T2 □ Dunnett's T3 □ Games-Howell □ Dunnett's C

Significance level: 0.05

Figure 33: Post Hoc Tests

The Post Hoc button allows one to specify the post hoc tests to be displayed. Once it has been determined that differences exist among the means, post hoc range tests and pairwise multiple comparisons can determine which means differ. Range tests identify homogeneous subsets of means that are not different from each other. Pairwise multiple comparisons test the difference between each pair of means, and yield a matrix where asterisks indicate significantly different group means at an alpha level of 0.05. The available tests are shown in the Figure 33. In the example, we chose to conduct a LSD (Fisher's Least Significant Difference) multiple comparison test.

The output of the above example is presented in Figure 34. First shown is the ANOVA table in which the small p-value .000 indicates *mpg* (miles per gallon) for a car is different among the three product *origin*s; the second table displays the contrast coefficients; the third table contains the test results of the contrast based on whether equal variances are assumed or not; the last table shows the LSD multiple pairwise comparison test result.



Oneway

ANOVA

Miles per Gallon

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7984.957	2	3992.479	97.969	.000
Within Groups	16056.415	394	40.752		
Total	24041.372	396			

Contrast Coefficients

	Country of Origin		
Contrast	American	European	Japanese
1	.5	.5	-1

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Miles per Gallon	Assume equal variances	1	-6.44	.838	-7.685	394	.000
	Does not assume equal variances	1	-6.44	.820	-7.857	140.636	.000

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Miles per Gallon
LSD

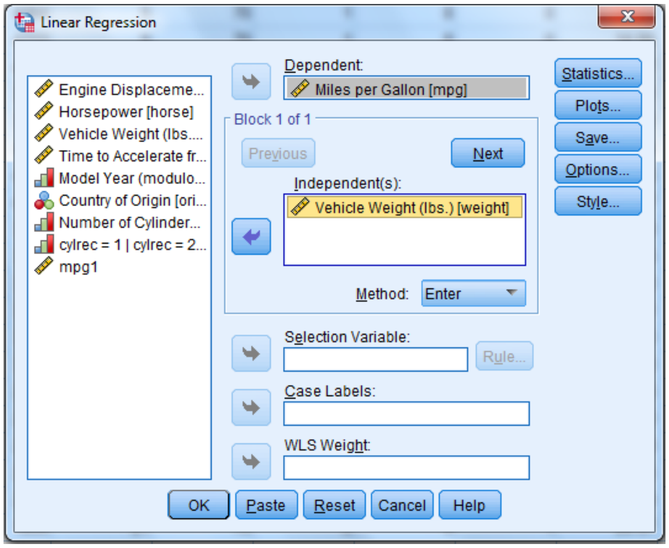
(I) Country of Origin	(J) Country of Origin	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
American	European	-7.763*	.864	.000	-9.46	-6.06
	Japanese	-10.322*	.825	.000	-11.94	-8.70
European	American	7.763*	.864	.000	6.06	9.46
	Japanese	-2.559*	1.048	.015	-4.62	-.50
Japanese	American	10.322*	.825	.000	8.70	11.94
	European	2.559*	1.048	.015	.50	4.62

*. The mean difference is significant at the 0.05 level.

Figure 34: One-Way ANOVA Output with Contrasts and Post-Hoc LSD Test

Regression

Linear regression can be found under the ‘Linear’ in the ‘Regression’ submenu under the ‘Analyze’ menu. Fill in the ‘Dependent’ and ‘Independent(s)’ fields with the appropriate variables. Underneath the ‘Independent(s)’ field a box labeled ‘Method’ says ‘enter.’ That can be changed to stepwise, forward, or backward selection for model selection purposes. To keep the full model, use the default option - ‘enter’. ‘Plots’ allows diagnostic plots to be created. ‘Statistics’ can be used to get more detailed information on the model. ‘Save’ allows you to select data to save back to the data set, including predicted values, various types of residuals, and influence statistics. ‘Options’ provides choices for model selection and the handling of missing values.



Dependent: Miles per Gallon [mpg]

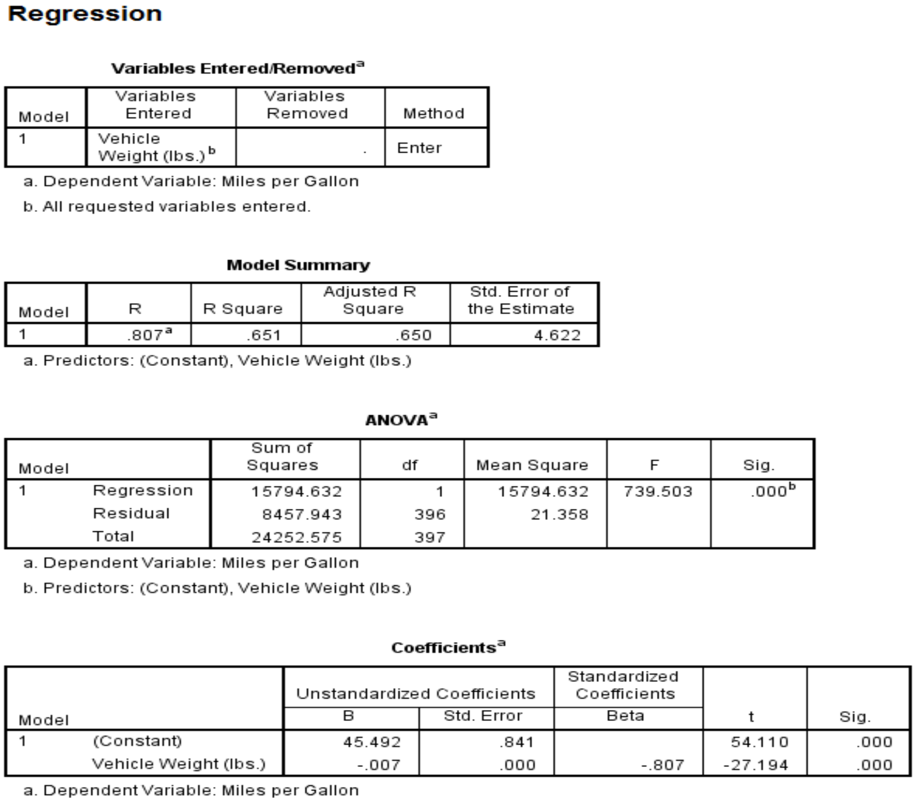
Independent(s): Vehicle Weight (lbs.) [weight]

Method: Enter

Figure 35: Regression

The results of the regression are shown in Figure 36. In the last table, titled “Coefficients”, we can see that the coefficient for vehicle weight is -0.007, indicating that there is negative relationship between vehicle weight and Miles per gallon. Since the p value for the ANOVA model (3rd table shown in Figure 36) is 0.000, there is evidence of a significant relationship between vehicle weight and miles per gallon.

Also, the from the *Coefficients* table, you can get the prediction line for the regression model. In this example it is $\widehat{mpg} = 45.492 + -0.007 * weight$



Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Vehicle Weight (lbs.)[b]	.	Enter

a. Dependent Variable: Miles per Gallon

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.807[a]	.651	.650	4.622

a. Predictors: (Constant), Vehicle Weight (lbs.)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15794.632	1	15794.632	739.503	.000[b]
	Residual	8457.943	396	21.358		
	Total	24252.575	397			

a. Dependent Variable: Miles per Gallon

b. Predictors: (Constant), Vehicle Weight (lbs.)

Coefficients[a]

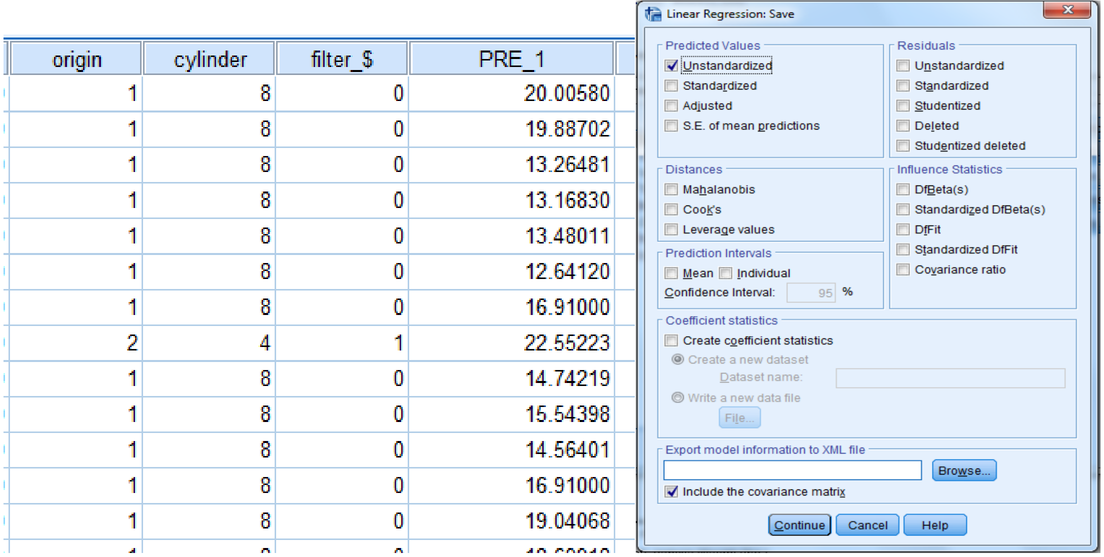
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	45.492	.841		54.110	.000
	Vehicle Weight (lbs.)	-.007	.000	-.807	-27.194	.000

a. Dependent Variable: Miles per Gallon

Figure 36: Regression output

Predictions

By clicking *Save* and checking *Unstandardized* under *Predicted Values*, one can get the predicted values for each of the values of the independent variables. If you need the prediction value for an x-value not already in your data set, just type in the x-value in the independent variable column of your data sheet. Don't type in anything for the y-value. Re-run the regression, asking for the prediction value.



origin	cylinder	filter_$	PRE_1
1	8	0	20.00580
1	8	0	19.88702
1	8	0	13.26481
1	8	0	13.16830
1	8	0	13.48011
1	8	0	12.64120
1	8	0	16.91000
2	4	1	22.55223
1	8	0	14.74219
1	8	0	15.54398
1	8	0	14.56401
1	8	0	16.91000
1	8	0	19.04068

Linear Regression: Save

Predicted Values
- ☑ Unstandardized
- □ Standardized
- □ Adjusted
- □ S.E. of mean predictions

Residuals
- □ Unstandardized
- □ Standardized
- □ Studentized
- □ Deleted
- □ Studentized deleted

Distances
- □ Mahalanobis
- □ Cook's
- □ Leverage values

Influence Statistics
- □ DfBeta(s)
- □ Standardized DfBeta(s)
- □ DfFit
- □ Standardized DfFit
- □ Covariance ratio

Prediction Intervals
- □ Mean □ Individual
- Confidence Interval: 95 %

Coefficient statistics
- □ Create coefficient statistics
- ● Create a new dataset
- Dataset name:
- ○ Write a new data file

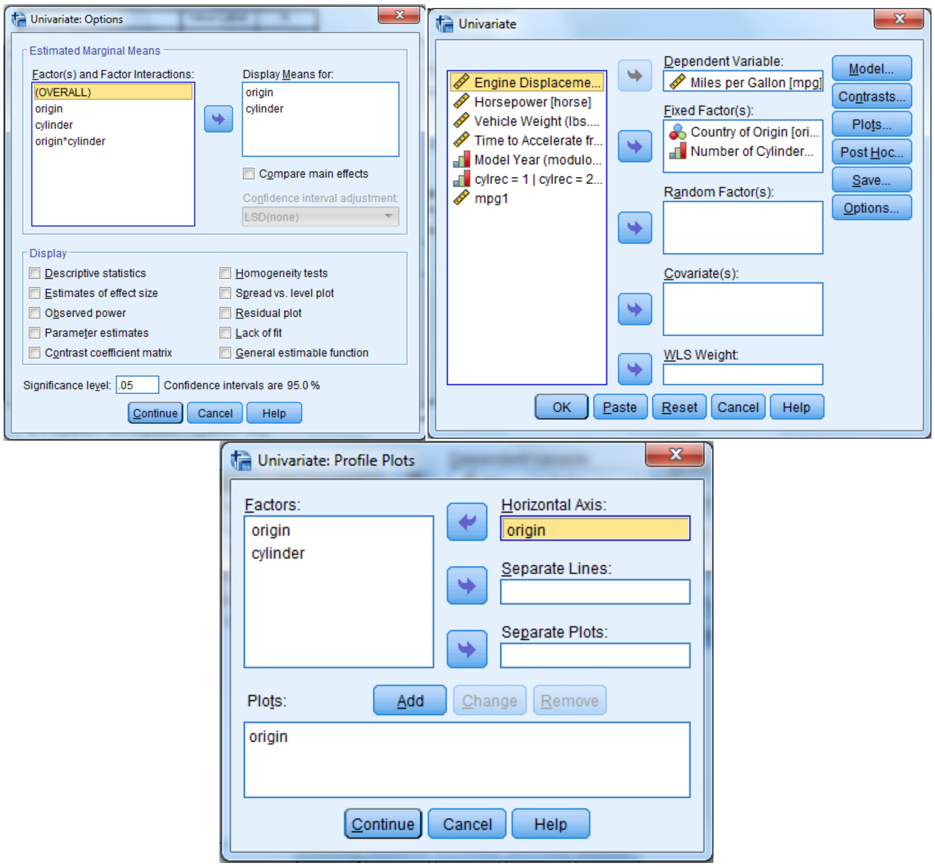
Export model information to XML file

- ☑ Include the covariance matrix

Figure 37: Prediction for Regression

General Linear Model

Only the Univariate part is introduced here. The GLM Univariate procedure (Figure 38) provides regression analysis and analysis of variance for one dependent variable by one or more factors and/or variables. The factor variables divide the population into groups. Using this General Linear Model procedure, you can test null hypotheses about the effects of other variables on the means of various groupings of a single dependent variable. You can investigate interactions between factors as well as the effects of individual factors, some of which may be random. In addition, the effects of covariates and covariate interactions with factors can be included. For regression analysis, the independent (predictor) variables are specified as covariates. The CARS example fits a regression model of *mpg* against the covariate *horse*, the factors *origin* and *cylinder*.



Univariate: Options

Estimated Marginal Means

Factor(s) and Factor Interactions:
(OVERALL)
origin
cylinder
origin*cylinder

Display Means for:
origin
cylinder

☐ Compare main effects

Confidence interval adjustment: LSD(none)

Display

☐ Descriptive statistics	☐ Homogeneity tests
☐ Estimates of effect size	☐ Spread vs. level plot
☐ Observed power	☐ Residual plot
☐ Parameter estimates	☐ Lack of fit
☐ Contrast coefficient matrix	☐ General estimable function

Significance level: .05 Confidence intervals are 95.0 %

Univariate

Engine Displaceme...
Horsepower [horse]
Vehicle Weight (lbs....
Time to Accelerate fr...
Model Year (modulo...
cylrec = 1 | cylrec = 2...
mpg1

Dependent Variable: Miles per Gallon [mpg]

Fixed Factor(s):
Country of Origin [ori...
Number of Cylinder...

Random Factor(s):

Covariate(s):

WLS Weight:

Univariate: Profile Plots

Factors:
origin
cylinder

Horizontal Axis: origin

Separate Lines:

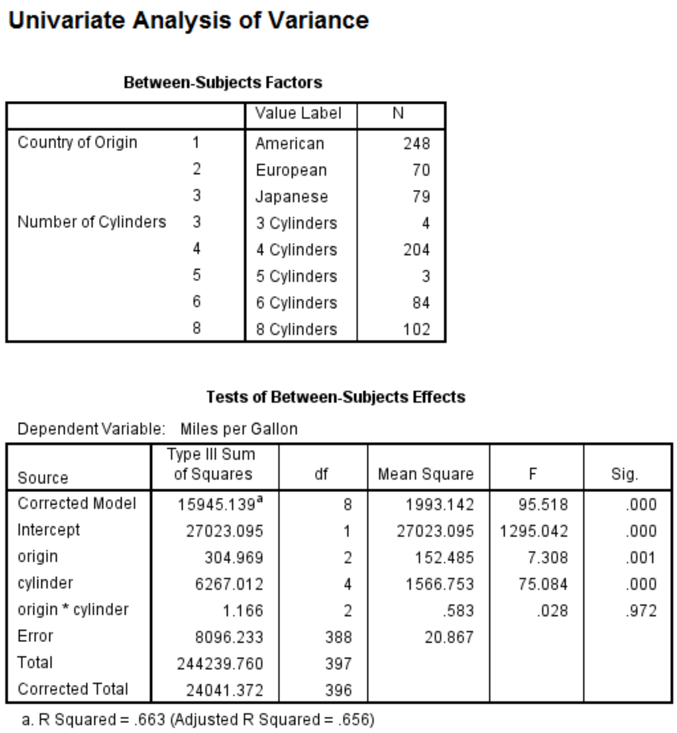
Separate Plots:

Plots:
origin

Figure 38: Univariate General Linear Models Procedure

The *Model* button allows you to select the effects that you wanted to include in the model. The default is full factorial, which includes all the main effects and interactions. Contrasts allow one to display tests on specified contrasts, which are used to compare marginal means between multiple groups. The *Plots* button can provide profile plots (interaction plots) which are useful for comparing marginal means in your model. The *Save* button allows you to save the values predicted by the model, the residuals, and the related measures as new variables in the *Data Editor*. The *Options* button provides options to display marginal means and their confidence intervals, descriptive statistics, residual plot, parameter estimates, observed power, etc.

Figure 38 on the next page shows the results of the Generalized linear model.



Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
Country of Origin	1	American	248
	2	European	70
	3	Japanese	79
Number of Cylinders	3	3 Cylinders	4
	4	4 Cylinders	204
	5	5 Cylinders	3
	6	6 Cylinders	84
	8	8 Cylinders	102

Tests of Between-Subjects Effects

Dependent Variable: Miles per Gallon

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	15945.139[a]	8	1993.142	95.518	.000
Intercept	27023.095	1	27023.095	1295.042	.000
origin	304.969	2	152.485	7.308	.001
cylinder	6267.012	4	1566.753	75.084	.000
origin * cylinder	1.166	2	.583	.028	.972
Error	8096.233	388	20.867		
Total	244239.760	397			
Corrected Total	24041.372	396			

a. R Squared = .663 (Adjusted R Squared = .656)

Figure 39: Univariate General Linear Models Result

Generalized linear model gives the result of between subject effects and within-subject (in this example, this is no within subject factors) result and the corresponding p-values. It also gives the result of estimated margin of means of each factor and the interaction plots for them depending on your set up in the Options and Plots in the model set up.

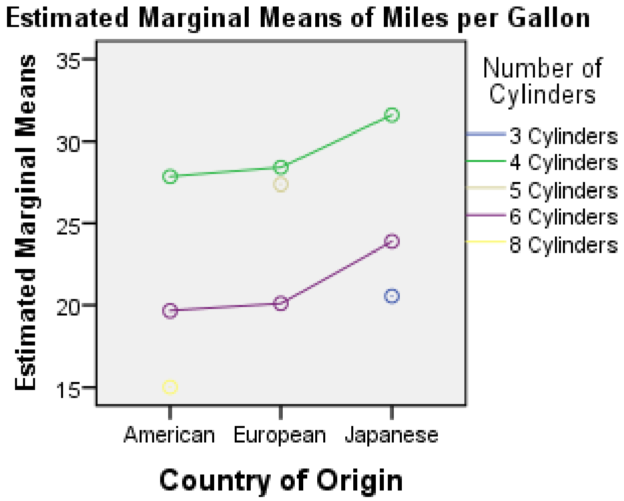


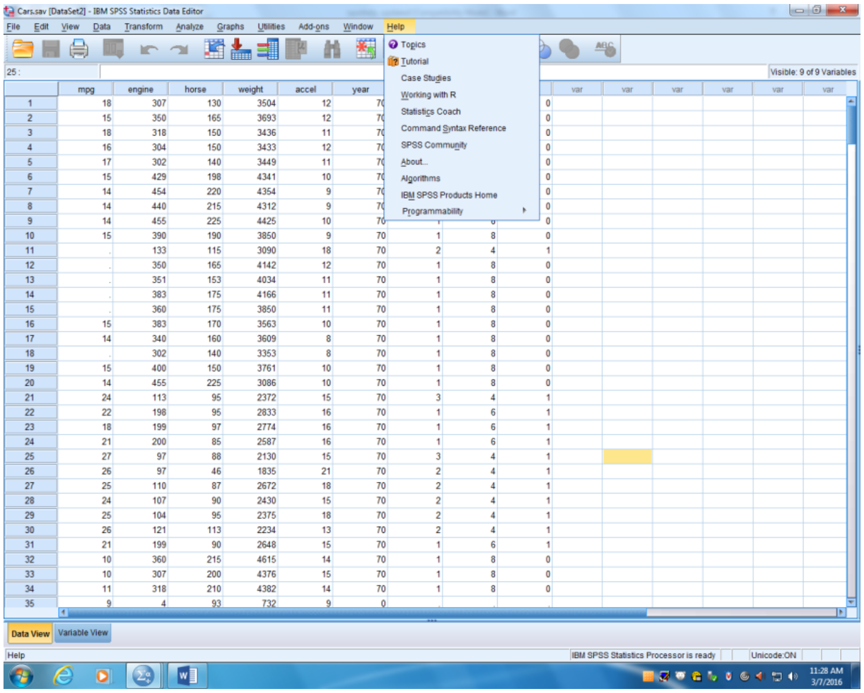
Figure 40: Univariate General Linear Models Plot

From the interaction plot (Figure 40) we can see that since the lines are parallel to each other, there is no significant interaction between factor origin and cylinder in this model, which agrees with the large p-value of 0.972 for interaction effect in between subject effects table.

Help in SPSS

To learn more about SPSS, you can look at the on-line tutorial that comes with the software: click *Help, Tutorial*. The preceding sections have provided an overview of commonly used procedures to get you started in using SPSS. SPSS has very good help documents – often containing examples and tutorials to make the point clear. The help docs can be referenced two different ways. First, almost every window in SPSS contains a *Help* button. Clicking it takes you immediately to the help documentation specific to that window. Second, the last menu in SPSS is a help menu. The first option, *Topics* brings up the help documentation. Use the index or a search command to look up the options that will give you the right results. Sometimes the help docs have the words *Show me* highlighted in blue. Clicking on *Show me* will walk you through an example of the procedure.

The *Statistics Coach* option is useful when you have an idea of what you're looking for, but do not know the name. Select an option on the right, and examples will appear on the left. If you find what you are looking for, hit the next button. You'll be guided through one or more screens asking you to describe the data. When the program has the information it needs, it will open a window telling you what the procedure's called and where to find it in the future. It will also go ahead and open up the menu for you. The Help facility in SPSS can be used on other issues such as Working with R, Algorithms, Programmability, etc.



	mpg	engine	horse	weight	accel	year				var	var	var	var	var	var
1	18	307	130	3504	12	70			0						
2	15	350	165	3693	12	70			0						
3	18	318	150	3436	11	70			0						
4	16	304	150	3433	12	70			0						
5	17	302	140	3449	11	70			0						
6	15	429	198	4341	10	70			0						
7	14	454	220	4354	9	70			0						
8	14	440	215	4312	9	70			0						
9	14	455	225	4425	10	70	1	8	0						
10	15	390	190	3850	9	70	1	8	0						
11	.	133	115	3090	18	70	2	4	1						
12	.	350	165	4142	12	70	1	8	0						
13	.	351	153	4034	11	70	1	8	0						
14	.	383	175	4166	11	70	1	8	0						
15	.	360	175	3850	11	70	1	8	0						
16	15	383	170	3563	10	70	1	8	0						
17	14	340	160	3609	8	70	1	8	0						
18	.	302	140	3353	8	70	1	8	0						
19	15	400	150	3761	10	70	1	8	0						
20	14	455	225	3086	10	70	1	8	0						
21	24	113	95	2372	15	70	3	4	1						
22	22	198	95	2833	16	70	1	6	1						
23	18	199	97	2774	16	70	1	6	1						
24	21	200	85	2587	16	70	1	6	1						
25	27	97	88	2130	15	70	3	4	1						
26	26	97	46	1835	21	70	2	4	1						
27	25	110	87	2672	18	70	2	4	1						
28	24	107	90	2430	15	70	2	4	1						
29	25	104	95	2375	18	70	2	4	1						
30	26	121	113	2234	13	70	2	4	1						
31	21	199	90	2648	15	70	1	6	1						
32	10	360	215	4615	14	70	1	8	0						
33	10	307	200	4376	15	70	1	8	0						
34	11	318	210	4382	14	70	1	8	0						
35	9	4	93	732	9	0	.	.	.						

Figure 41: SPSS Help options