SPSS Tutorials

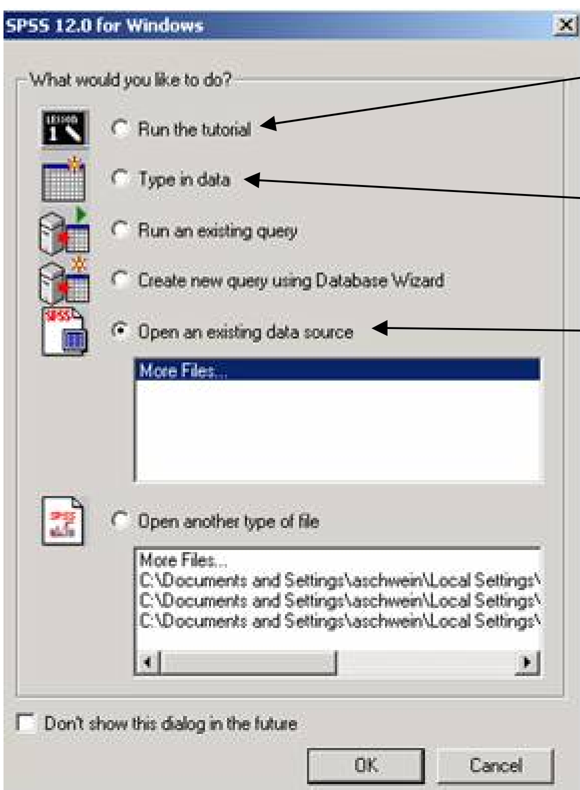
Getting Started and Entering Data

In this tutorial you will learn:

1. How to start an SPSS session
2. How to type in data
3. How to define variables and identify variable names
4. How to save a data file
5. How to open an existing data file
6. Using the SPSS toolbar

Starting SPSS

When you first open SPSS, you will be presented with the opening window. This window allows you to select from several options concerning how you would like to begin your session. If you do not want to start from this window in the future, select the box next to "Don't show this dialog in the future." The most likely options you will select are to type in data and to open an existing data source.

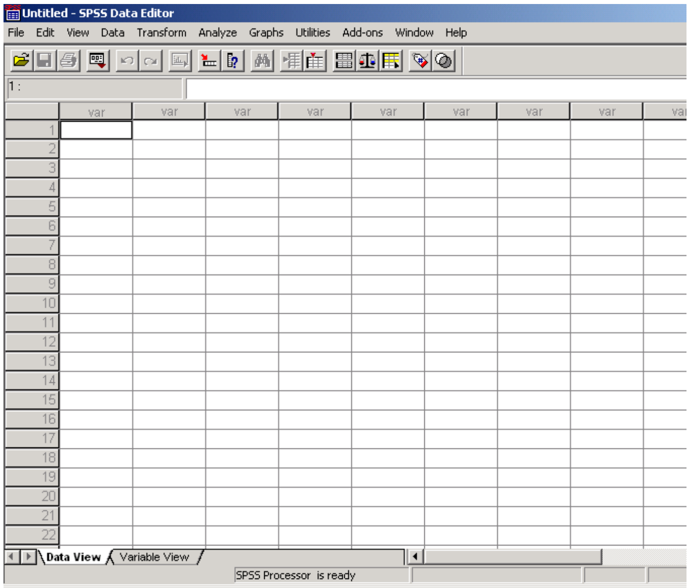


Walk through the tutorial designed by SPSS. It includes starting SPSS, entering and manipulating data and running basic statistics.

Open the data view consisting of a blank spreadsheet into which you can enter raw data by hand.

Open data that has already been saved. This data is likely to be saved from a prior use in which you typed in your own data. Later tutorials will discuss how to open data files other than SPSS data files (i.e. Excel files or database files). Check this button, and then select the file to open in the window below. If the preferred file is not shown, select more files and identify the data file to open.

Type in Data



If you opt to type in data from the opening window, a blank spreadsheet will appear. You many type in one data point per cell. Each column represents one variable (e.g., subject identification, gender, test score, etc.). It is a good idea to keep one subject per row. So, row 1 represents the first subject, row 2 represents the second subject and so on.

Data may include numbers or letter strings.

To begin, let's assume that we administered a test to 6 individuals. We recorded their gender, age and test score. Each individual also received an identification number (so we don't use any other identifying information per the human subjects requirements). The data are as follows:

Subject	Gender	Age	Test Score
1	m	18	95
2	f	21	80
3	m	20	75
4	f	19	79
5	f	18	88
6	m	22	62

To enter this data, select the top left cell and enter a "1." You may use the tab key to move across the row to a new variable for that same subject and enter an "m." You could also use the arrow keys to change cells.

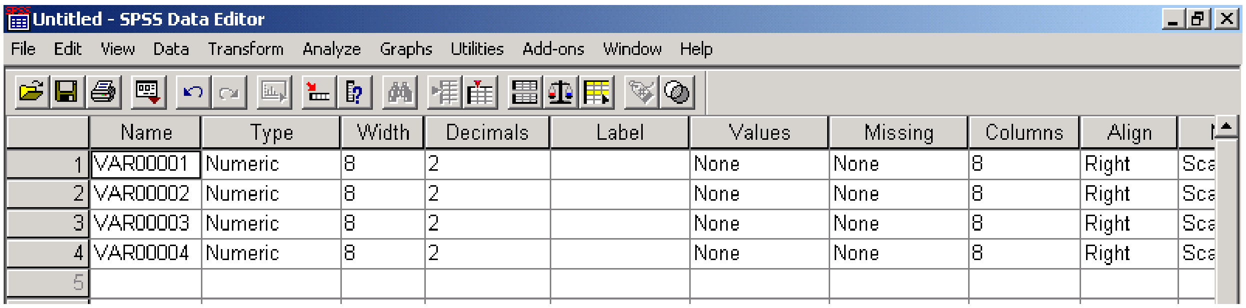
Depending on the settings for the version of SPSS you are using, you might notice that the letters, "m" and "f", do not appear. Instead, SPSS places periods. This is because it is expecting numerical values to correct this, you may either recode gender to numerical values, such as a "1" for males and a "2" for females, or you could edit the variable characteristics in Variable View (see the next section).

You may also notice that SPSS automatically names the variables (e.g., VAR00001). You can change the variable names in Variable View (see next section).

Defining Variables and Variable Names

In the previous section, we noted that the character strings did not appear in the spreadsheet and the variable names did not actually describe the data in each column. To define the variables, we will use Variable View. At the bottom of the page, notice two tabs. One says “Data View,” the other says “Variable View.” We have already used Data View to type in our data. Variable View will allow us to further define the nature of our variables.

Click the Variable View tab.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	N
1	VAR00001	Numeric	8	2		None	None	8	Right	Sca
2	VAR00002	Numeric	8	2		None	None	8	Right	Sca
3	VAR00003	Numeric	8	2		None	None	8	Right	Sca
4	VAR00004	Numeric	8	2		None	None	8	Right	Sca
5										

Each row represents a different variable. We had four variables (subject ID, gender, age, and test score) that SPSS named for us as VAR00001 through VAR00004.

Name. Type in the variable name you would like to use to describe the data in that column. I will use the names: Subject, Sex, Age, and Test. Variable names must be no more than 64 characters long and must begin with a letter. They may not end with a period or contain spaces, but other characters are allowed ($, #, _, @). Special characters may not be used, including *, !, ‘, and ?. Reserved words are not allowed. These include ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH. Reserved words correspond with SPSS functions (e.g., EQual, Greater Than).

Type. Select the type of variable: numeric, comma, dot, scientific notation, date, dollar, custom currency, or string. String represents letter strings. To get the gender data to appear as “m” and “f” in our data set, we would need to select string as the variable type. One drawback is that several statistical procedures require that all data be numerical. So, if we want to run analyses comparing males to females and use gender as the blocking variable then we would need to recode the data to numeric (e.g., 1 for males and 2 for females). (I opted to recode the data.)

Width. Select how many spaces could be maximally occupied by a data point in that column. SPSS selects 8 by default. However, if you have long values such as last names or social security numbers, you would need to raise this value.

Decimals. By default, SPSS inserts 2 decimals for each numerical value. If you prefer to change that, select the number of decimal places here, from 0 to 99. I changed the decimal places for Subject and Age to 0.

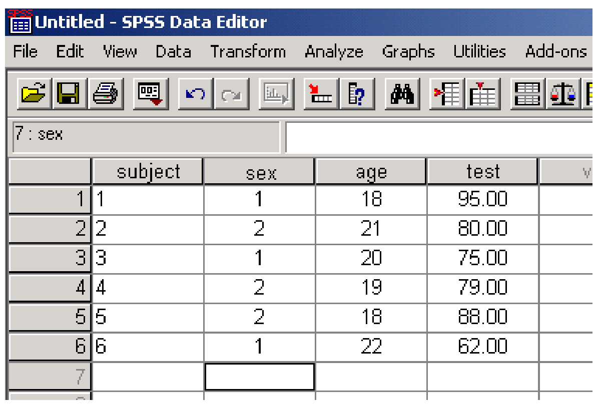
Label: Type in a longer definition for your variable. The variable name is a brief, one-word descriptor. The variable label is often a sentence or phrase. For example, we could further define the variable test as “Percent of correct answer out of 1000 total questions.”

Values: If you recoded variables in the data set (as we did with Sex), you can define what the data values mean here. For example, in Sex, click in the corresponding cell under Values. Then select the gray box on the right-hand side of the cell. This will open a new window. In the Value box, type in the value from the data view. In the Value Label box, type in the label you want to correspond with that value. In this case, we could type “1” in the Value box and “male” in the Value Label box, then select “Add.” Then we could type “2” and “female” followed by “Add.” When we are finished defining the values, then select OK.

Align: In the data view, where do you want the data points aligned (Left, Center, or Right).

Measure: Define the measurement scale for each variable (Nominal, Ordinal, Interval, or Ratio). This will have bearing for custom tables.

Once you make the appropriate adjustments, move back to Data View by clicking on the Data View tab on the bottom left of the screen. Notice that the variable names have changed, and the alignment and decimals have changed to meet your specifications.



Saving data

Select File, then Save.
Select the folder in which to save it using the “Save In” drop-down menu.
Name your data file in the box labeled “File Name.”
Identify the type of file you want to save:

SPSS Can only be opened and used by SPSS, but all your formatting is maintained

Excel Saved as an Excel spreadsheet. Some formatting maintained. Can be imported into SPSS.

Tab delimited A text file in which the data is separated by tabs. This is convenient if it is to be opened and used by applications that cannot read Excel or SPSS files.

Opening an Existing Data File

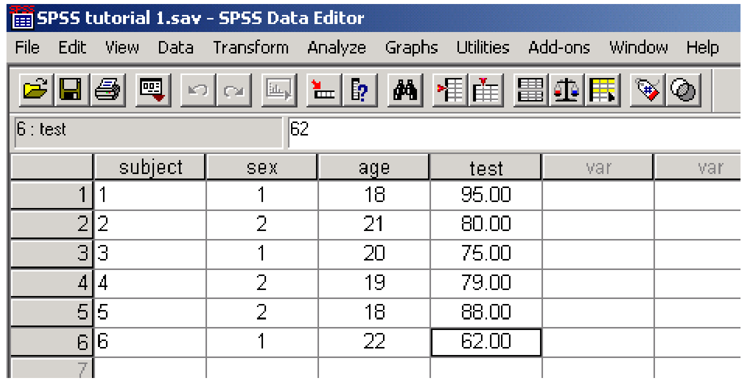
Now that you have saved your data, if you want to use it again at a later date, you may do so. To open the data file either:

1. Select “Open an existing data source” from the opening window and select your data file from the window or by using the Browse button.
2. From the data window, in data view, select File, then Open. Search the folders for your file and click Open.

SPSS Toolbar

The toolbar contains the following menus:

File	Edit
View	Data
Transform	Analyze
Graphs	Utilities
Add-Ons	Window
Help	



You have used some of these menus in previous tutorials. Some of the procedures within these menus are reflected as short-cuts on the second toolbar.

<u>File</u>

New Open a new data window, a new output window, etc.
Open Open existing data, output, etc.
Save Save the data set in the window.
Save As Save the data set, but with allowances to save it as other than an SPSS data set.
Print Preview View a preview of what your page would look like if printed.
Print Print the open page.

<u>Edit</u>

Undo Undo the last action performed.
Redo Redo the last action performed.
Cut Cut the selected cells and save them to the clipboard in case you want to paste them.
Copy Copy the selected cells and save them to the clipboard.
Paste Paste the contents of the clipboard to selected cells.

<u>View</u>

Status Bar Located at the bottom of the screen. Identifies what SPSS is currently running and whether or not you are currently using a filter or a weight or if the data has been split into groups.
Fonts Change the fonts that are used.
Value Labels If you used value labels, they will appear in data rather than the recoded values. For example, if you used 1 and 2 to stand for “Male” and “Female”, and labeled those values as such in Variable View, then you could display “Male” and “Female” rather than 1 and 2.
Variables Switch to Variable View.

<u>Data</u> Tools to manage and work with your data. These will be discussed in more detail in future tutorials.

<u>Transform</u> Alter and adjust your data, or create new variables.

<u>Analyze</u> Compute descriptive and inferential statistics.

<u>Graphs</u> Create a variety of graphs and pictorial depictions of the data.

<u>Utilities, Add-Ons, Window</u> Will not be discussed further.

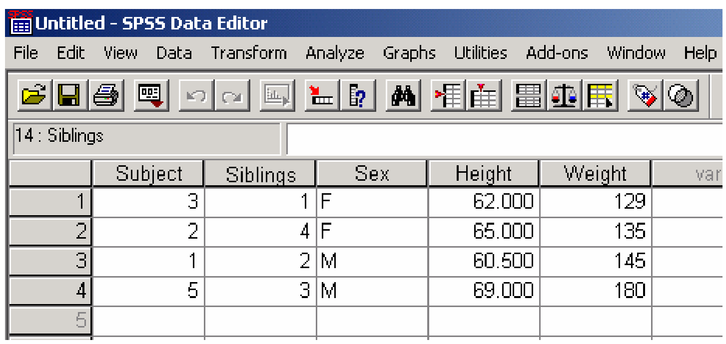
Assignment

1. Type the following data into SPSS. Identify variable names as shown. Define Sex as string.

Subject	Sex	Height	Weight
1	M	60.5	145
2	F	65	135
3	F	62	129
4	M	70.275	198
5	M	69	180

2. Change the decimal places so that Subject and Weight have no decimal places and Height has the appropriate number of spaces.
3. Save your data to a disk and label it “SPSS Assignment 1.”

The completed data set should look like:



	Subject	Siblings	Sex	Height	Weight	var
1	3	1	F	62.000	129	
2	2	4	F	65.000	135	
3	1	2	M	60.500	145	
4	5	3	M	69.000	180	
5						

Editing Data

In this tutorial you will learn:

7. How to insert and delete variables
8. How to edit and sort data
9. How to print data

Inserting and Deleting Variables

Inserting a Variable
Once your data set has been entered, you may decide to further edit it by inserting additional variables or deleting current ones.

1. Select the column to the right of where you want to enter the new variable. Do this by clicking on the variable name.
2. Click on Data, Insert Variable.
3. A new column will appear, named by SPSS, with periods as data points.
4. You may now enter new data in this column to replace the periods.

Alternatively, you could right click on the column (variable name) to the right of where you want to enter the new variable. Select Insert Variable.

Deleting a Variable

1. Select the column you wish to delete.
2. Click Edit, Clear.
3. The column will be removed.

Alternatively, you could right-click the column (variable name) and select Clear.

Inserting and Deleting Cases/Rows
Follow similar steps to insert or delete rows.
To *insert* a row, highlight the row above where you want to insert a new one, select Data, then Insert Case. Alternatively, you could right click on the row above, then select Insert Case.

To *delete* a row, highlight the row to be deleted, select Edit, then Clear, or right click on the row and select Clear.

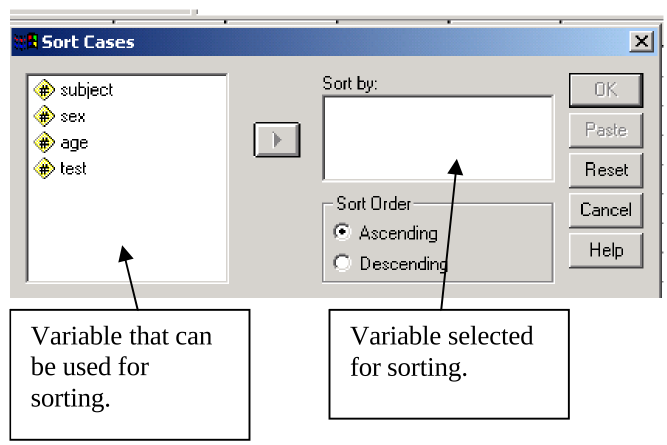
Editing and Sorting Data

Editing
It is common to enter some data incorrectly. If this occurs, simply select the cell with the incorrect data point and type the correct data. This will replace the existing data point.

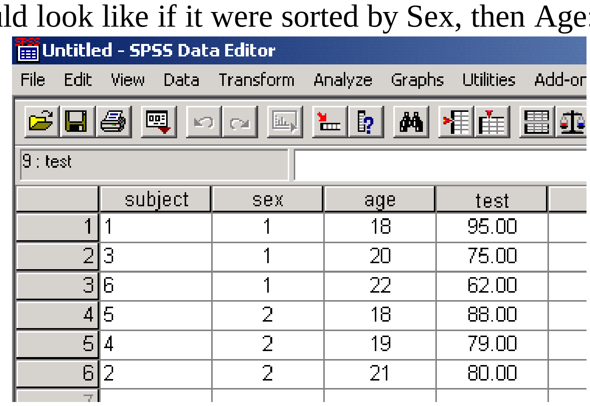
Also, note that at the top of the page is an information bar. You can change parts of an existing cell or add a character without overwriting what is currently in the cell by using the information bar. If you wanted to change mall to male, simply click on the cell with "mall" in it. Then click on the information bar. Move the cursor to the end of the word by clicking at the end or using the arrow keys. Delete the last *l* and type *e*. Then hit Enter once.

Sorting
To sort your data:

1. Data, Sort Cases. This will bring up a new window.
2. Highlight the first variable to sort by in the left-hand box.
3. Click the right arrow to move it to the Sort By box.
4. You may now determine if you want to sort in ascending or descending order of this variable.
5. If you want a second variable to sort by, follow steps 2-4. The data will first be sorted by the variable on the top of the Sort By box. Within that variable, the data will be sorted by the next variable in the list.



This is what the data would look like if it were sorted by Sex, then Age:



	subject	sex	age	test
1	1	1	18	95.00
2	3	1	20	75.00
3	6	1	22	62.00
4	5	2	18	88.00
5	4	2	19	79.00
6	2	2	21	80.00

Printing Data

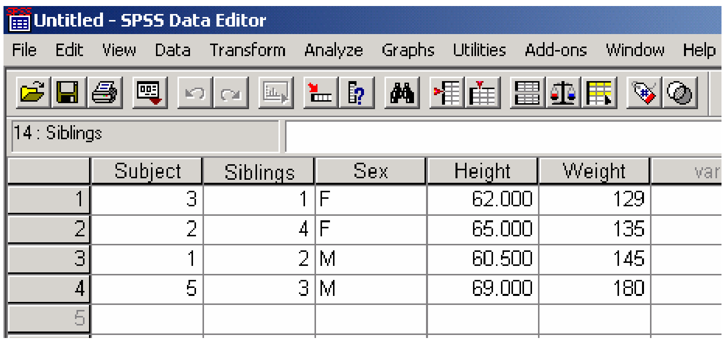
To print all of your data, select File, then Print.

To print some of your data, highlight the data to be printed, select File, then Print, then click Selection.

Assignment

4. Open the data saved as SPSS Assignment 1 from the last assignment.
5. Insert a new variable titled siblings. Enter the number of siblings of each individual as follows: 2, 4, 1, 0, 3.
6. You discovered subject number 4 lied when reporting height and weight. Delete this case.
7. Sort the data first by sex then by number of siblings within sex.
8. Print your data.
9. Save your data to a disk and label it "SPSS Assignment 2."

The completed data set should look like:



	Subject	Siblings	Sex	Height	Weight	var
1	3	1	F	62.000	129	
2	2	4	F	65.000	135	
3	1	2	M	60.500	145	
4	5	3	M	69.000	180	
5						

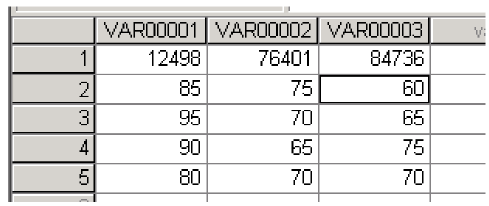
Manipulating Data

In this tutorial you will learn:

1. How to transpose data
2. How to select cases and filter data
3. How to transform and recode data

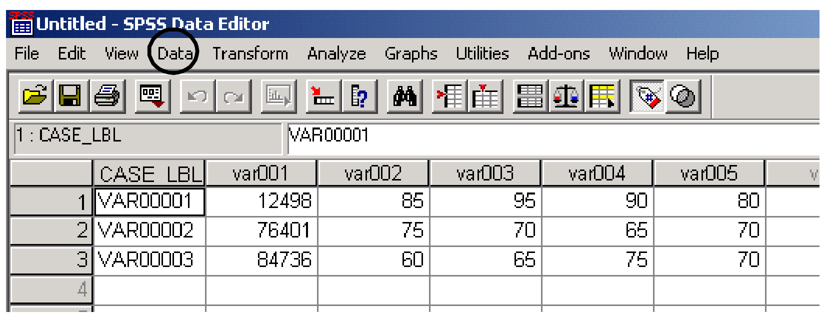
Transpose Data

There are times when data are entered in rows and you want them entered in columns. For example, some people list, in a column, their students' scores on tests. In the following example, the first row lists the students' ID numbers and the following rows reflect their test scores for the year. Unfortunately, you will not be able to calculate average scores and compare tests with the data entered this way. To make it easier to work with, you need to transpose the columns to rows and rows to columns so that the rows reflect students and the columns reflect variables.



	VAR00001	VAR00002	VAR00003
1	12498	76401	84736
2	85	75	60
3	95	70	65
4	90	65	75
5	80	70	70

To do this, go to Data → Transpose. A new window will appear asking which variables you want to transpose. Move the variables from the box on the left to the box on the right using the arrow button. I selected all three variables (columns). SPSS will create a new data set including the transposed values of the variables you selected. All other variables will be lost!



	CASE_LBL	var001	var002	var003	var004	var005
1	VAR00001	12498	85	95	90	80
2	VAR00002	76401	75	70	65	70
3	VAR00003	84736	60	65	75	70
4						

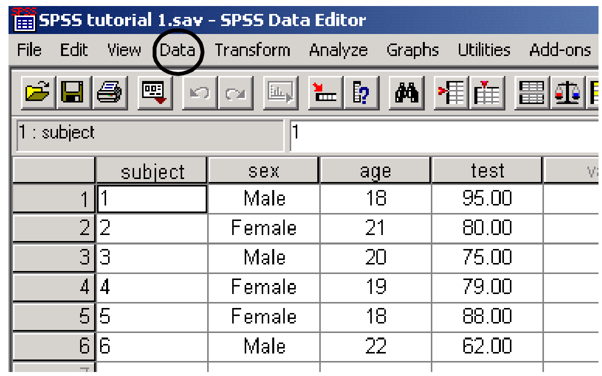
The new data set looks like this. Notice that the first column consists of the variable names from the prior data set. You can now change the variable names and formatting using Variable View.

Select Cases and Filter Data

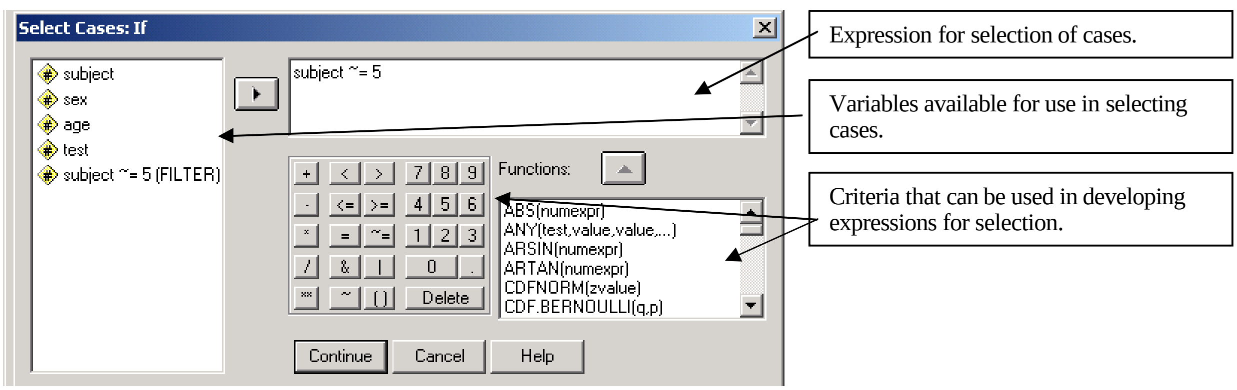
There are times in which you do not want to analyze all the data in your data set. You may want to filter out certain subgroups (e.g., by age, gender, score, or other value) or certain rows (e.g., if you found out one individual lied on the test).

With the following data from Tutorial 1, we will demonstrate how to do this.

Filter out certain rows.
Let's say we want to filter out subject number 5 because she was drunk when she took the test.



Go to Data → Select Cases.
Select If Condition is Satisfied.
This brings up a new window allowing you to identify the conditions under which cases will be selected. Here we want to select all cases that are not equal to subject number 5. Notice how useful it is to have an identification variable.



In this case, we want to select all subjects that are not equal to 5, so we build an expression stating "subject~=5." To do this, highlight subject in the right-hand box. Move it to the expression box using the right-arrow button. Then, either type or click the corresponding button to build your selection expression.
Some of the terms are:

~	Not
**	Exponent
&	And
|	Logical Or. (True if the expression before or after the | is true.)

Filter out subgroups
To filter out subgroups of data, follow the same procedures, except identify the subgroup(s) to maintain in the selection expression. For example, if we only want to analyze data belonging to females, we would enter the following expression: "Sex=2".

Whenever we filter out variables, SPSS creates a new variable (filter_$) that identifies whether or not a case is selected for use or not. Those cases that were filtered out (not selected) are also identified by a slash across their row number.

	subject	sex	age	test	filter_$
1	1	Male	18	95.00	Not Select
2	2	Female	21	80.00	Selected
3	3	Male	20	75.00	Not Select
4	4	Female	19	79.00	Selected
5	5	Female	18	88.00	Selected
6	6	Male	22	62.00	Not Select

Other Options

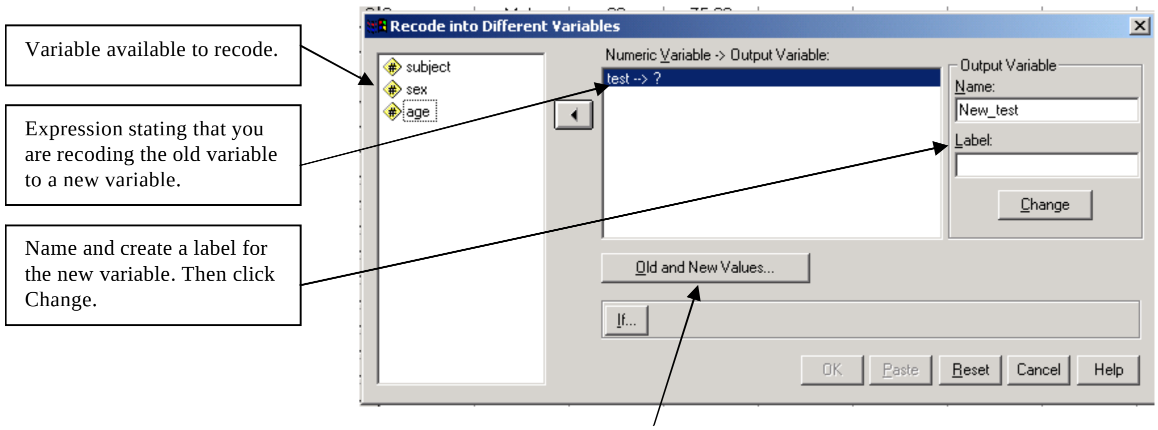
SPSS will also filter out random cases (either a given percentage or a given number of cases). SPSS will also use a filter variable that you have selected.

Transform and Recode Data

SPSS allows one to recode or transform existing data into different forms.

Recode Data

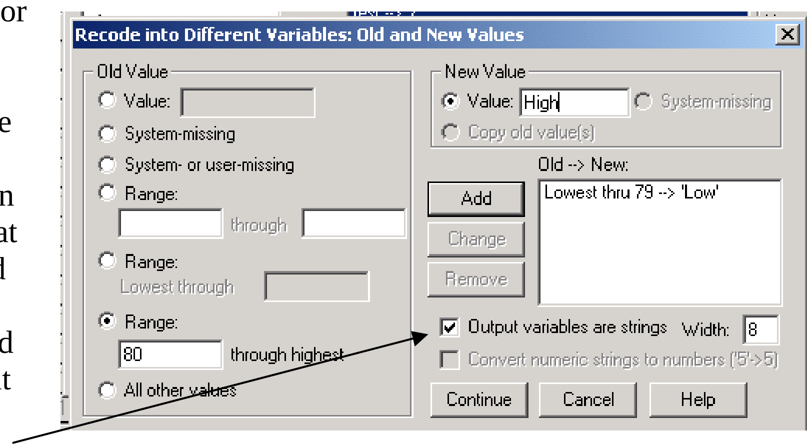
Recoding data assigns new values to existing data, or collapses subsets of data into new values. For example, we might want to group our six students by whether they scored high or low on the test. To do this, select Transform → Recode. At this point you have the option of recoding to the same or different variables. To recode to the same variable would replace the existing data with the new codes. To recode to a different variable would create a new variable with the new codes. The second is almost always my preference because it allows you to retain your original data.



We selected **test** to recode into a new variable called New_test. To identify the new values, click Old and New Values.

You must identified the old value or range of values and the new, recoded value. You can either recode individual values or a range of values.

In this case, all test scores less than 80 were considered low and all that were 80 or higher were considered high. For the low range, select Range, Lowest Through 79 for Old Value. For new value, if the output is a string variable as it is here, check, "Output variables are strings," and enter the new value to match the old range (Low). Then click Add.



For the high range, the parameters are listed in the window above. Then click Add. Click Continue when done.

SPSS creates a new variable with the recoded values:

	subject	sex	age	test	New_test
1	1	Male	18	95.00	High
2	2	Female	21	80.00	High
3	3	Male	20	75.00	Low
4	4	Female	19	79.00	Low
5	5	Female	18	88.00	High
6	6	Male	22	62.00	Low

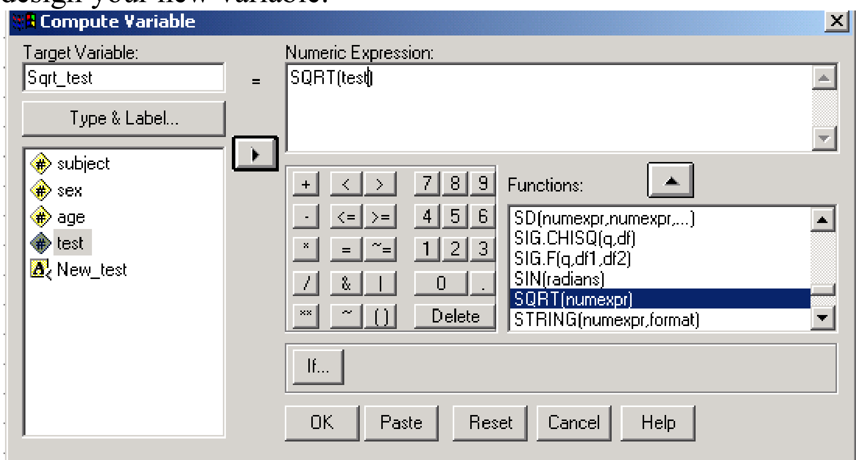
Transform Data

You can compute new variables as a function of old variables. For example, a new variable could be the sum of two old variables. The new variable could be the square root of an old variable. There are almost infinite possibilities for transforming data.

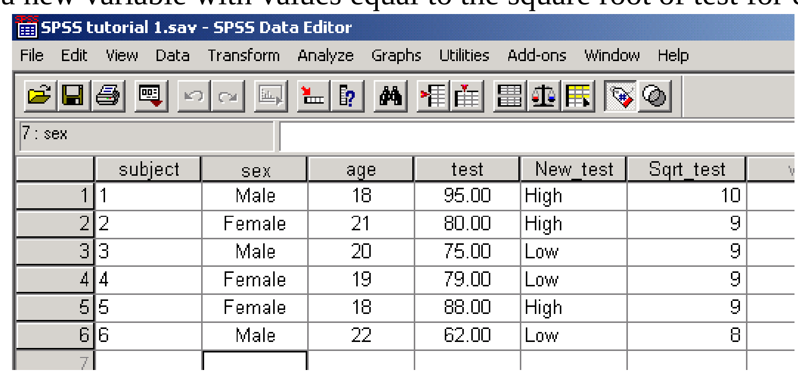
To transform data, select Transform → Compute.

Use the expression builder to design your new variable:

In Target Variable, type the name of the new variable that will contain the transformed values. The Numeric Expression box identifies how to compute the transformed values. Use this box similarly to that used in Selecting Cases. Here we have created a new variable that is equal to the square root of **test**.



SPSS will create a new variable with values equal to the square root of test for each case.



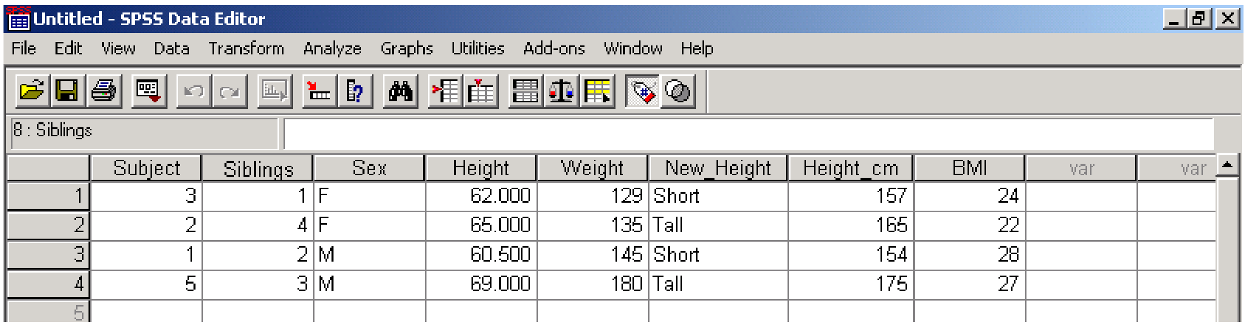
Assignment

1. Open the data saved as SPSS Assignment 2 from the last assignment.
2. Recode the data into tall and short individuals, where tall is 65 inches or taller. (Recode into a different variable.)
3. Create a new variable that transforms height in inches to height in centimeters. Multiply height by 2.54.
4. Create a new variable that is equal to body mass index (BMI).

$$BMI = \frac{\text{Weight in pounds}}{(\text{Height in inches})^2} * 703$$

5. Save the new data set as SPSS Assignment 3.

The final data set should look like:



	Subject	Siblings	Sex	Height	Weight	New_Height	Height_cm	BMI	var	var
1	3	1	F	62.000	129	Short	157	24		
2	2	4	F	65.000	135	Tall	165	22		
3	1	2	M	60.500	145	Short	154	28		
4	5	3	M	69.000	180	Tall	175	27		
5										

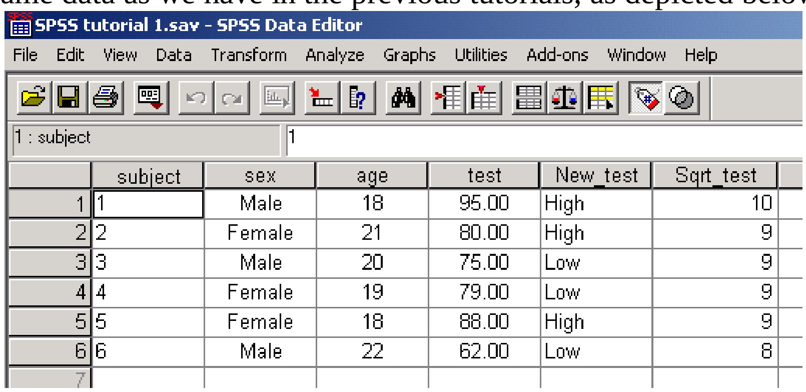
Descriptive Statistics and Frequency Tables

In this tutorial you will learn:

1. How to compute basic descriptive statistics
2. How to split files
3. How to read SPSS Output
4. How to create simple frequency tables
5. How to create frequency tables with two variables

For these procedures, we will use the ANALYZE menu on the toolbar.

We will use the same data as we have in the previous tutorials, as depicted below.



Descriptive Statistics

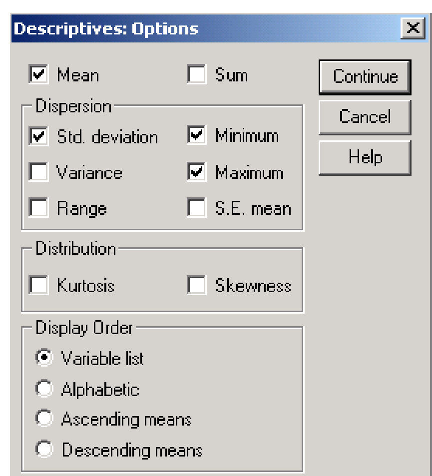
Descriptive statistics convert large sets of data to more meaningful, easier to interpret, chunks or values. They summarize the data. Examples include the mean, median, variance, and range.

SPSS contains a function that will compute many of these statistics, easily.
Simply select Analyze → Descriptive Statistics → Descriptives.
A new window will appear with two boxes. As before, the box on the left contains variables with which descriptive statistics may be calculated (i.e., numeric variables). Move the variables of interest to the right-hand box with the arrow button.

By default, SPSS will provide the following: (a) mean, (b) standard deviation, (c) sample size (d) minimum value, and (e) maximum value.

To select additional options or de-select default options, click the Options button.

Any item with a check-mark will be computed, for each variable selected in the previous step.



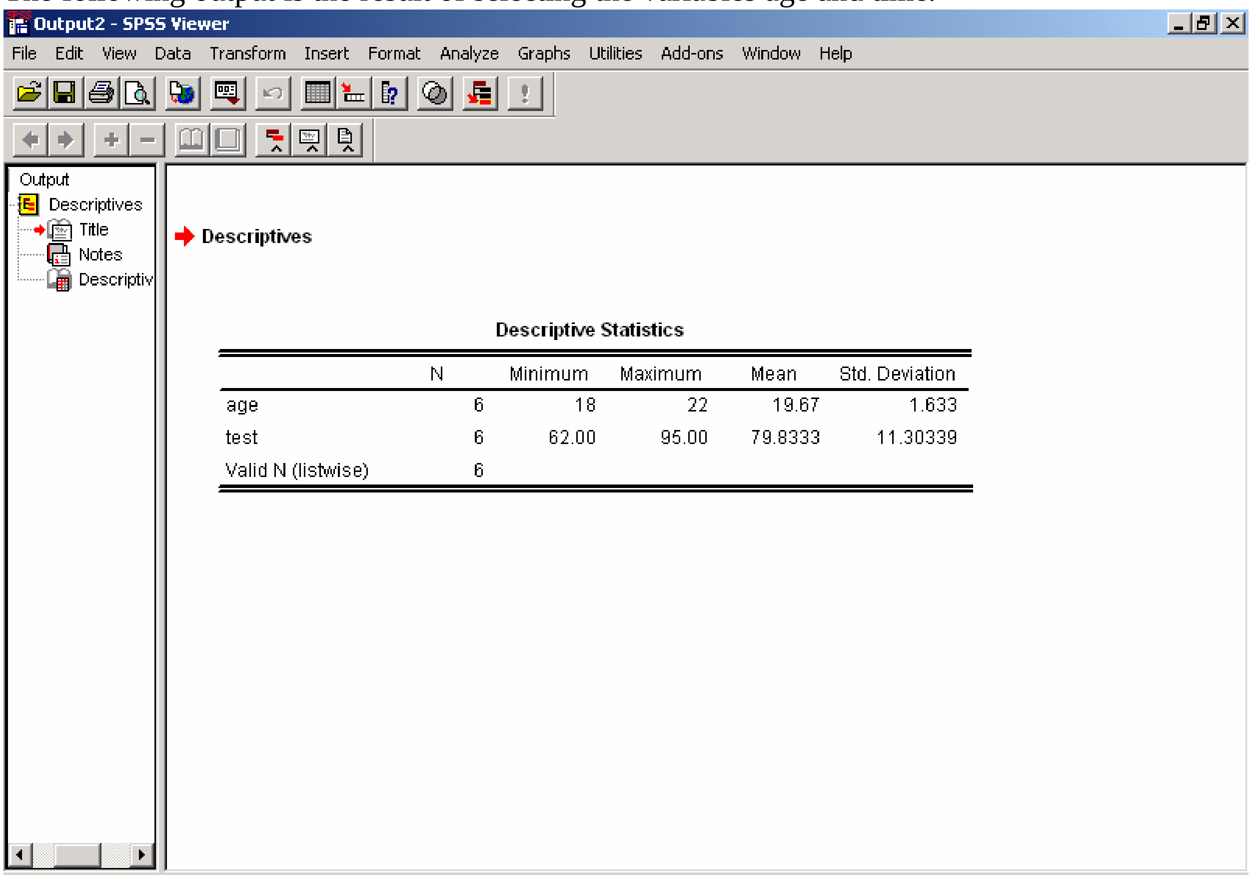
Under display order, you may select the order in which the variables appear in the output.

When you are finished, click Continue to return to the variable-selection window and click OK.

SPSS will compute the statistics and open a new window (Output) with the descriptive statistics.

Reading SPSS Output

The following output is the result of selecting the variables age and time.



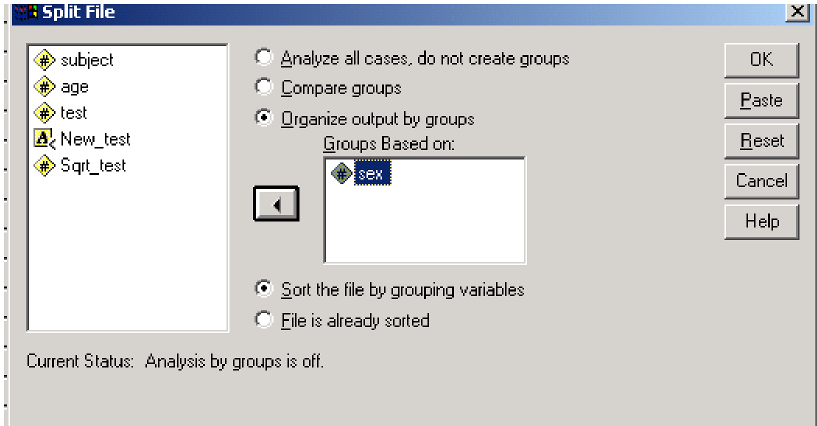
Notice there are two frames: the one on the left lists all the analyses available for viewing, as well as any notes and titles relating to those analyses; the one on the right depicts the results of the analyses.

Typically, SPSS reports results in tabular form. The variables are listed on the left followed by the statistics we selected earlier. We can see that the average age is 19.67 with a standard deviation of 1.633.

SPSS will report variable names or labels or both. To change the default setting, go to Edit → Options → Output labels.

How to split groups

If you want to calculate descriptive statistics on sub-groups (such as males and females separately), you may split your files. To do this:
Data → Split File



Determine how you want to split your data set.

To separate by sex, select Organize output by groups and move the variable sex to the right-hand window. Then select OK.

When you run descriptive statistics on any variable, they will be reported for each level of sex. See below:

Descriptives

sex = 1 Male

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
age	3	18	22	20.00	2.000
Valid N (listwise)	3				

a. sex = Male

sex = 2 Female

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
age	3	18	21	19.33	1.528
Valid N (listwise)	3				

a. sex = Female

You will have to turn off this feature if you want to compute statistics for the whole group, by selecting Analyze all cases, do not select groups.

Frequency Tables

Frequency tables include lists of values (categories) within each selected variables and the number of times each category occurs.

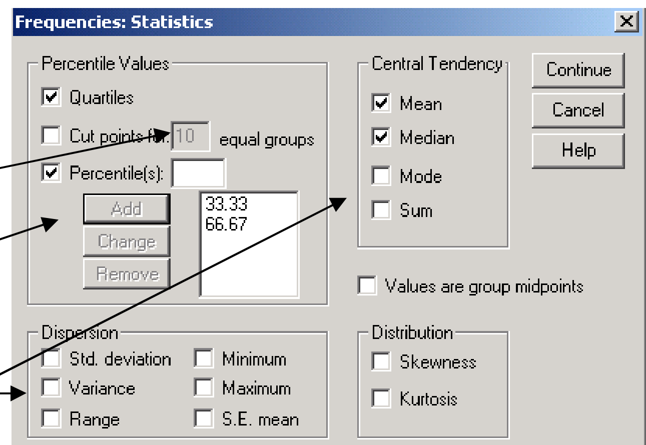
To create a table of frequencies (number of occurrences of given categories), select Analyze → Descriptive Statistics → Frequencies.

Select the variables to be depicted in the frequency table by moving them from the left- to the right-hand box. SPSS provides the user additional options, including statistics, charts, and format.

Statistics

SPSS will, by default, print the values of the selected variables and the frequencies of each. If you prefer additional information, click Statistics:

Options include percentile values. SPSS will print quartiles (fourths) or the values that divide the data into X equal groups (cut points). The number of groups is defined by the user. SPSS will also print selected percentiles. Simply, select Percentile(s), then type in the percentile of interest and click Add. We have selected thirds.



You may also select descriptive statistics, like measures of central tendency and dispersion, as well as statistics describing the distributions.

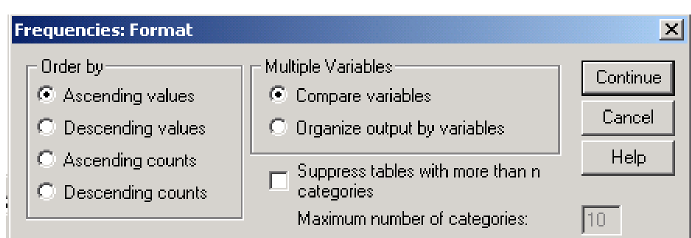
When finished, select Continue.

Charts

This option allows users the opportunity to view bar charts, pie charts, or histograms in addition to the frequency table. This might be useful if there are many categories for each variable or if two or more variables are to be compared. The charts may contain frequencies or percentages.

Format

With this option, users can determine the order in which categories will appear and whether or not multiple variables should be compared. This will impact how results are presented.



To cut back on the amount of output, users may choose not to view tables with many categories (categories).

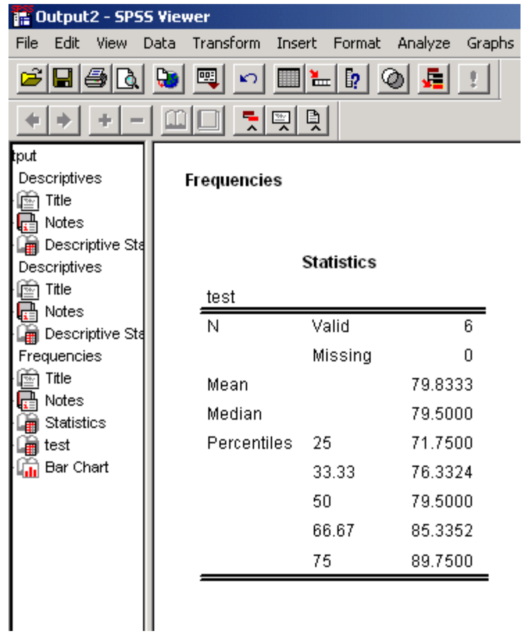
When finished click Continue to return to the variable-selection window. Then click OK.

Output

The new analyses are added to the descriptive statistics. Notice the addition in the left-hand frame.

The following statistics are for the variable test. Notice there are a total of six cases, and none are missing.

The mean test score is 79.833, the median is 79.5. The value of 71.75 cuts off the 25th percentile (25% of cases fall at or below this value), and so on.



If we scroll down the page, we will find additional results:

This table lists all the values of the variable test and the frequency of occurrence of each. Notice there is only one occurrence of each value.

test

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	62.00	1	16.7	16.7	16.7
	75.00	1	16.7	16.7	33.3
	79.00	1	16.7	16.7	50.0
	80.00	1	16.7	16.7	66.7
	88.00	1	16.7	16.7	83.3
	95.00	1	16.7	16.7	100.0
	Total	6	100.0	100.0	

Next, SPSS provides a bar chart depicting these frequency results, as selected under Charts. However, it is not of great interest because the frequency is 1 for each value.

The results look different for the variable of sex…

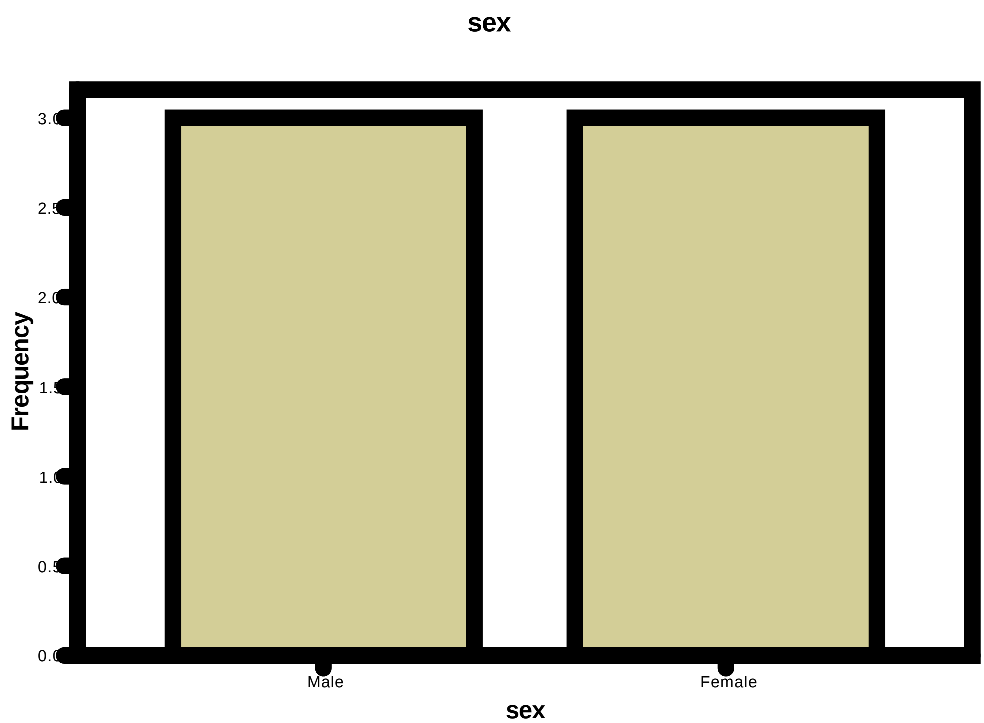
Statistics

sex

N	Valid	6
	Missing	0
Mean		1.50
Median		1.50
Percentiles	25	1.00
	33.33	1.00
	50	1.50
	66.67	2.00
	75	2.00

sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	3	50.0	50.0	50.0
	Female	3	50.0	50.0	100.0
	Total	6	100.0	100.0	

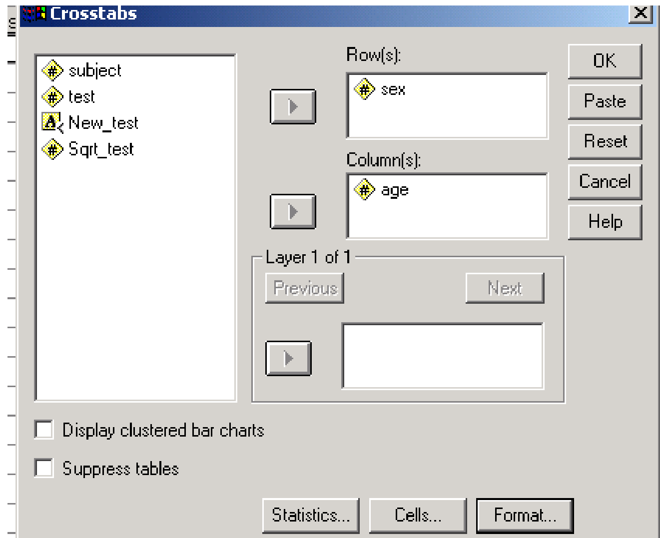


We can see from both the bar chart and the table that there are an equal number of males and females in the data set.

Frequency Tables with Two Variables

If you want to create a frequency table with two variables (crossed variables), such as the number of males and females at each age, use the Crosstabs procedure.

Analyze → Descriptive Statistics → Crosstabs



This will allow the user to create a table with one variable representing rows and another representing columns. Select the appropriate variables and move them to the correct box.

SPSS will create tables with more than two variables. Simply move the additional variables to the Layer box.

The Statistics option allows for statistics evaluating the association between variables. Cells allows the user to define what values to include in the cells. Format provides the option to report categories in ascending or descending order.

The output is as follows:

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
sex * age	6	100.0%	0	.0%	6	100.0%

sex * age Crosstabulation

Count

		age					Total
		18	19	20	21	22	
sex	Male	1	0	1	0	1	3
	Female	1	1	0	1	0	3
Total		2	1	1	1	1	6

Notice that there are two eighteen-year-olds – one male, one female. There is one nineteen-year-old, a female.

Assignment

1. Open the data saved from the previous assignment (SPSS Assignment 3).
2. Calculate the mean, standard deviation and range of body mass index, height and weight across all subjects.
3. Split the file by sex and calculate the mean and standard deviation of BMI for males and females.
4. Remove the split and create a frequency table for sex.
5. Create a frequency table that includes both sex and height (short or tall).

Output should look like:

Descriptives

Descriptive Statistics

	N	Range	Mean	Std. Deviation
Height	4	8.500	64.12500	3.750000
Weight	4	51	147.25	22.809
BMI	4	5	25.12	2.515
Valid N (listwise)	4			

Descriptives
Sex = F

Descriptive Statistics[a]

	N	Range	Mean	Std. Deviation
BMI	2	1	23.03	.798
Valid N (listwise)	2			

a. Sex = F

Sex = M

Descriptive Statistics[a]

	N	Range	Mean	Std. Deviation
BMI	2	1	27.21	.899
Valid N (listwise)	2			

a. Sex = M

Frequencies

Statistics

Sex

N	Valid	4
	Missing	0

Sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	F	2	50.0	50.0	50.0
	M	2	50.0	50.0	100.0
	Total	4	100.0	100.0	

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex * Height	4	100.0%	0	.0%	4	100.0%

Sex * Height Crosstabulation

Count

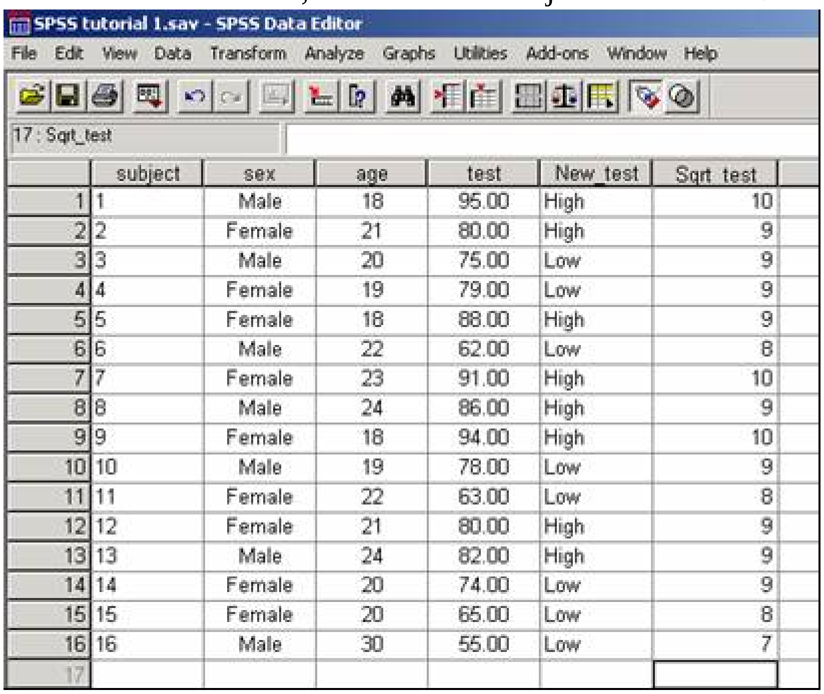
		Height				Total
		60.500	62.000	65.000	69.000	
Sex	F	0	1	1	0	2
	M	1	0	0	1	2
Total		1	1	1	1	4

Graphs and Charts

In this tutorial you will learn:

6. Identification of items in the Graphs menu
7. How to create bar graphs and line graphs
8. How to create scatterplots and histograms
9. How to create and manipulate interactive graphs
10. How to use the Gallery

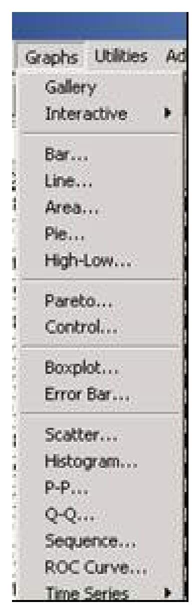
It is necessary to use a larger data set from those previously used for this and future tutorials. The data set from prior tutorials will be used, but additional subjects were added.



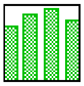
	subject	sex	age	test	New_test	Sqrt_test
1	1	Male	18	95.00	High	10
2	2	Female	21	80.00	High	9
3	3	Male	20	75.00	Low	9
4	4	Female	19	79.00	Low	9
5	5	Female	18	88.00	High	9
6	6	Male	22	62.00	Low	8
7	7	Female	23	91.00	High	10
8	8	Male	24	86.00	High	9
9	9	Female	18	94.00	High	10
10	10	Male	19	78.00	Low	9
11	11	Female	22	63.00	Low	8
12	12	Female	21	80.00	High	9
13	13	Male	24	82.00	High	9
14	14	Female	20	74.00	Low	9
15	15	Female	20	65.00	Low	8
16	16	Male	30	55.00	Low	7
17						

Graphs Menu

The Graphs menu includes many options. It allows you to specifically choose the type of graph or chart to create, including scatterplots, histograms, bar charts, line graphs, area graphs, pie charts, and box plots. It also contains an option to create interactive graphs that can be manipulated by rotating them or changing parameters.



Bar Graphs and Line Graphs

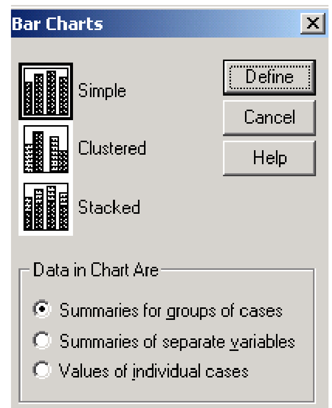


Bar Graphs

Graphs → Bar

Data or summaries of data are reflected on bars on a graph. This is appropriate when the variables on the X axis (horizontal axis) are categorical – either nominal or ordinal scale.

To create a bar graph, select Graphs → Bar. This opens a new window with options for the type of bar graph to be created:



<u>Simple</u> Each bar is a solitary piece of information. Bars may represent individual cases, frequencies, or means.

<u>Clustered</u> Categories of one variable (or two or more variables) can be represented within categories of another variable. Alternatively, two or more variables can be summarized for each individual case.

<u>Stacked</u> Similar to clustered, except bars are placed on top of each other rather than next to each other.

Within each of these types, there are three options for data to display.

<u>Summaries for groups of cases</u> Summarizes categories of an individual variable (or a variable within a variable).

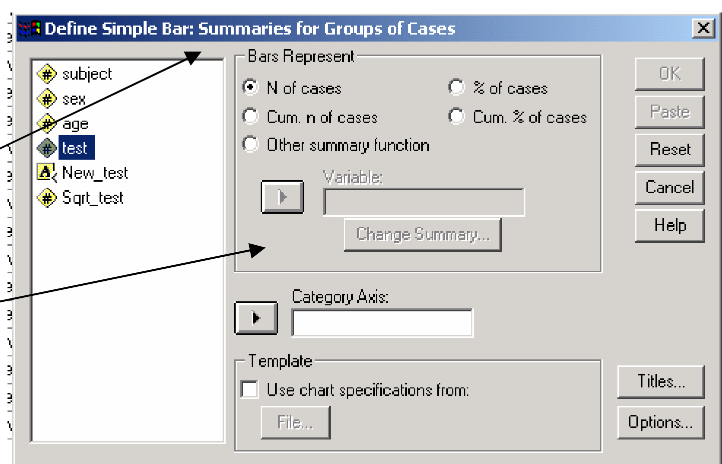
<u>Summaries of separate variables</u> Summarizes multiple variables with a bar for each variable (or each variable within categories of another variable).

<u>Values of individual cases</u> Summarizes a single variable (or variables within a variable) by individual case.

Once you have made your selection, click Define to further define the graph.

Simple Bar Graphs

Click Simple and determine the type of data to portray. For example, options for a simple bar chart with summaries for groups of cases, allows the following options:



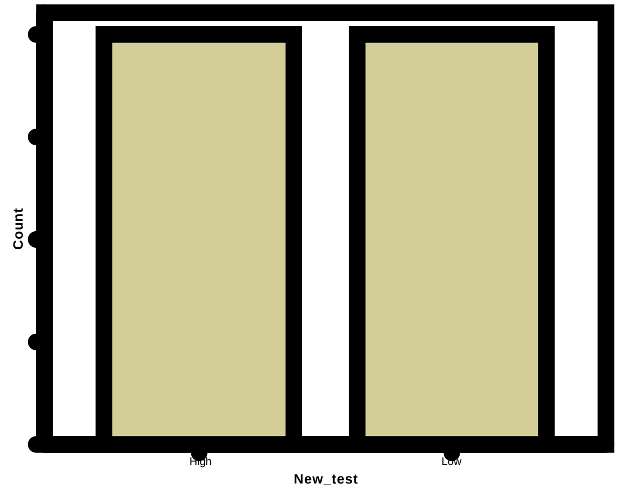
Determine what the bars represent: frequency (number of cases), percent of cases, cumulative percentage, etc. These values can be viewed in frequency tables. This will represent the Y axis (the vertical axis). Then, determine what variable will represent the X axis (the horizontal axis).

The Titles option allows the user to define a title and subtitle for the graph and footnotes as well.

Options allows the user to determine how to treat missing cases. When finished, click OK.

The following simple bar chart uses number of cases for the variable New_test.

Notice there are equal numbers of low and high test scores. This may tell an instructor that he/she is not inflating grades and is not grading too harsh, either. However, depending on the type of test, the attribute being tested, and the implications of the results, this graph may depict negative results. Interpretation is dependent upon many factors.



Using the Summaries of Separate Variables Option, the user may select from the following options to create a bar graph:

Mean of Values	Sum of Values	Minimum Value
Median of Values	Standard Deviation	Maximum Value
Mode of Values	Variance	Cumulative Sum
Number of Cases		

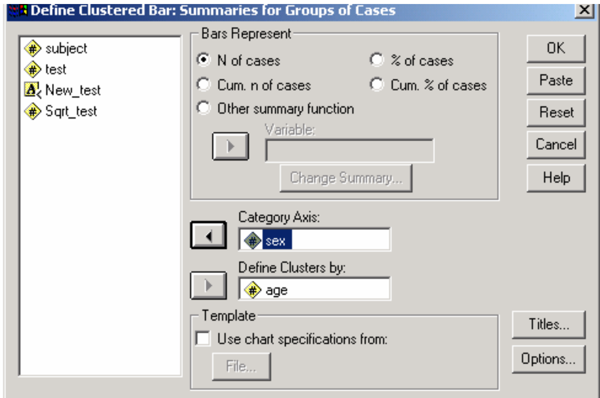
The user may also opt to identify a specific value and have bars reflect:

Percentage above	Percentile	Number below
Percentage below	Number above	

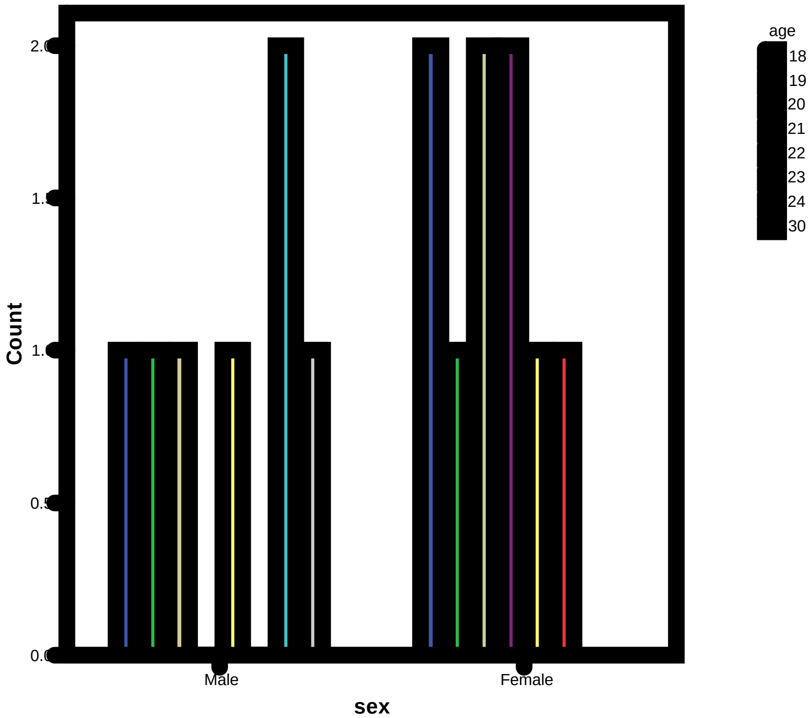
This is done by selecting the Change Summary option in the definition window. At least two variables must be selected in the Bars Represent window. This might be handy if the students took two or more tests and the average scores would be compared.

Clustered and Stacked Bar Graphs
The procedure for clustered bar graphs is similar to simple bar graphs, except a variable needs to be selected for clustering. The same procedures are used for stacked bar graphs. (Clustered are shown below.) One variable will be nested within another variable, as shown below:

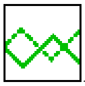
Here, sex provides the cluster and age is nested within sex. The final graph will depict sex on the X axis, number of cases on the Y axis. For the males and for the females, individual bars will reflect the number of cases for each level of age. See output.



Graph

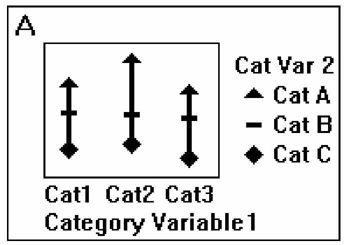


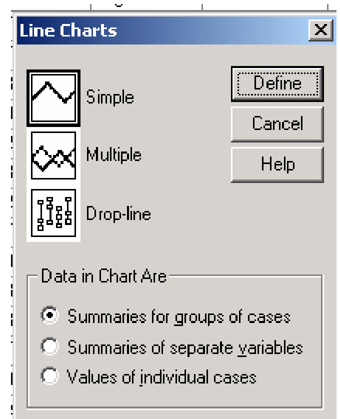
Managing Output
Resultant graphs can be adjusted by double clicking on the graph. A new, editing, window will open. In this way, colors, fonts, sizes of axes and titles can be changed. The legend may be deleted or inserted; text boxes can be inserted. Many other options are available.

*Line Charts*
Graphs → Line
Line charts differ from bar graphs in that the data are reflected by points, connected by a line. Line charts are appropriate when the variable on the X axis is numerical and interval or ratio scale data.

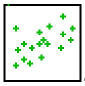
Simple, multiple and drop-line options are available. These are similar in definition to the definitions in bar graphs. Instead of clusters, multiple lines reflect variables within other variables in multiple line charts. Drop-line line charts are equivalent to stacked bar graphs, see below:

Drop-line Chart



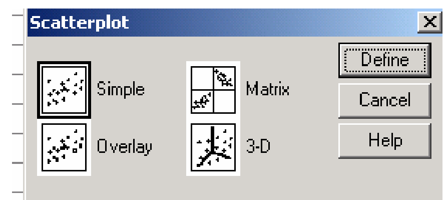


Scatterplots and Histograms

*Scatterplots*

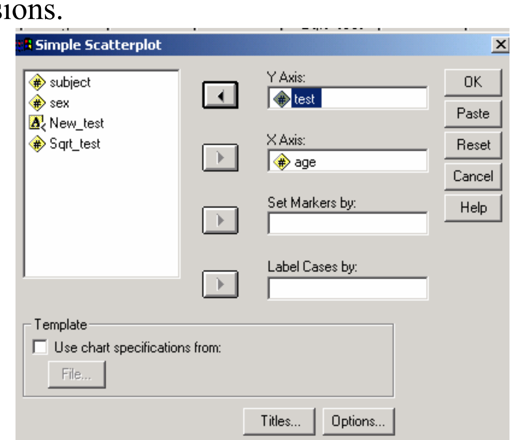
Graphs → Scatter

In a scatterplot, values of one numeric variable are depicted on the X axis and corresponding values of another numeric variable are depicted on the Y axis. Data points are presented on the graph depicting their corresponding scores on each variable. So, to create a scatterplot, two paired variables must be available. Often the independent variable is depicted on the X axis and the dependent variable is presented on the Y axis. This is especially useful in correlation and regression.



There are four options for type of scatterplot:

Simple	One X-Y variable pair
Overlay	More than one X-Y variable pairs (each depicted by a different marker)
Matrix	All possible pairs of variables are depicted by a cell with a scatterplot within the cell
3-D	Three variables are plotted in three dimensions.



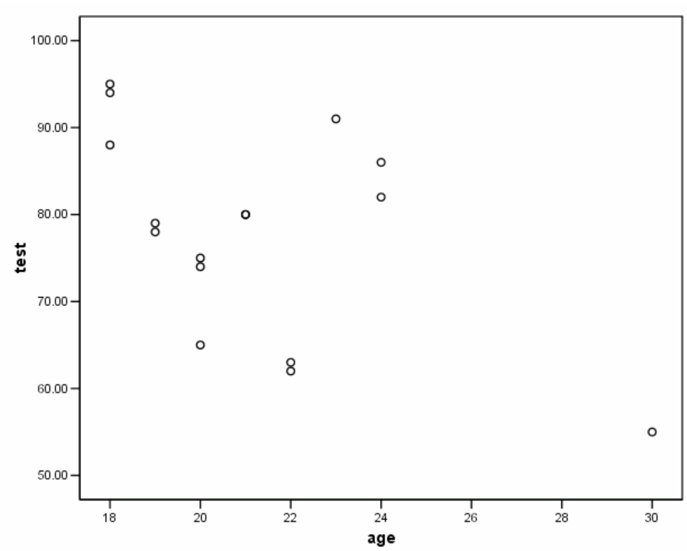
For our example, we might want to graph data points by age and test score. We go to

Graphs → Scatter

Select Simple, then click Define.

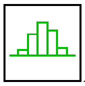
A new window will open in which one may select the X and Y variables.

Here, age will be on the X axis and test score on the Y axis.



We might have expected that test score would change as a function of age. That is, we might have expected that as people get older, they score higher (or lower). That does not appear to be the case here.

The scatterplot is also useful for identifying outliers, data points that are wildly different from the rest of the data. Notice the maverick data point in the bottom right corner. This reflects an individual that is older than the others and also scored much lower than everyone else. It would be worthwhile to make sure the data was entered correctly and to determine if there were other abnormalities with this case, to determine if it should be eliminated from the data set.

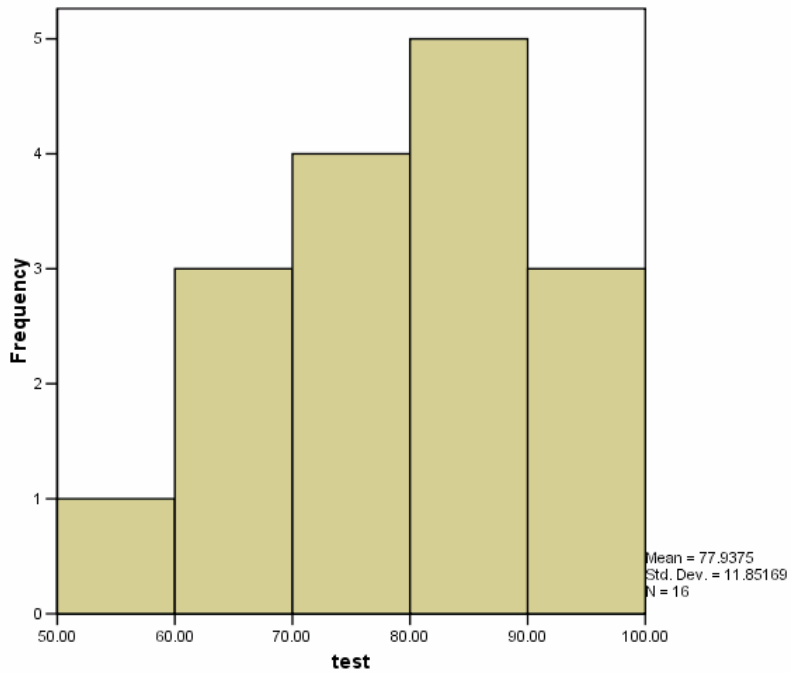
*Histograms*

Graphs → Histogram

Histograms show the distribution of a single variable.

Users may choose a variable to depict on the X axis. Frequency is depicted on the Y axis.

For the variable of test, the following histogram was created:



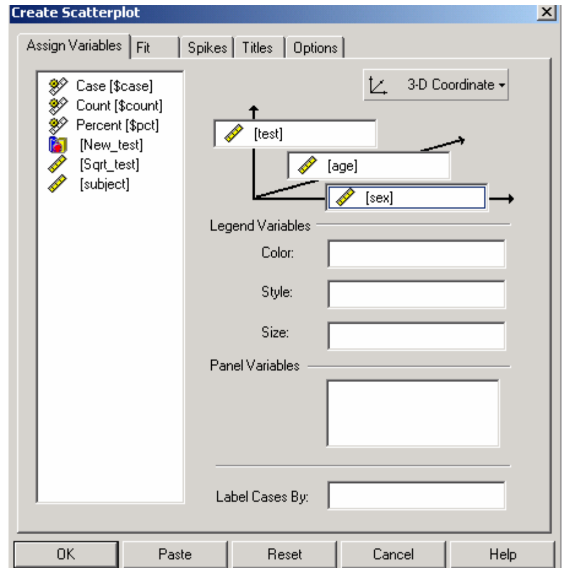
Notice that the highest number of test scores is between 80 and 90. The distribution seems to be negatively skewed, which means there is a tail to the left of the distribution.

Interactive Graphs

Graphs → Interactive → *Type of Graph*

The following types of graphs can be created interactively:

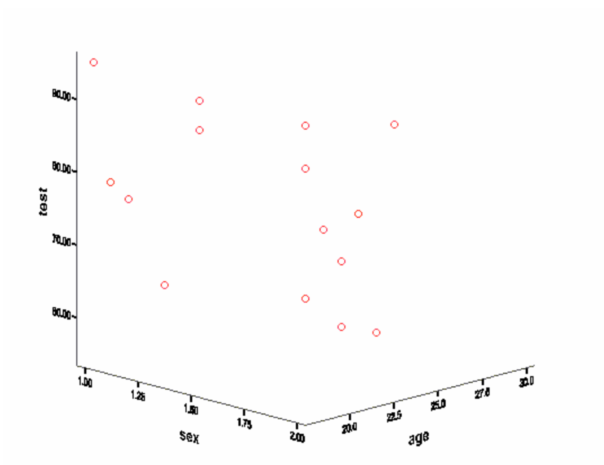
Bar	Line	Drop-line	Pie	Histogram
Dot	Ribbon	Area	Boxplot	Scatterplot



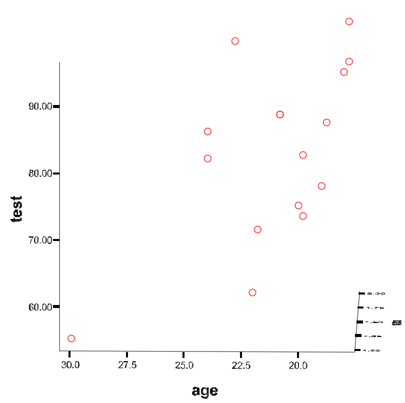
The example shown here will be for a scatterplot.

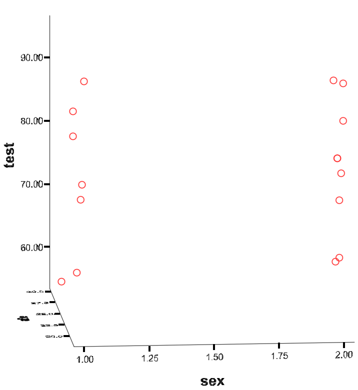
In this example, test will be on the Y axis. 3-D coordinate was selected, instead of 2-D, so there are 2 X-axes. On one X-axis is placed age and on the other, sex.

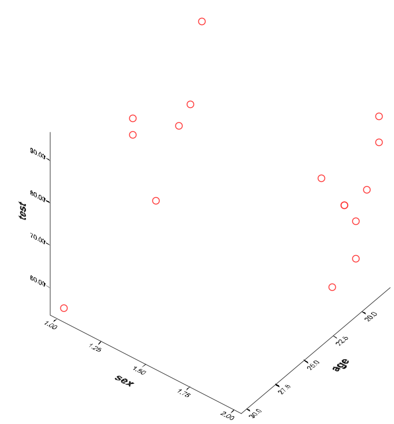
The output is printed below.



By double-clicking on the graph, we are able to manipulate the graph by adding symbols or text. We can also rotate the graph (see below for manipulated versions).



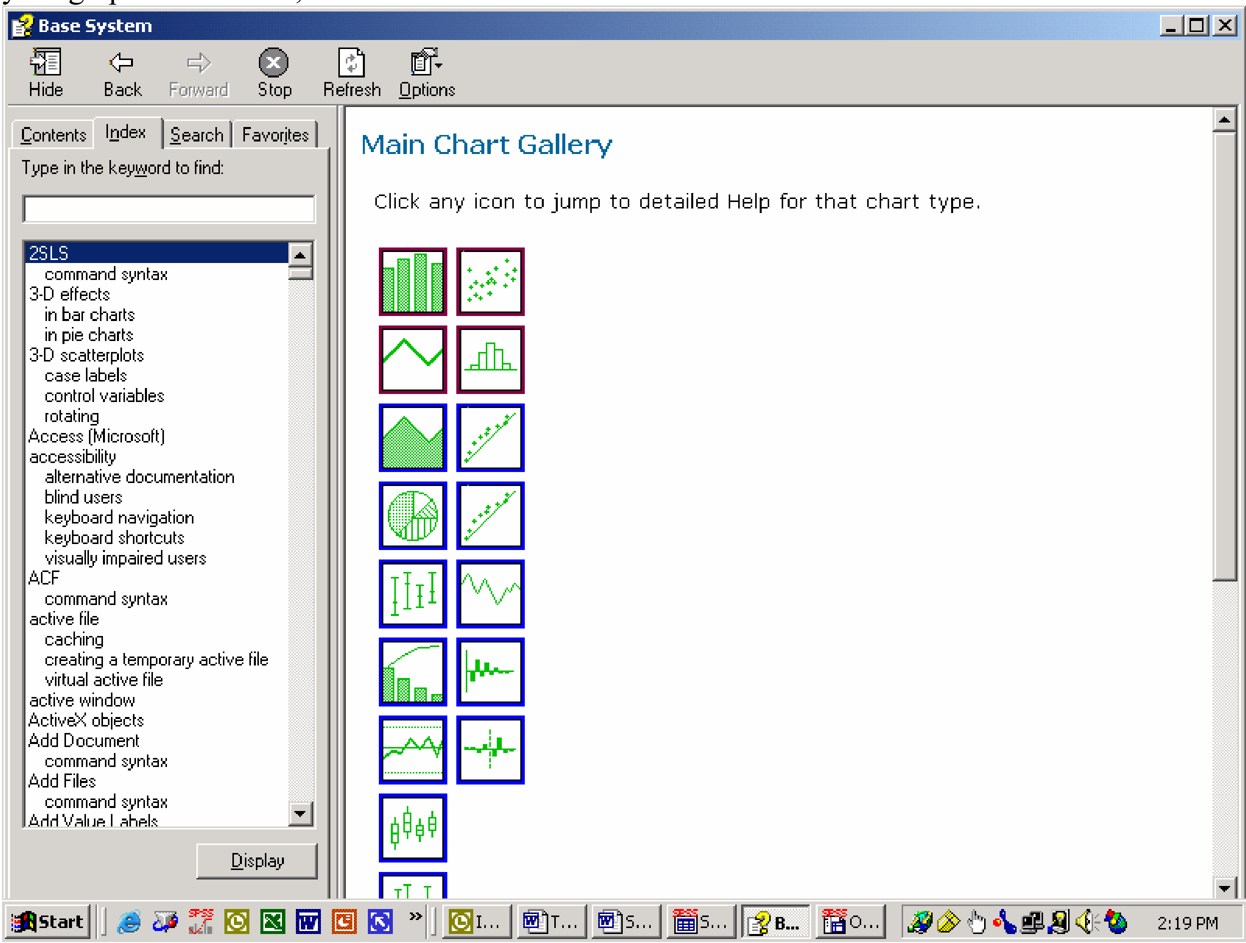




Gallery

Graphs → Gallery

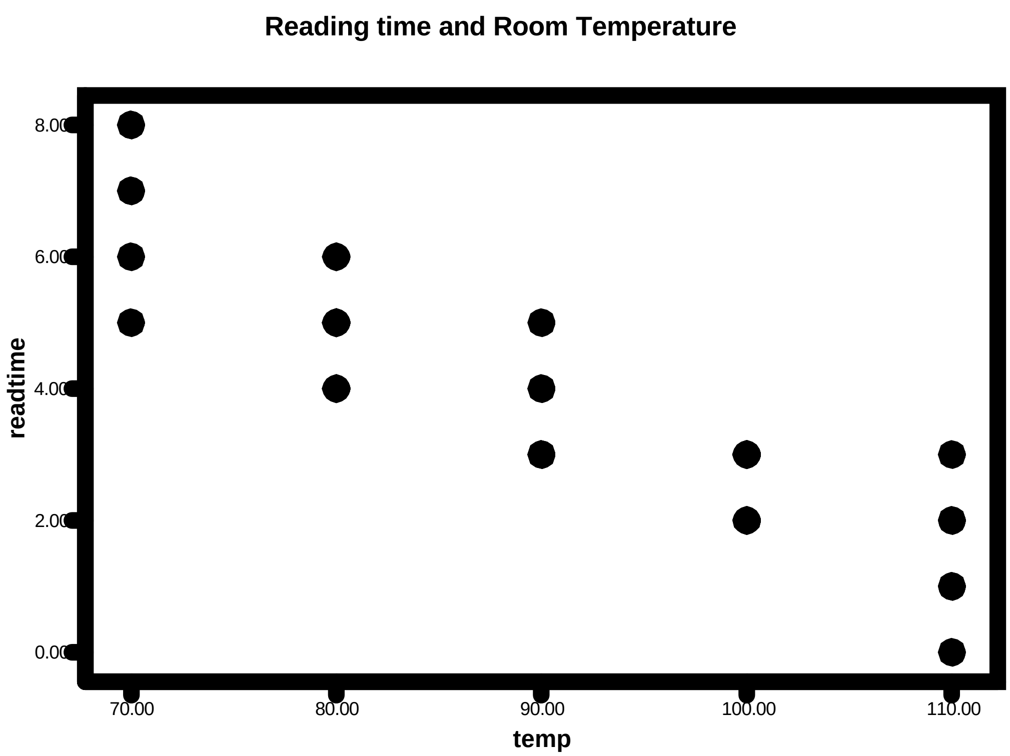
The Gallery is essentially a help page. It allows the user to click on a picture of a graph to obtain information about how to create it and what it is. This is useful if you do know what you want your graph to look like, but do not know what it is called.



Assignment

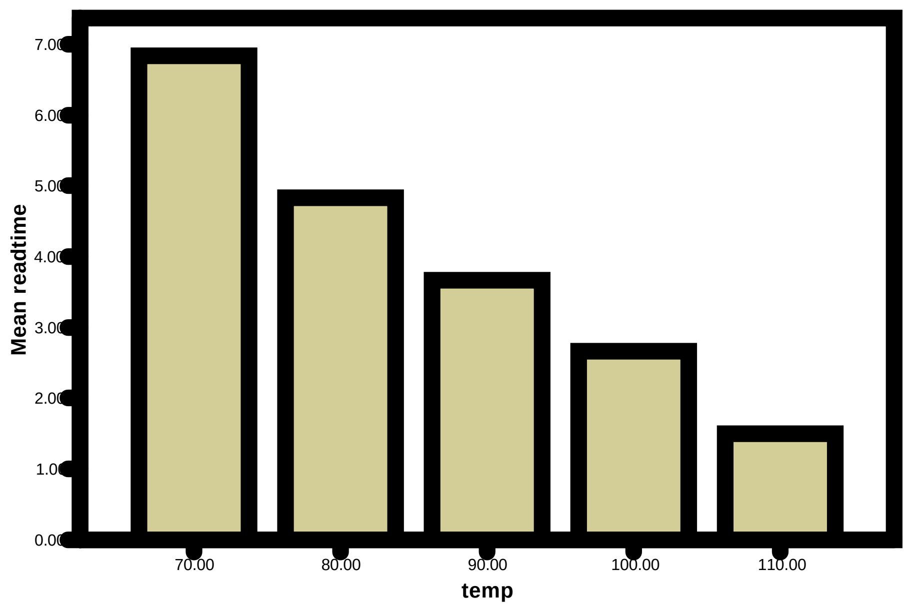
1. Open the data set entitled Readingtime.sav. The data are subject number, temperature in the room, and amount of time children continued to read. The data are fictitious.
2. Create a scatterplot with temperature on the X axis and reading time on the Y axis. Interpret the results.
3. Create a bar graph with temp on the X axis and mean reading time on the Y axis. You will have to use the Other Summary Function.

Output should look like:



As temperature increases, reading time decreases.

Reading time and temperature



One- and Two-Sample *t*-Tests

In this tutorial you will learn:

1. How to compute a one-sample *t*-test and interpret output
2. How to compute a two-sample dependent *t*-test and interpret output
3. How to compute a two-sample independent *t*-test and interpret output

The data from prior tutorials will be used here, with the following exceptions:

- Sqrt_test has been deleted
- A new test score, End_test, has been added to reflect the students' score on the test at the end of the year.

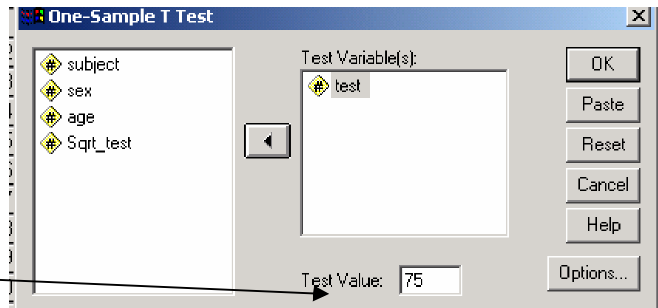
	subject	sex	age	test	New_test	End_test
1	1	Male	18	95.00	High	71.33
2	2	Female	21	80.00	High	61.33
3	3	Male	20	75.00	Low	58.00
4	4	Female	19	79.00	Low	60.67
5	5	Female	18	88.00	High	66.67
6	6	Male	22	62.00	Low	49.33
7	7	Female	23	91.00	High	68.67
8	8	Male	24	86.00	High	65.33
9	9	Female	18	94.00	High	70.67
10	10	Male	19	78.00	Low	60.00
11	11	Female	22	63.00	Low	50.00
12	12	Female	21	80.00	High	61.33
13	13	Male	24	82.00	High	62.67
14	14	Female	20	74.00	Low	57.33
15	15	Female	20	65.00	Low	51.33
16	16	Male	30	55.00	Low	44.67

One-sample *t*-test

Analyze → Compare Means → One-sample T Test

A one-sample *t*-test is used to test differences between a sample mean and a hypothesized (null) value. For example, it might be well-known that the mean score on the test (from the sample data used in the prior tutorials) is 75. This is a standard test that has been used and validated over several years. We hypothesize that our group of students will score differently than this due to a new training program in which they participated. A one-sample *t* -test comparing our sample mean to the population mean of 75 is appropriate. Remember, no statistic can make up for poor research design. So, review hypothesis testing and experimental design before running an experiment and analyzing data.

To test this,
Analyze → Compare Means → One-sample T Test



Move all variables to be analyzed to the Test Variable(s) box using the right arrow button.

Specify the hypothesized (null) value to which your sample mean will be compared. In this case, it is 75.

The Options button allows you to specify confidence intervals and to determine how to treat missing data. The options for missing data include:

<u>Exclude cases listwise</u>	If any case has a missing data point in a variable used in any of the *t*-tests requested, eliminate all data related to that case. Sample size is constant across all analyses.
<u>Exclude cases analysis by analysis</u>	If any case has a missing data point, eliminate data related to that case only for analyses that involve the variable with missing data. Sample size varies by analysis.

Output

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
test	16	77.9375	11.85169	2.96292

One-Sample Test

	Test Value = 75					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
test	.991	15	.337	2.93750	-3.3778	9.2528

The first table presents descriptive statistics. Double check this box to make sure all values are correct. If not, look for data entered incorrectly or determine if there are missing data points or other anomalies.

The second table presents results of the one-sample *t*-test. The *t* value is .991, with 15 degrees of freedom (*n*-1). The probability of collecting a sample with the mean value of 77.9375 given that the true population mean is 75 is .337 (Sig.). This is not significant at both α =.05 and α=.01. We fail to reject our null hypothesis.

The difference between our obtained sample mean and the hypothesized mean of 75 is 2.9375l. The 95% confidence interval for the difference in means is -3.3778 to 9.2528. We are 95% confident that the true population mean falls within this interval, as estimated with our sample data. Notice that the interval contains 2.93751. This corresponds with our decision to fail to reject the null hypothesis.

Two-sample dependent *t*-test

Analyze → Compare Means → Paired Samples T Test

Often we have two different groups whose means we want to compare. In this case a one-sample *t*-test is not appropriate (because we have two samples). There are two possible ways to design a two-sample experiment: dependent or independent. In dependent tests, either the same subjects or matched pairs contribute data. In the first case, a repeated measures design, subjects are tested at two separate occasions. One example is a pre-test, post-test design in which subjects take a test, receive a treatment, and then retake the test. Researchers are interested in change from the pre-test to the post-test and assume the change is due to the treatment (e.g., a study skills course, drug-therapy, etc.). In the second case, two individuals are matched on some attribute or list of attributes. For example, pairs of students may be matched on age and IQ but differ by gender to determine if males score differently on standardized tests than females. In another example, adults may be paired with their spouses to determine if the husbands' data differs from the wives' data.

Independent tests will be discussed in the next section.

Set up the data

SPSS requires the user to specify the two paired variables. The paired values must appear on the same row.

For example, we may want to compare pre-test and post-test scores for subjects. The data would be entered as follows:

Subject	Pretest	Posttest
1	30	40
2	40	50
3	50	60
4	60	70
5	70	80

Notice that each row represents a subject; the two paired variables (pre-test and post-test) appear as two columns.

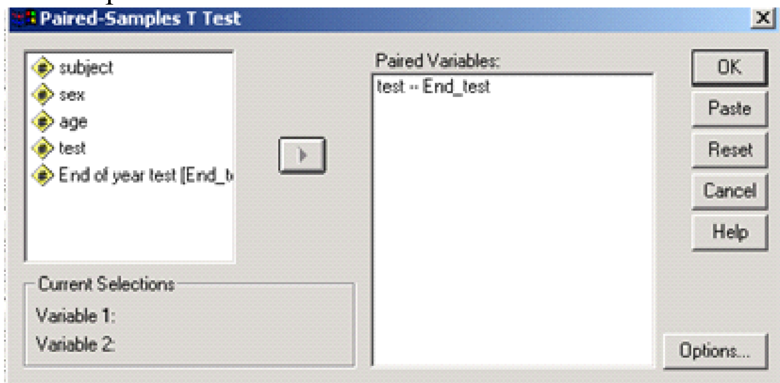
In a matched design, the data are entered the same way, except each row represents a pair. The two paired variables represent scores from each member of the pair:

Pair	Husband	Wife
1	7	6
2	8	5
3	9	4
4	7	6
5	6	7

For our example, let's assume that our students took the test at the beginning of the year. At the end of the year, they took an equivalent test. We want to test the null hypothesis that there is no change in test score from the beginning of the year (*test*) to the end of the year (*End_test*).

Analyses

Analyze → Compare Means → Paired-Samples T Test



Select the two paired variables in the box to the right by clicking on one then the other. They will appear under Current Selections as they are selected. Then use the right-arrow button to move the pair to the Paired Variables box. You may select any number of pairs.

The Options button allows the user to identify confidence intervals and to determine how to manage missing data (see above).

Output

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	test	77.9375	16	11.85169	2.96292
	End of year test	59.9583	16	7.90113	1.97528

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	test & End of year test	16	1.000	.000

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	test - End of year test	17.97917	3.95056	.98764	15.87406	20.08427	18.204	15	.000

The first table contains descriptive statistics for each variable.
The second presents correlations between the two variables. Here they are perfectly correlated (because the data for End_test was created as a function of test).
The third table reports the mean and standard deviation of the difference in value between the pairs. (Note that in a dependent *t*-test, the difference in score for each subject or pair is calculated and then determination is made to determine if the mean difference is significantly greater or less than zero.) The 85% confidence interval for the mean difference in presented. Notice it does not include zero. The *t* value for this test is 18.204 with 15 degrees of freedom (n_p-1). This is significant. Notice, the probability value under Sig. (2-tailed) is .000 (or reported as <.001)

which is less than .05 or .01. We can conclude that students scored significantly higher at the beginning of the year than at the end of the year.

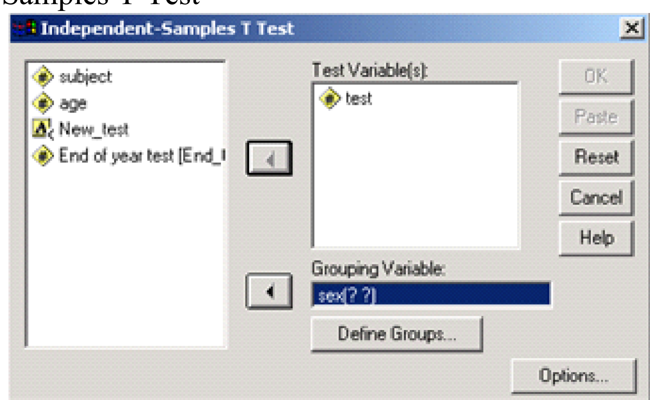
One-sample *t*-test

Analyze → Compare Means → Independent-Samples T Test

In an independent design, two samples are obtained, but members are not paired. The two samples are independent of one another. Hypotheses are concerned with difference between groups. For example, we might be interested in the difference between males and females on the test variable.

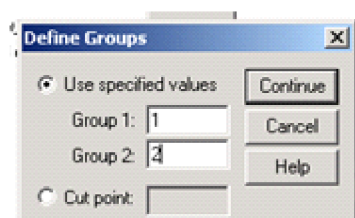
Set Up Data
The data must be arranged with one grouping variable (independent variable) and one test variable (dependent variable). Our data are arranged in such a way. The sex variable is our independent variable and the test variable is our dependent variable.

Analyses
Analyze → Compare Means → Independent-Samples T Test



Move the test variable to the Test Variable window with the right-arrow button. Move the grouping variable to the appropriate window with the associated right-arrow button.

You must also define the groups in the Grouping Variable. Select the Define Groups button.



Identify the values to define Group 1 and Group 2. Recall, we coded sex as 1 (male) and 2 (female). Then select continue.

An alternative is to use a numeric variable and identify a cut-off value that will divide the data into two groups (those above and those below the cut-off). This is useful when using median or mean splits.

Output

T-Test

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
test	Male	7	76.1429	13.77714	5.20727
	Female	9	79.3333	10.77033	3.59011

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
test	Equal variances assumed	.416	.529	-.521	14	.610	-3.19048	6.12324	-16.32351	9.94256
	Equal variances not assumed			-.504	11.167	.624	-3.19048	6.32491	-17.08612	10.70517

The first table includes descriptive statistics.
The second table includes, first, Levene's test for equality of variances. If one cannot assume the two groups have equal variance, it is not appropriate to use the pooled variance in computing a *t* value and degrees of freedom must be estimated, especially if the two groups have unequal sample sizes. Levene's test provides a test of the two sample variances. If significant, one cannot assume equal variance. If not significant, then make the assumption that the two groups have equal (or at least similar enough) variances.

If equal variance is a concern for you, first check to see if you can assume equal variance. If so, use the *t*-tests on the first line. If not, use the second line.

Here, we will assume equal variance. The *t* value is -.521 with associated degrees of freedom of 14 (n_1+n_2-2). This is not significant (sig. = .610). The mean difference between groups is -3.19048. This is followed by the standard error for the sampling distribution of mean differences and the 95% confidence interval for the difference in means.

Assignment

1. Open the data Readingtime.sav.
2. Filter out the data for temp=110.
3. Run the appropriate test of the null hypothesis that reading time is equal to 0.
4. Run the appropriate test of the null hypothesis that students read longer in cooler temperatures (70 and 80 degrees) than in warmer temperatures (90 and 100 degrees).

Output should look like:

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
readtime	24	4.5000	1.76930	.36116

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
readtime	12.460	23	.000	4.50000	3.7529	5.2471

Students read a significant amount of time greater than 0.

T-Test

Group Statistics

	temp	N	Mean	Std. Deviation	Std. Error Mean
readtime	>= 85.00	12	3.1667	.83485	.24100
	< 85.00	12	5.8333	1.40346	.40514

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
readtime	Equal variances assumed	5.037	.035	-5.657	22	.000	-2.66667	.47140	-3.64430	-1.68903
	Equal variances not assumed			-5.657	17.918	.000	-2.66667	.47140	-3.65737	-1.67596

The difference in reading time between high and low temperatures is significant. They read longer under cooler temperatures.

One-Way ANOVA

In this tutorial you will learn:

4. How to compute a one-way ANOVA and interpret output
5. How to run planned contrasts and interpret output
6. How to compute post-hoc comparisons and interpret output

The data for this tutorial include:

Drug Group — Indicates to which group the individual was randomly assigned: the old, commonly used drug (1), a new experimental drug (2), or a placebo (3).

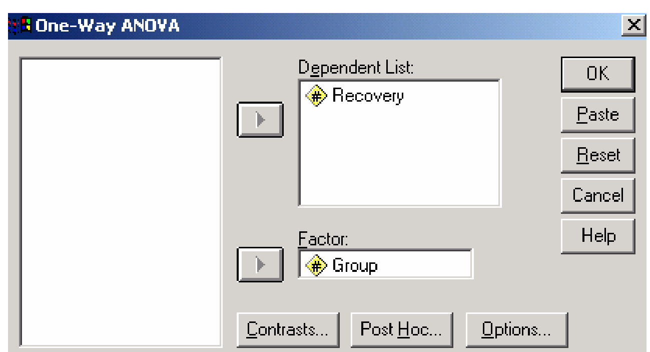
Recovery — Amount of time to recover from minor surgery (in days)

	Group	Recovery
1	1	5
2	1	8
3	1	5
4	1	7
5	1	6
6	2	8
7	2	8
8	2	9
9	2	7
10	2	6
11	3	3
12	3	3
13	3	5
14	3	2
15	3	4

The researcher wonders if there is a difference in recovery rate among groups and whether or not the new drug improves recovery. A one-way ANOVA will evaluate the variance among the group means as a function of overall variance. It is appropriate to test differences among 2 or more groups. A significant result can be interpreted to mean that all the group means are not equal (or close to equal).

One-way ANOVA

Analyze → Compare Means → One-Way ANOVA



Move the variables of interest from the left-hand window to the appropriate right-hand windows with the right-arrow buttons. The dependent variable goes in the window labeled Dependent List. The independent variable (grouping variable) goes in the window marked Factor. Then click OK.

The Options button provides the opportunity to include descriptive statistics, tests of homogeneity of variance (Levene test) in the output. It also allows the user to identify how to treat missing data.

Output

Oneway

Descriptives

Amount of time to recover from surgery (days)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
New Drug	5	6.20	1.304	.583	4.58	7.82	5	8
Old Drug	5	7.60	1.140	.510	6.18	9.02	6	9
Placebo	5	3.40	1.140	.510	1.98	4.82	2	5
Total	15	5.73	2.120	.547	4.56	6.91	2	9

ANOVA

Amount of time to recover from surgery (days)

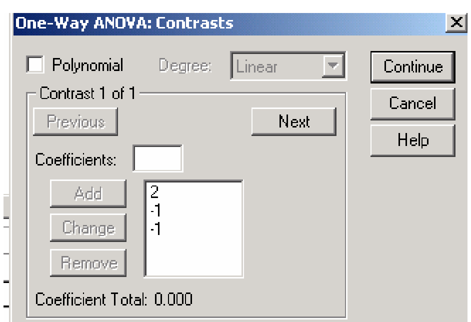
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	45.733	2	22.867	15.953	.000
Within Groups	17.200	12	1.433		
Total	62.933	14			

The first table presents descriptive statistics because that option was selected in the Options window. The second table presents the results of the ANOVA. We can see that the overall F is significant with 2 and 12 degrees of freedom, $F(2, 12) = 15{,}953$, $p < .001$. We can conclude that at least two group means are significantly different.

To learn which means are significantly different, further tests must be run: planned or post hoc.

Planned Contrasts

Planned contrasts are analyses that are conducted when they were explicitly planned for before data was collected. They take the form of special linear combinations in which all the weights sum to 0.



In the one-way ANOVA window, select Contrasts: Polynomial contrasts, also known as trends, including linear, quadratic, cubic, 4^{th} and 5^{th} are allowed with the Polynomial option.

Users may also define their own contrasts by supplying weights (that must sum to 0). Type each weight in the Coefficient box, then click add. Do this for each weight in order from first to last. The contrast selected here compares the new drug to the mean of old drug and placebo. It tests if the new drug results in different recovery times than the other 2 methods. More than one contrast can be selected by clicking the Next button.

Output

The following output is added to the one-way ANOVA:

Contrast Coefficients

	Drug Group		
Contrast	New Drug	Old Drug	Placebo
1	2	-1	-1

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Amount of time to recover from surgery (days)	Assume equal variances	1	1.40	1.311	1.067	12	.307
	Does not assume equal	1	1.40	1.371	1.021	7.123	.341

The first table lists the groups and the associated weights we selected. Double check to make sure it is correct.

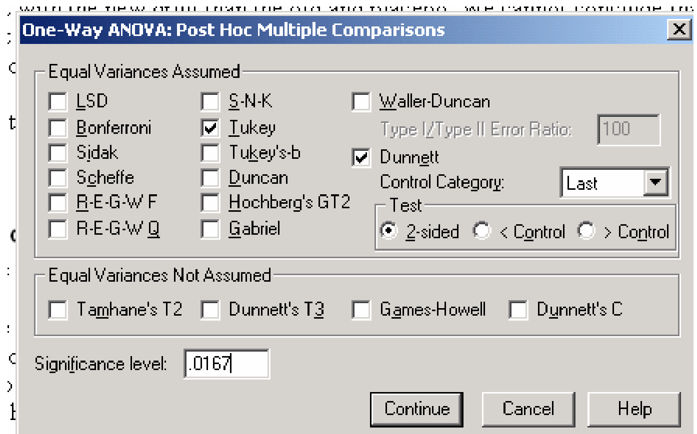
The second table reports the results of the contrast as *t*-tests. We can see that our contrast is not significant, $t(12) = 1.067$, $p = .307$. We cannot conclude, from this data, that recovery is different with the new drug than the old and placebo. We cannot conclude that new drug is no different than old drug and that new drug is no different than placebo. First, that is *accepting* the null hypothesis (bad). Second, remember, that this contrast lumps together the old drug and the placebo. Looking at the group means we can see that the mean with Old drug is higher than New drug, but Placebo is lower.

Post hoc Comparisons

Post hoc comparisons are comparisons of group means made after data have been collected. They do not assume any prior hypotheses. Most are pairwise comparisons, meaning they compare all pairs of means, to determine if they are significantly different. If there are a large number of group means, resulting in many pairwise comparisons, it may be wise to use a correction for alpha to control for familywise (or overall) Type I error rates. The more pairs you test, the higher your familywise alpha.

To run post hoc analyses on SPSS, in the one-way ANOVA window, select Post Hoc.

Tests are grouped by equal variance assumption. If one can assume equal variances, then the top group should be used. If not, the bottom group is available. For this data, we have selected Tukey and Dunnett. Because Dunnett is specifically geared to test each group mean against a control group mean, the control group must be selected. In this case, it is the last group, Placebo.



We also adjusted the significance level to .0167 (.05/3) to control the Type I error rate.

Output

The following output is added to the one-way ANOVA.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Amount of time to recover from surgery (days)

	(I) Drug Group	(J) Drug Group	Mean Difference (I-J)	Std. Error	Sig.	98.33% Confidence Interval Lower Bound	Upper Bound
Tukey HSD	New Drug	New Drug					
		Old Drug	-1.400	.757	.196	-3.89	1.09
		Placebo	2.800*	.757	.008	.31	5.29
	Old Drug	New Drug	1.400	.757	.196	-1.09	3.89
		Old Drug					
		Placebo	4.200*	.757	.000	1.71	6.69
	Placebo	New Drug	-2.800*	.757	.008	-5.29	-.31
		Old Drug	-4.200*	.757	.000	-6.69	-1.71
		Placebo					
Dunnett t (2-sided)[a]	New Drug	New Drug					
		Old Drug					
		Placebo	2.800*	.757	.006	.45	5.15
	Old Drug	New Drug					
		Old Drug					
		Placebo	4.200*	.757	.000	1.85	6.55
	Placebo	New Drug					
		Old Drug					
		Placebo					

*. The mean difference is significant at the .0167 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Homogeneous Subsets

Amount of time to recover from surgery (days)

	Drug Group	N	Subset for alpha = .0167	
			1	2
Tukey HSD[a]	Placebo	5	3.40	
	New Drug	5		6.20
	Old Drug	5		7.60
	Sig.		1.000	.196

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

The first table presents the results of the post hoc comparisons. The top part presents the results of Tukey's HSD. Each cluster compares the I group to each of the J groups. For example, the first cluster compares the new drug to the old drug and to placebo. New drug is significantly different from placebo, but not old drug. The next cluster compares old drug to the other groups. It is significantly different from placebo. From these results, we can conclude that the drugs, while not significantly different from each other, result in significantly longer recovery times than placebo.

The bottom part of the table presents the results of Dunnett t. It is read the same way, but recall that it will only compare each group to the control group (placebo). The same interpretations can be drawn.

The second table arranges like groups together. It is shown that the new drug and the old drug are not significantly different from each other, but placebo differs from both.

Assignment

1. Open the data titled Readingtime.sav. Remove all filters (from the last tutorial).
2. Run the appropriate test to determine if there is a difference in reading time among the different room temperatures. Additionally,
   a. Run the appropriate planned contrast to compare 100 and 110 degrees to 70 and 80 degrees.
   b. Conduct a post-hoc analysis of your choosing.

Output should look like:

Oneway

Descriptives

readtime

					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
70.00	6	6.8333	1.16905	.47726	5.6065	8.0602	5.00	8.00
80.00	6	4.8333	.75277	.30732	4.0433	5.6233	4.00	6.00
90.00	6	3.6667	.81650	.33333	2.8098	4.5235	3.00	5.00
100.00	6	2.6667	.51640	.21082	2.1247	3.2086	2.00	3.00
110.00	6	1.5000	1.04881	.42817	.3993	2.6007	.00	3.00
Total	30	3.9000	2.04011	.37247	3.1382	4.6618	.00	8.00

ANOVA

readtime

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	100.867	4	25.217	31.786	.000
Within Groups	19.833	25	.793		
Total	120.700	29			

Contrast Coefficients

	temp				
Contrast	70.00	80.00	90.00	100.00	110.00
1	1	1	0	-1	-1

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
readtime	Assume equal variances	1	7.5000	.72725	10.313	25	.000
	Does not assume equal	1	7.5000	.74162	10.113	15.692	.000

Post Hoc Tests

Multiple Comparisons

Dependent Variable: readtime
LSD

(I) temp	(J) temp	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
70.00	70.00					
	80.00	2.00000*	.51424	.001	.9409	3.0591
	90.00	3.16667*	.51424	.000	2.1076	4.2258
	100.00	4.16667*	.51424	.000	3.1076	5.2258
	110.00	5.33333*	.51424	.000	4.2742	6.3924
80.00	70.00	-2.00000*	.51424	.001	-3.0591	-.9409
	80.00					
	90.00	1.16667*	.51424	.032	.1076	2.2258
	100.00	2.16667*	.51424	.000	1.1076	3.2258
	110.00	3.33333*	.51424	.000	2.2742	4.3924
90.00	70.00	-3.16667*	.51424	.000	-4.2258	-2.1076
	80.00	-1.16667*	.51424	.032	-2.2258	-.1076
	90.00					
	100.00	1.00000	.51424	.063	-.0591	2.0591
	110.00	2.16667*	.51424	.000	1.1076	3.2258
100.00	70.00	-4.16667*	.51424	.000	-5.2258	-3.1076
	80.00	-2.16667*	.51424	.000	-3.2258	-1.1076
	90.00	-1.00000	.51424	.063	-2.0591	.0591
	100.00					
	110.00	1.16667*	.51424	.032	.1076	2.2258
110.00	70.00	-5.33333*	.51424	.000	-6.3924	-4.2742
	80.00	-3.33333*	.51424	.000	-4.3924	-2.2742
	90.00	-2.16667*	.51424	.000	-3.2258	-1.1076
	100.00	-1.16667*	.51424	.032	-2.2258	-.1076
	110.00					

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

All comparisons are significant except 90 and 100.