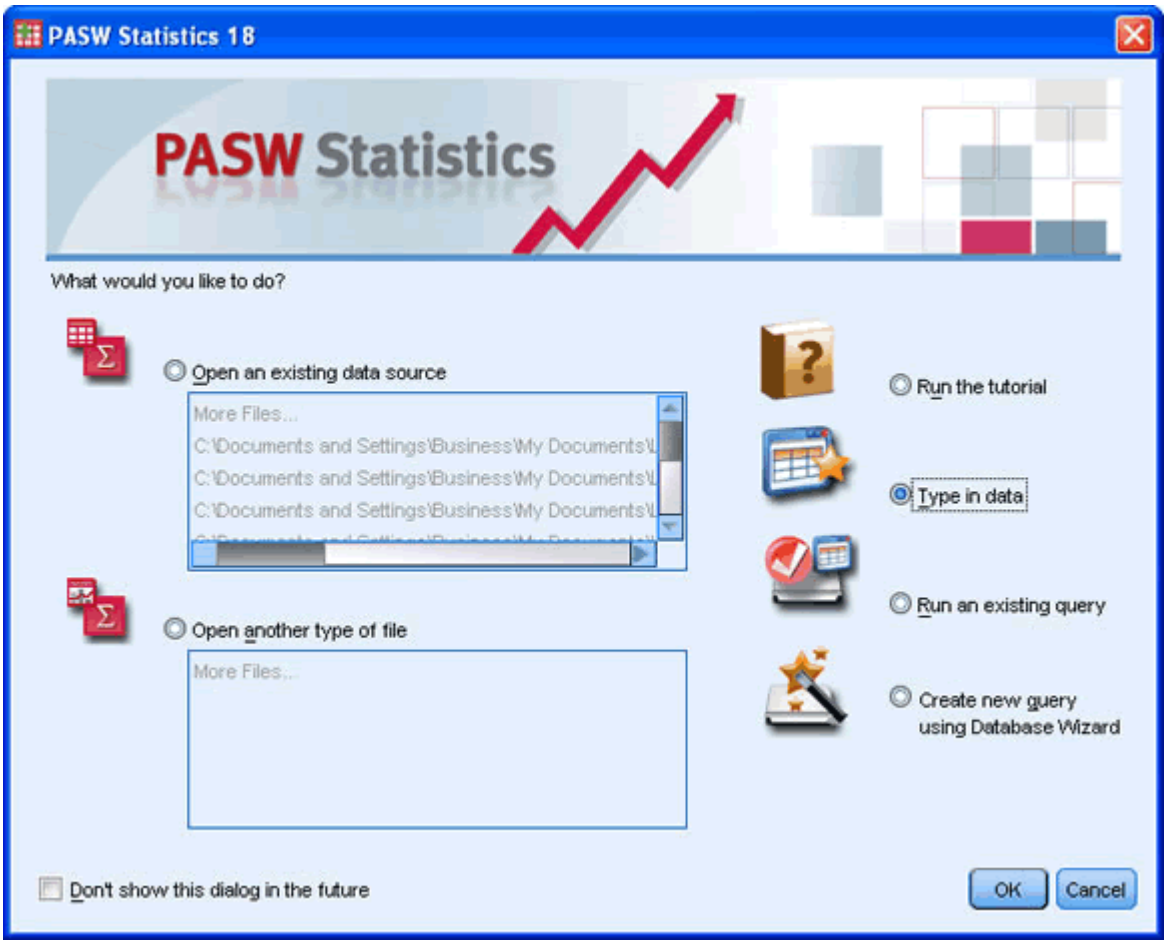
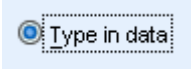
Getting to Know SPSS

Creating a New File in SPSS

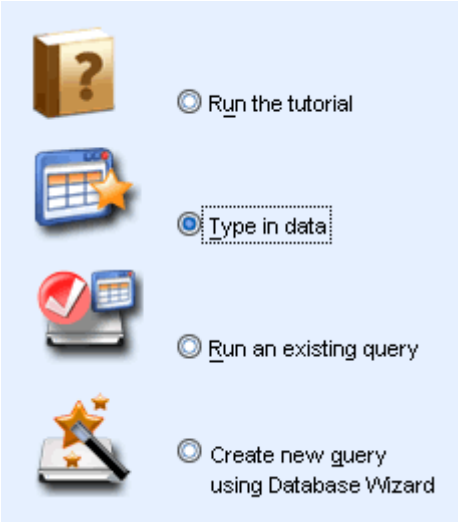
Creating a new file in SPSS is an essential, basic step before you can get fully started using the more complex operations available with this software. Please follow the guide below on how to create a new file in SPSS Statistics 18.0.

1. There are two ways to create a new file in SPSS - either when you start SPSS or when SPSS is already open. When you start SPSS you will be presented with the following screen:

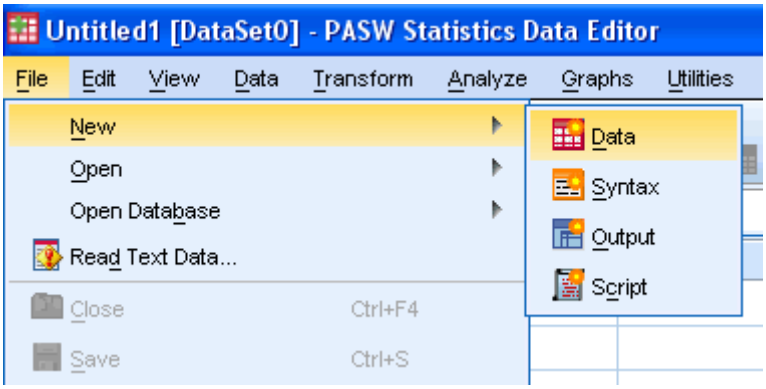


2. Make sure that 

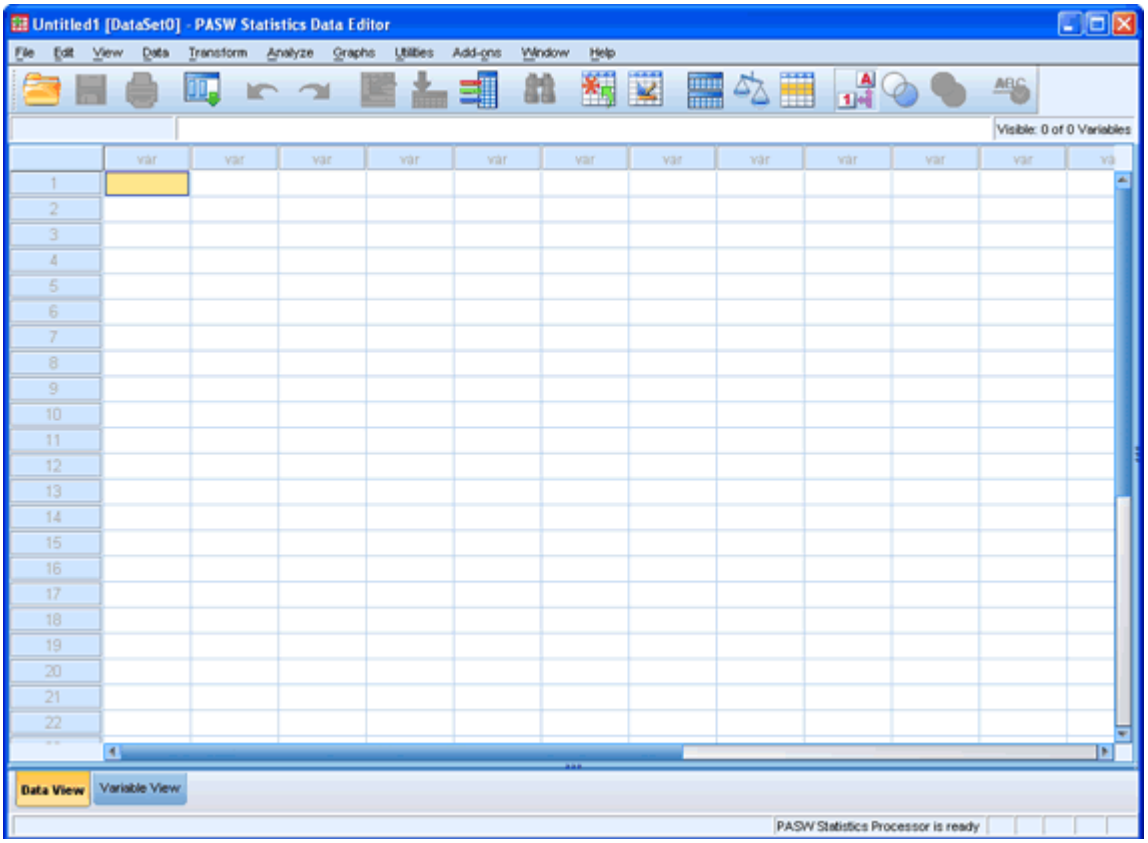
is selected as show below:



3. Click OK the button.
4. Otherwise, if SPSS is already open then click **File > New > Data** as shown below:

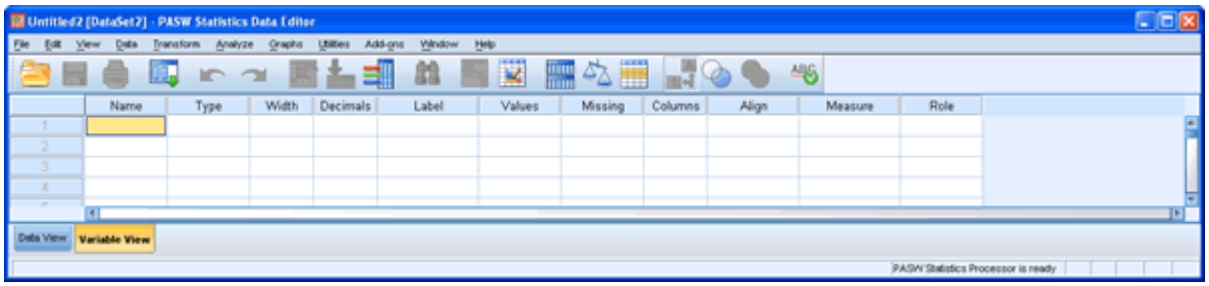


5. Both these options will lead to the following screen:



Working with Variables in SPSS

In SPSS, you need to define your variables, which occur in the Variables View. To access the Variable View you need to click the Variables View tab as shown below:



Here you can adjust the properties of each of your variables under 10 categories - Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align and Measure.

Name To change the name of a column (variable) in the Data View sheet, click in the appropriate cell and type in the new name. The names in this column must not start with a number. They also cannot contain special characters such as / * $, space etc. You will be given an error message if your name is in illegal format.

Type To define the type of data contained in the column (e.g. characters, strings, numbers etc.) click on the appropriate row in the Type column.

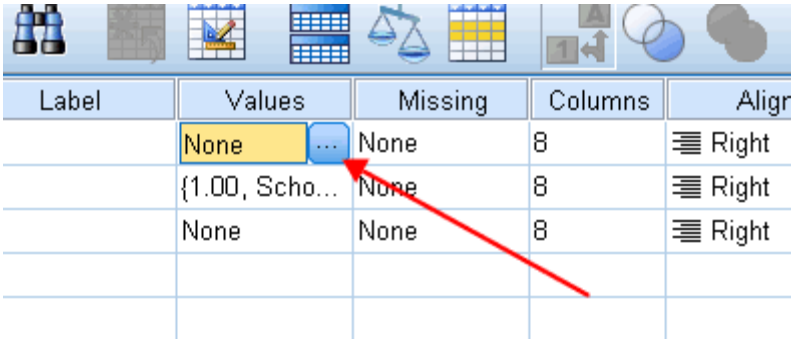
Width You can alter the number of digits displayed in the column by clicking in the appropriate Width cell.

Decimals You can alter the number of digits after the decimal place by clicking in the appropriate Width cell.

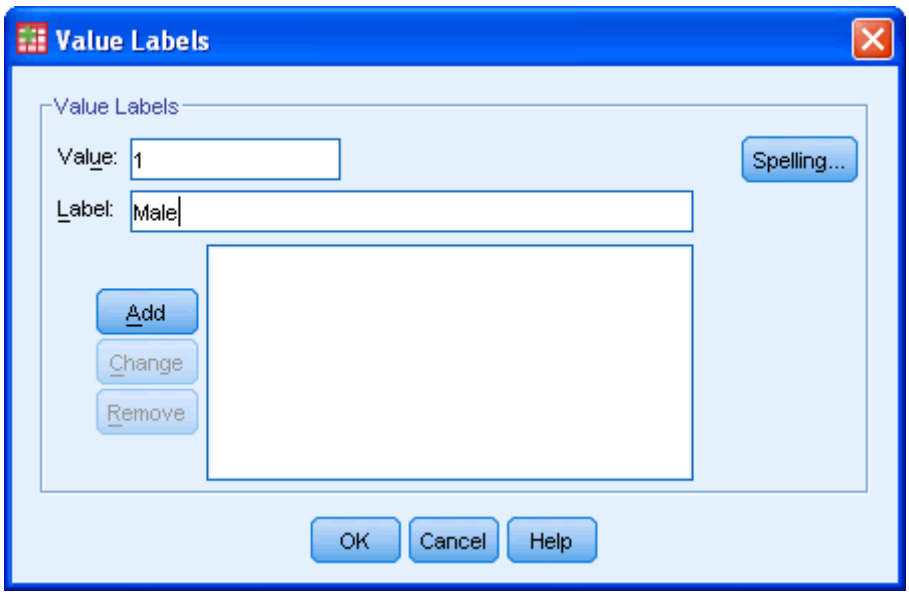
Label Click on the appropriate cell in the Labels column to give a long description which will help you to understand the variable you are referring to.

Values You can give a text label for category codes by clicking on the appropriate cell in the column values. Default is None.

First click on the "None" cell box and then on the ... button as show in the diagram below:



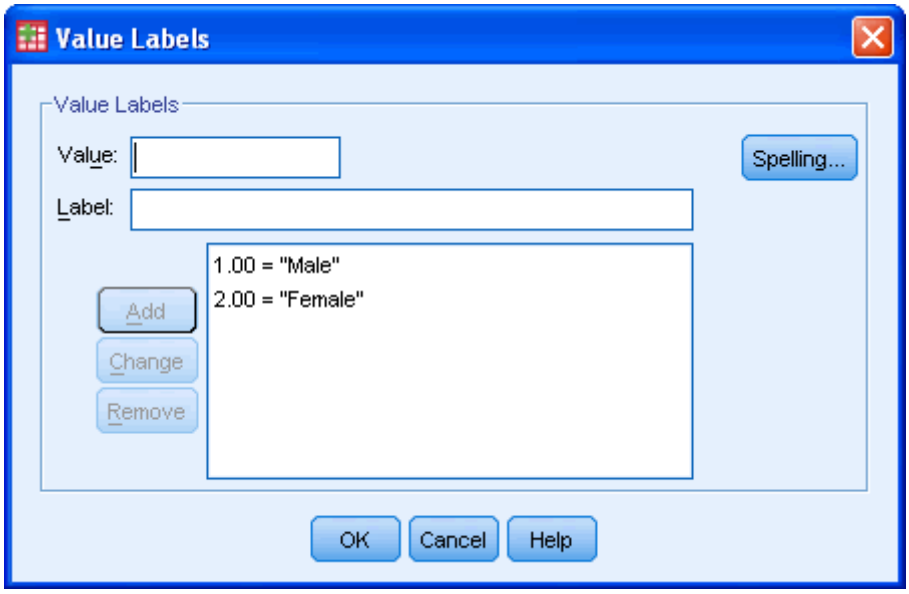
Enter the numeric code value 1 in the text box for Value:. This value corresponds to the group representing "males", so enter "males" in the *Label:* box.



Click on the Add button.

Repeat the above entering "2" in the Value: box and "female" in the *Label:* box and click on the Add button.

You will be presented with the following screen:



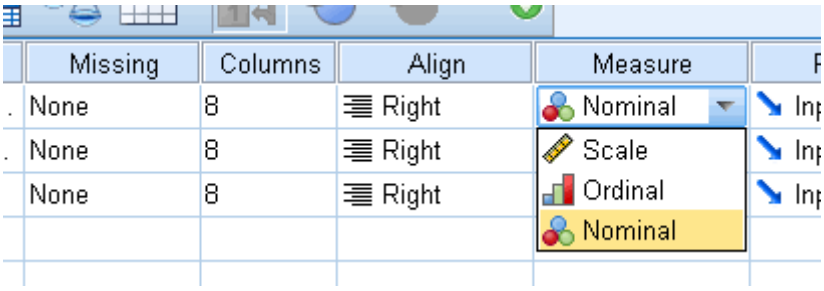
Click on the OK button.

Missing To define different type of missing vale code click on the appropriate cell in the Missing column.

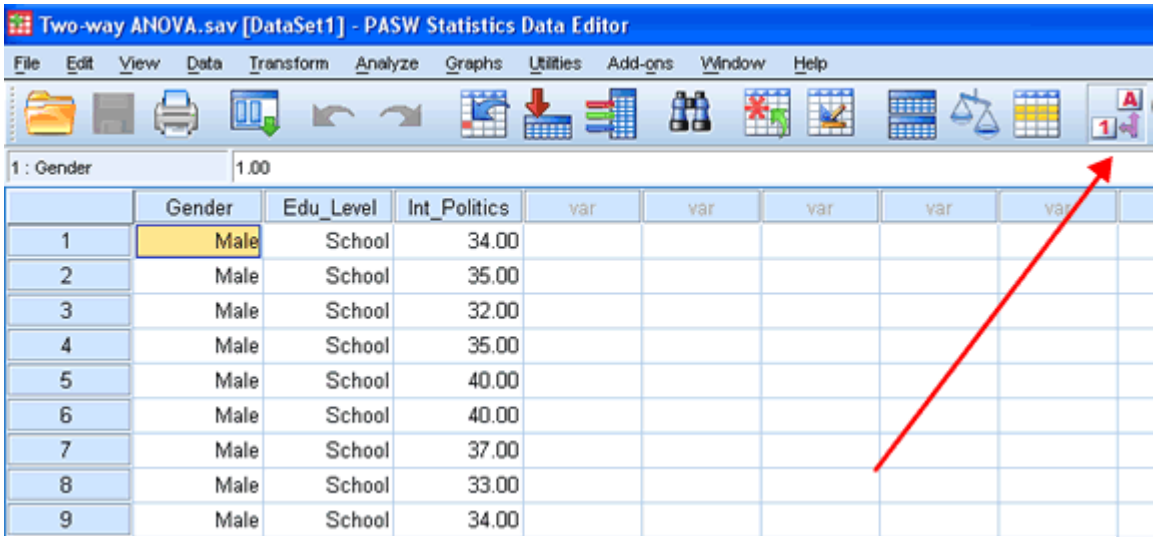
Columns The width of column in the data view sheet can be altered by clicking in the appropriate Column cell.

Align To align the text or numbers in Data View, click in the appropriate Align cell.

Measure To label the scale of data, click in the appropriate Measure cell. In our example, we need to select the "Nominal" value.



You can now see the results of your labeling by clicking the button in the Data Editor View as shown below:



Entering Data in SPSS

The "one person, one row" Rule

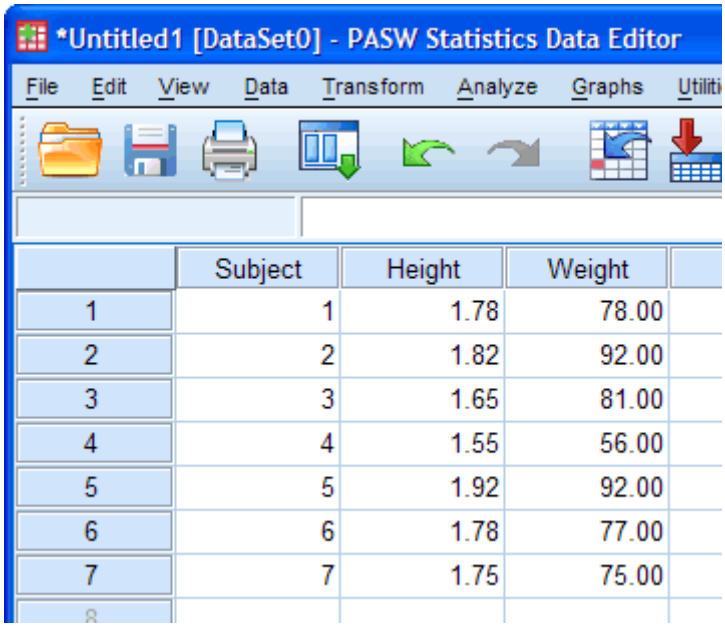
SPSS sets out its data in a spreadsheet-like manner. The principle behind entering data in almost all cases in SPSS is to enter each unique case on a new row. A case is the "object" which you are measuring in someway. Usually, a case is an individual, but it can also be a commercial product or a biological cell or something else entirely. For the purposes of this explanation, we shall assume that a case is an individual. Therefore, when entering data into SPSS you must put one person's data on one row only. If you find that you have an individual's data on more than one row then you have made a mistake. Equally, if a row contains more than one person's data, then you have also made a mistake.

We shall now look at the three most common tasks you face when entering data into SPSS plus two more advanced setups:

- Entering variables, e.g. height, weight
- Defining separate groups (between-subject factors), e.g. gender, level of education
- Entering repeated measures (within-subjects factors), e.g. time course
- Multiple separate groups, e.g. gender and level of education
- Separate groups and repeated measures, e.g. gender and time course.

Entering Variables

If you do not have repeated measures then SPSS treats each column as a separate variable. Thus, each variable goes in a separate column. For example, if we had measured the height and weight of a group of individuals then the data in SPSS would look like the following:

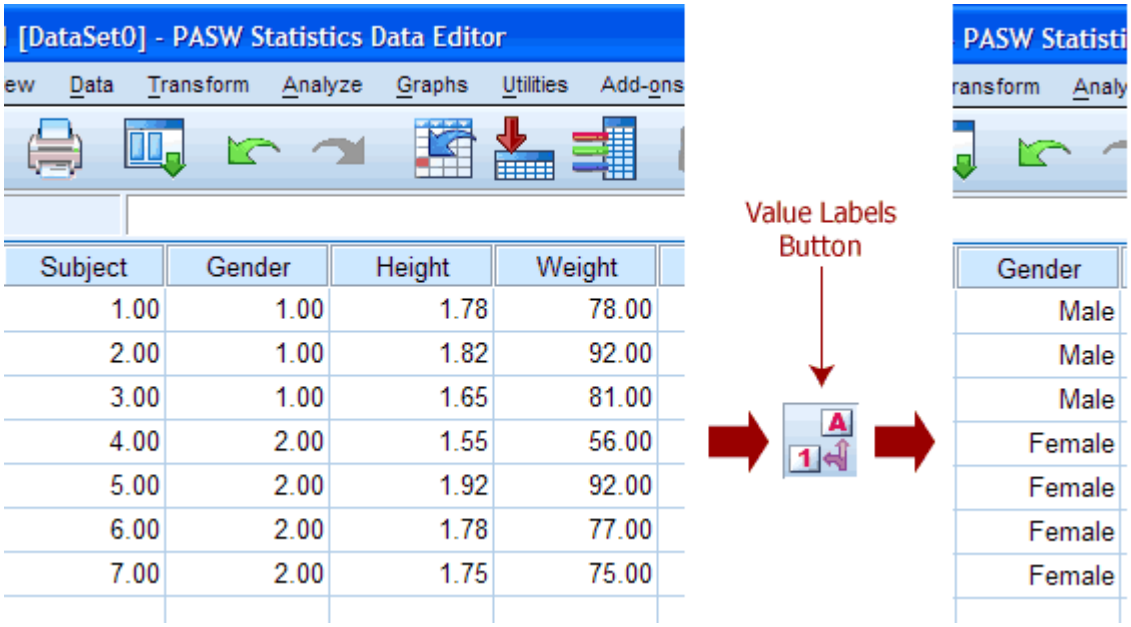


	Subject	Height	Weight
1	1	1.78	78.00
2	2	1.82	92.00
3	3	1.65	81.00
4	4	1.55	56.00
5	5	1.92	92.00
6	6	1.78	77.00
7	7	1.75	75.00

The **Subject** column has been added so that it is clear that each individual is placed on a separate row; however, SPSS does not need you to enter this column and it is mostly for you to be able to better visualize your data. So, even if we ignored the **Subject** column, we can see that one individual was 1.55 m tall and weighed 56 kg, looking at the **Height** and **Weight** columns, respectively. How to label variable columns is in our Working with Variables guide. To add more variables, simply add more columns - one column per variable. The only variation to this is discussed later in this guide when we have to enter repeated measures.

Defining Separate Groups

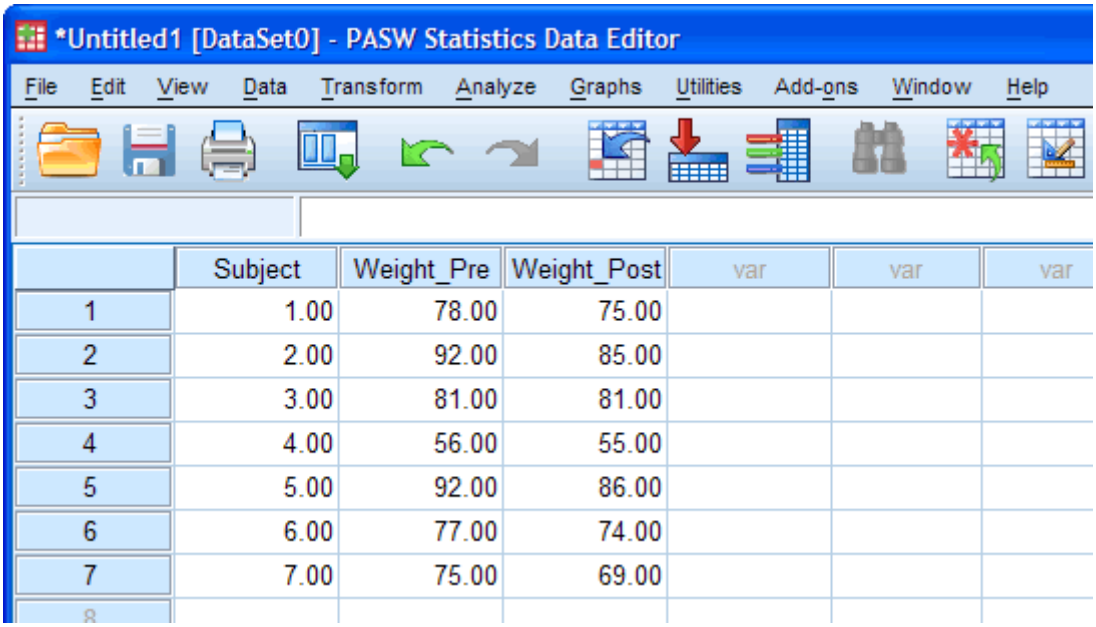
Separate groups are more commonly called between-subjects factors or independent groups. They are groups where the individuals in each group are unique, i.e. no person is in more than one group. In this sense, you could call the groups "mutually-exclusive". A common example is when differentiating between genders. You want to label some of your individuals as female and others as male. To identify which subjects were males and which were females, you need to create a "grouping variable" in SPSS. This is a separate column which includes information on which group a subject belongs to. We do this by labeling our groups numerically. So, for example, we label males as "**1**" and females as "**2**". By using the value attribute we can label these numbers as representing males and females, respectively. An example is shown below:



Looking at the columns on the left we can see that we have created a "grouping variable" called Gender that has two categories - "**1**" and "**2**". Because we labelled the numbers using the value attribute we can use the Value Label Button to switch to the text version of the "grouping variable" categories. In this example, we can see that "1" and "2" are replaced by "**Male**" and "**Female**", respectively. How to do this is explained in our guide on Working with Variables. You do not need to add text labels – SPSS will work fine without them – but it can provide extra clarity when analysing your data (especially as text labels are often used in the output instead of the numbers – this helps greatly). We can see in this example that the first three subjects were males and the last four subjects were females. What if you have more than two categories of your "grouping variable"? Simple, just add more numbers with, we recommend, corresponding text labels.

Entering Repeated Measures

Repeated measures, also called within-subject factors or related groups, are variables that are measured on more than one occasion. This can occur when you have measured the same subject for the same variable at more than one time point or under more than one condition. For example, you measured body weight at the beginning and end of a weight-loss programme. To enter this into SPSS you must ignore the "one-variable-one-column" rule and put each time point or condition in a new column as follows:

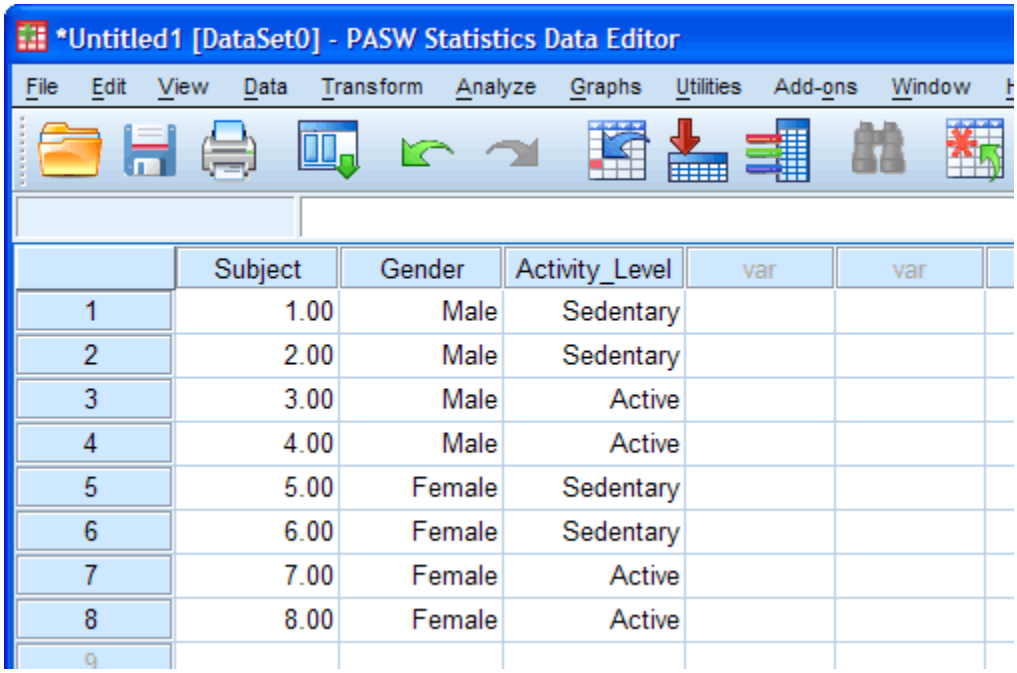


	Subject	Weight_Pre	Weight_Post	var	var	var
1	1.00	78.00	75.00			
2	2.00	92.00	85.00			
3	3.00	81.00	81.00			
4	4.00	56.00	55.00			
5	5.00	92.00	86.00			
6	6.00	77.00	74.00			
7	7.00	75.00	69.00			

Here we have labeled their weight at the beginning of the weight-loss programme as "**Weight_Pre**" and their weight after the weight-loss programme as "**Weight_Post**". It does not matter what you call these "related" columns (you could have called them weight1 and weight2, for example) as long the columns make sense to you. If you have a lot of time points, and/or conditions, then labeling the variables logically is important as otherwise it can become very confusing determining which variable is which. This is important as SPSS itself cannot tell the difference between columns that contain different variables and columns that contain a repeated variable and, therefore, cannot help you.

Multiple Separating Groups

Sometimes, such as when running a two-way ANOVA or when entering in your whole study data, you need to separate your subjects twice - i.e on two separate variables. For example, you need to separate subjects by their gender (male/female) and their physical activity level (sedentary/active). This will require two columns that act as "grouping variables" as shown below:

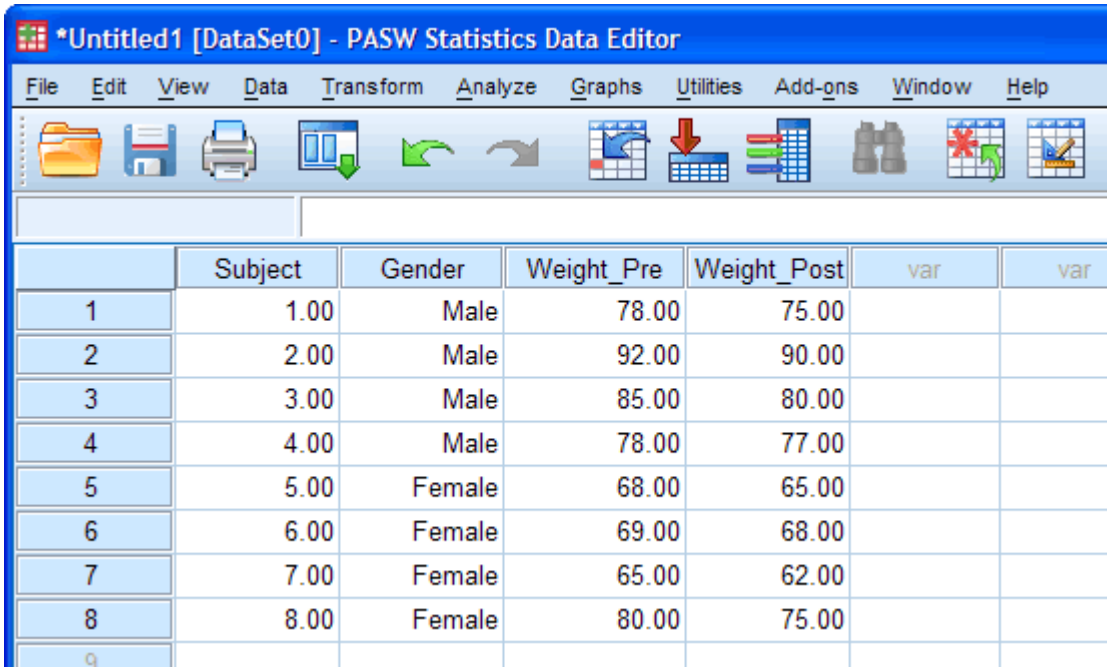


	Subject	Gender	Activity_Level	var	var
1	1.00	Male	Sedentary		
2	2.00	Male	Sedentary		
3	3.00	Male	Active		
4	4.00	Male	Active		
5	5.00	Female	Sedentary		
6	6.00	Female	Sedentary		
7	7.00	Female	Active		
8	8.00	Female	Active		

Here we can see that, for example, Subject 1 was male and sedentary and that Subject 7 was female and active. Notice that we are using the text labels as described earlier in this guide for added clarity.

Mixing Separate Groups and Repeated Measures

Sometimes, we have separated subjects into groups and then measured them repeatedly on the same dependent variable. Such data might be analysed using a mixed ANOVA. If we had males and females undertake a weight-loss programme and we weighted them pre- and post-intervention then we would have the following setup in SPSS:

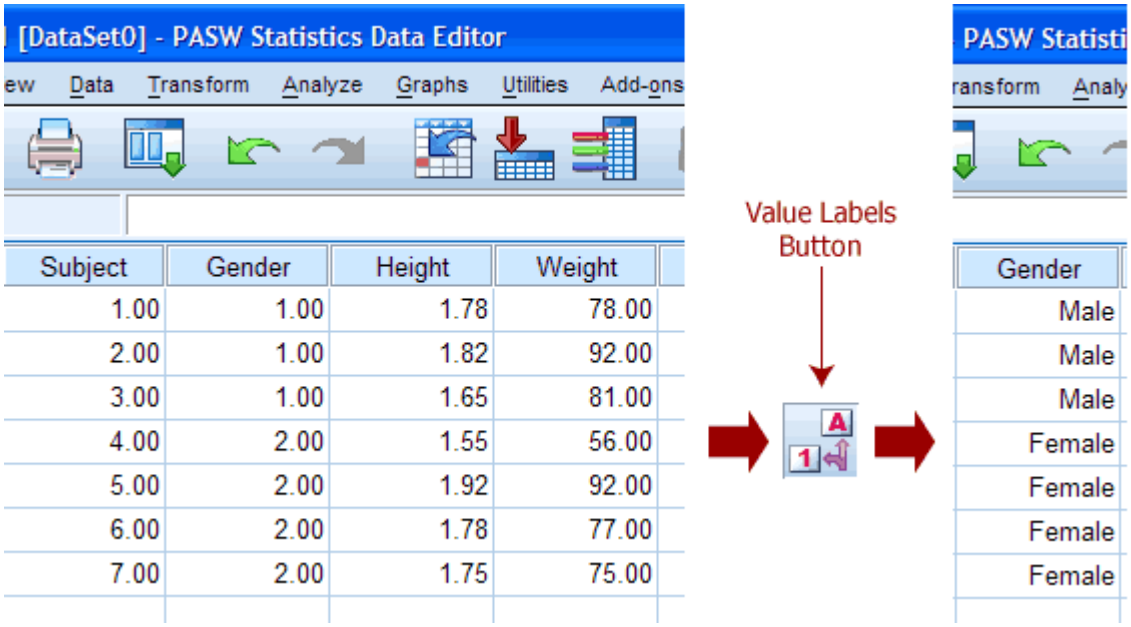


	Subject	Gender	Weight_Pre	Weight_Post	var	var
1	1.00	Male	78.00	75.00		
2	2.00	Male	92.00	90.00		
3	3.00	Male	85.00	80.00		
4	4.00	Male	78.00	77.00		
5	5.00	Female	68.00	65.00		
6	6.00	Female	69.00	68.00		
7	7.00	Female	65.00	62.00		
8	8.00	Female	80.00	75.00		

To generate this type of setup simply used the rules you have learnt in this guide under the Defining Separate Groups and Entering Repeated Measures sections.

Defining Separate Groups

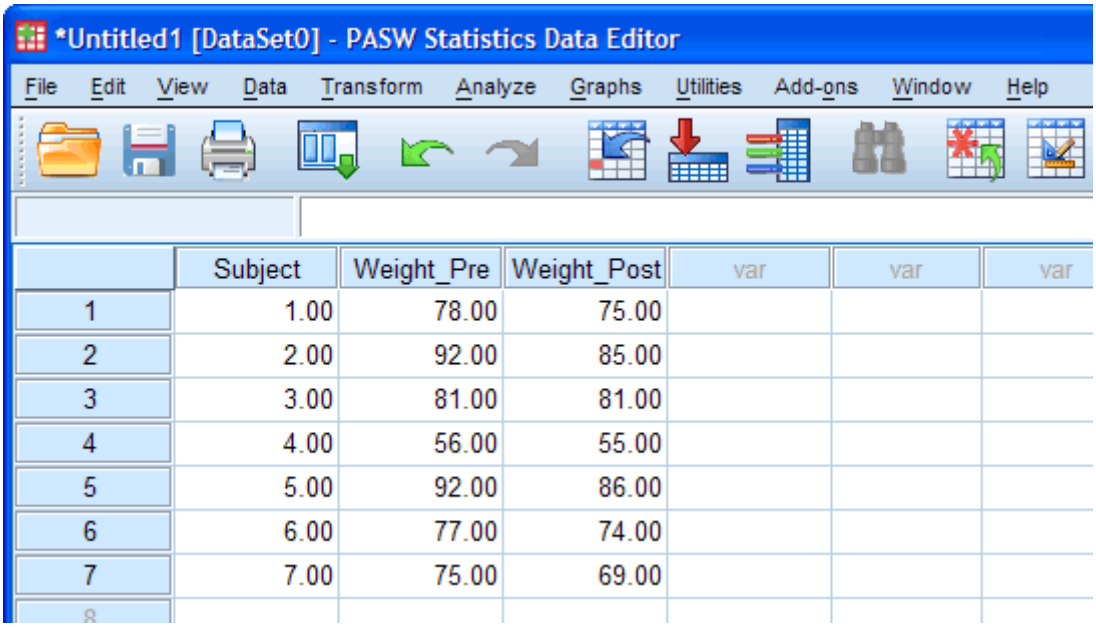
Separate groups are more commonly called between-subjects factors or independent groups. They are groups where the individuals in each group are unique, i.e. no person is in more than one group. In this sense, you could call the groups "mutually-exclusive". A common example is when differentiating between genders. You want to label some of your individuals as female and others as male. To identify which subjects were males and which were females, you need to create a "grouping variable" in SPSS. This is a separate column which includes information on which group a subject belongs to. We do this by labelling our groups numerically. So, for example, we label males as "**1**" and females as "**2**". By using the value attribute we can label these numbers as representing males and females, respectively. An example is shown below:



Looking at the columns on the left we can see that we have created a "grouping variable" called Gender that has two categories - "**1**" and "**2**". Because we labelled the numbers using the value attribute we can use the Value Label Button to switch to the text version of the "grouping variable" categories. In this example, we can see that "1" and "2" are replaced by "**Male**" and "**Female**", respectively. How to do this is explained in our guide on Working with Variables. You do not need to add text labels – SPSS will work fine without them – but it can provide extra clarity when analysing your data (especially as text labels are often used in the output instead of the numbers – this helps greatly). We can see in this example that the first three subjects were males and the last four subjects were females. What if you have more than two categories of your "grouping variable"? Simple, just add more numbers with, we recommend, corresponding text labels.

Entering Repeated Measures

Repeated measures, also called within-subject factors or related groups, are variables that are measured on more than one occasion. This can occur when you have measured the same subject for the same variable at more than one time point or under more than one condition. For example, you measured body weight at the beginning and end of a weight-loss programme. To enter this into SPSS you must ignore the "one-variable-one-column" rule and put each time point or condition in a new column as follows:

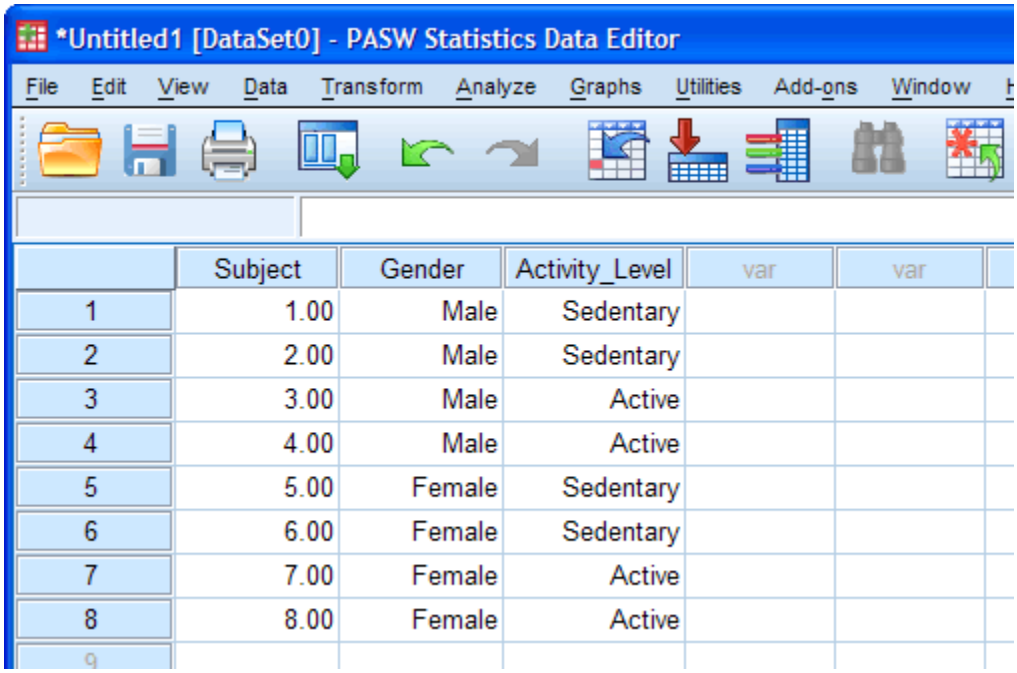


	Subject	Weight_Pre	Weight_Post	var	var	var
1	1.00	78.00	75.00			
2	2.00	92.00	85.00			
3	3.00	81.00	81.00			
4	4.00	56.00	55.00			
5	5.00	92.00	86.00			
6	6.00	77.00	74.00			
7	7.00	75.00	69.00			

Here we have labelled their weight at the beginning of the weight-loss programme as "**Weight_Pre**" and their weight after the weight-loss programme as "**Weight_Post**". It does not matter what you call these "related" columns (you could have called them weight1 and weight2, for example) as long the columns make sense to you. If you have a lot of time points, and/or conditions, then labelling the variables logically is important as otherwise it can become very confusing determining which variable is which. This is important as SPSS itself cannot tell the difference between columns that contain different variables and columns that contain a repeated variable and, therefore, cannot help you.

Multiple Separating Groups

Sometimes, such as when running a two-way ANOVA or when entering in your whole study data, you need to separate your subjects twice - i.e on two separate variables. For example, you need to separate subjects by their gender (male/female) and their physical activity level (sedentary/active). This will require two columns that act as "grouping variables" as shown below:

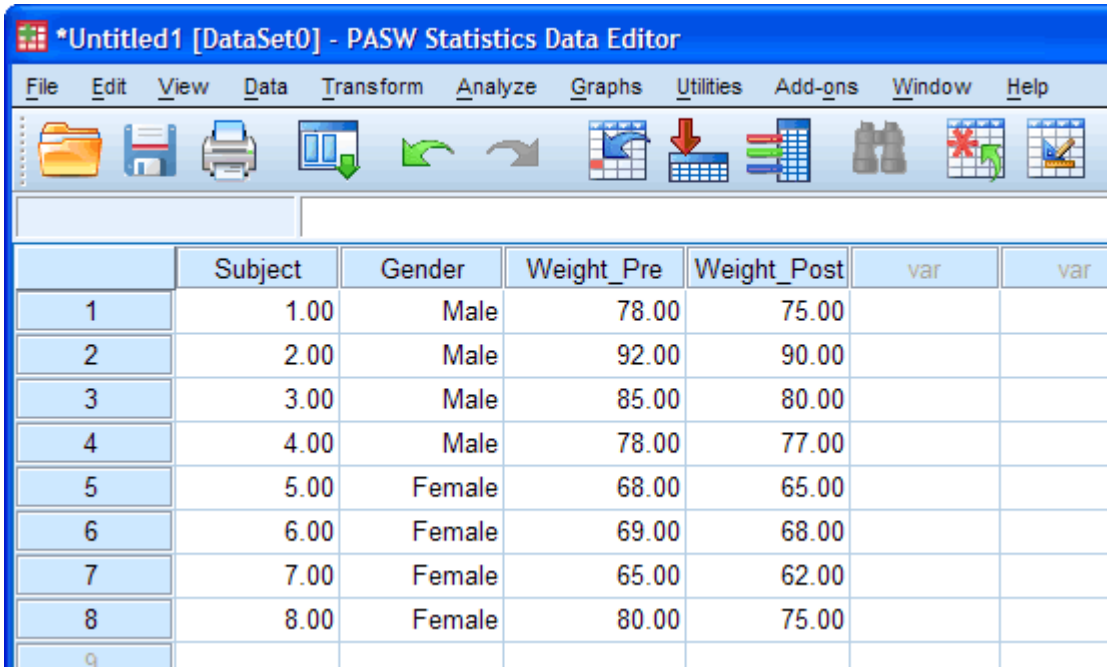


	Subject	Gender	Activity_Level	var	var
1	1.00	Male	Sedentary		
2	2.00	Male	Sedentary		
3	3.00	Male	Active		
4	4.00	Male	Active		
5	5.00	Female	Sedentary		
6	6.00	Female	Sedentary		
7	7.00	Female	Active		
8	8.00	Female	Active		

Here we can see that, for example, Subject 1 was male and sedentary and that Subject 7 was female and active. Notice that we are using the text labels as described earlier in this guide for added clarity.

Mixing Separate Groups and Repeated Measures

Sometimes, we have separated subjects into groups and then measured them repeatedly on the same dependent variable. Such data might be analyzed using a mixed ANOVA. If we had males and females undertake a weight-loss programme and we weighted them pre- and post-intervention then we would have the following setup in SPSS:

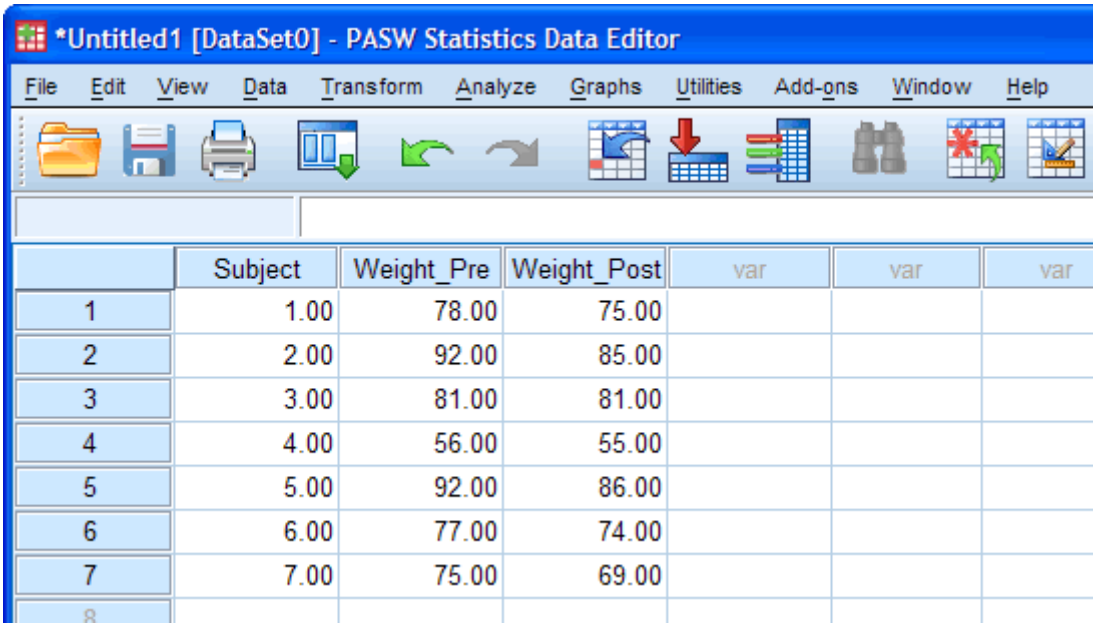


	Subject	Gender	Weight_Pre	Weight_Post	var	var
1	1.00	Male	78.00	75.00		
2	2.00	Male	92.00	90.00		
3	3.00	Male	85.00	80.00		
4	4.00	Male	78.00	77.00		
5	5.00	Female	68.00	65.00		
6	6.00	Female	69.00	68.00		
7	7.00	Female	65.00	62.00		
8	8.00	Female	80.00	75.00		

To generate this type of setup simply used the rules you have learnt in this guide under the Defining Separate Groups and Entering Repeated Measures sections.

Entering Repeated Measures

Repeated measures, also called within-subject factors or related groups, are variables that are measured on more than one occasion. This can occur when you have measured the same subject for the same variable at more than one time point or under more than one condition. For example, you measured body weight at the beginning and end of a weight-loss programme. To enter this into SPSS you must ignore the "one-variable-one-column" rule and put each time point or condition in a new column as follows:

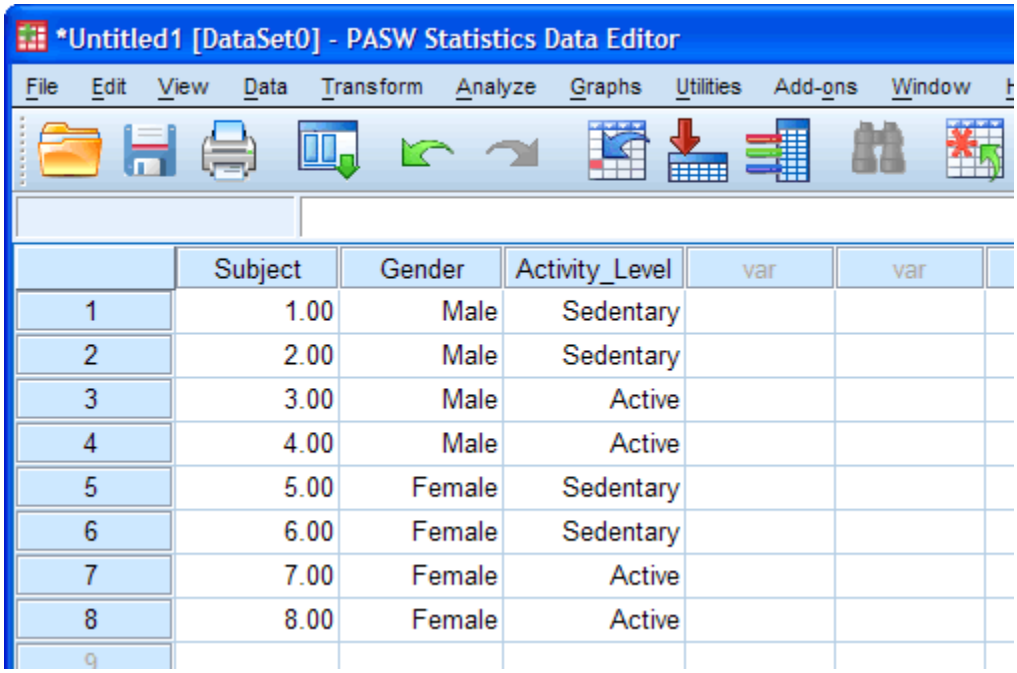


	Subject	Weight_Pre	Weight_Post	var	var	var
1	1.00	78.00	75.00			
2	2.00	92.00	85.00			
3	3.00	81.00	81.00			
4	4.00	56.00	55.00			
5	5.00	92.00	86.00			
6	6.00	77.00	74.00			
7	7.00	75.00	69.00			

Here we have labelled their weight at the beginning of the weight-loss programme as "**Weight_Pre**" and their weight after the weight-loss programme as "**Weight_Post**". It does not matter what you call these "related" columns (you could have called them weight1 and weight2, for example) as long the columns make sense to you. If you have a lot of time points, and/or conditions, then labelling the variables logically is important as otherwise it can become very confusing determining which variable is which. This is important as SPSS itself cannot tell the difference between columns that contain different variables and columns that contain a repeated variable and, therefore, cannot help you.

Multiple Separating Groups

Sometimes, such as when running a two-way ANOVA or when entering in your whole study data, you need to separate your subjects twice - i.e on two separate variables. For example, you need to separate subjects by their gender (male/female) and their physical activity level (sedentary/active). This will require two columns that act as "grouping variables" as shown below:

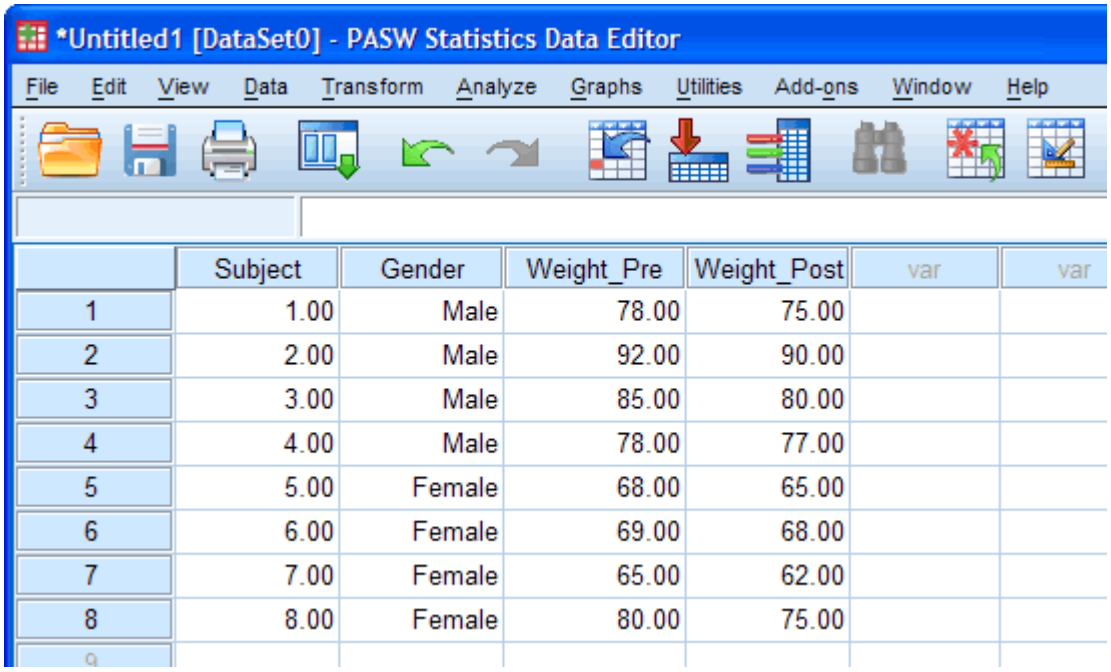


	Subject	Gender	Activity_Level	var	var
1	1.00	Male	Sedentary		
2	2.00	Male	Sedentary		
3	3.00	Male	Active		
4	4.00	Male	Active		
5	5.00	Female	Sedentary		
6	6.00	Female	Sedentary		
7	7.00	Female	Active		
8	8.00	Female	Active		

Here we can see that, for example, Subject 1 was male and sedentary and that Subject 7 was female and active. Notice that we are using the text labels as described earlier in this guide for added clarity.

Mixing Separate Groups and Repeated Measures

Sometimes, we have separated subjects into groups and then measured them repeatedly on the same dependent variable. Such data might be analysed using a mixed ANOVA. If we had males and females undertake a weight-loss programme and we weighted them pre- and post-intervention then we would have the following setup in SPSS:



	Subject	Gender	Weight_Pre	Weight_Post	var	var
1	1.00	Male	78.00	75.00		
2	2.00	Male	92.00	90.00		
3	3.00	Male	85.00	80.00		
4	4.00	Male	78.00	77.00		
5	5.00	Female	68.00	65.00		
6	6.00	Female	69.00	68.00		
7	7.00	Female	65.00	62.00		
8	8.00	Female	80.00	75.00		

To generate this type of setup simply used the rules you have learnt in this guide under the Defining Separate Groups and Entering Repeated Measures sections. 1

Transforming Data in SPSS

8

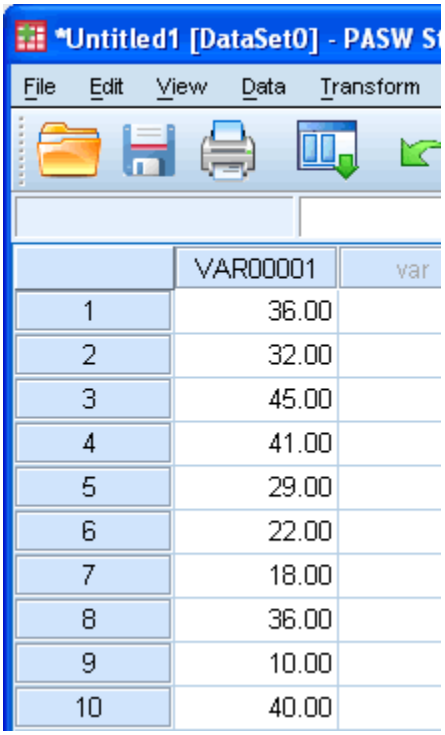
In this guide we will enter some data and then perform a transformation of the data. Transforming data is performed for a whole host of different reasons but one of the most common is to apply a transformation to data that is not normally distributed so that the new, transformed data is normally distributed. Transforming a non-normal distribution into a normal distribution is performed in a number of different ways depending on the original distribution of data but a common technique is to take the log of the data. In this example we will show you how SPSS allows you to do this.

There are an infinite possible ways to transform data, although there are some approaches that are much more common than others. We will add extra examples of transforming data in the near future.

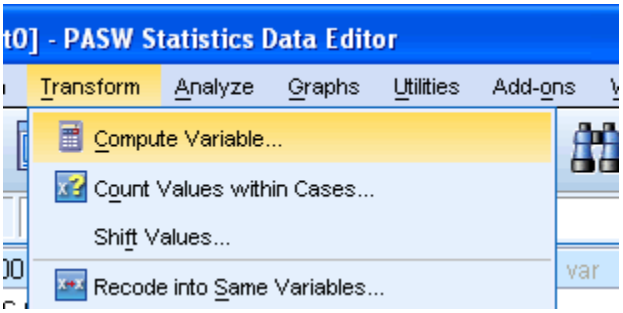
Data to Transform									
36	32	45	41	29	22	18	36	10	40

1. Your data should end up looking like the following:

   Remember that each individual's results go on a separate line (row) in SPSS.



2. Rename the variable "**Data**" instead of the default "**VAR00001**". *You can name it something else if you wish.*
3. Click on **Transform > Compute Variable...** in the top menu.

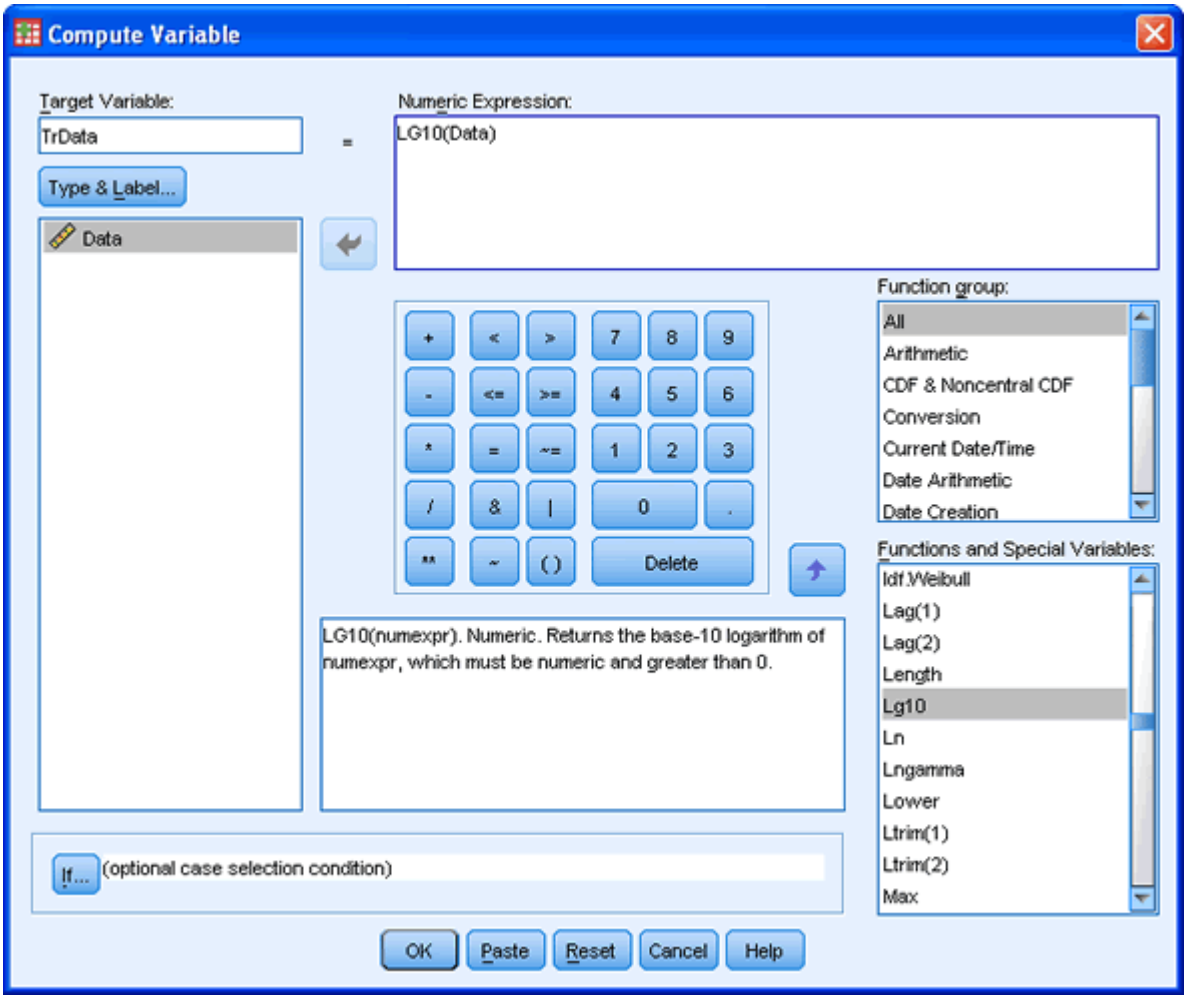


4. You need to first select the function you would like to use. To do this click "**All**" in the "*Function group:*" box then scroll down in the "*Functions and Special Variables:*" box and click "**Lg10**" to select it.

   Transfer the Lg10 function into the "*Numeric Expression:*" box by pressing the button.

   Click the Data variable in the lefthand box and then click the button which will result in the expression you see in the "*Numeric Expression:*" box below.

   All to do now is to give this new variable a name. We have called the new variable "**TrData**". Type this name into the "*Target Variable:*" box in the top lefthand corner.



5. Click on the OK button.
6. You will be presented with the SPSS Data Editor which will now show the log transformed data under the new variable name "**TrData**" that you defined.

Recoding Variables in SPSS

The instructions below will show you how to recode variables in SPSS 14.0. You can use recoding to produce different values or codes for a variable. Recoding can be done in one of two ways:

- Recoding into the same variable
- Recoding into a different variable

In this guide we will concentrate on recoding into a different variable for which there are 3 main types of recoding:

- Recode single values

- Recode a given range of values
- Recode data into two categories

We will use an example of each type in order to demonstrate how to recode variables in SPSS.

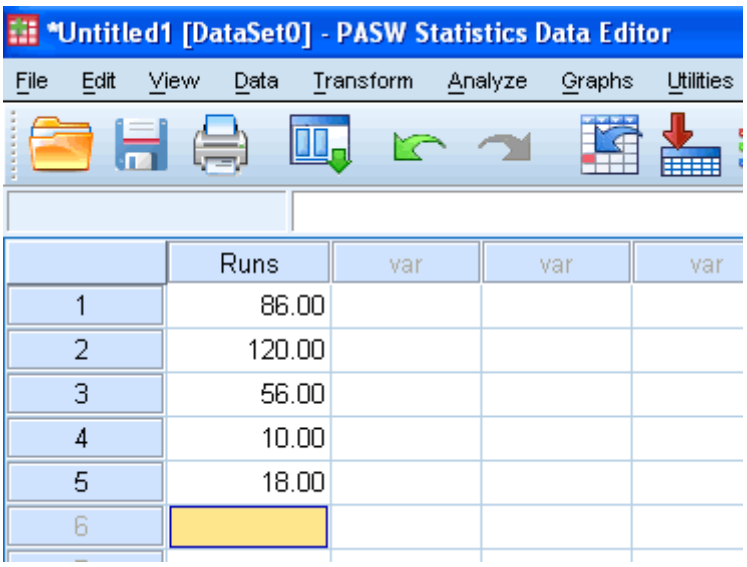
Recode single values

Example: The data given below represents runs scored by 5 batsmen in a national-level match. Recode the data so that the batsmen are rank ordered by their number of runs, with the batsman with the highest runs given a code of "1" and the batsman with the lowest runs given a "5".

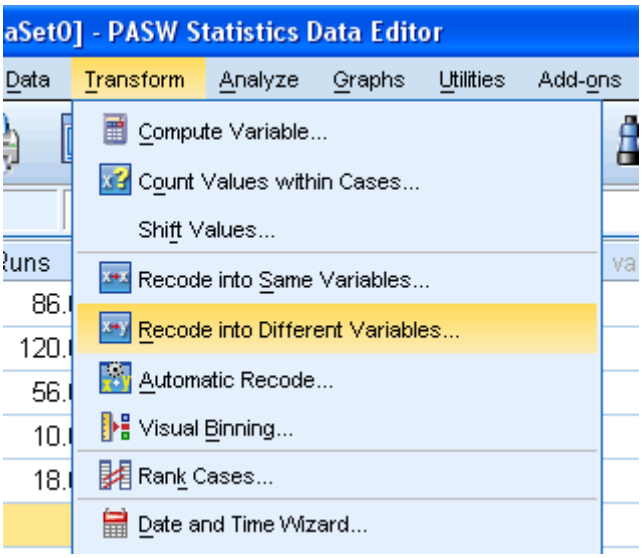
Number of runs by batsmen					
Batsmen	1	2	3	4	5
Runs	86	120	56	10	18

1. Enter the data in the SPSS Data Editor and name the variable "**Runs**".

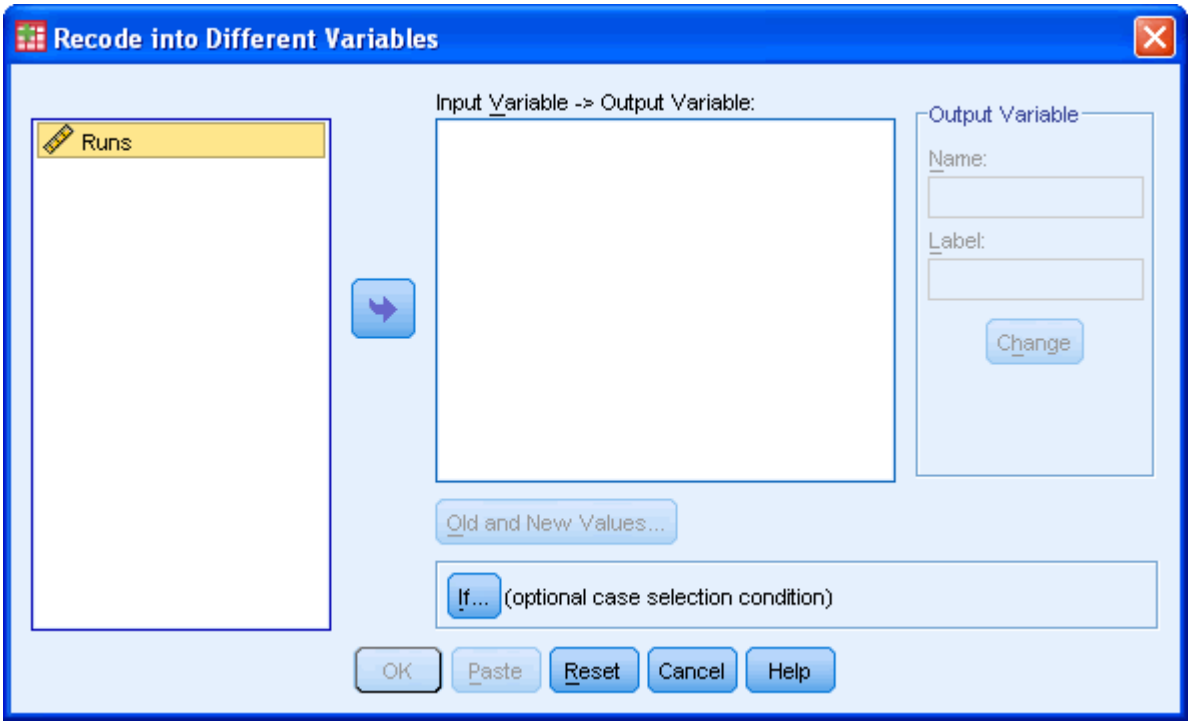
   Remember that each individual's results go on a separate line (row) in SPSS.



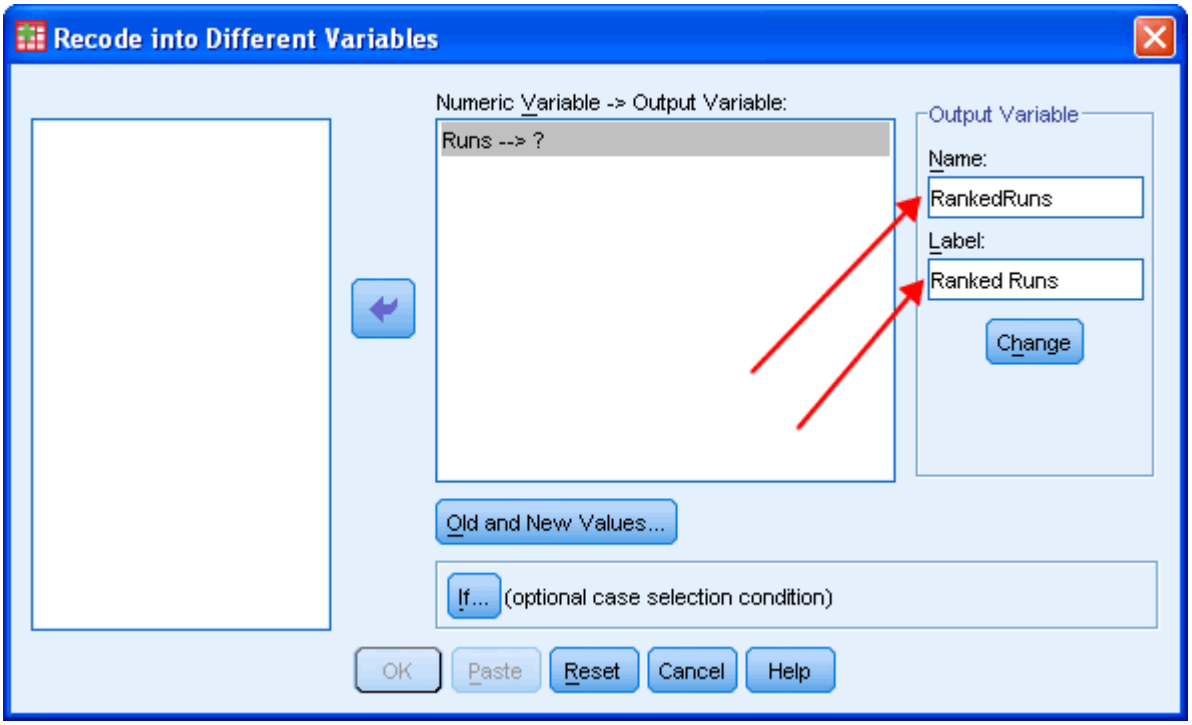
2. Click on **Transform > Recode Into Different Variables...** in the top menu.



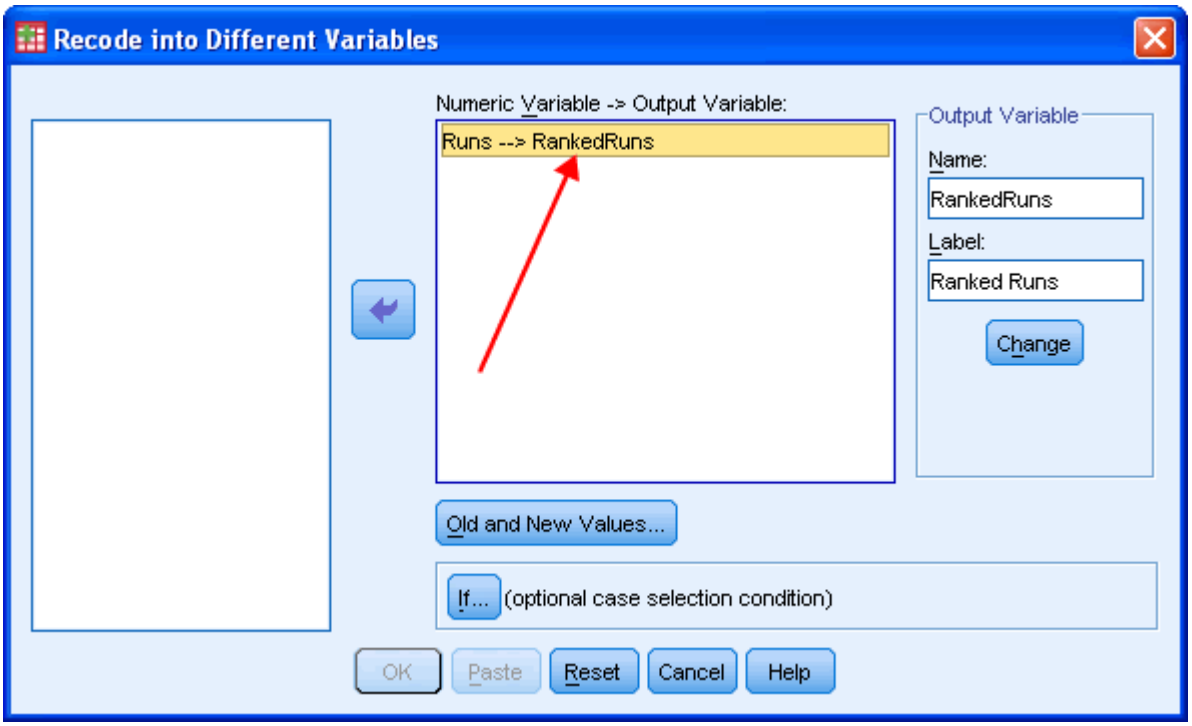
3. You need to highlight the "**Runs**" variable in the left-hand box by clicking on it and then click the button to move it across to the "*Numeric Variable -> Output Variable:*" box.



4. In the Output Variable grouping show below, give the new variable you are about to create a name and label. In this example, we have named the new variable "**Ranked-Runs**" and the label as "**Ranked Runs**" (see the).

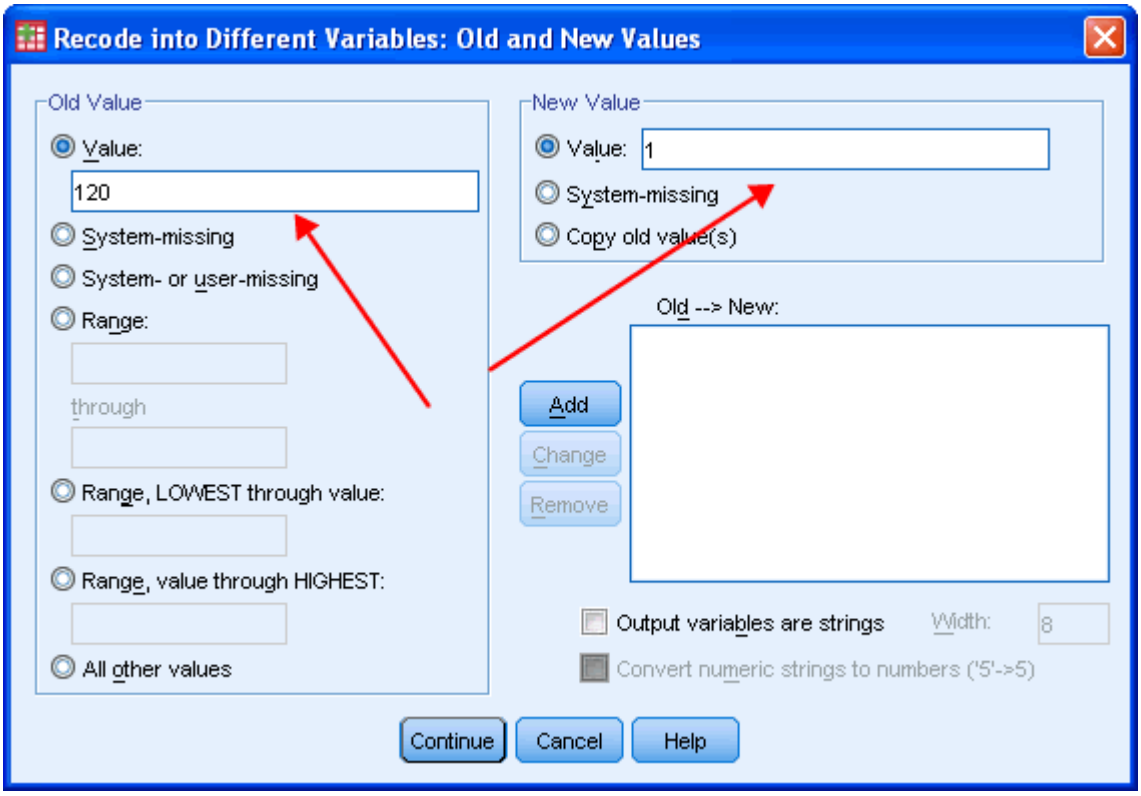


5. Click the Change button.
6. You should now see that the "*Numeric Variable -> Output Variable:*" box has now changed to reflect these changes (see the).

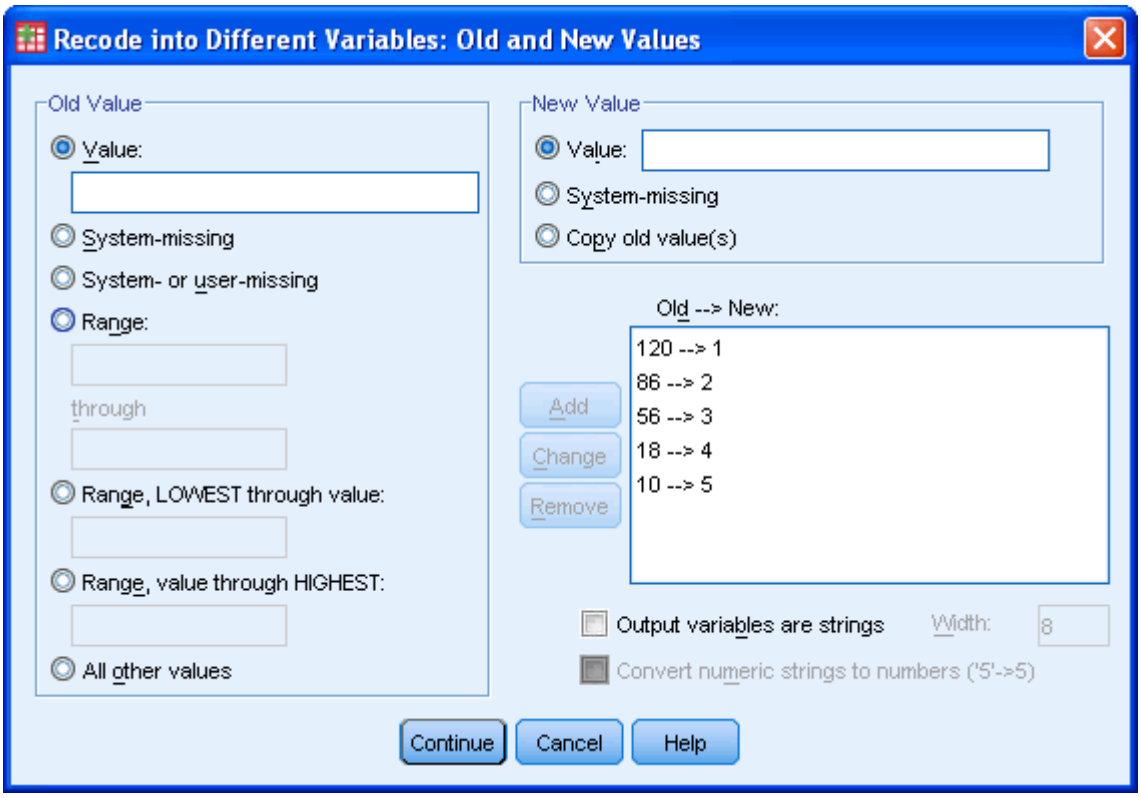


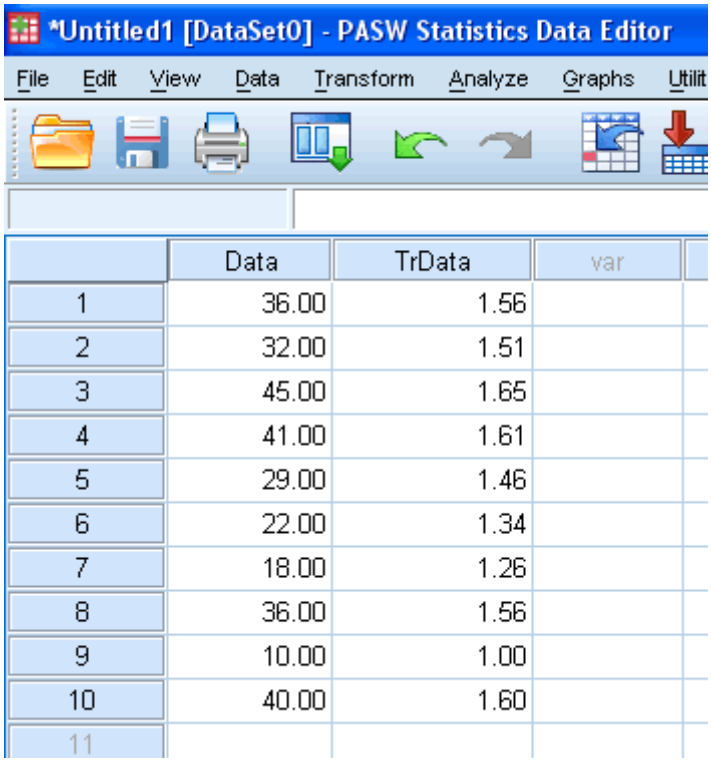
7. Click the Old and New Values... button.
8. Enter the first score (120) into the "*Old Value*" box and set the new code of "1" into the "*New Value*".
9. Click the Add button.

10. Repeat for all other values such that you are presented with the following screen:



11. Click the Continue button.
12. You will have returned to the previous screen and will now have to click the OK button. This will take you to the Data View Editor which will now show that you have created a new variable that is based on recoding your existing variable:





	Data	TrData	var
1	36.00	1.56	
2	32.00	1.51	
3	45.00	1.65	
4	41.00	1.61	
5	29.00	1.46	
6	22.00	1.34	
7	18.00	1.26	
8	36.00	1.56	
9	10.00	1.00	
10	40.00	1.60	
11			

Recoding Variables in SPSS (cont...)

Recode a given range

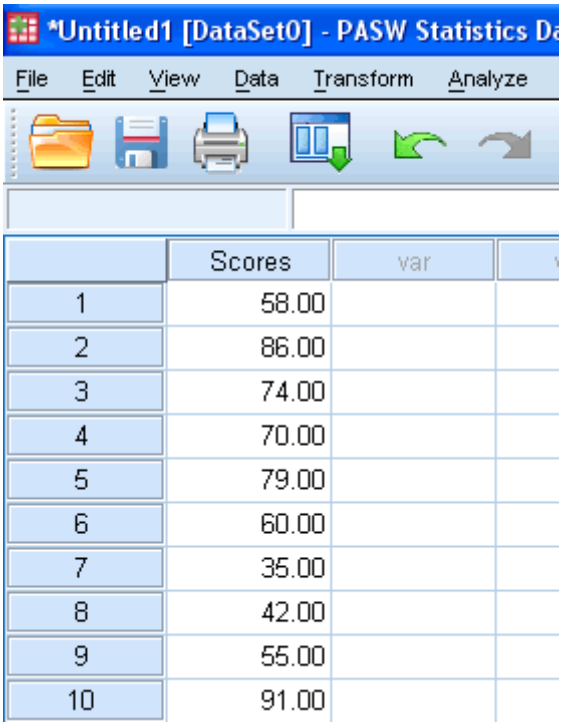
Example: The data given below represents the scores of 10 students in a final examination. Recode the data giving code "1" to scores between 75 - 100, code 2 to scores between 61 - 75, code 3 to scores between 41 - 60 and code 4 to scores between 0 - 40.

Final examination scores of 10 students

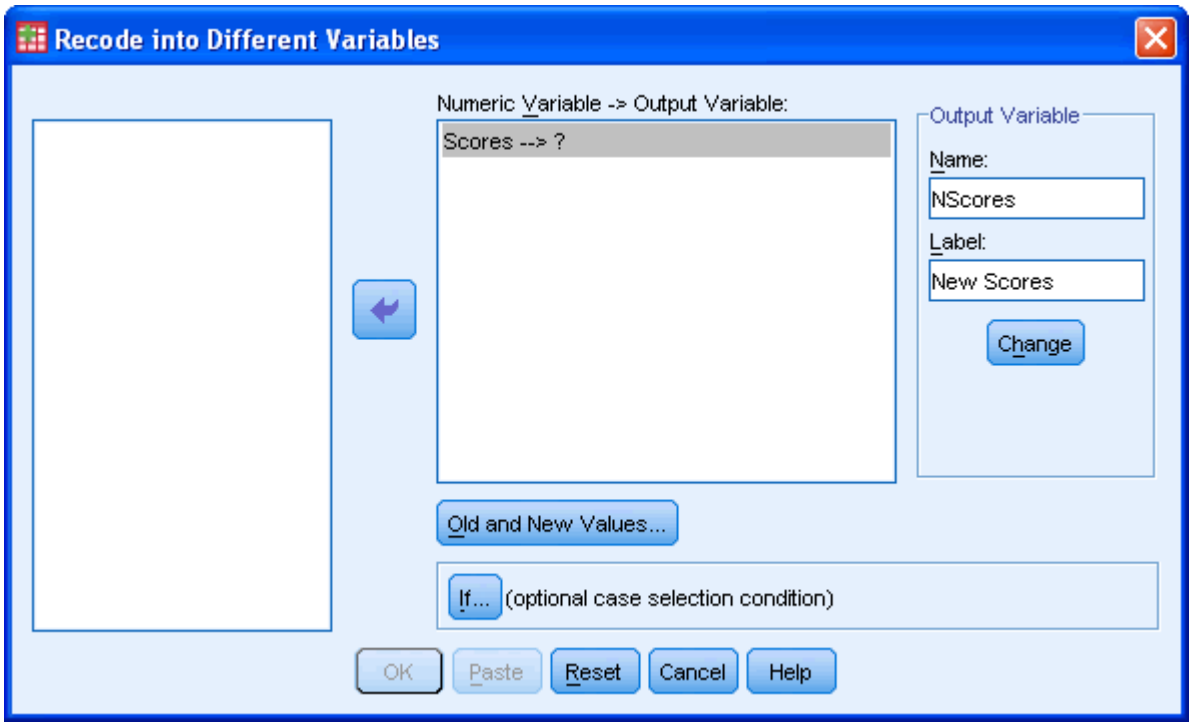
Scores	58	86	74	70	79	60	35	42	55	91

1. Enter the data in the SPSS Data Editor and name the variable "**Scores**".

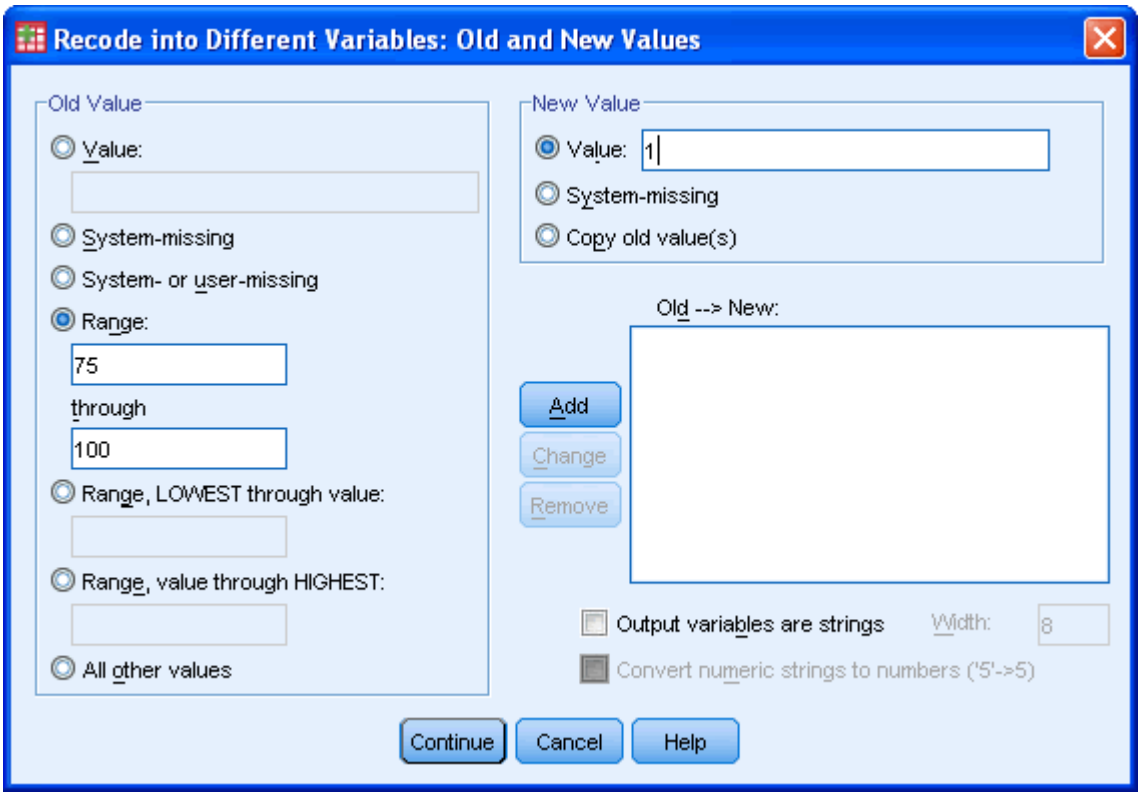
   Remember that each individual's results go on a separate line (row) in SPSS.



2. Click on **Transform > Recode Into Different Variable...** in the top menu.
3. Transfer the variable you want to recode by selected it and pressing the button and give the new variable a name and label. In this example, we have given the new variable a name of "**NScores**" and label of "**New Scores**" as shown below:

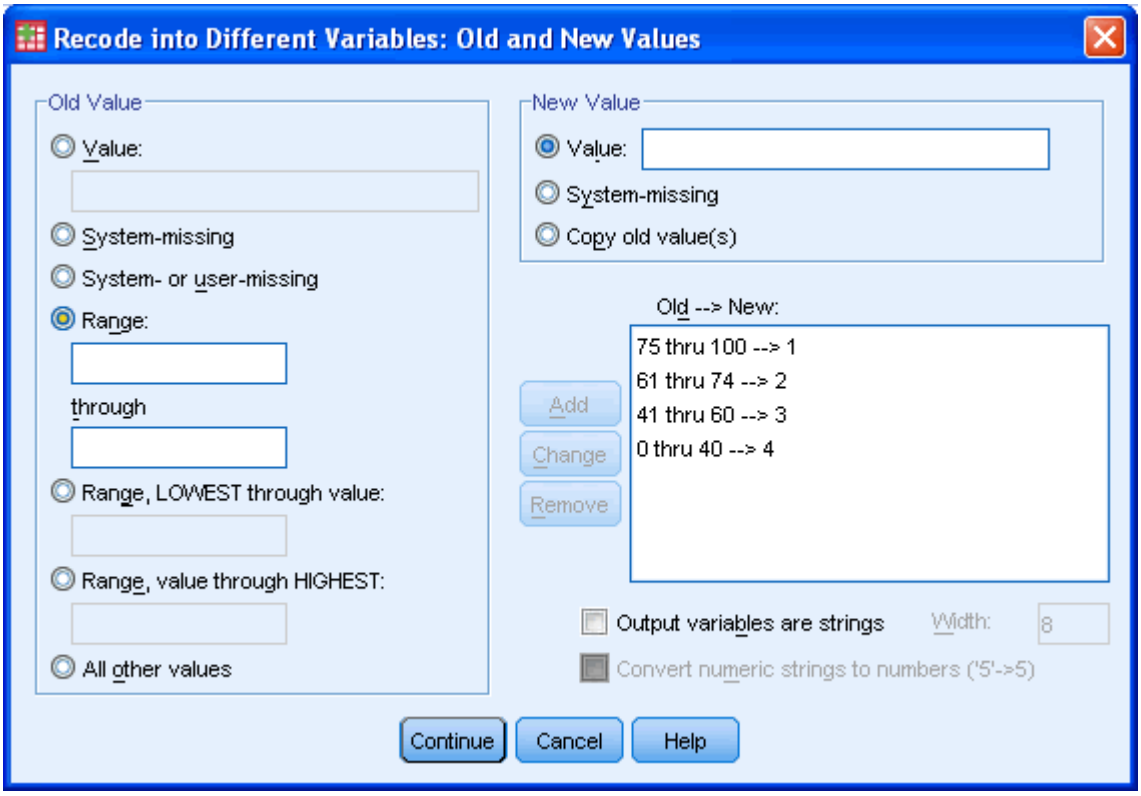


4. Click on the Change button.
5. Click on the Old and New Values... button.
6. Enter the first range of "**75 - 100**" into the "*Old Value*" box and set the new code to "**1**" into the "*New Value*" box as shown below:

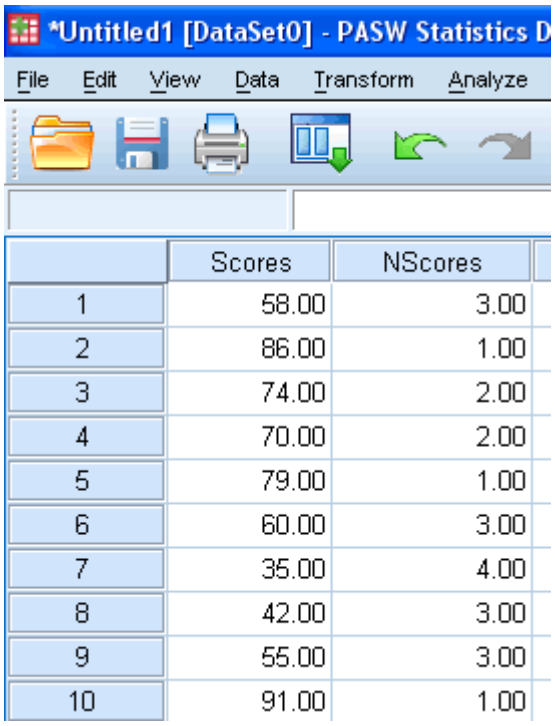


7. Click on the Add button.

8. Repeat for all other values such that you are presented with the following screen:



9. Click the Continue button.
10. You will have returned to the previous screen and will now have to click the OK button. This will take you to the Data View Editor which will now show that you have created a new variable that is based on recoding your existing variable:



	Scores	NScores
1	58.00	3.00
2	86.00	1.00
3	74.00	2.00
4	70.00	2.00
5	79.00	1.00
6	60.00	3.00
7	35.00	4.00
8	42.00	3.00
9	55.00	3.00
10	91.00	1.00

Recoding Variables in SPSS (cont...)

Recoding data into two categories

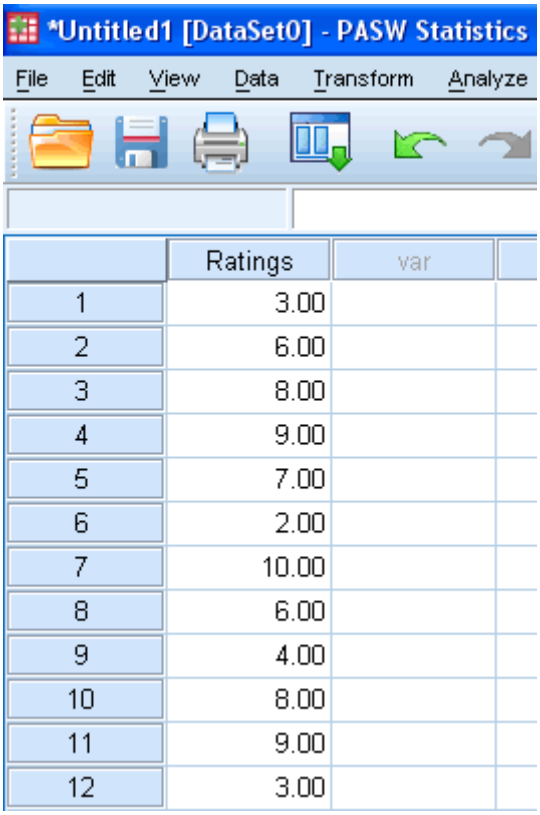
Example: The data given below represents a satisfaction rating out of 10 for a new service offered by a company. The company would like to code all those who responded by giving ratings above 5 a "Satisfactory" code and those below 5 a "Dissatisfactory" code.

Satisfaction scores for a new service

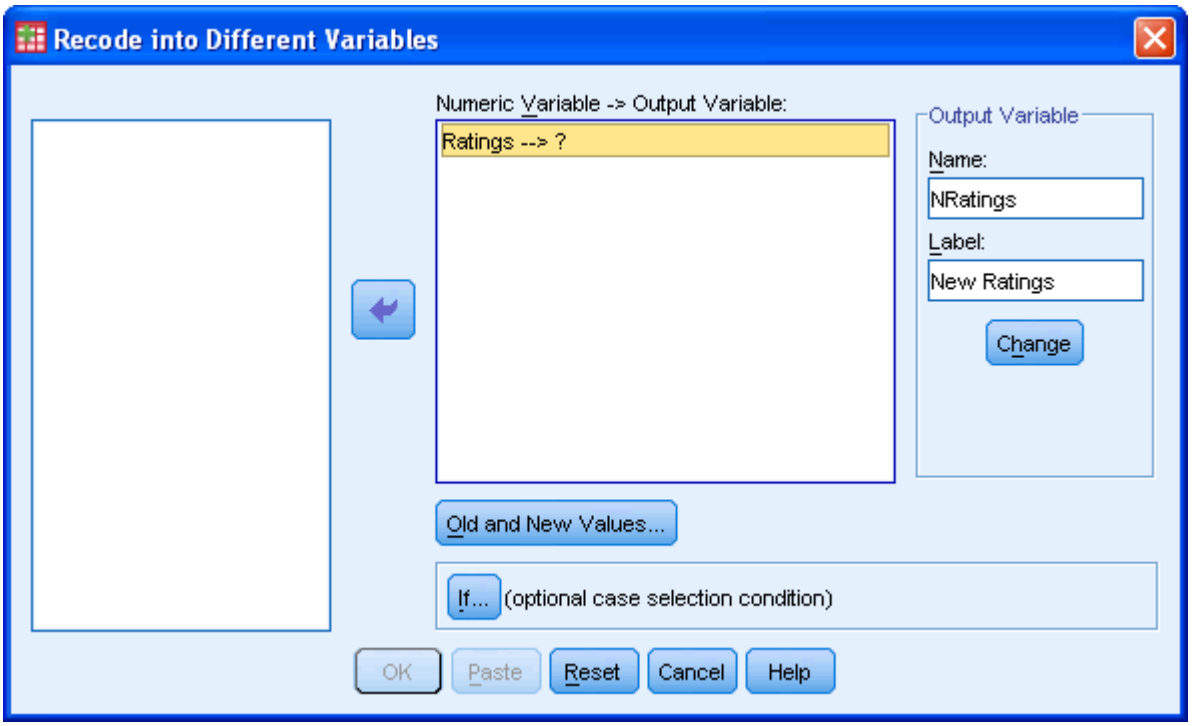
Scores	3	6	8	9	7	2	10	6	4	8	9	3

1. Enter the data in the SPSS Data Editor and name the variable "**Ratings**".

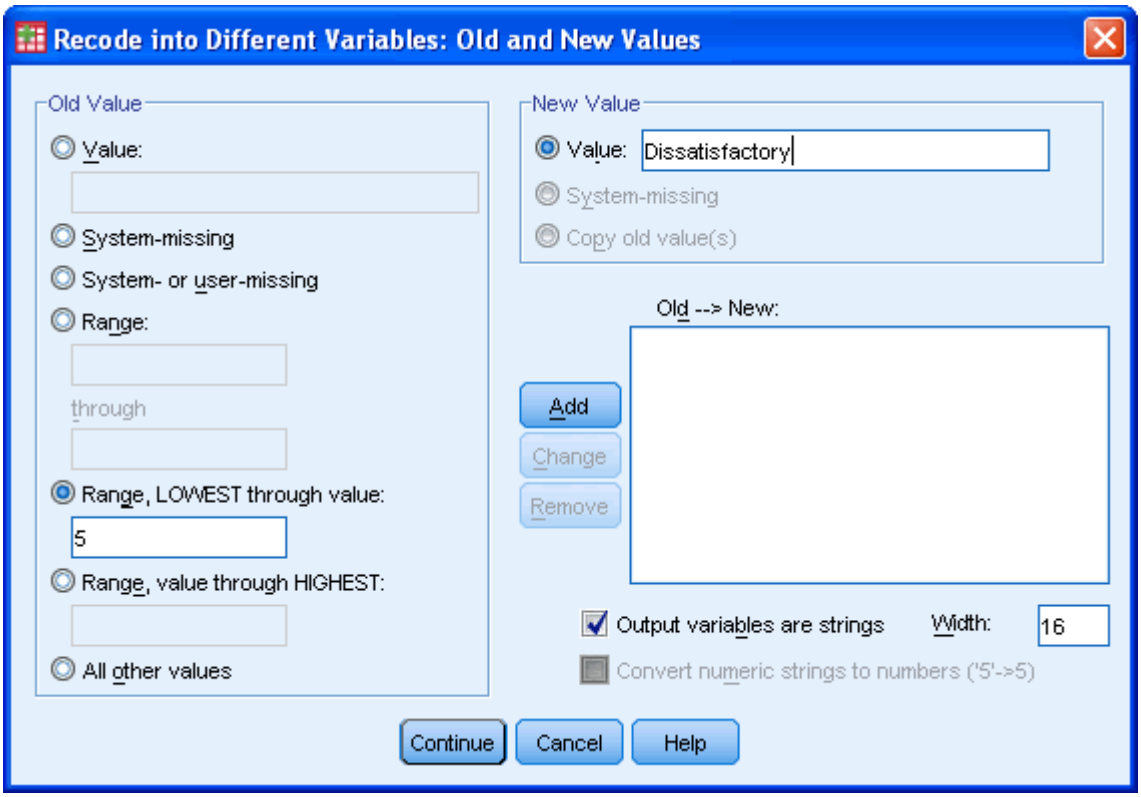
   Remember that each individual's results go on a separate line (row) in SPSS.



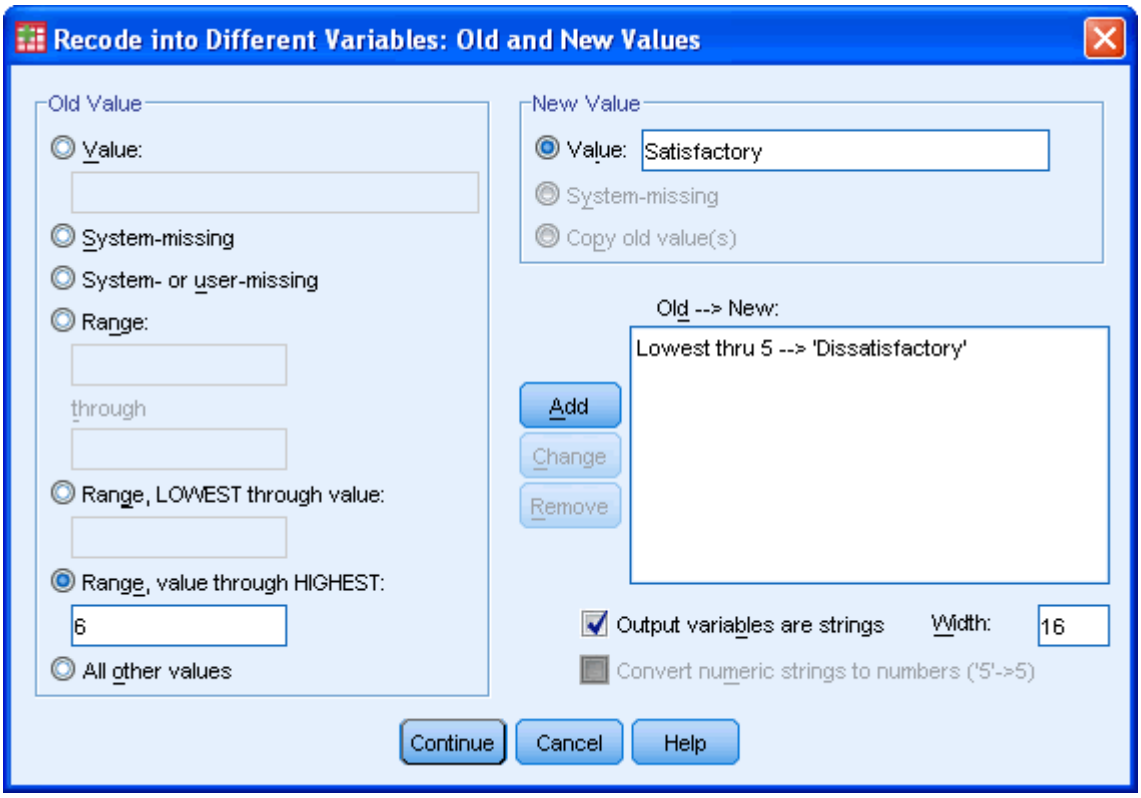
2. Click on **Transform > Recode Into Different Variable...** in the top menu.
3. Transfer the variable you want to recode by selected it and pressing the button and give the new variable a name and label. In this example, we have given the new variable a name of "**NRatings**" and label of "**New Ratings**" as shown below:



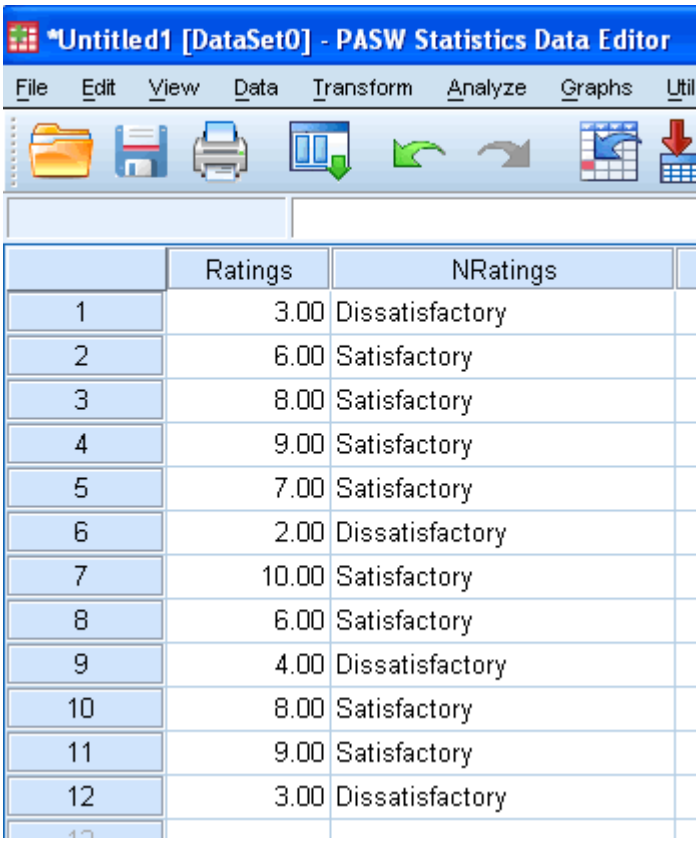
4. Click the Change button.
5. Click the Old and New Values... button.
6. Enter the value of "**5**" into the "*Range, LOWEST through value:*" box and set the new code to "**1**" into the "*New Value*" box. Click the "*Output variables are strings*" checkbox so that it will except values that are not numbers. You need to increase the "*Width:*" from 8 to 16 also, so that the string labels are not truncated.



7. Click the button.
8. Enter the value of "**6**" into the "*Range, value through HIGHEST:*" box and set the new code to "**1**" into the "*New Value*" box.



9. Click the Add button.
10. Click the Continue button.
11. You will have returned to the previous screen and will now have to click the OK button. This will take you to the Data View Editor which will now show that you have created a new variable that is based on recoding your existing variable:



	Ratings	NRatings
1	3.00	Dissatisfactory
2	6.00	Satisfactory
3	8.00	Satisfactory
4	9.00	Satisfactory
5	7.00	Satisfactory
6	2.00	Dissatisfactory
7	10.00	Satisfactory
8	6.00	Satisfactory
9	4.00	Dissatisfactory
10	8.00	Satisfactory
11	9.00	Satisfactory
12	3.00	Dissatisfactory

Recoding Variables in SPSS (cont...)

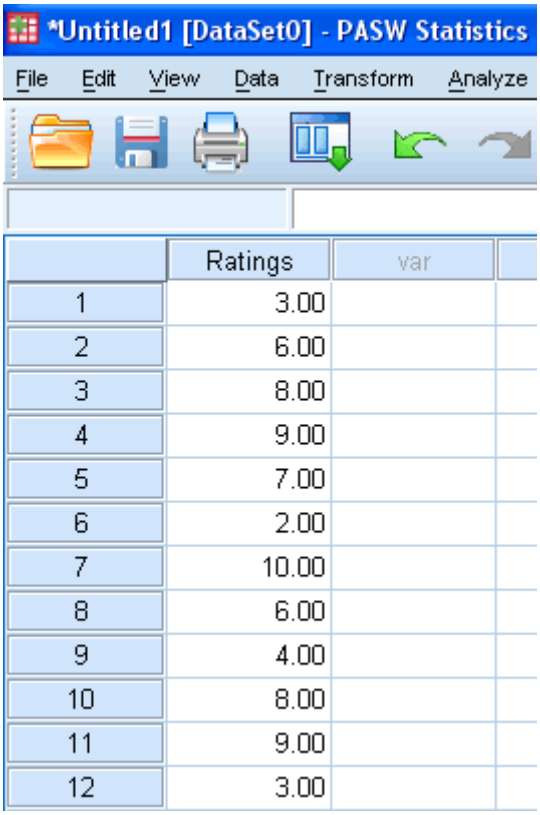
Recoding data into two categories

Example: The data given below represents a satisfaction rating out of 10 for a new service offered by a company. The company would like to code all those who responded by giving ratings above 5 a "Satisfactory" code and those below 5 a "Dissatisfactory" code.

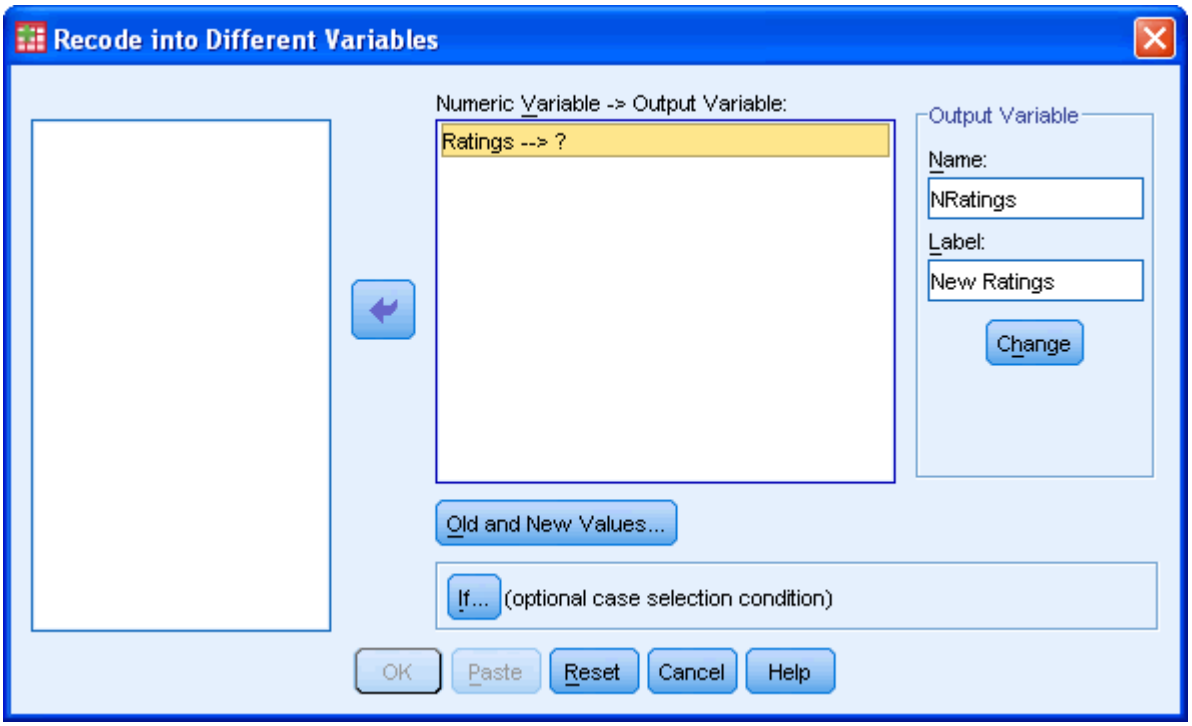
Satisfaction scores for a new service												
Scores	3	6	8	9	7	2	10	6	4	8	9	3

1. Enter the data in the SPSS Data Editor and name the variable "**Ratings**".

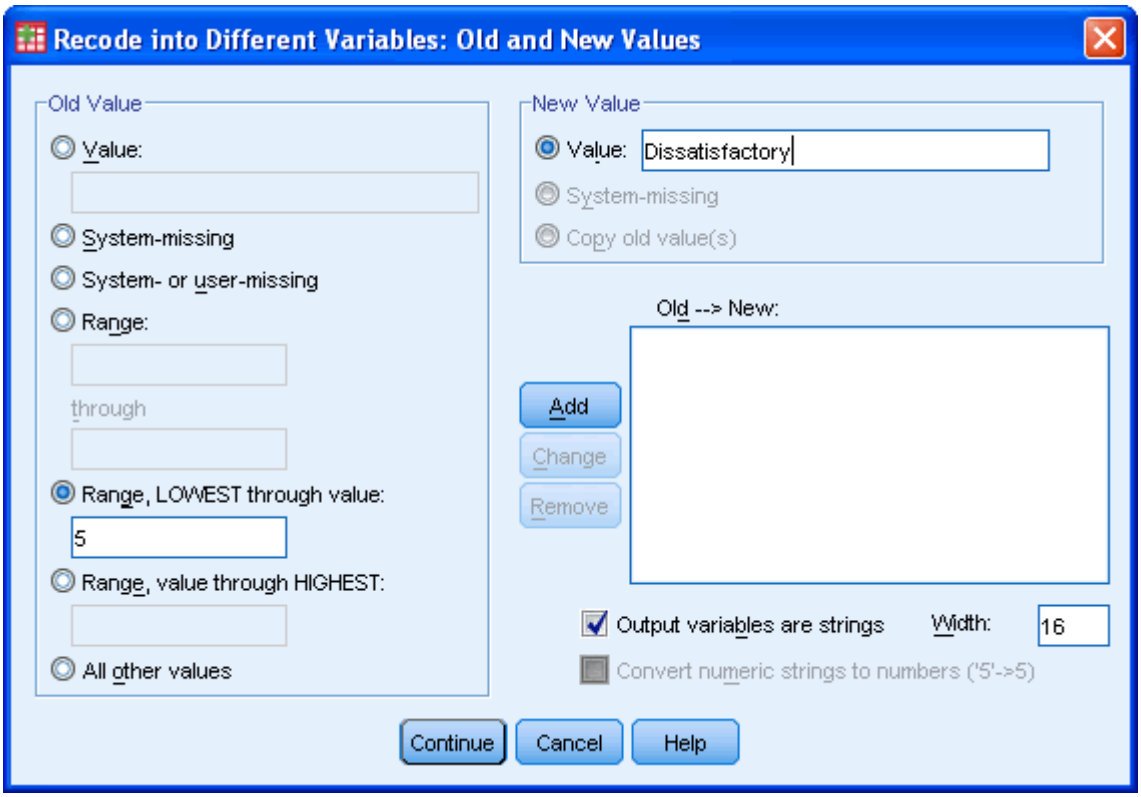
   Remember that each individual's results go on a separate line (row) in SPSS.



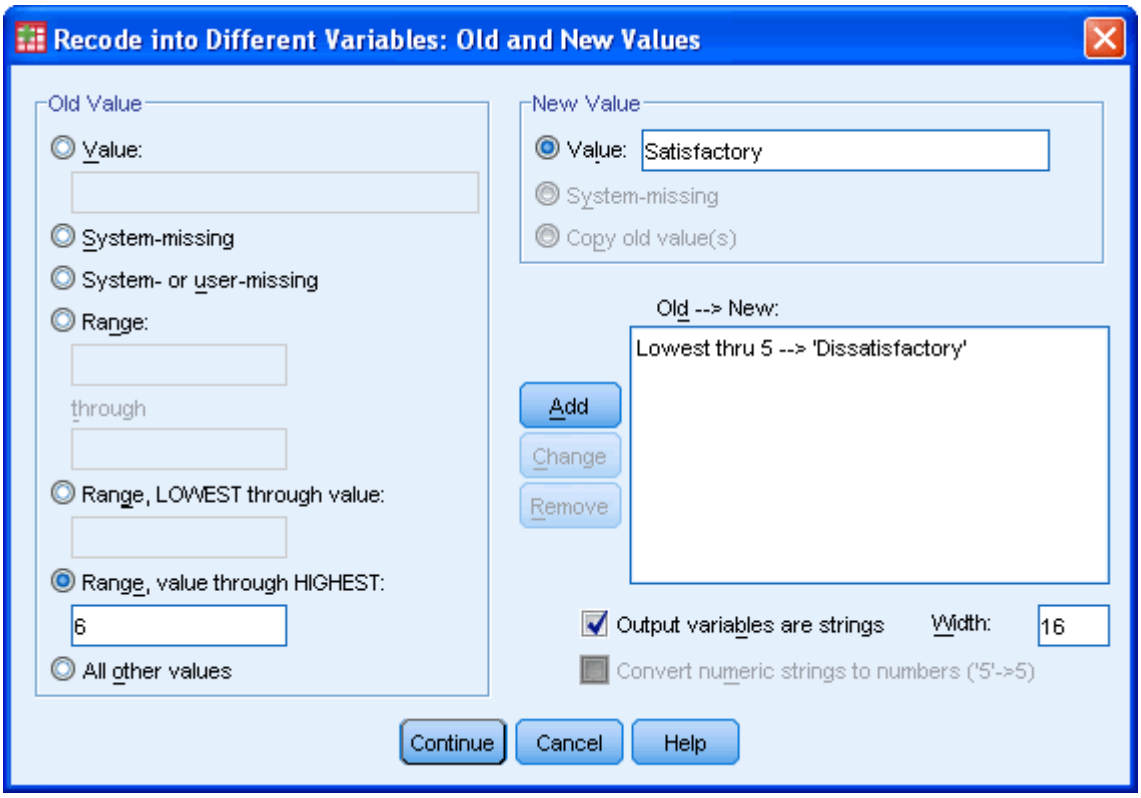
2. Click on **Transform > Recode Into Different Variable...** in the top menu.
3. Transfer the variable you want to recode by selected it and pressing the button and give the new variable a name and label. In this example, we have given the new variable a name of "**NRatings**" and label of "**New Ratings**" as shown below:



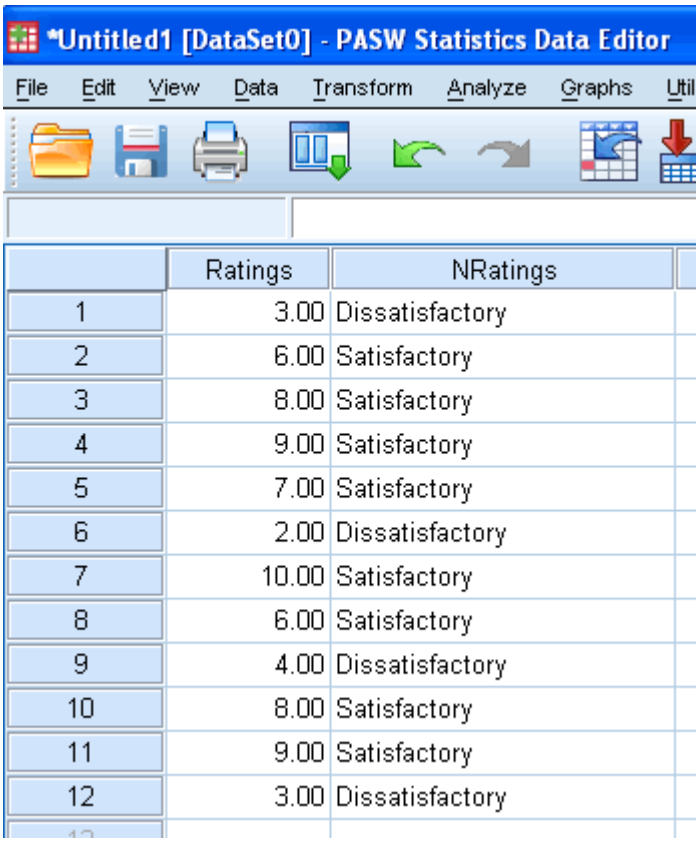
4. Click the Change button.
5. Click the Old and New Values... button.
6. Enter the value of "**5**" into the "*Range, LOWEST through value:*" box and set the new code to "**1**" into the "*New Value*" box. Click the "*Output variables are strings*" checkbox so that it will except values that are not numbers. You need to increase the "*Width:*" from 8 to 16 also, so that the string labels are not truncated.



7. Click the button.
8. Enter the value of "**6**" into the "*Range, value through HIGHEST:*" box and set the new code to "**1**" into the "*New Value*" box.



9. Click the Add button.
10. Click the Continue button.
11. You will have returned to the previous screen and will now have to click the OK button. This will take you to the Data View Editor which will now show that you have created a new variable that is based on recoding your existing variable:



	Ratings	NRatings
1	3.00	Dissatisfactory
2	6.00	Satisfactory
3	8.00	Satisfactory
4	9.00	Satisfactory
5	7.00	Satisfactory
6	2.00	Dissatisfactory
7	10.00	Satisfactory
8	6.00	Satisfactory
9	4.00	Dissatisfactory
10	8.00	Satisfactory
11	9.00	Satisfactory
12	3.00	Dissatisfactory

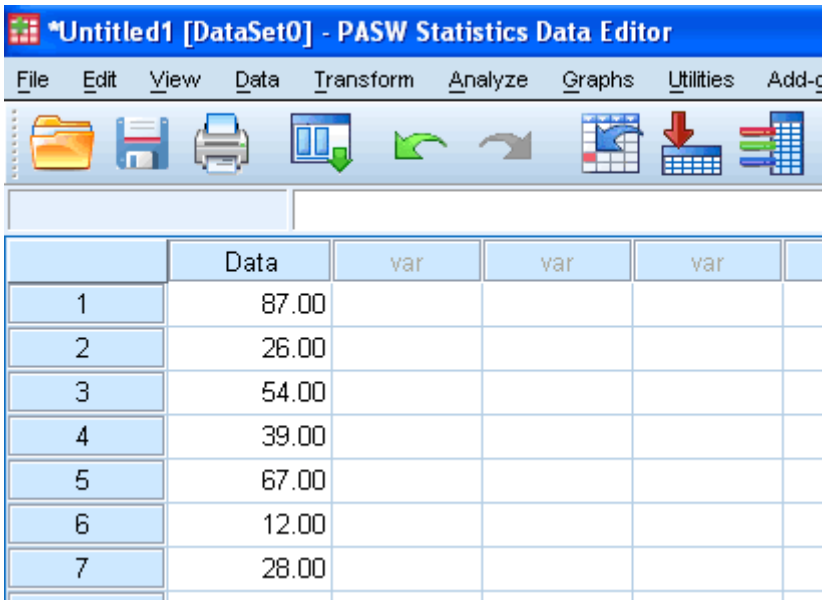
Ranking Data in SPSS

Ranking is used to recode the data into their rank ordering from smallest to largest or largest to smallest. We will demonstrate this by entering in some data and ranking it in SPSS.

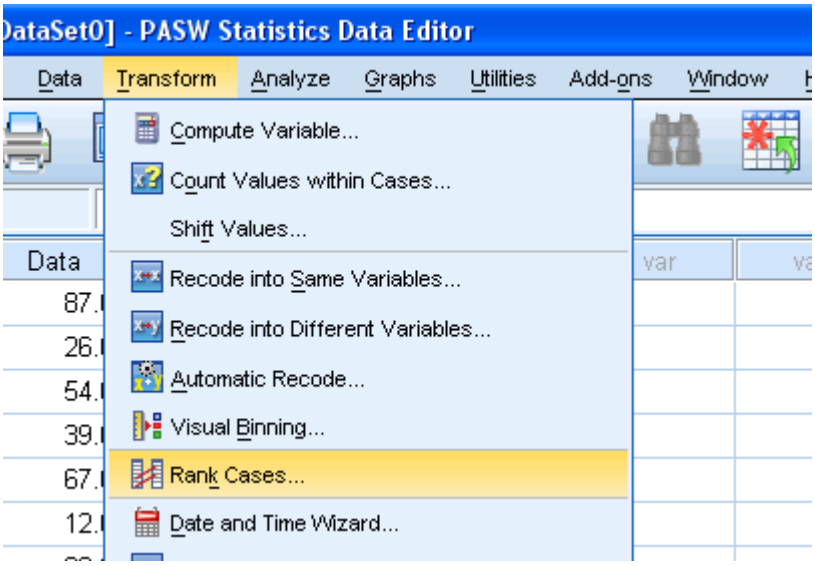
Data to Rank														
87	26	54	39	67	12	28	98	54	68	23	64	28	43	77

1. Your data should end up looking like the following (we have named the variable "Data"):

   Remember that each individual's results go on a separate line (row) in SPSS.



2. Click on **Transform > Rank Cases...** in the top menu.

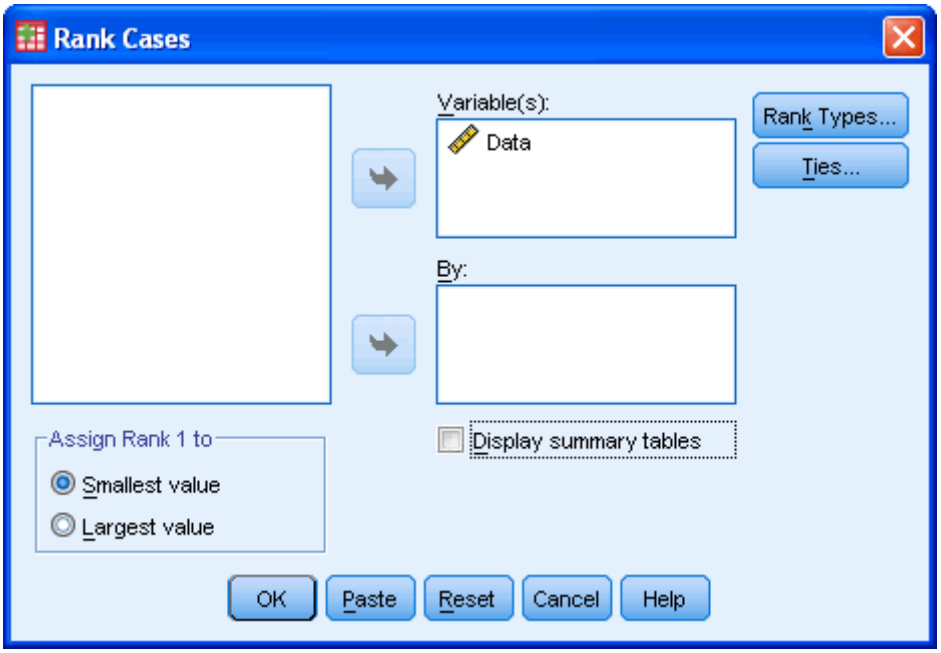


.

3. Click on the Data variable in the lefthand box and click the button to move it to the "*Variable(s):*" box.

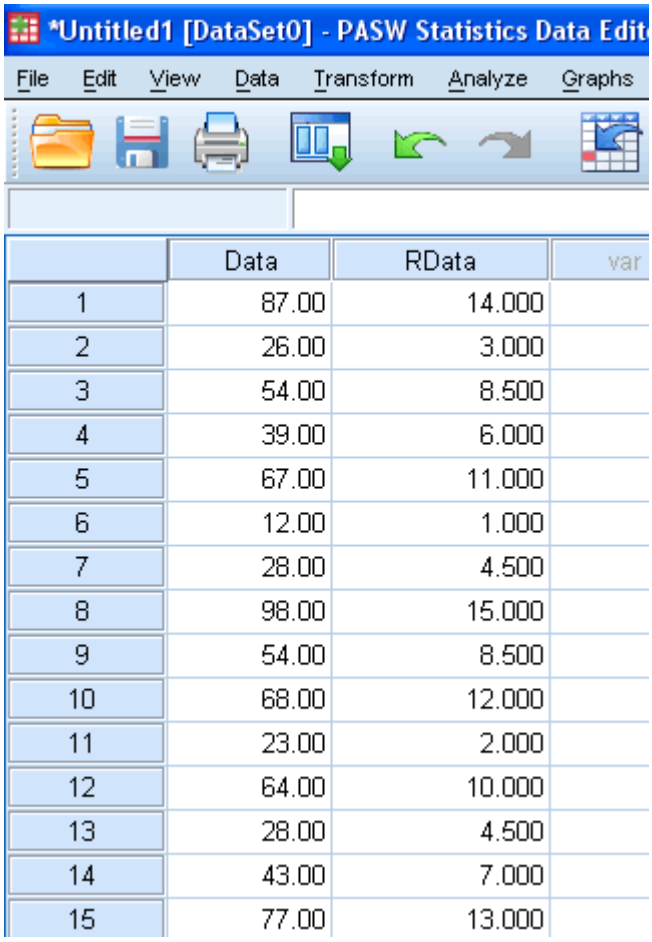
   Uncheck the "*Display summary tables*" checkbox.

   If you wish to have the largest value have a rank of "1" then select the radio box "*Largest value*" from the "*Assign Rank 1 to*" box. We will stick with the default in this example, which is "*Smallest value*".



4. Click on the

button.
5. You will be presented with the SPSS Data Editor which will now show the ranked data under a new variable name. This new variable name will have the same variable name as the old name but with the addition of the letter "R" at the beginning. Hence, in this case, the ranked data variable is called "RData".



	Data	RData	var
1	87.00	14.000	
2	26.00	3.000	
3	54.00	8.500	
4	39.00	6.000	
5	67.00	11.000	
6	12.00	1.000	
7	28.00	4.500	
8	98.00	15.000	
9	54.00	8.500	
10	68.00	12.000	
11	23.00	2.000	
12	64.00	10.000	
13	28.00	4.500	
14	43.00	7.000	
15	77.00	13.000	

Graphs & Charts

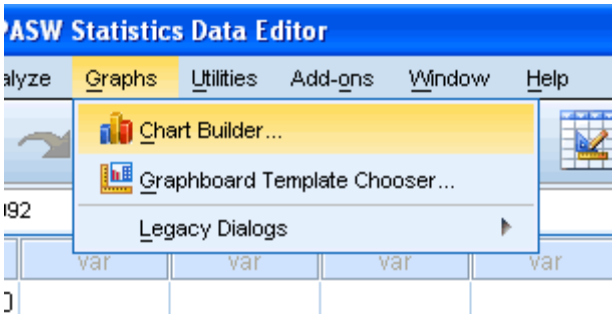
Creating a Bar Chart using SPSS

Objectives

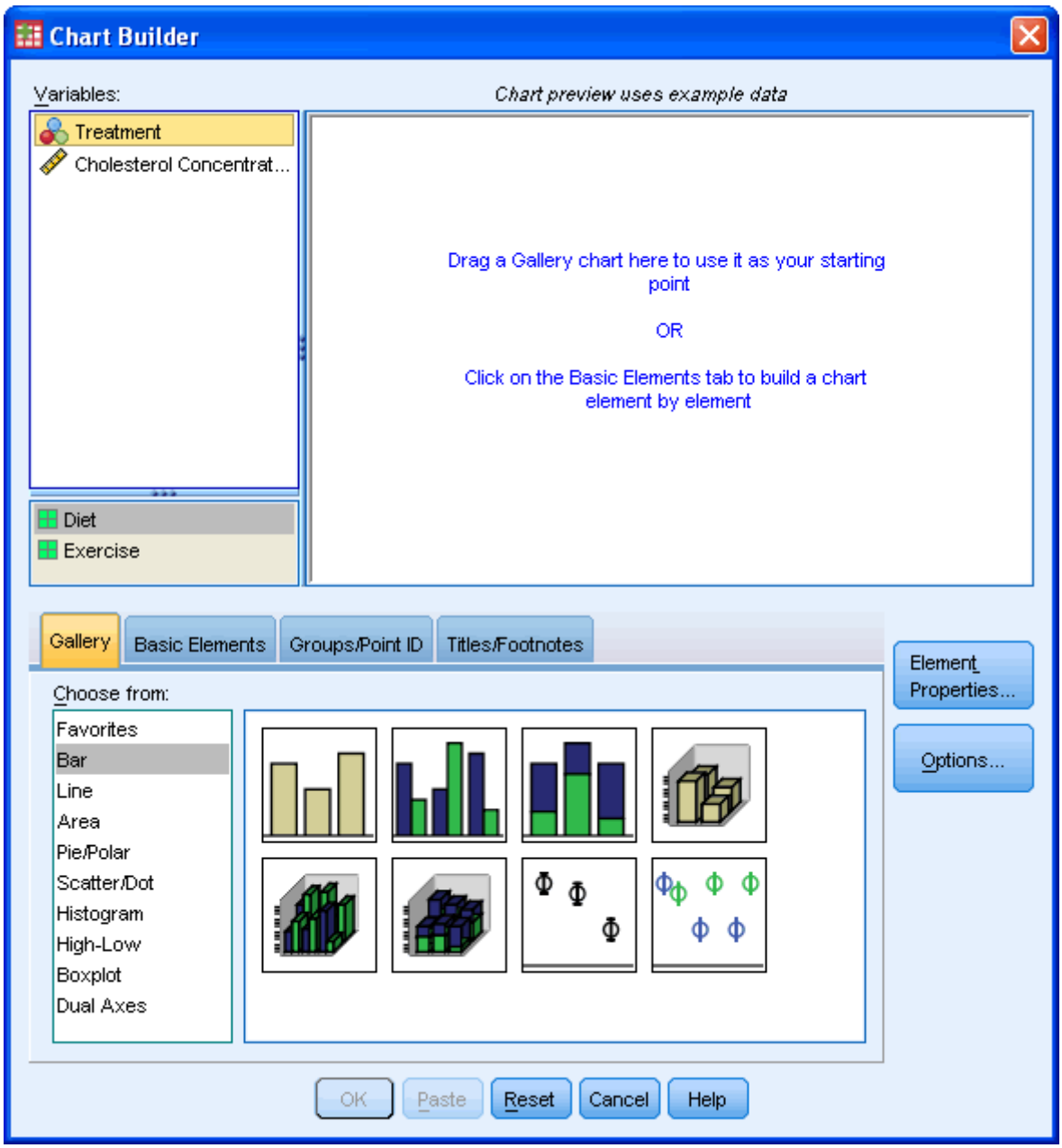
A bar chart is helpful in graphically describing (visualizing) your data; it will often be used in addition to inferential statistics (see our Descriptive and Inferential Statistics guide). A bar chart can be appropriate if you are running an Independent T-Test or Dependent T-Test. The example we will use is based on the data from our Independent T-Test guide.

Test Procedure in SPSS

1. Click **Graphs > Chart Builder...** on the top menu as shown below:

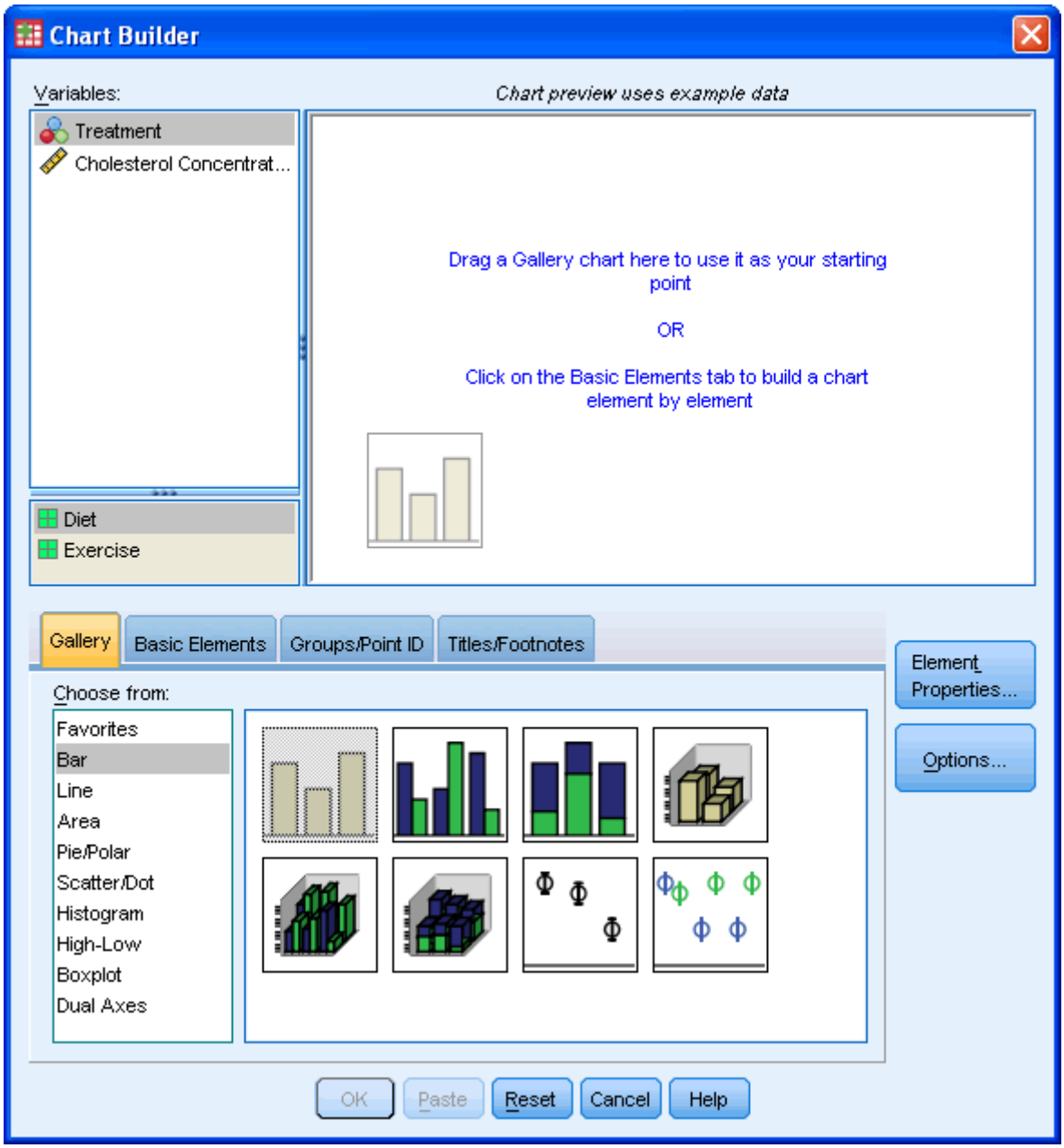


2. You will be presented with the following screen:



3. Under the Gallery Tab (

) select the **Bar** option and the **simple bar chart** icon (top-left icon). Drag-and-drop this icon into the **Chart Preview Area**.

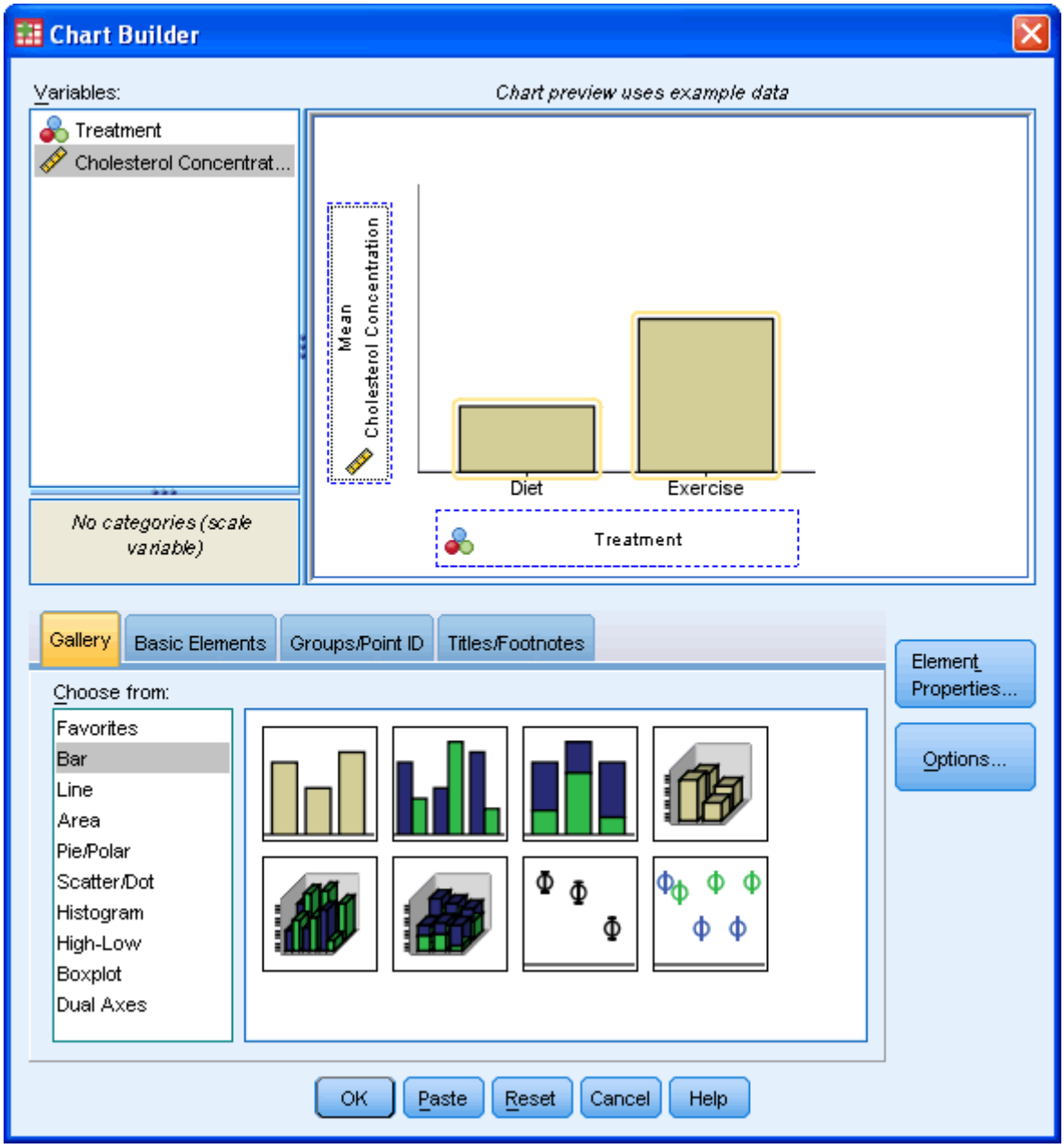


4. You will be presented with the following dialog boxes: **Chart Builder** and **Element Properties**. As you can see, the **Chart Preview Area** has been populated with a template of a simple bar chart.

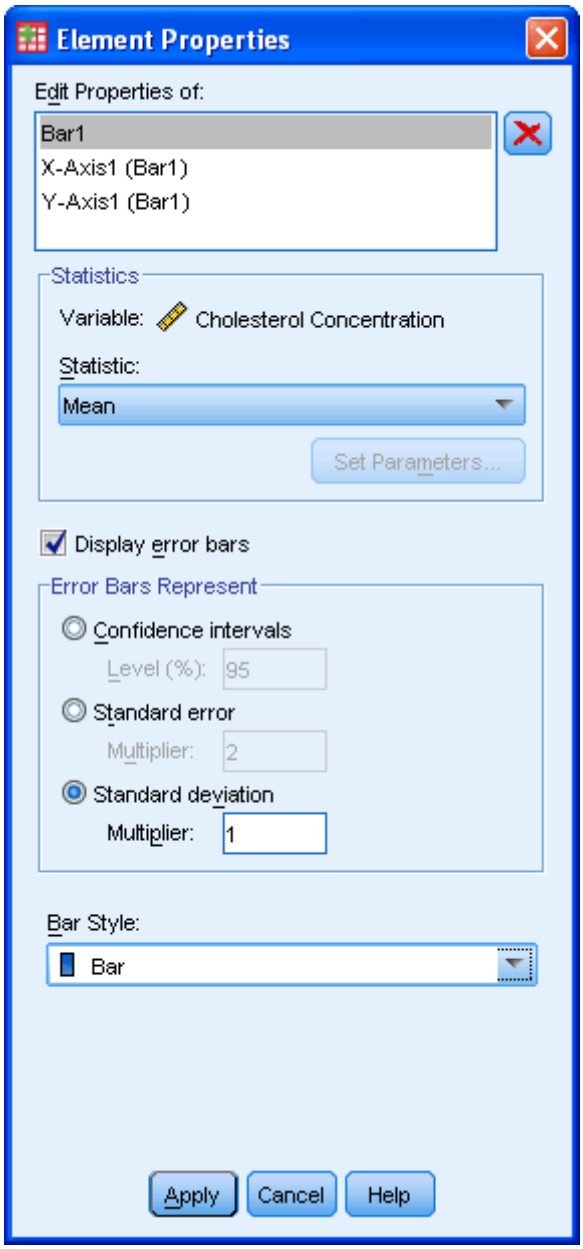
Creating a Bar Chart using SPSS (cont...)

5. Transfer the independent variable, Treatment, into the "**X-Axis?**" box and the dependent (outcome) variable, Cholesterol Concentration, into the "**Y-**

Axis?" box within the **Preview Chart Area** by drag-and-dropping the variables from the "Variables:" box.



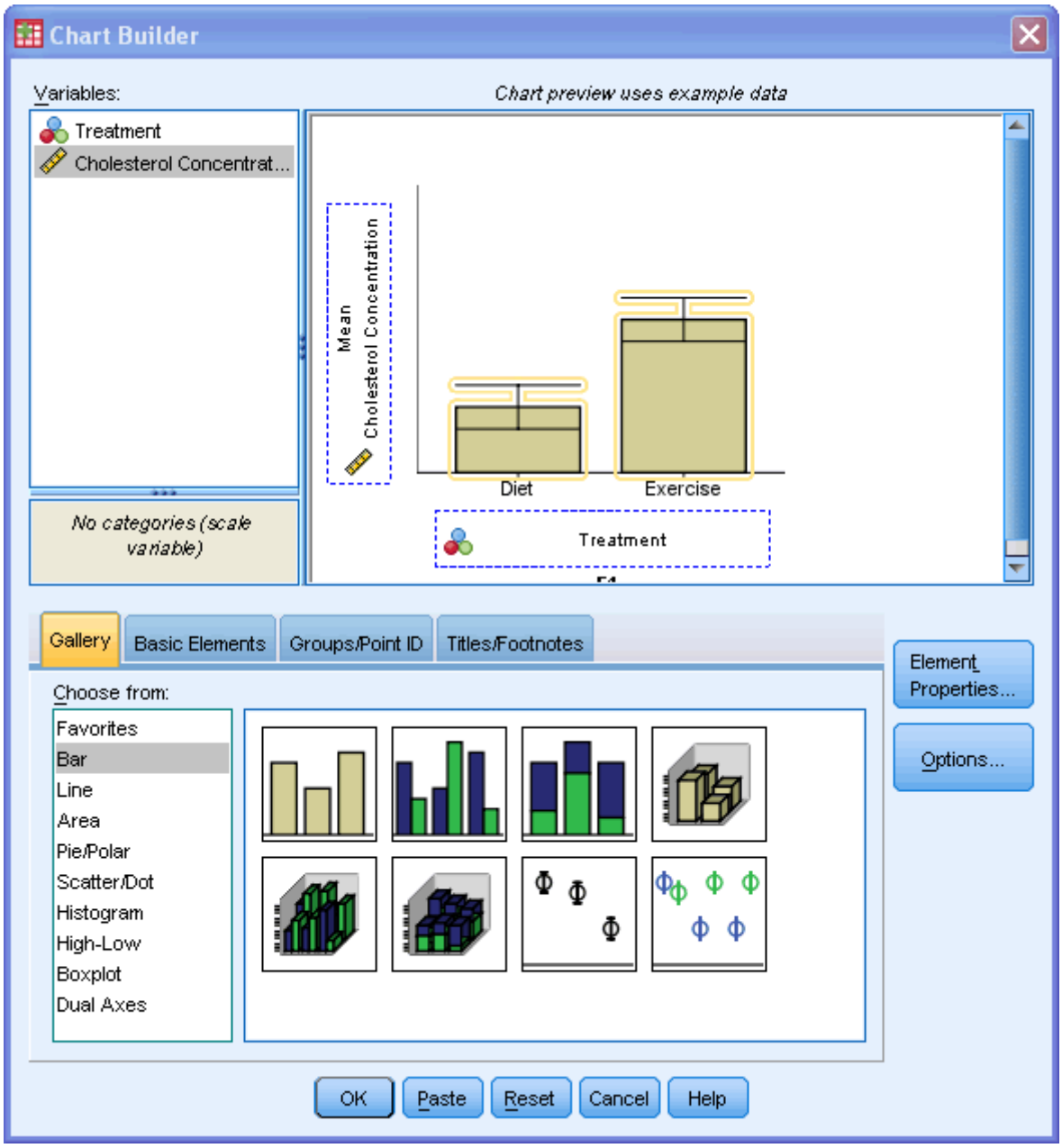
6. Ideally, we want to be able to show a measure of the spread of the data. In this case, we wish to have error bars that represent ± 1 standard deviations. To do this we tick the "**Dispay error bars**" checkbox and then, under the "**Error Bars Represent**" area we check the radio box entitled "**Standard deviation, Multiplier:**" and enter "1".



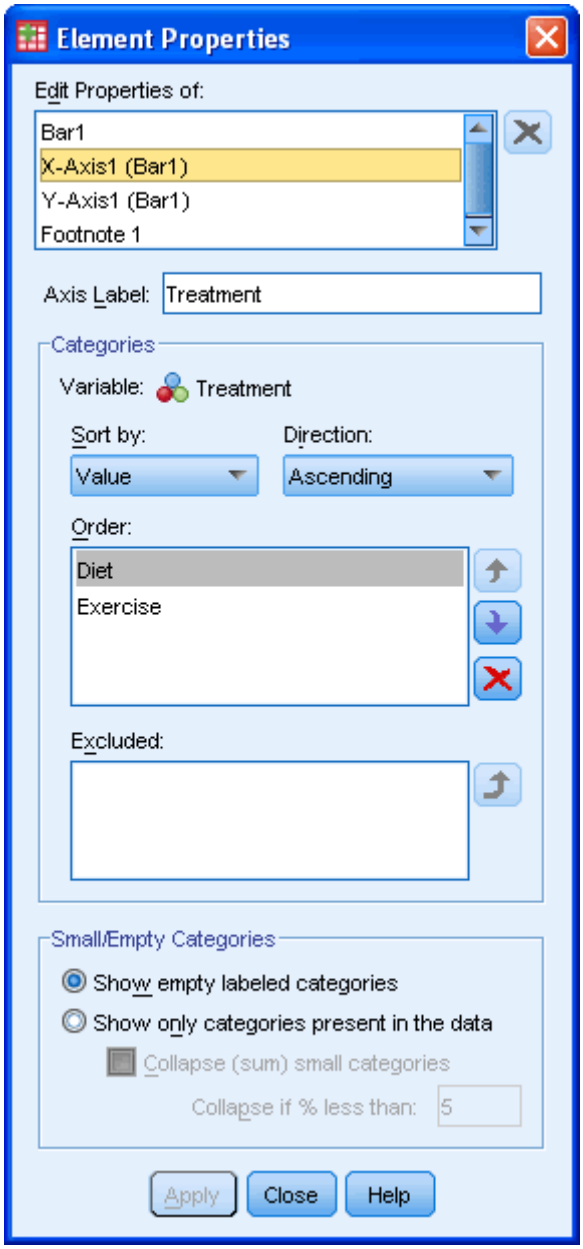
Click the

button.

7. You will be presented with the following screen (showing the error bars added in the "**Chart Preview Area**"):

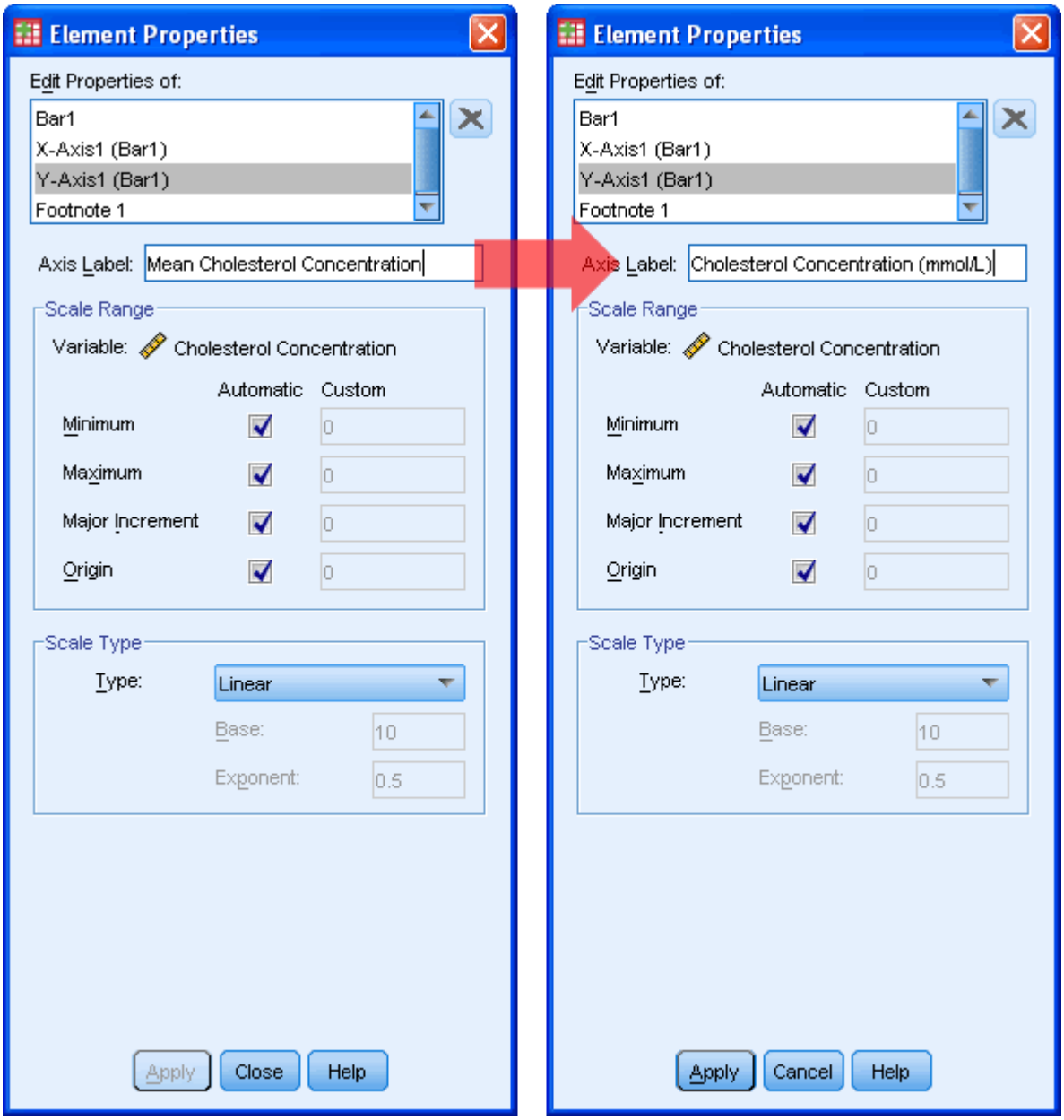


8. We do not need to do anything in the following screen, in this example. However, it does present some options which you might find useful. You can use the and to rearrange the order of the categories and the button to exclude a category. If you make a mistake and exclude a variable you later want to include then you can simply click the button in the "Excluded:" area.



If you make any changes remember to click the Apply button.

9. We want to change the y-axis label so that we can remove the "mean" text and add in some units of measurement. We do this by selecting "Y-Axis (Bar1)" in the "Edit Properties of:" area and then change the "Axis Label:" as below:

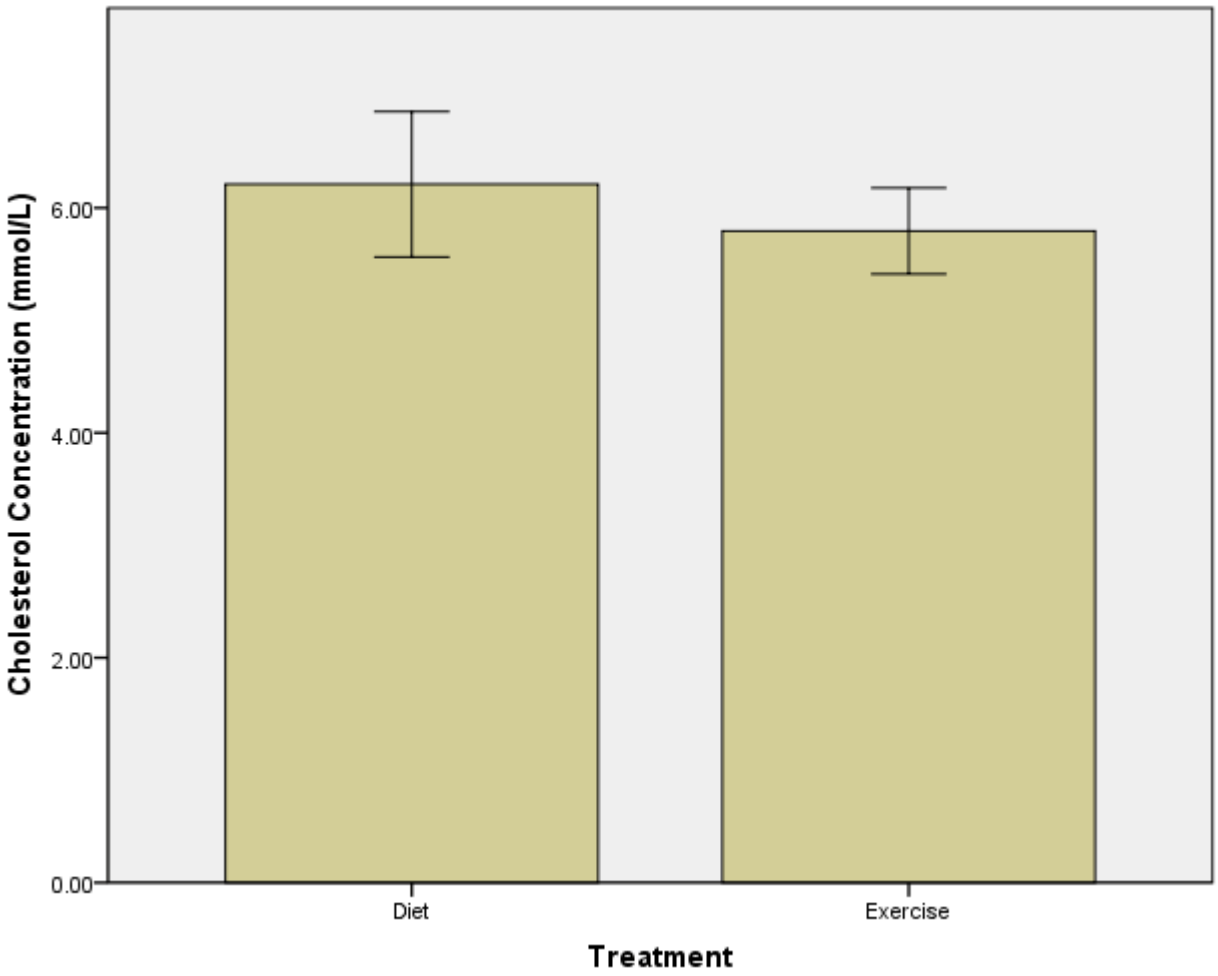


Click the Apply button.

10. Click the OK button.

Output

You will be presented with the following:



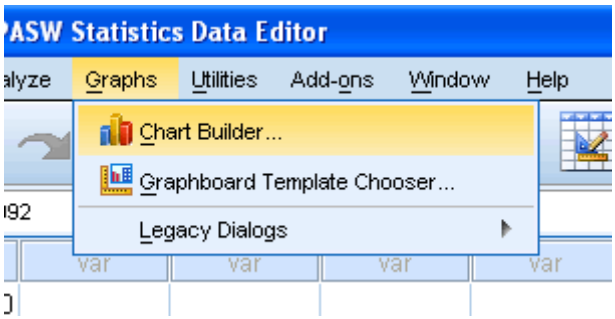
Creating a Bar Chart using SPSS

Objectives

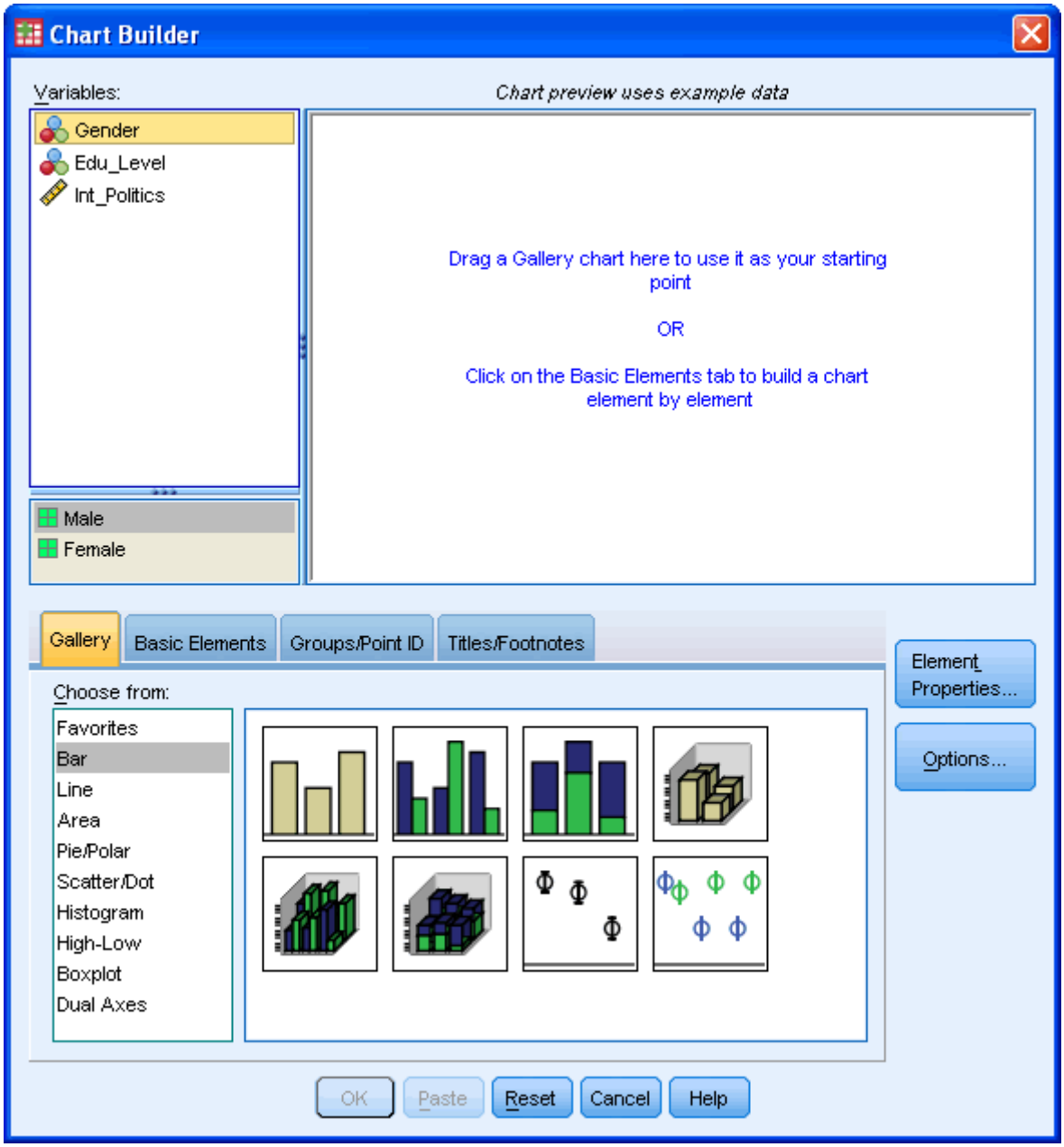
A clustered bar chart is helpful in graphically describing (visualizing) your data; it will often be used in addition to inferential statistics (see our guide on Descriptive and Inferential Statistics). A clustered bar chart can be appropriate if you are running a two-way ANOVA.

Test Procedure in SPSS

1. Click **Graphs > Chart Builder...** on the top menu as shown below:

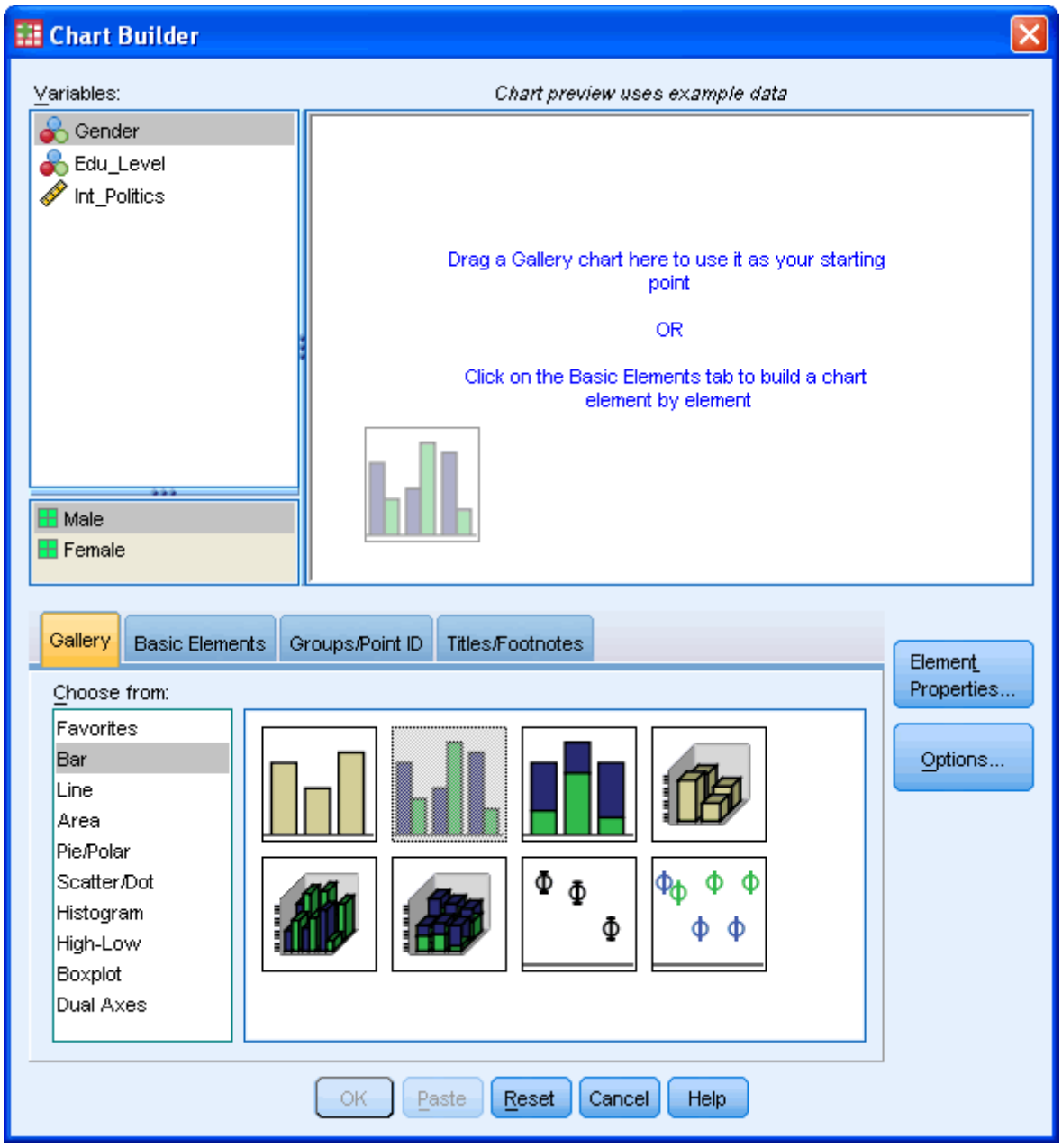


2. You will be presented with the following screen:

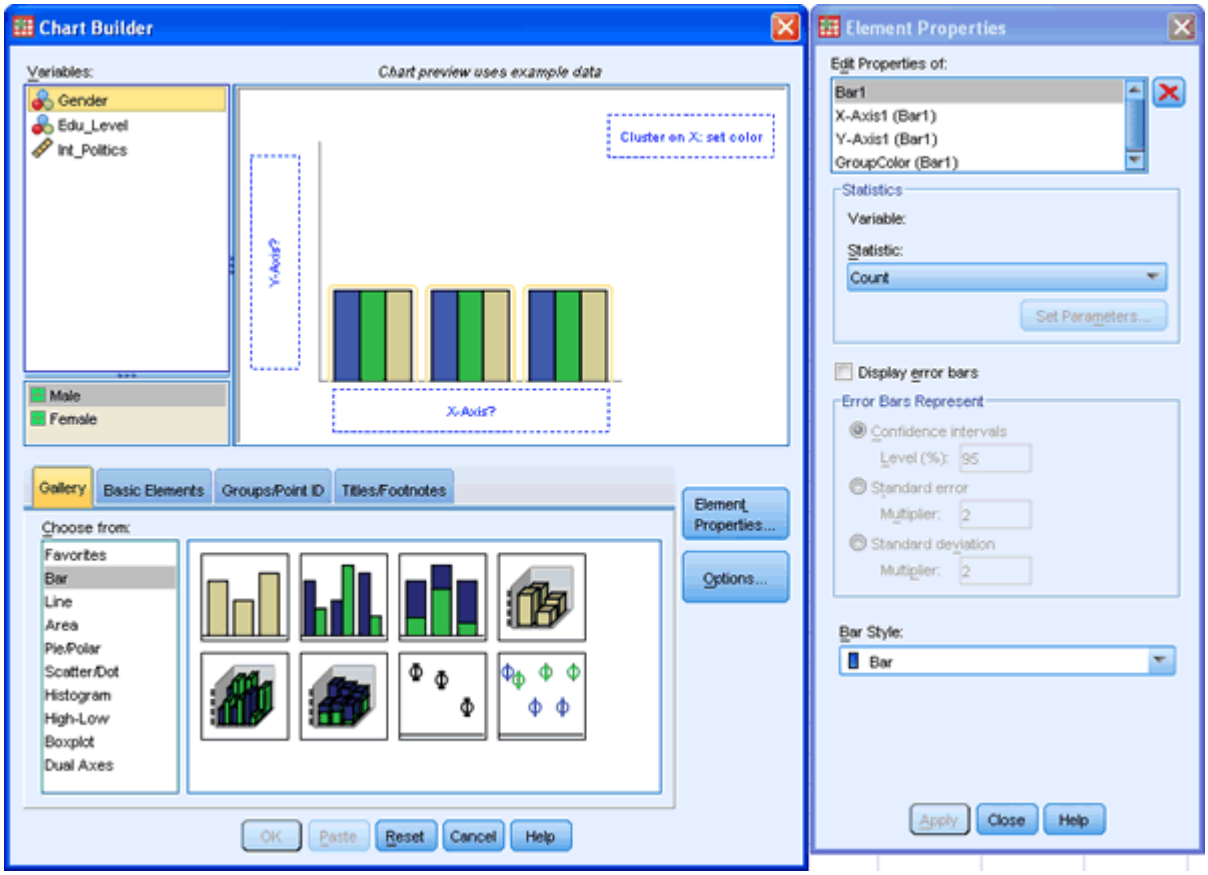


3. Under the Gallery Tab (

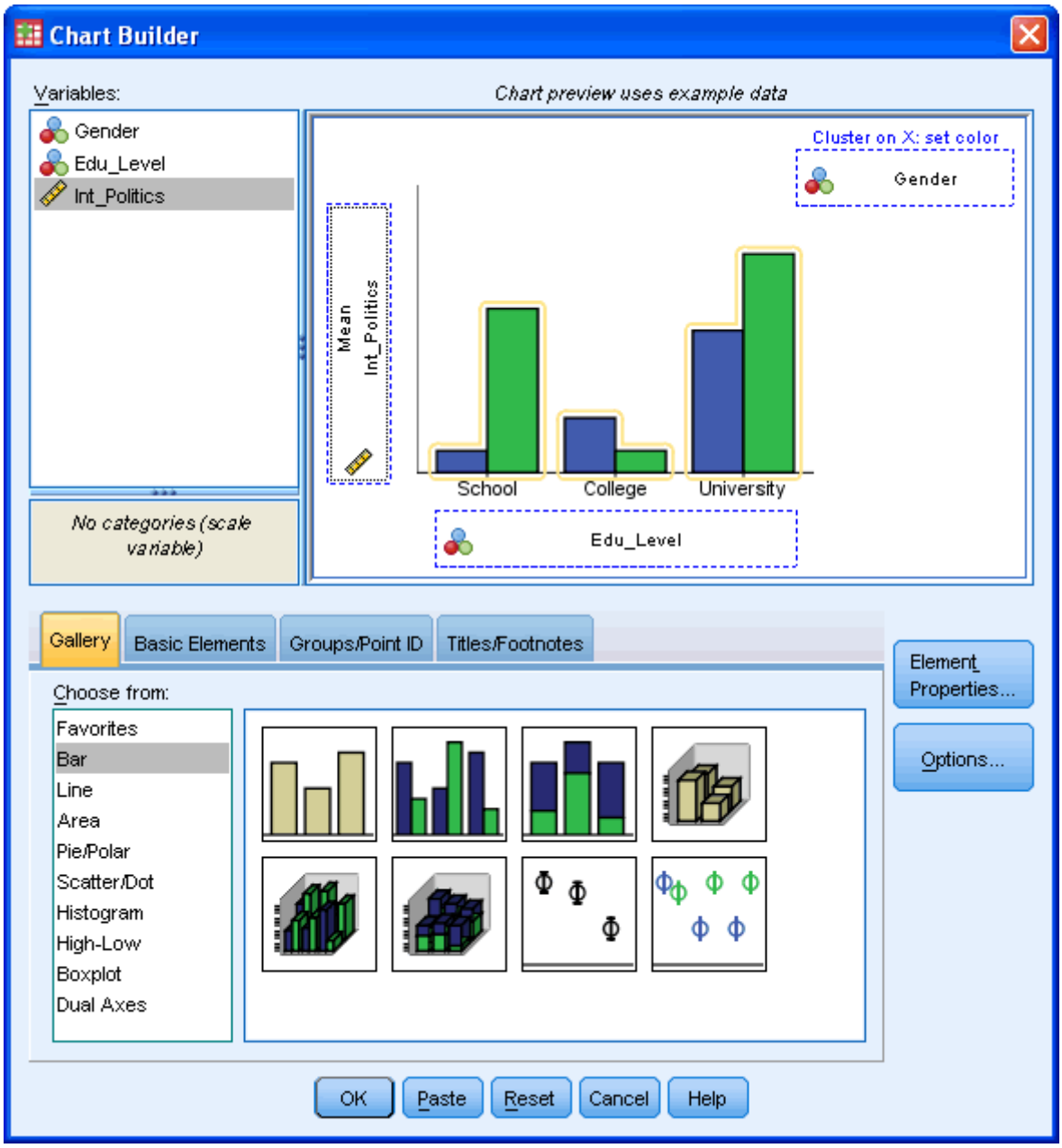
) select the **Bar** option and the **clustered bar chart** icon (top row, second from left). Drag-and-drop this icon into the **Chart Preview Area** (as shown below).



4. You will be presented with the following dialog boxes: **Chart Builder** and **Element Properties**. As you can see, the **Chart Preview Area** has been populated with a template of a clustered bar chart.

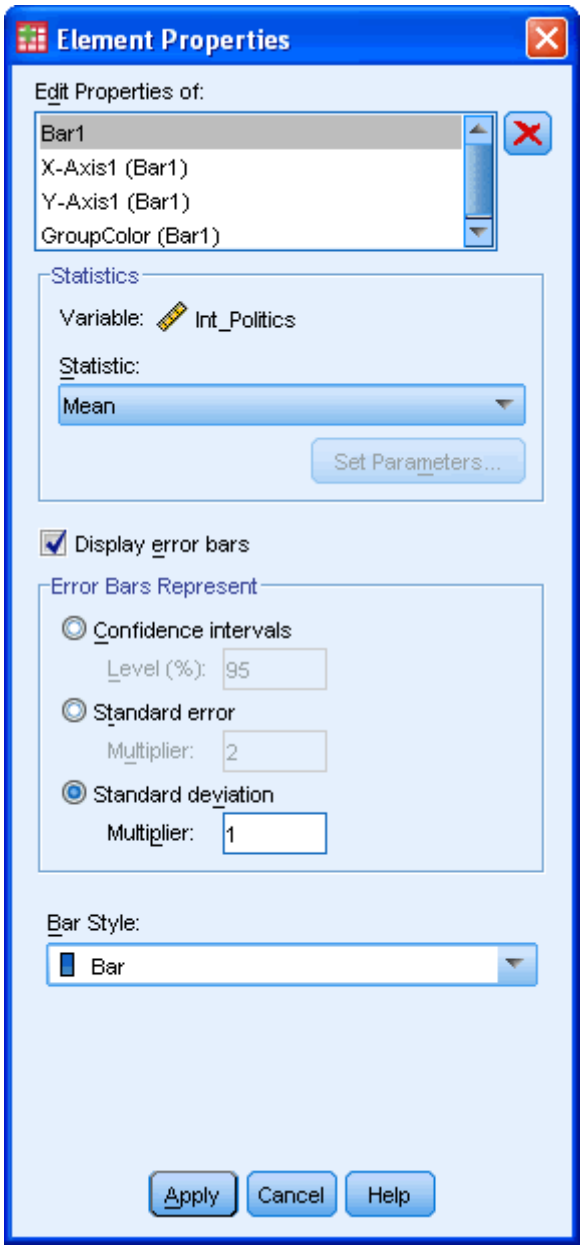


5. Transfer the independent variable "Edu_Level" into the "**X-Axis?**" box, the other independent variable, "Gender", into the "**Cluster on X: set color**" (top-right corner of the **Chart Preview Area**) and the dependent variable "Int_Politics" into the "**Y-Axis?**" box.



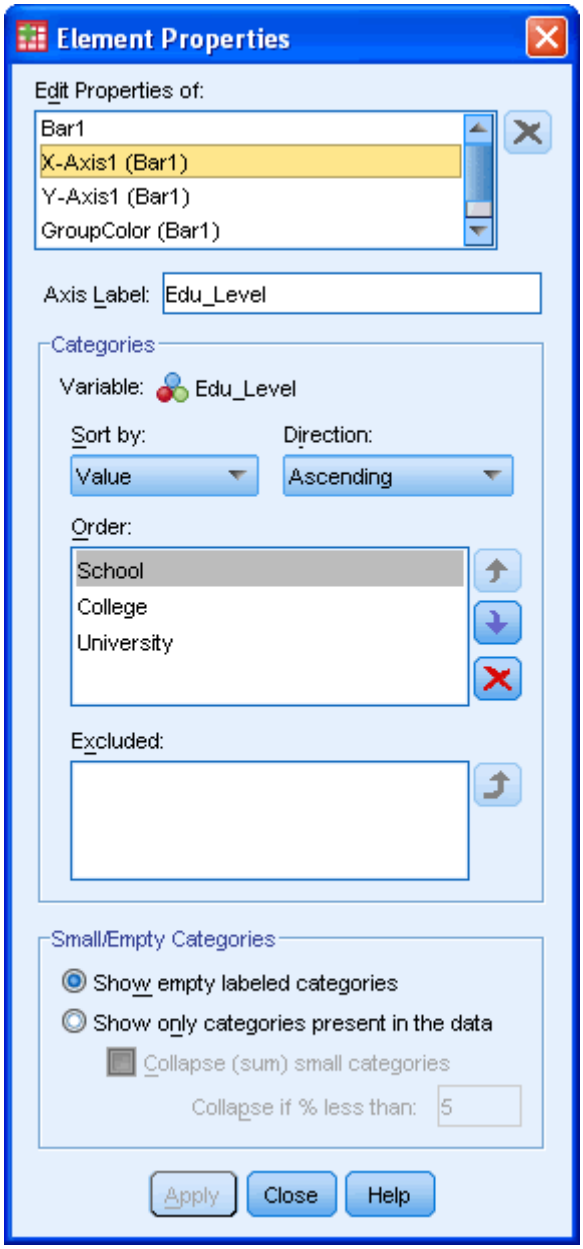
Creating a Bar Chart using SPSS (cont...)

6. Ideally, we want to be able to show a measure of the spread of the data. In this case, we wish to have error bars that represent +/- 1 standard deviations. To do this we tick the "**Dispay error bars**" checkbox and then, under the "**Error Bars Represent**" area we check the radio box entitled "**Standard deviation, Multiplier:**" and enter "1".



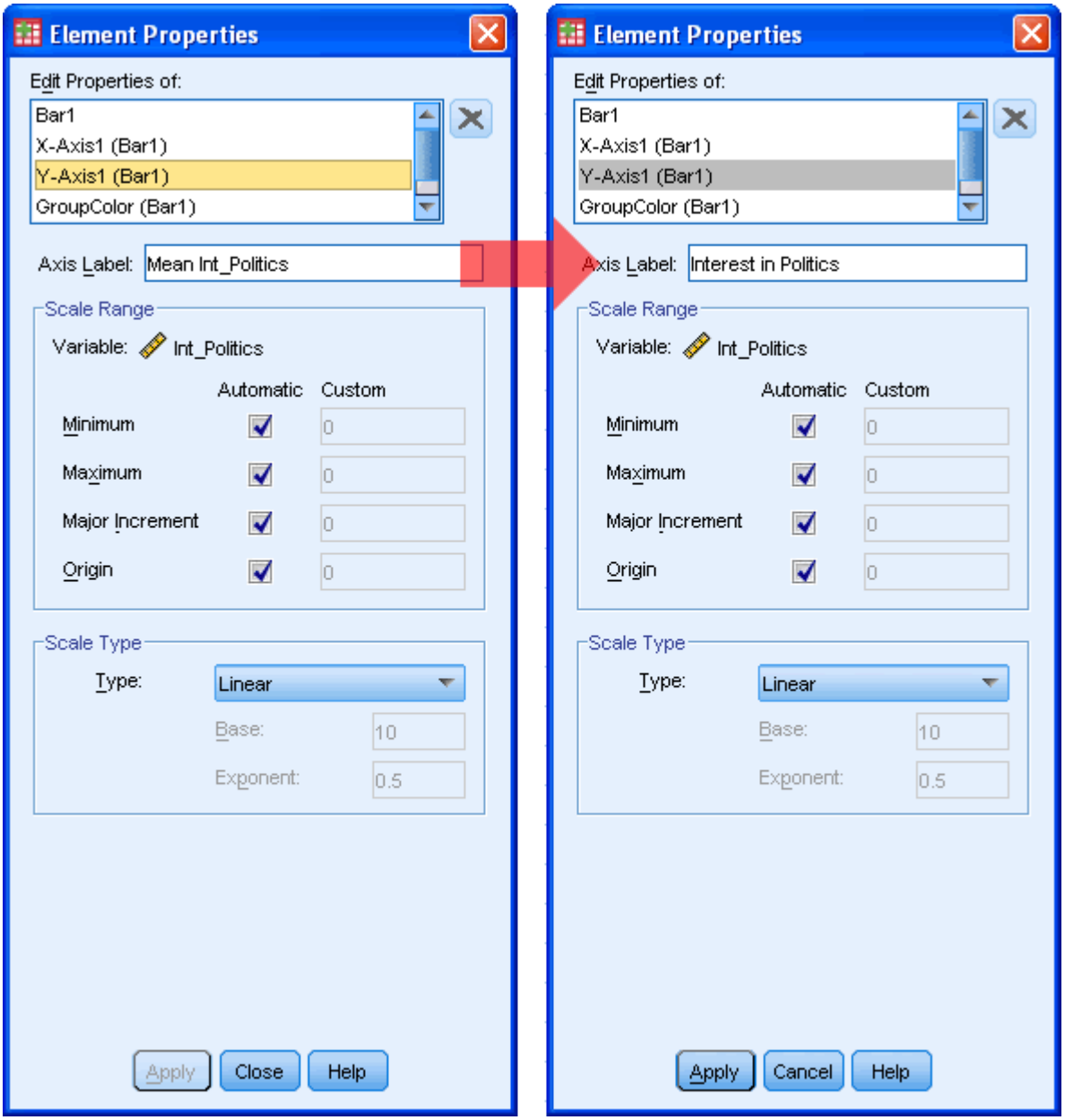
Click the Apply button.

7. We do not need to do anything in the following screen, in this example. However, it does present some options which you might find useful. You can use the and button to rearrange the order of the categories and the button to exclude a category. If you make a mistake and exclude a variable you later want to include then you can simply click the button in the "**Excluded:**" area.



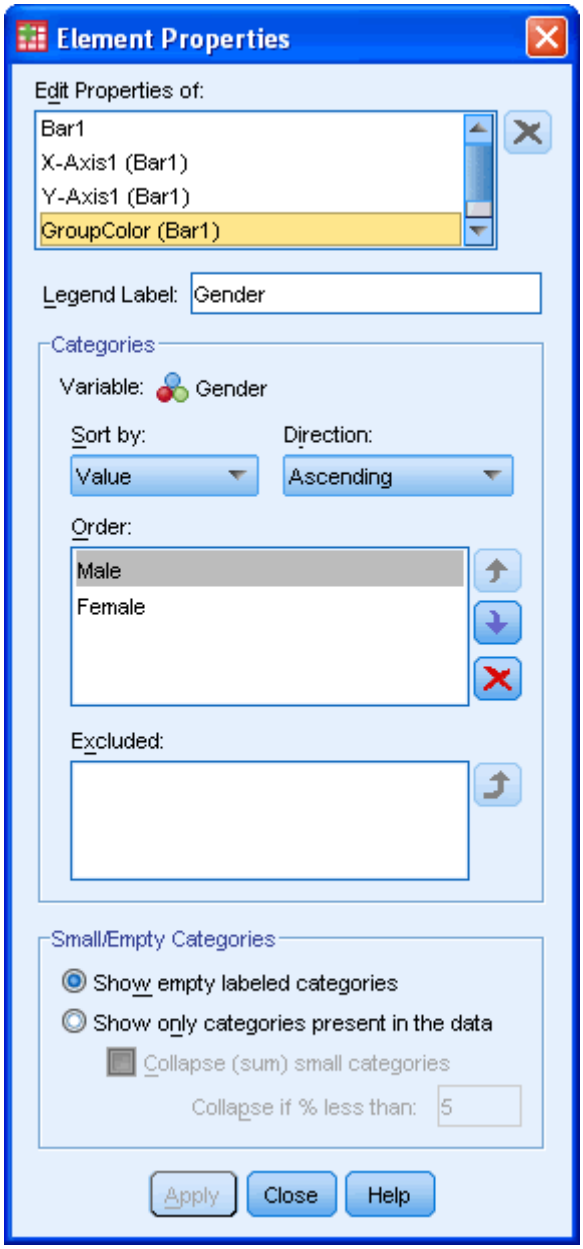
If you make any changes remember to click the Apply button.

8. We want to change the y-axis label so that we can remove the "mean" text and make the title more meaningful. We do this by selecting "**Y-Axis1 (Bar1)**" in the "**Edit Properties of:**" area and then change the "**Axis Label:**" as below:



Click the Apply button.

9. We do not need to do anything in the following screen, in this example. However, it does present some options which you might find useful. You can use the and button to rearrange the order of the categories and the button to exclude a category. If you make a mistake and exclude a variable you later want to include then you can simply click the button in the "**Excluded:**" area.

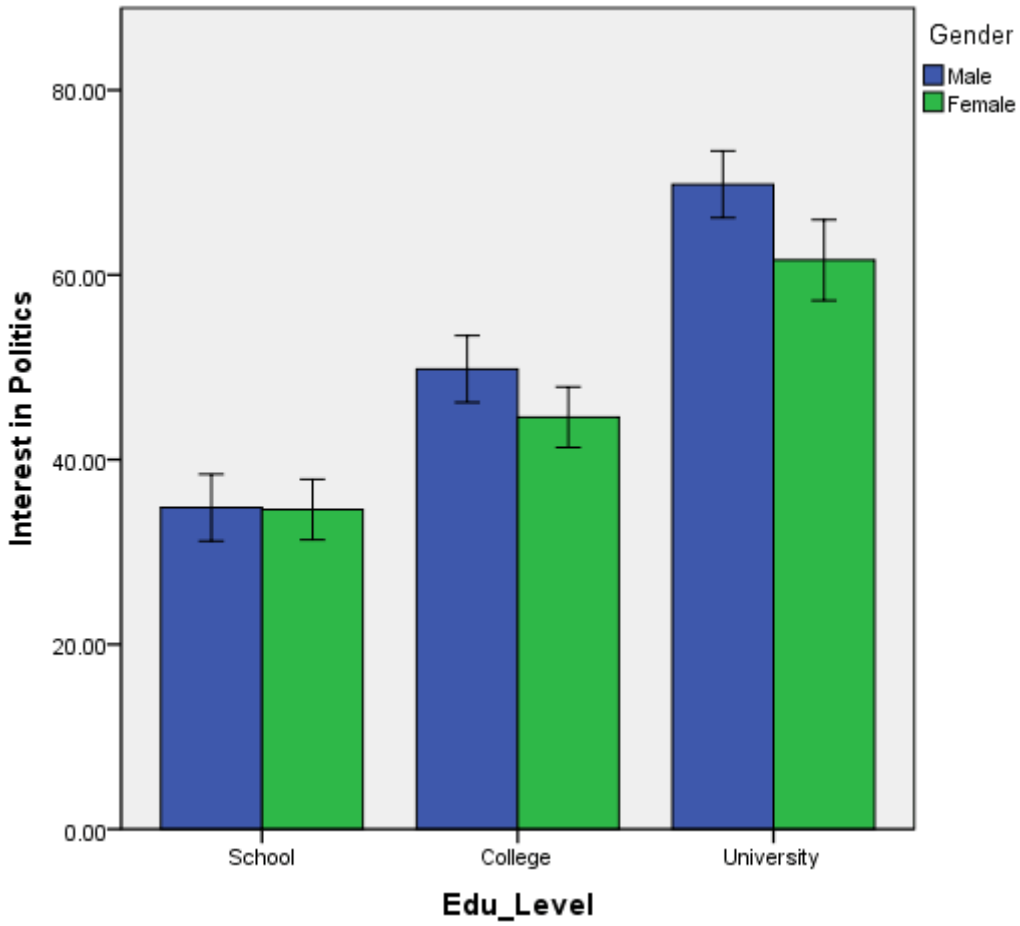


If you make any changes remember to click the Apply button.

10. Click the OK button.

Output

You will be presented with the following output:



A Simple Scatterplot using SPSS

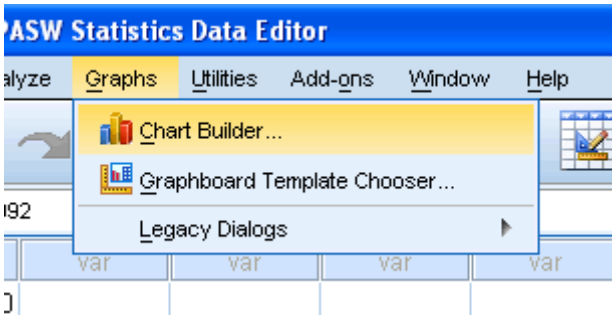
7

Objectives

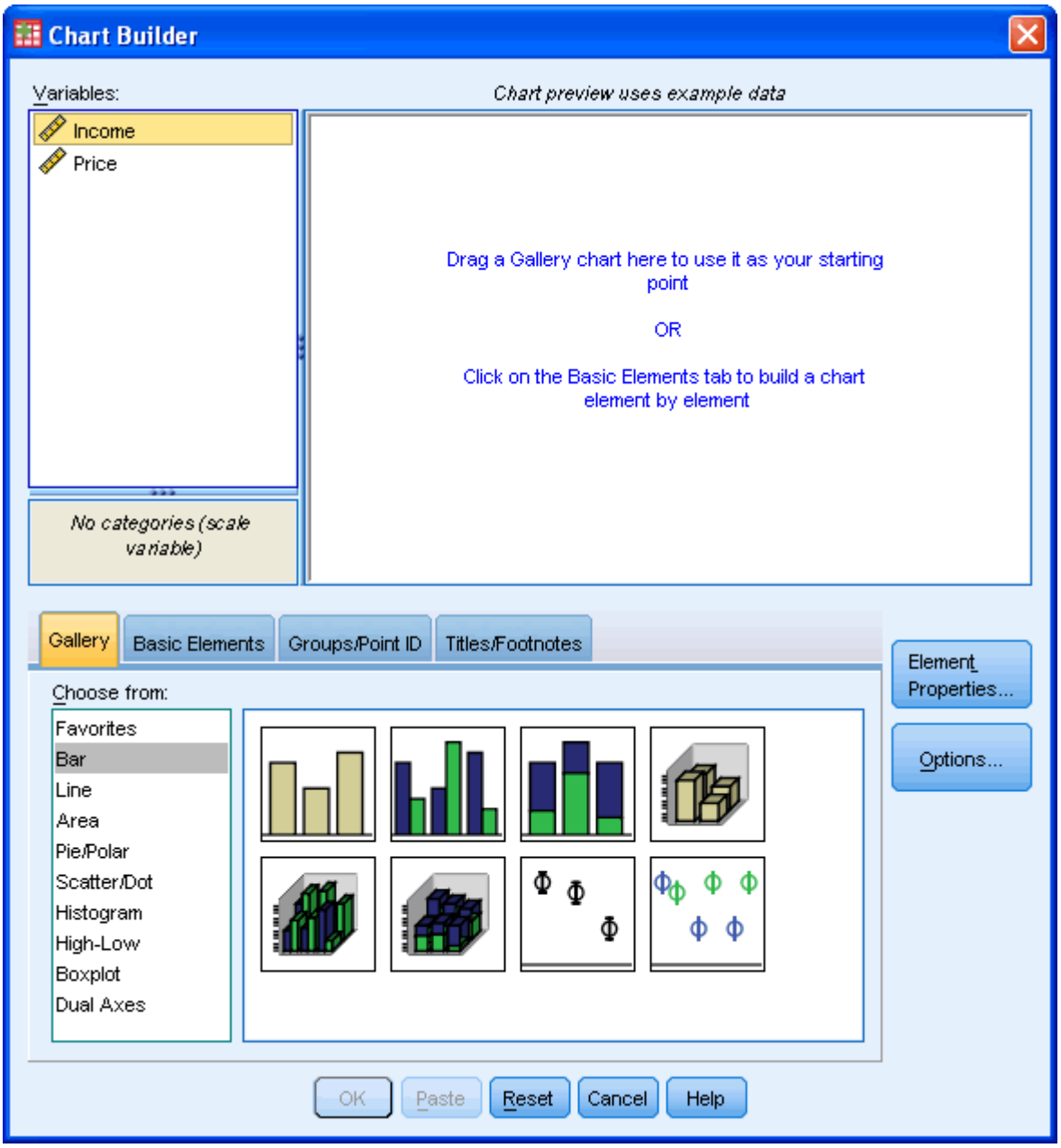
A scatterplot can be used to detemine whether a relationship is linear, detect outliers and graphically present a relationship. Determining whether a relationship is linear is an important assumption of correlation and simple regression. The example presented here is the same as in our guide on simple regression.

Procedure

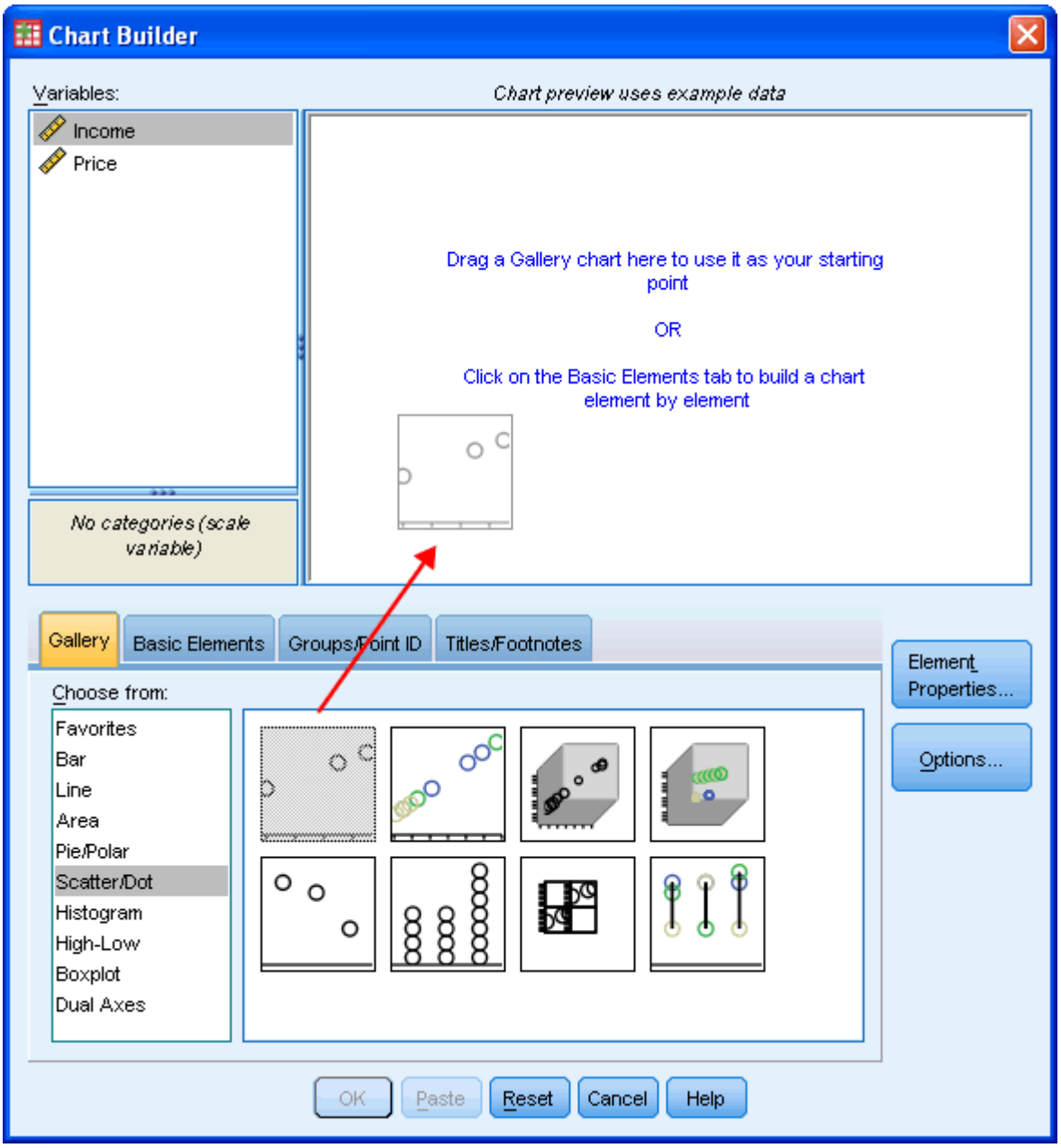
1. Click **Graphs > Chart Builder...** on the top menu as shown below:



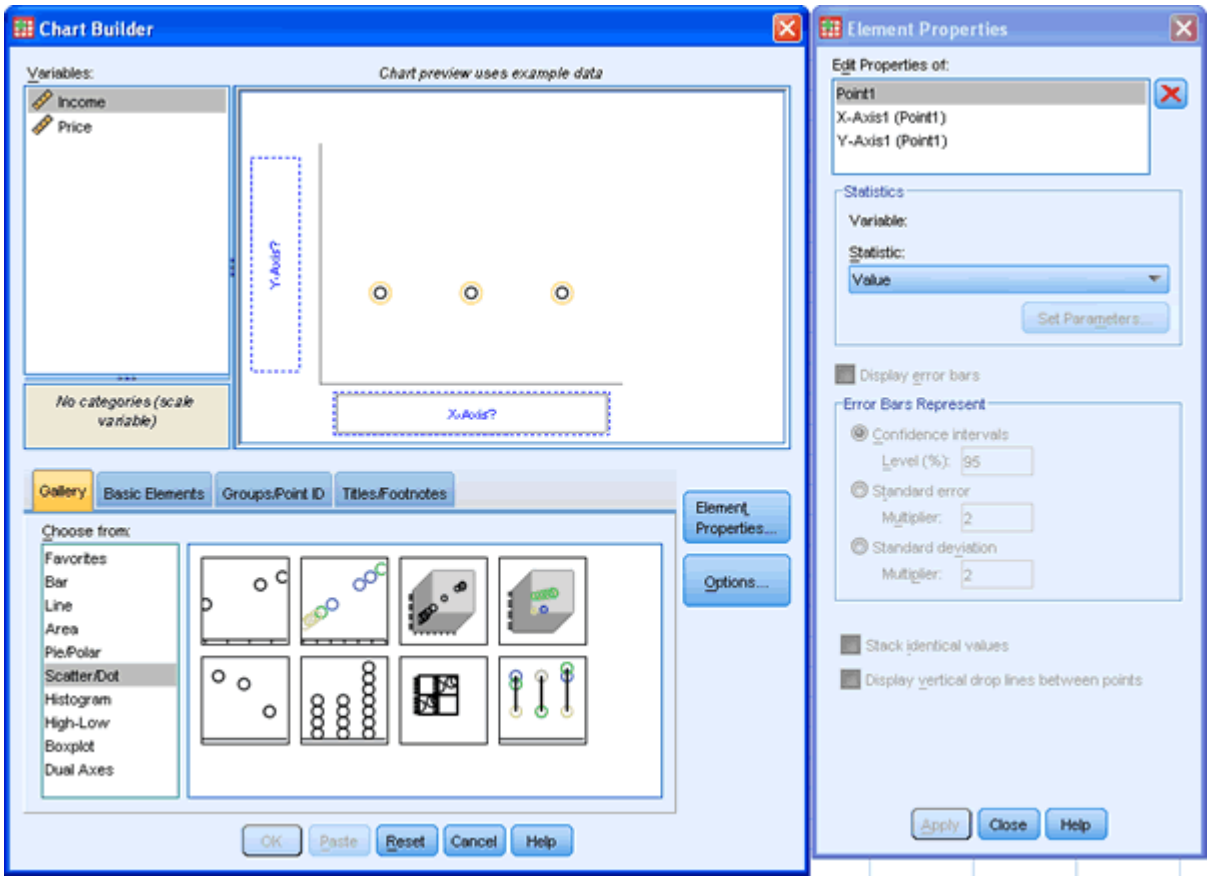
2. You will be presented with the following screen:



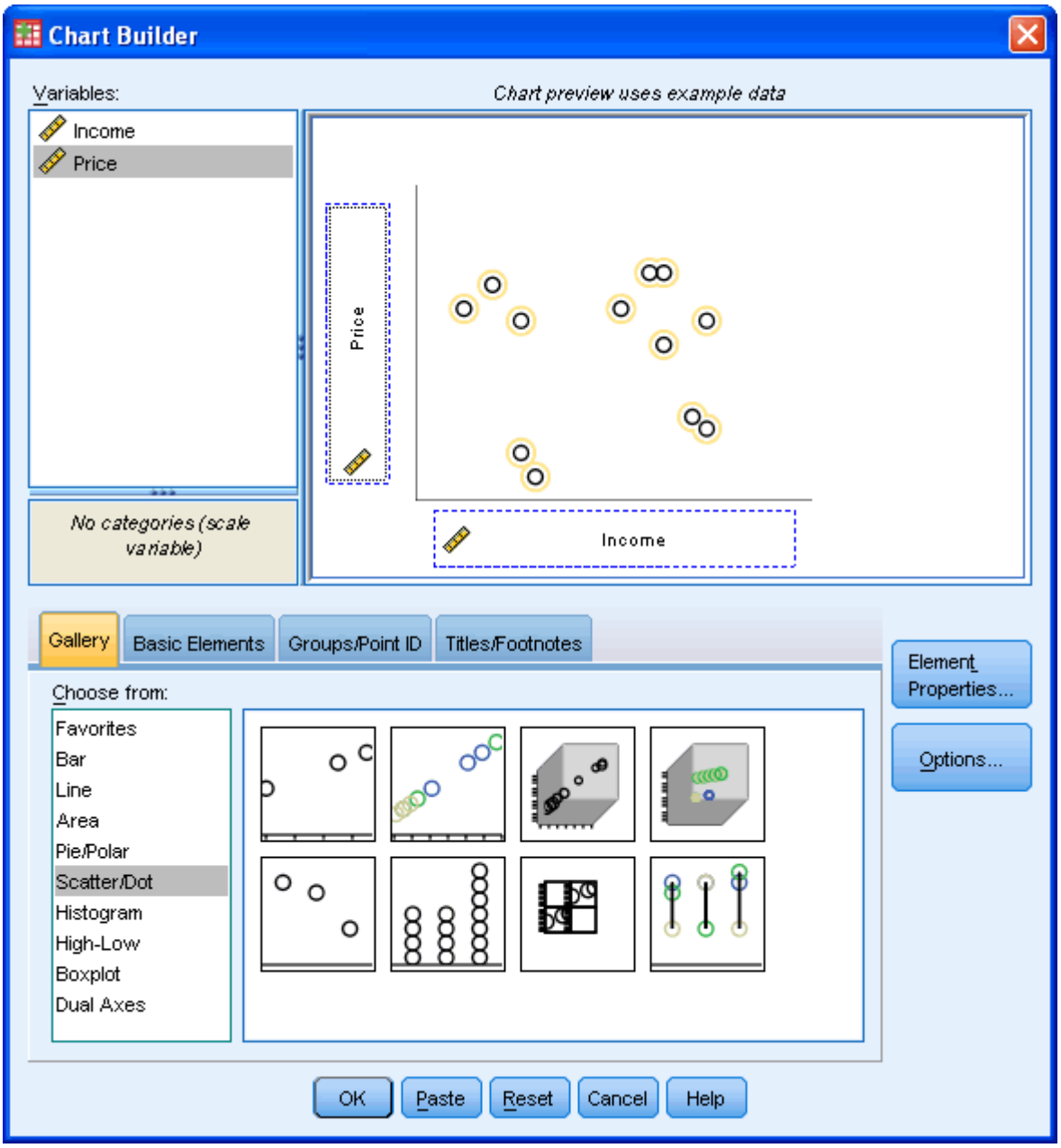
3. Under the Gallery Tab (Gallery) select the **Scatter/Dot** option and the **simple scatterplot icon** (top-left icon). Drag-and-drop this icon into the **Chart Preview Area** (as shown by the → in the diagram below).



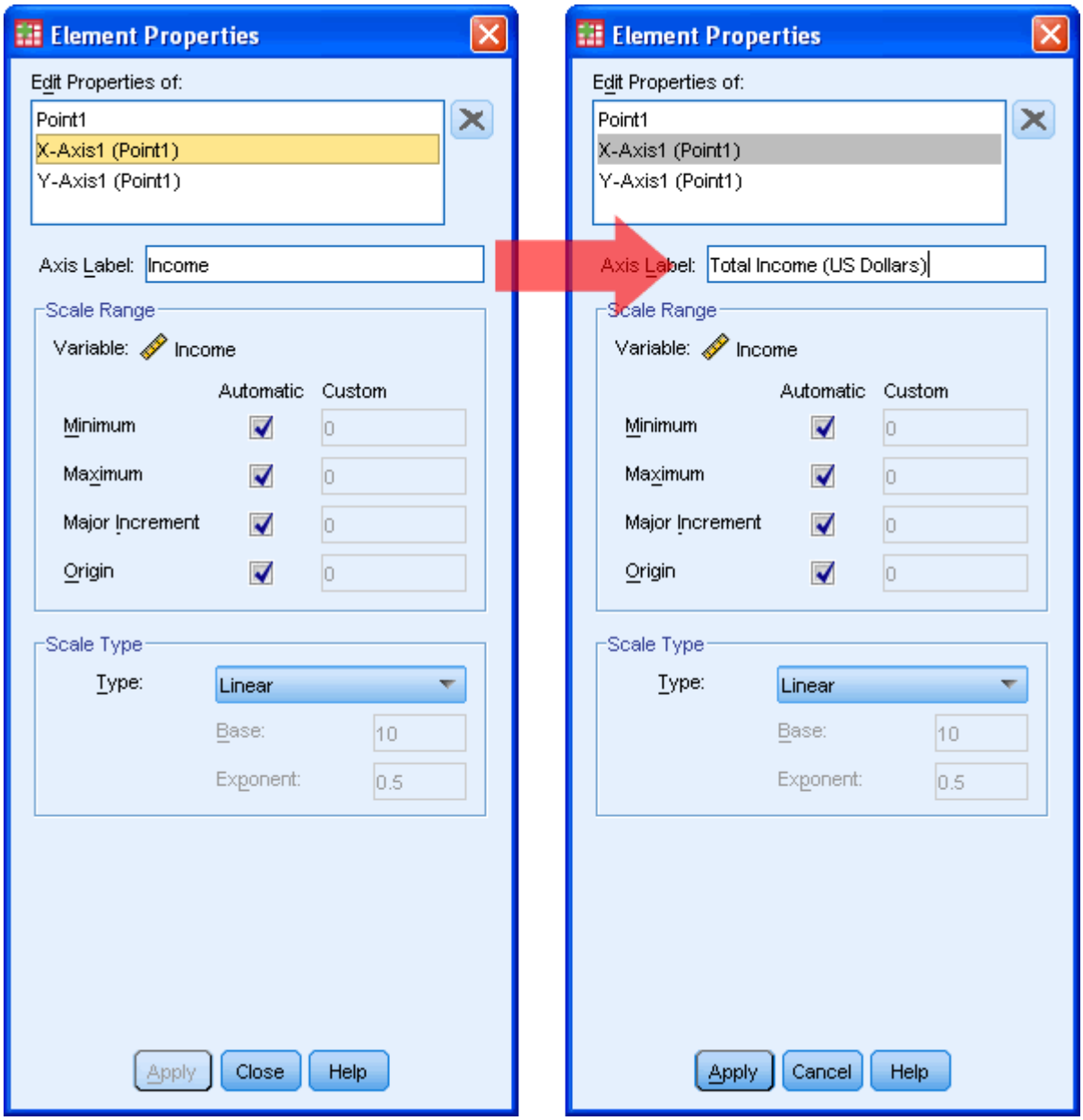
4. You will be presented with the following dialog boxes: **Chart Builder** and **Element Properties**. As you can see, the **Chart Preview Area** has been populated with a template of a simple scatterplot.



5. Transfer the independent (predictor) variable, Income, into the "**X-Axis?**" box and the dependent (outcome) variable into the "**Y-Axis?**" box within the **Preview Chart Area** by drag-and-dropping the variables from the "Variables:" box.



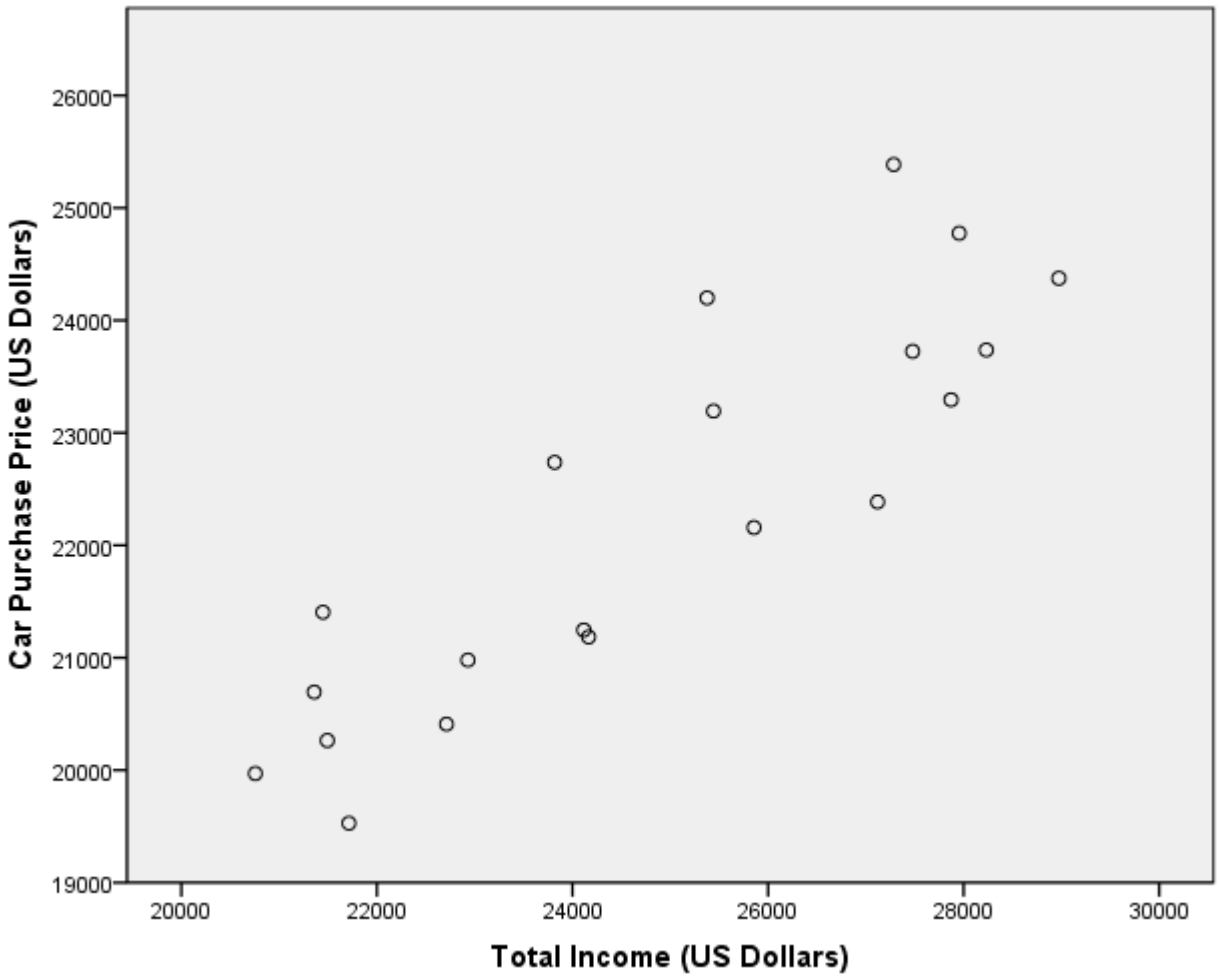
6. If you wish to change the axis labels then select "**X-Axis1 (Point1)**" and/or "**Y-Axis1 (Point1)**" in the **Element Properties** dialog box and type in the new axis title. Below we demonstrate changing the X-Axis label from "Income" to "Total Income (US Dollars)". Remember to click the Apply button after each label change. Repeat with the Y-Axis by changing the label from "Price" to "Car Purchase Price (US Dollars)" and, again, click the Apply button.

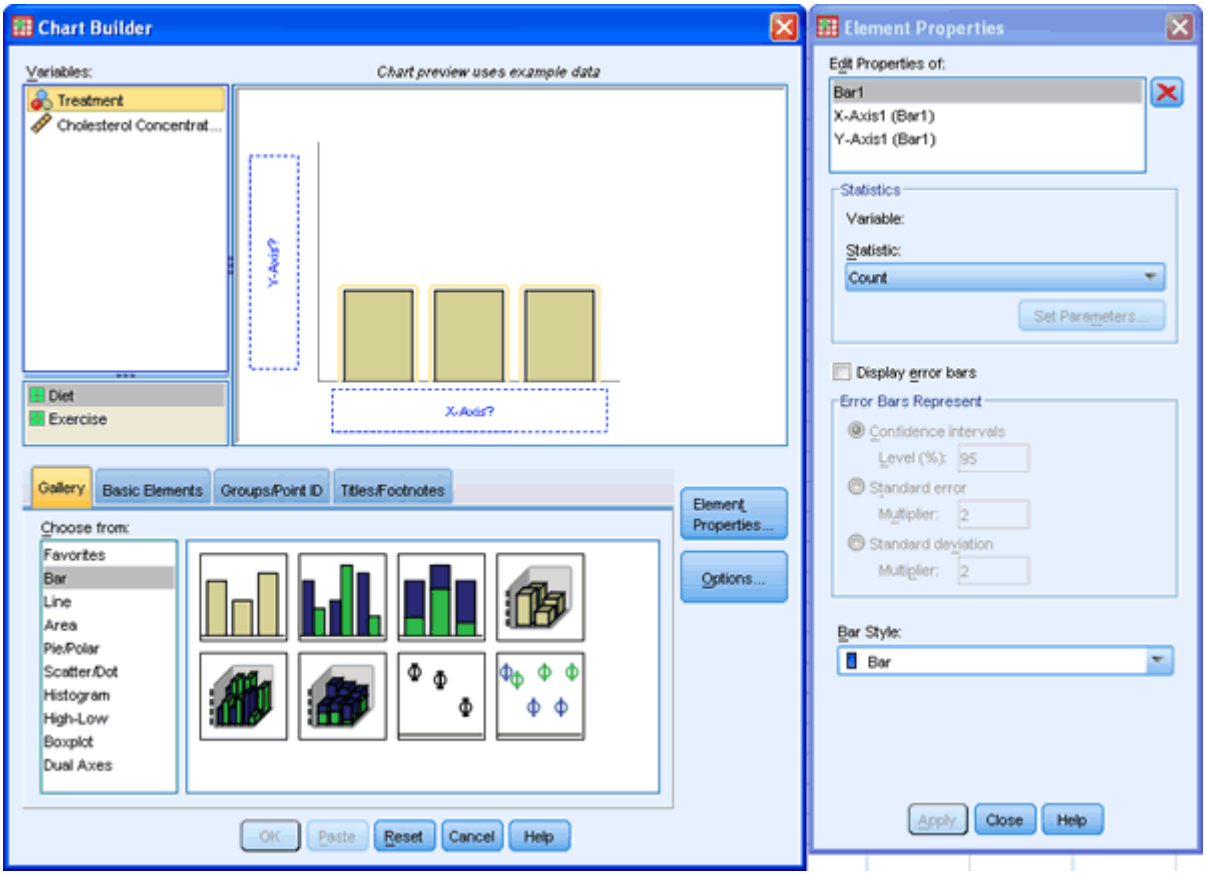


7. Click the OK button.

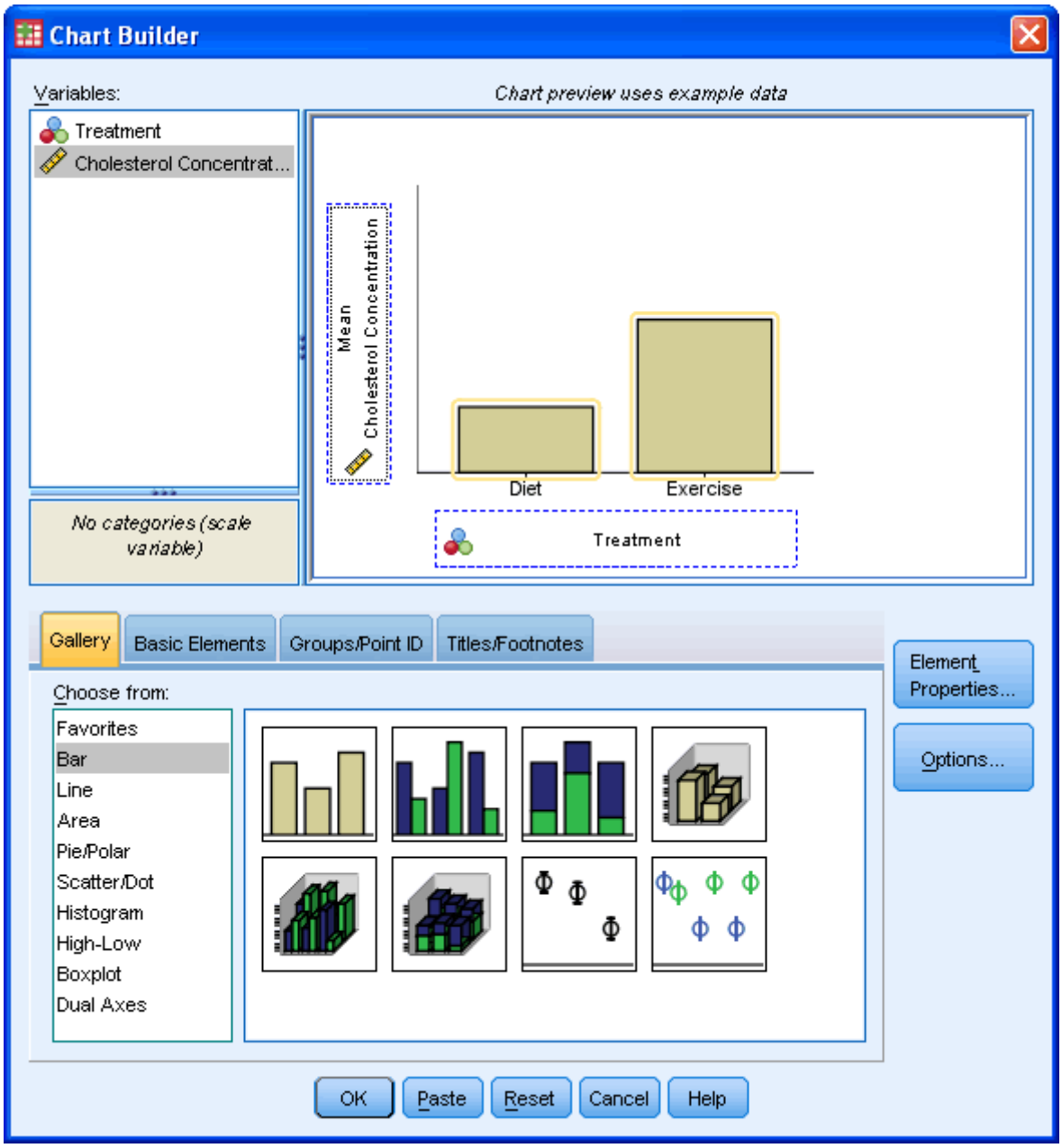
Output

You will be presented with the following output:


Car Purchase Price (US Dollars)
26000
25000
24000
23000
22000
21000
20000
19000
20000
22000
24000
26000
28000
30000
Total Income (US Dollars)



5. Transfer the independent variable, Treatment, into the "**X-Axis?**" box and the dependent (outcome) variable, Cholesterol Concentration, into the "**Y-Axis?**" box within the **Preview Chart Area** by drag-and-dropping the variables from the "Variables:" box.



Checking Assumptions

Testing for Normality using SPSS

Introduction

An assessment of the normality of data is a prerequisite for many statistical tests as normal data is an underlying assumption in parametric testing. There are two main methods of assessing normality - graphically and numerically.

This guide will help you to determine whether your data is normal and, therefore, that this assumption is met in your data for statistical tests. The approaches can be divided into two main themes - relying on statistical tests or visual inspection. Statistical tests have the advantage of making an objective judgement of normality but are disadvantaged by sometimes not being sensitive enough at low

sample sizes or overly sensitive to large sample sizes. As such, some statisticians prefer to use their experience to make a subjective judgement about the data from plots/graphs. Graphical interpretation has the advantage of allowing good judgement to assess normality in situations when numerical tests might be over or under sensitive but graphical methods do lack objectivity. If you do not have a great deal of experience interpreting normality graphically then it is probably best to rely on the numerical methods.

If you would like to see an example of how to test for normality in SPSS as part of a complete statistical analysis, you can see this in action as part of our enhanced Independent-samples t-test in SPSS guide. To go straight to the relevant section, click here. This is a free, complete example of an enhanced guide in our Premium section. You can check out our low prices for access to all the enhanced content in our Premium section here.

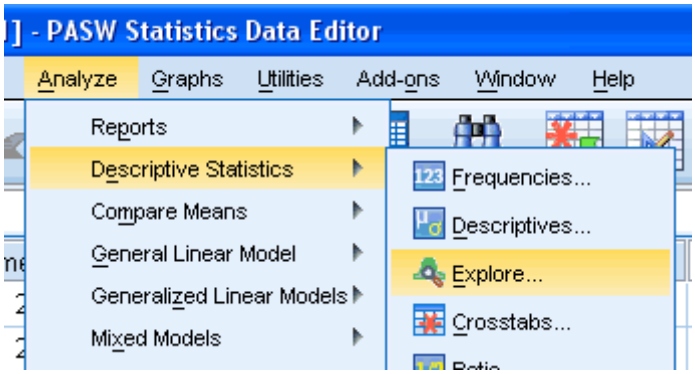
Methods of assessing normality

SPSS allows you to test all of these procedures within **Explore...** command. The **Explore...** command can be used in isolation if you are testing normality in one group or splitting your dataset into one or more groups. For example, if you have a group of participants and you need to know if their height is normally distributed then everything can be done within the **Explore...** command. If you split your group into males and females (i.e. you have a categorical independent variable) then you can test for normality of height within both the male group and the female group using just the **Explore...** command. This applies even if you have more than two groups. However, if you have 2 or more categorical, independent variables then the **Explore...** command on its own is not enough and you will have to use the **Split File...** command also.

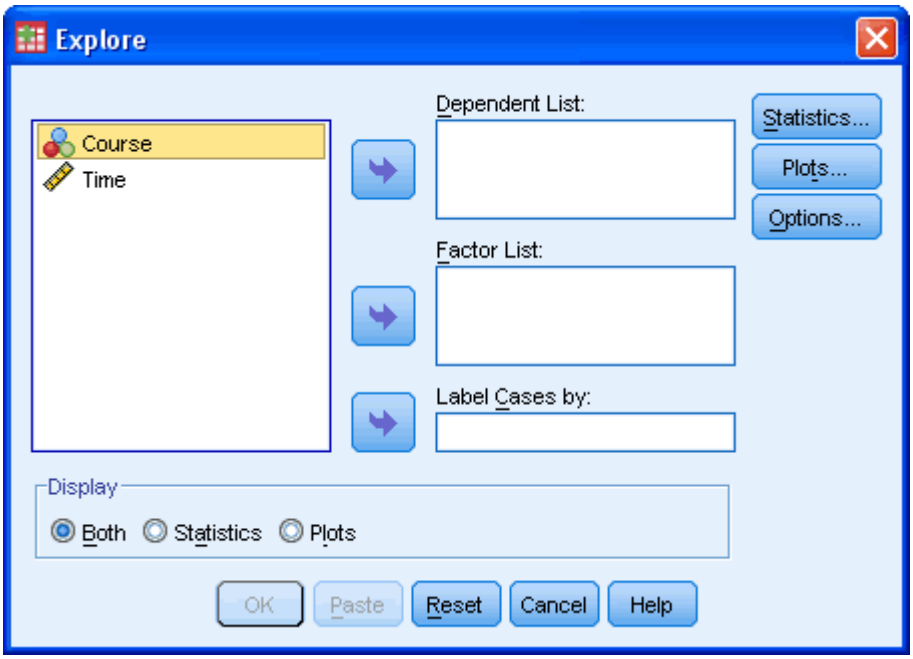
Procedure for none or one grouping variable

The following example comes from our guide on how to perform a one-way ANOVA in SPSS.

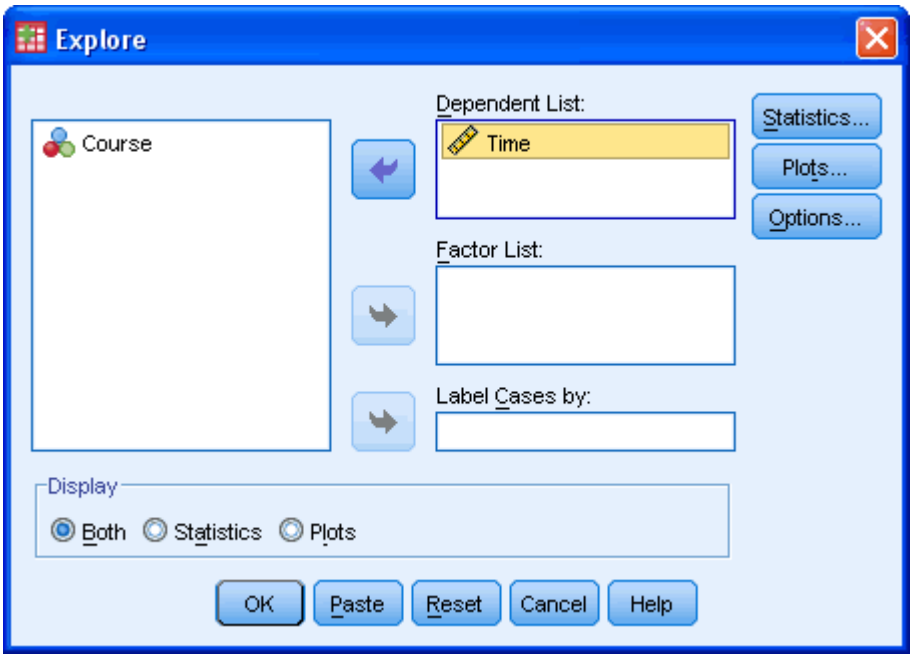
1. Click **Analyze > Descriptive Statistics > Explore...** on the top menu as shown below:



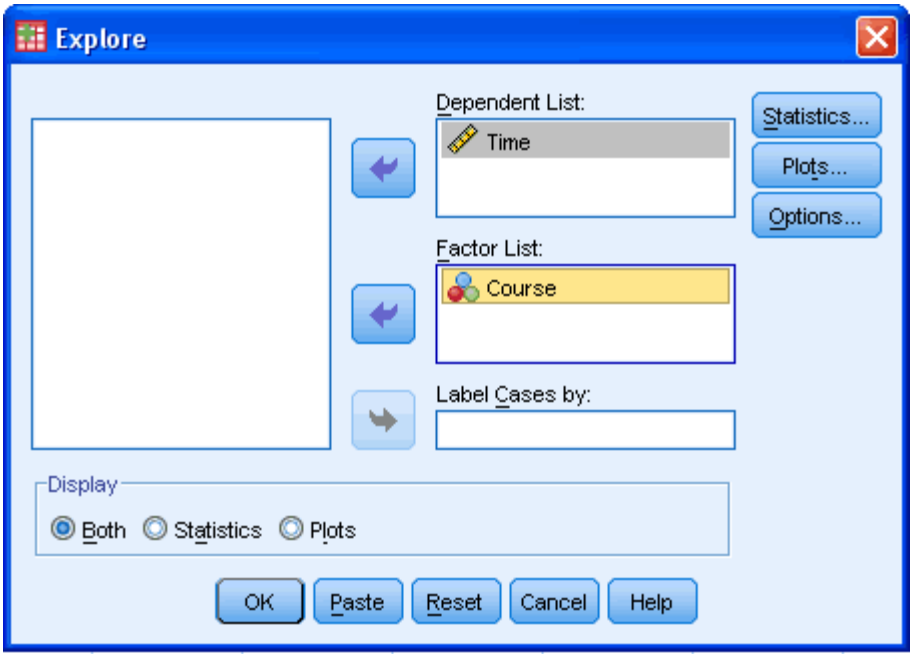
2. You will be presented with the following screen:



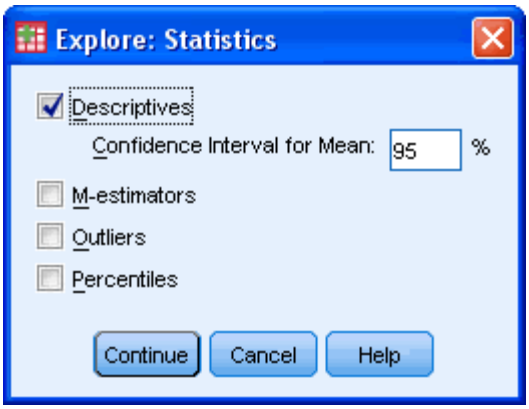
3. Transfer the variable that needs to be tested for normality into the "Dependent List:" box by either drag-and-dropping or using the button. In this example, we transfer the "Time" variable into the "Dependent List:" box. You will then be presented with the following screen:



4. [Optional] If you need to establish if your variable is normally distributed for each level of your independent variable then you need to add your independent variable to the "Factor List:" box by either drag-and-dropping or using the button. In this example, we transfer the "Course" variable into the "Factor List:" box. You will be presented with the following screen:

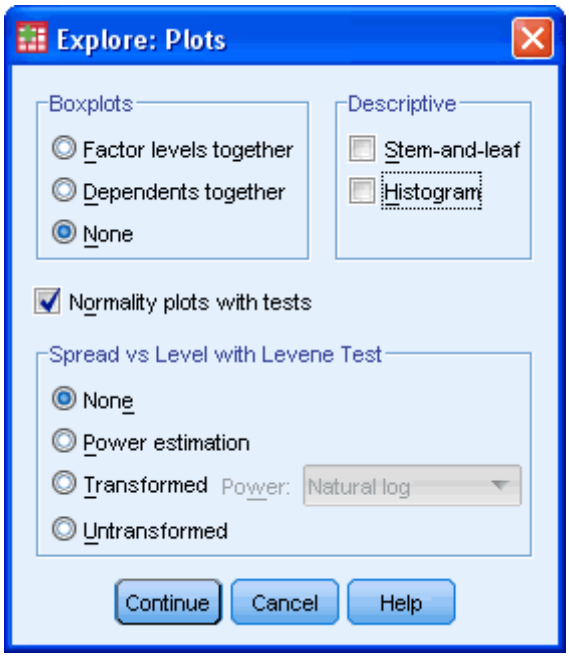


5. Click the Statistics... button. You will be presented with the following screen:



Leave the above options unchanged and click the Continue button.

6. Click the Plots... button. Change the options so that you are presented with the following screen:



Click the Continue button.

7. Click the OK button.

Output

SPSS outputs many table and graphs with this procedure. One of the reasons for this is that the **Explore...** command is not used solely for the testing of normality but in describing data in many different ways. When testing for normality, we are mainly interested in the **Tests of Normality** table and the **Normal Q-Q Plots**, our numerical and graphical methods to test for the normality of data, respectively.

Shapiro-Wilk Test of Normality

Tests of Normality

Course		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Time	Beginner	.177	10	.200*	.964	10	.827
	Intermediate	.166	10	.200*	.969	10	.882
	Advanced	.151	10	.200*	.965	10	.837

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

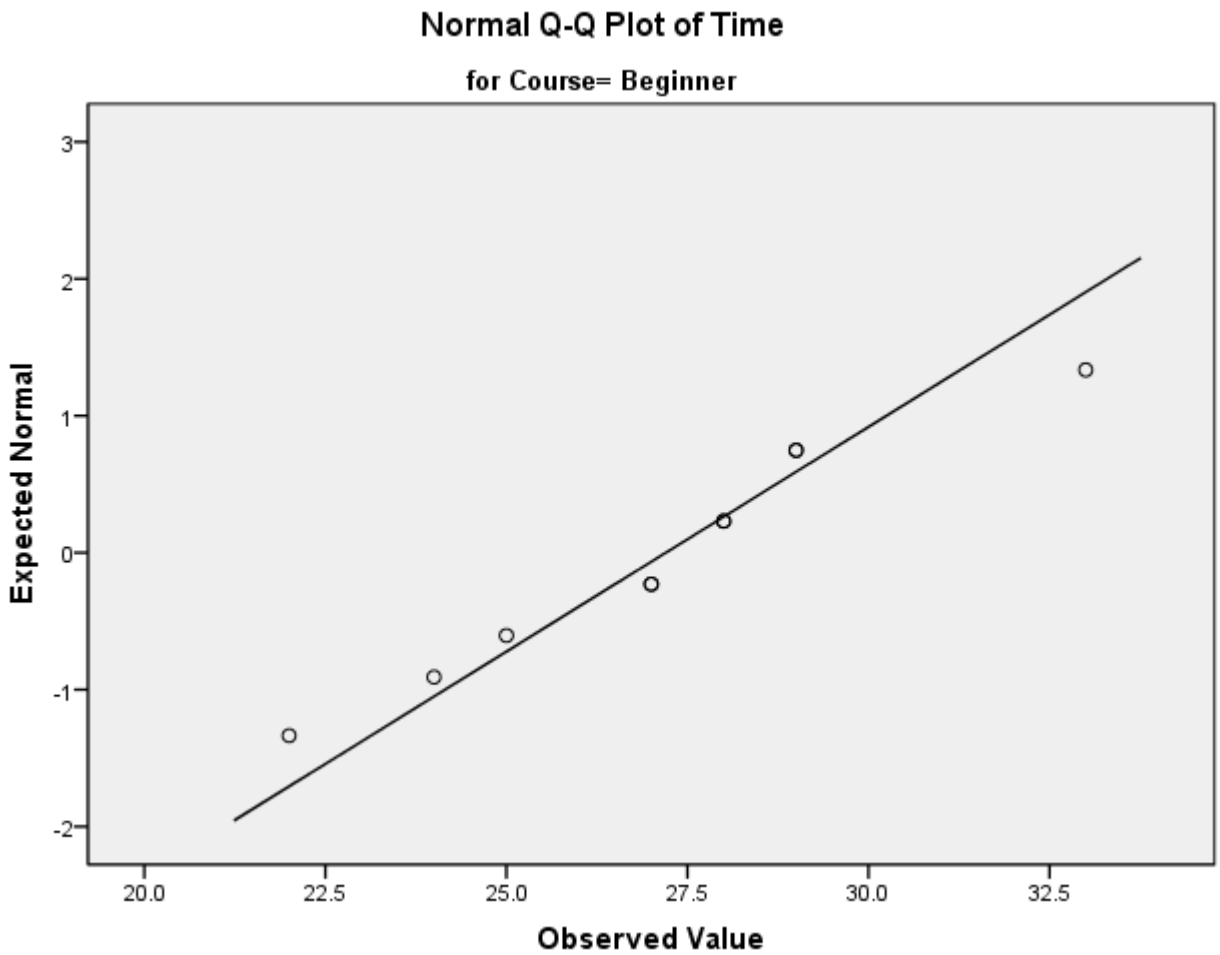
The above table presents the results from two well-known tests of normality, namely the Kolmogorov-Smirnov Test and the Shapiro-Wilk Test. We Shapiro-Wilk Test is more appropriate for small sample sizes (< 50 samples) but can also handle sample sizes as large as 2000. For this reason, we will use the Shapiro-Wilk test as our numerical means of assessing normality.

We can see from the above table that for the "Beginner", "Intermediate" and "Advanced" Course Group the dependent variable, "Time", was normally distributed. How do we know this? If the **Sig.** value of the Shapiro-Wilk Test is greater the 0.05 then the data is normal. If it is below 0.05 then the data significantly deviate from a normal distribution.

If you need to use skewness and kurtosis values to determine normality, rather the Shapiro-Wilk test, you will find these in our upgraded Premium SPSS guide. Check out our low prices here.

Normal Q-Q Plot

In order to determine normality graphically we can use the output of a normal Q-Q Plot. If the data are normally distributed then the data points will be close to the diagonal line. If the data points stray from the line in an obvious non-linear fashion then the data are not normally distributed. As we can see from the normal Q-Q plot below the data is normally distributed. If you at all unsure of being able to correctly interpret the graph then rely on the numerical methods instead as it can take a fair bit of experience to correctly judge the normality of data based on plots.



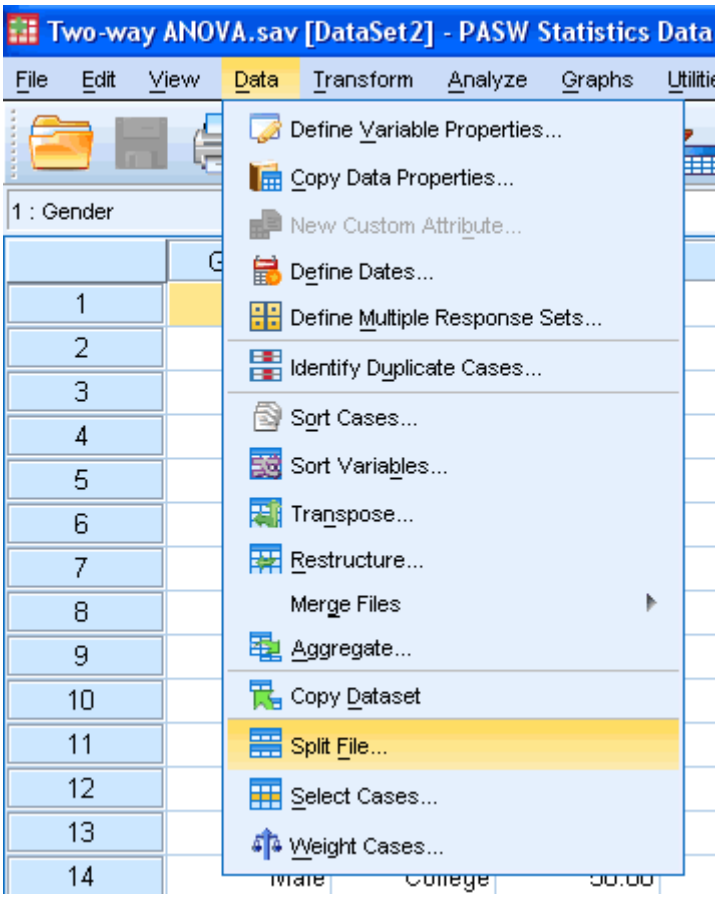
If you need to know what Normal Q-Q Plots look like when distributions are not normal (e.g. negatively skewed), you will find these in our upgraded Premium SPSS guide. Check out our low prices here.

Testing for Normality using SPSS (cont...)

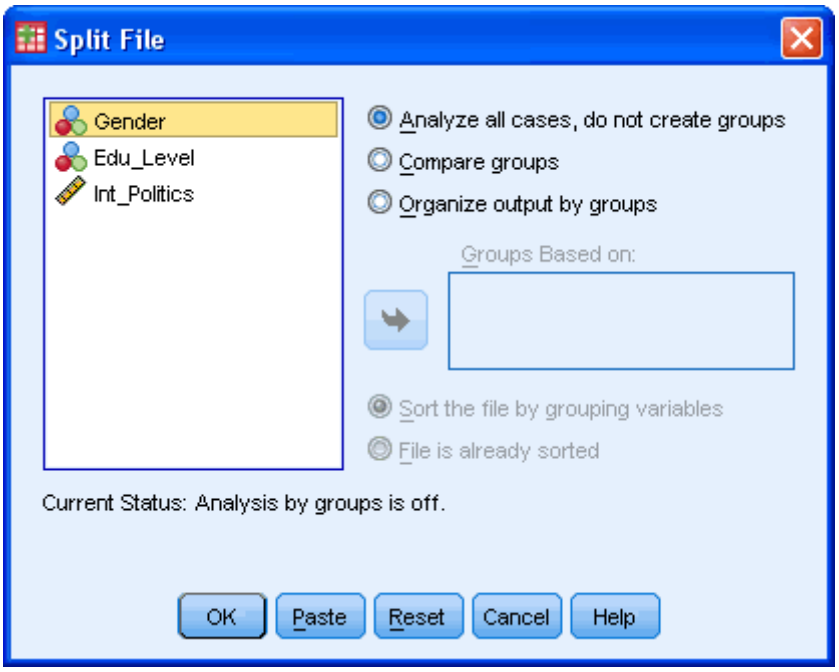
Procedure when there are two or more independent variables

The **Explore...** command on its own cannot separate the dependent variable into groups based on not one but two or more independent variables. However, we can perform this feat by using the **Split File...** command.

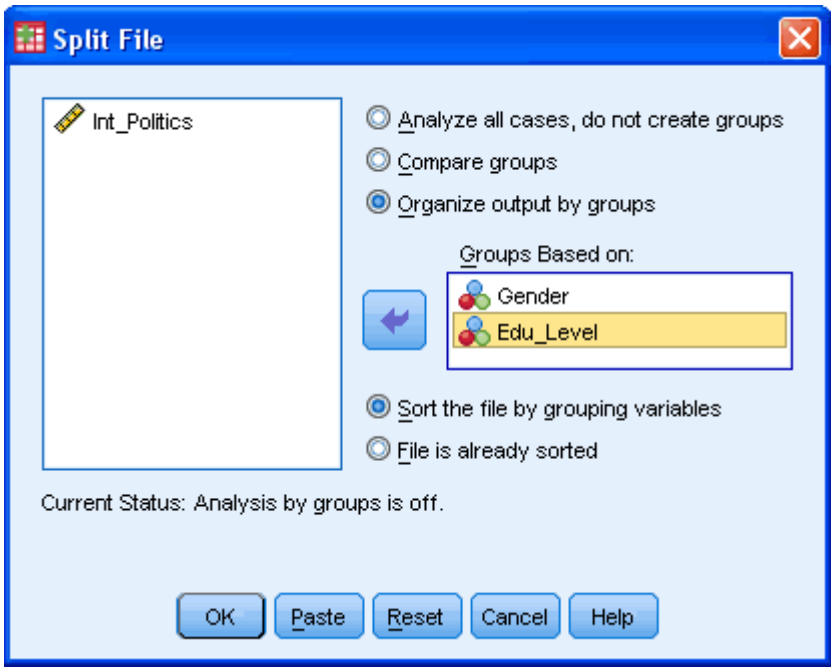
1. Click **Data > Split File...** on the top menu as shown below:



2. You will be presented with the following screen:



3. Click the radio option, "Organize output by groups". Transfer the independent variables you wish to categorize the dependent variable on into the "Groups Based on:". In this example, we want to know whether interest in politics (Int_Politics) is normally distributed when grouped/categorized by Gender AND Edu_Level (education level). You will be presented with the following screen:

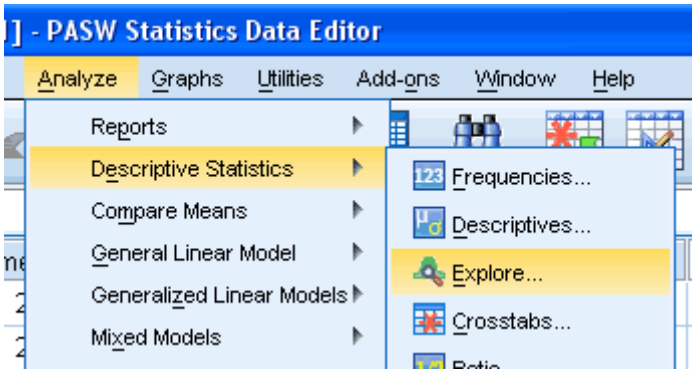


Click the

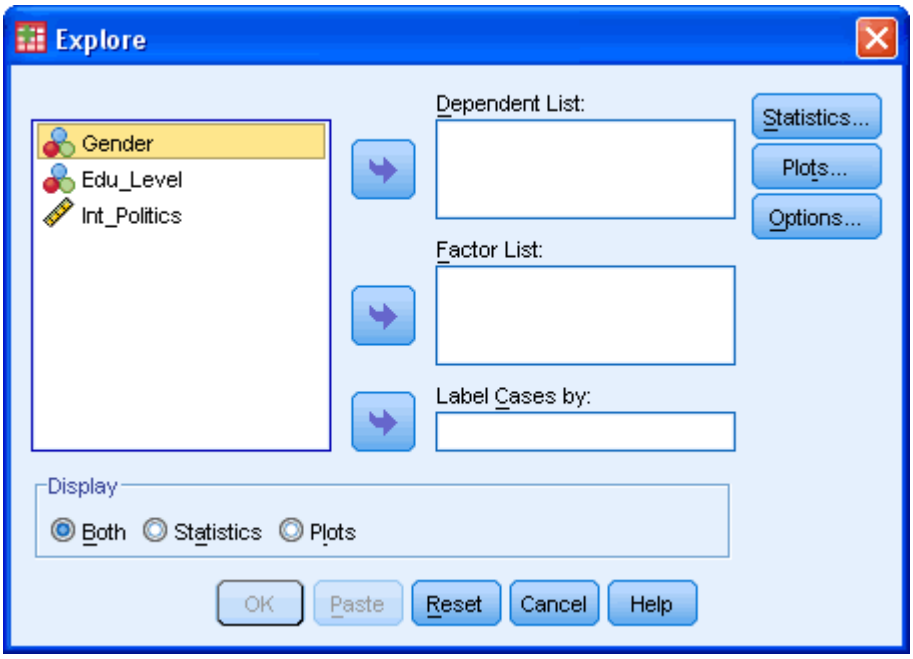
button.

[Your file is now split and the output from any tests will be organized into the groups you have selected.]

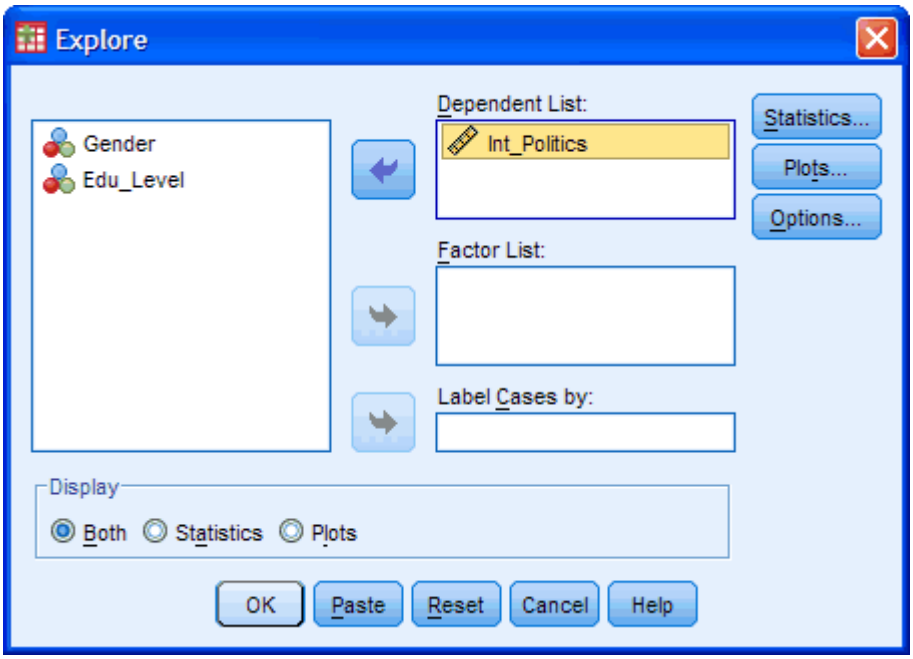
4. Click **Analyze > Descriptive Statistics > Explore...** on the top menu as shown below:



5. You will be presented with the following screen:

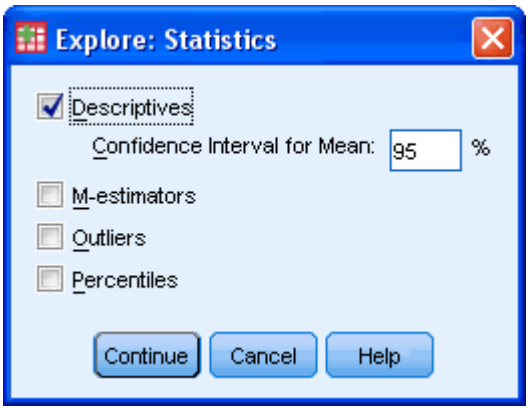


6. Transfer the variable that needs to be tested for normality into the "Dependent List:" box by either drag-and-dropping or using the button. In this example, we transfer the "Int_Politics" variable into the "Dependent List:" box. You will then be presented with the following screen:



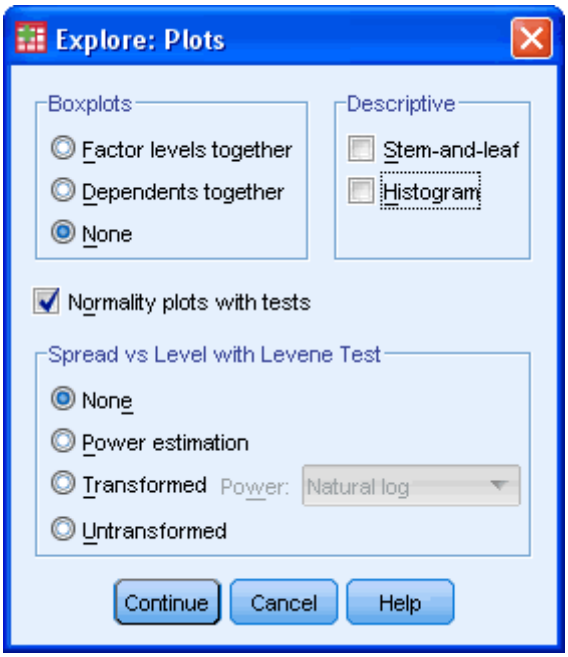
[*There is no need to transfer the independent variables "Gender" and "Edu_Level" into the "Factor List:" box as this has been accomplished with the* ***Split File...*** *command. Why not simply transfer these two independent variables into the "Factor List:" box? Because this will not achieve the desired result. It will first analyse "Int_Politics" for normality with respect to "Gender" and then with respect to "Edu_Level". It does NOT analyse "Int_Politics" for normality by grouping individuals into both "Gender" and "Edu_Level" AT THE SAME TIME.*]

7. Click the Statistics... button. You will be presented with the following screen:



Leave the above options unchanged and click the Continue button.

8. Click the Plots... button. Change the options so that you are presented with the following screen:



Click the Continue button.

9. Click the OK button.

Output

You will now see that the output has been split into separate sections based on the combination of groups of the two independent variables. As an example we show the tests of normality when the dependent variable, "**Int_Politics**", is categorized into the first "**Gender**" group (male) and first "**Edu_Level**" group (School). All other possible combinations are also presented in the full output but we will not shown them here for clarity.

Gender = Male, Edu_Level = School

Under this above category you are presented with the **Tests of Normality** table as shown below:

Tests of Normality[b]

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Int_Politics	.178	10	.200*	.944	10	.599

a. Lilliefors Significance Correction

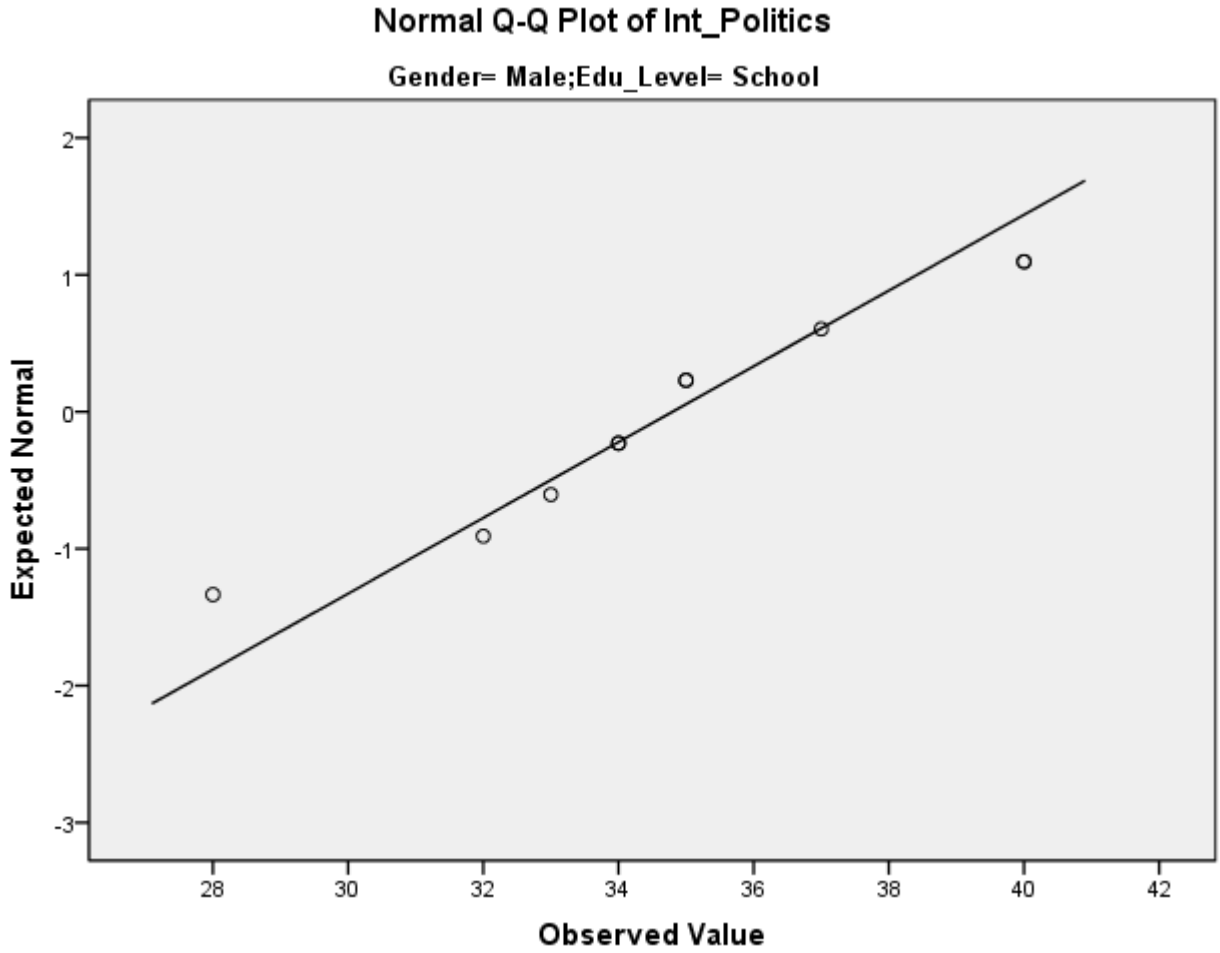
*. This is a lower bound of the true significance.

b. Gender = Male, Edu_Level = School

The Shapiro-Wilk test is now analyzing the normality of "Int_Politics" on the data of those individuals that are classified as both "male" in the independent variable "Gender" and "school" in the independent variable "Edu_Level". As the **Sig.** value

under the Shapiro-Wilk column is greater than 0.05 we can conclude that "Int_Politics" for this particular subset of individuals is normally distributed.

The same data from the same individuals are now also being analyzed to produce a Normal Q-Q Plot as below. From this graph we can conclude that the data appears to be normally distributed as it follows the diagonal line closely and does not appear to have a non-linear pattern.



Predicting Scores

Linear Regression Analysis using SPSS

Objectives

Regression analysis is the next step up after correlation; it is used when we want to predict the value of a variable based on the value of another variable. In this case, the variable we are using to predict the other variable's value is called the independent variable or sometimes the predictor variable. The variable we are wishing to predict is called the dependent variable or sometimes the outcome variable.

Example

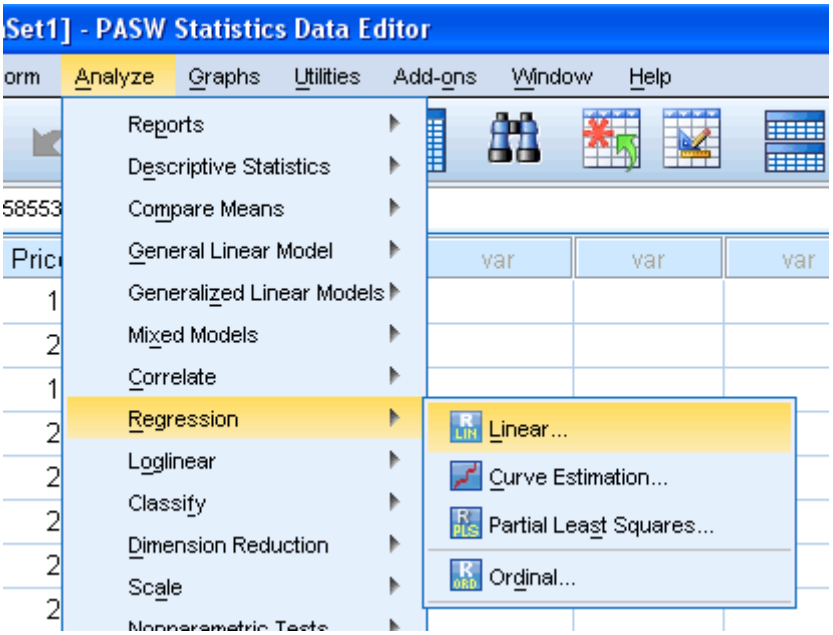
A saleman for a large car brand is interested in determining whether there is a relationship between an individual's income and the price they pay for a car. They will use this information to determine which cars to offer potential customers in new areas where average income is known.

Assumptions

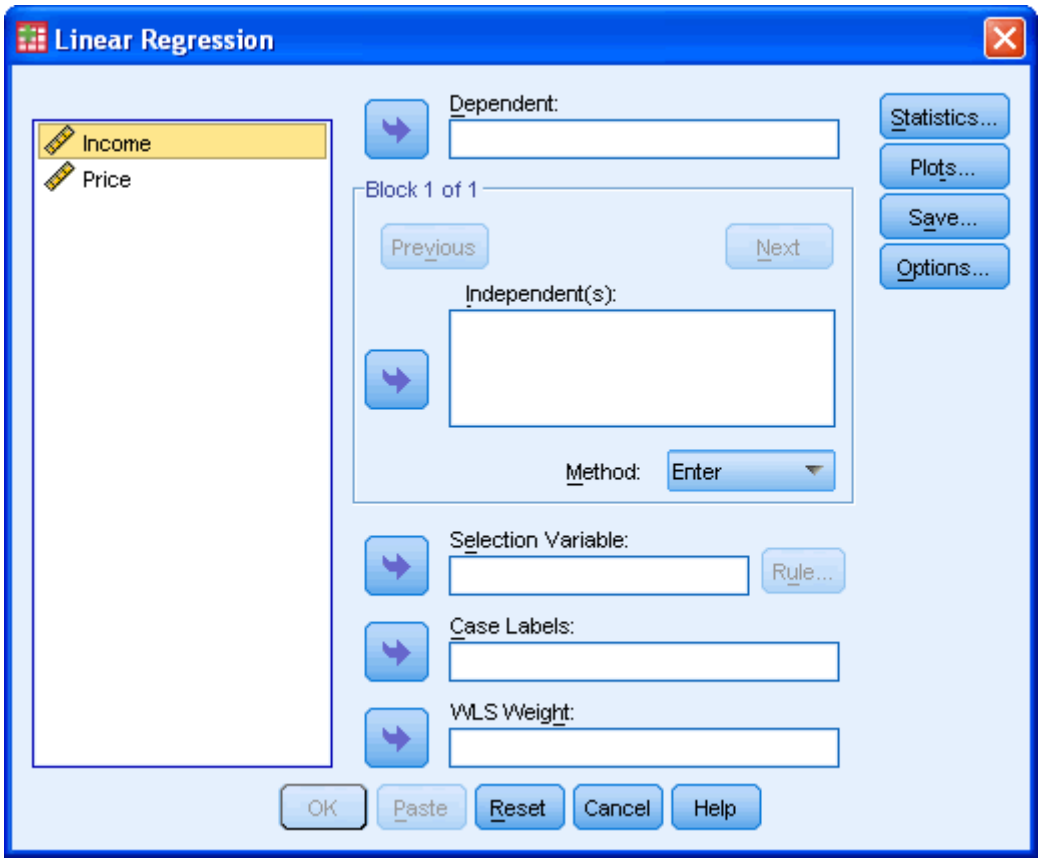
- Variables are measured **at the interval or ratio level (continuous)** (see Types of Variable guide).
- Variables are **approximately normally distributed** (see Testing for Normality guide).
- There is a **linear relationship** between the two variables.

Procedure

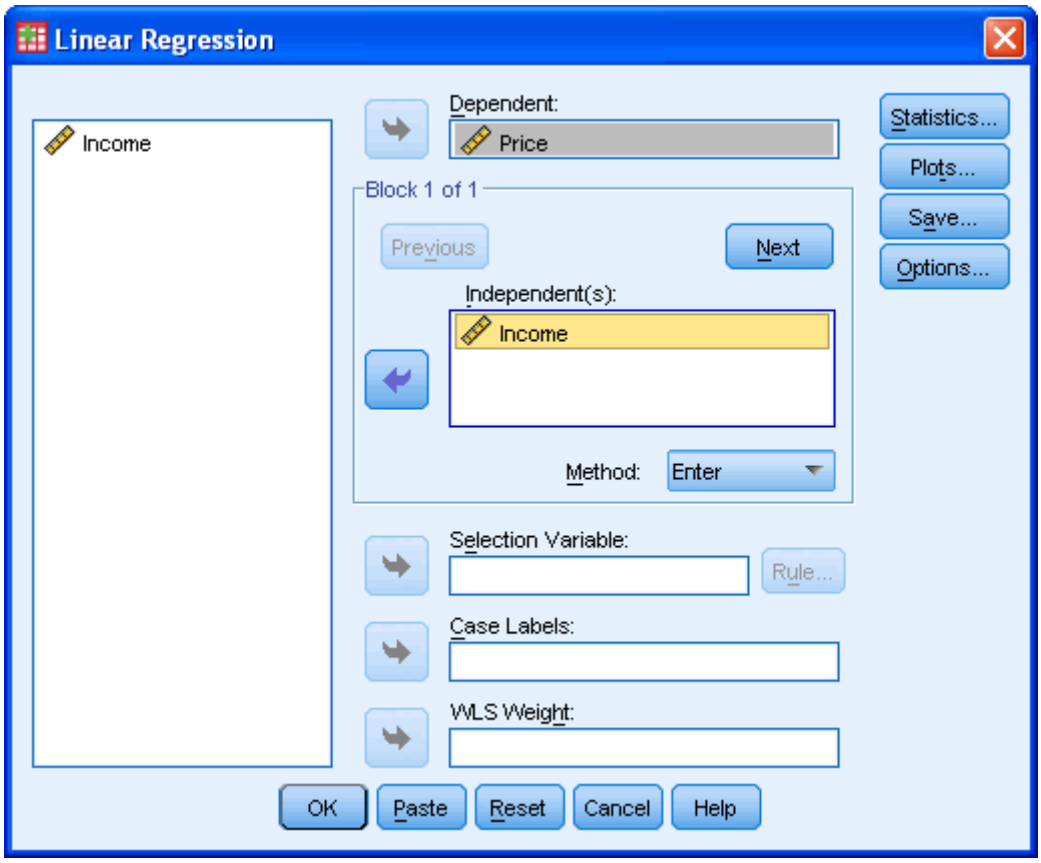
1. Click **Analyze > Regression > Linear...** on the top menu.



2. You will be presented with the following dialog box:



3. Transfer the independent (predictor) variable, Income, into the "Independent(s):" box and the dependent (outcome) variable, Price, into the "Dependent:" box. You can do this by either drag-and-dropping or by using the buttons.



4. Click the button.

Output of Linear Regression Analysis

SPSS will generate quite a few tables in its results section for a linear regression. In this session, we are going to look at the important tables. The first table of interest is the **Model Summary** table. This table provides the R and R^2 value. The R value is 0.873, which represents the simple correlation and, therefore, indicates a high degree of correlation. The R^2 value indicates how much of the dependent variable, price, can be explained by the independent variable, income. In this case, 76.2% can be explained, which is very large.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.873[a]	.762	.749	874.779

a. Predictors: (Constant), Income

The next table is the **ANOVA** table. This table indicates that the regression model predicts the outcome variable significantly well. How do we know this? Look at the "Regression" row and go to the **Sig.** column. This indicates the statistical significance of the regression model that was applied. Here, $P < 0.0005$ which is less than 0.05 and indicates that, overall, the model applied is significantly good enough in predicting the outcome variable.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.418E7	1	4.418E7	57.737	.000[a]
	Residual	1.377E7	18	765238.393		
	Total	5.796E7	19			

a. Predictors: (Constant), Income

b. Dependent Variable: Price

The table below, **Coefficients**, provides us with information on each predictor variable. This provides us with the information necessary to predict price from income. We can see that both the constant and income contribute significantly to the model (by looking at the **Sig.** column). By looking at the **B** column under the **Unstandardized Coefficients** column we can present the regression equation as:

Price = 8287 + 0.564(Income)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8286.786	1852.256		4.474	.000
	Income	.564	.074	.873	7.598	.000

a. Dependent Variable: Price

Published with written permission from SPSS

Associations

Pearson's Product-Moment Correlation using SPSS

34

Objectives

The Pearson product-moment correlation coefficient is a measure of the strength and direction of association that exists between two variables measured on at least an interval scale. It is denoted by the symbol *r*. An introductory guide to this test is provided in our Statistical Guides section here and we recommend you read it if you are not familiar with this test.

Assumptions

- Variables are measured **at the interval or ratio level (continuous)** (see Types of Variable guide).
- Variables are **approximately normally distributed** (see Testing for Normality guide).
- There is a **linear relationship** between the two variables.
- Pearsons's *r* is sensitive to outliers so it is best if **outliers are kept to a minimum or there are no outliers**.

How to test the assumptions for a Pearson's correlation and what to do when you have any violations of these assumptions is fully explained in our upgraded Premium SPSS guide. Check out our low prices here.

Example

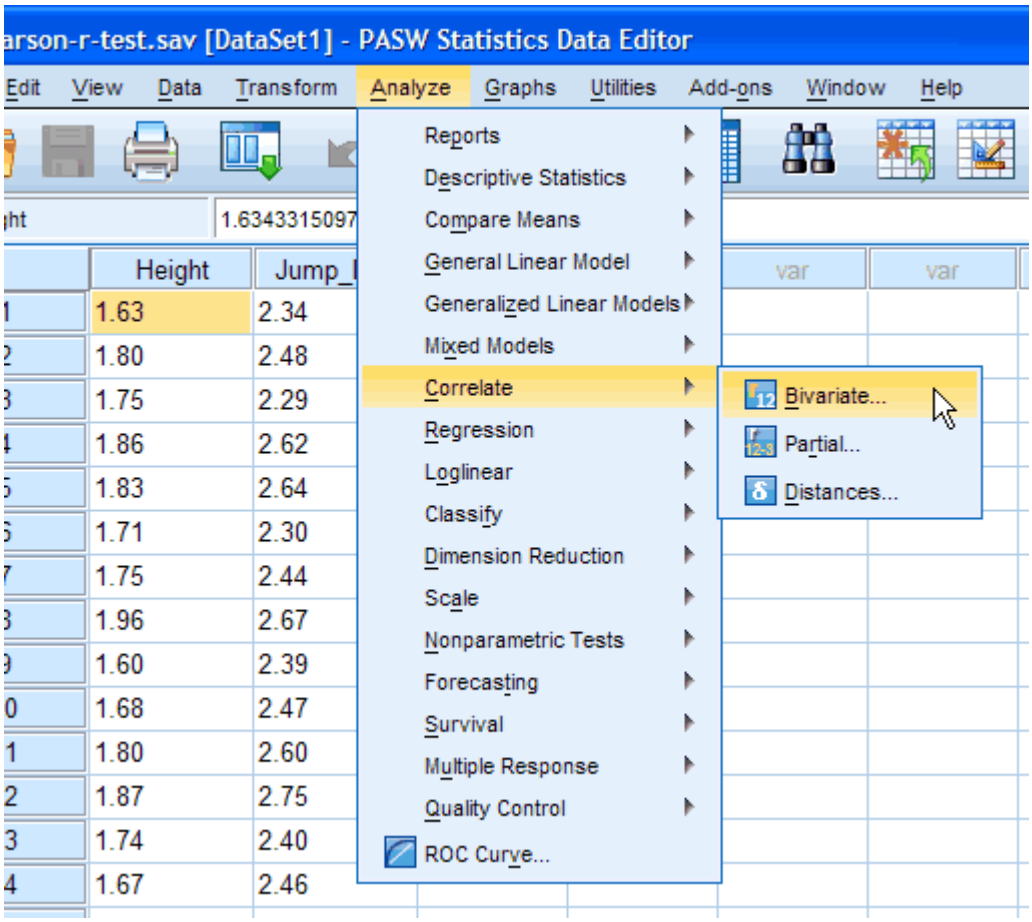
A researcher wishes to know whether a person's height is related to how well they perform in a long jump. The researcher recruited untrained individuals from the general population, measured their height and had them perform a long jump. The researcher then investigates whether there is an association between height and long jump performance.

Testing assumptions

Your variables need to be normally distributed. To determine whether your samples are normally distributed read our guide on Testing for Normality in SPSS. Pearson's r is also very susceptible to outliers in the data so you need to test for outliers. What if your samples are not normally distributed or there are outliers? If your samples violate the assumption of normality or have outliers then you might need to consider using a non-parametric test such as Spearman's Correlation.

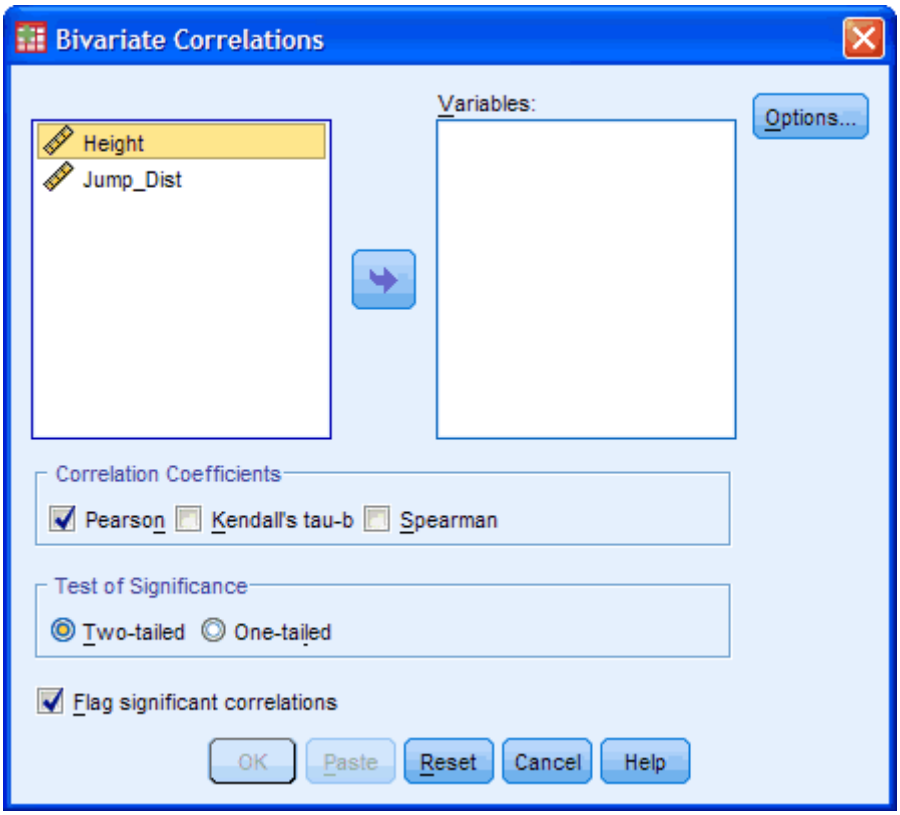
Test Procedure in SPSS

1. Click **Analyze > Correlate > Bivariate...** on the menu system as shown below:



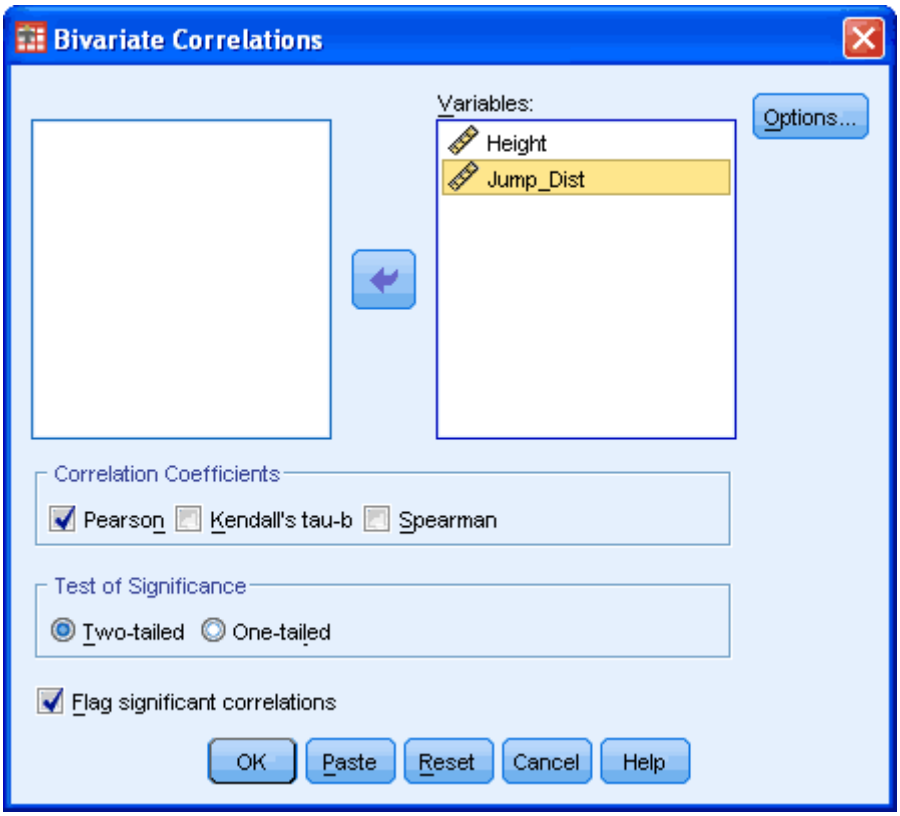
Published with written permission from SPSS Inc, an IBM Company.

You will be presented with the following screen:



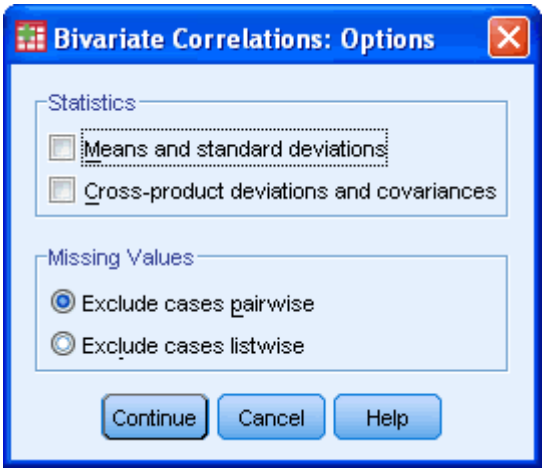
Published with written permission from SPSS Inc, an IBM Company.

2. Transfer the variables "Height" and "Jump_Dist" into the "*Variables:*" box by dragging-and-dropping or by clicking the button. You will end up with a screen similar to the one below:



Published with written permission from SPSS Inc, an IBM Company.

3. Make sure that the Pearson tickbox is checked under the "Correlation Coefficients" group (although it is selected by default in SPSS).
4. Click the Options... button. If you wish to generate some descriptives you can do it here by clicking on the particular tickbox.



Published with written permission from SPSS Inc, an IBM Company.

Then click the Continue button.

5. Click the OK button.

Output

You will be presented with the **Correlations** table in the output viewer as below:

Correlations

		Height	Jump_Dist
Height	Pearson Correlation	1	.777**
	Sig. (2-tailed)		.000
	N	27	27
Jump_Dist	Pearson Correlation	.777**	1
	Sig. (2-tailed)	.000	
	N	27	27

**. Correlation is significant at the 0.01 level (2-tailed).

Published with written permission from SPSS Inc, an IBM Company.

The results are presented in a matrix such that, as can be seen above, the correlations are replicated. Nevertheless, the table presents the Pearson correlation coefficient, the significance value and the sample size that the calculation is based on. In this example, we can see that the Pearson correlation coefficient, r, is 0.777 and that this is statistically significant ($p < 0.0005$).

Understanding the Output

In our example you might present the results are follows:

A Pearson product-moment correlation was run to determine the relationship between an individual's height and their performance in a long jump (distance jumped). The data showed no violation of normality, linearity or homoscedasticity (you will need to have checked for these). There was a strong, positive correlation between height and distance jumped, which was statistically significant ($r = .777$, $n = 27$, $p < .0005$).

Spearman's Rank Order Correlation using SPSS

25

Objectives

The Spearman Rank Order Correlation coefficient, r_s, is a non-parametric measure of the strength and direction of association that exists between two variables measured on at least an ordinal scale. It is denoted by the symbol r_s (or the greek letter ρ,pronounced rho). The test is used for either ordinal variables or for interval data that has failed the assumptions necessary for conducting the Pearson's product-moment correlation. If you would like some more background on this test you can find it in our statistical guide here.

Assumptions

- Variables are measured on an ordinal, interval or ratio scale (see out Types of Variable article).

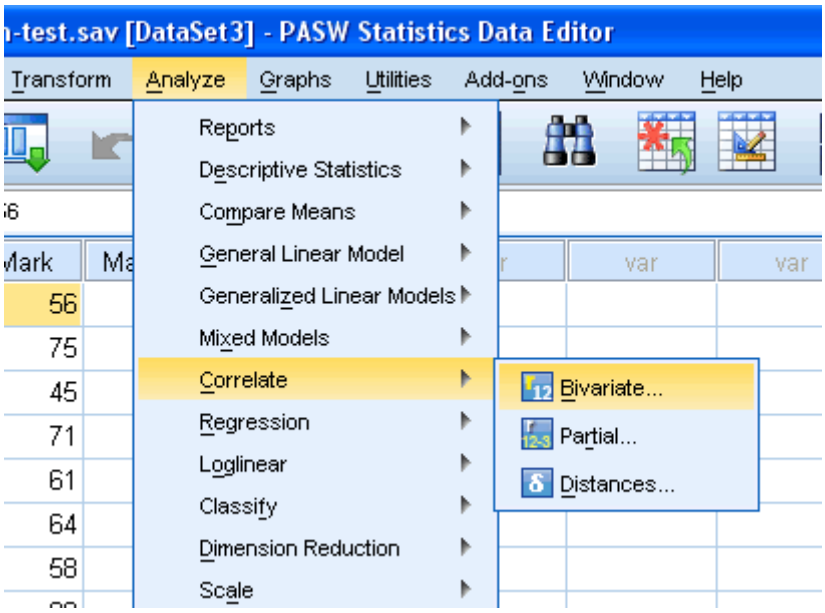
- Variables need **NOT** be normally distributed.
- There is a **monotonic relationship** between the two variables, i.e. either the variables increase in value together or as one variable value increases the other variable value decreases (see guide here for more information).
- This type of correlation is **NOT** very sensitive to outliers.

Example

A teacher is interested in those who do the best at English also do better in Maths (assessed by exam) students in English are also the best performers in Maths. She records the scores of her 10 students as they performed in end-of-year examinations for both English and Maths.

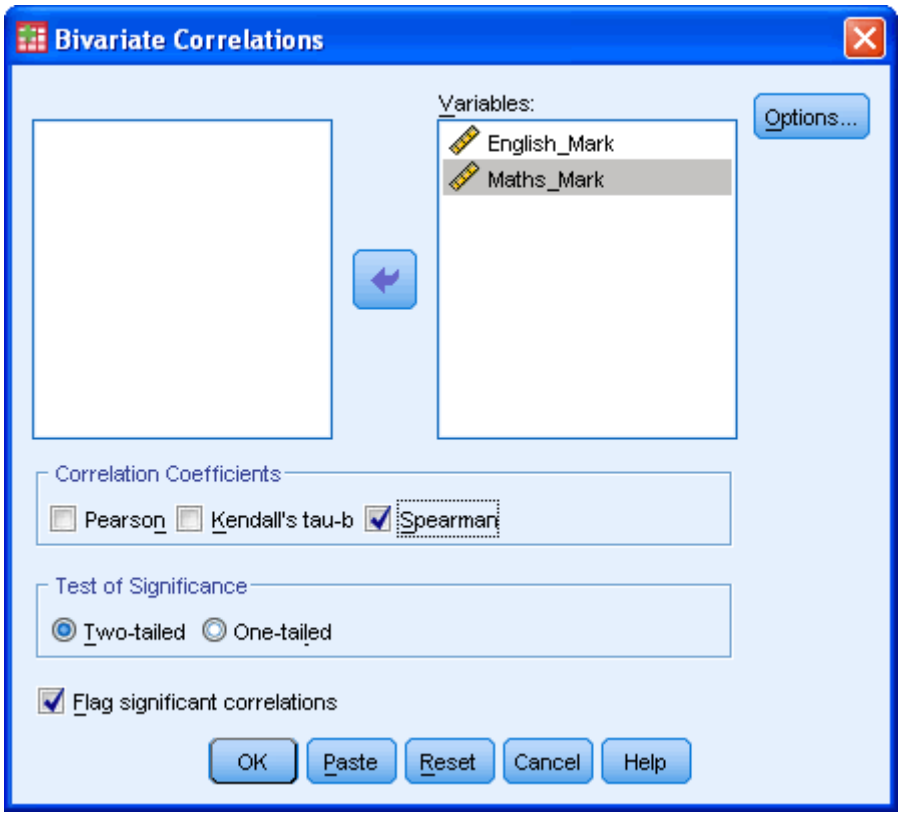
Test Procedure in SPSS

1. Click **Analyze > Correlate > Bivariate...** on the menu system as shown below:



Published with written permission from SPSS Inc, an IBM Company.

2. Transfer the variables "English_Mark" and "Maths_Mark" into the "Variables" box by dragging-and-dropping or by clicking the button. You will end up with a screen similar to the one below:



Published with written permission from SPSS Inc, an IBM Company.

3. Make sure that you uncheck the Pearson tickbox (it is selected by default in SPSS) and check the Spearman tickbox under the "Correlation Coefficients" group.
4. Click the OK button.

Output

You will be presented with 3 tables in output viewer under the title "Correlations" as below:

Correlations

			English_Mark	Maths_Mark
Spearman's rho	English_Mark	Correlation Coefficient	1.000	.669*
		Sig. (2-tailed)	.	.035
		N	10	10
	Maths_Mark	Correlation Coefficient	.669*	1.000
		Sig. (2-tailed)	.035	.
		N	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

Published with written permission from SPSS Inc, an IBM Company.

The results are presented in a matrix such that, as can be seen, the correlations are replicated. Nevertheless, the table presents Spearman's Rank Order Correlation, its significance value and the sample size that the calculation was

based on. In this example, we can see that Spearman's correlation coefficient, r_s, is 0.669 and that this is statistically significant (P = 0.035).

Reporting the Output

In our example you might present the results are follows: A Spearman's Rank Order correlation was run to determine the relationship between 10 students' English and maths exam marks. There was a strong, positive correlation between English and maths marks, which was statistically significant ($r_s(8)$ = .669, P = .035).

Chi-Square Test for Association using SPSS

48

Objective

The Chi-Square test for independence, also called Pearson's Chi-square test or the Chi-square test of association is used to discover if there is a relationship between two categorical variables.

Example

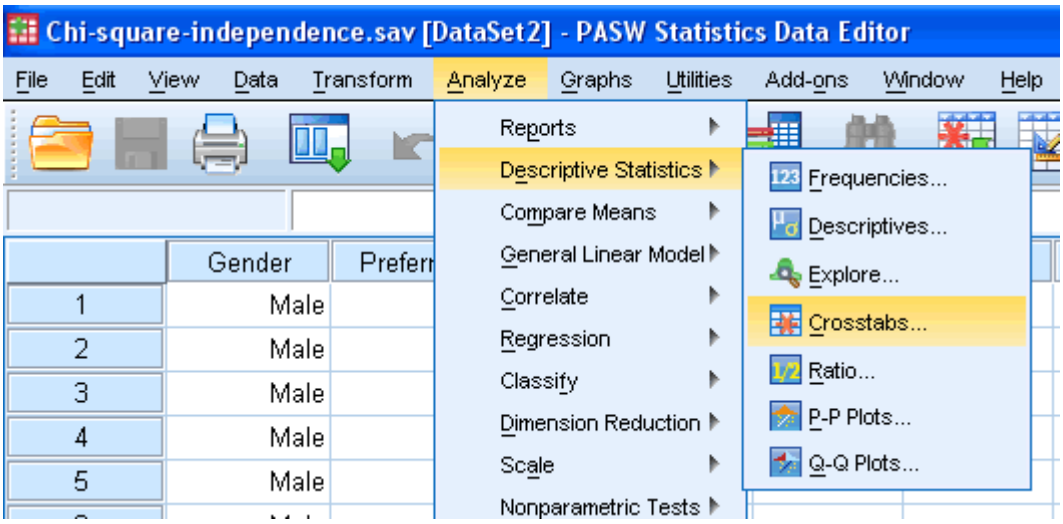
Educators are always looking for novel ways in which to teach statistics to undergraduates as part of a non-statistics degree course, e.g. psychology. With current technology it is possible to present how-to guides for statistical programs online instead of in a book. However, different people learn in different ways. An educator would like to know whether gender (male/female) is associated with the preferred type of learning medium (online vs. books). We therefore have two nominal variables: Gender(male/female) and Preferred Learning Medium (online/books).

Assumptions

- Two variables that are **ordinal or nominal (categorical data)**. (see our guide on Types of Variable)
- There are **two or more groups in each variable**.

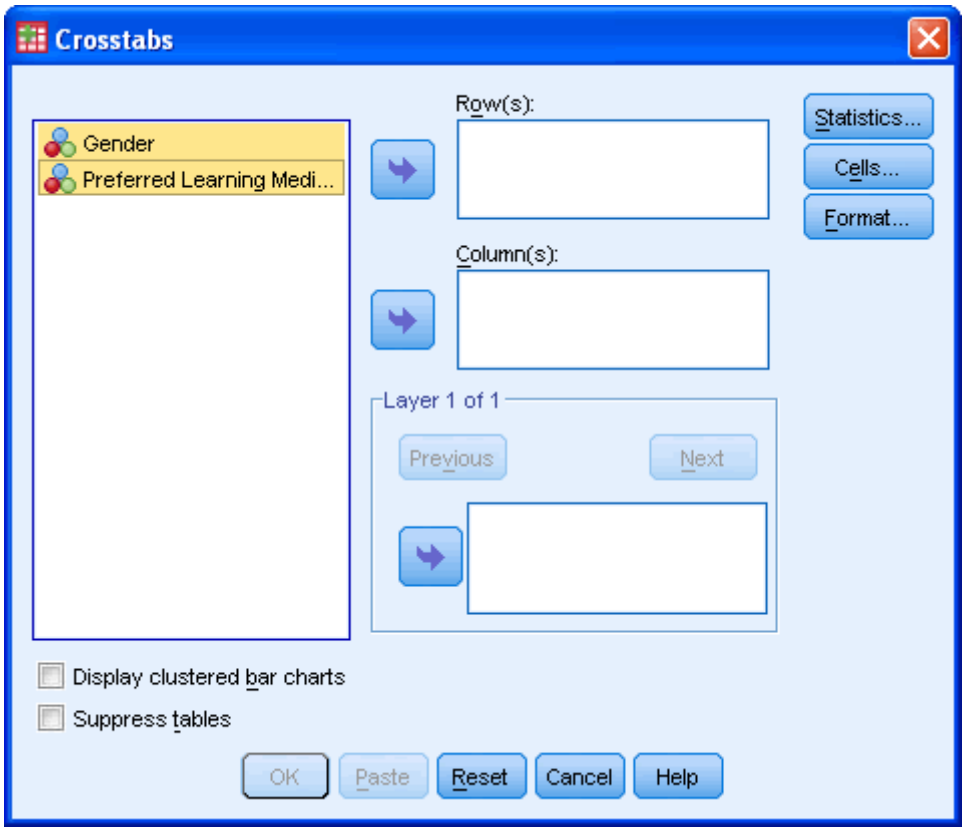
Test Procedure in SPSS

1. Click **Analyze > Descriptives Statistics > Crosstabs...** on the to menu as shown below:



Published with written permission from SPSS Inc, an IBM Company.

2. You will be presented with the following:



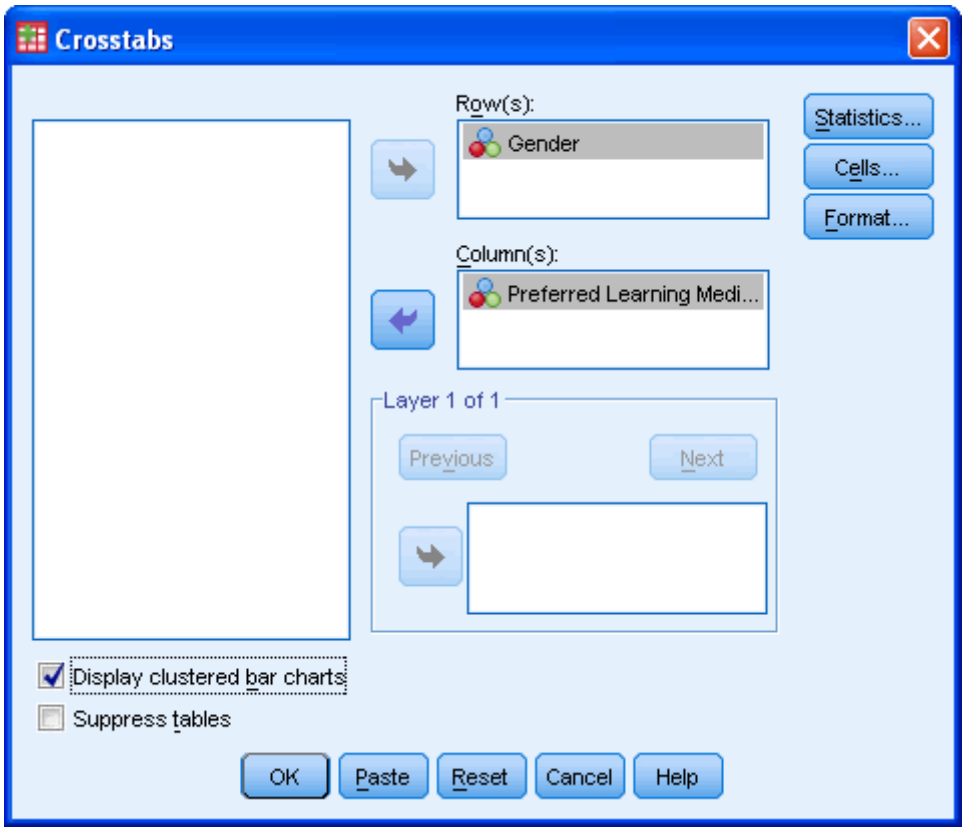
Published with written permission from SPSS Inc, an IBM Company.

3. Transfer one of the variables into the "Row(s):" box and the other variable into the "Column(s):" box. In our example we will transfer the "Gender" variable into the "Row(s):" box and "Preferred_Learning" into the "Column(s):" box. There are two ways to do this. You can highlight the variable with your mouse and then use the relevant buttons to transfer

the variables or you can drag-and-drop the variables. How do you know which variable goes in the row or column box? There is no right or wrong way. It will depend on how you want to present your data.

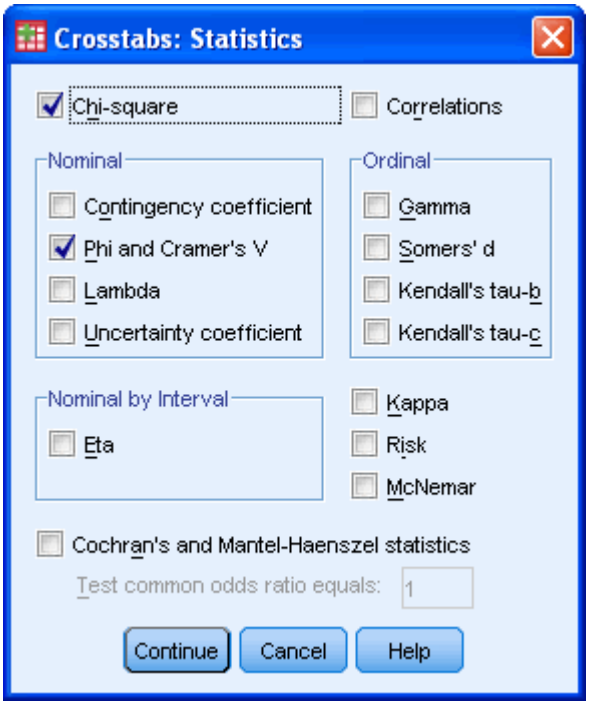
If you want to display clustered bar charts (recommended) then make sure that "Display clustered bar charts" checkbox is ticked.

You will end up with a screen similar to the one below:



Published with written permission from SPSS Inc, an IBM Company.

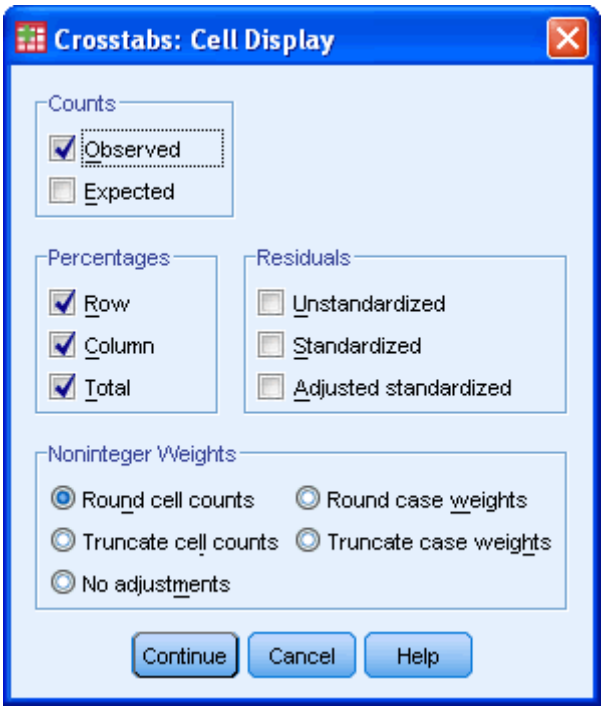
4. Click on the Statistics... button. Select the "Chi-square" and "Phi and Cramer's V" options as shown below:



Published with written permission from SPSS Inc, an IBM Company.

Click the Continue button.

5. Click the Cells... button. Select "Observed" from the "Counts" area and "Row", "Column" and "Total" from the "Percentages" area as shown below:

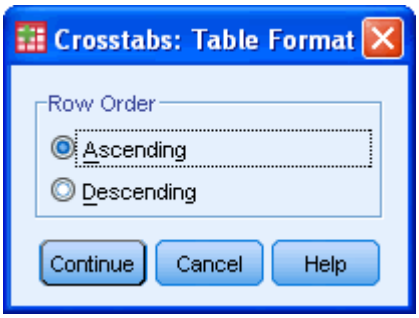


Published with written permission from SPSS Inc, an IBM Company.

Click the Continue button.

6. Click the Format... button. [This next option is only really useful if you have more than two categories in one of your variables but we will show it here in case you have]

   You will be presented with the following:



Published with written permission from SPSS Inc, an IBM Company.

   This option allows you to change the order of the values to either ascending or descending.

   Once you have made your choice click the Continue button.

7. Click the OK button to generate your output.

Output

You will be presented with some tables in the Output Viewer under the title "Crosstabs". The tables of note are presented below:

The Crosstabulation Table (Gender*Preferred Learning Medium Crosstabulation)

Gender * Preferred Learning Medium Crosstabulation

			Preferred Learning Medium		
			Books	Online	Total
Gender	Male	Count	16	24	40
		% within Gender	40.0%	60.0%	100.0%
		% within Preferred Learning Medium	55.2%	47.1%	50.0%
		% of Total	20.0%	30.0%	50.0%
	Female	Count	13	27	40
		% within Gender	32.5%	67.5%	100.0%
		% within Preferred Learning Medium	44.8%	52.9%	50.0%
		% of Total	16.3%	33.8%	50.0%
Total		Count	29	51	80
		% within Gender	36.3%	63.8%	100.0%
		% within Preferred Learning Medium	100.0%	100.0%	100.0%
		% of Total	36.3%	63.8%	100.0%

Published with written permission from SPSS Inc, an IBM Company.

This table allows us to understand that both males and females prefer to learn using online materials *vs.* books.

The Chi-Square Tests Table

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.487[a]	1	.485		
Continuity Correction[b]	.216	1	.642		
Likelihood Ratio	.487	1	.485		
Fisher's Exact Test				.642	.321
Linear-by-Linear Association	.481	1	.488		
N of Valid Cases	80				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 14.50.

b. Computed only for a 2x2 table

Published with written permission from SPSS Inc, an IBM Company.

When readings this table we are interested in the results for the Continuity correction. We can see here that Chi-square(1) = 0.487, *P* = 0.485. This tells us that there is no statistically significant association between Gender and Preferred Learning Medium. That is, both Males and Females equally prefer online learning vs. books. If you had a 2 x 2 contingency table and small numbers then .....

The Symmetric Measures Table

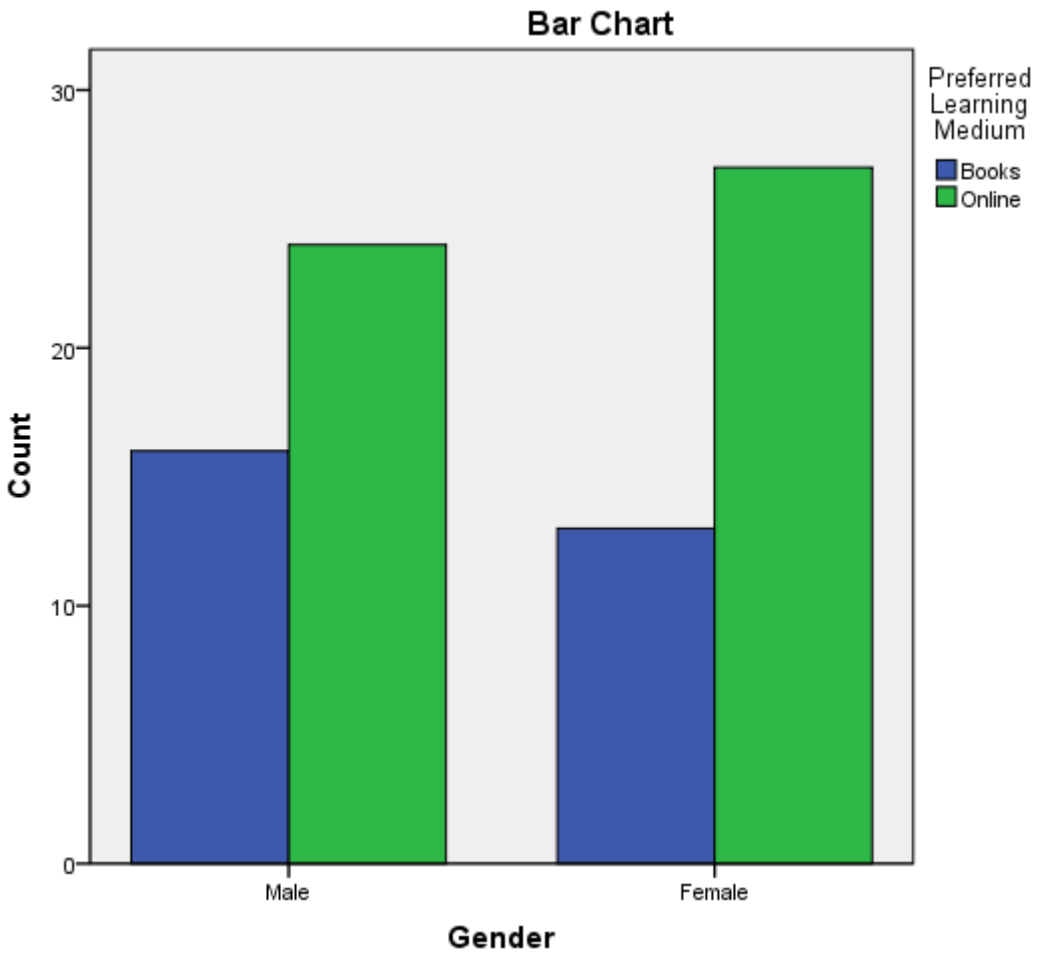
Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.078	.485
	Cramer's V	.078	.485
N of Valid Cases		80	

Published with written permission from SPSS Inc, an IBM Company.

Phi and Cramer's V are both tests of the strength of association. We can see that the strength of association between the variables is very weak.

Bar chart



Published with written permission from SPSS Inc, an IBM Company.

It can be easier to visualize data than read tables. The clustered bar chart option allows a relevant graph to be produced that highlights the group categories and the frequency of counts in these groups.

Reliability

Cronbach's Alpha (α) using SPSS

88

Introduction

Cronbach's alpha is the most common measure of internal consistency ("reliability"). It is most commonly used when you have multiple Likert questions in a survey/questionnaire that form a scale and you wish to determine if the scale is reliable.

Example

A researcher has devised a nine-question questionnaire with which they hope to measure how safe people feel at work at an industrial complex. Each question was a 5-point Likert item from "strongly disagree" to "strongly agree". In order to understand whether the questions in this questionnaire all reliably measure the

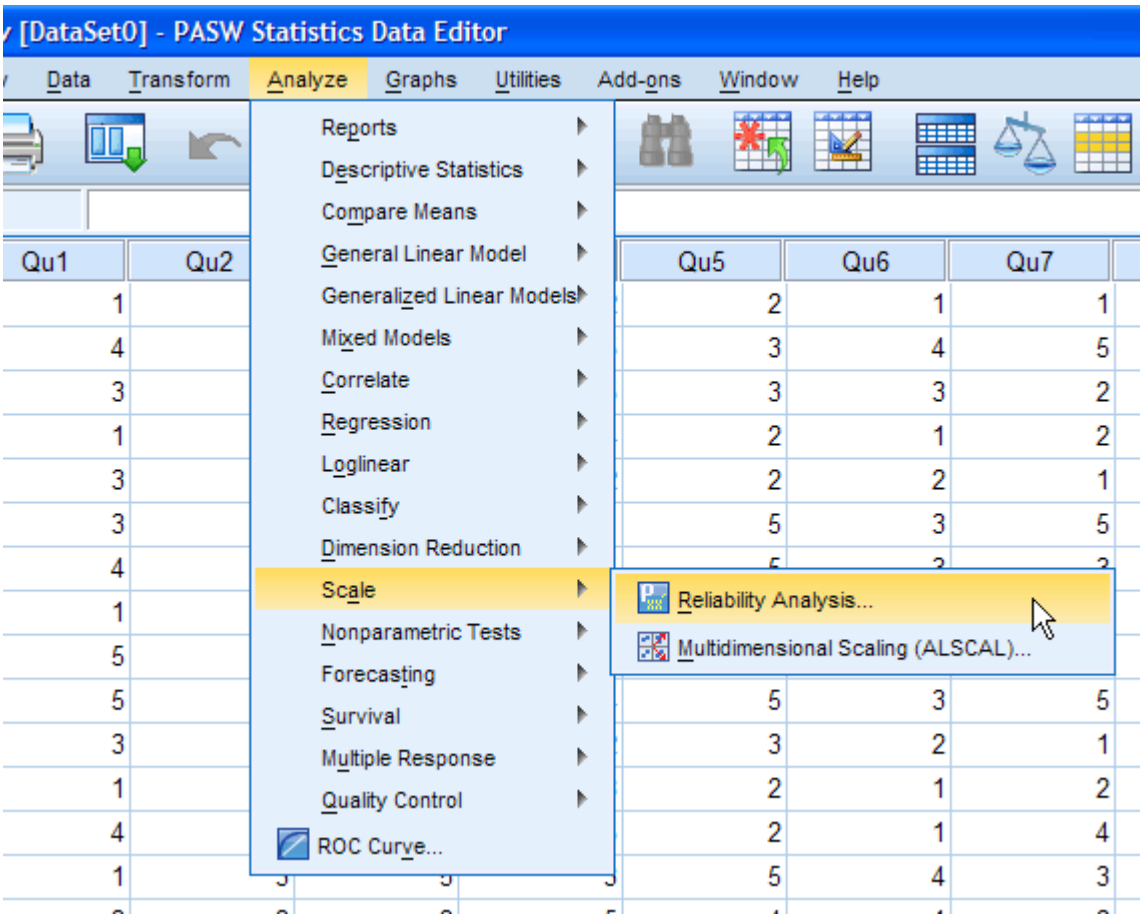
same latent variable (feeling of safety) (so a Likert scale could be constructed), a Cronbach's alpha was run on a sample size of 15 workers.

Setup in SPSS

The nine questions have been labelled "**Qu1**" through to "**Qu9**". To know how to correctly enter your data into SPSS in order to run a Cronbach's alpha test please read our Entering Data into SPSS tutorial.

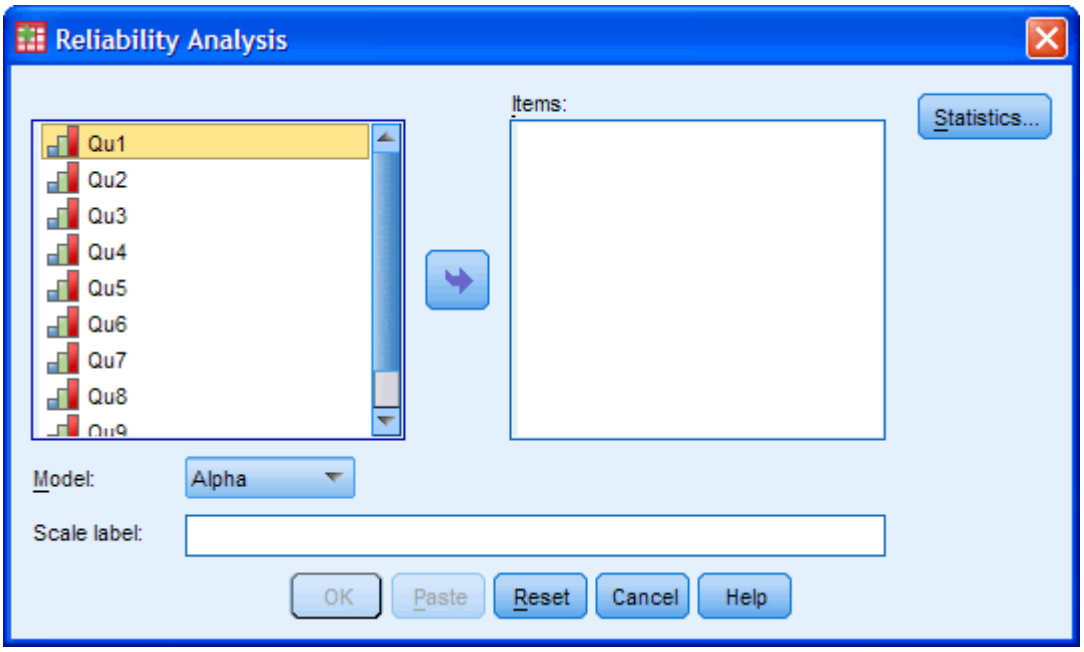
Test Procedure in SPSS

1. Click **Analyze > Scale > Reliability Analysis...** on the top menu as shown below:



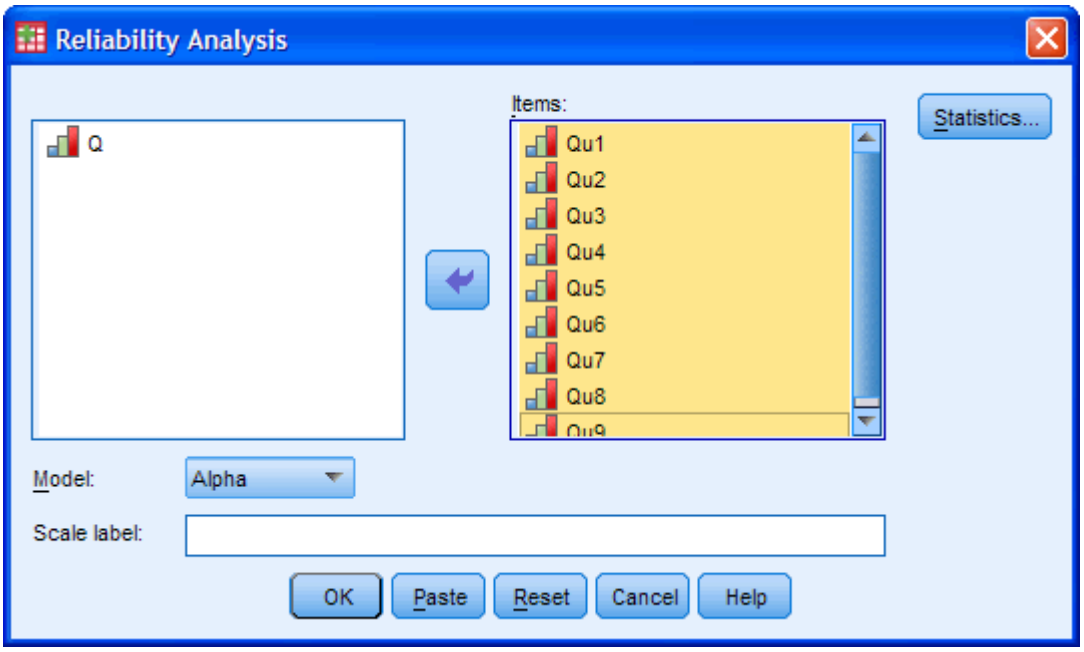
Published with written permission from SPSS Inc, an IBM Company.

2. You will be presented with the **Reliability Analysis** dialogue box:



Published with written permission from SPSS Inc, an IBM Company.

3. Transfer the variables "**Qu1**" to "**Qu9**" into the "*Items:*" box. You can do this by drag-and-dropping the variables into their respective boxes or by using the button. You will be presented with the following screen:

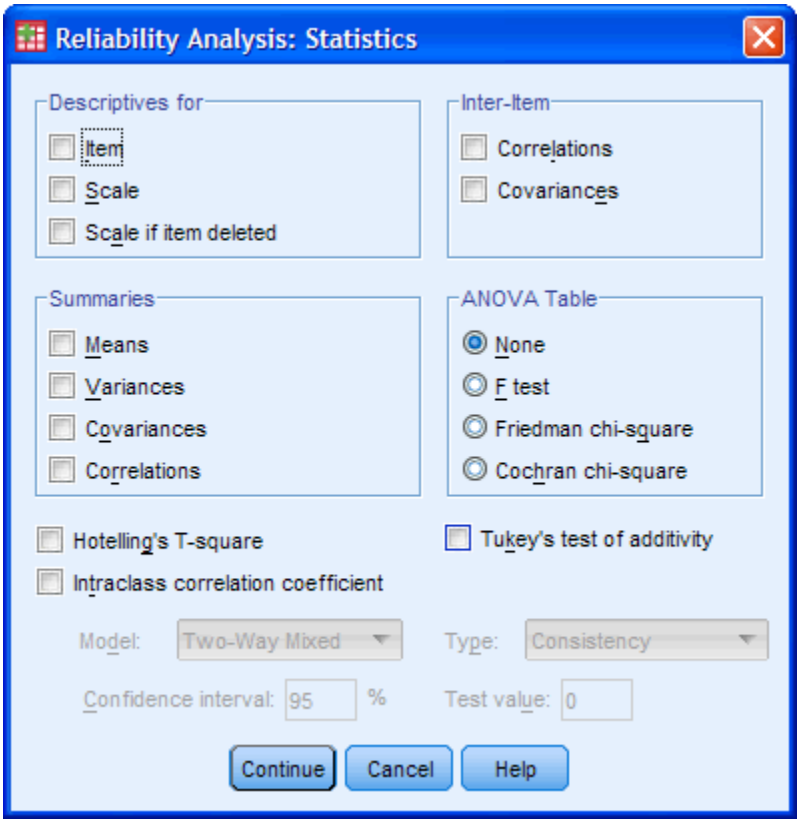


Published with written permission from SPSS Inc, an IBM Company.

4. Leave the "*Model:*" set as "Alpha", which represents Cronbach's alpha in SPSS. If you want to provide a name for the scale enter it in the "*Scale label:*" box. Since this only prints the name you enter at the top of the

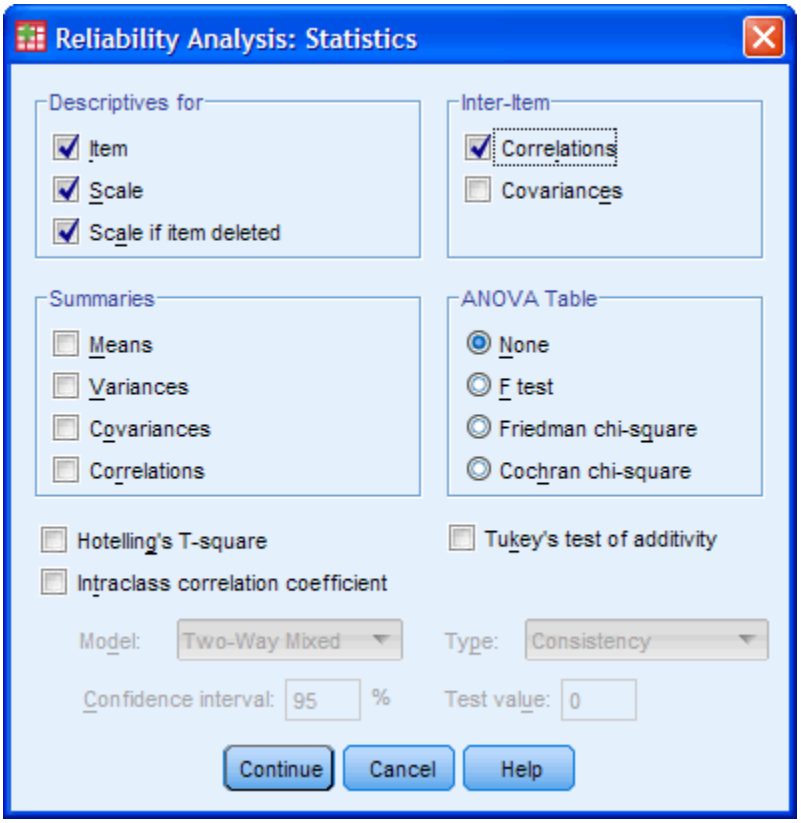
SPSS output, it is certainly not essential that you do; and in this case we will leave it blank.

5. Click on the Statistics... button, which will present the **Reliability Analysis: Statistics** dialogue box, as shown below:



Published with written permission from SPSS Inc, an IBM Company.

6. Select the "*Item*", "*Scale*" and "*Scale if item deleted*" in the "Descriptives for" box and "Correlations" in the "Inter-Item" box, as shown below:



Published with written permission from SPSS Inc, an IBM Company.

7. Click the Continue button. This will return you to the **Reliability Analysis** dialogue box.
8. Click the OK button to generate the output.

SPSS Output for Cronbach's Alpha

SPSS produces many different tables. The first important table is the **Reliability Statistics** table that provides the actual value for **Cronbach's alpha**, as shown below:

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.805	.796	9

Published with written permission from SPSS Inc, an IBM Company.

We can see that in our example, Cronbach's alpha is **0.805**, which indicates a high level of internal consistency for our scale with this specific sample.

Item-Total Statistics

The **Item-Total Statistics** table presents the **Cronbach's Alpha if Item Deleted** in the final column , as shown below:

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Qu1	24.20	45.029	.633	.588	.767
Qu2	23.93	47.352	.520	.651	.783
Qu3	24.07	46.638	.654	.899	.767
Qu4	23.40	47.114	.551	.823	.779
Qu5	23.60	51.257	.389	.573	.799
Qu6	24.47	50.695	.372	.693	.802
Qu7	24.07	45.210	.615	.777	.770
Qu8	24.20	56.457	.128	.791	.823
Qu9	24.07	45.210	.589	.610	.774

Published with written permission from SPSS Inc, an IBM Company.

This column presents the value that Cronbach's alpha would be if that particular item was deleted from the scale. We can see that removal of any question except question 8, would result in a lower Cronbach's alpha. Therefore, we would not want to remove these questions. Removal of question 8 would lead to a small improvement in Cronbach's alpha and we can also see that the **Corrected Item-Total Correlation** value was low (0.128) for this item. This might lead us to consider whether we should remove this item.

Cronbach's alpha simply provides you with an overall reliability coefficient for a set of variables, e.g. questions. If your questions reflect different underlying personal qualities (or other dimensions), for example, employee motivation and employee commitment, then Cronbach's alpha will not be able to distinguish between these. In order to do this and then check their reliability (using Cronbach's alpha), you will first need to run a test such as a principal components analysis (PCA). If this sounds like something you would want to do, check out the features of our Premium PCA guide, here.

Differences Between Groups

Getting to Know SPSS

Creating a New File
Working with Variables

Checking Assumptions

Testing for Normality

Differences Between Groups

Independent T-Test

Entering Data
Transforming Data
Recoding Variables
Ranking Data

Graphs & Charts

Bar Chart
Clustered Bar Chart
Scatterplot

Predicting Scores

Linear Regression

Associations

Pearson's Correlation
Spearman's Correlation
Chi-Square for Association

Reliability

Cronbach's Alpha

Dependent T-Test
One-way ANOVA
ANOVA with Repeated Measures
Two-way ANOVA
One-way MANOVA
Mann-Whitney U Test
Wilcoxon Signed Rank Test
Kruskal-Wallis H Test
Friedman Test

One-Sample Tests

Goodness of Fit Test

SPSS Tutorials

Essentials

Descriptive & Inferential Statistics
Measures of Central Tendency
Measures of Spread
Absolute Deviation & Variance
Standard Deviation
Hypothesis Testing
Types of Variable
Normal Distribution Calculations
Histograms
Standard Score

Statistical Tests

Independent T-Test
Dependent T-Test
One-way ANOVA
Repeated Measures ANOVA
Pearson Correlation
Spearman's Correlation

Testing Assumptions

Sphericity

Calculators

Essentials

Standard Deviation
Normal Distribution

Groups

Dependent T-Test

Relationships

Pearson's Correlation
Spearman's Correlation

Independent T-Test using SPSS

83

Objectives

The independent t-test compares the means between two unrelated groups on the same continuous, dependent variable. The SPSS t-test procedure allows the testing of equality of variances (Levene's test) and the *t*-value for both equal- and unequal-variance. It also provides the relevant descriptive statistics. A statistical guide on the independent t-test is provided here.

You can access our much more comprehensive version of this guide for FREE, normally only available in our Laerd Statistics Premium section, here. We have made the guide available to illustrate the difference between our free and enhanced guides.

Assumptions

- Independent variable consists of **two independent groups**.
- Dependent variable is either **interval or ratio** (see our guide on Types of Variable).
- Dependent variable is **approximately normally distributed** (see Testing for Normality article)
- Similiar variances between the two groups (**homogeneity of variances**) (tested for in this t-test procedure).

Background to Example

The concentration of cholesterol (a type of fat) in the blood is associated with the risk of developing heart disease, such that higher concentrations of cholesterol indicate a higher level of risk and lower concentrations indicate a lower level of risk. If you lower the concentration of cholesterol in the blood then your risk for developing heart disease can be reduced. Being overweight and/or physically inactive increases the concentration of cholesterol in your blood. Both exercise and weightloss can reduce cholesterol concentration. However, it is not known whether exercise or weightloss is best for lowering blood cholesterol concentration.

Example

A random sample of inactive male individuals that were classified as overweight were recruited to a study to investigate whether an exercise or weight loss intervention is more effective in lowering cholesterol levels. To this end, they randomly split the group into two sub-groups; one group underwent an exercise training programme and the other group undertook a calorie-controlled diet. In order to determine which treatment programme was more effective, the mean cholesterol concentrations were compared between the two groups at the end of the treatment programmes.

Setup in SPSS

In SPSS we separated the groups for analysis by creating a grouping variable called "**Group**" and gave the exercise group a value of "**1**" and the diet group a value of "**2**". Cholesterol concentrations were entered under the variable name "**Cholesterol**". How to correctly enter data in SPSS to run an independent t-test is explained in our guide here.

Descriptives

Unless you have other reasons to do so, it would be considered normal to present information on the mean and standard deviation for this data. You might also state the number of participants you had in each group that were to be analysed.

This is in case you have, for example, missing values and the number of recruited participants is larger than that which can be analysed.

You might also wish to present a diagram so that you can show your results visually in order that a reader might understand them better. You could present a bar chart with error bars (for example, SD or 95% CI) (see our guide here).

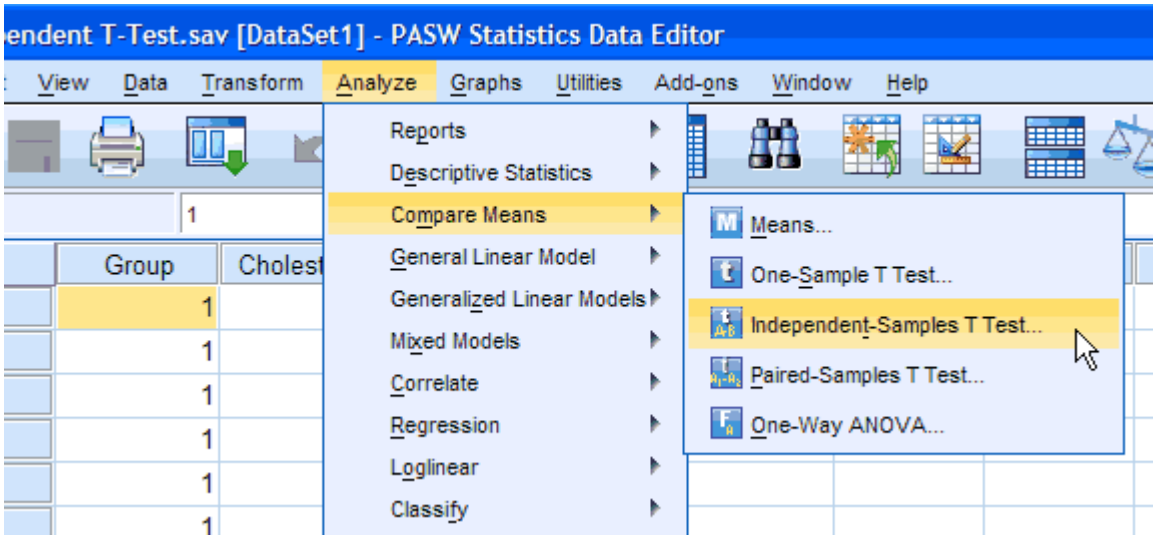
Testing assumptions

To determine whether your samples are normally distributed read our Testing for Normality article. What if your samples are not normally distributed? Well, if your data set is large then small deviations are generally tolerable. However, if your samples are small or your data set is largely non-normal then you need to consider a non-parametric test instead, such as the Mann-Whitney U Test.

The assumption of equal variances is tested in SPSS by Levene's Test for Equality of Variances. The result of this test is presented in the output when running an independent t-test and is discussed later in this guide.

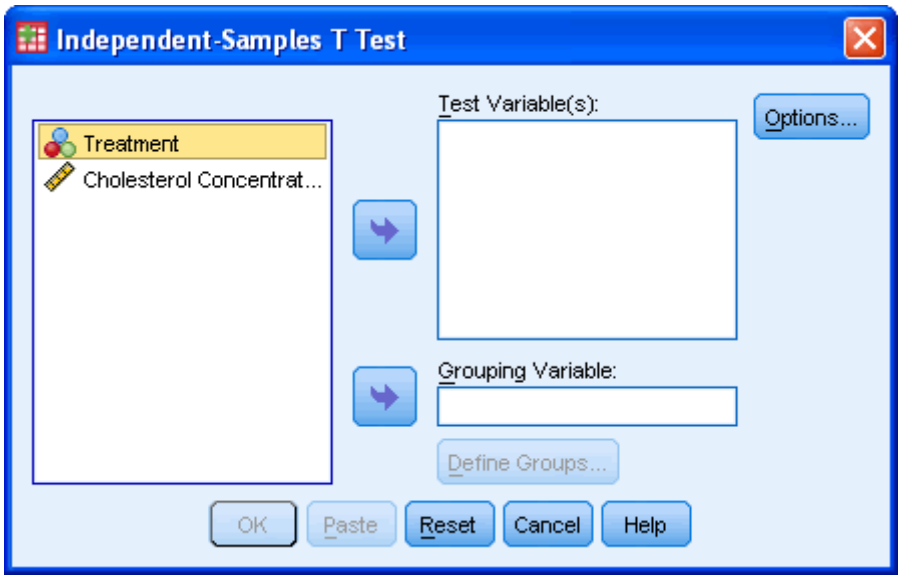
Test Procedure in SPSS

1. Click **Analyze > Compare Means > Independent-Samples T Test...** on the top menu as shown below.



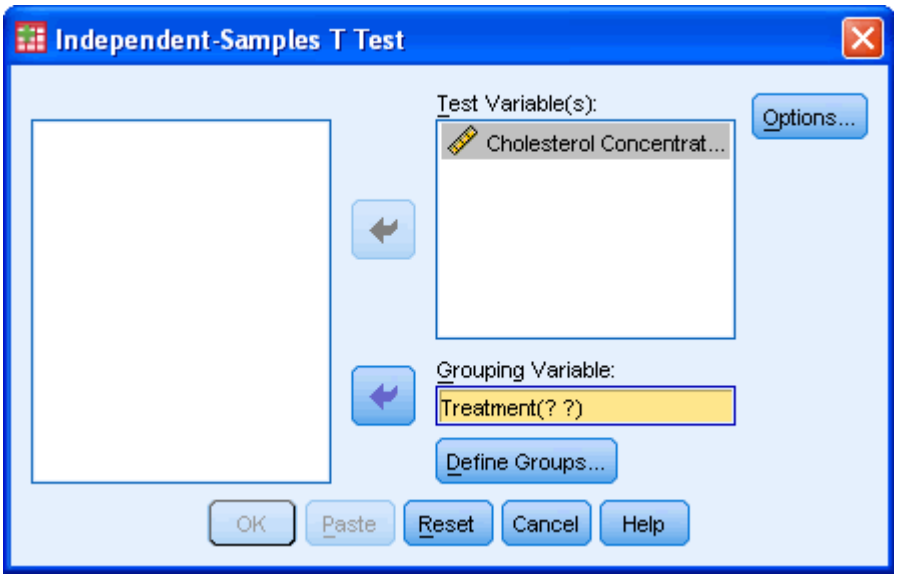
Published with written permission from SPSS Inc, an IBM company.

You will be presented with the following:



Published with written permission from SPSS Inc, an IBM company.

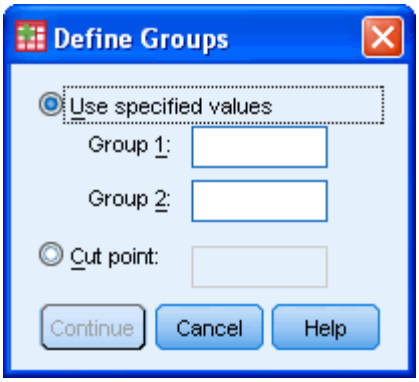
2. Put the "**Cholesterol Concentration**" variable into the "Test Variable(s):" box and the "**Treatment**" variable into the "Grouping Variable:" box by highlighting the relevant variables and pressing the buttons.



Published with written permission from SPSS Inc, an IBM company.

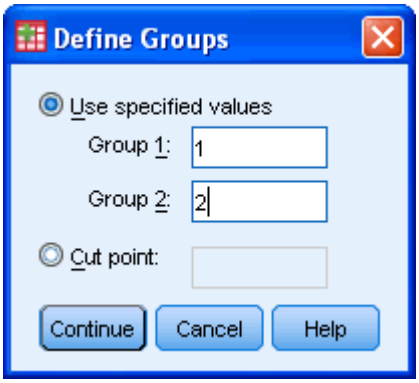
3. You then need to define the groups (treatments). Press the Define Groups... button.

You will be presented with the following screen:



Published with written permission from SPSS Inc, an IBM company.

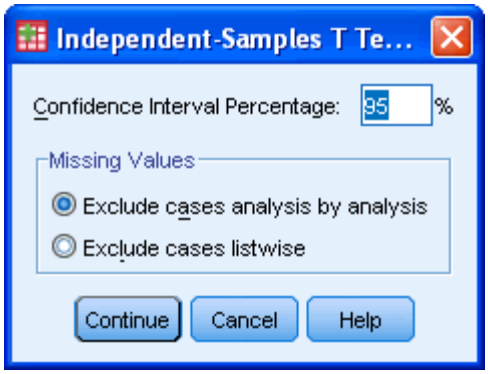
4. Enter "**1**" into the "Group 1:" box and enter "**2**" into the "Group 2:" box. Remember that we labelled the Diet Treatment group as "**1**" and the Exercise Treatment group as "**2**". If you have more than 2 treatment groups, e.g. a diet, exercise and drug treatment group, then you could type in "1" to "Group 1:" box and "3" to "Group 2:" box if you wished to compare the diet with drug treatment.



Published with written permission from SPSS Inc, an IBM company.

5. Press the Continue button
6. If you need to change the confidence level limits, or change how to exclude cases then press the Options... button.

   You will be presented with the following:



Published with written permission from SPSS Inc, an IBM company.

7. Click the Continue button.
8. Click the OK button.

Output of the independent t-test in SPSS

You will be presented with two tables containing all the data generated by the independent t-test procedure in SPSS.

Group Statistics Table

This table provides useful descriptive statistics for the two groups that you compared including the mean and standard deviation.

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Cholesterol Concentration	Diet	20	6.1450	.51959	.11618
	Exercise	20	5.7950	.38179	.08537

Published with written permission from SPSS Inc, an IBM company.

Independent Samples Test Table

This table provides the actual results from the independent t-test and Levine's Test for Equality of Variances. [In order to fit the image of this table onto this webpage we have swapped the rows and columns in the table so that the table is arranged vertically not horizontally. Therefore, you will be presented with a slightly differently looking table but the data and interpretation remain the same. By clicking on the image below you will see the original, horizontally arranged table so that you are able to compare the two.]

Independent Samples Test

			Cholesterol Concentration	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		.314	
	Sig.		.579	
t-test for Equality of Means	t		2.428	2.428
	df		38	34.886
	Sig. (2-tailed)		.020	.021
	Mean Difference		.35000	.35000
	Std. Error Difference		.14418	.14418
	95% Confidence Interval of the Difference	Lower	.05813	.05727
		Upper	.64187	.64273

Published with written permission from SPSS Inc, an IBM company.

The first thing you need to do is check to see if you have similar variances in the two groups by checking the result of Levine's Test for Equality of Variances. To check this, look at the "**Sig.**" row within **Levine's Test for Equality of Variances** row as highlighted below:

Independent Samples Test

		Cholesterol Concentration	
		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	.314	
	Sig.	.579	

Published with written permission from SPSS Inc, an IBM company.

If the variances are equal in both groups, then the *p*-value ("**Sig.**") will be greater than 0.05. However, if the *p*-value is less than 0.05, the variances are unequal. If you have unequal variances then you need to use the **Equal variances not assumed** column otherwise you use the **Equal variances assumed** column.

In this case, we have a *p*-value of 0.579 for Levene's test so we can conclude that we have equal variances and we look at the **Equal variances assumed** column. Looking down this column we can see that the group means are significantly different as the value in the "**Sig. (2-tailed)**" row is less than 0.05. Looking at the **Group Statistics** table we can see that those people that undertook the exercise trial had lower cholesterol levels at the end of the programme than those that underwent a calorie-controlled diet.

Reporting the output of the independent t-test

We might report the statistics in the following format: *t*(degrees of freedom[*df*]) = *t*-value, *p* = significance level. In our case this would be: *t*(38) = 2.470, *p* = 0.018. Therefore, we might report the results of the study as follows:

This study found that overweight and physically inactive male participants had statistically significant lower cholesterol levels (5.78 ± 0.38 mmol/L) at the end of an exercise training programme vs. after a calorie-controlled diet (6.21 ± 0.65 mmol/L) (*t*(38) = 2.470, *p* = 0.018).

Dependent T-Test using SPSS

17

Objectives

The dependent t-test (called the Paired-Samples T Test in SPSS) compares the means between two related groups on the same continuous variable. The SPSS paired-samples t-test procedure also provides relevant descriptive statistics. For an easy-to-follow guide on the dependent t-test please see our statistical guide.

Assumptions

- Dependent variable is **interval or ratio (continuous)** (see our Types of Variable guide).
- The differences in the dependent variable between the two related groups are **approximately normally distributed**.
- Independent variable consists of **two related groups or "matched-pairs"**.
- **No outliers** in the differences between the two related groups.

It is vitally important to check these assumptions because if they are violated the result of the dependent t-test can be invalid. How to first calculate the difference scores, and then to check the above assumptions on these scores, is presented in the enhanced version of this guide, available as part of our Laerd Statistics Premium content. To get a sense of the advantages of purchasing access to Laerd Statistics Premium, you can view our enhanced Independent-samples t-test in SPSS guide for free (normally Premium). To go straight to the relevant section for testing assumptions, click here. This enhanced guide also explain what to do if you violate any of the assumptions. You can check out our low prices for access to all the enhanced content in our Premium section here.

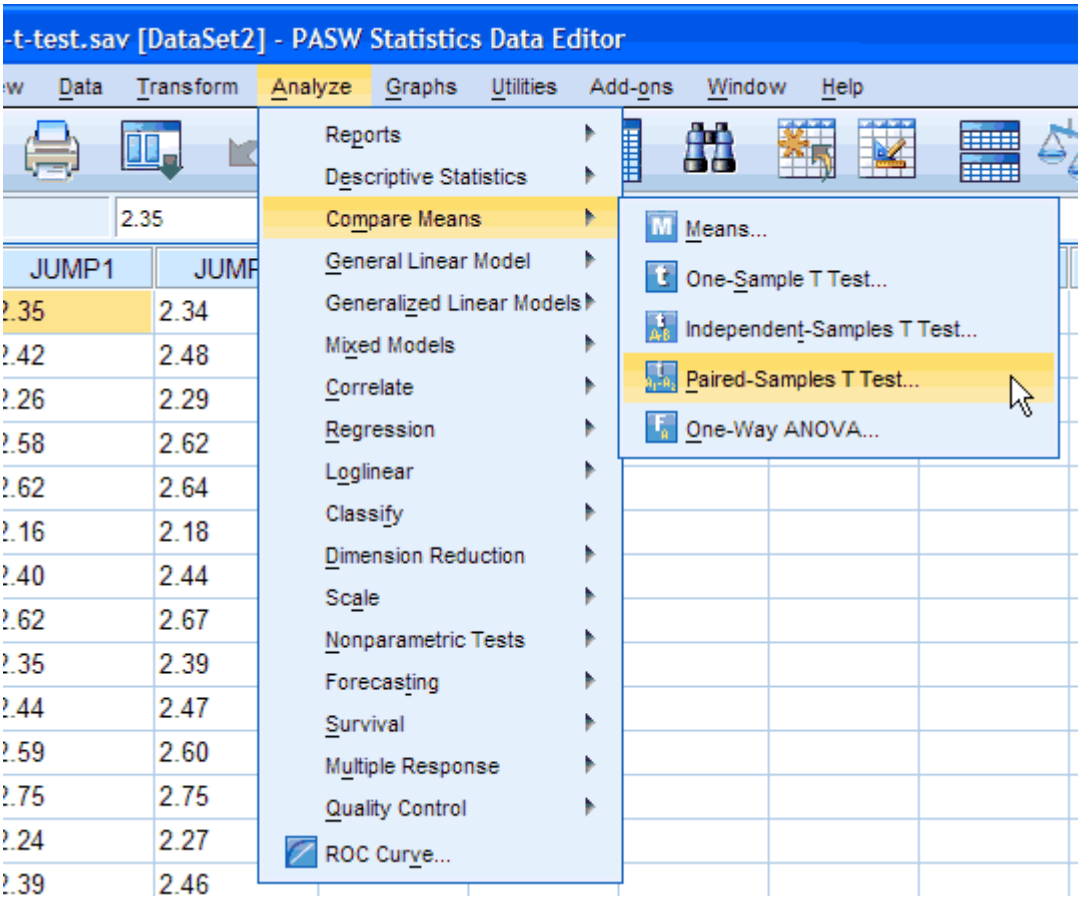
Example

A group of Sports Science students (n = 20) are selected from the population to investigate whether a 12 week plyometric training programme improves their standing long jump performance. In order to test whether this training improves performance, the sample group are tested for their long jump performance before they undertake a plyometric training programme and then again at the end of the programme.

Test Procedure in SPSS

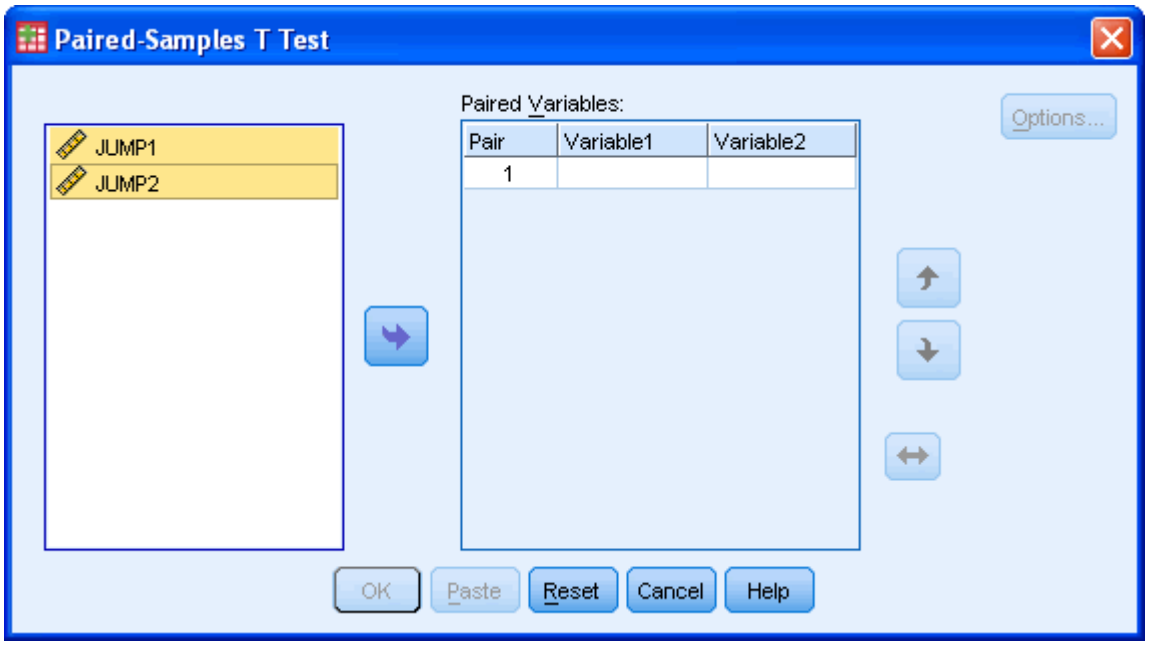
[If you are unsure of how to correctly enter your data into SPSS in order to run a dependent t-test then read our guide on how to do it here. Our enhanced guide includes a description of the file set-up and the ability to download the SPSS file for the guide.]

1. Click **Analyze > Compare Means > Paired-Samples T Test...** on the top menu.



Published with written permission from SPSS Inc, an IBM company.

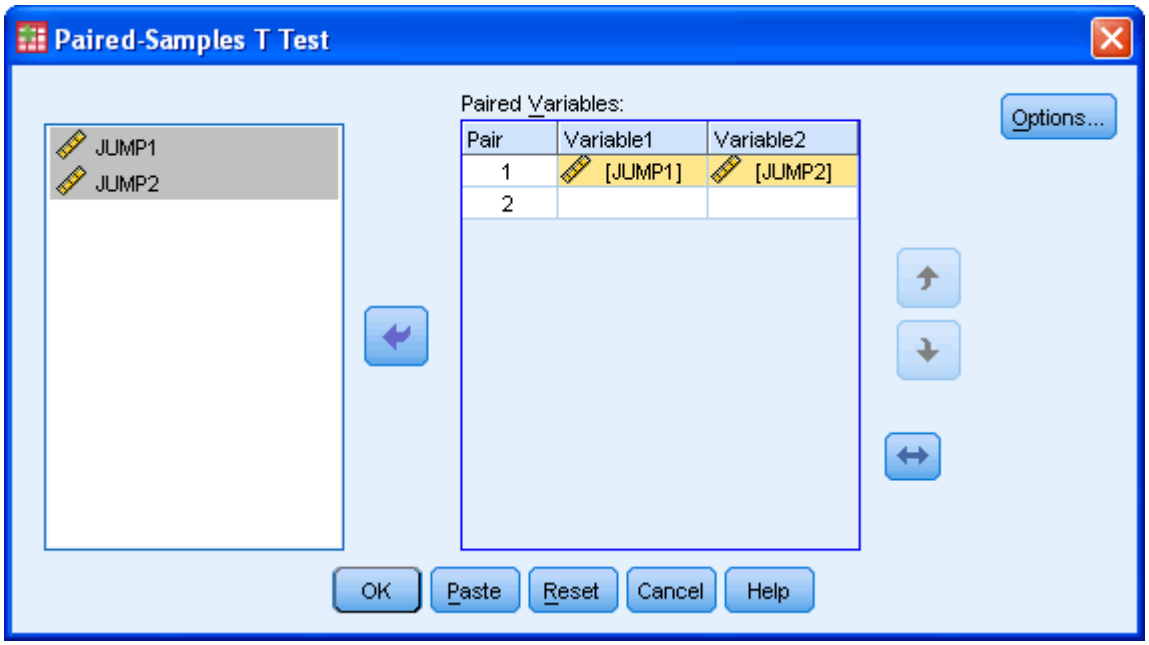
2. You will be presented with the following:



Published with written permission from SPSS Inc, an IBM company.

3. You need to transfer the variables "**JUMP1**" and "**JUMP2**" into the "Paired Variables:" box. There are two ways to do this. You can either highlight both variables (use the cursor and hold down the shift key and press the button, or you can drag and drop each variable into the boxes). If you are using older versions of SPSS, you will need to transfer the variables using the former method.

   You will end up with a screen similar to the one below:



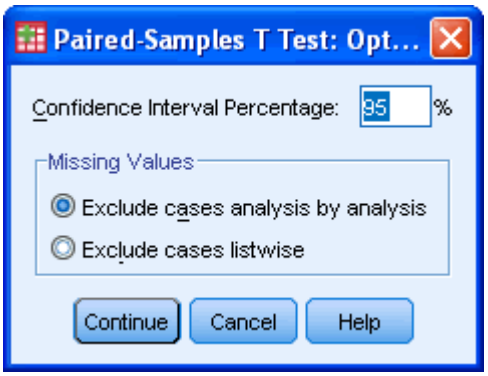
Published with written permission from SPSS Inc, an IBM company.

button shifts the pair of variables you have highlighted down one level.

button shifts the pair of variables you have highlighted up one level.

button shifts the order of the variables with a variable pair itself.

4. If you need to change the confidence level limits or to exclude cases then press the Options... button:



Published with written permission from SPSS Inc, an IBM company.

Click on the Continue button.

5. Click the OK button to generate the output.

SPSS Output of the Dependent T-Test

You will be presented with three tables in the **Output Viewer** under the title "T-Test" but you only need to look at two tables - the **Paired Sample Statistics** table and the **Paired Samples Test** table, as discussed below:

Paired Sample Statistics Table

The first table titled **Paired Sample Statistics** is where SPSS has generated descriptive statistics for your variables. You can use the data here to describe the characteristics of the first and second jumps in your results.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	JUMP1	2.4815	20	.16135	.03608
	JUMP2	2.5155	20	.15982	.03574

Published with written permission from SPSS Inc, an IBM company.

Paired Samples Test Table

The **Paired Samples Test** table is where the results of the dependent t-test are presented. A lot of information is presented here and it is important to remember

that this information refers to the **differences** between the two jumps (the subtitle reads "Paired Differences"). As such, the columns of the table labelled "Mean", "Std. Deviation", "Std. Error Mean", 95% CI refer to the mean difference between the two jumps and the standard deviation, standard error and 95% CI of this mean difference, respectively. The last 3 columns express the results of the dependent t-test, namely the *t*-value, the degrees of freedom and the significance level.

Paired Samples Test

			Pair 1
			JUMP1 - JUMP2
Paired Differences	Mean		-.03400
	Std. Deviation		.03185
	Std. Error Mean		.00712
	95% Confidence Interval of the Difference	Lower	-.04891
		Upper	-.01909
t			-4.773
df			19
Sig. (2-tailed)			.000

Published with written permission from SPSS Inc, an IBM company.

Reporting the output of the Dependent T-Test

We might report the statistics in the following format: *t*(degrees of freedom[df]) = *t*-value, *p* = significance level. In our case this would be: $t(19) = -4.773$, $p < 0.0005$. Due to the means of the two jumps and the direction of the *t*-value we can conclude that there was a statistically significant improvement in jump distance following the plyometric training programme from 2.48 ± 0.16 m to 2.52 ± 0.16 m ($p < 0.0005$); an improvement of 0.03 ± 0.03 m.

N.B. SPSS will output many results to many decimal places but you should understand your measuring scale to know whether it is appropriate to report your results in such accuracy.

A full explanation of the results for a dependent t-test is included in our enhanced guide in Laerd Statistics Premium. You can find out about pricing here.

Dependent T-Test for Paired Samples

17

What does this test do?

The dependent t-test (also called the paired t-test or paired-samples t-test) compares the means of two related groups to detect whether there are any statistically significant differences between these means.

If you wish to learn how to calculate the dependent t-test then we have a dependent t-test calculator that also generates all the working involved in getting to the answer. The calculator can be found here.

What variables do you need for a dependent t-test?

You need one dependent variable that is measured on an interval or ratio scale (see our Types of Variable guide if you need clarification). You also need one categorical variable that has only two related groups.

What is meant by "related groups"?

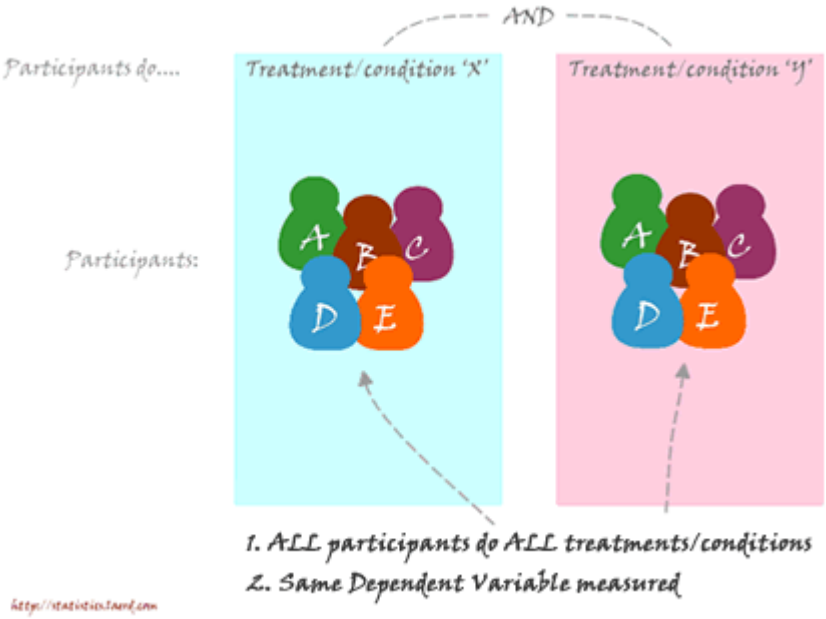
A dependent t-test is an example of a "within-subjects" or "repeated-measures" statistical test. This indicates that the same subjects are tested more than once. Thus, in the dependent t-test, "related groups" indicates that the same subjects are present in both groups. The reason that it is possible to have the same subjects in each group is because each subject has been measured on two occasions on the same dependent variable. For example, you might have measured 10 individuals' (subjects') performance in a spelling test (the dependent variable) before and after they underwent a new form of computerised teaching method to improve spelling. You would like to know if the computer training improved their spelling performance. Here, we can use a dependent t-test as we have two related groups. The first related group consists of the subjects at the beginning (prior to) the computerised spell training and the second related group consists of the same subjects but now at the end of the computerised training.

Does the dependent t-test test for "changes" or "differences" between related groups?

The dependent t-test can be used to test either a "change" or a "difference" in means between two related groups but not both at the same time. Whether you are measuring a "change" or "difference" between the means of the two related groups depends on your study design. The two types of study design are indicated in the following diagrams.

How do you detect differences between experimental conditions using the dependent t-test?

The dependent t-test can look for "differences" between means when subjects are measured on the same dependent variable under two different conditions. For example, you might have tested subjects' eyesight (dependent variable) when wearing two different types of spectacle (independent variable). See the diagram below for a general schematic of this design approach (click the image to enlarge):



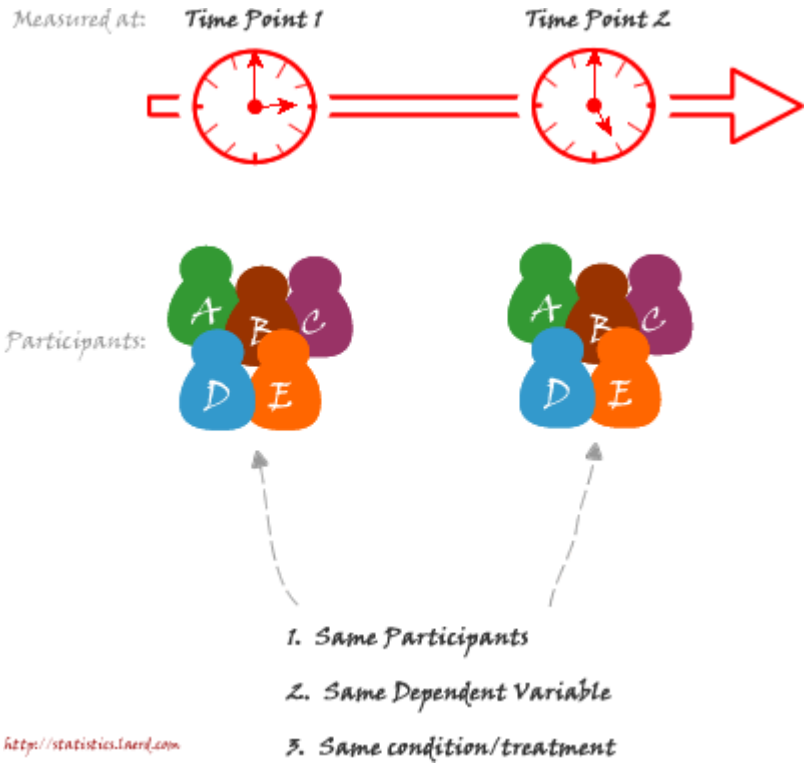
Find out more about the dependent t-test on the next page.

Dependent T-Test for Paired Samples (cont...)

2

How do you detect changes in time using the dependent t-test?

The dependent t-test can also look for "changes" between means when the subjects are measured on the same dependent variable but at two time points. A common use of this is in a pre-post study design. In this type of experiment we measure subjects at the beginning and at the end of some intervention, e.g. an exercise-training programme or business-skills course. A general schematic is provided below (click image to enlarge):



How else might you use the dependent t-test?

You can also use the dependent t-test to study more complex study designs although it is not normally recommended. The most common, more complex study design where you might use the dependent t-test is where you have a crossover design with two different interventions that are both performed by the same subjects. One example of this design is where you have one of the interventions act as a control. For example, you might want to investigate whether a course of diet counselling can help people lose weight. To study this you could simply measure subjects' weight before and after the diet counselling course for any changes in weight using a dependent t-test. However, to improve the study design you also include want to include a control trial. During this control trial, the subjects could either receive "normal" counselling or do nothing at all or something else you deem appropriate. In order to assess this study using a dependent t-test you would use the same subjects for the control trial as the diet counselling trial. You then measure the differences between the interventions at the end, and only at the end, of the two interventions. Remember, however, that this is unlikely to be the preferred statistical analysis for this study design.

What are the assumptions of the dependent t-test?

The types of variable needed for the dependent t-test have already been discussed earlier in this guide. In addition, the distribution of the differences between the scores of the two related groups needs to be normally distributed. We do this by simply subtracting each individuals' score in one group from their score in the other related group and then testing for normality in the normal way (see our guide on how to test for normality in SPSS here). It is important to note that the two related groups do not need to be normally distributed themselves - just the differences between the groups.

Dependent T-Test for Paired Samples (cont...)

3

What hypothesis is being tested?

The dependent t-test is testing the null hypothesis that there are no differences between the means of the two related groups. If we get a significant result then we can reject the null hypothesis that there are no significant differences between the means and accept the alternative hypothesis that there are statistically significant differences between the means. We can express this as follows:

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

What is the advantage of a dependent t-test over an independent t-test?

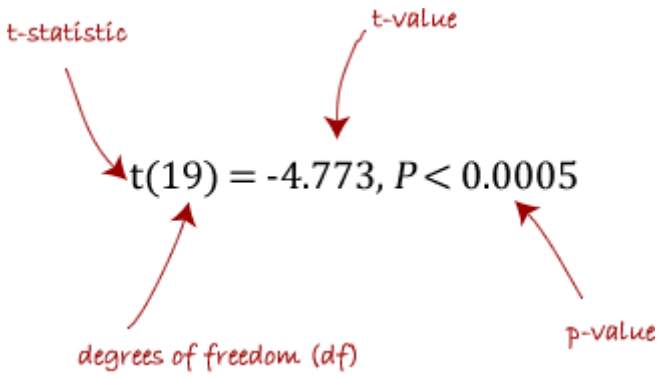
Before we answer this question, we need to point out that you cannot choose one test over the other unless your study design allows it. What we are discussing here is whether it is advantageous to design a study that uses one set of subjects whom are measured twice or two separate groups of subjects measured once each. The major advantage of choosing a repeated-measures design (and therefore running a dependent t-test) is that you get to eliminate the individual differences that occur between subjects - the concept that no two people are the same - and this increases the power of the test. What this means is that you are more likely to detect any significant differences, if they do exist, using the dependent t-test versus the independent t-test.

Can the dependent t-test be used to compare different subjects?

Yes, but this does not happen very often. You can use the dependent t-test instead of using the usual independent t-test when each subject in one of the independent groups is closely related to another subject in the other group on many individual characteristics. This approach is called a "matched-pairs" design. The reason we might want to do this is that the major advantage of running a within-subject (repeated-measures) design is that you get to eliminate between-groups variation from the equation (each individual is unique and will react slightly differently than someone else), thereby increasing the power of the test. Hence, the reason why we use the same subjects - we expect them to react in the same way as they are, after all, the same person. The most obvious case of when a "matched-pairs" design might be implemented is when using identical twins. Effectively you are choosing parameters to match your subjects on which you believe will result in each pair of subjects reacting in a similar way.

How do I report the result of a dependent t-test?

You need to report the test as follows:



where df is N - 1, where N = number of subjects.

Should I report confidence levels?

Confidence intervals (CI) are a useful statistic to include as they indicate the direction and size of a result. It is common to report 95% confidence intervals, which you will most often see reported as 95% CI. Programmes such as SPSS will automatically calculate these confidence intervals for you otherwise you need to calculate them by hand. You will want to report the mean and 95% confidence levels for the differences between the two related groups.

If you wish to run a dependent t-test in SPSS then you can find out how to do this in our guide here.

One-way ANOVA

35

What is this test for?

The one-way analysis of variance (ANOVA) is used to determine whether there are any significant differences between the means of three or more independent (unrelated) groups. This guide will provide a brief introduction to the one-way ANOVA including the assumptions of the test and when you should use this test. If you are familiar with the one-way ANOVA then you can skip this guide and go straight to how to run this test in SPSS by clicking here.

What does this test do?

The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are significantly different from each other. Specifically, it tests the null hypothesis:

$$H_O: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

where μ = group mean and k = number of groups. If, however, the one-way ANOVA returns a significant result then we accept the alternative hypothesis (H_A), which is that there are at least 2 group means that are significantly different from each other.

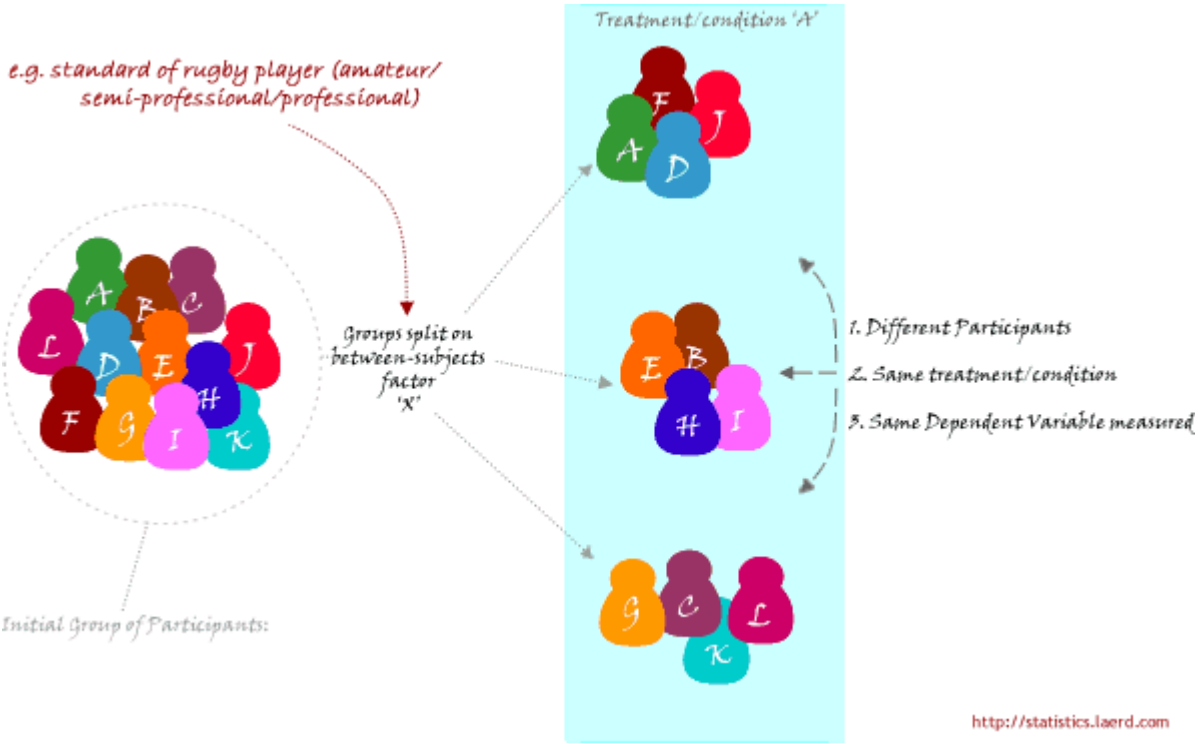
At this point, it is important to realise that the one-way ANOVA is an *omnibus* test statistic and cannot tell you which specific groups were significantly different from

each other, only that at least two groups were. To determine which specific groups differed from each other you need to use a *post-hoc* test. *Post-hoc* tests are described later in this guide.

When might you need to use this test?

If you are dealing with individuals, you are likely to encounter this situation using two different types of study design:

One study design is to recruit a group of individuals and then randomly split this group into 3 or more smaller groups, i.e. each subject is allocated to one, and only one, group. You then get each group to undertake different tasks (or put them under different conditions) and measure the outcome/response on the same dependent variable. For example, a researcher wishes to know whether different pacing strategies affect the time to complete a marathon. The researcher randomly assigns a group of volunteers to either a group that (a) starts slow and then increases their speed, (b) starts fast and slows down or (c) runs at a steady pace throughout. The time to complete the marathon is the outcome (dependent) variable. This study design is illustrated schematically in the Figure below:



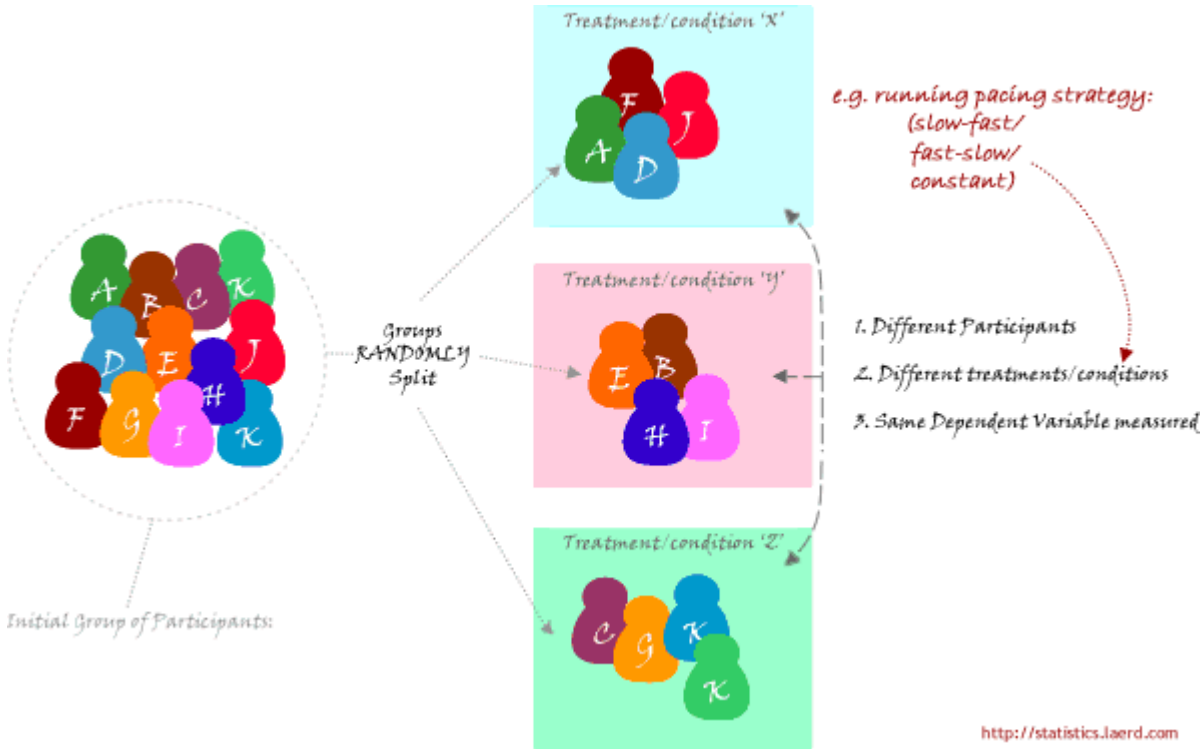
When you might use this test is continued on the next page.

One-way ANOVA (cont...)

When might you need to use this test? (cont...)

A second study design is to recruit a group of individuals and then split them into groups based on some independent variable. Again, each individual will be assigned to one group only. This independent variable is sometimes called an

attribute independent variable because you are splitting the group based on some attribute that they possess, e.g. their level of education; every individual has a level of education, even if it is "none". Each group is then measured on the same dependent variable having undergone the same task or condition (or none at all). For example, a researcher is interested in determining whether there are differences in leg strength between amateur, semi-professional and professional rugby players. The force/strength measured on an isokinetic machine is the dependent variable. This type of study design is illustrated schematically in the Figure below:



Why not compare groups with multiple t-tests?

Every time you conduct a t-test there is a chance that you will make a Type 1 error. An ANOVA controls for these errors so that the Type 1 error remains at 5% and you can be more confident that any significant result you find is not just down to chance. See our guide on hypothesis testing for more information on Type I errors.

What assumptions does the test make?

There are three main assumptions, listed here:

1. The dependent variable is normally distributed in each group that is being compared in the one-way ANOVA. So, for example, if we were comparing three groups; amateur, semi-professional and professional rugby players; on their leg strength, then their leg strength values (dependent variable) would have to be normally distributed for the amateur group of players, normally distributed for the semi-professionals and normally distributed for

the professional players. You can test for normality in SPSS (see our guide here).

2. There is homogeneity of variances. This means that the population variances in each group are equal. If you use SPSS, Levene's Test for Homogeneity of Variances is included in the output when you run a one-way ANOVA in SPSS (see our One-way ANOVA using SPSS guide).
3. This is a study design issue that you will so you will need to examine your study design to determine whether this could have occurred.

What to do when the assumptions are not met is dealt with on the next page.

One-way ANOVA (cont...)

5

What happens if my data fail these assumptions?

Firstly, don't panic! The first two of these assumptions are easily fixable, even if the last assumption is not. Lets go through the options as above:

1. The one-way ANOVA is considered a robust test against the normality assumption. This means that it tolerates violations to its normality assumption rather well. As regards the normality of group data, the one-way ANOVA can tolerate data that is non-normal (skewed or kurtotic distributions) with only a small effect on the Type I error rate. However, platykurtosis can have a profound effect when your group sizes are small. This leaves you with two options: (1) transform your data using various algorithms so that the shape of your distributions become normally distributed (see our normality guide here) or (2) choose the non-parametric Kruskal-Wallis H Test which does not require the assumption of normality (read our guide on this test here).
2. There are two tests that you can run that are applicable when the assumption of homogeneity of variances has been violated: (1) Welch or (2) Brown and Forsythe test. Alternatively, you could run a Kruskal-Wallis H Test. For most situations it has been shown that the Welsh test is best. Both the Welch and Brown and Forsythe tests are available in SPSS (see our One-way ANOVA using SPSS guide).
3. A lack of independence of cases has been stated as the most important assumptions to fail. Often, there is little you can do that offers a good solution to this problem. A full explanation of this problem and all assumptions mentioned here, including numerical explanations, are provided in Intermediate Statistics: A Modern Approach by Dr James Stevens.

How do I run a one-way ANOVA?

There are numerous ways to run a one-way ANOVA, however, we provide a comprehensive, step-by-step guide on how to do this using SPSS.

How do I report the results of a one-way ANOVA?

You will have calculated the following results or obtained them from SPSS:

Structure of results:

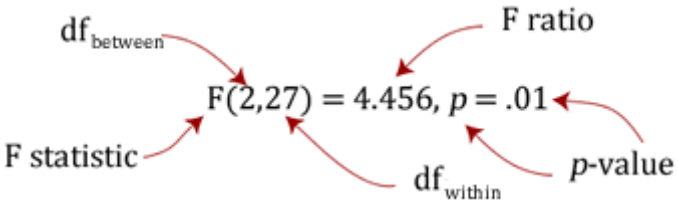
Source	*SS*	*df*	*MS*	*F*	*Sig.*
Between	SS_b	k-1	MS_b	MS_b/MS_w	*p* value
Within	SS_w	N-k	MS_w		
Total	$SS_b + SS_w$	N-1			

An example:

Source	*SS*	*df*	*MS*	*F*	*Sig.*
Between	91.476	2	45.733	4.467	.021
Within	276.400	27	10.237		
Total	367.867	29			

You will want to report this as follows:

There was a statistically significant difference between groups as determined by one-way ANOVA ($F(2,27) = 4.467$, $P = .021$). This is all you will need to write for the one-way ANOVA per se. However, in reality you will want probably also want to report means ± SD for your groups as well as follow-up a significant result with *post-hoc* tests. If you use SPSS then these descriptive statistics will be reported in the output along with the result from the one-way ANOVA. The general form of writing the result of a one-way ANOVA is as follows:



where df = degrees of freedom.

It is very important that you do not report the result as "significant difference" but that you report it as "statistically significant difference". This is because your decision as to whether the result is significant or not should not be based solely on your statistical test. Therefore, to indicate to readers that this "significance" is a statistical one, include this is your sentence.

Find out what else you have to do when you have a significant or a not-significant ANOVA result on the next page.

One-way ANOVA (cont...)

12

My p-value is greater than 0.05, what do I do now?

Report the result of the one-way ANOVA, e.g. "there were no statistically significant differences between group means as determined by one-way ANOVA ($F(2,27) = 1.397$, $P = .15$)". Not achieving a statistically significant result does not mean you should not report group means +/- SD also. However, running post-hoc tests is not warranted and should not be carried out.

My p-value is less than 0.05, what do I do now?

Firstly, you need to report your results as highlighted in the "How do I report the results?" section above. You then need to follow-up the one-way ANOVA by running *post-hoc* tests.

Homogeneity of variances was violated; how to continue?

You need to perform the same procedures as in the above three sections but add into your results section that this assumption was violated and you needed to run a Welch F test.

What are post-hoc tests?

Recall from earlier that the ANOVA test tells you whether you have an overall difference between your groups but it does not tell you which specific groups differed - post-hoc tests do. Because *post-hoc* tests are run to confirm where the differences occurred between groups, they should, therefore, only be run when you have a shown an overall significant difference in group means (i.e. a significant one-way ANOVA result). Post-hoc tests attempt to control the experimentwise error rate usually alpha = 0.05) in the same manner that the one-way ANOVA is used instead of multiple t-tests. *Post-hoc* tests are termed a posteriori tests - that is, performed after the event (the event in this case being a study).

Which *post-hoc* test should I use?

There are a great number of different *post-hoc* tests that you can use, however, you should only run one *post-hoc* test - do not run multiple *post-hoc* tests. For a one-way ANOVA, you will probably find that just one of four tests need to be considered. If your data meet the assumption of homogeneity of variances then either use the Tukey's honestly significant difference (HSD) or Scheffé *post-hoc* tests. Often, Tukey's HSD test is recommended by statisticians as it is not as conservative as the Scheffe test (which means that you are more likely to detect differences if they exist with Tukey's HSD test). Note that if you use SPSS, Tukey's HSD test is simply referred to as "Tukey" in the post-hoc multiple comparisons dialogue box). If your data did not meet the homogeneity of variances assumption then you should consider running either the Games Howell or Dunnett's C *post-hoc* test. The Games Howell test is generally recommended.

How should I graphically present my results?

First off, it is not essential that you present your results in a graphical form. However, it can add a lot of clarity to your results. There are a few key points to producing a good graph. Firstly, you need to present error bars for each group mean. It is customary to use the standard deviation of each group but standard

error and confidence limits are also used in the literature. You should also make sure that the scale is appropriate for what you are measuring. These points and more are discussed in our guide on selecting an appropriate graph (guide here). Generally, if graphically presenting data from an ANOVA, we recommend using a bar chart with standard deviation bars.

What to do now?

Now that you understand the one-way ANOVA, go to our guide on how to run the test in SPSS here.

ANOVA with Repeated Measures using SPSS

72

Overview

An ANOVA with repeated measures is for comparing three or more group means where the participants are the same in each group. This usually occurs in two situations - when participants are measured multiple times to see changes to an intervention or when participants are subjected to more than one condition/trial and the response to each of these conditions wants to be compared. For a complete guide on ANOVA with Repeated Measures, please go to our guide here.

Example

Heart disease is one of the largest causes of premature death and it is now known that chronic, low-level inflammation is a cause of heart disease. Exercise is known to have many benefits including protection against heart disease. A researcher wished to know whether this protection against heart disease might be afforded by exercise reducing inflammation. The researcher was also curious as to whether this protection might be gained over a short period of time or whether it took longer. In order to investigate this idea the researcher recruited 20 participants who underwent a 6-month exercise training program. In order to determine whether inflammation had been reduced, he measured the inflammatory marker called CRP pre-training, 2 weeks into training and post-6-months-training.

Upgrade to our Premium content »

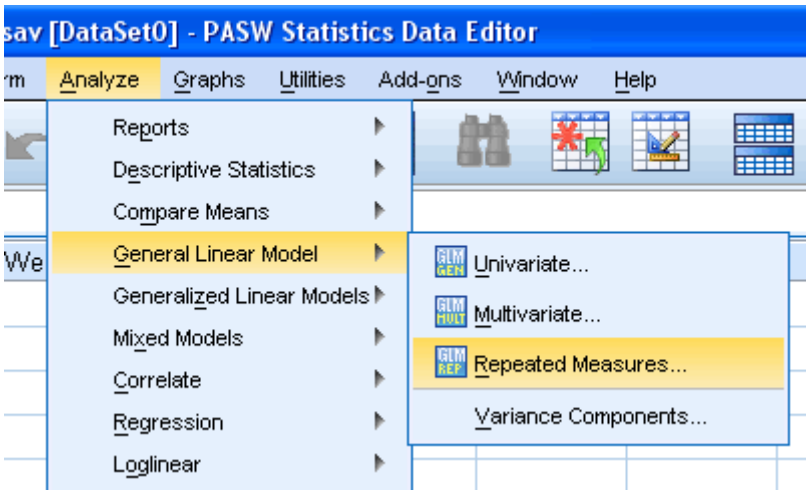
Assumptions

- The dependent variable is **interval or ratio** (**continuous**). (see our guide on Types of Variable)
- Dependent variable is **approximately normally distributed**. (see SPSS article on Testing for Normality)
- **Sphericity** (see our guide on Sphericity).
- One independent variable where participants are tested on the same dependent variable at least 2 times.

Test Procedure in SPSS

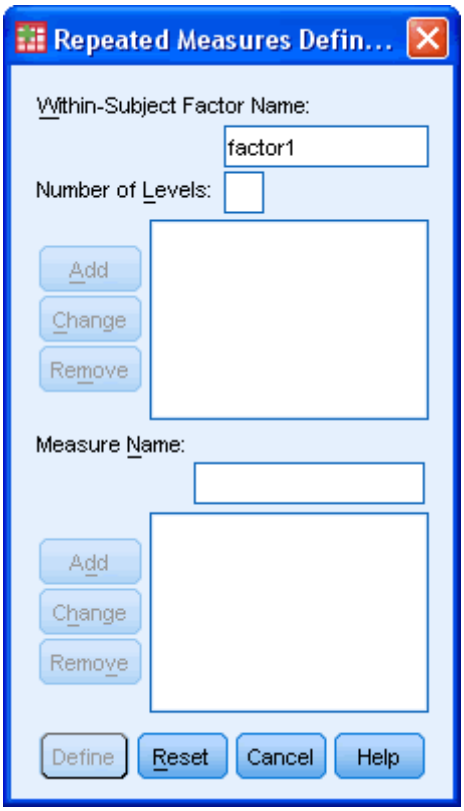
[To know how to correctly enter your data into SPSS in order to run a repeated measures ANOVA please read our Entering Data in SPSS tutorial.]

1. Click **Analyze > General Linear Model > Repeated Measures...** on the top menu as shown below:



Published with written permission from SPSS Inc, an IBM Company.

2. You will be presented with the following screen:



Published with written permission from SPSS Inc, an IBM Company.

3. In the "Within-Subject Factor Name:" replace "factor1" with a name that is more meaningful name for your independent variable. In our example, we will call our within-subject factor name "**time**" as it represents the different times that we took CRP measurements from our participants (pre-training, pre + 2 weeks and post-training).

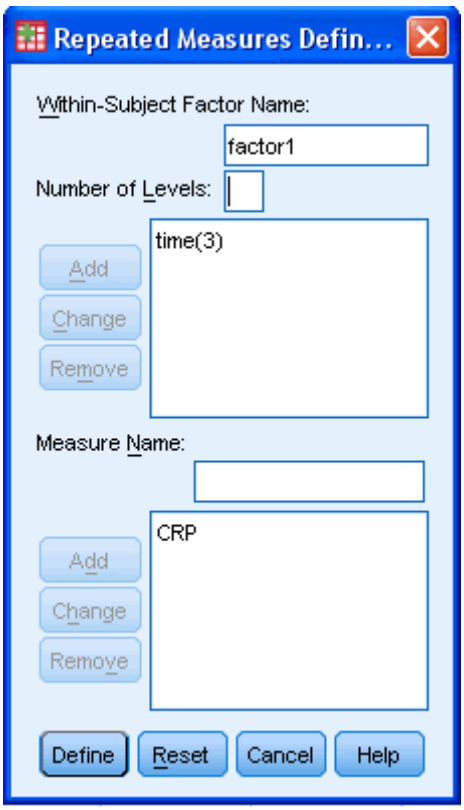
   Enter into the "Number of Levels:" box the number of times the dependent variable has been measured. In this case, enter "**3**", representing pre-training, pre + 2 weeks and post-training.

   Click the Add button.

   Put an appropriate name for your dependent variable in the "Measure Name:" box. In this case we have labelled our dependent variable CRP.

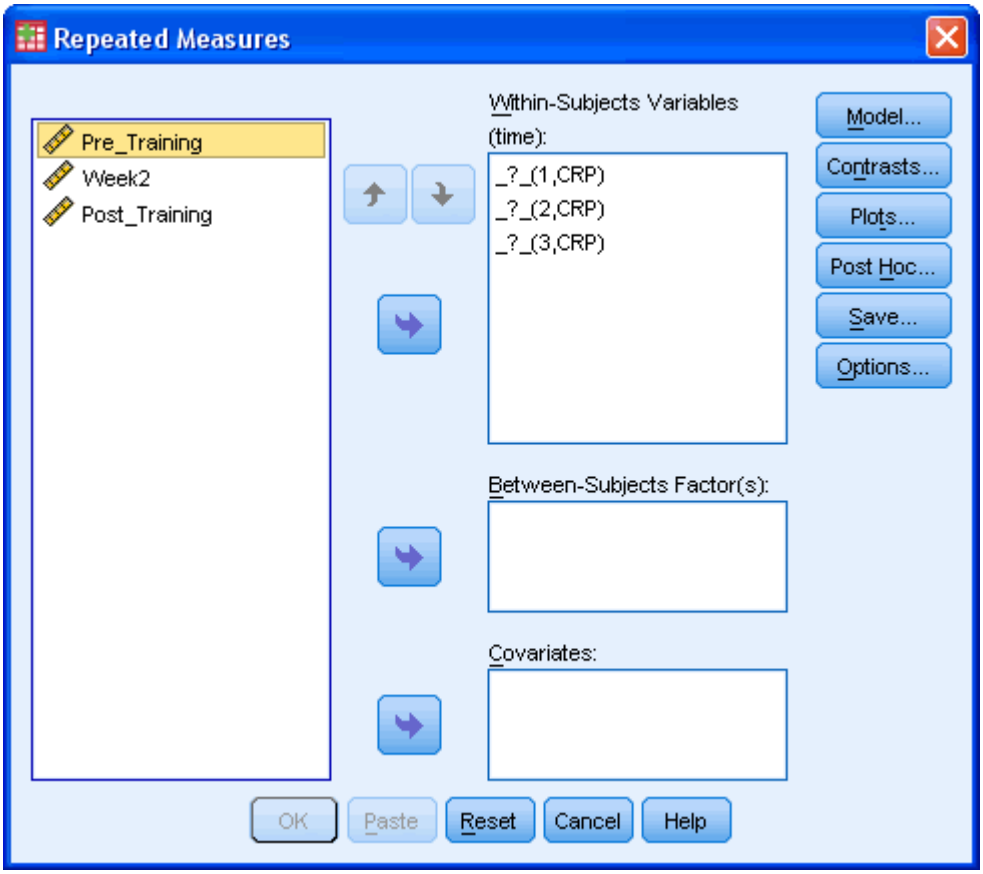
   Click the Add button.

   You will be presented with the diagram screen below:



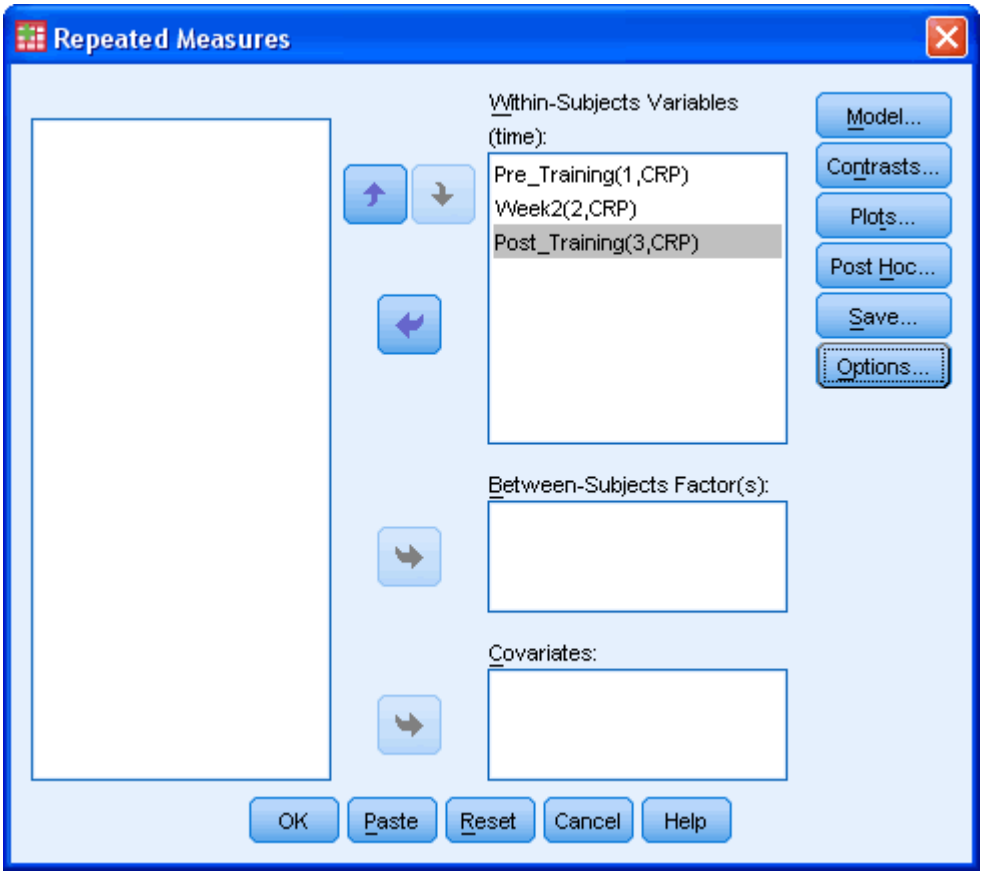
Published with written permission from SPSS Inc, an IBM Company.

4. Click the Define button and you will be presented with the following screen:



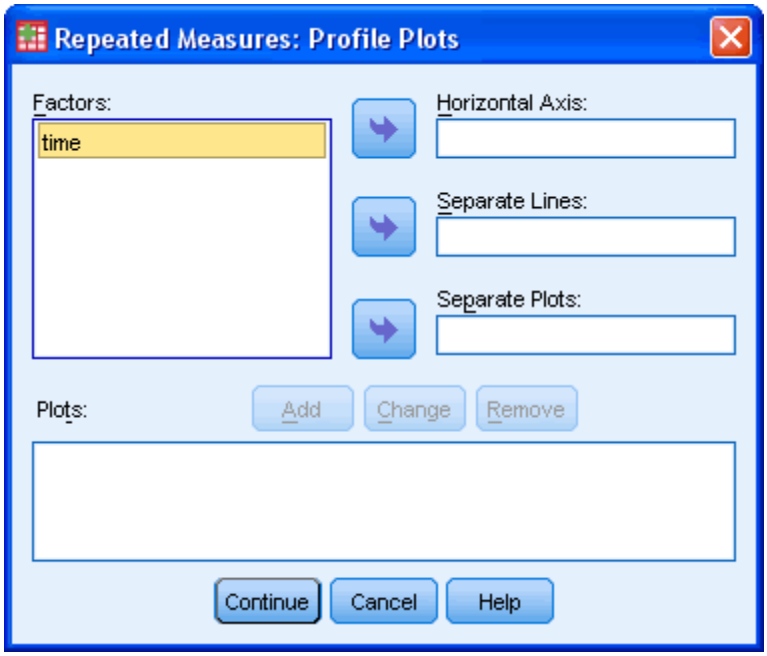
Published with written permission from SPSS Inc, an IBM Company.

5. Transfer "**Pre_Training**", "**Week2**" and "**Post_Training**" into the "Within-Subjects Variables (time):" box by either drag-and-dropping or using the button. If you make a mistake you can use the and buttons to reorder your variables.



Published with written permission from SPSS Inc, an IBM Company.

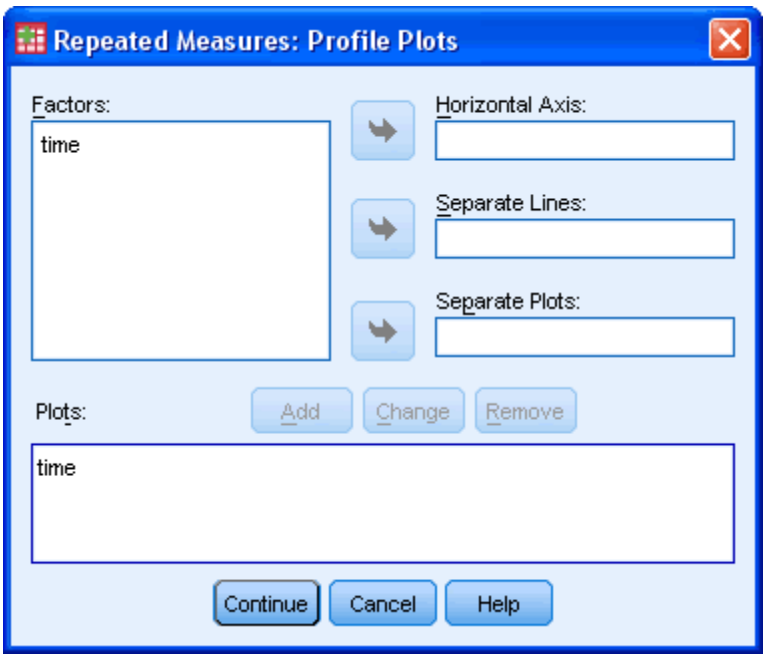
6. Click the Plots... button. You will be presented with the following screen:



Published with written permission from SPSS Inc, an IBM Company.

7. Transfer the "**time**" factor from the "Factors:" box into the "Horizontal Axis:" box by either drag-and-drop or the button.

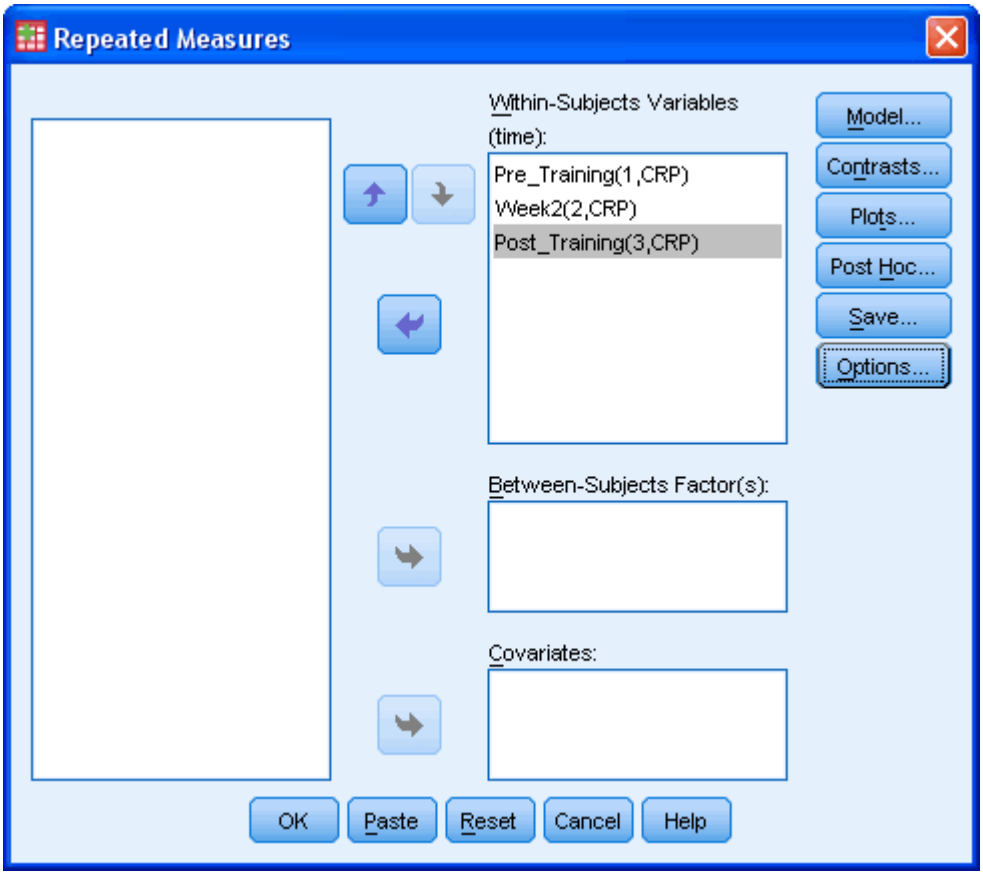
Click the Add button. You will be presented with the following screen:



Published with written permission from SPSS Inc, an IBM Company.

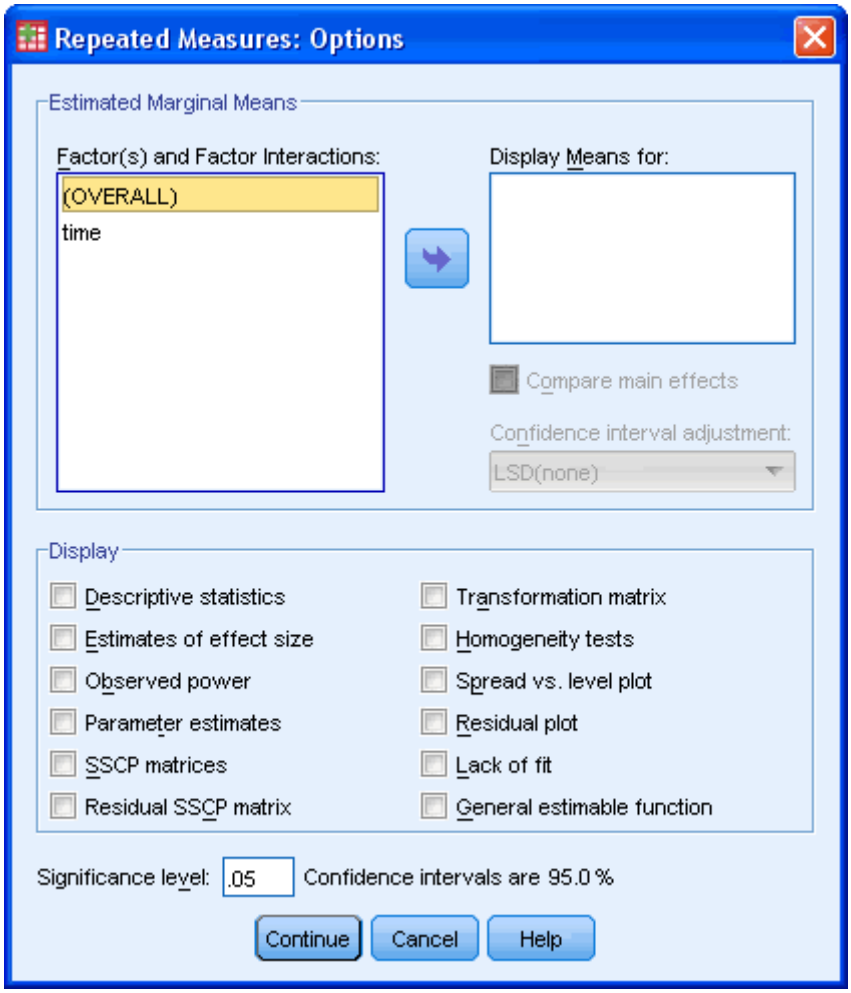
Click the Continue button.

You will be presented with the following screen:



Repeated Measures ANOVA using SPSS (cont...)

8. Click the Options... button. You will be presented with the following screen:

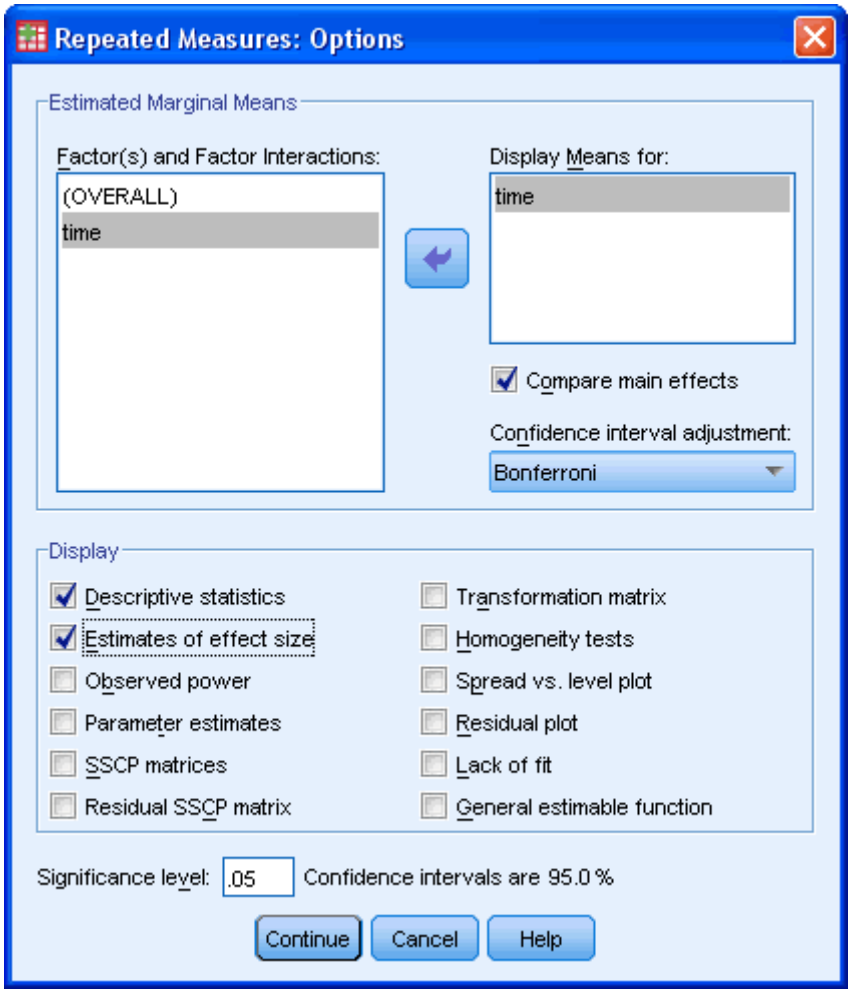


Published with written permission from SPSS Inc, an IBM Company.

9. Transfer the factor "time" from the "Factor(s) and Factor Interactions:" box to the "Display Means for:" box.

   Tick the "Compare main effects" checkbox and select "Bonferroni" from the drop-down menu under "Confidence interval adjustment:".

   Tick the "Descriptive statistics" and "Estimates of effect size" checkboxes in the "Display" area. If you have done all this you will be presented with the following screen:



Published with written permission from SPSS Inc, an IBM Company.

Click the Continue button.

10. Click the OK button.

SPSS Output

SPSS will output a great many tables but for this analysis there are only four tables and one graph that we are interested in. We will go through these tables in order:

Within-Subjects Factors Table

This table reminds us of the levels of our independent variable (within-subject factor) and labels the time points 1, 2 and 3. We will need these labels later on when analysing our results in the **Pairwise Comparisons** table. Take care not to get confused with the "Dependent Variable" column in this table as it seems to suggest that the different time points are our dependent variable. This is not true - the column label is referring to fact that the dependent variable "CRP" is measured at each of these time points.

Within-Subjects Factors

Measure:CRP

time	Dependent Variable
1	Pre_Training
2	Week2
3	Post_Training

Published with written permission from SPSS Inc, an IBM Company.

Descriptive Statistics Table

This table simply provides important descriptive statistics for this analysis as shown below:

Descriptive Statistics

	Mean	Std. Deviation	N
Pre_Training	3.0875	.97763	20
Week2	2.9675	.89388	20
Post_Training	2.2450	.49892	20

Published with written permission from SPSS Inc, an IBM Company.

Mauchly's Test of Sphericity Table

This table shows the results of Mauchly's Test of Sphericity which tests for one of the assumptions of the ANOVA with repeated measures, namely sphericity. It is important to look at this table as this assumption is commonly violated.

Mauchly's Test of Sphericity[b]

Measure:CRP

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	.293	22.115	2	.000	.586	.601	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: time

Published with written permission from SPSS Inc, an IBM Company.

We are interested in the **Approx. Chi-Square** value and its associated **Sig.** value. We can see from this example that the significance level is below 0.05 (it is < 0.0005). A significant value for Mauchly's Test of Sphericity indicates that the assumption of sphericity has been violated. Luckily, if this happens, methods have been developed that mean that we can still proceed with the test by using a correctional adjustment called Greenhouse-Geisser (a guide to sphericity can be found here).

Tests of Within-Subjects Effects Table

This table tells us if there was an overall significant difference between the means at the different time points.

Tests of Within-Subjects Effects

Measure:CRP

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
time	Sphericity Assumed	8.308	2	4.154	21.032	.000	.525
	Greenhouse-Geisser	8.308	1.171	7.092	21.032	.000	.525
	Huynh-Feldt	8.308	1.202	6.912	21.032	.000	.525
	Lower-bound	8.308	1.000	8.308	21.032	.000	.525
Error(time)	Sphericity Assumed	7.505	38	.198			
	Greenhouse-Geisser	7.505	22.257	.337			
	Huynh-Feldt	7.505	22.837	.329			
	Lower-bound	7.505	19.000	.395			

Published with written permission from SPSS Inc, an IBM Company.

From this table we are able to discover the *F* value for the "**time**" factor, its associated significance level and effect size (**Partial Eta Squared**). As our data violated the assumption of sphericity we look at the values in the **Greenhouse-Geisser** row (as indicated in red in the screenshot). Had sphericity not been violated we would have looked under the **Sphericity Assumed** row. We can report that when using an ANOVA with repeated measures with a Greenhouse-Geisser correction, the mean scores for CRP concentration were statistically significantly different ($F(1.171, 22.257) = 21.032$, $P < 0.0005$).

<u>Pairwise Comparisons Table</u>

The results presented in the previous table informed us that we have an overall significant difference in means but we do not know where those differences occurred. This table presents the results of the Bonferroni post-hoc test, which allows us to discover which specific means differed. Remember, if your overall ANOVA result was not significant then you should not examine the **Pairwise Comparisons** table.

Pairwise Comparisons

Measure:CRP

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	.120	.057	.149	-.030	.270
	3	.842*	.173	.000	.389	1.296
2	1	-.120	.057	.149	-.270	.030
	3	.723*	.162	.001	.298	1.147
3	1	-.842*	.173	.000	-1.296	-.389
	2	-.723*	.162	.001	-1.147	-.298

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

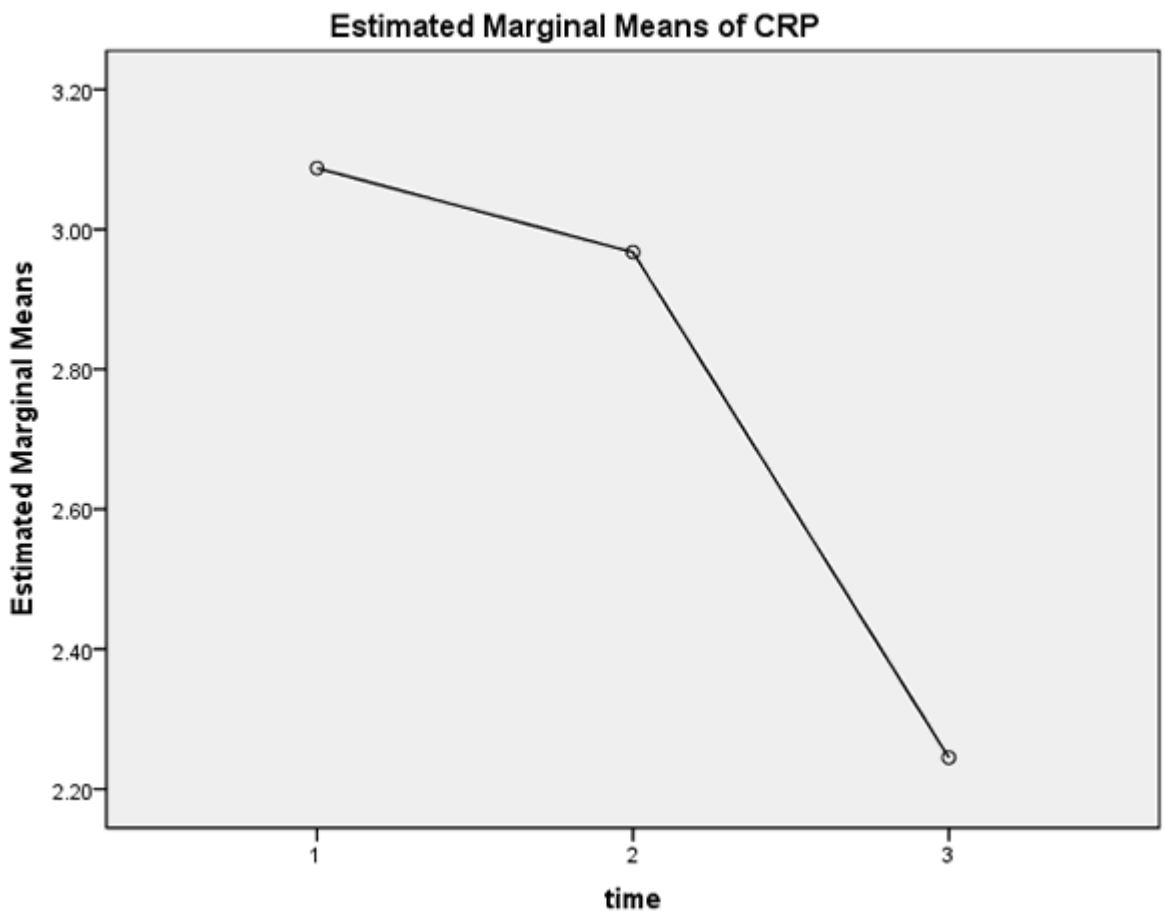
*. The mean difference is significant at the .05 level.

Published with written permission from SPSS Inc, an IBM Company.

Looking at the table above we need to remember the labels associated with the time points in our experiment from the **Within-Subject Factors** table. This table gives us the significance level for differences between the individual time points. We can see that there was a significant difference in CRP concentration between post-training and pre-training (P = 0.0005) and between post-training and after 2 weeks of training (P = 0.001) but no significant differences between pre-training and after 2 weeks of training (P = 0.149). From the **Mean Difference (I-J)** column we can see that CRP concentration was significantly reduced at this time point.

Profile Plot

This plot is the last element to this analysis. We are only including it so that you can see some of the limitations of doing so in its current format. SPSS has means of altering graphs axes. This is important as these profile plots always tend to exaggerate the differences between means but choosing a y-axis range of values that is too narrow. In this case, it is known that most people has CRP concentrations ranging from 0 to 3 so the profile plot that you should produce should take this into consideration. However, this plot can be useful in gaining an easy understanding of the tabular results.



Published with written permission from SPSS Inc, an IBM Company.

Reporting the Output

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that mean CRP concentration differed statistically significantly between time points ($F(1.171, 22.257) = 21.032$, $P < 0.0005$). Post hoc tests using the Bonferroni correction revealed that exercise training elciited a slight reduction in CRP concentration from pre-training to 2-weeks of training (3.09 ± 0.98 mg/L *vs.* 2.97 ± 0.89 mg/L, respectively) which was not statistically significant ($P = .149$). However, post-training CRP had been reduced to 2.24 ± 0.50 mg/L which was statistically significantly different to pre-training ($P < .0005$) and 2-weeks training ($P = .001$) concentrations. We can, therefore, conclude that a long-term exercise training program (6 months) elicits a statistically significant reduction in CRP concentration but not after only 2 weeks of training.

Two-way ANOVA using SPSS

121

Introduction

The two-way ANOVA compares the mean differences between groups that have been split on two independent variables (called factors). You need two independent, categorical variables and one continuous, dependent variable (see our guide on Types of Variable).

Assumptions

- Dependent variable is either **interval or ratio (continuous)** (see our guide on Types of Variable)
- The dependent variable is **approximately normally distributed** for each combination of levels of the two independent variables (see our Testing for Normality guide, which deals specifically with the two-way ANOVA).
- **Homogeneity of variances** of the groups formed by the different combinations of levels of the two independent variables.
- Independence of cases (this is a study design issue and is not addressed by SPSS).

You will need to run statistical tests in SPSS to check all of these assumptions when carrying out a two-way ANOVA. If you do not run the statistical tests on these assumptions correctly, the results you get when running a two-way ANOVA might not be valid. If you are unsure how to do this correctly, we show you how, step-by-step in our Premium guide. To learn about the Premium guide for two-way ANOVA, click here. To learn more about our Premium site in general, where access starts from just $3.99/£2.99/€3.99, Take the Tour.

Example

A researcher was interested in whether an individual's interest in politics was influenced by their level of education and their gender. They recruited a random sample of participants to their study and asked them about their interest in politics, which they scored from 0 - 100 with higher scores meaning a greater interest. The researcher then divided the participants by gender (Male/Female) and then again by level of education (School/College/University).

Setup in SPSS

In SPSS we separated the individuals into their appropriate groups by using two columns representing the two independent variables and labelled them "**Gender**" and "**Edu_Level**". For "**Gender**", we coded males as "**1**" and females as "**2**", and for "**Edu_Level**", we coded school as "**1**", college as "**2**" and university as "**3**". The participants interest in politics was entered under the variable name, "**Int_Politics**". To know how to correctly enter your data into SPSS in order to run a two-way ANOVA, please read our Entering Data in SPSS tutorial, where there is a specific example. The data setup can be seen in the diagram below (click image to see full data set). We have given our data text labels (see our Working with Variables guide).

	Gender	Edu_Level	Int_Politics	var
1	Male	School	34.00	
2	Male	School	35.00	
3	Male	School	32.00	
4	Male	School	35.00	
5	Male	School	40.00	
6	Male	School	40.00	
7	Male	School	37.00	
8	Male	School	33.00	
9	Male	School	34.00	
10	Male	School	28.00	
11	Male	College	49.00	
12	Male	College	50.00	
13	Male	College	47.00	

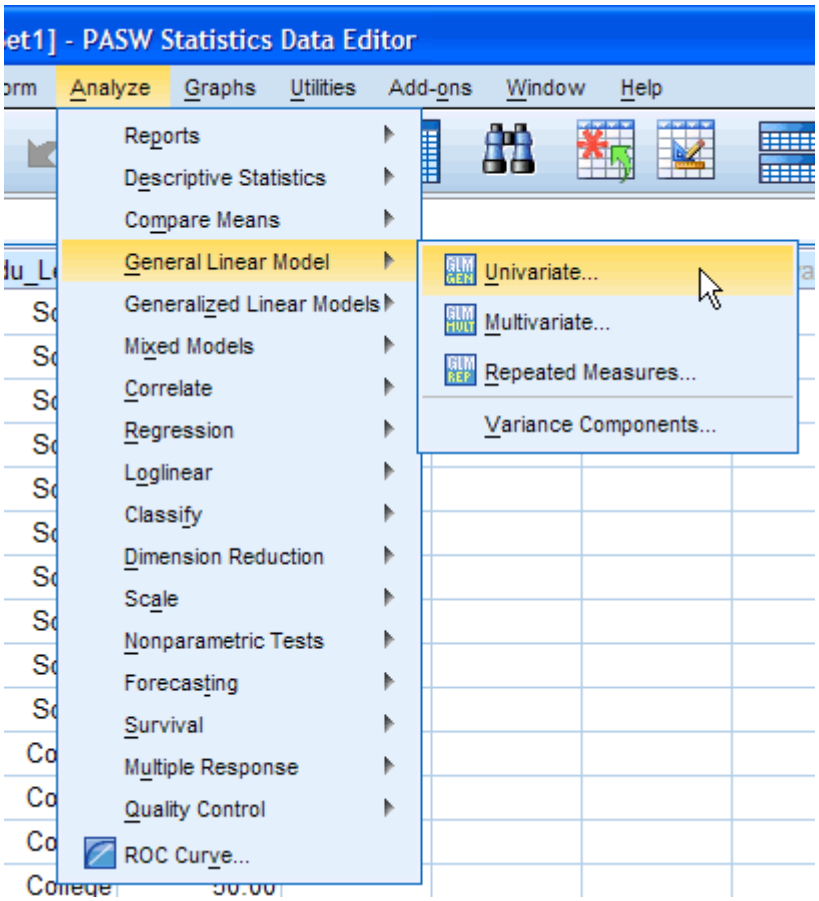
Published with written permission from SPSS Inc, an IBM Company.

Testing of Assumptions

To determine whether your dependent variable is normally distributed for each combination of the levels of the two independent variables see our Testing for Normality guide that runs through how to test for normality using SPSS using a specific two-way ANOVA example. In SPSS, homogeneity of variances is tested using Levene's Test for Equality of Variances. This is included in the main procedure for running the two-way ANOVA, so we get to evaluate whether there is homogeneity of variances at the same time as we get the results from the two-way ANOVA.

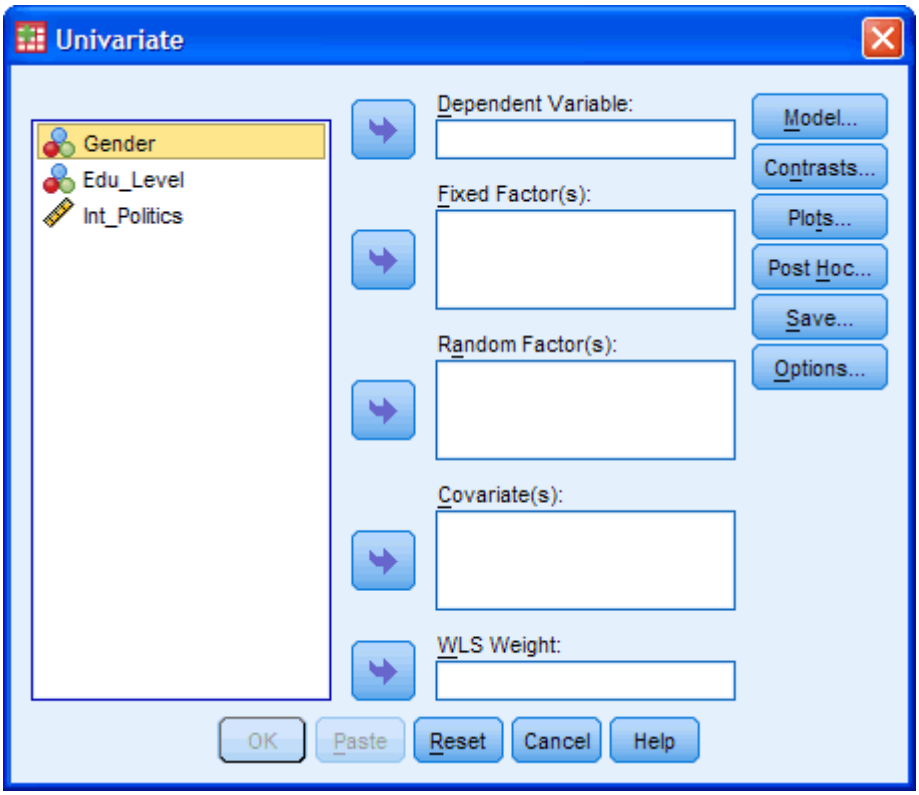
Test Procedure in SPSS

1. Click **Analyze > General Linear Model > Univariate...** on the top menu as shown below:



Published with written permission from SPSS Inc, an IBM Company.

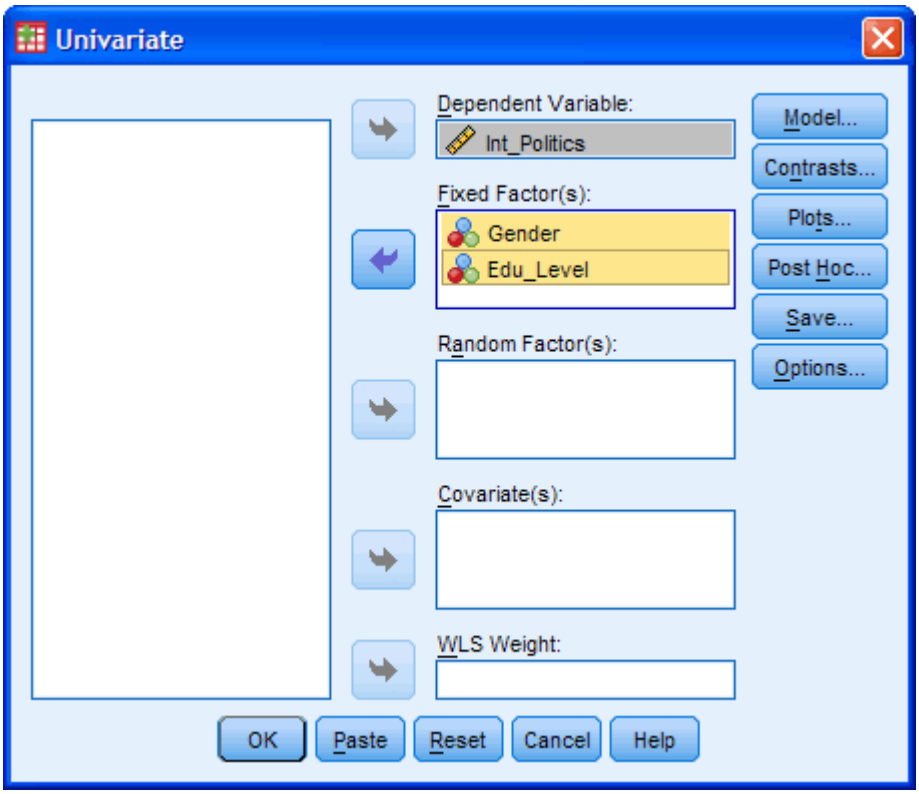
2. You will be presented with the "Univariate" dialogue box:



Published with written permission from SPSS Inc, an IBM Company.

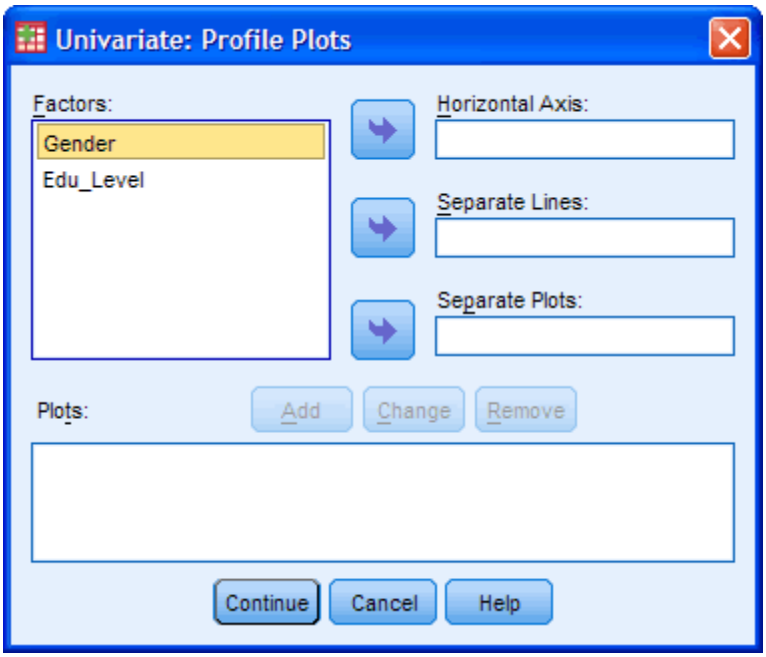
3. You need to transfer the dependent variable "**Int_Politics**" into the "*Dependent Variable:*" box and transfer both independent variables, "**Gender**" and "**Edu_Level**", into the "*Fixed Factor(s):*" box. You can do this by drag-and-dropping the variables into the respective boxes or by using the button. If you are using older versions of SPSS you will need to use the former method. The result is shown below:

   [For this analysis you will not need to worry about the "*Random Factor(s):*", "*Covariate(s):*" or "*WLS Weight:*" boxes.]



Published with written permission from SPSS Inc, an IBM Company.

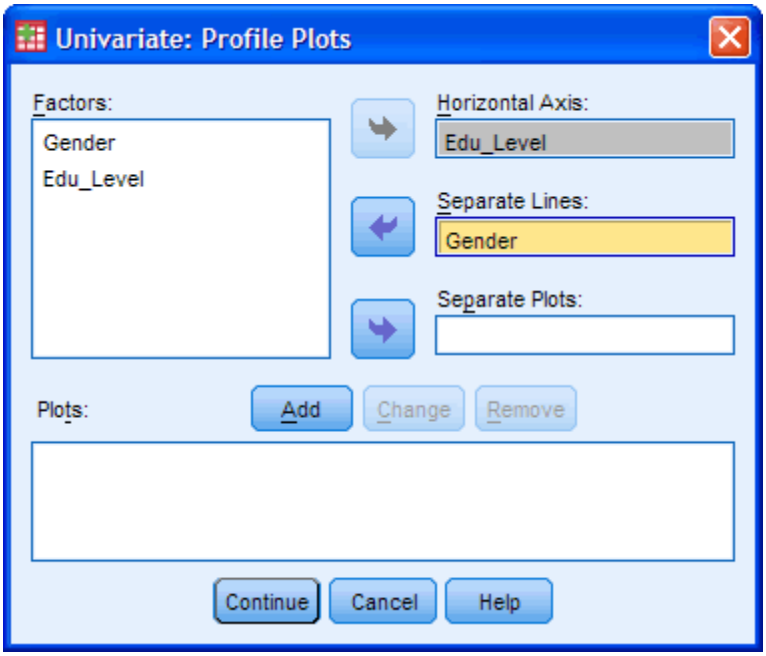
4. Click on the Plots... button. You will be presented with the "Univariate: Profile Plots" dialogue box:



Published with written permission from SPSS Inc, an IBM Company.

5. Transfer the independent variable "**Edu_Level**" from the "*Factors:*" box into the "*Horizontal Axis:*" box and transfer the "**Gender**" variable into the "*Separate Lines:*" box. You will be presented with the following screen:

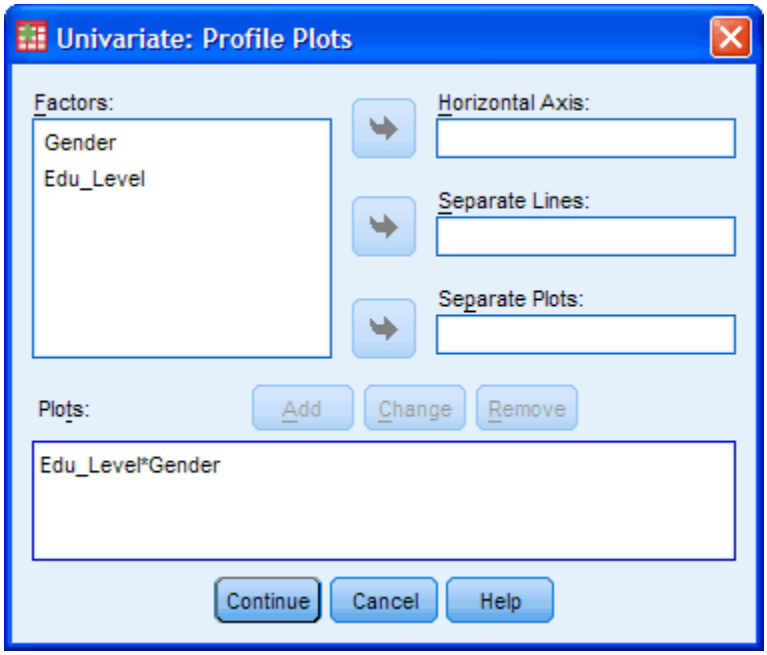
   [Tip: Put the independent variable with the greater number of levels in the "*Horizontal Axis:*" box.]



Published with written permission from SPSS Inc, an IBM Company.

6. Click the

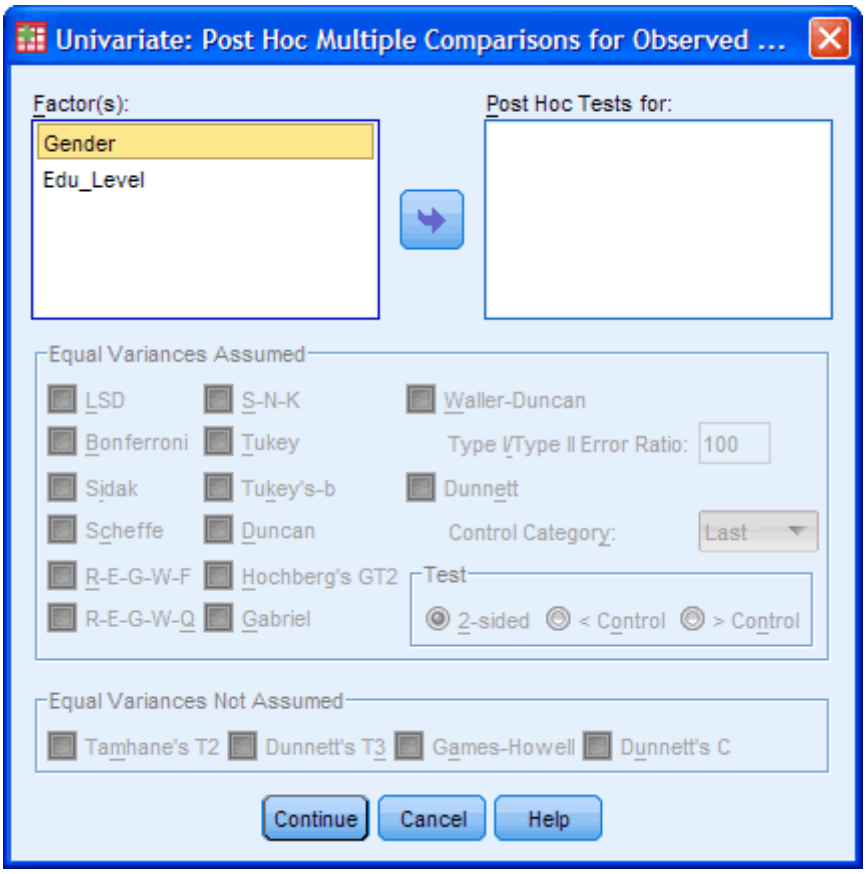
button.



Published with written permission from SPSS Inc, an IBM Company.

You will see that "**Edu_Level*Gender**" has been added to the "*Plots:*" box.

7. Click the Continue button. This will return you to the "Univariate" dialogue box.
8. Click the Post Hoc... button. You will be presented with the "Univariate: Post Hoc Multiple Comparisons for Observed..." dialogue box as shown below:

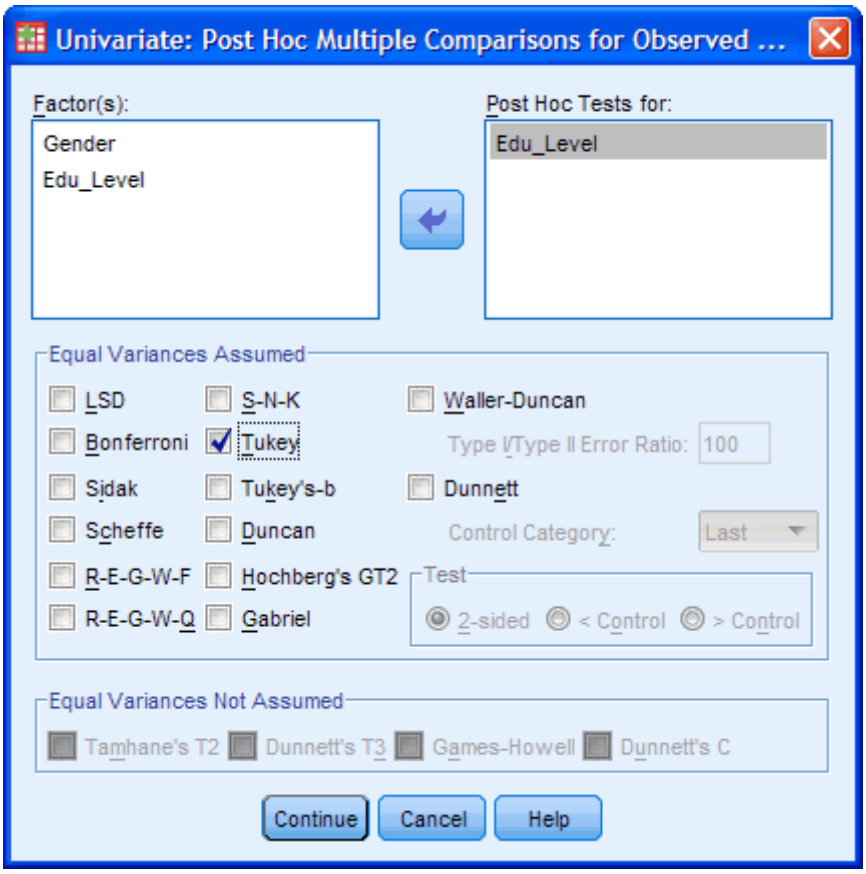


Published with written permission from SPSS Inc, an IBM Company.

Transfer "**Edu_Level**" from the "*Factor(s):*" box to the "*Post Hoc Tests for:*" box. This will make the "Equal Variances Assumed" section become active (loose the "grey sheen") and present you with some choices for which post-hoc test to use. For this example, we are going to select "Tukey", which is a good, all-round post-hoc test.

[You only need to transfer independent variables that have more than two levels into the "*Post Hoc Tests for:*" box. This is why we do not transfer "**Gender**".]

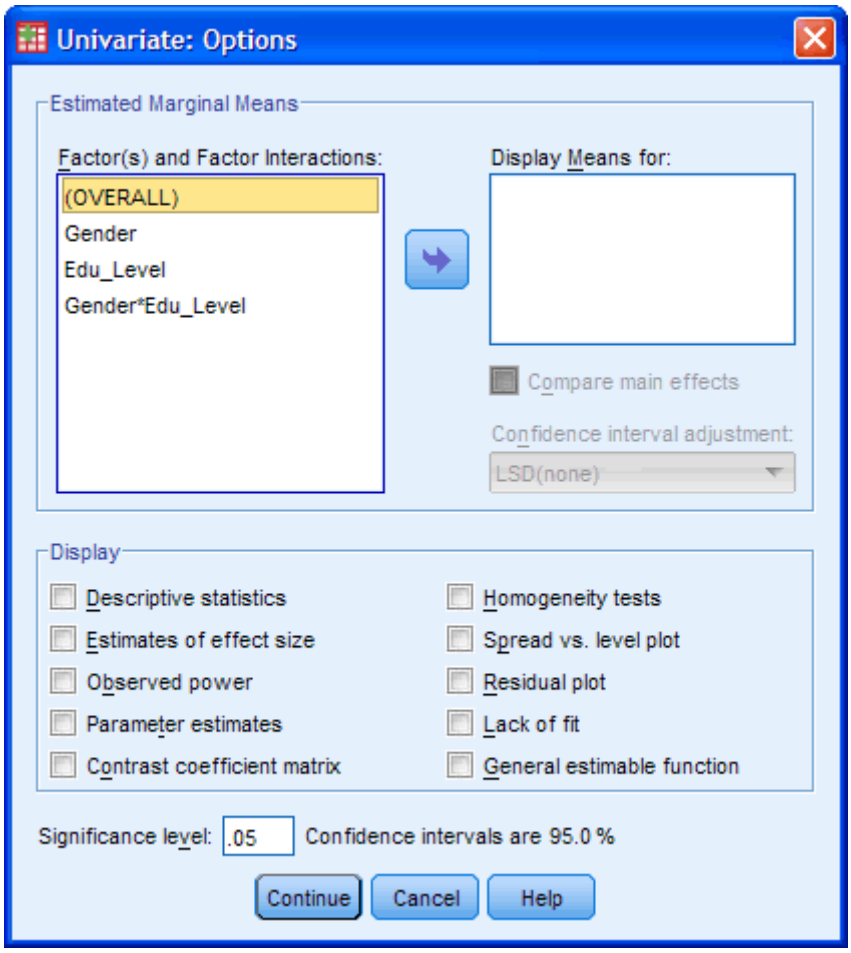
You will finish up with the following screen:



Published with written permission from SPSS Inc, an IBM Company.

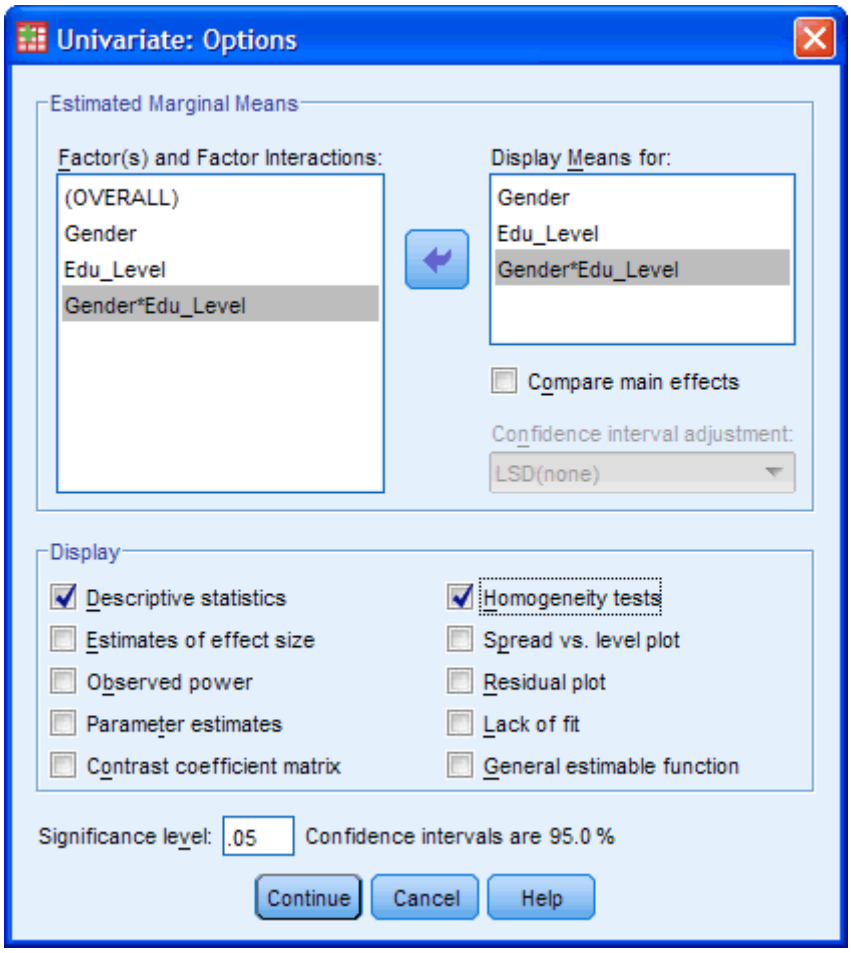
Click the Continue button to return to the "Univariate" dialogue box.

9. Click the Options... button. This will present you with the "Univariate: Options" dialogue box as shown below:



Published with written permission from SPSS Inc, an IBM Company.

10. Transfer "**Gender**", "**Edu_Level**" and "**Gender*Edu_Level**" from the "*Factor(s) and "Factor Interactions:*" box into the "*Display Means for:*" box. In the "Display" section, tick the "Descriptive Statistics" and "Homogeneity tests" options. You will presented with the following screen:



Published with written permission from SPSS Inc, an IBM Company.

Click the Continue button to return to the "Univariate" dialogue box.

11. Click the OK button to generate the output.

Go to the next page for the SPSS output, simple effects analysis and an explanation of the output.

Two-way ANOVA using SPSS (cont...)

49

SPSS Output of Two-way ANOVA

SPSS produces many tables in its output from a two-way ANOVA and we are going to start with the "Descriptives" table as shown below:

Descriptive Statistics

Dependent Variable:Int_Politics

Gender	Edu_Level	Mean	Std. Deviation	N
Male	School	38.2000	4.18463	10
	College	44.1000	4.26745	10
	University	64.1000	3.07137	10
	Total	48.8000	11.87841	30
Female	School	39.6000	3.27278	10
	College	44.6000	3.27278	10
	University	58.0000	6.46357	10
	Total	47.4000	9.05767	30
Total	School	38.9000	3.72615	20
	College	44.3500	3.71023	20
	University	61.0500	5.83524	20
	Total	48.1000	10.49649	60

Published with written permission from SPSS Inc, an IBM Company.

This table is very useful as it provides the mean and standard deviation for the groups that have been split by both independent variables. In addition, the table also provides "Total" rows, which allows means and standard deviations for groups only split by one independent variable or none at all to be known.

Levene's Test of Equality of Error Variances

The next table to look at is Levene's Test of Equality of Error Variances as shown below:

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Int_Politics

F	df1	df2	Sig.
1.106	5	54	.368

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Gender + Edu_Level + Gender * Edu_Level

Published with written permission from SPSS Inc, an IBM Company.

From this table we can see that we have homogeneity of variances of the dependent variable across groups. We know this as the **Sig.** value is greater than 0.05, which is the level we set for alpha. If the **Sig.** value had been less than 0.05 then we would have concluded that the variance across groups was significantly different (unequal).

Tests of Between-Subjects Effects Table

This table shows the actual results of the two-way ANOVA as shown below:

Tests of Between-Subjects Effects

Dependent Variable:Int_Politics

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5525.200[a]	5	1105.040	61.190	.000
Intercept	138816.600	1	138816.600	7686.727	.000
Gender	29.400	1	29.400	1.628	.207
Edu_Level	5328.100	2	2664.050	147.517	.000
Gender * Edu_Level	167.700	2	83.850	4.643	.014
Error	975.200	54	18.059		
Total	145317.000	60			
Corrected Total	6500.400	59			

a. R Squared = .850 (Adjusted R Squared = .836)

Published with written permission from SPSS Inc, an IBM Company.

We are interested in the Gender, Edu_Level and Gender*Edu_Level rows of the table as highlighted above. These rows inform us of whether we have significant mean differences between our groups for our two independent variables, Gender and Edu_Level, and for their interaction, Gender*Edu_Level. We must first look at the Gender*Edu_Level interaction as this is the most important result we are after. We can see from the **Sig.** column that we have a statistically significant interaction at the $P = .014$ level. You may wish to report the results of Gender and Edu_Level as well. We can see from the above table that there was no significant difference in interest in politics between Gender ($P = .207$) but there were significant differences between educational levels ($P < .0005$).

Multiple Comparisons Table

This table shows the Tukey post-hoc test results for the different levels of education as shown below:

Multiple Comparisons

Int_Politics
Tukey HSD

(I) Edu_Level	(J) Edu_Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
School	College	-5.4500*	1.34385	.000	-8.6887	-2.2113
	University	-22.1500*	1.34385	.000	-25.3887	-18.9113
College	School	5.4500*	1.34385	.000	2.2113	8.6887
	University	-16.7000*	1.34385	.000	-19.9387	-13.4613
University	School	22.1500*	1.34385	.000	18.9113	25.3887
	College	16.7000*	1.34385	.000	13.4613	19.9387

Based on observed means.
The error term is Mean Square(Error) = 18.059.

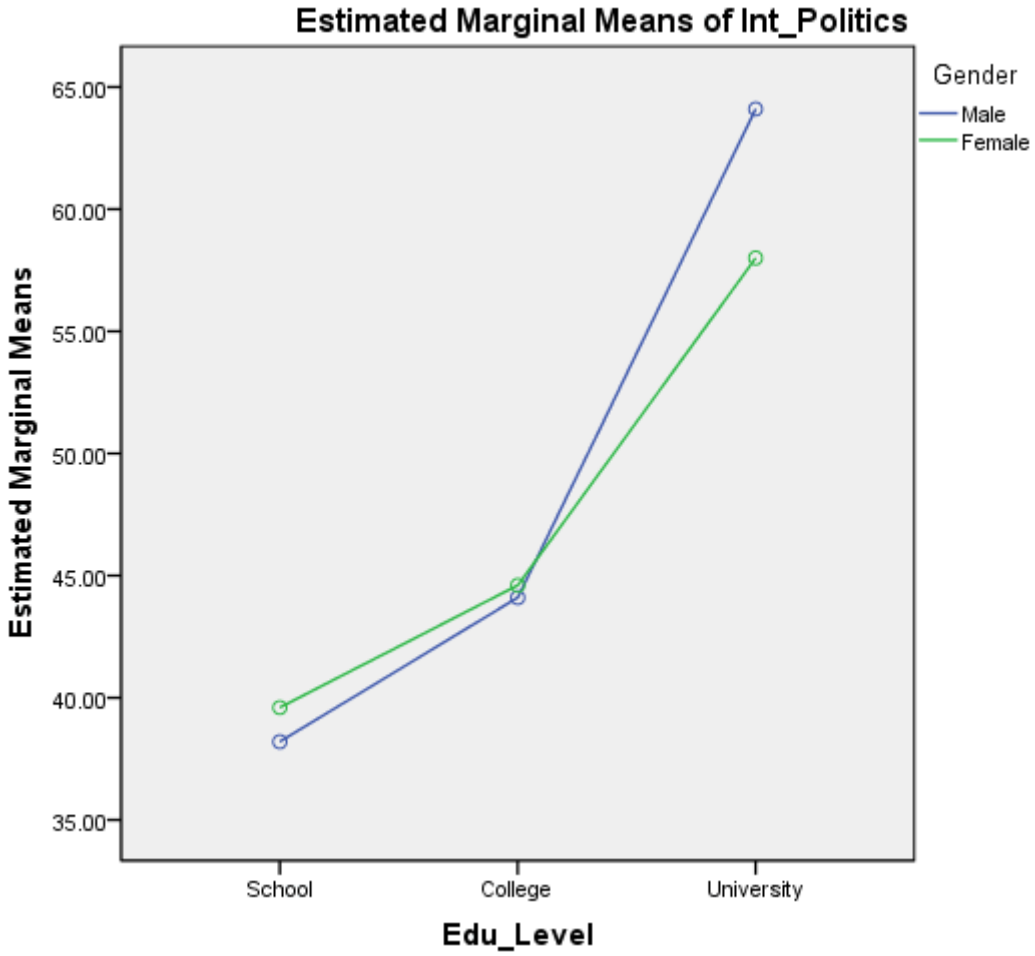
*. The mean difference is significant at the .05 level.

Published with written permission from SPSS Inc, an IBM Company.

We can see form the above table that there is some repetition of the results but, regardless of which row we choose to read from, we are interested in the differences between (1) School and College, (2) School and University, and (3) College and University. From the results we can see that there is a significant difference between all three different combinations of educational level (P < .0005).

Plot of the Results

The following plot is not of sufficient quality to present in your reports but provides a good graphical illustration of your results. In addition, we can get an idea of whether there is an interaction effect by inspecting whether the lines are parallel or not.



Published with written permission from SPSS Inc, an IBM Company.

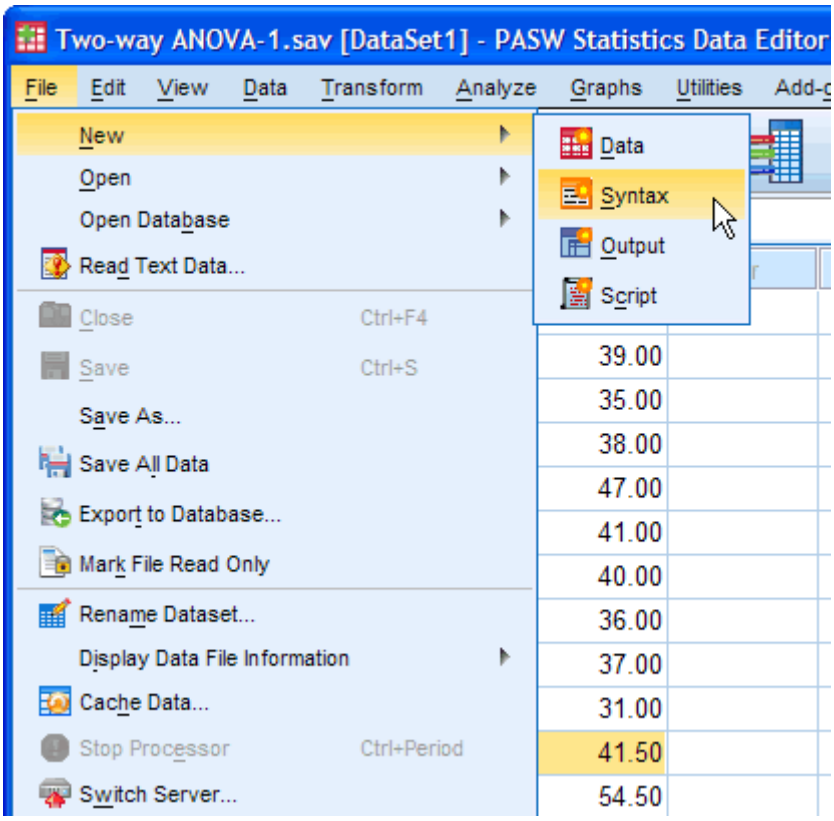
From this plot we can see how our results from the previous table might make sense. Remember that if the lines are not parallel then there is the possibility of an interaction taking place.

Procedure for Simple Main Effects in SPSS

You can follow up the results of a significant interaction effect by running tests for simple main effects - that is, the mean difference in interest in politics between genders at each education level. SPSS does not allow you to do this using the

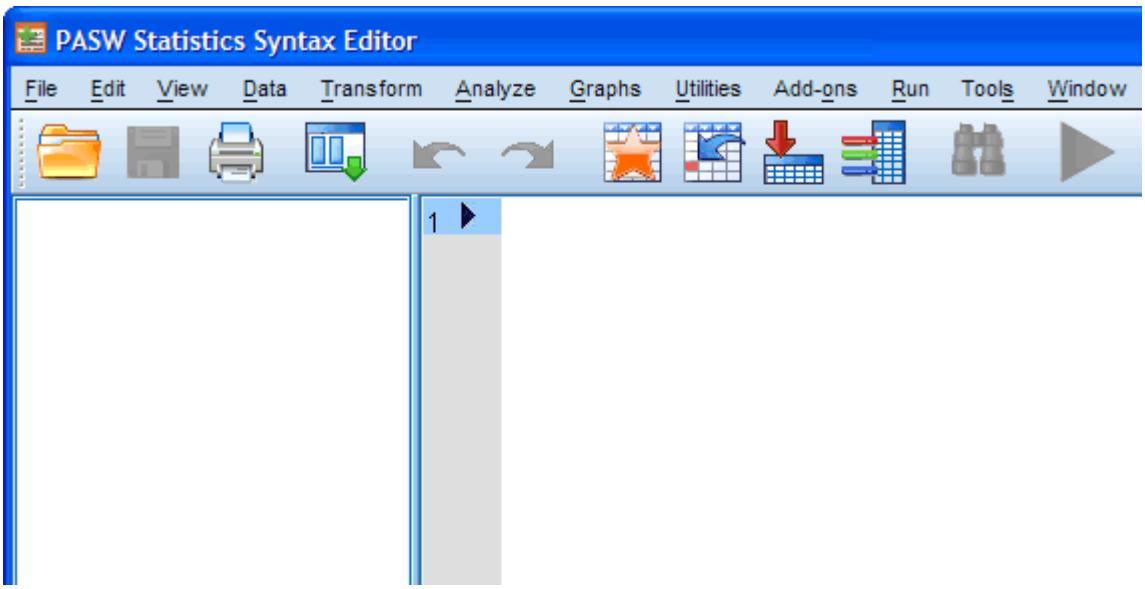
graphical interface you will be familiar with, but requires you to use syntax. We explain how to do this below:

1. Click **File > New > Syntax** from the main menu as shown below:



Published with written permission from SPSS Inc, an IBM Company.

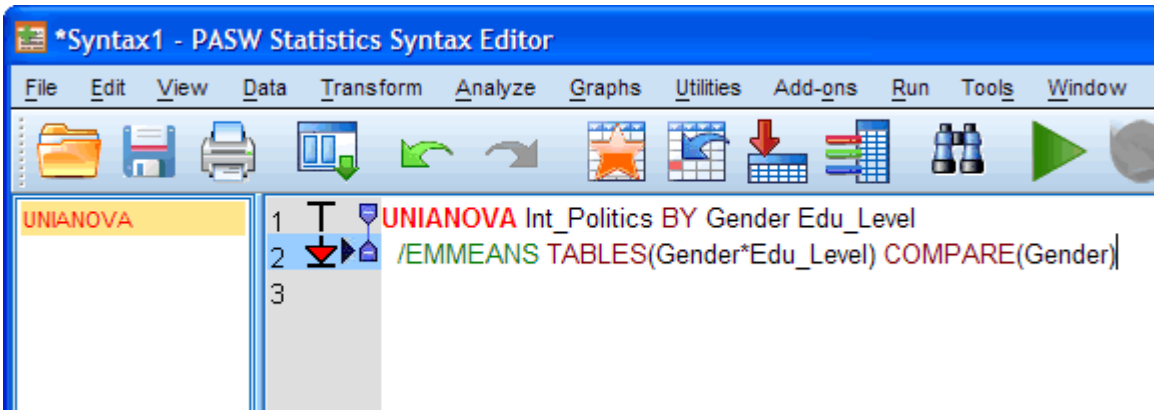
2. You will be presented with the Syntax Editor as shown below:



Published with written permission from SPSS Inc, an IBM Company.

3. Type text into the syntax editor so that you end up with the following (the colours are automatically added):

   [Depending on the version of SPSS you are using you might have suggestion boxes appear when you type in SPSS-recognised commands, such as, UNIANOVA. If you are familiar with using this type of auto-prediction then please feel free to do so, but otherwise simply ignore the pop-up suggestions and keep typing normally.]



Published with written permission from SPSS Inc, an IBM Company.

   Basically, all text you see above that is in CAPITALS, is required by SPSS and does not change when you enter your own data. Non-capitalised text represents your variables and will change when you use your own data. Breaking it all down, we have:

UNIANOVA	Tells SPSS to use the Univariate Anova command
Int_Politics BY Gender Edu_Level	Your dependent variable BY your two independent variables (with a space between them)
/EMMEANS	Tells SPSS to calculate estimated marginal means
TABLES(Gender*Edu_Level)	Generate statistics for the interaction term. Put your two independent variables here, separated by a * to denote an interaction
COMPARE(Gender)	Tells SPSS to compare the interaction term between genders

4. Making sure that the cursor is at the end of row 2 in the syntax editor click the button, which will run the syntax you have typed. Your results should appear in the Output Viewer below the results you have already generated.

SPSS Output of Simple Main Effects

The table you are interested in is the **Univariate Tests** table:

Univariate Tests

Dependent Variable:Int_Politics

Edu_Level		Sum of Squares	df	Mean Square	F	Sig.
School	Contrast	9.800	1	9.800	.543	.465
	Error	975.200	54	18.059		
College	Contrast	1.250	1	1.250	.069	.793
	Error	975.200	54	18.059		
University	Contrast	186.050	1	186.050	10.302	.002
	Error	975.200	54	18.059		

Each F tests the simple effects of Gender within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

Published with written permission from SPSS Inc, an IBM Company.

This table shows us whether there are statistical differences in mean political interest between gender for each educational level. We can see that there are no statistically significant mean differences between male and females' interest in politics when individuals are educated to school (P = .465) or college level (P = .793). However, when individuals are educated to University level, there are significant differences between males and females' interest in politics (P = .002).

Reporting the results of a two-way ANOVA

You should emphasize the results from the interaction first, before you mention the main effects. In addition, you should report whether your dependent variable was normally distributed for each group and how you measured it (we will provide an example below).

A two-way ANOVA was conducted that examined the effect of gender and education level on interest in politics. Our dependent variable, interest in politics, was normally distributed for the groups formed by the combination of the levels of education level and gender as assessed by the Shapiro-Wilk test. There was homogeneity of variance between groups as assessed by Levene's test for equality of error variances. There was a significant interaction between the effects of gender and education level on interest in politics, F (2, 54) = 4.643, P = .014. Simple main effects analysis showed that there males were significantly more interested in politics than females when educated to University level (P = .002) but there were no differences between gender when educated to school (P = .543) or college level (P = .793).

If you have found this article useful, but want to ensure that you haven't missed anything, or failed the assumptions vital to correctly run a two-way ANOVA, you can learn about the Premium guide for two-way ANOVA here. To learn more about our Premium site in general, which you can gain access to from just $3.99/£2.99/€3.99, Take the Tour.

One-way MANOVA using SPSS

38

Introduction

The one-way MANOVA is used to determine whether there are any differences between independent groups on more than one continuous dependent variable. In this regard, it differs from a one-way ANOVA only in measuring more than one dependent variable at the same time, unlike the one-way ANOVA that only measures one dependent variable.

Assumptions

- One independent variable consists of **two or more categorical independent groups**.
- Two or more dependent variables that are either **interval** or **ratio** (**continuous**) (see our guide on Types of Variable).
- Multivariate Normality
- Equality of variances between the independent groups (**homogeneity of variances**).
- Independence of cases.

Example

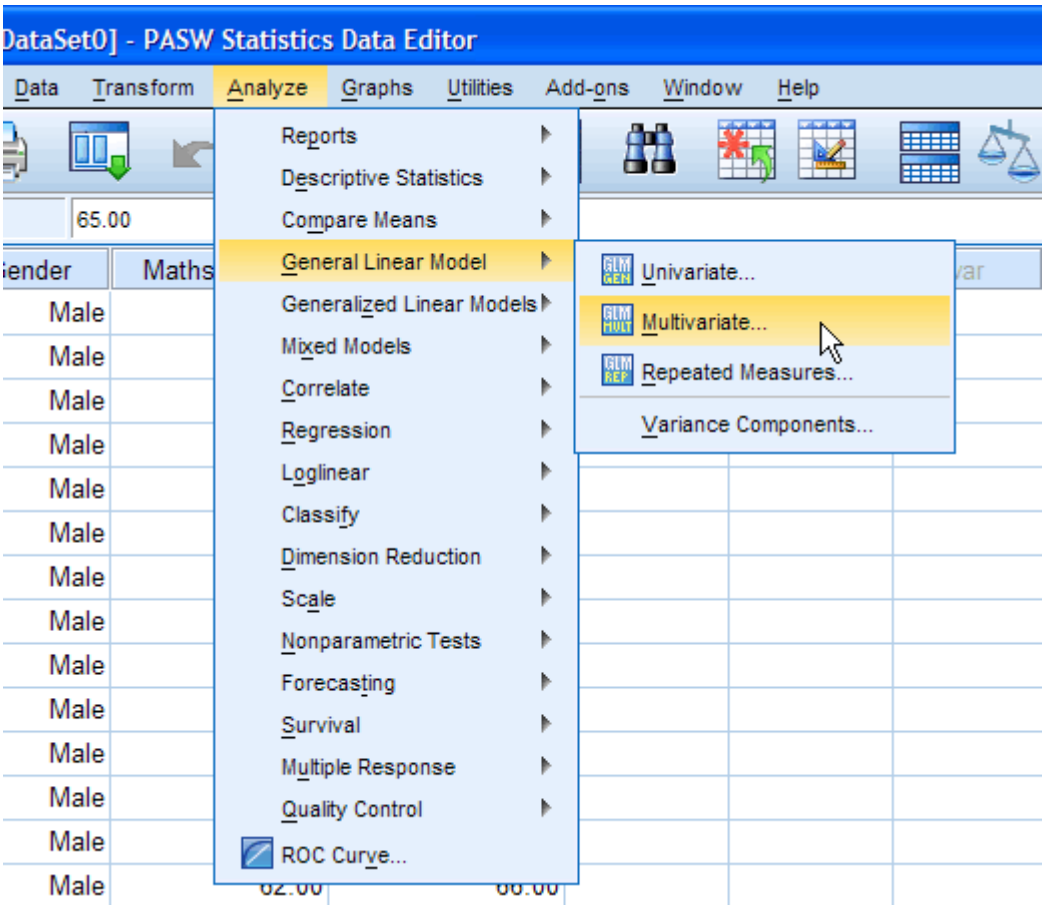
A high school takes its intake from three different primary schools. A teacher was concerned that, even after a few years, there were academic differences between the pupils from the different schools. As such, she randomly selected 20 pupils from School A, 20 pupils from School B and 20 pupils from School C and measured their academic performance as assessed by the marks they received for their end-of-year English and Maths exams.

Setup in SPSS

The independent variable separating boys and girls has been labelled "**Gender**" with "**Male**" and "**Female**" categories. The dependent variables, English and Maths exam mark, were labelled "**English_Score**" and "**Maths_Score**", respectively. To know how to correctly enter your data into SPSS in order to run a one-way MANOVA please read our Entering Data in SPSS tutorial.

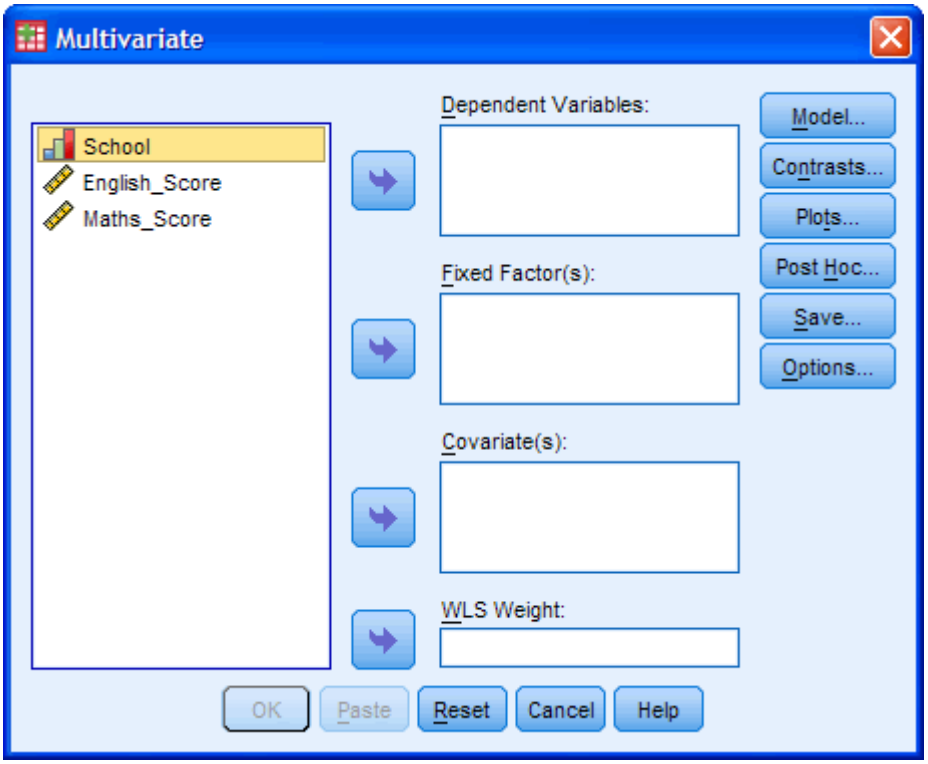
Test Procedure in SPSS

1. Click **Analyze > General Linear Model > Mulivariate...** on the top menu as shown below:



Published with written permission from SPSS Inc, an IBM Company.

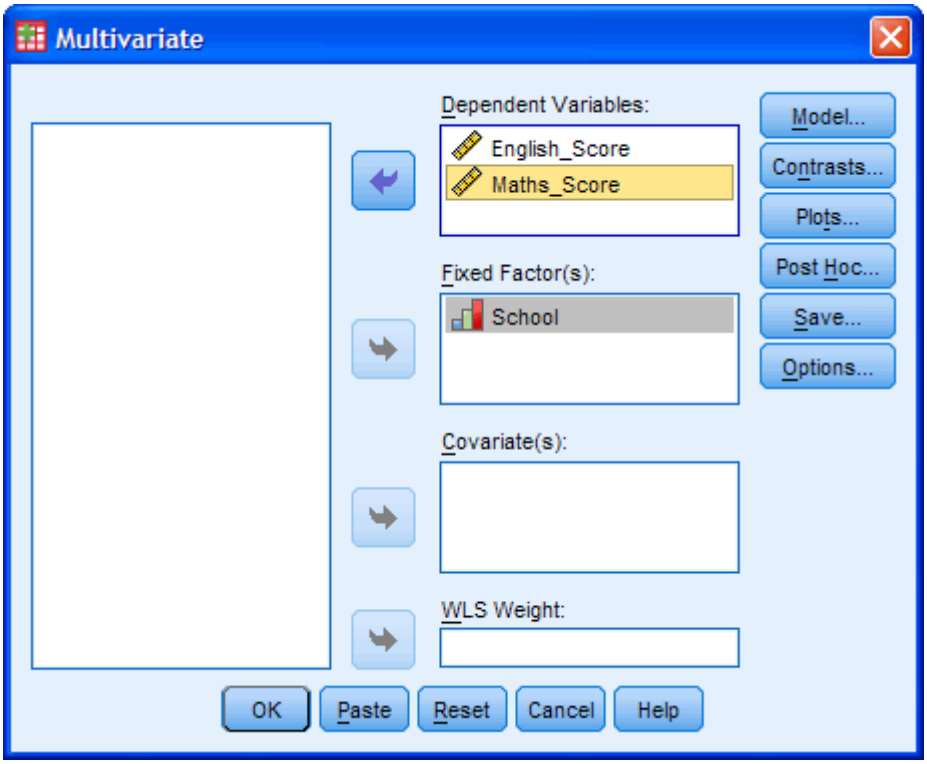
2. You will be presented with the **Multivariate** dialogue box:



Published with written permission from SPSS Inc, an IBM Company.

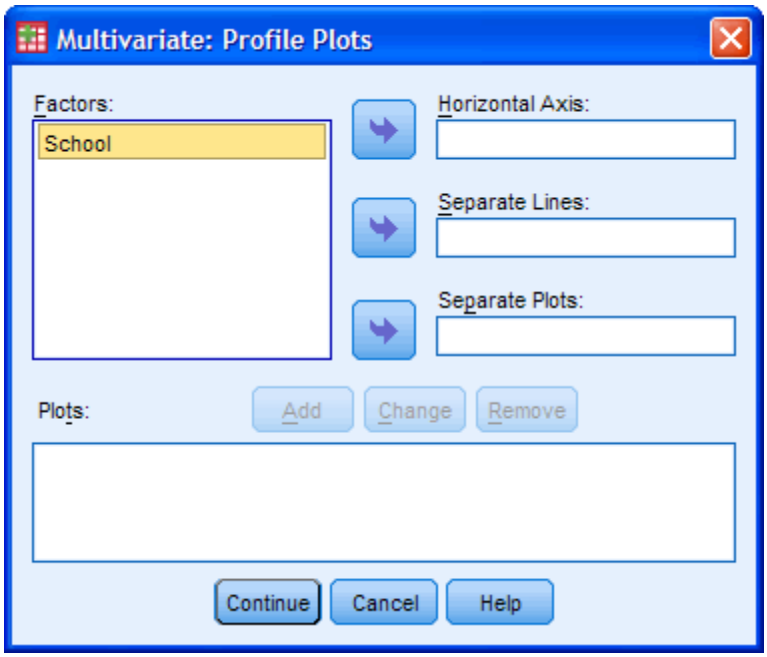
3. Transfer the independent variable, "**School**", into the "*Fixed Factor(s):*" box and transfer the dependent variables, "**English_Score**" and "**Maths_Score**", into the "*Dependent Variables:*" box. You can do this by drag-and-dropping the variables into their respective boxes or by using the button. For older versions of SPSS, you will need to use the latter method. The result is shown below: button. If you are using older versions of SPSS you will need to use the former method. The result is shown below:

   [For this analysis, you will not need to use the "*Covariate(s):*" (used for MANCOVA) or "*WLS Weight:*" boxes.]



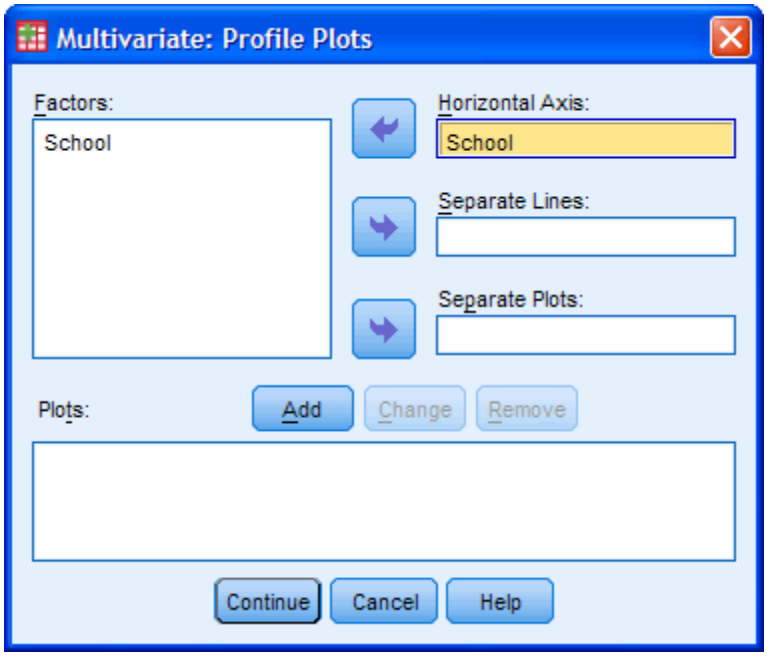
Published with written permission from SPSS Inc, an IBM Company.

4. Click on the Plots... button. You will be presented with the **Multivariate: Profile Plots** dialogue box:



Published with written permission from SPSS Inc, an IBM Company.

5. Transfer the independent variable, "**School**", into the "*Horizontal Axis:*" box, as shown below:



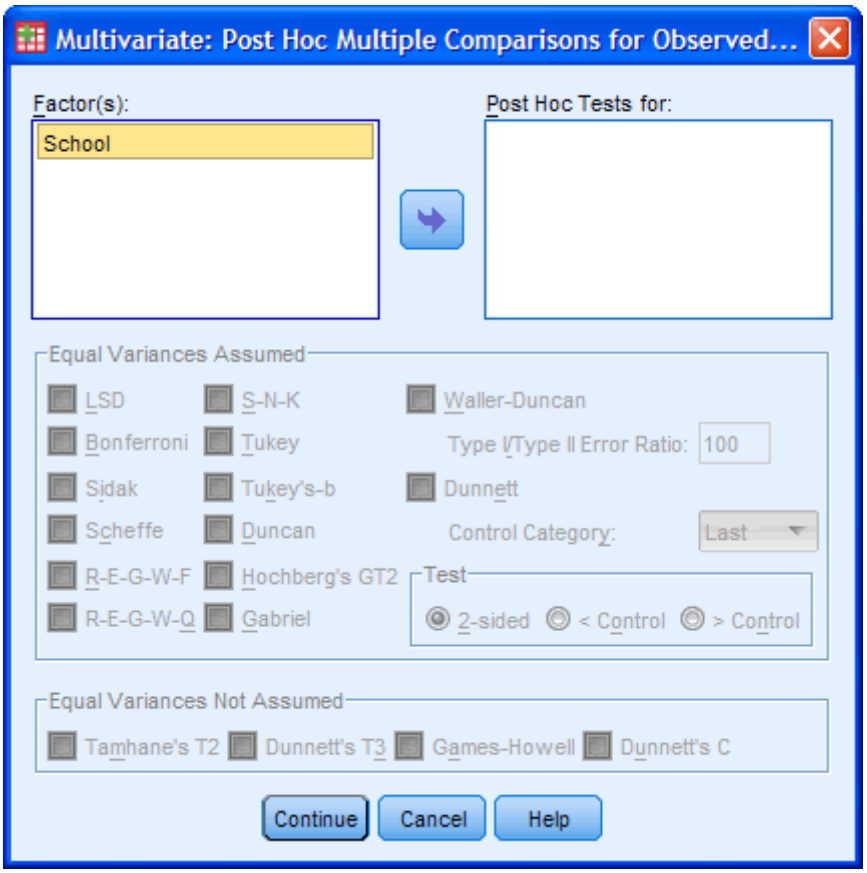
Published with written permission from SPSS Inc, an IBM Company.

6. Click the

button. You will see that "**School**" has been added to the "*Plots:*" box, as shown below:

Published with written permission from SPSS Inc, an IBM Company.

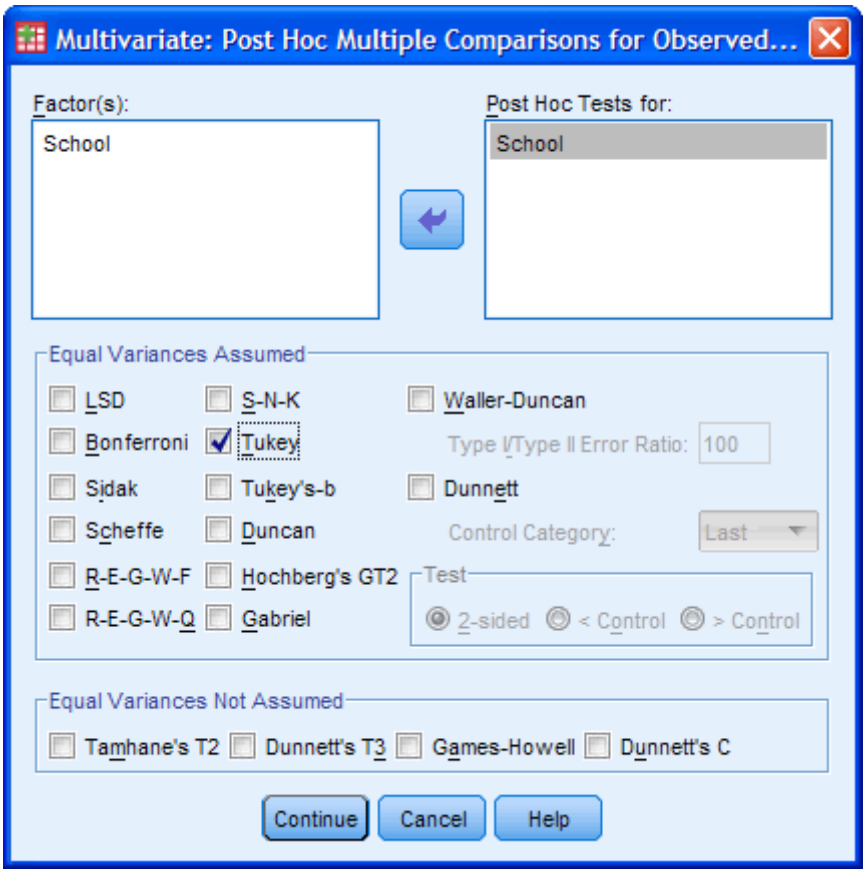
7. Click the Continue button. This will return you to the **Multivariate** dialogue box.
8. Click the Post Hoc... button. You will be presented with the **Multivariate: Post Hoc Multiple Comparisons for Observed...** dialogue box, as shown below:



Published with written permission from SPSS Inc, an IBM Company.

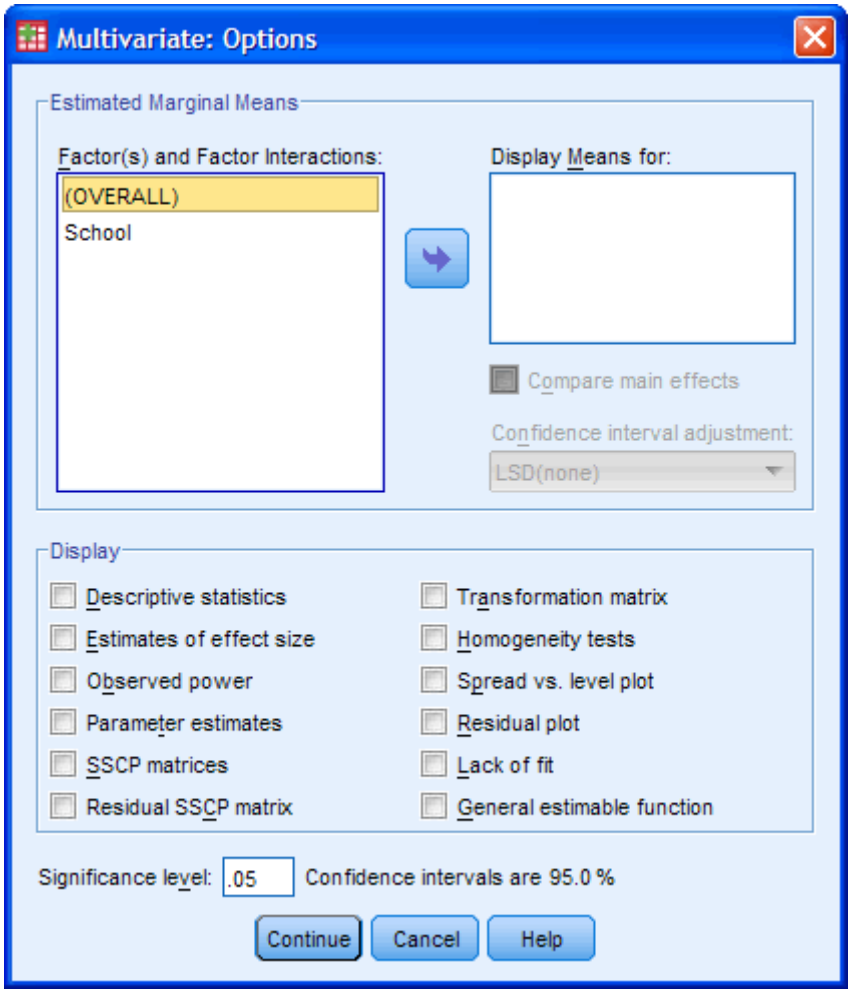
9. Transfer the independent variable, "**School**", into the "*Post Hoc Tests for:*" box and select the "*Tukey*" checkbox in the "Equal Variances Assumed" area, as shown below:

   [You can select other post-hoc tests depending on your data and study design. If your independent variable only has two levels/categories, you do not need to complete this post-hoc section.]



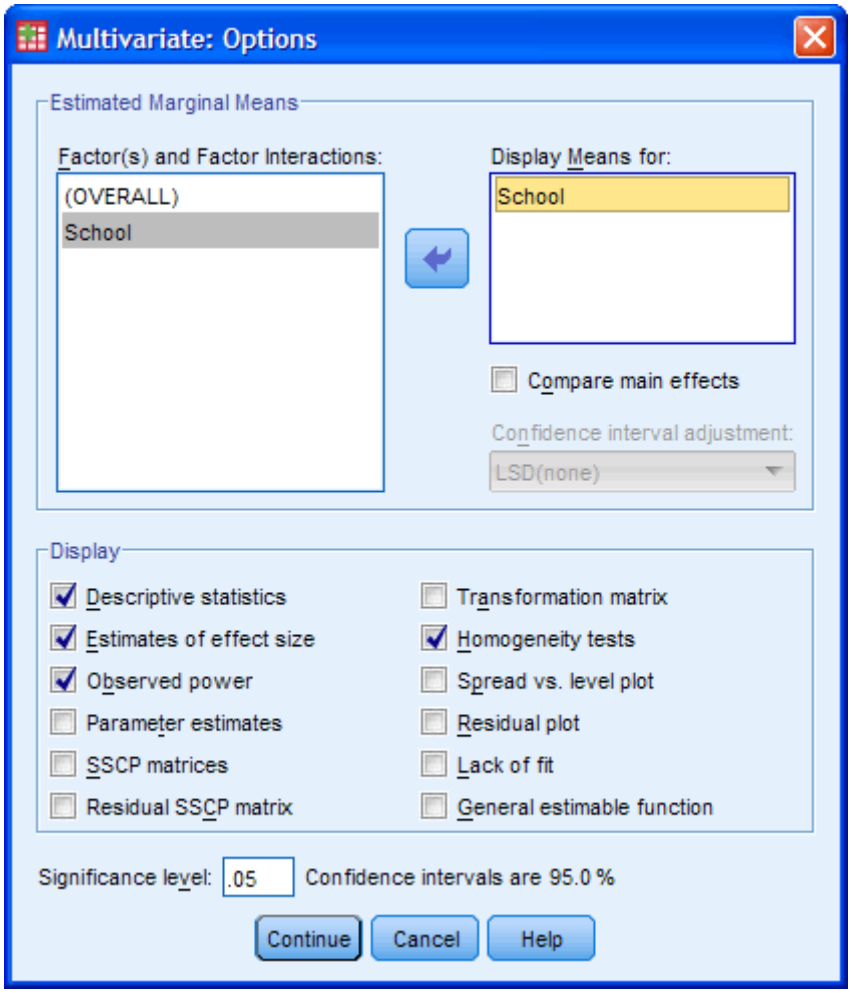
Published with written permission from SPSS Inc, an IBM Company.

10. Click the Continue button. This will return you to the **Multivariate** dialogue box.
11. Click the Options... button. This will present you with the **Multivariate: Options** dialogue box, as shown below:



Published with written permission from SPSS Inc, an IBM Company.

12. Transfer the independent variable, "**School**", from the "*Factor(s) and Factor Interactions:*" box into the "*Display Means for:*" box. Select the "Descriptive statistics", "Estimates of effect size", "Observed power" and "Homogeneity tests" checkboxes in the "Display" area. You will be presented with the following screen:



Published with written permission from SPSS Inc, an IBM Company.

13. Click the Continue button. This will return you to the **Multivariate** dialogue box.
14. Click the OK button to generate the output.

Go to the next page for the SPSS output and explanation of the output.

One-way MANOVA using SPSS (cont...)

33

SPSS Output of One-way MANOVA

SPSS produces many different tables. The first important one is the **Descriptive Statistics** table shown below. This table is very useful as it provides the mean and standard deviation for the two different dependent variables, that have been split by the independent variable. In addition, the table also provides "Total" rows, which allows means and standard deviations for groups only split by the dependent variable to be known.

Descriptive Statistics

	School	Mean	Std. Deviation	N
English_Score	School A	75.6000	8.22960	20
	School B	62.7000	9.10234	20
	School C	61.5500	7.14124	20
	Total	66.6167	10.30401	60
Maths_Score	School A	43.9000	8.46603	20
	School B	40.7500	8.16201	20
	School C	30.7500	7.71789	20
	Total	38.4667	9.78145	60

Published with written permission from SPSS Inc, an IBM Company.

Homogeneity of Covariances

One of the assumptions of the MANOVA is homogeneity of covariances, which is tested for by **Box's Test of Equality of Covariance Matrices**. If the "**Sig.**" value is less than .001 (*P* < .001) then the assumption of homogeneity of covariances was violated. However, we can see from the table below that, in this case, we did not violate this assumption (*P* = .064).

Box's Test of Equality of Covariance Matrices[a]

Box's M	12.556
F	1.986
df1	6
df2	80975.077
Sig.	.064

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + School

Published with written permission from SPSS Inc, an IBM Company.

[For this analysis, you will not need to use the "*Covariate(s):*" (used for MANCOVA) or "*WLS Weight:*" boxes.]

Multivariate Tests

The **Multivariate Tests** table is where we find the actual result of the one-way MANOVA. You need to look at the second Effect, labelled "**School**", and the Wilks' Lambda row (highlighted in red). To determine whether the one-way MANOVA was statistically significant you need to look at the "**Sig.**" column. We can see from the table that we have a "**Sig.**" value of .000, which means P < .0005. Therefore, we can conclude that this school's pupils academic performance was significantly dependent on which prior school they had attended (P < .0005).

Multivariate Tests[d]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Intercept	Pillai's Trace	.989	2435.089[a]	2.000	56.000	.000	.989	4870.177	1.000
	Wilks' Lambda	.011	2435.089[a]	2.000	56.000	.000	.989	4870.177	1.000
	Hotelling's Trace	86.967	2435.089[a]	2.000	56.000	.000	.989	4870.177	1.000
	Roy's Largest Root	86.967	2435.089[a]	2.000	56.000	.000	.989	4870.177	1.000
School	Pillai's Trace	.616	12.681	4.000	114.000	.000	.308	50.724	1.000
	Wilks' Lambda	.450	13.735[a]	4.000	112.000	.000	.329	54.938	1.000
	Hotelling's Trace	1.075	14.782	4.000	110.000	.000	.350	59.128	1.000
	Roy's Largest Root	.915	26.072[c]	2.000	57.000	.000	.478	52.144	1.000

a. Exact statistic

b. Computed using alpha = .05

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Design: Intercept + School

Published with written permission from SPSS Inc, an IBM Company.

Reporting the Result (without follow-up tests)

You could report the result of this test as follows: There was a statistically significant difference between a pupils prior school on their academic performance, F (3, 112) = 13.74, P < .0005; Wilk's λ = 0.450, partial ε^2 = .33.

If you had not achieved a statistically significant result then you would not perform any further follow-up tests. However, as our case shows that we did, we will continue with further tests.

Follow-up Tests - Univariate ANOVAs

To determine how the dependent variables differ for the independent variable we need to look at the **Tests of Between-Subjects Effects** Table but firstly we need to check for homogeneity of variances, which we can do by referring to **Levene's Test of Equality of Error Variances** Table, as shown below:

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
English_Score	.544	2	57	.583
Maths_Score	.188	2	57	.829

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + School

Published with written permission from SPSS Inc, an IBM Company.

We can see from the table above that both the Maths and English scores have homogeneity of variances (P > .05) (the last column in the table). The results of the ANOVAs for our dependent variables are presented in the **Tests of Between-Subjects Effects**, as shown below (highlighted in red):

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	English_Score	2434.233[a]	2	1217.117	18.114	.000	.389	36.228	1.000
	Maths_Score	1885.633[c]	2	942.817	14.295	.000	.334	28.591	.998
Intercept	English_Score	266266.817	1	266266.817	3962.769	.000	.986	3962.769	1.000
	Maths_Score	88781.067	1	88781.067	1346.134	.000	.959	1346.134	1.000
School	English_Score	2434.233	2	1217.117	18.114	.000	.389	36.228	1.000
	Maths_Score	1885.633	2	942.817	14.295	.000	.334	28.591	.998
Error	English_Score	3829.950	57	67.192					
	Maths_Score	3759.300	57	65.953					
Total	English_Score	272531.000	60						
	Maths_Score	94426.000	60						
Corrected Total	English_Score	6264.183	59						
	Maths_Score	5644.933	59						

a. R Squared = .389 (Adjusted R Squared = .367)

b. Computed using alpha = .05

c. R Squared = .334 (Adjusted R Squared = .311)

Published with written permission from SPSS Inc, an IBM Company.

We can see from this table that prior schooling has a statistically significant effect on both English (F (2, 57) = 18.11; $P < .0005$; partial $\varepsilon^2 = .39$) and Maths scores (F (2, 57) = 14.30; $P < .0005$; partial $\varepsilon^2 = .33$). It is important to note that you should make an alpha correction to account for multiple ANOVAs being run, such as a Bonferroni correction. As such, in this case, we accept statistical significance at $P < .025$. We can follow up these significant ANOVAs with Tukey's HSD post-hoc tests, as shown below in the **Multiple Comparisons** Table:

Multiple Comparisons

Tukey HSD

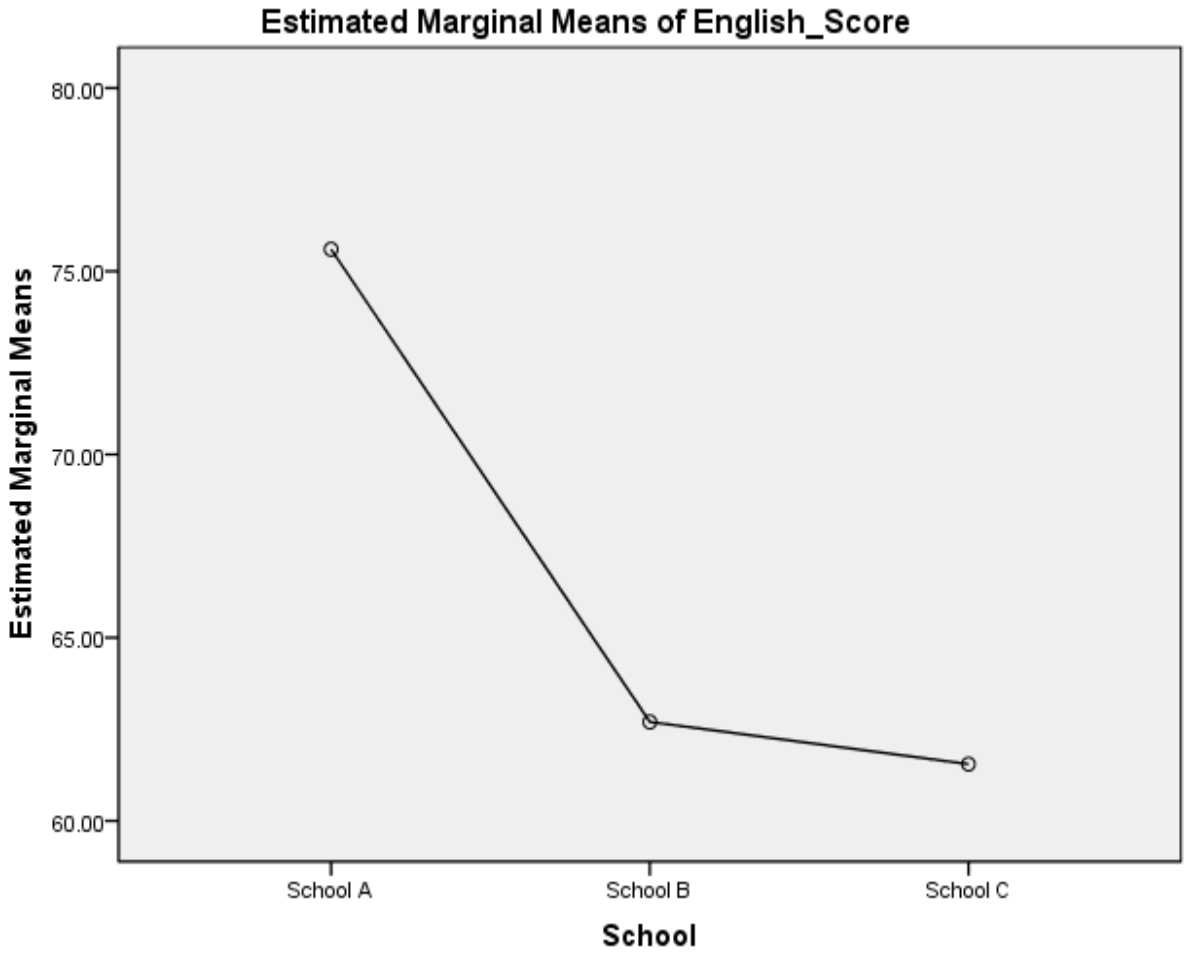
Dependent Variable	(I) School	(J) School	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
English_Score	School A	School B	12.9000*	2.59214	.000	6.6622	19.1378
		School C	14.0500*	2.59214	.000	7.8122	20.2878
	School B	School A	-12.9000*	2.59214	.000	-19.1378	-6.6622
		School C	1.1500	2.59214	.897	-5.0878	7.3878
	School C	School A	-14.0500*	2.59214	.000	-20.2878	-7.8122
		School B	-1.1500	2.59214	.897	-7.3878	5.0878
Maths_Score	School A	School B	3.1500	2.56812	.443	-3.0300	9.3300
		School C	13.1500*	2.56812	.000	6.9700	19.3300
	School B	School A	-3.1500	2.56812	.443	-9.3300	3.0300
		School C	10.0000*	2.56812	.001	3.8200	16.1800
	School C	School A	-13.1500*	2.56812	.000	-19.3300	-6.9700
		School B	-10.0000*	2.56812	.001	-16.1800	-3.8200

Based on observed means.
The error term is Mean Square(Error) = 65.953.

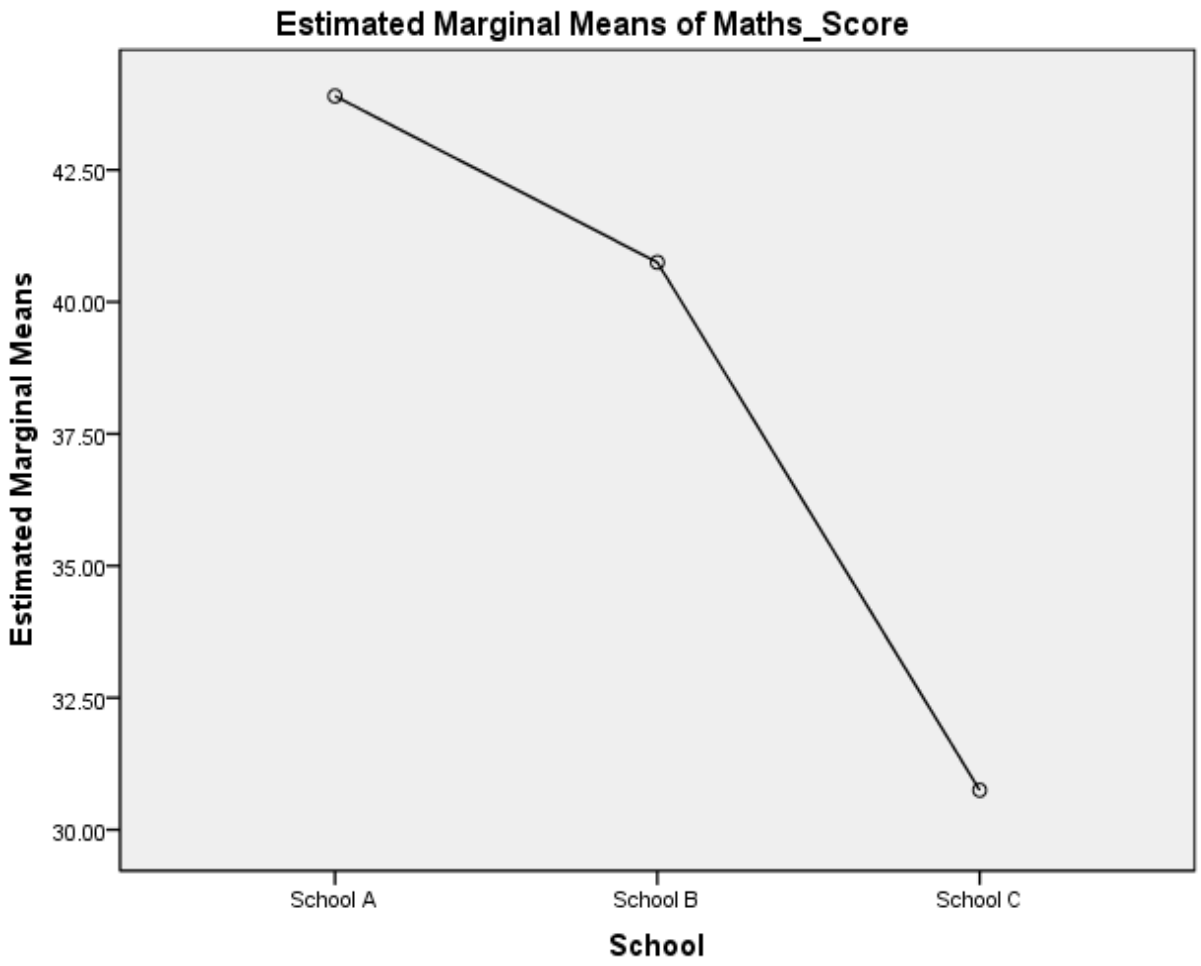
*. The mean difference is significant at the .05 level.

Published with written permission from SPSS Inc, an IBM Company.

The table above shows that for mean scores for English were statistically significantly different between School A and School B ($P < .0005$), and School A and School C ($P < .0005$) but not between School B and School C ($P = .897$). Mean Maths scores were statistically significantly different between School A and School C ($P < .0005$), and School B and School C ($P = .001$) but not between School A and School B ($P = .443$). These differences can be easily visualised by the plots generated by this procedure, as shown below:



Published with written permission from SPSS Inc, an IBM Company.



Published with written permission from SPSS Inc,

Mann-Whitney U Test using SPSS

46

Objectives

The Mann-Whitney U Test is used to compare differences between two independent groups when the dependent variable is either (a) ordinal or (b) interval but not normally distributed. It is the nonparametric alternative to the independent t-test.

Assumptions

- Random samples from populations
- The dependent variable is either **ordinal**, **interval** or **ratio** (see our Types of Variable guide for help)
- Samples do **NOT need to be normally distributed**

Example

A random sample of overweight, male individuals were recruited to a study to investigate whether an exercise or weight loss intervention would be more effective at lowering blood cholesterol levels in overweight men. To this end, researchers randomly split their sample group into two equally-sized, smaller

groups; one group underwent an exercise training programme and the other group undertook a calorie-controlled, weight-loss diet. In order to determine which treatment programme (exercise or diet) was more effective, cholesterol concentrations were compared between the two groups at the end of the treatment programmes.

Setup in SPSS

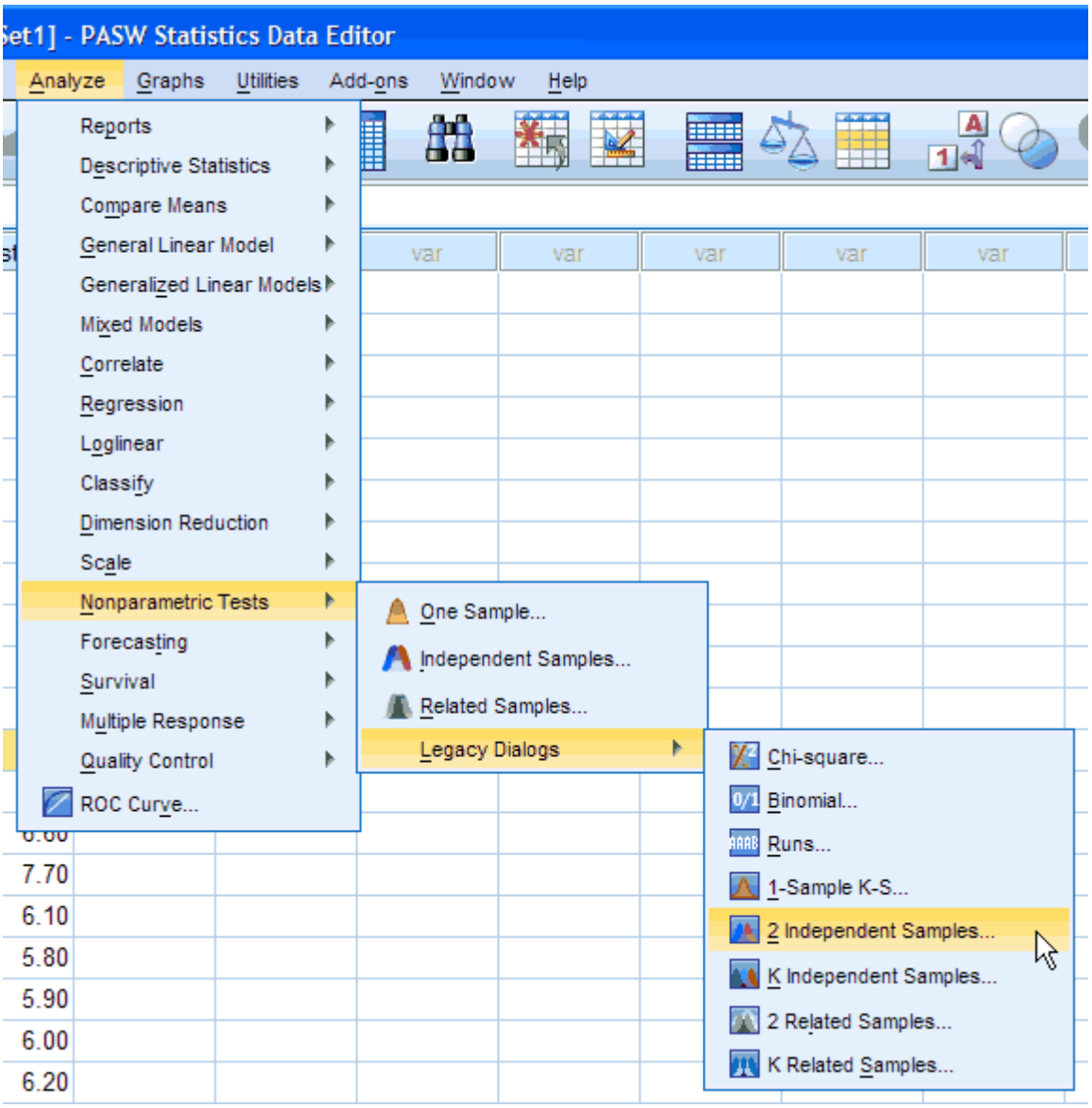
In SPSS we separated the groups for analysis by creating a grouping variable called "**Group**" and gave the exercise group a value of "**1**" and the diet group a value of "**2**". Cholesterol concentrations were entered under the variable name "**Cholesterol**". To correctly setup your data in SPSS to run a Mann-Whitney U Test please read our Entering Data in SPSS tutorial.

Testing Assumptions

Cholesterol concentration is measured on a continuous measurement scale (specifically, a ratio scale) and thus meets the variable requirements for this test. However, having tested the normality of the cholesterol data in the two treatments we have been able to conclude that one of the groups is not normally distributed. Therefore, we cannot use an Independent T-Test and have decided to run a Mann-Whitney U Test. The procedure for checking the normality of your data is provided in our Testing for Normality guide.

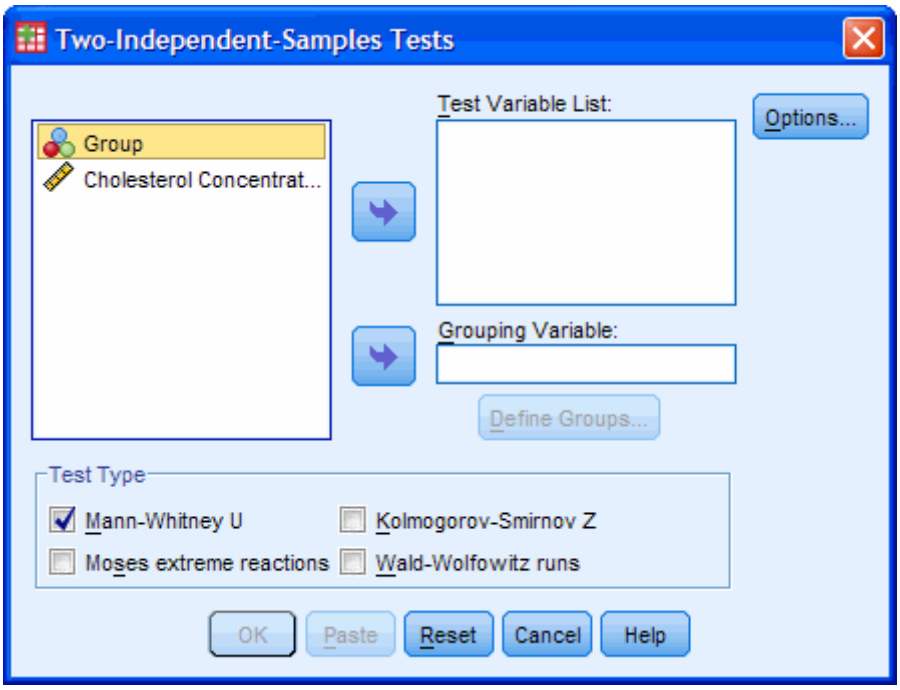
Test Procedure in SPSS

1. Click **Analyze > Nonparametric Tests > Legacy Dialogs > 2 Independent Samples...** on the top menu as shown below:



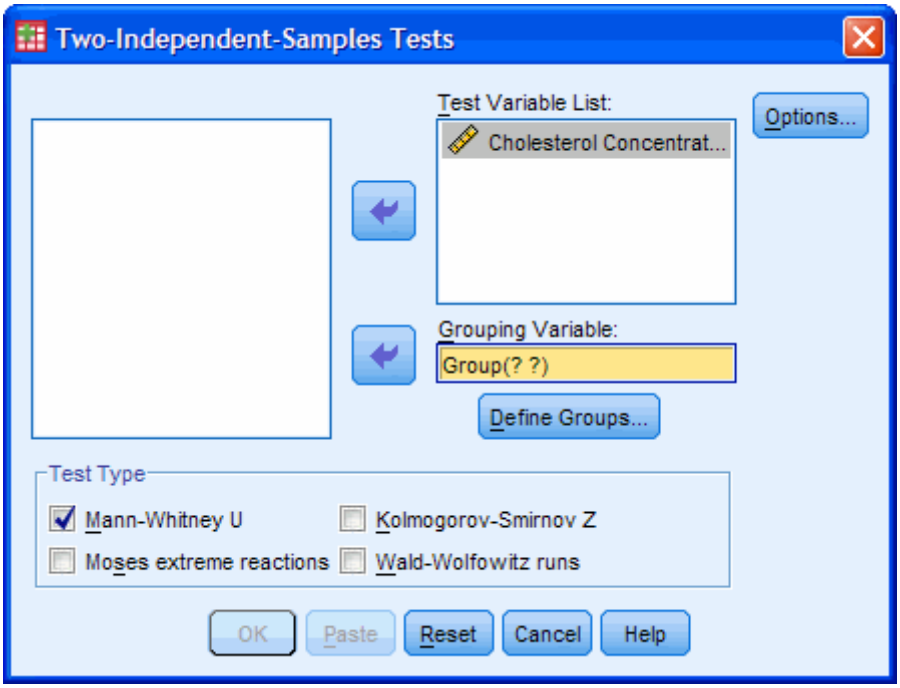
Published with written permission from IBM SPSS Inc.

2. You will be presented with the following screen:



Published with written permission from IBM SPSS Inc.

3. Move the dependent variable **"Cholesterol"** to the *"Test Variable List:"* box and the independent variable **"Group"** to the *"Grouping Variable:"* box by using the button or by dragging-and-dropping the variables into the boxes.



Published with written permission from IBM SPSS Inc.

4. Make sure that the "**Mann-Whitney U**" checkbox is ticked in the "Test Type" area and the *"Grouping Variable:"* box is highlighted in yellow (as

seen above). If not highlighted in yellow, simply click your cursor in the box.

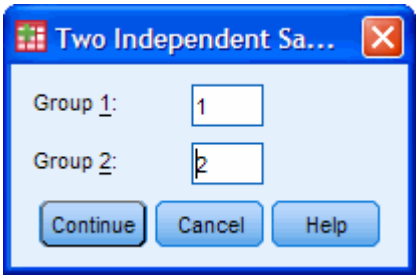
5. Click on the Define Groups... button. The button will not be clickable if you have not highlighted the *"Grouping Variable:"* box as instructed in Point 4. You will be presented with the following screen:

Published with written permission from IBM SPSS Inc.

Click through to the next page for the remaining procedure and how to interpret the output.

Mann-Whitney U Test using SPSS (cont...)

6. Enter the values "**1**" and "**2**" into the "*Group 1:*" and "*Group 2:*" boxes.

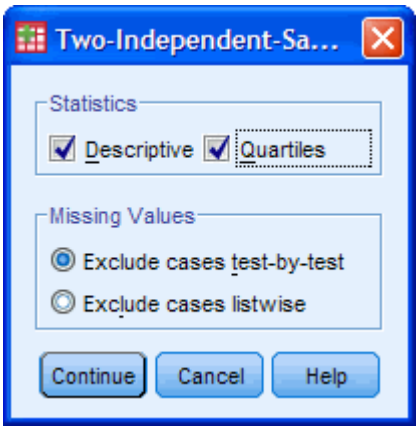


Published with written permission from IBM SPSS Inc.

Click on the Continue button.

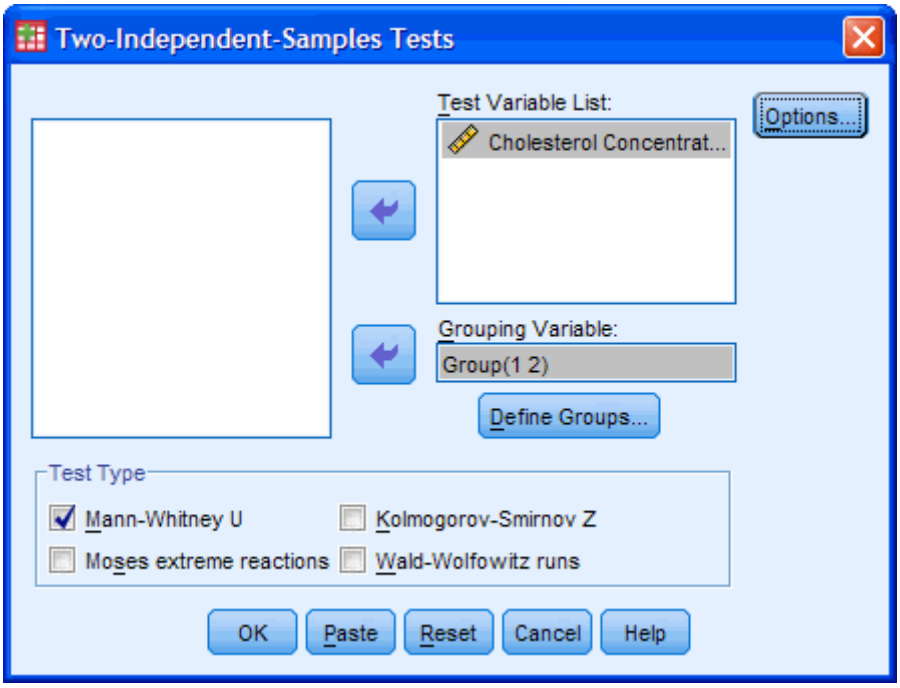
[What are these numbers and why do I have to enter them? The numbers represent the labels we gave to the exercise and diet treatment groups, respectively. The reason SPSS asks you to enter them is because you might have more than two groups and you want to analyse multiple combinations of groups (SPSS does not automatically analyse all combinations of all groups). For example, your categories could be exercise (label = 1), diet (label = 2) and drugs (label = 3) and, if you wanted to compare exercise to drugs you would enter "1" and "3" into the Group boxes.]

7. If you wish to use this procedure to generate some descriptive statistics then click on the Options... button and then tick "**Descriptive**" and "**Quartiles**" within the "Statistics" area. You will be presented with the dialog box below:



Published with written permission from IBM SPSS Inc.

Click the Continue button, which will bring you back to the main dialog box with the *"Grouping Variable:"* box now completed as shown below:



Published with written permission from IBM SPSS Inc.

8. Click the OK button, which will get SPSS to generate the output for the Mann-Whitney U Test.

Ouput and Interpretation

SPSS will produce three tables of output from running a Mann-Whitney U Test as will be described in the next three sections.

Descriptives

The Descriptives output table looks as follows:

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Cholesterol Concentration	40	5.9700	.48368	5.20	7.70	5.6250	5.9000	6.2000
Group	40	1.50	.506	1	2	1.00	1.50	2.00

Published with written permission from SPSS Inc., an IBM Company.

Although we have decided to show you how you can get SPSS to provide descriptive statistics for the Mann-Whitney U Test they are not actually very useful. The reason for this is two-fold. Firstly, in order to compare the groups we need the individual group values not the amalgamated ones. This table does not provide us with this vital information, so we cannot compare any possible differences between the exercise and diet groups. Secondly, we chose the Mann-Whitney U Test because one of the individual groups (exercise group) was not normally distrbuted. However, we have not tested to see if the amalgamation of the two groups results in the larger group being normally distributed. Therefore, we do not know whether to use the mean and standard deviation or the median and interquartile range (IQR). The IQR is the 25th to 75th percentile. This will act as a surrogate to the standard deviation we would otherwise report if the data were normally distributed. We recommend that you ignore this table.

Ranks Table

This is the first table that provides information regarding the ouput of the actual Mann-Whitney U Test. It shows mean rank and sum of ranks for the two groups tested (exercise and diet treatment groups) as shown below:

Mann-Whitney Test

Ranks

	Group	N	Mean Rank	Sum of Ranks
Cholesterol Concentration	Diet	20	25.00	500.00
	Exercise	20	16.00	320.00
	Total	40		

Published with written permission from SPSS Inc., an IBM Company.

The above table is very useful as it indicates which group had the highest cholesterol concentration; namely, the group with the highest mean rank. In this case, the diet group had the highest cholesterol concentrations.

Test Statistics Table

This table shows us the actual significance value of the test (see below). Specifically, the "Test Statistics" table provides the test statistic, *U*, value as well as the asymptotic significance (2-tailed) *p*-value.

Test Statistics[b]

	Cholesterol Concentration
Mann-Whitney U	110.000
Wilcoxon W	320.000
Z	-2.446
Asymp. Sig. (2-tailed)	.014
Exact Sig. [2*(1-tailed Sig.)]	.014[a]

a. Not corrected for ties.

b. Grouping Variable: Group

Published with written permission from SPSS Inc., an IBM Company.

From this data it can be concluded that there is a statistically significant difference between the exercise and diet treatment group's median cholesterol concentration at the end of both treatments (U = 110, p = 0.014). It can be further concluded that the exercise treatment elicited statistically significant lower cholesterol concentrations than the dietary group (p = 0.014).

How to run a Mann-Whitney test using SPSS's new nonparametric procedure is explained in our Premium articles; find out more here.

Wilcoxon Signed Rank Test using SPSS

51

Overview

The Wilcoxon Signed-Rank Test is the nonparametric test equivalent to the dependent t-test. It is used when we wish to compare two sets of scores that come from the same participants. This can occur when we wish to investigate any change in scores from one time point to another or individuals are subjected to more than one condition. As the Wilcoxon Signed-Ranks Test does not assume normality in the data it can be used when this assumption has been violated and the use of the dependent t-test is inappropriate.

Example

A pain researcher is interested in finding methods to reduce lower back pain in individuals without having to use drugs. The researcher thinks that having acupuncture in the lower back might reduce back pain. To investigate this, the researcher recruits 25 participants to their study. At the beginning of the study, the researcher asks the participants to rate their back pain on a scale of 1 to 10 with 10 indicating the greatest level of pain. After 4 weeks of twice weekly acupuncture the participants are asked again to indicate their level of back pain on a scale of 1 to 10 with 10 indicating the greatest level of pain. The researcher

wishes to understand whether the participants' pain levels changed after they had undergone the acupuncture so they run a Wilcoxon Signed-Rank Test.

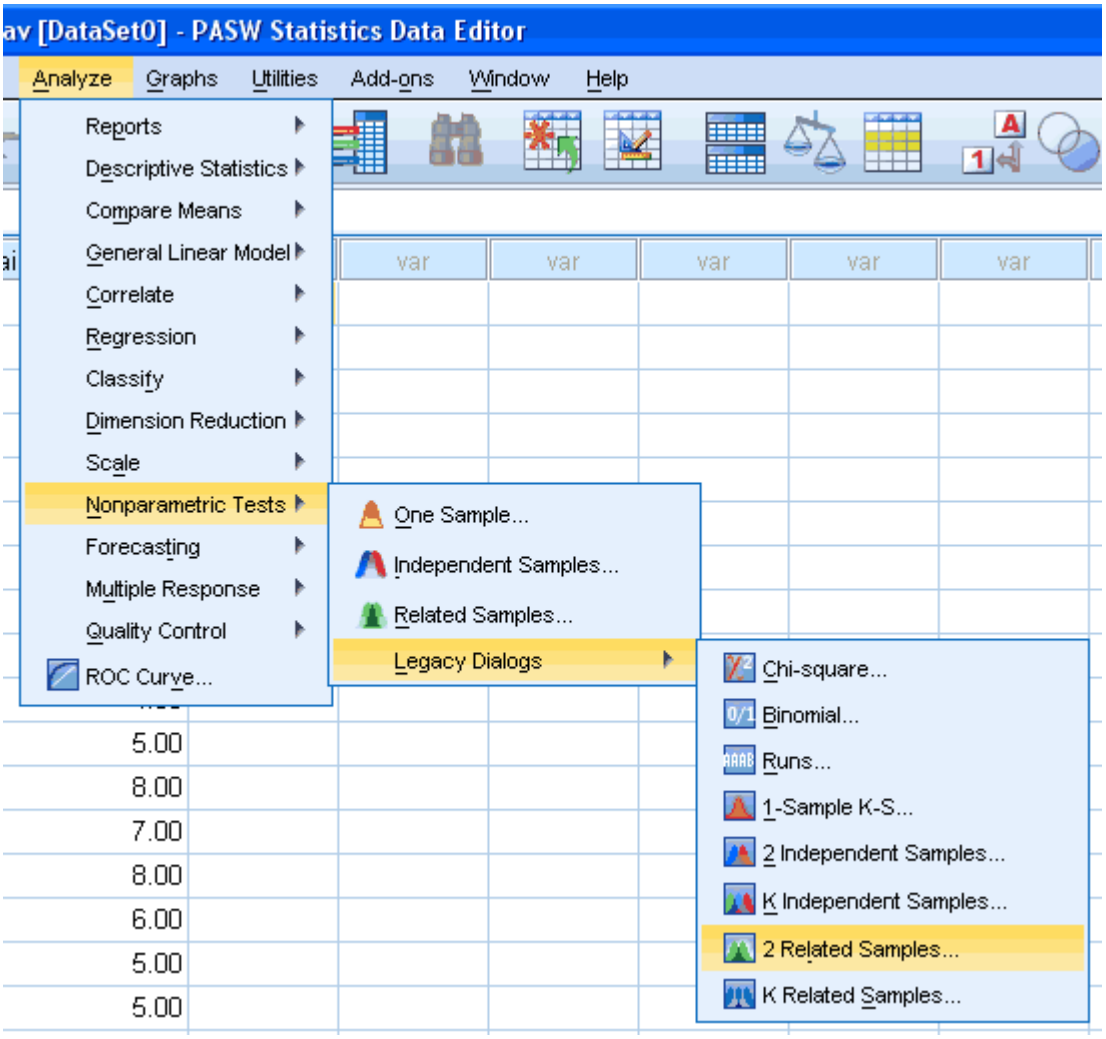
Assumptions

- One dependent variable that is either **ordinal**, **interval** or **ratio** (see our Types of Variable guide).
- One independent variable that consists of **one group** or **two "matched-pairs" groups**.

Test Procedure in SPSS

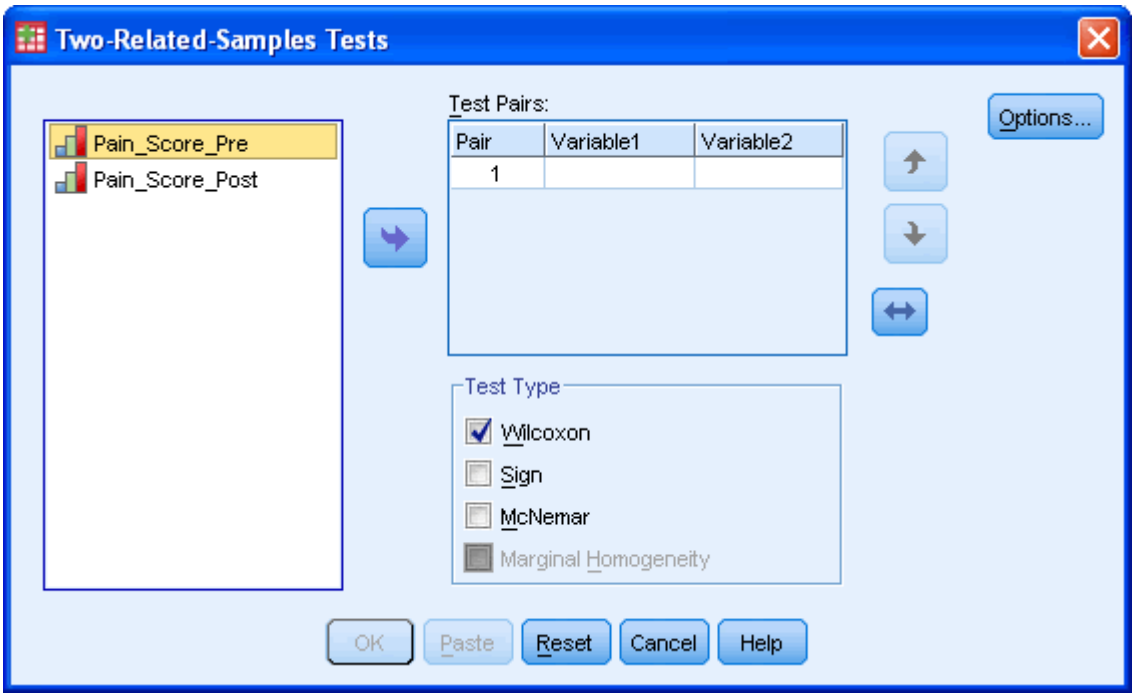
1. Click **Analyze > Nonparametric Tests > Legacy Dialogs > 2 Related Samples...** on the top menu in Version 18.0 of SPSS.

   For older versions of SPSS, click **Analyze > Nonparametric Tests > Related Samples...**.



Published with written permission from SPSS Inc, an IBM Company.

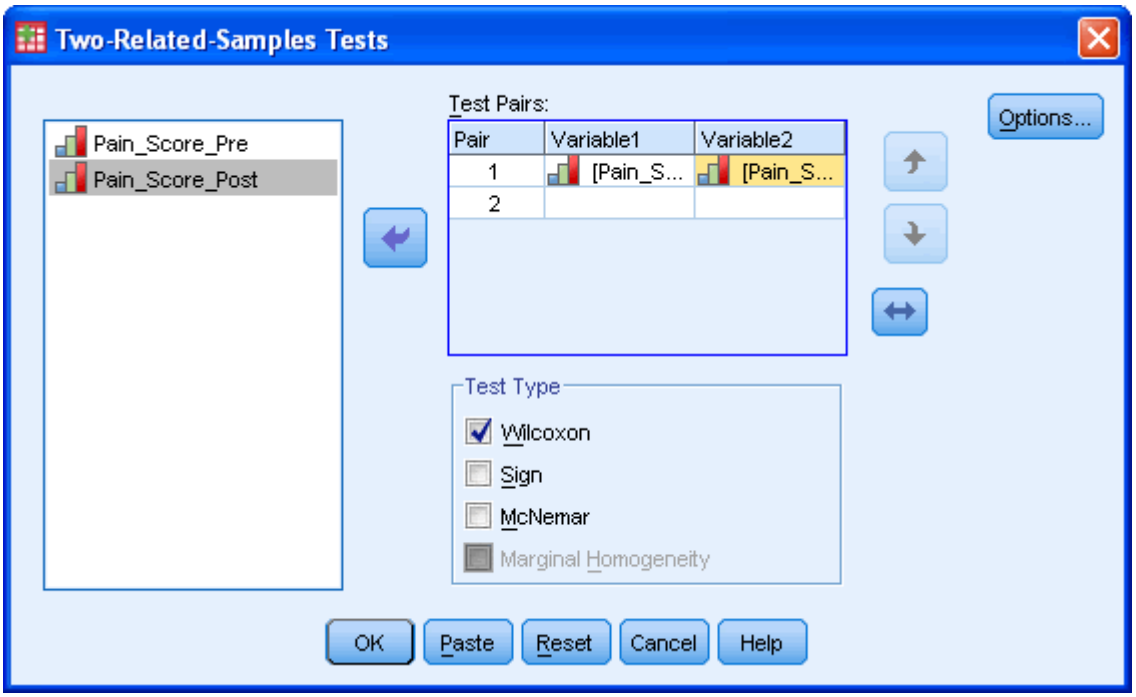
2. You will be presented with the following:



Published with written permission from SPSS Inc, an IBM Company.

3. Transfer the variables you are interested in analyzing into the "Test Pairs:" box. In our example, we need to transfer the variables **Pain_Score_Pre** and **Pain_Score_Post**, which represent the Pain Scores before and after the acupuncture intervention, respectively. There are two ways to do this. You can either (1) highlight both variables (use the cursor and hold down the shift key) and then press the button or (2) you can drag-and-drop each variable into the boxes. Make sure that the "Wilcoxon" checkbox is ticked in the "Test Type" box.

   You will end up with a screen similar to the one below:



Published with written permission from SPSS Inc, an IBM Company.

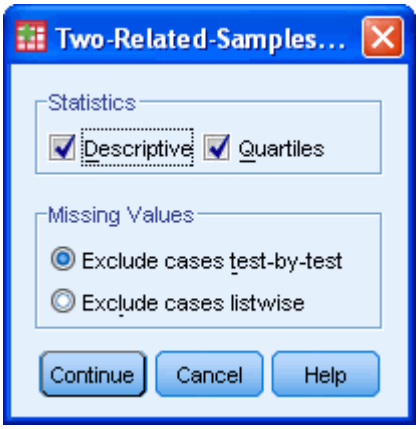
button shifts the pair of variables you have highlighted down one level.

button shifts the pair of variables you have highlighted up one level.

button shifts the order of the variables with a variable pair itself.

4. If you want to generate descriptives or quartiles for your variables then select them by clicking the Options... button and ticking the "Descriptive" and "Quartiles" checkboxes under the "Statistics" box. You can also decide how to deal with missing values.

   You will end up with a screen similar to the one below:



Published with written permission from SPSS Inc, an IBM Company.

Click on the Continue button.

5. Click on the OK button to generate the output.

SPSS Output of the Wilcoxon Signed Rank Test

You will be presented with some tables in the Output Viewer under the title **NPar Tests**.

Descriptives Table

The first table titled **Descriptive Statistics** is where SPSS has generated descriptive and quartile statistics for your variables if you selected these options. If you did not select these options, this table will not appear in your results. You can use the results from this table to describe the Pain Score scores before and after the acupuncture treatment. As you have used a nonparametric test it is most likely that you should use the quartiles information to describe both your groups.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Pain_Score_Pre	25	5.4400	1.78139	2.00	9.00	4.0000	5.0000	6.5000
Pain_Score_Post	25	5.1600	1.57268	2.00	8.00	4.0000	5.0000	6.0000

Published with written permission from SPSS Inc, an IBM Company.

Ranks Table

The **Ranks** table provides some interesting data on the comparison of participant's Before (Pre) and After (Post) Pain Score score. We can see from the table's legend that 11 participants had a higher pre-acupucture treatment Pain Score than after their treatment. However, 4 participants had a higher Pain Score after treatment and 10 participants saw no change in their Pain Score.

Ranks

		N	Mean Rank	Sum of Ranks
Pain_Score_Post - Pain_Score_Pre	Negative Ranks	11[a]	8.00	88.00
	Positive Ranks	4[b]	8.00	32.00
	Ties	10[c]		
	Total	25		

a. Pain_Score_Post < Pain_Score_Pre

b. Pain_Score_Post > Pain_Score_Pre

c. Pain_Score_Post = Pain_Score_Pre

Published with written permission from SPSS Inc, an IBM Company.

Test Statistics Table

By examining the final **Test Statistics** table we can discover whether these changes, due to acupuncture treatment, led overall to a statistically significant difference in Pain Scores. We are looking for the **Asymp. Sig. (2-tailed)** value,

which in this case is 0.071. This is the *P* value for the test. In statistics, the Wilcoxon Signed Ranks Test is denoted by the test statistic T although we can report the Z statistic instead.

Test Statistics[b]

	Pain_Score_Post - Pain_Score_Pre
Z	-1.807[a]
Asymp. Sig. (2-tailed)	.071

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Published with written permission from SPSS Inc, an IBM Company.

We could, therefore, report our results as follows:

A Wilcoxon Signed Ranks Test showed that a 4 week, twice weekly acupuncture treatment course did not elicit a statistically significant change in lower back pain in individuals with existing lower back pain (Z = -1.807, *P* = 0.071). Indeed, median Pain Score rating was 5.0 both pre- and post-treatment.

Kruskal-Wallis H Test using SPSS

44

Overview

The Kruskal-Wallis Test is the nonparametric test equivalent to the one-way ANOVA and an extension of the Mann-Whitney Test to allow the comparison of more than two independent groups. It is used when we wish to compare three or more sets of scores that come from different groups.

Common Uses

As the Kruskal-Wallis Test does not assume normality in the data and is much less sensitive to outliers it can be used when these assumption have been violated and the use of the one-way ANOVA is inappropriate. In addition, if your data is ordinal then you *cannot* use a one-way ANOVA but you can use this test.

Example

A medical researcher has heard anecdotal evidence that certain anti-depressive drugs can have the positive side-effect of lowering neurological pain in those individuals with chronic, neurological back pain when administered in doses lower than those prescribed for depression. The medical researcher would like to investigate this anecdotal evidence with a study. The researcher identifies 3 well-known, anti-depressive drugs which might have this positive side-effect and labels them Drug A, Drug B and Drug C. The researcher then recruits a group of 60 individuals with a similar level of back pain and randomly assigns them to one of three groups - Drug A, Drug B or Drug C treatment groups and prescribes the relevant drug for a 4 week period. At the end of the 4 week period, the researcher asks the participants to rate their back pain on a scale of 1 to 10 with 10 indicating the greatest level of pain. The researcher wishes to compare the

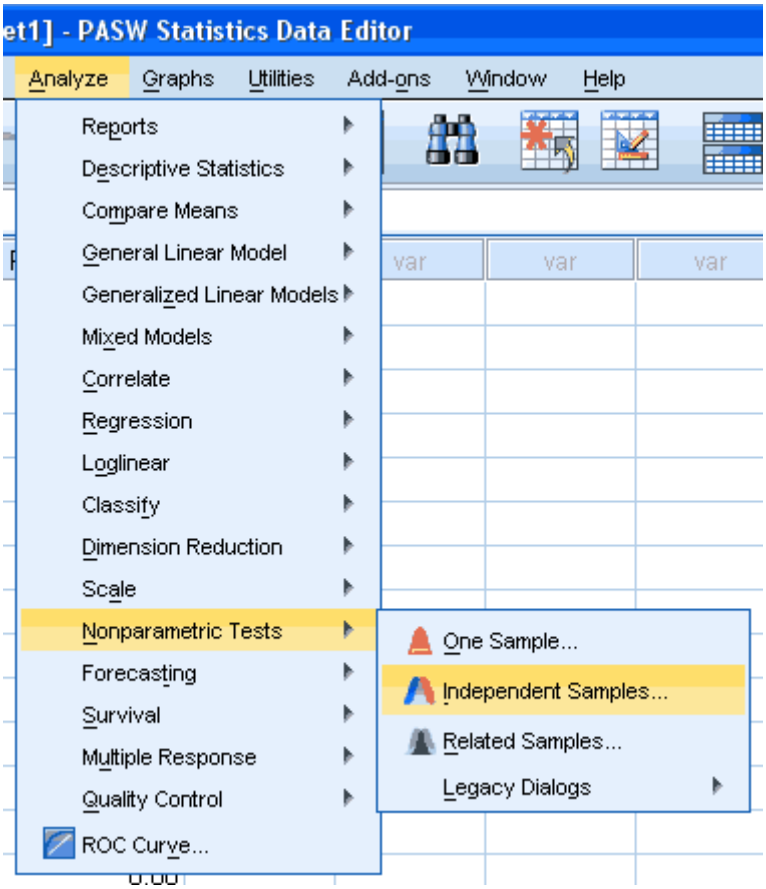
levels of pain experienced by the different groups at the end of the drug treatment period. The researcher runs a Kruskal-Wallis Test to compare this ordinal, dependent measure (Pain Score) between the three drug treatments (independent variable is the type of drug, with more than two groups).

Assumptions

- One dependent variable that is **ordinal**, **interval** or **ratio** (see our guide on types of variable).
- One independent variable that consists of **three or more independent groups**.

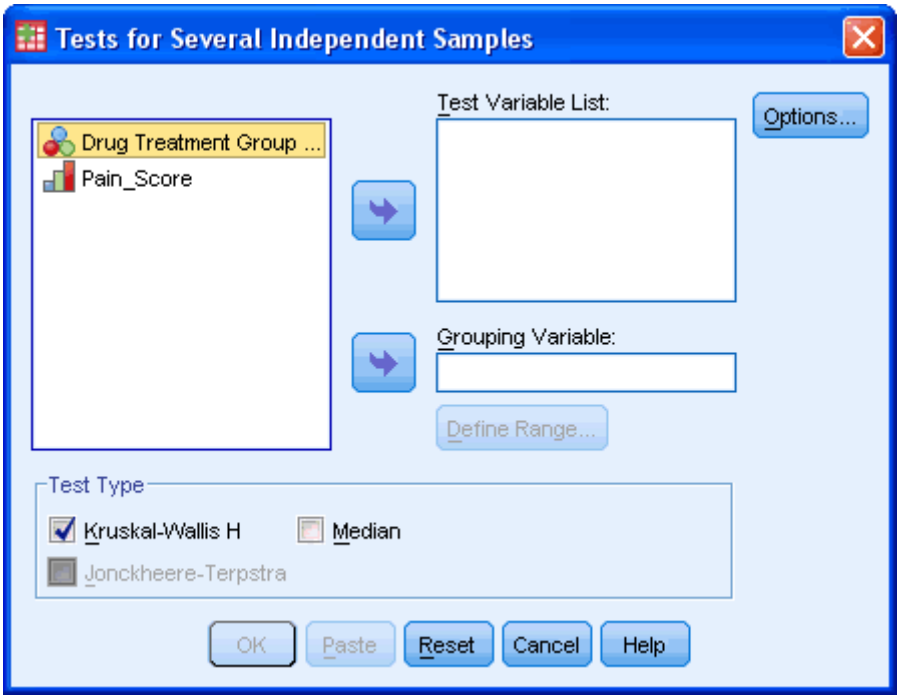
Test Procedure in SPSS

1. Click **Analyze > Nonparametric Tests > Legacy Dialogs > K Independent Samples...** on the top menu as shown below:



Published with written permission from SPSS Inc, an IBM Company.

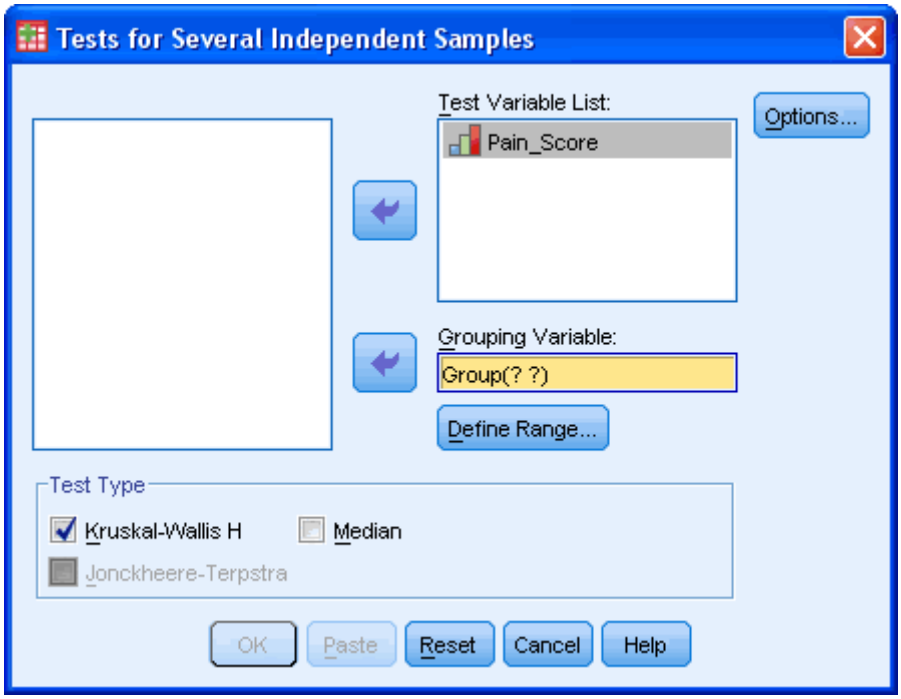
2. You will be presented with the following:



Published with written permission from SPSS Inc, an IBM Company.

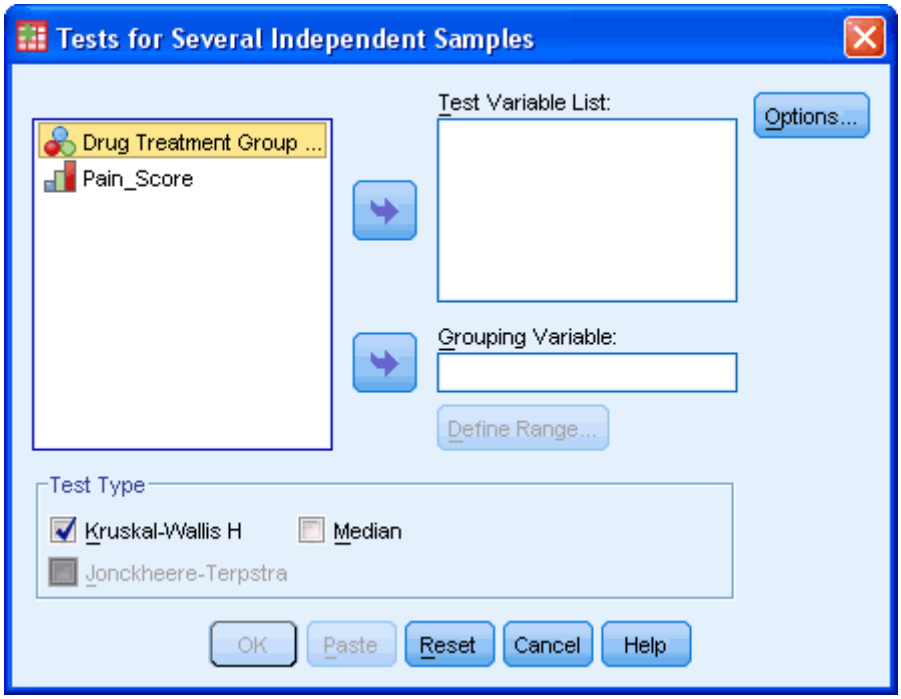
3. Transfer the dependent variable that you are interested in analyzing into the "Test Variable List:" box. In our example, we need to transfer the variable **Pain_Score** into this box. The independent variable needs to be transfered into the "Grouping Variable:" box. There are two ways transfer your variables. You can either highlight drag-and-drop each variable into the respective boxes or you highlight the variable by using the cursor and clicking the button. Make sure that the "Kruskal-Wallis H" checkbox is ticked in the "Test Type" box.

   You will end up with a screen similar to the one below:



Published with written permission from SPSS Inc, an IBM Company.

4. You will be presented with the following:

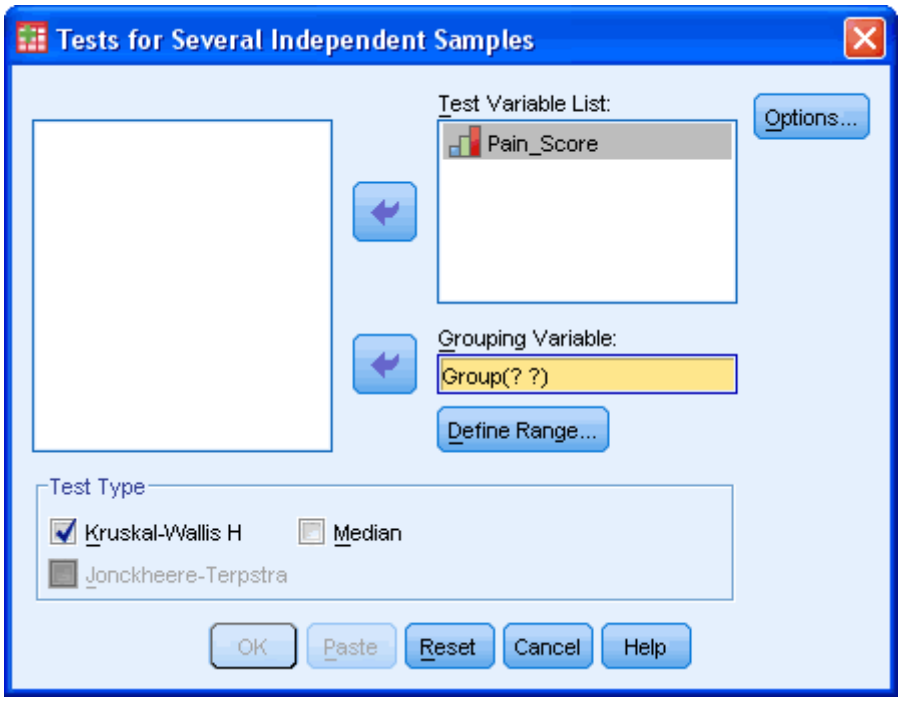


Published with written permission from SPSS Inc, an IBM Company.

5. Transfer the dependent variable that you are interested in analyzing into the "Test Variable List:" box. In our example, we need to transfer the variable **Pain_Score** into this box. The independent variable needs to be transfered into the "Grouping Variable:" box. There are two ways transfer

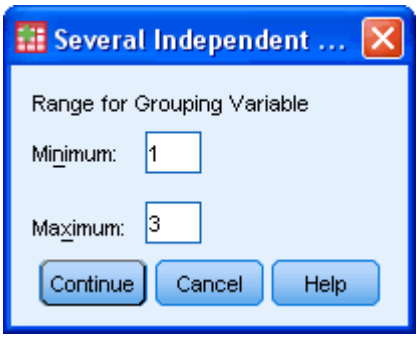
your variables. In this case, this means transfering the **Drug Treatment Group** variable into this box. You can either drag-and-drop each variable into the respective boxes or you can highlight the variable by using the cursor and clicking the button. Make sure that the "Kruskal-Wallis H" checkbox is ticked in the "Test Type" box.

You will end up with a screen similar to the one below:



Published with written permission from SPSS Inc, an IBM Company.

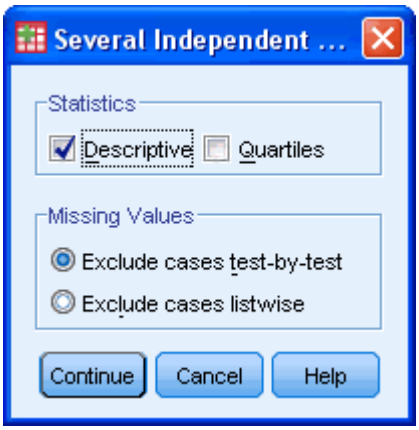
6. Press the Define Groups... button and type "**1**" into the "Minimum" box and "**3**" into the "Maximum" box. This is defining the range of the values for the categories of the independent variables. In this case there are 3 groups/categories called Drug A, Drug B and Drug C. If there had been 4 groups but you did not want to include the first group in the analysis then you would have entered "2" and "4" into the "Minimum" and "Maximum" boxes, respectively (assuming you ordered the groups numerically).



Published with written permission from SPSS Inc, an IBM Company.

Click the Continue button.

7. Click the Options... button. Tick the "Descriptive" checkbox if you want descriptives and/or "Quartiles" if you want quartiles. You will be presented with the following if you select Descriptives:



Published with written permission from SPSS Inc, an IBM Company.

Click the Continue button.

SPSS Output for the Kruskal-Wallis H Test

You will be presented with the following output (Descriptives excluded):

Kruskal-Wallis Test

Ranks

	Drug Treatment Group	N	Mean Rank
Pain_Score	Drug A	20	35.33
	Drug B	20	34.83
	Drug C	20	21.35
	Total	60	

Test Statistics[a,b]

	Pain_Score
Chi-square	8.520
df	2
Asymp. Sig.	.014

a. Kruskal Wallis Test

b. Grouping Variable: Drug Treatment Group

Published with written permission from SPSS Inc, an IBM Company.

The Ranks table shows the mean rank of the Pain_Score for each drug group. The Test Statistics table presents the Chi-square value (Kruskal-Wallis H), the degrees of freedom and the significance level.

Reporting the Output of the Kruskal-Wallis H Test

In our example, we can report that there was a statistically significant difference between the different drug treatments ($H(2) = 8.520$, $P = 0.014$) with a mean rank of 35.33 for Drug A, 34.83 for Drug B and 21.35 for Drug C.

How to run a Kruskal-Wallis test using SPSS's new nonparametric procedure along with *post-hoc* tests to determine where differences lie is explained in our Premium articles; find out more here.

Friedman Test in SPSS

40

Overview

The Friedman Test is the non-parametric alternative to the one-way ANOVA with repeated measures. It is used to test for differences between groups when the dependent variable being measured is ordinal. It can also be used for continuous data that has violated the assumptions necessary to run the one-way ANOVA with repeated measures; for example, marked deviations from normality.

Assumptions

- **One group** that is measured on **three or more different occasions**.
- Group is a random sample from the population.
- One dependent variable that is either **ordinal**, **interval** or **ratio** (see our Types of Variable guide).
- Samples do **NOT need to be normally distributed**.

Example

A researcher wishes to examine whether music has an effect on the perceived psychological effort required to perform an exercise session. To test this, the researcher recruits 12 runners who each run three times on a treadmill for 30 minutes long and conducted at the same running speed. Each subject runs once listening to no music at all, runs once listening to classical music and runs another listening to dance music, in a random order. At the end of each run, subjects are asked to record how hard the running session felt on a scale of 1 to 10, with 1 being easy and 10 extremely hard. A Friedman test is then run to see if there are differences between the music type on perceived effort.

Setting Up Your Data in SPSS

SPSS puts all repeated measures data on the same row in its Data View and, therefore, you will need as many variables as related groups. In our example, we need three variables, which we have labelled "none", "classical" and "dance" to represent the subjects' perceived effort of the run when listening to the different types of music.

	none	classical	dance
1	8.00	8.00	7.00
2	7.00	6.00	6.00
3	6.00	8.00	6.00
4	8.00	9.00	7.00
5	5.00	8.00	5.00
6	9.00	7.00	7.00
7	7.00	7.00	7.00
8	8.00	7.00	7.00
9	8.00	6.00	8.00
10	7.00	6.00	6.00
11	7.00	8.00	6.00
12	9.00	9.00	6.00

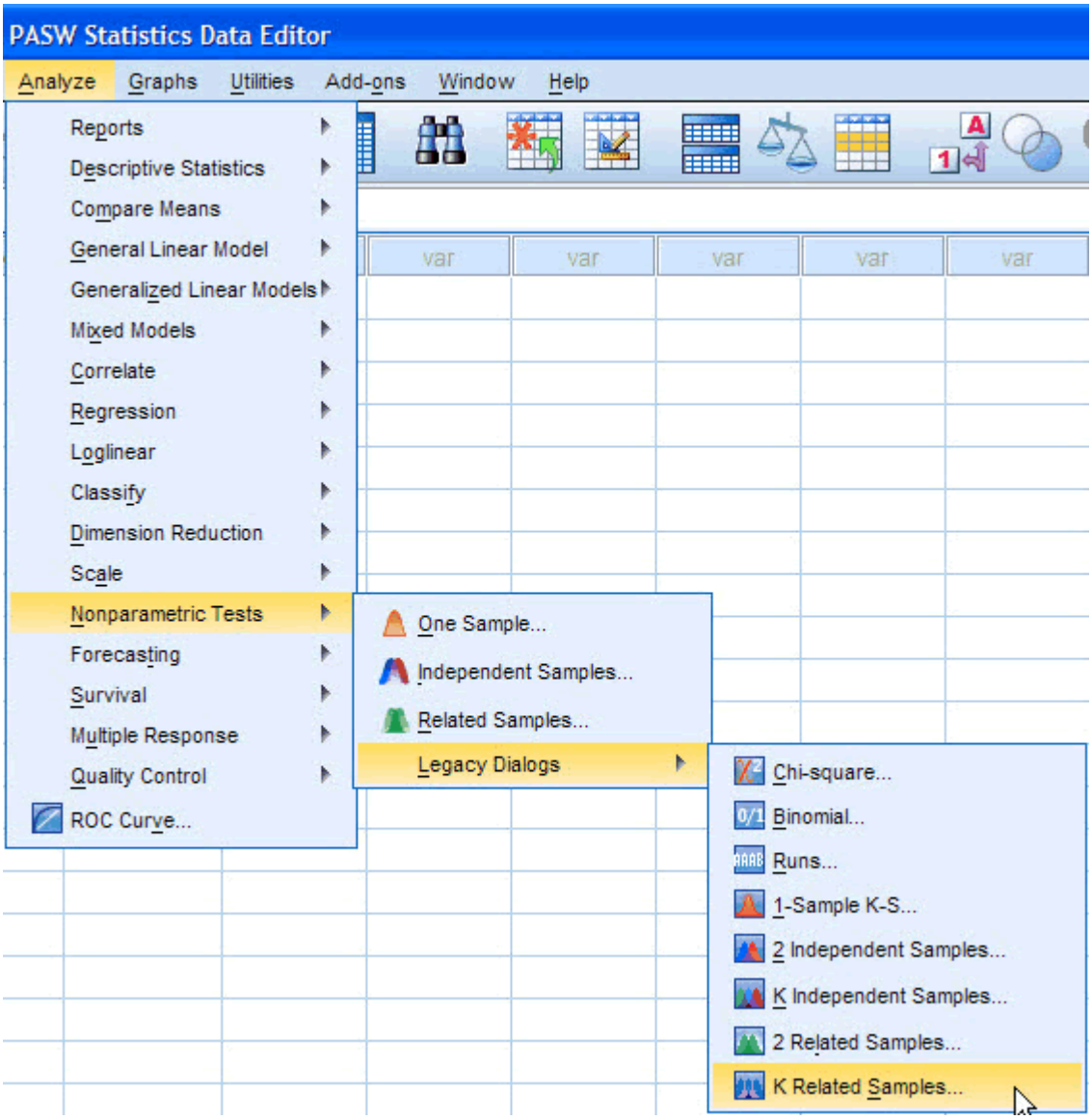
Published with written permission from SPSS Inc, an IBM Company.

Testing of Assumptions

The Friedman Test procedure in SPSS will not test any of the assumptions that are required for this test. In most cases this is because the assumptions are a methodological or study design issue and not what SPSS is designed for. In the case of assessing the types of variable you are using, SPSS will not provide you with any errors if you incorrectly label your variables as nominal.

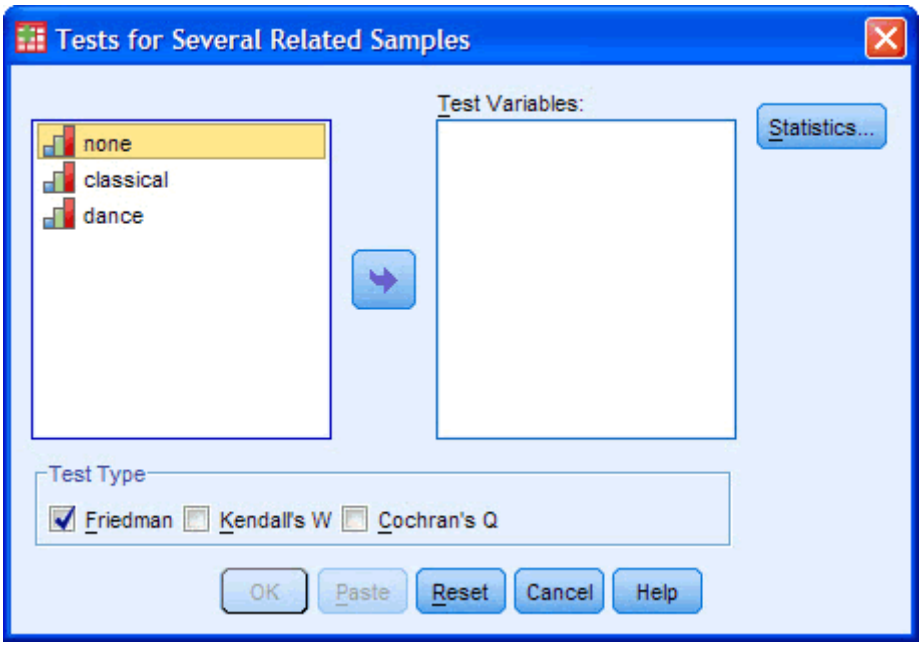
Friedman Test Procedure in SPSS

1. Click **Analyze > Nonparametric Tests > Legacy Dialogs > K Related Samples...** on the top menu as shown below: (ignore Legacy Dialogs if on a previous version to SPSS 18.0)



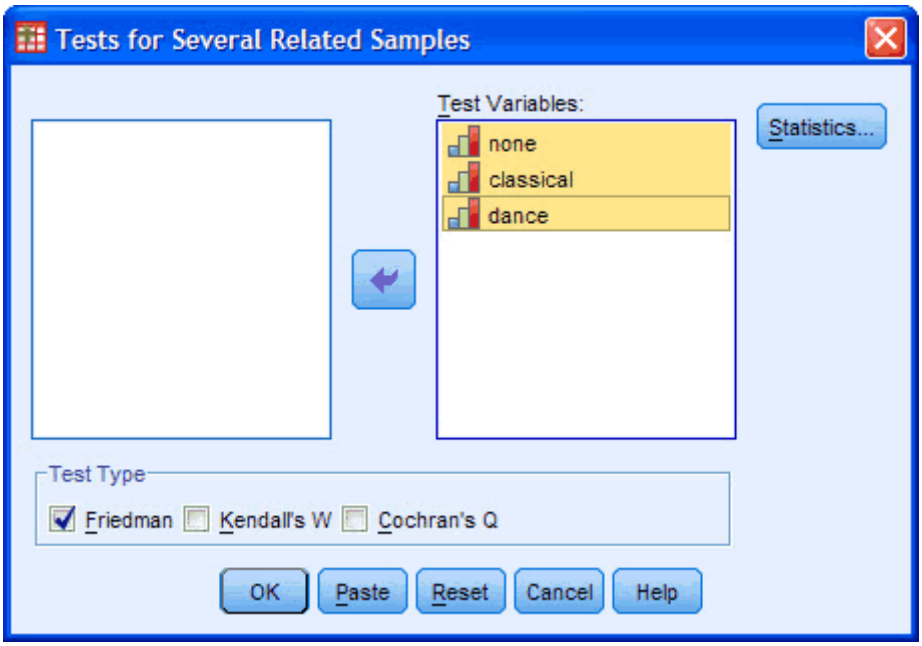
Published with written permission from SPSS Inc, an IBM Company.

2. You will be presented with the following screen:



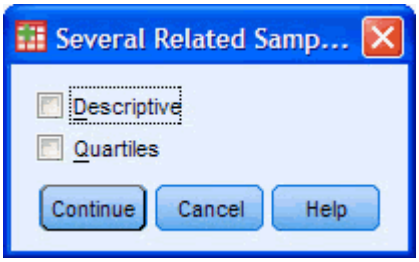
Published with written permission from SPSS Inc, an IBM Company.

3. Move the dependent variables "**none**", "**classical**" and "**dance**" to the "Test Variables:" box by using the button or by dragging-and-dropping the variables into the box. You will end up with the following:



Published with written permission from SPSS Inc, an IBM Company.

4. Make sure that "**Friedman**" is selected in the "*Test Type* " option area.
5. Click the Statistics... button. You will be presented with the following screen:



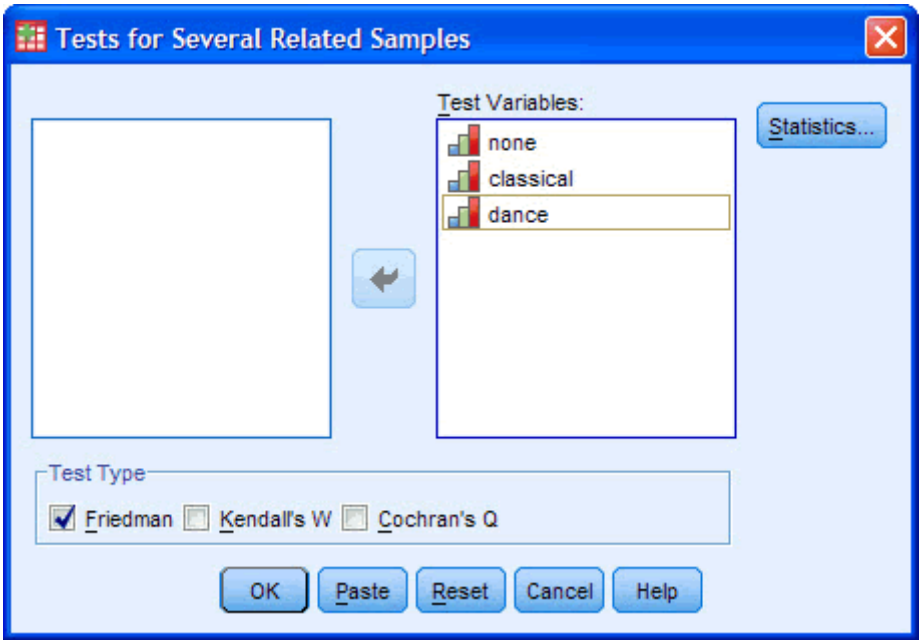
Published with written permission from SPSS Inc, an IBM Company.

6. Tick the "**Quartiles**" option: [It is most likely that you will only want to include the "Quartiles" option as your data is probably unsuitable for "Descriptives", hence why you are running a non-parametric test. However, SPSS includes this option anyway.]

Published with written permission from SPSS Inc, an IBM Company.

7. Click the

button. This will return you back to the main dialogue box:



Published with written permission from SPSS Inc, an IBM Company.

8. Click the

button to run the Friedman Test.

SPSS Output for the Friedman Test

SPSS will generate two or three tables depending on whether you selected to have descriptives and/or quartiles generated in addition to running the Friedman Test.

Descriptive Statistics Table

The following table will be produced if you selected the Quartiles option:

Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
none	12	7.0000	7.5000	8.0000
classical	12	6.2500	7.5000	8.0000
dance	12	6.0000	6.5000	7.0000

Published with written permission from SPSS Inc, an IBM Company.

This is a very useful table as it can be used to present descriptive statistics in your results section for each of the time points or conditions (depending on your study design) for your dependent variable. This usefulness will be presented later on in this guide, in the "Reporting the Output" section.

Ranks Table

The Ranks table shows the mean rank for each of the related groups, as shown below:

Ranks

	Mean Rank
none	2.38
classical	2.17
dance	1.46

Published with written permission from SPSS Inc, an IBM Company.

The Friedman Test compares the mean ranks between the related groups and indicates how the groups differed and it is included for this reason. However, you are not very likely to actually report these values in your results section but most likely will report the median value for each related group.

Test Statistics Table

This is the table which informs you of the actual result of the Friedman Test and whether there was an overall statistically significant difference between the mean ranks of your related groups. For the example used in this guide, the table looks as follows:

Test Statistics[a]

N	12
Chi-square	7.600
df	2
Asymp. Sig.	.022

a. Friedman Test

Published with written permission from SPSS Inc, an IBM Company.

The above table provides the test statistic (χ^2) value (**Chi-square**), degrees of freedom (**df**) and the significance level (**Asymp. Sig.**), which is all we need to report the result of the Friedman Test. We can see, from our example, that there is an overall statistically significant difference between the mean ranks of the related groups. It is important to note that the Friedman Test is an omnibus test like its parametric alternative - that is, it tells you whether there are overall differences but does not pinpoint which groups in particular differ from each other. To do this you need to run *post-hoc* tests, which will be discussed after the next section.

Reporting the Output of the Friedman Test (without *post-hoc* tests)

You can report the Friedman Test result as follows: There was a statistically significant difference in perceived effort depending on which type of music was listened to whilst running, $\chi^2(2) = 7.600$, $P = 0.022$.

You could also include the median values for each of the related groups. However, at this stage, you only know that there are differences *somewhere* between the related groups but you do not know *exactly* where those differences lie. Remember though, that if your Friedman Test result was not statistically significant then you should not run *post-hoc* tests.

Post-hoc Tests

To examine where the differences actually occur, you need to run separate Wilcoxon Signed-Rank Tests on the different combinations of related groups. So, in this example, you would compare the following combinations:

1. None to Classical
2. None to Dance
3. Classical to Dance

You need to use a Bonferroni adjustment on the results you get from the Wilcoxon tests as you are making multiple comparisons, which makes it more likely that you will declare a result significant when you should not (a Type I error). Luckily, the Bonferroni adjustment is very easy to calculate; simply take the significance level you were initially using (in this case 0.05) and divide it by the number of tests you are running. So in this example, we have a new significance level of 0.05/3 = 0.017. This means that if the *P* value is larger than 0.017 then we do not have a statistically significant result.

Running these tests (see how with our Wilcoxon Signed-Rank Test guide) on the results from this example then you get the following result:

Test Statistics[b]

	classical - none	dance - none	dance - classical
Z	-.061[a]	-2.636[a]	-1.811[a]
Asymp. Sig. (2-tailed)	.952	.008	.070

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Published with written permission from SPSS Inc, an IBM Company.

This table shows the output of the Wilcoxon Signed-Rank Test on each of our combinations. It is important to note that the significance values have not been adjusted in SPSS to compensate for multiple comparisons - you must manually compare the significance values produced by SPSS to the Bonferroni-adjusted significance level you have calculated. We can see that at the $P < 0.017$ significance level only perceived effort between no music and dance (dance-none, $P = 0.008$) was statistically significantly different.

Reporting the Output of the Friedman Test (with *post-hoc* tests)

You can report the Friedman Test with *post-hoc* tests results as follows: There was a statistically significant difference in perceived effort depending on which type of music was listened to whilst running, $\chi^2(2) = 7.600$, $P = 0.022$. *Post-hoc* analysis with Wilcoxon Signed-Rank Tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $P < 0.017$. Median (IQR) perceived effort levels for the no-music, classical and dance music running trial were 7.5 (7 to 8), 7.5 (6.25 to 8) and 6.5 (6 to 7), respectively. There were no significant differences between the no-music and classical music running trials ($Z = -0.061$, $P = 0.952$) or between the classical and dance music running trials ($Z = -1.811$, $P = 0.070$) despite an overall reduction in perceived effort in the dance vs. classical running trials. However, there was a statistically significant reduction in perceived effort in the dance music vs. no music trial ($Z = -2.636$, $P = 0.008$).

One-Sample Tests

Chi-Square Goodness-of-Fit Test in SPSS

15

Objective

The Chi-Square Goodness-of-Fit test is used to determine whether a sample of data came from a population with a specific distribution.

Assumptions

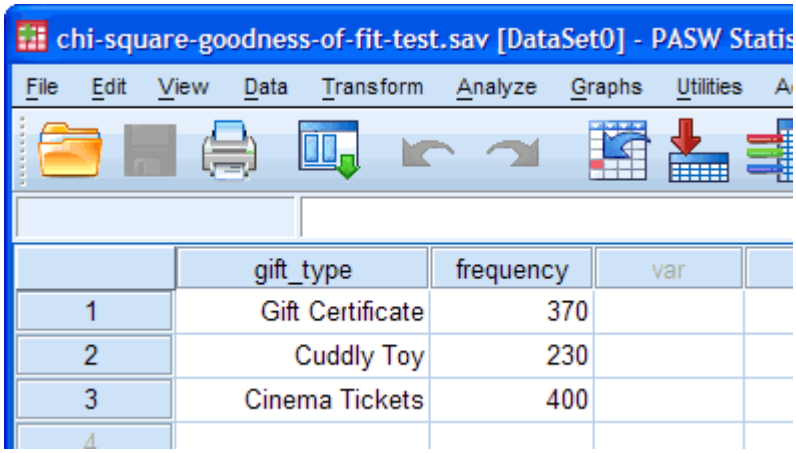
- One categorical variable, with two or more categories (see our guide on Types of Variable)
- A hypothesized proportion (equal or unequal)
- No more than 20% of expected frequencies have counts less than 5.

Example

A website owner, Christopher, wants to offer a free gift to people that purchase a subscription to his website. New subscribers can choose one of three gifts of equal value: an Amazon Gift Voucher, a cuddly toy, or free cinema tickets. After 1000 people have signed up, Christopher wants to review the figures to see if the three gifts offered were equally popular.

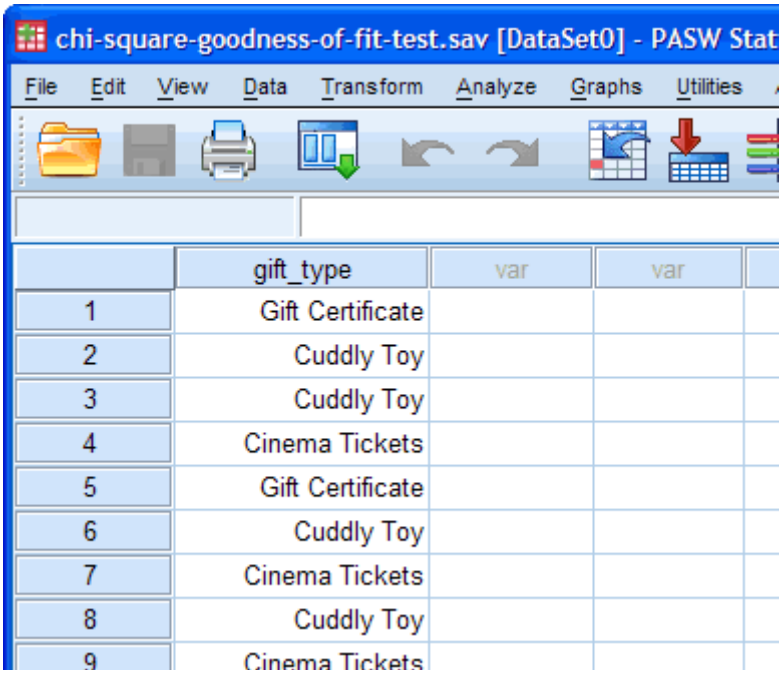
Set-up in SPSS

There are two methods of entering data into SPSS in order to run a Chi-Square Goodness-of-Fit test in SPSS. Common to both methods is a column in the SPSS data file for the categorical variable, which in this example, we shall name **gift_type**. We have assigned codes of "**1**" for the Amazon Gift Certificate and labelled it "**Gift Certificate**", "**2**" for the cuddly toy and labelled it "**Cuddly Toy**", and "**3**" for the free cinema tickets and labelled it "**Cinema Tickets**" (help on how to enter data can be found in our Entering Data in SPSS guide and how to code variables can be found in our Working with Variables guide). If the frequency data has already been summated for the various categories then we need to create a second column that contains the respective frequency counts; we have called this variable **frequency**. This type of data entry is shown below:



Published with written permission from SPSS Inc, an IBM Company.

Alternatively, you may have the data in raw form, i.e. you have not summated the frequencies. In this case, you do not need a second column as SPSS can calculate the frequencies of occurrence of each category for you. This would mean that, in this example, there are 1000 rows of data, of which the beginning of said data is shown below:

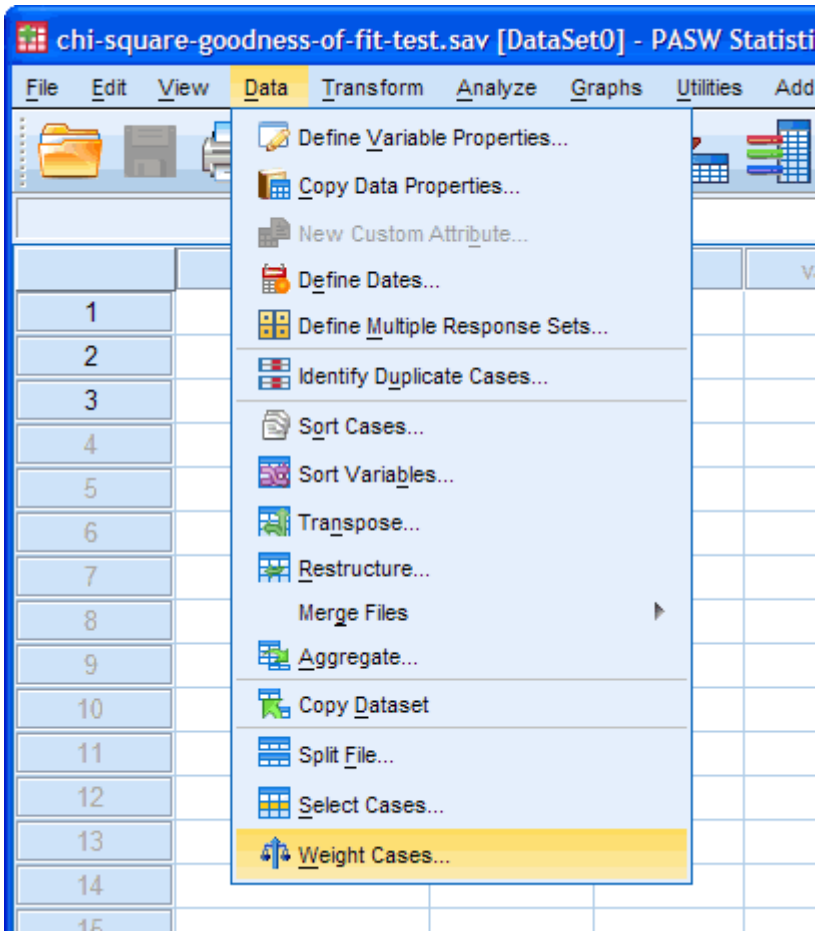


Published with written permission from SPSS Inc, an IBM Company.

Test Procedure in SPSS

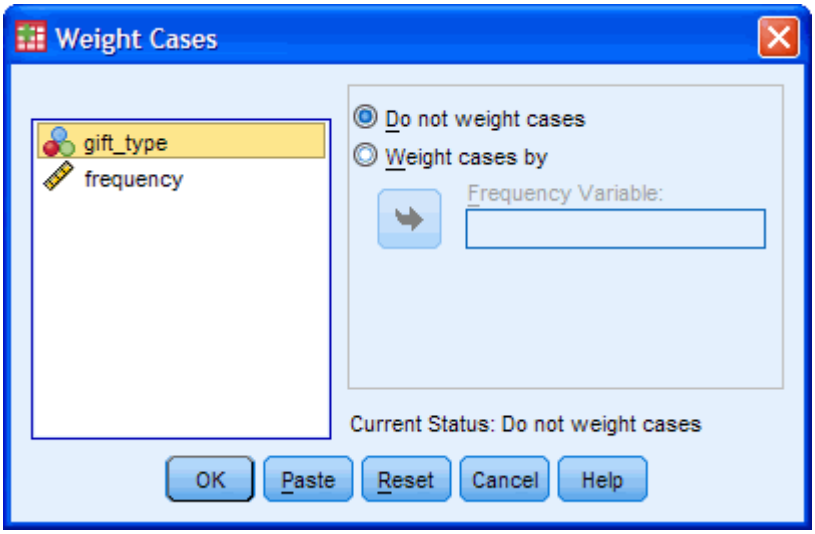
Weighting cases if summation used

a. Click **Data** > **Weight Cases...** on the top menu as shown below:



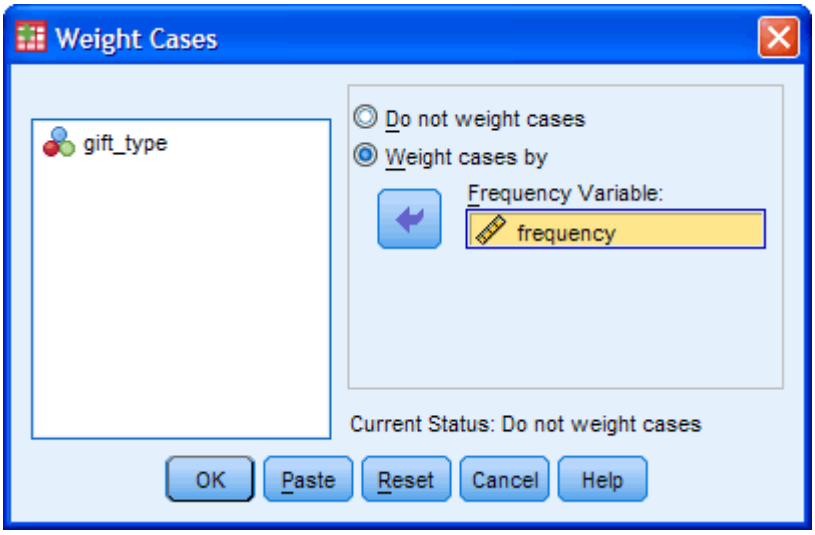
Published with written permission from SPSS Inc, an IBM Company.

b. You will be presented with the Weight Cases dialogue box as shown below:



Published with written permission from SPSS Inc, an IBM Company.

c. Select the "*Weight cases by*" radio box and transfer the "**frequency**" variable into the "*Frequency Variable:*" box, which has now become highlighted, by using the button. You will get the following screen:

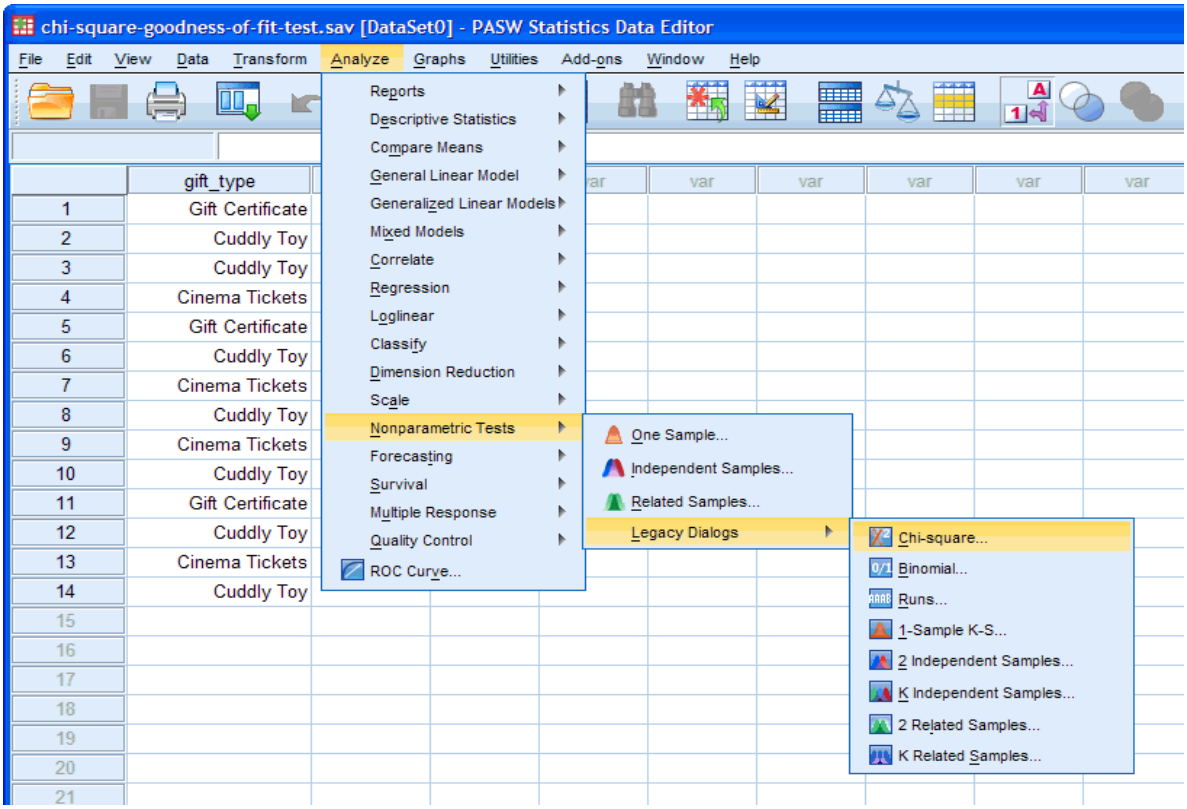


Published with written permission from SPSS Inc, an IBM Company.

d. Click the button.

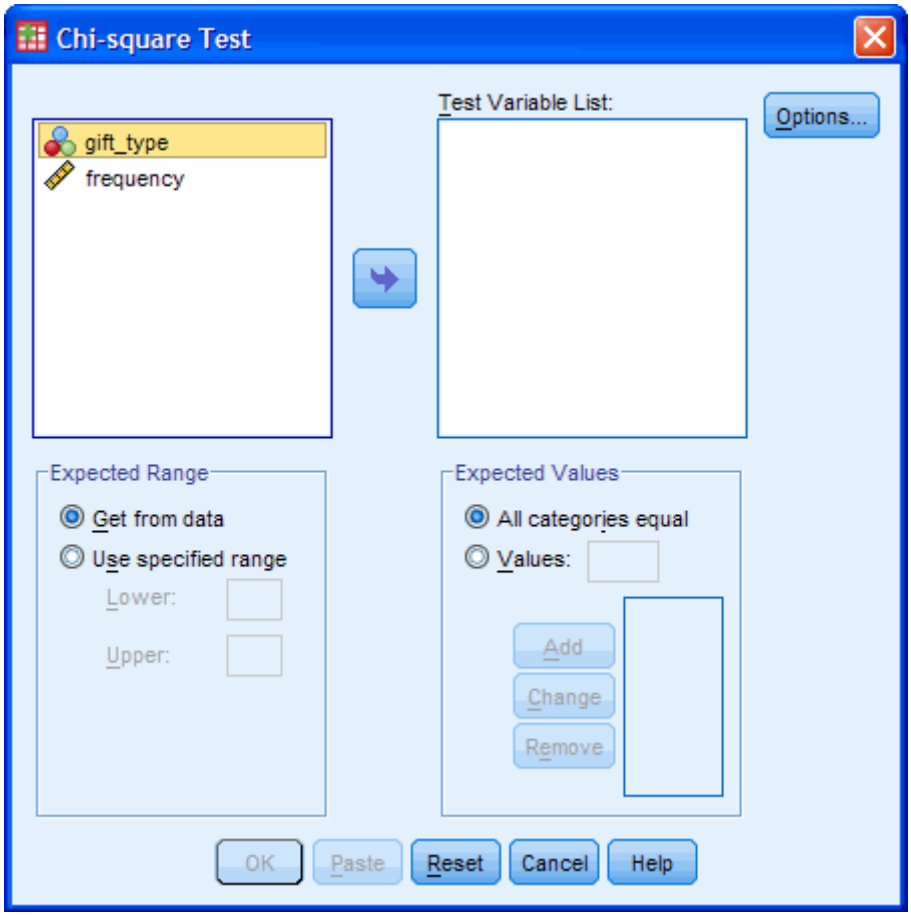
Procedure for both methods

1. Click **Analyze** > **Nonparametric Tests** > **Legacy Dialogs** > **Chi-square...** on the top menu as shown below: (If you are on older versions of SPSS you will not have to go through the **Legacy Dialogs** menu)



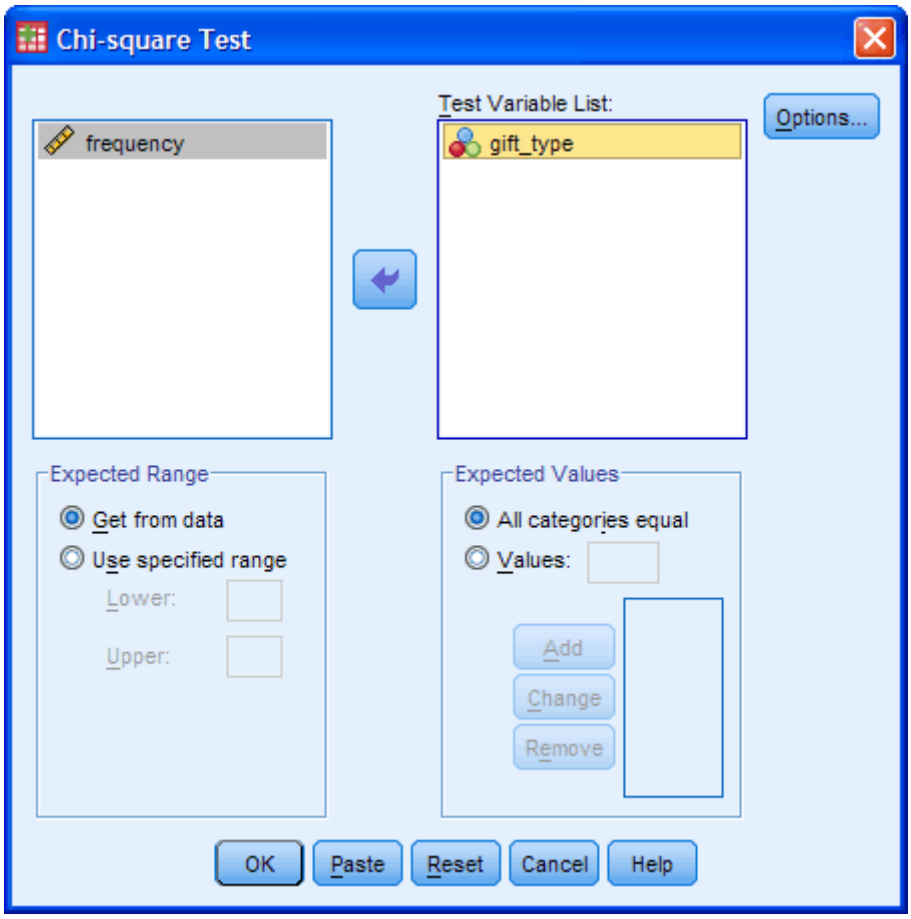
Published with written permission from SPSS Inc, an IBM Company.

2. You will be presented with the **Chi-square Test** dialogue box, as shown below: (The "**frequency**" variable will only be present if you entered in your data in summated form)



Published with written permission from SPSS Inc, an IBM Company.

3. Transfer the "**gift_type**" variable into the "*Test Variable List:*" box by using the button as shown below:



Published with written permission from SPSS Inc, an IBM Company.

[Keep the "*All categories equal*" option selecting in the "*Expected Values*" area as we are assuming equal proportions for each category.]

4. Click the OK button to generate the output.

SPSS Output for Chi-Square Goodness-of-Fit Test

The table below, **gift_type**, provides the observed frequencies (Observed N) for each gift as well as the expected frequencies (Expected N), which are the frequencies expected if the null hypothesis is true. The difference between the observed and expected frequencies is provided in the Residual column.

gift_type

	Observed N	Expected N	Residual
Gift Certificate	370	333.3	36.7
Cuddly Toy	230	333.3	-103.3
Cinema Tickets	400	333.3	66.7
Total	1000		

Published with written permission from SPSS Inc, an IBM Company.

The table below, **Test Statistics**, provides the actual result of the Chi-Square Goodness-of-Fit test. We can see from this table that our test statistic is

statistically significant: $\chi^2(2) = 49.4$, $p < .0005$. We can, therefore, reject the null hypothesis and conclude that there are statistically significant differences in the preference of the type of sign-up gift, with less people preferring the Cuddly Toy (N = 230) compared to either the Amazon Gift Certificate (N = 370) and Cinema Tickets (N = 400).

Test Statistics

	gift_type
Chi-square	49.400[a]
df	2
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 333.3.

Published with written permission from SPSS Inc, an IBM Company.

The footnotes to the table inform us that there are no expected frequencies less than a count of 5, which means that we have not violated one of the assumptions of the Chi-Square Goodness-of-Fit test. We could, of course, have determined this from the values in the **gift_type** table but SPSS conveniently calculates the percentage for us in the **Test Statistics** table.

Statistical Guide

Essentials

Descriptive and Inferential Statistics

205

When analysing data, for example, the marks achieved by 100 students for a piece of coursework, it is possible to use both descriptive and inferential statistics in your analysis of their marks. Typically, in most research conducted on groups of people, you will use both descriptive and inferential statistics to analyse your results and draw conclusions. *So what are descriptive and inferential statistics? And what are their differences?*

Descriptive Statistics

Descriptive statistics is the term given to the analysis of data that helps describe, show or summarize data in a meaningful way such that, for example, patterns might emerge from the data. Descriptive statistics do not, however, allow us to make conclusions beyond the data we have analysed or reach conclusions regarding any hypotheses we might have made. They are simply a way to describe our data.

Descriptive statistics are very important, as if we simply presented our raw data it would be hard to visulize what the data was showing, especially if there was a lot of it. Descriptive statistics therefore allow us to present the data in a more meaningful way which allows simpler interpretation of the data. For example, if we had the results of 100 pieces of students' coursework, we may be interested in the overall performance of those students. We would also be interested in the distribution or spread of the marks. Descriptive statistics allow us to do this. How to properly describe data through statistics and graphs is an important topic and discussed in other Laerd Statistics Guides. Typically, there are two general types of statistic that are used to describe data:

- Measures of central tendency: these are ways of describing the central position of a frequency distribution for a group of data. In this case, the frequency distribution is simply the distribution and pattern of marks scored by the 100 students from the lowest to the highest. We can describe this central position using a number of statistics, including the mode, median, and mean. You can read about measures of central tendency here.

- Measures of spread: these are ways of summarizing a group of data by describing how spread out the scores are. For example, the mean score of our 100 students may be 65 out of 100. However, not all students will have scored 65 marks. Rather, their scores will be spread out. Some will be lower and others higher. Measures of spread help us to summarize how spread out these scores are. To describe this spread, a number of statistics are available to us, including the range, quartiles, absolute deviation, variance and standard deviation.

When we use descriptive statistics it is useful to summarize our group of data using a combination of tabulated description (i.e. tables), graphical description (i.e. graphs and charts) and statistical commentary (i.e. a discussion of the results).

Inferential Statistics

We have seen that descriptive statistics provide information about our immediate group of data. For example, we could calculate the mean and standard deviation of the exam marks for the 100 students and this could provide valuable information about this group of 100 students. Any group of data like this, that includes all the data you are interested, in is called a **population**. A population can be small or large, as long as it includes all the data you are interested in. For example, if you were only interested in the exam marks of 100 students, then the 100 students would represent your population. Descriptive statistics are applied to populations and the properties of populations, like the mean or standard deviation, are called **parameters** as they represent the whole population (i.e. everybody you are interested in).

Often, however, you do not have access to the whole population you are interested in investigating but only have a limited number of data instead. For example, you might be interested in the exam marks of all students in the UK. It is not feasible to measure all exam marks of all students in the whole of the UK so you have to measure a smaller **sample** of students, for example, 100 students, that are used to represent the larger population of all UK students. Properties of samples, such as the mean or standard deviation, are not called parameters but **statistics**. Inferential statistics are techniques that allow us to use these samples to make generalizations about the populations from which the samples were drawn. It is, therefore, important the sample accurately represents the population. The process of achieving this is called sampling (sampling strategies are discussed in detail here on our sister site). Inferential statistics arise out of the fact that sampling naturally incurs sampling error and thus a sample is not expected to perfectly represent the population. The methods of inferential statistics are (1) the estimation of parameter(s) and (2) testing of statistical hypotheses.

We have provided some answers to common FAQs on the next page. Alternatively, why not now read our guide on Types of Variable?

FAQs - Descriptive and Inferential Statistics

33

What are the similarities between descriptive and inferential statistics?

Both descriptive and inferential statistics rely on the same set of data. Descriptive statistics rely solely on this set of data whilst inferential statistics also rely on this data in order to make generalisations about a larger population.

What are the strengths of using descriptive statistics to examine a distribution of scores?

Other than the clarity with which descriptive statistics can clarify large volumes of data, there are no uncertainties about the values you get (other than only measurement error, etc.).

What are the limitations of descriptive statistics?

Descriptive statistics are limited in so much that they only allow you to make summations about the people or objects that you have actually measured. You cannot use the data you have collected to generalize to other people or objects (i.e. using data from a sample to infer the properties/parameters of a population). For example, if you tested a drug to beat cancer and it worked in your patients, you cannot claim that it would work in other cancer patients only relying on descriptive statistics (but inferential statistics would give you this opportunity).

What are the limitations of inferential statistics?

There are two main limitations to the use of inferential statistics. The first, and most important, limitation, which is present in all inferential statistics, is that you

are providing data about a population that you have not fully measured and, therefore, cannot ever be completely sure that the values/statistics you calculate are correct. Remember, inferential statistics are based on the concept of using the values measured in a sample to estimate/infer the values that would be measured in a population; there will always be a degree of uncertainty in doing this. The second limitation is connected with the first limitation. Some, but not all, inferential tests require the user (i.e. you) to make educated guesses (based on theory) to run the inferential tests. Again, there will be some uncertainty in this process, which will have repercussions on the certainty of the results of some inferential statistics.

Why not now read our guide on Types of Variable?

Measures of Central Tendency

313

Introduction

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location. They are also classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as, the median and the mode.

The mean, median and mode are all valid measures of central tendency but, under different conditions, some measures of central tendency become more appropriate to use than others. In the following sections we will look at the mean, mode and median and learn how to calculate them and under what conditions they are most appropriate to be used.

Mean (Arithmetic)

The mean (or average) is the most popular and well known measure of central tendency. It can be used with both discrete and continuous data, although its use is most often with continuous data (see our Types of Variable guide for data types). The mean is equal to the sum of all the values in the data set divided by the number of values in the data set. So, if we have n values in a data set and they have values x_1, x_2, ..., x_n, then the sample mean, usually denoted by $\bar{x}$ (pronounced x bar), is:

$$\bar{x} = \frac{(x_1 + x_2 + \cdots + x_n)}{n}$$

This formula is usually written in a slightly different manner using the Greek capitol letter, $\sum$, pronounced "sigma", which means "sum of...":

$$\bar{x} = \frac{\sum x}{n}$$

You may have noticed that the above formula refers to the sample mean. So, why call have we called it a sample mean? This is because, in statistics, samples and populations have very different meanings and these differences are very important, even if, in the case of the mean, they are calculated in the same way. To acknowledge that we are calculating the population mean and not the sample mean, we use the Greek lower case letter "mu", denoted as μ:

$$\mu = \frac{\sum x}{n}$$

The mean is essentially a model of your data set. It is the value that is most common. You will notice, however, that the mean is not often one of the actual values that you have observed in your data set. However, one of its important properties is that it minimises error in the prediction of any one value in your data set. That is, it is the value that produces the lowest amount of error from all other values in the data set.

An important property of the mean is that it includes every value in your data set as part of the calculation. In addition, the mean is the only measure of central tendency where the sum of the deviations of each value from the mean is always zero.

When not to use the mean

The mean has one main disadvantage: it is particularly susceptible to the influence of outliers. These are values that are unusual compared to the rest of the data set by being especially small or large in numerical value. For example, consider the wages of staff at a factory below:

Staff	1	2	3	4	5	6	7	8	9	10
Salary	15k	18k	16k	14k	15k	15k	12k	17k	90k	95k

The mean salary for these ten staff is $30.7k. However, inspecting the raw data suggests that this mean value might not be the best way to accurately reflect the typical salary of a worker, as most workers have salaries in the $12k to 18k range. The mean is being skewed by the two large salaries. Therefore, in this situation we would like to have a better measure of central tendency. As we will find out later, taking the median would be a better measure of central tendency in this situation.

Another time when we usually prefer the median over the mean (or mode) is when our data is skewed (i.e. the frequency distribution for our data is skewed). If we consider the normal distribution - as this is the most frequently assessed in statistics - when the data is perfectly normal then the mean, median and mode are identical. Moreover, they all represent the most typical value in the data set. However, as the data becomes skewed the mean loses its ability to provide the best central location for the data as the skewed data is dragging it away from the typical value. However, the median best retains this position and is not as strongly influenced by the skewed values. This is explained in more detail in the skewed distribution section later in this guide.

Median

The median is the middle score for a set of data that has been arranged in order of magnitude. The median is less affected by outliers and skewed data. In order to calculate the median, suppose we have the data below:

65	55	89	56	35	14	56	55	87	45	92

We first need to rearrange that data into order of magnitude (smallest first):

14	35	45	55	55	**56**	56	65	87	89	92

Our median mark is the middle mark - in this case 56 (highlighted in bold). It is the middle mark because there are 5 scores before it and 5 scores after it. This works fine when you have an odd number of scores but what happens when you have an even number of scores? What if you had only 10 scores? Well, you simply have to take the middle two scores and average the result. So, if we look at the example below:

65	55	89	56	35	14	56	55	87	45

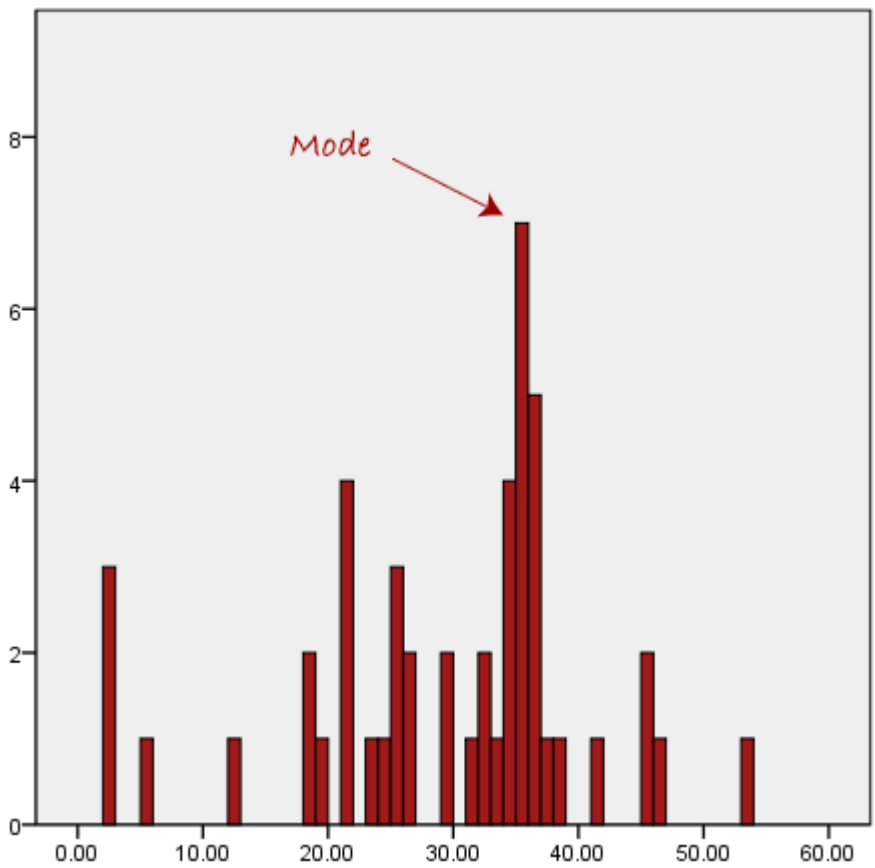
We again rearrange that data into order of magnitude (smallest first):

14	35	45	55	**55**	**56**	56	65	87	89	92

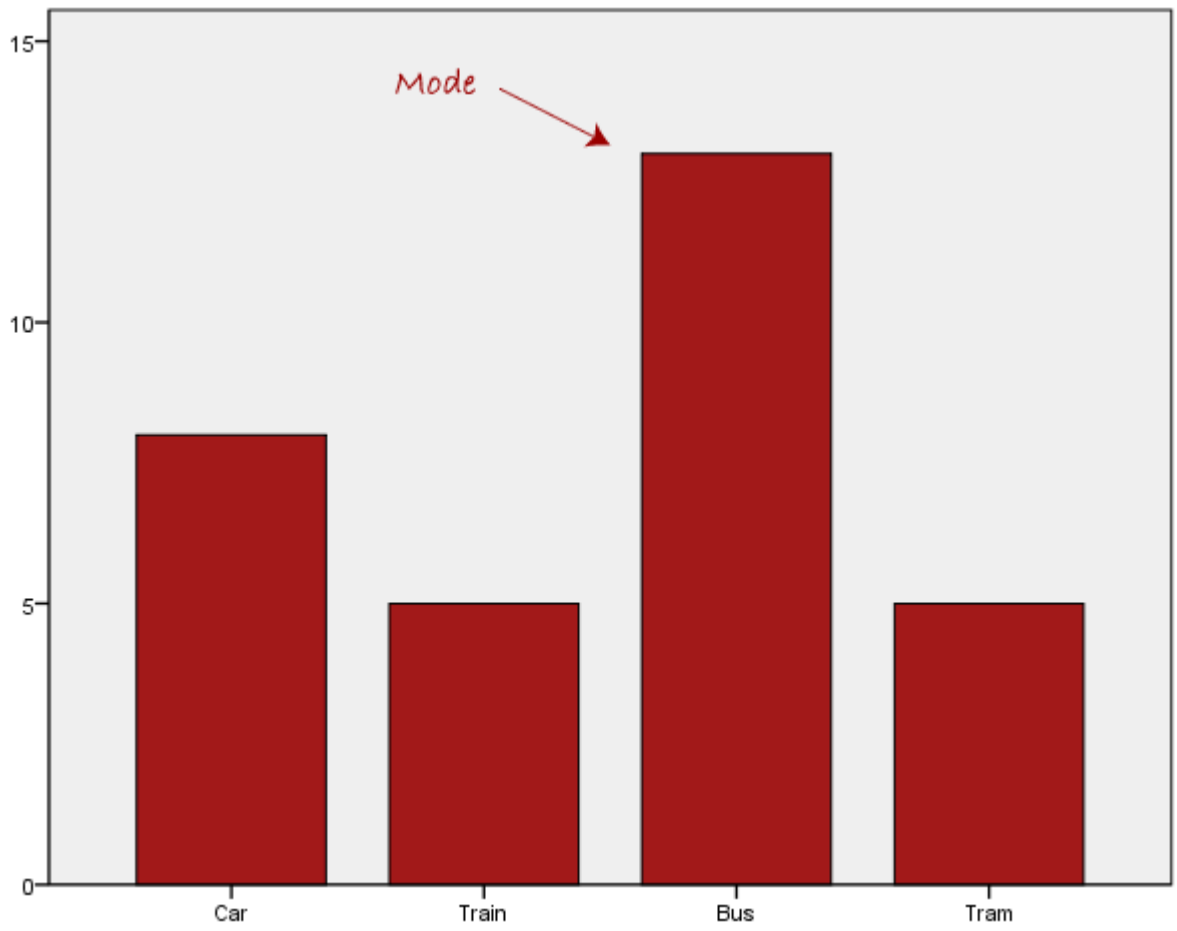
Only now we have to take the 5th and 6th score in our data set and average them to get a median of 55.5.

Mode

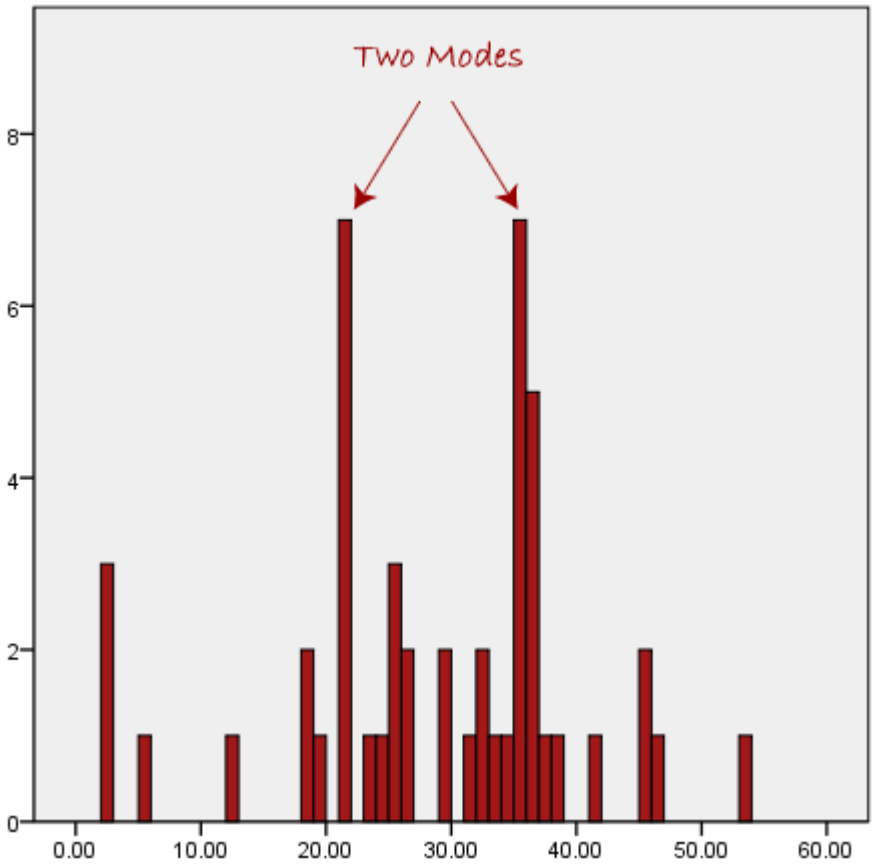
The mode is the most frequent score in our data set. On a histogram it represents the highest bar in a bar chart or histogram. You can, therefore, sometimes consider the mode as being the most popular option. An example of a mode is presented below:



Normally, the mode is used for categorical data where we wish to know which is the most common category as illustrated below:

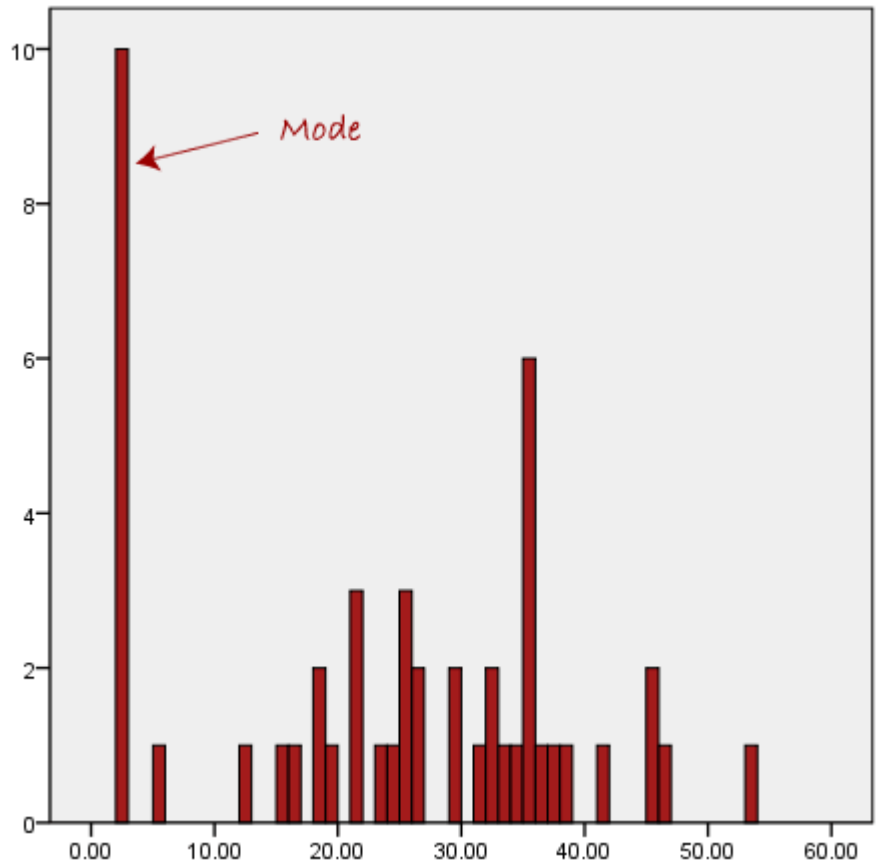


We can see above that the most common form of transport, in this particular data set, is the bus. However, one of the problems with the mode is that it is not unique, so it leaves us with problems when we have two or more values that share the highest frequency, such as below:



We are now stuck as to which mode best describes the central tendency of the data. This is particularly problematic when we have continuous data, as we are more likely not to have any one value that is more frequent than the other. For example, consider measuring 30 peoples' weight (to the nearest 0.1 kg). How likely is it that we will find two or more people with **exactly** the same weight, e.g. 67.4 kg? The answer, is probably very unlikely - many people might be close but with such a small sample (30 people) and a large range of possible weights you are unlikely to find two people with exactly the same weight, that is, to the nearest 0.1 kg. This is why the mode is very rarely used with continuous data.

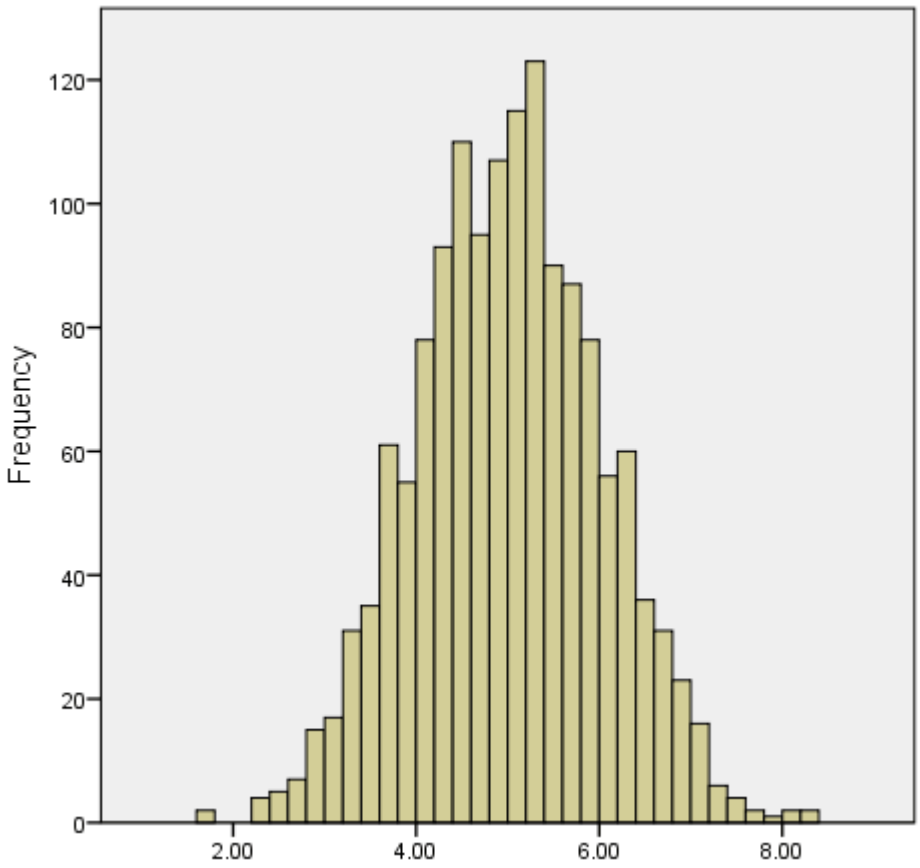
Another problem with the mode is that it will not provide us with a very good measure of central tendency when the most common mark is far away from the rest of the data in the data set, as depicted in the diagram below:



In the above diagram the mode has a value of 2. We can clearly see, however, that the mode is not representative of the data, which is mostly concentrated around the 20 to 30 value range. To use the mode to describe the central tendency of this data set would be misleading.

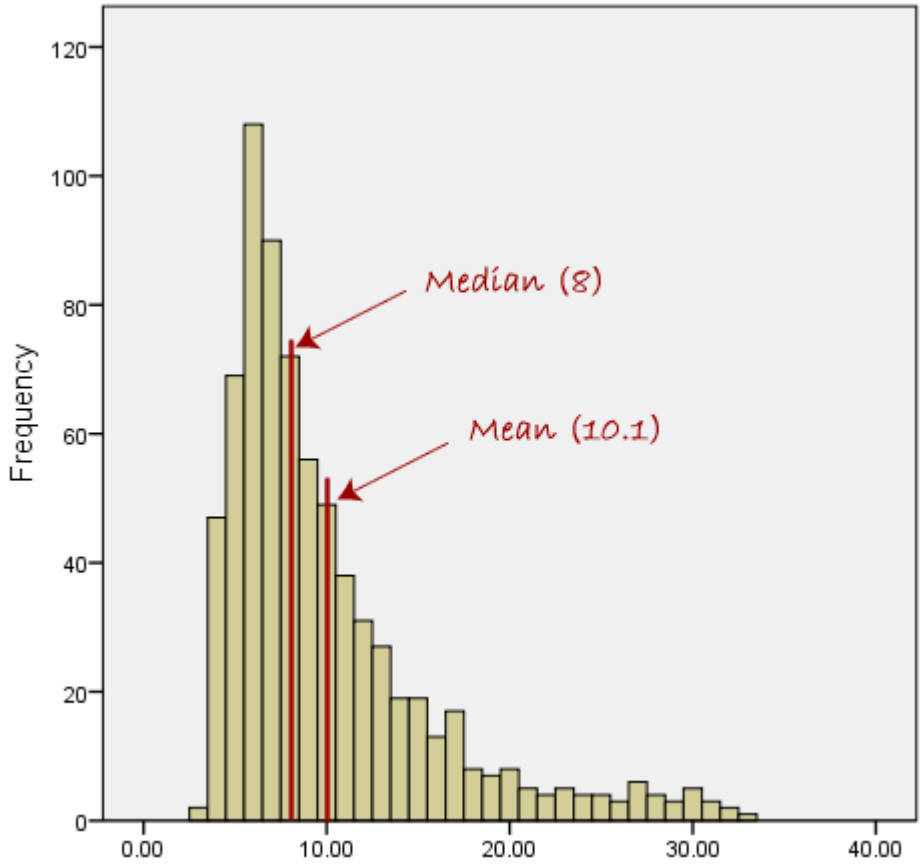
Skewed Distributions and the Mean and Median

We often test whether our data is normally distributed as this is a common assumption underlying many statistical tests. An example of a normally distributed set of data is presented below:



When you have a normally distributed sample you can legitimately use both the mean or the median as your measure of central tendency. In fact, in any symmetrical distribution the mean, median and mode are equal. However, in this situation, the mean is widely preferred as the best measure of central tendency as it is the measure that includes all the values in the data set for its calculation, and any change in any of the scores will affect the value of the mean. This is not the case with the median or mode.

However, when our data is skewed, for example, as with the right-skewed data set below:



we find that the mean is being dragged in the direct of the skew. In these situations, the median is generally considered to be the best representative of the central location of the data. The more skewed the distribution the greater the difference between the median and mean, and the greater emphasis should be placed on using the median as opposed to the mean. A classic example of the above right-skewed distribution is income (salary), where higher-earners provide a false representation of the typical income if expressed as a mean and not a median.

If dealing with a normal distribution, and tests of normality show that the data is non-normal, then it is customary to use the median instead of the mean. This is more a rule of thumb than a strict guideline however. Sometimes, researchers wish to report the mean of a skewed distribution if the median and mean are not appreciably different (a subjective assessment) and if it allows easier comparisons to previous research to be made.

Summary of when to use the mean, median and mode

Please use the following summary table to know what the best measure of central tendency is with respect to the different types of variable.

Type of Variable	Best measure of central tendency
Nominal	Mode
Ordinal	Median
Interval/Ratio (not skewed)	Mean
Interval/Ratio (skewed)	Median

For answers to frequently asked questions about measures of central tendency, please go the next page.

FAQs - Measures of Central Tendency

83

Please find below some common questions that are asked regarding measures of central tendency, along with their answers. These FAQs are in addition to our article on measures of central tendency found on the previous page.

What is the best measure of central tendency?

There can often be a "best" measure of central tendency with regards to the data you are analysing but there is no one "best" measure of central tendency. This is because whether you use the median, mean or mode will depend on the type of data you have (see our Types of Variable guide), such as nominal or continuous data; whether your data has outliers and/or is skewed; and what you are trying to show from your data. Further considerations of when to use each measure of central tendency is found in our guide on the previous page.

In a strongly skewed distribution, what is the best indicator of central tendency?

It is usually inappropriate to use the mean in such situations where your data is skewed. You would normally choose the median or mode with the median usually preferred. This is discussed on the previous page under the subtitle, "When not to use the mean".

Does all data have a median, mode and mean?

Yes and no. All continuous data has a median, mode and mean. However, strictly speaking, ordinal data has a median and mode only, and nominal data has only a mode. However, a consensus has not been reached among statisticians about whether the mean can be used with ordinal data and you can often see a mean reported for Likert data in research.

When is the mean the best measure of central tendency?

The mean is usually the best measure of central tendency to use when your data distribution is continuous and symmetrical, such as when your data is normally distributed. However, it all depends on what you are trying to show from your data.

When is the mode the best measure of central tendency?

The mode is the least used of the measures of central tendency and can only be used when dealing with nominal data. For this reason, the mode will be the best measure of central tendency (as it is the only one appropriate to use) when dealing with nominal data. The mean and/or median are usually preferred when dealing with all other types of data but this does not mean it is never used with these data types.

When is the median the best measure of central tendency?

The median is usually preferred to other measures of central tendency when your data set is skewed (i.e. forms a skewed distribution) or you are dealing with ordinal data. However, the mode can also be appropriate in these situations but is not as commonly used as the median.

What is the most appropriate measure of central tendency when the data has outliers?

The median is usually preferred in these situations as the value of the mean can be distorted by the outliers. However, it will depend on the how influential the outliers are. If they do not significantly distort the mean then using the mean as the measure of central tendency will usually be preferred.

In a normally distributed data set, which is greatest: mode, median or mean?

If the data set is perfectly normal then the mean, median and mean are equal to each other (i.e. the same value).

For any data set, which measures of central tendency have only one value?

The median and mean can only have one value for a given data set. The mode can have more than one value (see Mode section on previous page).

Measures of Spread

25

Introduction

A measure of spread, sometimes also called a measure of dispersion, is used to describe the variability in a sample or population. It is usually used in conjunction with a measure of central tendency, such as, the mean or median, to provide an overall description of a set of data.

Why is it important to measure the spread of data?

There are many reasons why the measure of the spread of data values is important but one of the main reasons regards its relationship with measures of central tendency. A measure of spread gives us an idea of how well the mean, for example, represents the data. If the spread of values in the data set is large then the mean is not as representative of the data as if the spread of data is small. This is because a large spread indicates that there are probably large differences between individual scores. Additionally, in research, it is often seen as positive if there is little variation in each data group as it indicates that the similar.

We will be looking at the range, quartiles, variance, absolute deviation and standard deviation.

Range

The range is the difference between the highest and lowest scores in a data set and is the simplest measure of spread. So we calculate range as:

Range = maximum value - minimum value

For example, let us consider the following data set:

23 56 45 65 59 55 62 54 85 25

The maximum value is 85 and the minimum value is 23. This results in a range of 62, which is 85 minus 23. Whilst using the range as a measure of spread is limited, it does set the boundaries of the scores. This can be useful if you are measuring a variable that has either a critical low or high threshold (or both) that should not be crossed. The range will instantly inform you whether at least one value broke these critical thresholds. In addition, the range can be used to detect any errors when entering data. For example, if you have recorded the age of school children in your study and your range is 7 to 123 years old you know you have made a mistake!

Quartiles and Interquartile Range

Quartiles tell us about the spread of a data set by breaking the data set into quarters, just like the median breaks it in half. For example, consider the marks of the 100 students below, which have been ordered from the lowest to the highest scores, and the quartiles highlighted in red.

Order	Score	Order	Score	Order	Score	Order	Score	Order	Score
1st	35	21st	42	41st	53	61st	64	81st	74
2nd	37	22nd	42	42nd	53	62nd	64	82nd	74
3rd	37	23rd	44	43rd	54	63rd	65	83rd	74
4th	38	24th	44	44th	55	64th	66	84th	75
5th	39	25th	45	45th	55	65th	67	85th	75
6th	39	26th	45	46th	56	66th	67	86th	76
7th	39	27th	45	47th	57	67th	67	87th	77
8th	39	28th	45	48th	57	68th	67	88th	77
9th	39	29th	47	49th	58	69th	68	89th	79
10th	40	30th	48	50th	58	70th	69	90th	80
11th	40	31st	49	51st	59	71st	69	91st	81
12th	40	32nd	49	52nd	60	72nd	69	92nd	81
13th	40	33rd	49	53rd	61	73rd	70	93rd	81
14th	40	34th	49	54th	62	74th	70	94th	81
15th	40	35th	51	55th	62	75th	71	95th	81
16th	41	36th	51	56th	62	76th	71	96th	81
17th	41	37th	51	57th	63	77th	71	97th	83
18th	42	38th	51	58th	63	78th	72	98th	84
19th	42	39th	52	59th	64	79th	74	99th	84
20th	42	40th	52	60th	64	80th	74	100th	85

The **first quartile** (Q1) lies between the 25th and 26th student's marks, the **second quartile** (Q2) between the 50th and 51st student's marks, and the **third quartile** (Q3) between the 75th and 76th student's marks. Hence:

First quartile (Q1) = 45 + 45 ÷ 2 = **45**
Second quartile (Q2) = 58 + 59 ÷ 2 = **58.5**
Third quartile (Q3) = 71 + 71 ÷ 2 = **71**

In the above example, we have an even number of scores (100 students rather than an odd number such as 99 students). This means that when we calculate the quartiles, we take the sum of the two scores around each quartile and then half them (hence Q1= 45 + 45 ÷ 2 = 45) . However, if we had an odd number of scores (say, 99 students), then we would only need to take one score for each quartile (that is, the 25th, 50th and 75th scores). You should recognize that the second quartile is also the median.

Quartiles are a useful measure of spread because they are much less affected by outliers or a skewed data set than the equivalent measures of mean and standard deviation. For this reason, quartiles are often reported along with the median as the best choice of measure of spread and central tendency, respectively, when dealing with skewed and/or data with outliers. A common way of expressing quartiles is as an interquartile range. The interquartile range describes the difference between the third quartile (Q3) and the first quartile (Q1), telling us about the range of the middle half of the scores in the distribution. Hence, for our 100 students:

Interquartile range = Q3 - Q1
= 71 - 45
= 26

However, it should be noted that in journals and other publications you will usually see the interquartile range reported as 45 to 71, rather than the calculated range.

A slight variation on this is the semi-interquartile range, which is half the interquartile range = ½ (Q3 - Q1). Hence, for our 100 students, this would be 26 ÷ 2 = 13.

Absolute Deviation & Variance

12

Variation

Quartiles are useful but they are also somewhat limited because they do not take into account every score in our group of data. To get a more representative idea of spread we need to take into account the actual values of each score in a data set. The absolute deviation, variance and standard deviation are such measures.

The absolute and mean absolute deviation show the amount of deviation (variation) that occurs around the mean score. To find the total variability in our group of data, we simply add up the deviation of each score from the mean. The average deviation of a score can then be calculated by dividing this total by the number of scores. How we calculate the deviation of a score from the mean

depends on our choice of statistic, whether we use absolute deviation, variance or standard deviation.

Absolute Deviation and Mean Absolute Deviation

Perhaps the simplest way of calculating the deviation of a score from the mean is to take each score and minus the mean score. For example, the mean score for the group of 100 students we used earlier was 58.75 out of 100. Therefore, if we took a student that scored 60 out of 100, the deviation of a score from the mean is 60 - 58.75 = 1.25. It is important to note that scores above the mean have positive deviations (as demonstrated above) whilst that scores below the mean will have negative deviations.

To find out the total variability in our data set, we would perform this calculation for all of the 100 students' scores. However, the problem is that because we have both positive and minus signs, when we add up all of these deviations, they cancel each other out, giving us a total deviation of zero. Since we are only interested in the deviations of the scores and not whether they are above or below the mean score, we can ignore the minus sign and take only the absolute value, giving us the **absolute deviation**. Adding up all of these absolute deviations and dividing them by the total number of scores then gives us the mean absolute deviation (see below). Therefore, for our 100 students the mean absolute deviation is 12.81, as shown below:

$$mean\ absolute\ deviation = \frac{\sum|X-\mu|}{N}$$

$$= \frac{1281}{100}$$

$$= 12.81$$

Where μ = mean, X = score, ∑= the sum of, N = number of scores, ∑X = "add up all the scores", | | = take the absolute value (i.e. ignore the minus sign).

Variance

Another method for calculating the deviation of a group of scores from the mean, such as the 100 students we used earlier, is to use the variance. Unlike the absolute deviation, which uses the absolute value of the deviation in order to "rid itself" of the negative values, the variance achieves positive values by squaring each of the deviations instead. Adding up these squared deviations gives us the sum of squares, which we can then divide by the total number of scores in our group of data (in other words, 100 because there are 100 students) to find the variance (see below). Therefore, for our 100 students, the variance is 211.89, as shown below:

$$variance = \frac{\sum(X-\mu)^2}{N}$$

$$= \frac{21188.75}{100}$$

$$= 211.89$$

Where μ = mean, X = score, ∑= the sum of, N = number of scores, ∑X = "add up all the scores"

As a measure of variability, the variance is useful. If the scores in our group of data are spread out then the variance will be a large number. Conversely, if the scores are spread closely around the mean, then the variance will be a smaller number. However, there are two potential problems with the variance. First, because the deviations of scores from the mean are 'squared', this gives more weight to extreme scores. If our data contains outliers (in other words, one or a small number of scores that are particularly far away from the mean and perhaps do not represent well our data as a whole) this can give undo weight to these scores. Secondly, the variance is not in the same units as the scores in our data set: variance is measured in the units squared. This means we cannot place it on our frequency distribution and cannot directly relate its value to the values in our data set. Therefore, the figure of 211.89, our variance, appears somewhat arbitrary. Calculating the standard deviation rather than the variance rectifies this problem. Nonetheless, analysing variance is extremely important in some statistical analyses, discussed in other statistical guides.

Standard Deviation

54

Introduction

The standard deviation is a measure of the spread of scores within a set of data. Usually, we are interested in the standard deviation of a population. However, as we are often presented with data from a sample only, we can estimate the population standard deviation from a sample standard deviation. These two standard deviations, sample and population standard deviations, are calculated differently. In statistics we are usually presented with having to calculate sample standard deviations, and so this is what this article will focus on, although the formula for a population standard deviation will also be shown.

When to use the sample or population standard deviation

We are normally interested in knowing the population standard deviation as our population contains all the values we are interested in. Therefore, you would normally calculate the population standard deviation if: (1) you have the entire population or (2) you have a sample of a larger population but you are only interested in this sample and do not wish to generalize your findings to the population. However, in statistics, we are usually presented with a sample from which we wish to estimate (generalize to) a population, and the standard deviation is no exception to this. Therefore, if all you have is a sample but you

wish to make a statement about the population standard deviation from which the sample is drawn, then you need to use the sample standard deviation. Confusion can often arise as to which standard deviation to use due to the name "sample" standard deviation incorrectly being interpreted as meaning the standard deviation of the sample itself and not as the estimate of the population standard deviation based on the sample.

What type of data should you use when you calculate a standard deviation?

The standard deviation is used in conjunction with the mean, to summarise continuous data not categorical data. In addition, the standard deviation, like the mean, is normally only appropriate when the continuous data is not significantly skewed or has outliers.

Examples of when to use the sample or population standard deviation

Q. A teacher sets an exam for their pupils. The teacher wants to summarize the results the pupils attained as a mean and standard deviation. Which standard deviation should be used?

A. Population standard deviation. Why? Because the teacher is only interested in this class of pupils' scores and nobody else.

Q. A researcher has recruited males aged 45 to 65 years old for an exercise training study to investigate risk markers for heart disease, e.g. cholesterol. Which standard deviation would most likely be used?

A. Sample standard deviation. Although not explicitly stated, a researcher investigating health related issues will not be simply concerned with just the participants of their study; they will want to show how their sample results can be generalised to the whole population (in this case, males aged 45 to 65 years old). Hence, the use of the sample standard deviation.

Q. One of the questions on a national consensus survey asks for respondent's age. Which standard deviation would be used to describe the variation in all ages received from the consensus?

A. Population standard deviation. A national consensus is used to find out information about the nation's citizens. By definition, it includes the whole population, therefore, a population standard deviation would be used.

What are the formulas for the standard deviation?

The **sample standard deviation formula** is:

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}}$$

where,

s = sample standard deviation
$\sum$= sum of...

$\bar{X}$ = sample mean
n = number of scores in sample.

The **population standard deviation** formula is:

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$$

where,

σ = population standard deviation
$\sum$ = sum of...
μ = population mean
n = number of scores in sample.

Is there an easy way to calculate the standard deviation?

Yes, we have a sample and population standard deviation calculator that shows you all the working as well! It can be found in our calculator section here.

Absolute Deviation & Variance

12

Variation

Quartiles are useful but they are also somewhat limited because they do not take into account every score in our group of data. To get a more representative idea of spread we need to take into account the actual values of each score in a data set. The absolute deviation, variance and standard deviation are such measures.

The absolute and mean absolute deviation show the amount of deviation (variation) that occurs around the mean score. To find the total variability in our group of data, we simply add up the deviation of each score from the mean. The average deviation of a score can then be calculated by dividing this total by the number of scores. How we calculate the deviation of a score from the mean depends on our choice of statistic, whether we use absolute deviation, variance or standard deviation.

Absolute Deviation and Mean Absolute Deviation

Perhaps the simplest way of calculating the deviation of a score from the mean is to take each score and minus the mean score. For example, the mean score for the group of 100 students we used earlier was 58.75 out of 100. Therefore, if we took a student that scored 60 out of 100, the deviation of a score from the mean is 60 - 58.75 = 1.25. It is important to note that scores above the mean have positive deviations (as demonstrated above) whilst that scores below the mean will have negative deviations.

To find out the total variability in our data set, we would perform this calculation for all of the 100 students' scores. However, the problem is that because we have both positive and minus signs, when we add up all of these deviations, they cancel each other out, giving us a total deviation of zero. Since we are only

interested in the deviations of the scores and not whether they are above or below the mean score, we can ignore the minus sign and take only the absolute value, giving us the **absolute deviation**. Adding up all of these absolute deviations and dividing them by the total number of scores then gives us the mean absolute deviation (see below). Therefore, for our 100 students the mean absolute deviation is 12.81, as shown below:

$$mean\ absolute\ deviation = \frac{\sum|X-\mu|}{N}$$

$$= \frac{1281}{100}$$

$$= 12.81$$

Where μ = mean, X = score, ∑= the sum of, N = number of scores, ∑X = "add up all the scores", | | = take the absolute value (i.e. ignore the minus sign).

Variance

Another method for calculating the deviation of a group of scores from the mean, such as the 100 students we used earlier, is to use the variance. Unlike the absolute deviation, which uses the absolute value of the deviation in order to "rid itself" of the negative values, the variance achieves positive values by squaring each of the deviations instead. Adding up these squared deviations gives us the sum of squares, which we can then divide by the total number of scores in our group of data (in other words, 100 because there are 100 students) to find the variance (see below). Therefore, for our 100 students, the variance is 211.89, as shown below:

$$variance = \frac{\sum(X-\mu)^2}{N}$$

$$= \frac{21188.75}{100}$$

$$= 211.89$$

Where μ = mean, X = score, ∑= the sum of, N = number of scores, ∑X = "add up all the scores"

As a measure of variability, the variance is useful. If the scores in our group of data are spread out then the variance will be a large number. Conversely, if the scores are spread closely around the mean, then the variance will be a smaller number. However, there are two potential problems with the variance. First, because the deviations of scores from the mean are 'squared', this gives more weight to extreme scores. If our data contains outliers (in other words, one or a small number of scores that are particularly far away from the mean and perhaps do not represent well our data as a whole) this can give undo weight to these scores. Secondly, the variance is not in the same units as the scores in our data set: variance is measured in the units squared. This means we cannot place it on our frequency distribution and cannot directly relate its value to the values in our

data set. Therefore, the figure of 211.89, our variance, appears somewhat arbitrary. Calculating the standard deviation rather than the variance rectifies this problem. Nonetheless, analysing variance is extremely important in some statistical analyses, discussed in other statistical guides.

Absolute Deviation & Variance

12

Variation

Quartiles are useful but they are also somewhat limited because they do not take into account every score in our group of data. To get a more representative idea of spread we need to take into account the actual values of each score in a data set. The absolute deviation, variance and standard deviation are such measures.

The absolute and mean absolute deviation show the amount of deviation (variation) that occurs around the mean score. To find the total variability in our group of data, we simply add up the deviation of each score from the mean. The average deviation of a score can then be calculated by dividing this total by the number of scores. How we calculate the deviation of a score from the mean depends on our choice of statistic, whether we use absolute deviation, variance or standard deviation.

Absolute Deviation and Mean Absolute Deviation

Perhaps the simplest way of calculating the deviation of a score from the mean is to take each score and minus the mean score. For example, the mean score for the group of 100 students we used earlier was 58.75 out of 100. Therefore, if we took a student that scored 60 out of 100, the deviation of a score from the mean is 60 - 58.75 = 1.25. It is important to note that scores above the mean have positive deviations (as demonstrated above) whilst that scores below the mean will have negative deviations.

To find out the total variability in our data set, we would perform this calculation for all of the 100 students' scores. However, the problem is that because we have both positive and minus signs, when we add up all of these deviations, they cancel each other out, giving us a total deviation of zero. Since we are only interested in the deviations of the scores and not whether they are above or below the mean score, we can ignore the minus sign and take only the absolute value, giving us the **absolute deviation**. Adding up all of these absolute deviations and dividing them by the total number of scores then gives us the mean absolute deviation (see below). Therefore, for our 100 students the mean absolute deviation is 12.81, as shown below:

$$mean\ absolute\ deviation = \frac{\sum|X-\mu|}{N}$$

$$= \frac{1281}{100}$$

$$= 12.81$$

Where μ = mean, X = score, ∑= the sum of, N = number of scores, ∑X = "add up all the scores", | | = take the absolute value (i.e. ignore the minus sign).

Variance

Another method for calculating the deviation of a group of scores from the mean, such as the 100 students we used earlier, is to use the variance. Unlike the absolute deviation, which uses the absolute value of the deviation in order to "rid itself" of the negative values, the variance achieves positive values by squaring each of the deviations instead. Adding up these squared deviations gives us the sum of squares, which we can then divide by the total number of scores in our group of data (in other words, 100 because there are 100 students) to find the variance (see below). Therefore, for our 100 students, the variance is 211.89, as shown below:

$$variance = \frac{\sum(X-\mu)^2}{N}$$

$$= \frac{21188.75}{100}$$

$$= 211.89$$

Where μ = mean, X = score, ∑= the sum of, N = number of scores, ∑X = "add up all the scores"

As a measure of variability, the variance is useful. If the scores in our group of data are spread out then the variance will be a large number. Conversely, if the scores are spread closely around the mean, then the variance will be a smaller number. However, there are two potential problems with the variance. First, because the deviations of scores from the mean are 'squared', this gives more weight to extreme scores. If our data contains outliers (in other words, one or a small number of scores that are particularly far away from the mean and perhaps do not represent well our data as a whole) this can give undo weight to these scores. Secondly, the variance is not in the same units as the scores in our data set: variance is measured in the units squared. This means we cannot place it on our frequency distribution and cannot directly relate its value to the values in our data set. Therefore, the figure of 211.89, our variance, appears somewhat arbitrary. Calculating the standard deviation rather than the variance rectifies this problem. Nonetheless, analysing variance is extremely important in some statistical analyses, discussed in other statistical guides.

Standard Deviation

Introduction

The standard deviation is a measure of the spread of scores within a set of data. Usually, we are interested in the standard deviation of a population. However, as we are often presented with data from a sample only, we can estimate the population standard deviation from a sample standard deviation. These two standard deviations, sample and population standard deviations, are calculated differently. In statistics we are usually presented with having to calculate sample standard deviations, and so this is what this article will focus on, although the formula for a population standard deviation will also be shown.

When to use the sample or population standard deviation

We are normally interested in knowing the population standard deviation as our population contains all the values we are interested in. Therefore, you would normally calculate the population standard deviation if: (1) you have the entire population or (2) you have a sample of a larger population but you are only interested in this sample and do not wish to generalize your findings to the population. However, in statistics, we are usually presented with a sample from which we wish to estimate (generalize to) a population, and the standard deviation is no exception to this. Therefore, if all you have is a sample but you wish to make a statement about the population standard deviation from which the sample is drawn, then you need to use the sample standard deviation. Confusion can often arise as to which standard deviation to use due to the name "sample" standard deviation incorrectly being interpreted as meaning the standard deviation of the sample itself and not as the estimate of the population standard deviation based on the sample.

What type of data should you use when you calculate a standard deviation?

The standard deviation is used in conjunction with the mean, to summarise continuous data not categorical data. In addition, the standard deviation, like the mean, is normally only appropriate when the continuous data is not significantly skewed or has outliers.

Examples of when to use the sample or population standard deviation

Q. A teacher sets an exam for their pupils. The teacher wants to summarize the results the pupils attained as a mean and standard deviation. Which standard deviation should be used?

A. Population standard deviation. Why? Because the teacher is only interested in this class of pupils' scores and nobody else.

Q. A researcher has recruited males aged 45 to 65 years old for an exercise training study to investigate risk markers for heart disease, e.g. cholesterol. Which standard deviation would most likely be used?

A. Sample standard deviation. Although not explicitly stated, a researcher investigating health related issues will not be simply concerned with just the participants of their study; they will want to show how their sample results can be

generalised to the whole population (in this case, males aged 45 to 65 years old). Hence, the use of the sample standard deviation.

Q. One of the questions on a national consensus survey asks for respondent's age. Which standard deviation would be used to describe the variation in all ages received from the consensus?

A. Population standard deviation. A national consensus is used to find out information about the nation's citizens. By definition, it includes the whole population, therefore, a population standard deviation would be used.

What are the formulas for the standard deviation?

The **sample standard deviation formula** is:

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}}$$

where,

s = sample standard deviation
$\sum$= sum of...
$\bar{X}$= sample mean
n = number of scores in sample.

The **population standard deviation** formula is:

$$\sigma = \sqrt{\frac{\sum(X - \mu)^2}{n}}$$

where,

σ= population standard deviation
$\sum$= sum of...
μ= population mean
n = number of scores in sample.

Is there an easy way to calculate the standard deviation?

Yes, we have a sample and population standard deviation calculator that shows you all the working as well! It can be found in our calculator section here.

Standard Deviation

54

Introduction

The standard deviation is a measure of the spread of scores within a set of data. Usually, we are interested in the standard deviation of a population. However, as we are often presented with data from a sample only, we can estimate the population standard deviation from a sample standard deviation. These two standard deviations, sample and population standard deviations, are calculated

differently. In statistics we are usually presented with having to calculate sample standard deviations, and so this is what this article will focus on, although the formula for a population standard deviation will also be shown.

When to use the sample or population standard deviation

We are normally interested in knowing the population standard deviation as our population contains all the values we are interested in. Therefore, you would normally calculate the population standard deviation if: (1) you have the entire population or (2) you have a sample of a larger population but you are only interested in this sample and do not wish to generalize your findings to the population. However, in statistics, we are usually presented with a sample from which we wish to estimate (generalize to) a population, and the standard deviation is no exception to this. Therefore, if all you have is a sample but you wish to make a statement about the population standard deviation from which the sample is drawn, then you need to use the sample standard deviation. Confusion can often arise as to which standard deviation to use due to the name "sample" standard deviation incorrectly being interpreted as meaning the standard deviation of the sample itself and not as the estimate of the population standard deviation based on the sample.

What type of data should you use when you calculate a standard deviation?

The standard deviation is used in conjunction with the mean, to summarise continuous data not categorical data. In addition, the standard deviation, like the mean, is normally only appropriate when the continuous data is not significantly skewed or has outliers.

Examples of when to use the sample or population standard deviation

Q. A teacher sets an exam for their pupils. The teacher wants to summarize the results the pupils attained as a mean and standard deviation. Which standard deviation should be used?

A. Population standard deviation. Why? Because the teacher is only interested in this class of pupils' scores and nobody else.

Q. A researcher has recruited males aged 45 to 65 years old for an exercise training study to investigate risk markers for heart disease, e.g. cholesterol. Which standard deviation would most likely be used?

A. Sample standard deviation. Although not explicitly stated, a researcher investigating health related issues will not be simply concerned with just the participants of their study; they will want to show how their sample results can be generalised to the whole population (in this case, males aged 45 to 65 years old). Hence, the use of the sample standard deviation.

Q. One of the questions on a national consensus survey asks for respondent's age. Which standard deviation would be used to describe the variation in all ages received from the consensus?

A. Population standard deviation. A national consensus is used to find out information about the nation's citizens. By definition, it includes the whole population, therefore, a population standard deviation would be used.

What are the formulas for the standard deviation?

The **sample standard deviation formula** is:

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}}$$

where,

s = sample standard deviation
$\sum$= sum of...
$\bar{X}$= sample mean
n = number of scores in sample.

The **population standard deviation** formula is:

$$\sigma = \sqrt{\frac{\sum(X - \mu)^2}{n}}$$

where,

σ= population standard deviation
$\sum$= sum of...
μ= population mean
n = number of scores in sample.

Is there an easy way to calculate the standard deviation?

Yes, we have a sample and population standard deviation calculator that shows you all the working as well! It can be found in our calculator section here.

Hypothesis Testing

55

When we conduct a piece of quantitative research we are inevitably attempting to answer a hypothesis that we have set. Since hypothesis testing has many facets, let's look at an example.

The lecturer's dilemma

Two statistics lecturers, Sarah and Mike, think that they use the best method to teach their students. Each lecturer has 50 statistics students that are studying a graduate degree in management. In Sarah's class, students have to attend one lecture and one seminar class every week, whilst in Mike's class students only have to attend one lecture. Sarah thinks that seminars, in addition to lectures, are an important teaching method in statistics, whilst Mike believes that lectures are sufficient by themselves and thinks that students are better off solving

problems by themselves in their own time. This is the first year that Sarah has given seminars, but since they take up a lot of her time, she wants to make sure that she is not wasting her time and that seminars improve her students' performance.

The structure of hypothesis testing

Whilst all pieces of quantitative research have some dilemma, issue or problem that they are trying to investigate, the focus in hypothesis testing is to find ways to structure these in such a way that we can test them effectively. Typically, it is important to:

1. Define the **research hypothesis** and set the **parameters** for the study.
2. Set out the **null** and **alternative hypothesis** (or more than one hypothesis; in other words, a number of hypotheses).
3. Explain how you are going to **operationalise** (that is, **measure** or **operationally define**) what you are studying and set out the **variables** to be studied.
4. Set the **significance level**.
5. Make a **one-** or **two-tailed prediction**.
6. Determine whether the distribution that you are studying is **normal** (this has implications for the types of statistical tests that you can run on your data).
7. **Select** an appropriate **statistical test** based on the variables you have defined and whether the distribution is normal or not.
8. **Run** the **statistical tests** on your data and **interpret** the **output**.
9. **Accept** or **reject** the **null hypothesis**.

Whilst there are some variations on this structure, it is adopted by most thorough quantitative research studies. Throughout this guide we discuss this hypothesis testing process using our example of the two statistics lecturers, Sarah and Mike, and their alternative teaching methods. We focus on the first fives steps in the process, as well as the decision to either accept or reject the null hypothesis. A discussion of normality, selecting statistical tests, and running these statistical tests is discussed in the statistical guide, Selecting Statistical Tests. At the end, we highlight some of the other statistical guides that we think you would benefit from reading before going on to design your own quantitative piece of research.

The research hypothesis

This study aims to examine the effect that two different teaching methods – providing lectures and seminar classes (Sarah) and providing lectures by themselves (Mike) – had on the performance of Sarah's 50 students and Mike's 50 students. By establishing that we are not only interested in these 100 students, we set the parameters for the study. This is important because if we were interested in the effect that these teaching methods had on students' performance in general, there would be wider sampling implications (see the statistical guide, Sampling, for more information).

Whilst Mike is sceptical about the effectiveness of seminars, Sarah clearly believes that giving seminars, in addition to lectures, helps her students do better than those in Mike's class. This leads us to the following research hypothesis:

Research Hypothesis: When students attend seminar classes, in addition to

lectures, their performance increases.

The null and alternative hypothesis

Whilst only providing a research hypothesis like the above is sometimes adequate, it is good practice in quantitative research to re-state this as a null and alternative hypothesis.

Null Hypotheses (H_o):	Undertaking seminar classes has no effect on students' performance.
Alternative Hypothesis (H_a):	Undertaking seminar class has a positive effect on students' performance.

The null hypothesis predicts that the distributions that we are comparing are the same. This brings us back to the core of hypothesis testing; **comparing distributions** (see the statistical guide, **Frequency Distributions**, for more information on distributions). In this study, there are two distributions that we are comparing.

Distribution 1 (seminar)	The distribution of exam marks for the 50 students in Sarah's class that attended both lectures and seminars.
Distribution 2 (lecture only)	The distribution of exam marks for the 50 students in Mike's class that attended only lectures.

If the two distributions are the same this would mean that the addition of seminars to lectures as a teaching method did not have an effect on students' performance and we would accept the null hypothesis. Alternatively, if there was a difference in the distributions and this difference was **statistically significant**, we would reject the null hypothesis. The question then arises: Do we accept the alternative hypothesis?

Before we answer this question and three related concepts (the alternative hypothesis, one- and two-tailed predictions, and statistical significance) it is worth addressing the issue of operationally defining our study.

Hypothesis Testing (cont...)

8

Operationally defining (measuring) the study

The question arises: What do we mean by performance? Clearly, performance could mean how students score in a piece of coursework, how many times they can answer questions in class, what marks they get in their exams, and so on. There are two obvious reasons why we should be clear about how we operationalise (that is, measure) what we are studying. First, we simply need to be clear so that people reading our work are in no doubt about what we are studying. This makes it easier for them to repeat the study in future to see if they also get the same (or similar) results; something called **internal validity**. Second, one of the criteria by which quantitative research is assessed, perhaps by an examiner if you are a student, is how you define what your are measuring (in this case, performance) and how you choose to measure it.

Nonetheless, it is worth noting that these choices will sometimes be personal choices and at other times they will be guided by something else. For example, if we were to measure intelligence, there may be a number of characteristics that we could use, such as IQ, emotional intelligence, and so forth. What we choose here will likely be a personal choice because all these variables are **proxies** for intelligence; that is, they are variables used to infer an individual's intelligence but not everyone would agree that IQ alone is an accurate measure of intelligence. In comparison, if we were measuring firm performance, there may be an established number of measures in the academic and practitioner literature that determine what we should test, such as Return on Assets, etc… Therefore, to know what you should measure, it is always worth looking at the literature first to see what other studies have done, whether you use the same measures or not. It is then a matter of making an educated decision whether the variables you choose to examine are accurate proxies for what you are trying to study, as well as discussing the potential limitations of these proxies.

In the case of measuring a student's performance there are a number of proxies that could be used, such as class participation, coursework marks and exam marks, since these are all good measures of performance. However, in this case, we choose exam marks as our measure of performance for two reasons. First, as a statistics tutor, we feel that Sarah's main job is to help her students get the best grade possible since this will affect her students' overall grades in their graduate management degree. Second, the assessment for the statistics course is a single 2 hour exam. Since there is no coursework and class participation is not assessed in this course, exam marks seem to be the most appropriate proxy for performance. However, it is worth noting that if the assessment for the statistics course was not only a 2 hour exam but also a piece of coursework, we would probably have chosen to measure both exam marks and coursework marks as proxies of performance.

Variables

The next step is to define the variables that we are using in our study (see the statistical guide, Types of Variable, for more information). Since the study aims to examine the effect that two different teaching methods – providing lectures and seminar classes (Sarah) and providing lectures by themselves (Mike) – had on the performance of Sarah's 50 students and Mike's 50 students, the variables being measured are:

Dependent variable:	Exam marks
Independent variable:	Teaching method ("seminar" vs. "lecture only")

By using a very straightforward example, we have only one dependent variable and one independent variable although studies can examine any number of dependent and independent variables.

One- and two-tailed predictions

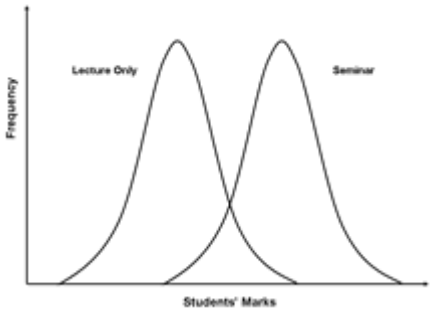
Returning to our discussion of the null and alternative hypothesis, it is worth re-stating that if the two distributions (the seminar distribution and the lecture only

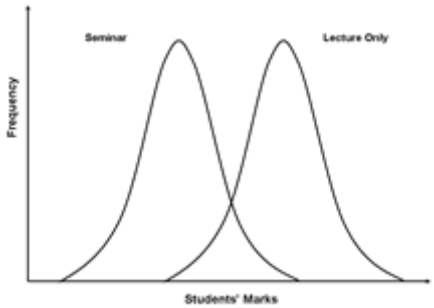
distribution) are the same, this would mean that the addition of seminars to lectures as a teaching method did not have an effect on students' performance and we would accept the null hypothesis. Alternatively, if there was a difference in the distributions and this difference was **statistically significant**, we would reject the null hypothesis. The question then arises: Do we accept the alternative hypothesis?

Alternative Hypothesis (H_a): Undertaking seminar class has a positive effect on students' performance.

The alternative hypothesis tells us two things. First, what **predictions** did we make about the effect of the independent variable(s) on the dependent variable(s)? Second, what was the predicted **direction** of this effect? Let's use our example to highlight these two points.

Sarah predicted that her teaching method (independent variable: teaching method), whereby she not only required her students to attend lectures (distribution: lecture only), but also seminars (distribution: seminar), had a positive effect (that is, increased) students' performance (dependent variable: exam marks). The direction of her prediction is what we call one-tailed, which can be illustrated using the diagrams below.





Hypothesis Testing (cont...)

7

A one-tailed prediction indicates that you believe that your distribution (the "seminar" distribution) is either higher up (diagram A) or lower down (diagram B) the scale (in this case, exam marks) compared with the alternative distribution (the "lectures only" distribution).

Alternately, a two-tailed prediction means that we do not make a choice over the direction that our distribution moves. Rather, it simply implies that our distribution could be either higher up or lower down the scale. If Sarah had made a two-tailed prediction, our alternative hypothesis might have been:

Alternative Hypothesis (H_a): Undertaking seminar class has an effect on students' performance.

In other words, we simply take out the word "positive", which implies the direction of our effect. In our example, making a two-tailed prediction may seem strange, which would be a fair comment. After all, it would be logical to expect that "extra" tuition (going to seminar classes as well as lectures) would either

have a positive effect on students' performance or no effect at all, but certainly not a negative effect. Therefore, the two distributions would either be the same or the seminar distribution would move higher up the scale. Nonetheless, there are cases when we do not know what the effect might be. For example, rather than using seminars in addition to teaching, Sarah could have used an untested, experimental teaching method. Since this teaching method was experimental, we may not be able to assume whether it would have a positive or negative effect on students' performance, or simply no effect at all.

Significance levels

Whilst we are close to being able to either accept or reject the null hypothesis and if we reject it, either accept or reject the alternative hypothesis, rigour requires that we first set the significance level for our study.

Statistical significance is about probability. It asks the question: What is the probability that a score could have arisen by chance? In terms of our two distributions, statistically analysing the differences in the distribution may, for example, suggest that there are no differences in the two distributions; hence we should accept the null hypothesis. However, how confident are we that there really are no differences between the distributions?

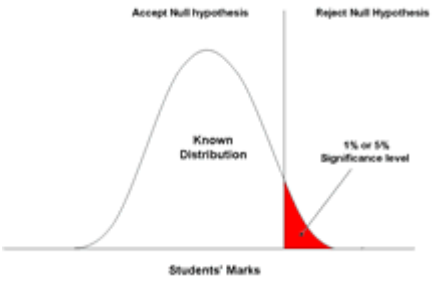
Typically, if there was a 5% or less chance (5 times in 100 or less) that a score from the focal distribution (the "seminar" distribution") could not have come from the comparison distribution (the "lectures only" distribution), we would accept the null hypothesis. Alternately, if the chance was greater than 5% (6 times in 100 or more) we would reject the null hypothesis. We do not reject the null hypothesis because our statistical analysis did not show that the two distributions were the same. We reject it because at a significance level of 0.05 (that is, 5% or less chance) we could not be confident enough that this result did not simply happen by chance.

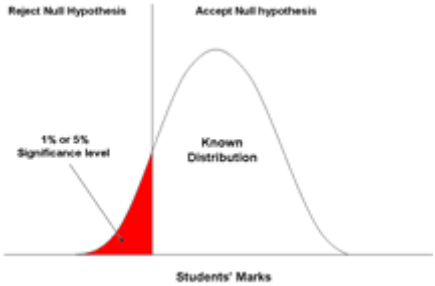
Whilst there is relatively little justification why a significance level of 0.05 is used rather than 0.04 or 0.10, for example, it is widely accepted in academic research. However, if we want to be particularly confident in our results, we set a more stringent level of 0.01 (a 1% chance or less; 1 in 100 chance or less).

Accepting or rejecting the null hypothesis

Therefore, let's return finally to the question of whether we (a) reject or accept the null hypothesis; and (b) if we reject the null hypothesis, do we accept the alternative hypothesis.

If our statistical analysis shows that the two distributions are the same at the significance level (either 0.05 or 0.01) that we have set, we simply accept the null hypothesis. Alternatively, if the two distributions are different, we need to either accept or reject the alternative hypothesis. This will depend on whether we made a one- or two-tailed prediction (see below).

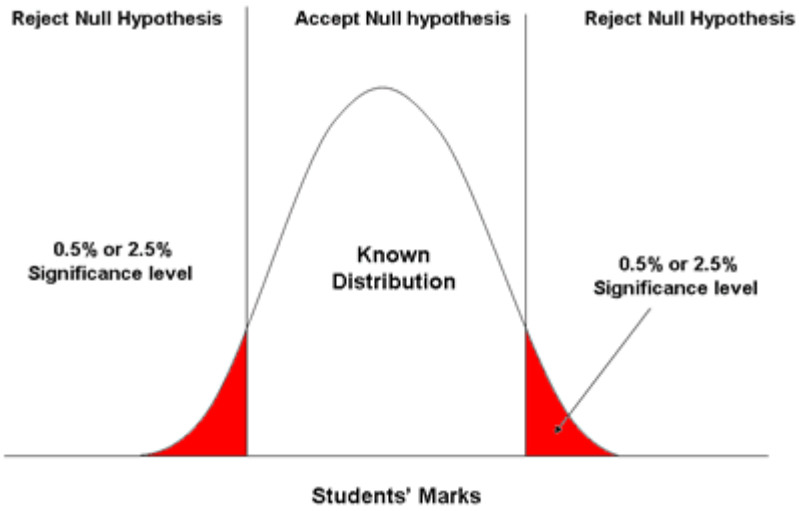




Hypothesis Testing (cont...)

[8](#)

If we (a) made a one-tailed prediction, (b) predicted that our distribution (the "seminar" distribution) moved in the correct direction (either up or down the scale), and (c) the result was statistically significant at the level we selected (either 0.05 or 0.01), then we accept the alternative hypothesis. If we were wrong about either the direction of our prediction or the result was not statistically significant at the selected level, then we reject the alternative hypothesis.



If we made a two-tailed prediction, this has a bearing on our significance level because we are unsure in which direction our distribution (the "seminar" distribution) will move (up or down). As a result, both ends of our distribution are relevant. Allowing a 5% or less chance that a score from our distribution (the "seminar" distribution") came from the comparison distribution (the "lectures only" distribution) would result in a 10% or less chance of making an error because we have to add together the 5% or less significance level from both ends of our distribution (see diagram A). Therefore, we can only allow at 2.5% chance or less (0.025) that a score from our distribution came from the comparison distribution (see diagram B), which added together would ensure that our overall chance of making an error remained at the 0.05 significance level.

On this basis, if we (a) made a two-tailed prediction, and (b) the result was statistically significant at the level we selected (either 0.025 or 0.005), then we accept the alternative hypothesis, that "undertaking seminar class has an effect

on students' performance", but we would also state the direction of the effect that was indicated by our results. In other words, we would state that undertaking seminar class has a "positive" effect on students' performance.

Moving forward

Other key aspects of hypothesis testing include a discussion of normality, selecting statistical tests, and running these statistical tests. These are discussed in the statistical guide, Selecting Statistical Tests. However, we would recommend that before you read this guide, you first read the guide on Sampling.

Types of Variable

151

All experiments examine some kind of variable(s). A variable is not only something that we measure, but also something that we can manipulate and something we can control for. To understand the characteristics of variables and how we use them in research, this guide is divided into three main sections. First, we illustrate the role of dependent and independent variables. Second, we discuss the difference between experimental and non-experimental research. Finally, we explain how variables can be characterised as either categorical or continuous.

Dependent and Independent Variables

An independent variable, sometimes called an experimental or predictor variable, is a variable that is being manipulated in an experiment in order to observe the effect on a dependent variable, sometimes called an outcome variable.

Imagine that a tutor asks 100 students to complete a maths test. The tutor wants to know why some students perform better than others. Whilst the tutor does not know the answer to this, she thinks that it might be because of two reasons: (1) some students spend more time revising for their test; and (2) some students are naturally more intelligent than others. As such, the tutor decides to investigate the effect of revision time and intelligence on the test performance of the 100 students. The dependent and independent variables for the study are:

Dependent Variable: **Test Mark** (measured from 0 to 100)

Independent Variables: **Revision time** (measured in hours) **Intelligence** (measured using IQ score)

The dependent variable is simply that, a variable that is dependent on an independent variable(s). For example, in our case the test mark that a student achieves is dependent on revision time and intelligence. Whilst revision time and intelligence (the independent variables) may (or may not) cause a change in the test mark (the dependent variable), the reverse is implausible; in other words, whilst the number of hours a student spends revising and the higher a student's IQ score may (or may not) change the test mark that a student achieves, a

change in a student's test mark has no bearing on whether a student revises more or is more intelligent (this simply doesn't make sense).

Therefore, the aim of the tutor's investigation is to examine whether these independent variables - revision time and IQ - result in a change in the dependent variable, the students' test scores. However, it is also worth noting that whilst this is the main aim of the experiment, the tutor may also be interested to know if the independent variables - revision time and IQ - are also connected in some way.

In the section on experimental and non-experimental research that follows, we find out a little more about the nature of independent and dependent variables.

Experimental and Non-Experimental Research

- Experimental research: In experimental research, the aim is to manipulate an independent variable(s) and then examine the effect that this change has on a dependent variable(s). Since it is possible to manipulate the independent variable(s), experimental research has the advantage of enabling a researcher to identify a cause and effect between variables. For example, take our example of 100 students completing a maths exam where the dependent variable was the exam mark (measured from 0 to 100) and the independent variables were revision time (measured in hours) and intelligence (measured using IQ score). Here, it would be possible to use an experimental design and manipulate the revision time of the students. The tutor could divide the students into two groups, each made up of 50 students. In "group one", the tutor could ask the students not to do any revision. Alternately, "group two" could be asked to do 20 hours of revision in the two weeks prior to the test. The tutor could then compare the marks that the students achieved.
- Non-experimental research: In non-experimental research, the researcher does not manipulate the independent variable(s). This is not to say that it is impossible to do so, but it will either be impractical or unethical to do so. For example, a researcher may be interested in the effect of illegal, recreational drug use (the dependent variable(s)) on certain types of behaviour (the independent variable(s)). However, whilst possible, it would be unethical to ask individuals to take illegal drugs in order to study what effect this had on certain behaviours. As such, a researcher could ask both drug and non-drug users to complete a questionnaire that had been constructed to indicate the extent to which they exhibited certain behaviours. Whilst it is not possible to identify the cause and effect between the variables, we can still examine the association or relationship between them.In addition to understanding the difference between dependent and independent variables, and experimental and non-experimental research, it is also important to understand the different characteristics amongst variables. This is discussed next.

Categorical and Continuous Variables

Categorical variables are also known as discrete or qualitative variables. Categorical variables can be further categorized as either nominal, ordinal or dichotomous.

- Nominal variables are variables that have two or more categories but which do not have an intrinsic order. For example, a real estate agent could classify their types of property into distinct categories such as houses, condos, co-ops or bungalows. So "type of property" is a nominal variable with 4 categories called houses, condos, co-ops and bungalows. Of note, the different categories of a nominal variable can also be referred to as groups or levels of the nominal variable. Another example of a nominal variable would be classifying where people live in the USA by state. In this case there will be many more levels of the nominal variable (50 in fact).
- Dichotomous variables are nominal variables which have only two categories or levels. For example, if we were looking at gender, we would most probably categorize somebody as either "male" or "female". This is an example of a dichotomous variable (and also a nominal variable). Another example might be if we asked a person if they owned a mobile phone. Here, we may categorise mobile phone ownership as either "Yes" or "No". In the real estate agent example, if type of property had been classified as either residential or commercial then "type of property" would be a dichotomous variable.
- Ordinal variables are variables that have two or more categories just like nominal variables only the categories can also be ordered or ranked. So if you asked someone if they liked the policies of the Democratic Party and they could answer either "Not very much", "They are OK" or "Yes, a lot" then you have an ordinal variable. Why? Because you have 3 categories, namely "Not very much", "They are OK" and "Yes, a lot" and you can rank them from the most positive (Yes, a lot), to the middle response (They are OK), to the least positive (Not very much). However, whilst we can rank the levels, we cannot place a "value" to them; we cannot say that "They are OK" is twice as positive as "Not very much" for example.

Continuous variables are also known as quantitative variables. Continuous variables can be further categorized as either interval or ratio variables.

- Interval variables are variables for which their central characteristic is that they can be measured along a continuum and they have a numerical value (for example, temperature measured in degrees Celsius or Fahrenheit). So the difference between 20C and 30C is the same as 30C to 40C. However, temperature measured in degrees Celsius or Fahrenheit is NOT a ratio variable.
- Ratio variables are interval variables but with the added condition that 0 (zero) of the measurement indicates that there is none of that variable. So, temperature measured in degrees Celsius or Fahrenheit is not a ratio

variable because 0C does not mean there is no temperature. However, temperature measured in Kelvin is a ratio variable as 0 Kelvin (often called absolute zero) indicates that there is no temperature whatsoever. Other examples of ratio variables include height, mass, distance and many more. The name "ratio" reflects the fact that you can use the ratio of measurements. So, for example, a distance of ten metres is twice the distance of 5 metres.

Ambiguities in classifying a type of variable

In some cases, the measurement scale for data is ordinal but the variable is treated as continuous. For example, a Likert scale that contains five values - strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree - is ordinal. However, where a Likert scale contains seven or more value - strongly agree, moderately agree, agree, neither agree nor disagree, disagree, moderately disagree, and strongly disagree - the underlying scale is sometimes treated as continuous although where you should do this is a cause of great dispute.

It is worth noting that how we categorise variables is somewhat of a choice. Whilst we categorised gender as a dichotomous variable (you are either male or female), social scientists may disagree with this, arguing that gender is a more complex variable involving more than two distinctions, but also including measurement levels like genderqueer, intersex, and transgender. At the same time, some researchers would argue that a Likert scale, even with seven values, should never be treated as a continuous variable.

How to do Normal Distributions Calculations

42

This guide will show you how to calculate the probability (area under the curve) of a standard normal distribution. It will first show you how to interpret a Standard Normal Distribution Table. It will then show you how to calculate the:

- probability less than a z-value
- probability greater than a z-value
- probability between z-values
- probability outside two z-values.

We have a calculator that calculates probabilities based on z-values for all the above situations. In addition, it also outputs all the working to get to the answer, so you know the logic of how to calculate the answer. The calculator can be found here.

How to Use the Standard Normal Distribution Table

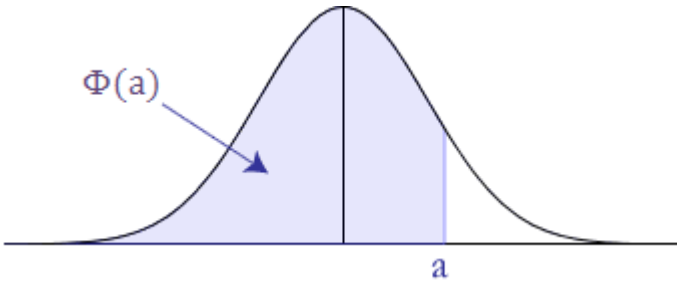
The most common form of standard normal distribution table that you see is a table similar to the one below (click image to enlarge):

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.

The Standard Normal Distribution Table

The standard normal distribution table provides the probability that a normally distributed random variable Z, with mean equal to 0 and variance equal to 1, is less than or equal to z. It does this for positive values of z only, i.e. z-values on the right-hand side of the mean. What this means in practice is that if someone asks you to find the probability of a value being less than a specific, positive z-value, you can simply look that value up in the table. We call this area Φ. Thus, for this table, $P(Z < a) = \Phi(a)$, where a is positive.

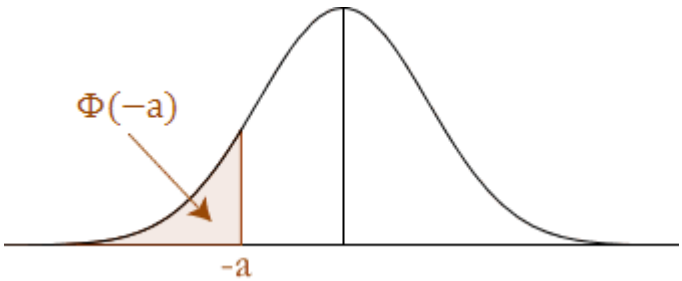
Diagrammatically, the probability of Z less than 'a' being Φ(a), as determined from the standard normal distribution table, is shown below:



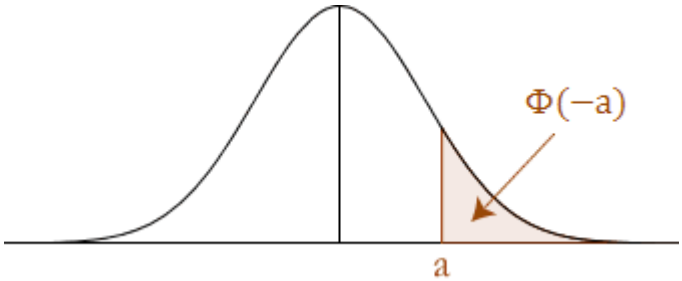
Probability less than a z-value

P(Z < –a)

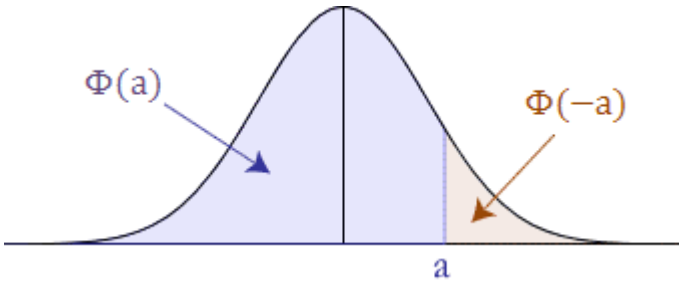
As explained above, the standard normal distribution table only provides the probability for values less than a positive z-value, i.e. z-values on the right-hand side of the mean. So how do we calculate the probability below a negative z-value (as illustrated below)?



We start by remembering that the standard normal distribution has a total area (probability) equal to 1 and it is also symmetrical about the mean. Thus, we can do the following to calculate negative z-values: we need to appreciate that the area under the curve covered by P(Z > a) is the same as the probability less than –a {P(Z < –a)} as illustrated below:



Making this connection is very important as, from the standard normal distribution table, we can calculate the probability less than 'a', as 'a' is now a positive value. Imposing P(Z < a) on the above graph is illustated below:



From the above illustration, and from our knowledge that the area under the standard normal distribution is equal to 1, we can conclude that the two areas add up to 1. We can, therefore, make the following statements:

$$\Phi(a) + \Phi(-a) = 1$$

$$\therefore\ \Phi(-a) = 1 - \Phi(a)$$

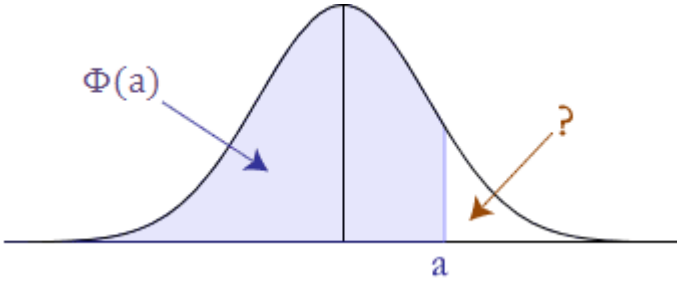
Thus, we know that to find a value less than a negative z-value we use the following equation:

$$\Phi(-a) = 1 - \Phi(a), \qquad \text{e.g. } \Phi(-1.43) = 1 - \Phi(1.43)$$

Probability greater than a z-value

<u>P(Z > a)</u>

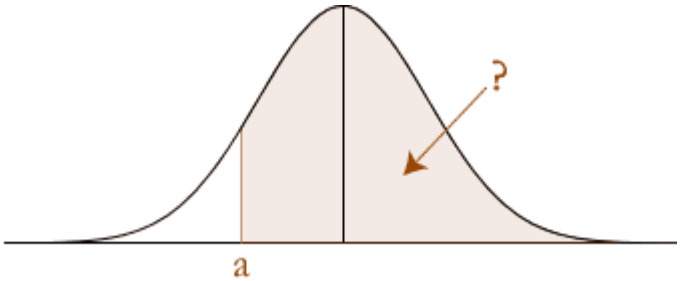
The probability of P(Z > a) is: 1 – Φ(a). To understand the reasoning behind this look at the illustration below:



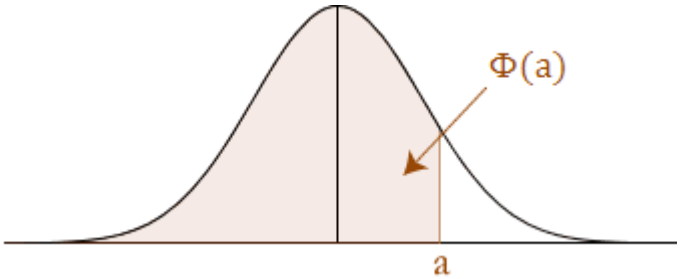
You know Φ(a) and you know that the total area under the standard normal curve is 1 so by mathematical deduction: P(Z > a) is: 1 - Φ(a).

P(Z > –a)

The probability of P(Z > –a) is P(a), which is Φ(a). To understand this we need to appreciate the symmetry of the standard normal distribution curve. We are trying to find out the area below:



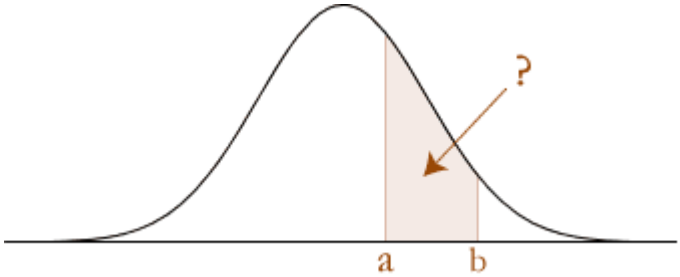
But by reflecting the area around the centre line (mean) we get the following:



Notice that this is the same size area as the area we are looking for, only we already know this area, as we can get it straight from the standard normal distribution table: it is P(Z < a). Therefore, the P(Z > –a) is P(Z < a), which is Φ(a).

Probability between z-values

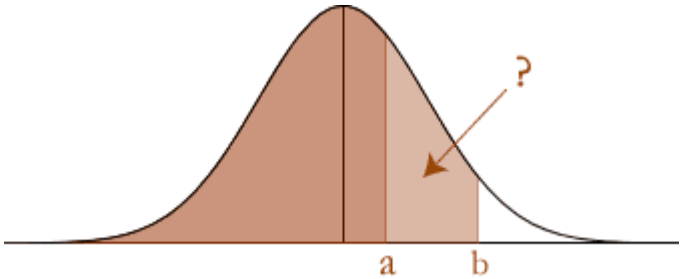
You are wanting to solve the following:



The key requirement to solve the probability between z-values is to understand that the probability between z-values is the difference between the probability of the greatest z-value and the lowest z-value:

$P(a < Z < b) = P(Z < b) - P(Z < a)$

which is illustrated below:



<u>P(a < Z < b)</u>

The probability of $P(a < Z < b)$ is calculated as follows.

First separate the terms as the difference between z-scores:

$P(a < Z < b) = P(Z < b) - P(Z < a)$ (explained in the section above)

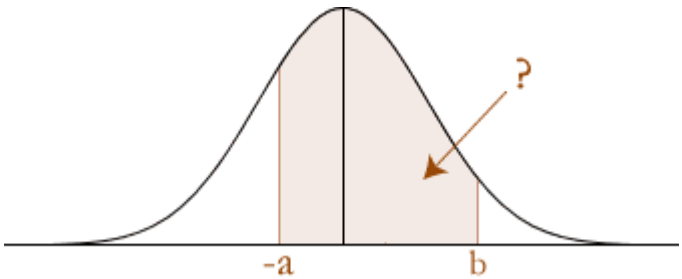
Then express these as their respective probabilities under the standard normal distribution curve:

$$P(Z < b) - P(Z < a) = \Phi(b) - \Phi(a).$$

Therefore, $P(a < Z < b) = \Phi(b) - \Phi(a)$, where a and b are positive.

<u>P(−a < Z < b)</u>

The probability of $P(-a < Z < b)$ is illustrated below:



First separate the terms as the difference between z-scores:

$P(-a < Z < b) = P(Z < b) - P(Z < -a)$

Then express these as their respective probabilities under the standard normal distribution curve:

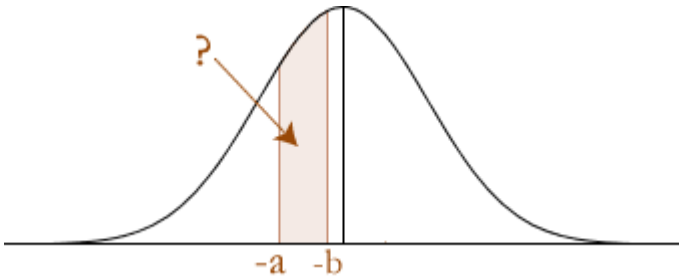
$P(Z < b) - P(Z < -a) = \Phi(b) - \Phi(-a)$

$= \Phi(b) - \{1 - \Phi(a)\}P(Z < -a)$ explained above.

$\therefore\ P(-a < Z < b) = \Phi(b) - \{1 - \Phi(a)\}$, where a is negative and b is positive.

P(−a < Z < −b)

The probability of P(−a < Z < −b) is illustrated below:



First separate the terms as the difference between z-scores:

P(−a < Z < −b) = P(Z < −b) − P(Z < −a)

Then express these as their respective probabilities under the standard normal distribution curve:

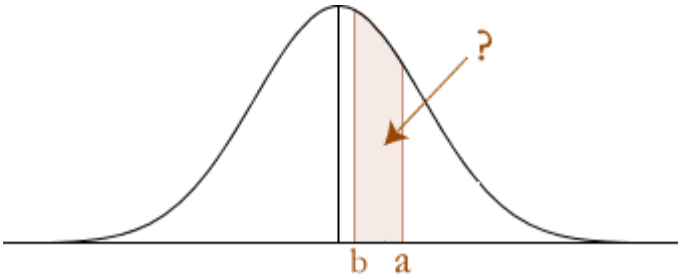
$$P(Z < b) - P(Z < -a) = \Phi(-b) - \Phi(-a)$$

$$= \{1 - \Phi(b)\} - \{1 - \Phi(a)\}$$ P(Z < −a) explained above.

$$= 1 - \Phi(b) - 1 + \Phi(a)$$

$$= \Phi(a) - \Phi(b)$$

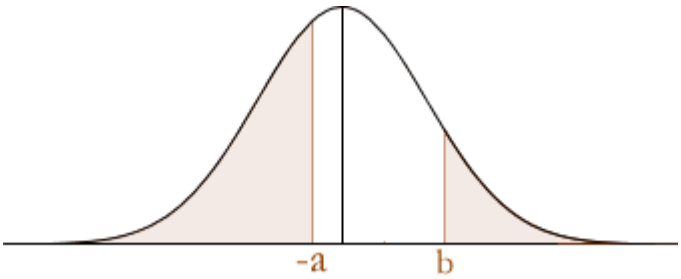
The above calculations can also be seen clearly in the diagram below:



Notice that the reflection results in a and b "swapping positions".

Probability outside of a range of z-values

An illustration of this type of problem is found below:



To solve these types of problems you simply need to work out each separate area under the standard normal distribution curve and then add the probabilities together. This will give you the total probability.

When a is negative and b is positive (as above) the total probability is:

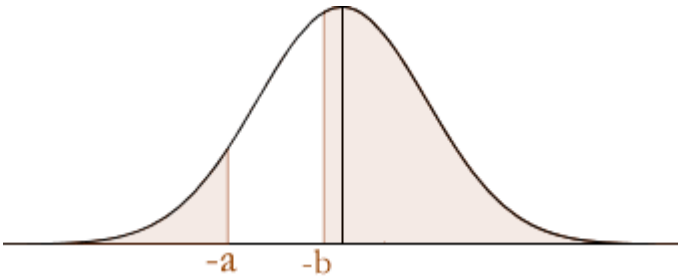
$P(Z < -a) + P(Z > b) = \Phi(-a) + \{1 - \Phi(b)\}\ P(Z > b)$ explained above.

$= \{1 - \Phi(a)\} + \{1 - \Phi(b)\}\ P(Z < -a)$ explained above.

$= 1 - \Phi(a) + 1 - \Phi(b)$

$= 2 - \Phi(a) - \Phi(b)$

When a and b are negative as illustrated below:



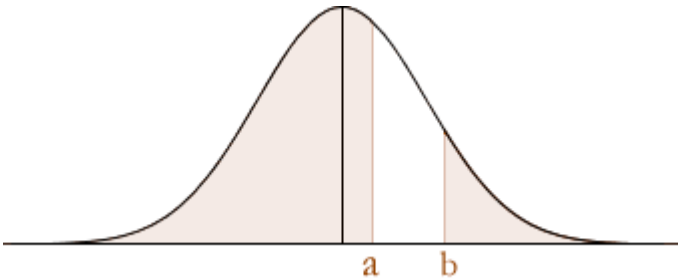
The total probability is:

$P(Z < -a) + P(Z > -b) = \Phi(-a) + \Phi(b)P(Z > -b)$ explained above.

$= \{1 - \Phi(a)\} + \Phi(b)P(Z < -a)$ explained above.

$= 1 + \Phi(b) - \Phi(a)$

When a and b are positive as illustrated below:



The total probability is:

$P(Z < a) + P(Z > b) = \Phi(a) + \{1 - \Phi(b)\}P(Z > b)$ explained above.

$= 1 + \Phi(a) - \Phi(b)$

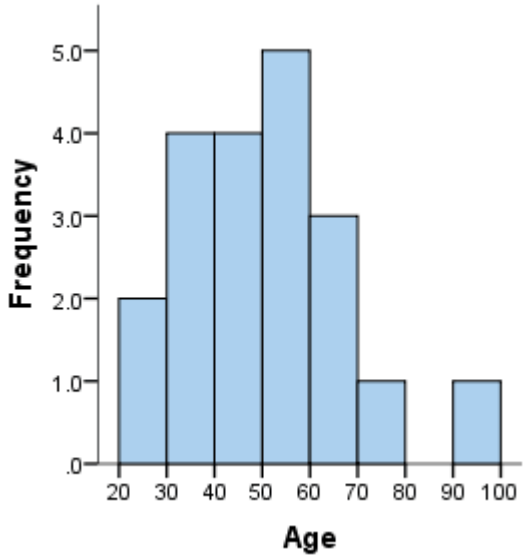
Check out our calculator here to get some practice in!

Histograms

9

What is a histogram?

A histogram is a plot that lets you discover, and show, the underlying frequency distribution (shape) of a set of continuous data. This allows the inspection of the data for its underlying distribution (e.g. normal distribution), outliers, skewness, etc. An example of a histogram, and the raw data it was constructed from, is shown below:



36	25	38	46	55	68	72	55	36	38
67	45	22	48	91	46	52	61	58	55

How do you construct a histogram from a continuous variable?

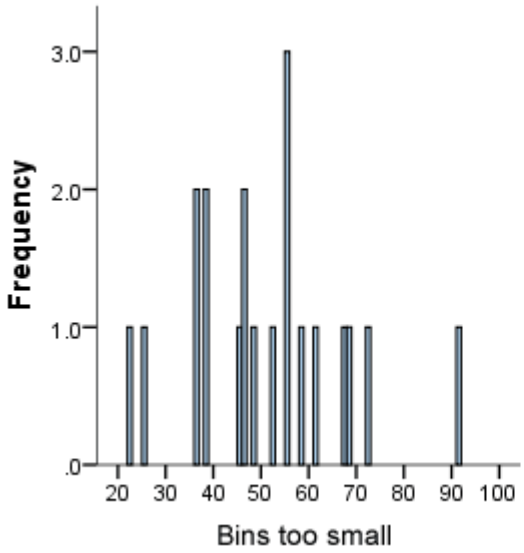
To construct a histogram from a continuous variable you first need to split the data into intervals, called **bins**. In the example above, **age** has been split into bins, with each bin representing a 10-year period starting at 20 years. Each bin contains the number of occurrences of scores in the data set that are contained within that bin. For the above data set, the frequencies in each bin have been tabulated along with the scores that contributed to the frequency in each bin (see below):

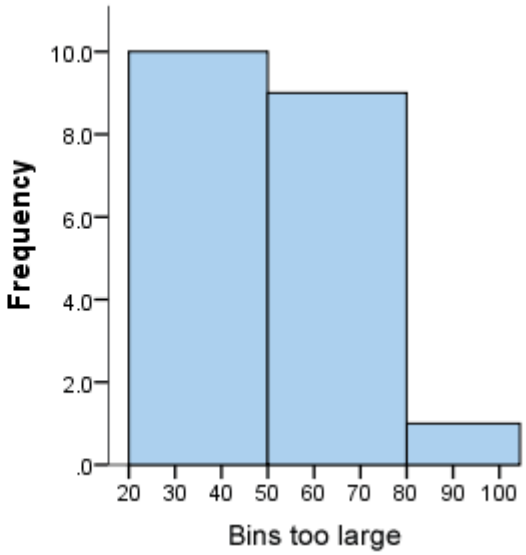
Bin	Frequency	Scores Included in Bin
20-30	2	25,22
30-40	4	36,38,36,38
40-50	4	46,45,48,46
50-60	5	55,55,52,58,55
60-70	3	68,67,61
70-80	1	72
80-90	0	-
90-100	1	91

Notice that, unlike a bar chart, there are no "gaps" between the bars (although some bars might be "absent" reflecting no frequencies). This is because a histogram represents a continuous data set, and as such, there are no gaps in the data. (Although you will have to decide whether you round up or round down scores on the boundaries of bins)

Choosing the correct bin width

There is no right or wrong answer as to how wide a bin should be, but there are rules of thumb. You need to make sure that the bins are not too small or too large. Consider the histogram we produced earlier (see above): the following histograms use the same data but have either much smaller or larger bins, as shown below:





We can see from the histogram on the left, that the bin width is too small as it shows too much individual data and does not allow the underlying pattern (frequency distribution) of the data to be easily seen. At the other end of the scale, is the diagram on the right, where the bins are too large and, again, we are unable to find the underlying trend in the data.

Histograms are based on area not height of bars

In a histogram, it is the area of the bar that indicates the frequency of occurrences for each bin. This means that the height of the bar does not necessarily indicate how many occurrences of scores there were within each individual bin. It is the product of height multiplied by the width of the bin that indicates the frequency of occurrences within that bin. One of the reasons that the height of the bars is often incorrectly assessed as indicating frequency and not the area of the bar is due to the fact that a lot of histograms often have equally spaced bars (bins) and, under these circumstances, the height of the bin does reflect the frequency.

What is the difference between a bar chart and a histogram?

The major difference is that a histogram is only used to plot the frequency of score occurrences in a continuous data set that has been divided into classes,

called bins. Bar charts, on the other hand, can be used for a great deal of other types of variables including ordinal and nominal data sets.

Standard Score

59

The standard score (more commonly referred to as a z-score) is a very useful statistic because (a) it allows us to calculate the probability of a score occurring within our normal distribution and (b) it enables us to compare two scores that are from different normal distributions. The standard score does this by converting (in other words, standardizing) scores in a normal distribution to z-scores in what becomes a standard normal distribution. To explain what this means in simple terms, let's use an example (if needed, see our statistical guide, Normal Distribution Calculations, for background information on normal distribution calculations).

Setting the scene: Part 1

A tutor sets a piece of English Literature coursework for the 50 students in his class. We make the assumption that when the scores are presented on a histogram the data is found to be normally distributed. The mean score is 60 out of 100 and the standard deviation (in other words, the variation in the scores) is 15 marks (see our statistical guides, Measures of Central Tendency and Standard Deviation, for more information about the mean and standard deviation).

Having looked at the performance of the tutor's class, one student, Sarah, has asked the tutor if, by scoring 70 out of 100, she has done well. Bearing in mind that the mean score was 60 out of 100 and that Sarah scored 70, then at first sight it may appear that since Sarah has scored 10 marks above the 'average' mark, she has achieved one of the best marks. However, this does not take into consideration the variation in scores amongst the 50 students (in other words, the standard deviation). After all, if the standard deviation is 15, then there is a reasonable amount of variation amongst the scores when compared with the mean.

Whilst Sarah has still scored much higher than the mean score, she has not necessarily achieved one of the best marks in her class. The question arises: How well did Sarah perform in her English Literature coursework compared to the other 50 students? Before answering this question, let us look at another problem.

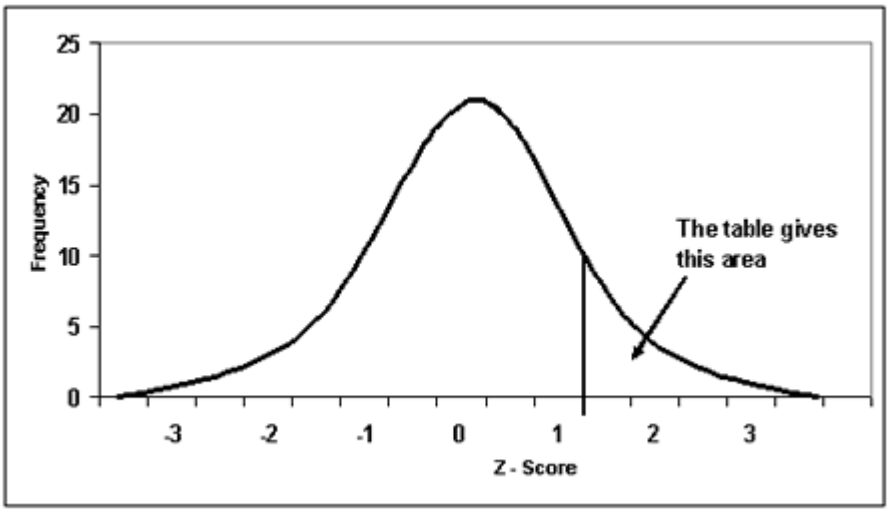
The tutor has a dilemma. In the next academic year, he must choose which of his students have performed well enough to be entered into an advanced English Literature class. He decides to use the coursework scores as an indicator of the performance of his students. As such, he feels that only those students that are in the top 10% of the class should be entered into the advanced English Literature class. The question arises: Which students came in the top 10% of the class?

Therefore, we are left with two questions to answer. First, how well did Sarah perform in her English Literature coursework compared to the other 50 students? Second, which students came in the top 10% of the class?

Whilst it is possible to calculate the answer to both of these questions using the existing mean score and standard deviation, this is very complex. Therefore, statisticians have come up with **probability distributions**, which are ways of calculating the probability of a score occurring for a number of common distributions, such as the normal distribution. In our case, we make the assumption that the students' scores are normally distributed. As such, we can use something called the **standard normal distribution** and its related **z-scores** to answer these questions much more easily.

Standard Normal Distribution and Standard Score (z-score)

When a frequency distribution is normally distributed we can find out the probability of a score occurring by standardising the scores, known as standard scores (or z scores). The standard normal distribution simply converts the group of data in our frequency distribution such that the mean is 0 and the standard deviation is 1 (see below).



Standard Score (cont...)

[14](#)

Z-scores are expressed in terms of standard deviations from their means. Resultantly, these z-scores have a distribution with a mean of 0 and a standard deviation of 1. The formula for calculating the standard score is given below.

Calculating the Standard Score (Z-Score)

$$\text{Standard Score, } z = \frac{X - \mu}{\sigma}$$

TERMS:
μ = mean (pronounced 'mu')
X = score
σ = standard deviation (pronounced 'sigma')

As the formula shows, the standard score is simply the score, minus the mean score, divided by the standard deviation. Therefore, let's return to our two questions.

1. How well did Sarah perform in her English Literature coursework compared to the other 50 students?

To answer this question, we can re-phrase it as: What percentage (or number) of students scored higher than Sarah and what percentage (or number) of students scored lower than Sarah? First, let's reiterate that Sarah scored 70 out of 100, the mean score was 60, and the standard deviation was 15 (see below).

	Score (X)	Mean μ	Standard Deviation s
English Literature	70	60	15

In terms of z-scores, this gives us:

$$\text{Standard Score, } z = \frac{X - \mu}{\sigma} = \frac{70 - 60}{15} = \frac{10}{15} = \mathbf{0.6667}$$

The z-score is 0.67 (to 2 decimal places), but now we need to work out the percentage (or number) of students that scored higher and lower than Sarah. To do this, we need to refer to the standard normal distribution table.

This table helps us to identify the probability that a score is greater or less than our z-score score. To use the table, which is easier than it might look at first sight, we start with our z-score, **0.67** (if our z-score had more than two decimal places, for example, ours was 0.6667, we would round it up or down accordingly; hence, 0.6667 would become 0.67). The y-axis in the table highlights the first two digits of our z-score and the x-axis the second decimal place. Therefore, we start with the y-axis, finding 0.6, and then move along the x-axis until we find 0.07, before finally reading off the appropriate number, in this case, **0.2514**. This means that the probability of a score being greater than 0.67 is 0.2514. If we look at this as a percentage, we simply times the score by 100; hence 0.2514 x 100 = 25.14%. In other words, around 25% of the class got a better mark than Sarah (roughly 25 students since there is no such thing as part of a student!).

Going back to our question, "How well did Sarah perform in her English Literature coursework compared to the other 50 students?", then clearly we can see that Sarah did better than a large proportion of students, with **74.86%** of the class scoring lower than her (100% - 25.14% = 74.86%). We can also see how well she performed relative to the mean score by subtracting her score from the mean (0.5 - 0.2514 = 0.2486), Hence, **24.86%** of the scores (0.2486 x 100 = 24.86%) were lower than Sarah's but above the mean score. However, the key finding is that Sarah's score was not one of the best marks. It wasn't even in the top 10% of scores in the class, even though at first sight we may have expected it to be. This leads us onto the second question.

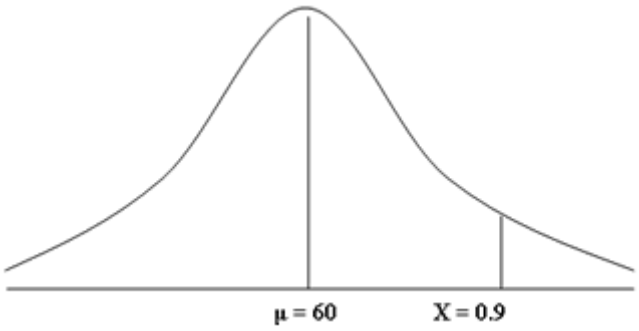
2. Which students came in the top 10% of the class?

A better way of phrasing this would be to ask: What mark would a student have to achieve to be in the top 10% of the class and qualify for the advanced English Literature class?

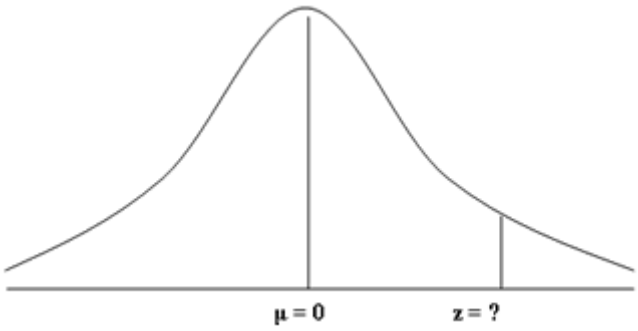
Standard Score (cont...)

6

To answer this question, we need to find the mark (which we call "X") on our frequency distribution that reflects the top 10% of marks. Since the mean score was 60 out of 100, we immediately know that the mark will be greater than 60. After all, if we refer to our frequency distribution below, we are interested in the area to the right of the mean score of 60 that reflects the top 10% of marks (shaded in red). As a decimal, the top 10% of marks would be those marks above 0.9 (that is, 100% - 90% = 10% or 1 - 0.9 = 0.1).



First, we should convert our frequency distribution into a standard normal distribution as discussed in the opening paragraphs of this guide. As such, our mean score of 60 becomes 0 and the score (X) we are looking for, 0.9, becomes our z-score, which is currently unknown. Note the changes to the labelling of the x-axis.



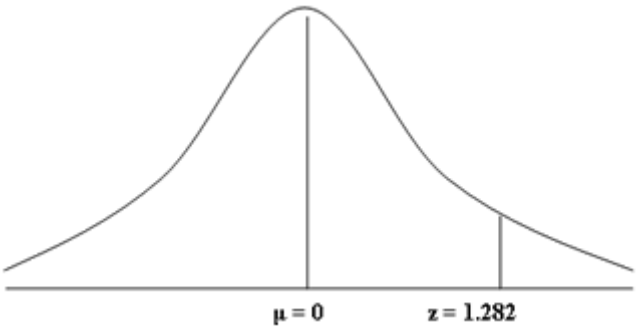
The next step involves finding out the value for our z-score. To do this, we refer back to the standard normal distribution table.

In answering the first question in this guide, we already knew the z-score, 0.67, which we used to find the appropriate percentage (or number) of students that scored higher than Sarah, 0.2514 (that is, 25.14% or roughly 25 students achieve a higher mark than Sarah). Using the z-score, 0.67, and the y-axis and

x-axis of the standard normal distribution table, this guided us to the appropriate value, 0.2514. In this case, we need to do the exact reverse to find our z-score.

We know the percentage we are trying to find, the top 10% of students, corresponds to 0.9. As such, we first need to find the value 0.9 in standard normal distribution table. When looking at the table, you may notice that the closest value to 0.9 is 0.8997. If we take the 0.8997 value as our starting point and then follow this row across to the left, we are presented with the first part of the z-score. You will notice that the value on the y-axis for 0.8997 is 1.2. We now need to do the same for the x-axis, using the 0.8997 value as our starting point and following the column up. This time, the value on the x-axis for 0.8997 is 0.08. This forms the second part of the z-score. Putting these two values together, the z-score for 0.8997 is 1.28 (that is, 1.2 + 0.08 = 1.28).

There is only one problem with this z-score; that is, it is based on a value of 0.8997 rather than the 0.9 value we are interested in. This is one of the difficulties of refer to the standard normal distribution table because it cannot give every possible z-score value (that we require a quite enormous table!). Therefore, you can either take the closest two values, 0.8997 and 0.9015, to your desired value, 0.9, which reflect the z-scores of 1.28 and 1.29, and then calculate the exact value of "z" for 0.9, or you can use a z-score calculator. If we use a z-score calculator, our value of 0.9 corresponds with a z-score of 1.282. In other words, P (z > 1.282) = 0.1.



Now that we have the key information (that is, the mean score, μ, the standard deviation, s , and z-score, z) we can answer our question directly, namely: What mark would a student have to achieve to be in the top 10% of the class and qualify for the advanced English Literature class? First, let us reiterate the facts:

Score	Mean	Standard Deviation	z-score
(X)	μ	s	z
?	60	15	1.282

To find out the relevant score, we apply the following formula:

$$(1) \quad P\left(Z \geq \frac{X - 60}{15}\right) = 0.10$$

$$(2) \quad P(z > 1.282) = 0.1$$

$$(3) \quad \frac{X - 60}{15} = 1.282$$

$$(4) \quad X = 1.282 \times 15 + 60$$

$$(5) \quad \mathbf{X = 79.23}$$

Therefore, students that scored above 79.23 marks out of 100 came in the top 10% of the English Literature class, qualifying for the advanced English Literature class as a result.

Setting the Scene: Part II

Clearly, the z-score statistic is helpful in highlighting how Sarah performed in her English Literature coursework and what mark a student would have to achieve to be in the top 10% of the class and qualify for the advanced English Literature class. However, we have only been talking about one distribution here, namely the distribution of scores amongst 50 students that completed a piece of English Literature coursework. What if Sarah wanted to compare how well she performed in her Maths coursework compared with her English Literature coursework?

In this case, Sarah achieved a higher mark in her Maths coursework, 72 out of 100. However, as we have already learnt, just because her Maths score (72) is higher than her English Literature score (70), we shouldn't assume that she performed better in her Maths coursework compared to her English Literature coursework. The question therefore arises: How well did Sarah perform in her Maths coursework compared to her English Literature coursework?

Standard Score (cont...)

17

Clearly, the two scores (her English Literature and Maths coursework marks) come from different distributions. The distribution of 50 students that completed the English Literature coursework has a mean of 60 and standard deviation of 15. The distribution of 50 students that completed the Maths coursework, on the other hand, has a mean of 68 and a standard deviation of 6. This gives us the following:

	Score (X)	Mean μ	Standard Deviation s
English Literature	70	60	15
Maths	72	68	6

Since these scores are from two different distributions, we need to standardise them into z-scores so that they can be directly compared. This gives us:

English Literature $z = \frac{X - \mu}{\sigma} = \frac{70 - 60}{15} = \mathbf{0.67}$

Maths $z = \frac{X - \mu}{\sigma} = \frac{72 - 68}{6} = \mathbf{0.67}$

The z-scores highlight that the student is two thirds (z = 0.67) of a standard deviation above the mean in English Literature, but also two thirds (z = 0.67) of a standard deviation above the mean in Maths. Using the standard normal distribution table, we can see that Sarah clearly performed above 'average' in both subjects although again, around 25% of the class got a better mark than her. However, the key point her is that the standard score showed that Sarah performed equally well in her English Literature and Maths coursework, even though her marks were different in both pieces. This shows the usefulness of the standard score statistic.

In the following statistical guide, Hypothesis Testing, we start to set out some of the key aspects of quantitative research.

Statistical Tests

Independent T-Test for Two Samples

42

Introduction

The independent t-test, also called the two sample t-test or student's t-test is an inferential statistical test that determines whether there is a statistically significant difference between the means in two unrelated groups.

Hypothesis for the independent t-test

The null hypothesis for the independent t-test is that the population means from the two unrelated groups are equal:

H_0: $u_1 = u_2$

In most cases, we are looking to see if we can show that we can reject the null hypothesis and accept the alternative hypothesis, which is that the population means are not equal:

H_A: $u_1 \neq u_2$

To do this we need to set a significance level (alpha) that allows us to either reject or accept the alternative hypothesis. Most commonly, this value is set at 0.05.

What do you need to run an independent t-test?

In order to run an independent t-test you need the following:

- One independent, categorical variable that has two levels.

- One dependent variable

Unrelated groups

Unrelated groups, also called unpaired groups or independent groups, are groups in which the cases in each group are different. Often we are investigating differences in individuals, which means that when comparing two groups, an individual in one group cannot also be a member of the other group and vice versa. An example would be gender - an individual would have to be classified as either male or female - not both.

Assumption of normality of the dependent variable

The independent t-test requires that the dependent variable is approximately normally distributed within each group. We can test for this using a multitude of tests but the Shapiro-Wilks Test or a graphical method, such as a Q-Q Plot, are very common. You can run these tests using SPSS, the procedure for which can be found in our Testing for Normality guide. However, the t-test is described as a robust test with respect to the assumption of normality. This means that even deviations away from normality do not have a large influence on Type I error rates. The exception to this is if the difference in the size of the groups is greater than 1.5 (largest compared to smallest).

What to do when you violate the normality assumption

If you find that either one or both of your group's data is not approximately normally distributed and groups sizes differ greatly then you have two options: (1) transform your data so that the data becomes normally distributed (to do this in SPSS see our guide on Transforming Data), or (2) run the Mann-Whitney U Test which is a non-parametric test that does not require the assumption of normality (to run this test in SPSS see our guide on the Mann-Whitney U Test).

Assumption of Homogeneity of Variance

The independent t-test assumes the variances of the two groups you are measuring to be equal. If your variances are unequal then this can affect the Type I error rate. The assumption of homogeneity of variance can be tested using Levene's Test of Equality of Variances, which is produced in SPSS when running the independent t-test. If you have run Levene's Test of Equality of Variances, in SPSS or by another means, then you will get a result similar to that below:

Levene's Test for Equality of Variances	F	.314	
	Sig.	.579	

This test for homogeneity of variance provides an *F* statistic and a significance value (*P*-value). We are primarily concerned with the significance level - if it is greater than 0.05 then our group variances can be treated as equal. However, if $P < 0.05$, then we have unequal variances and we have violated the assumption of homogeneity of variance.

Overcoming a Violation of the Assumption of Homogeneity of Variance

If the Levene's Test for Equality of Variances is statistically significant and, therefore, indicates unequal variances, we can correct for this violation by not using the pooled estimate for the error term for the t-statistic and also making adjustments to the degrees of freedom using the Welch-Satterthwaite method. In all reality, you will probably never have heard of these adjustments as SPSS hides this information and simply labels the two options as "Equal variances assumed" and "Equal variances not assumed" without explicitly stating the underlying tests used. However, you can see the evidence of these tests as below:

		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	23.636	
	Sig.	.000	
t-test for Equality of Means	t	2.516	2.233
	df	16	7.001
	Sig. (2-tailed)	.023	.061

From the result of Levene's Test for Equality of Variances we can reject the null hypothesis that there is no difference in the variances between the groups and accept the alternative hypothesis that there is a significant difference in the variances between groups. The effect of not being able to assume equal variances is evident in the final column of the above figure where we see a reduction in the value of the *t*-statistic and a large reduction in the degrees of freedom (df). This has the effect of increasing the *P*-value above the critical significance level of 0.05. In this case, we therefore do not accept the alternative hypothesis and accept that there are no statistically significant differences between means. This would not have been our conclusion had we not tested for homogeneity of variances.

Reporting the Result of an Independent T-Test

When reporting the result of an independent t-test, you need to include the *t*-statistic value, the degrees of freedom (df) and the significance value of the test (*P*-value). The format of the test result is: $t(\text{df}) = t\text{-statistic}$, P = significance value. For the example above you could, therefore, report the result as $t(7.001) = 2.233$, $P = 0.061$.

Fully Reporting Your Results

In order to provide enough information for readers to fully understand the results when you have run an independent t-test you should include the result of normality tests, Levene's Equality of Variances test, the two group means and standard deviations, the actual t-test result and the direction of the difference (if any). In addition, you might also wish to include the difference between the groups along with the 95% confidence intervals. For example:

Inspection of Q-Q Plots revealed that cholesterol concentration was normally distributed for both groups and that there was homogeneity of variance as assessed by Levene's Test for Equality of Variances. Therefore, an independent t-

test was run on the data as well as 95% confidence intervals (CI) for the mean difference. It was found that after the two interventions, cholesterol concentrations in the dietary group (6.15 ± 0.52 mmol/L) were significantly higher than the exercise group (5.80 ± 0.38 mmol/L) ($t(38) = 2.470$, $P = 0.018$) with a difference of 0.35 (95% CI, 0.06 to 0.64) mmol/L.

Dependent T-Test for Paired Samples

17

What does this test do?

The dependent t-test (also called the paired t-test or paired-samples t-test) compares the means of two related groups to detect whether there are any statistically significant differences between these means.

If you wish to learn how to calculate the dependent t-test then we have a dependent t-test calculator that also generates all the working involved in getting to the answer. The calculator can be found here.

What variables do you need for a dependent t-test?

You need one dependent variable that is measured on an interval or ratio scale (see our Types of Variable guide if you need clarification). You also need one categorical variable that has only two related groups.

What is meant by "related groups"?

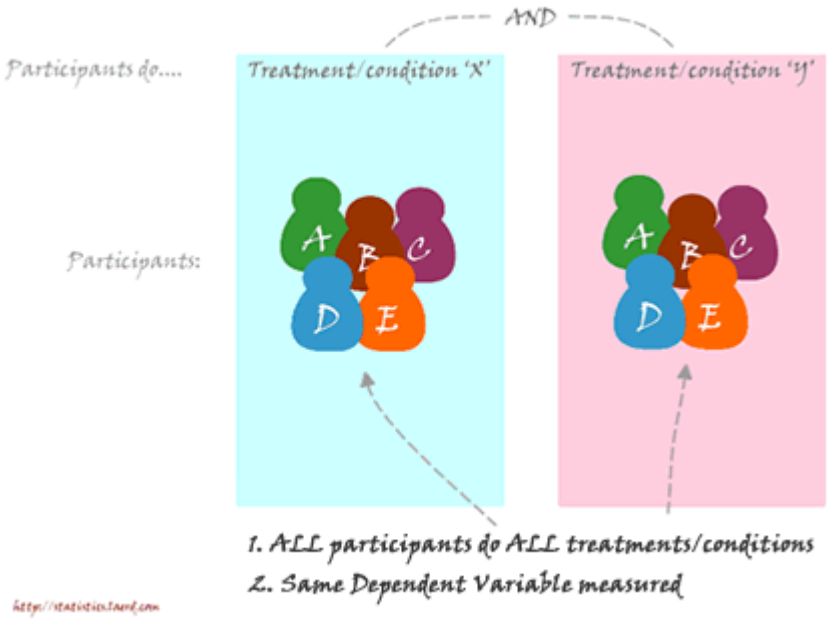
A dependent t-test is an example of a "within-subjects" or "repeated-measures" statistical test. This indicates that the same subjects are tested more than once. Thus, in the dependent t-test, "related groups" indicates that the same subjects are present in both groups. The reason that it is possible to have the same subjects in each group is because each subject has been measured on two occasions on the same dependent variable. For example, you might have measured 10 individuals' (subjects') performance in a spelling test (the dependent variable) before and after they underwent a new form of computerised teaching method to improve spelling. You would like to know if the computer training improved their spelling performance. Here, we can use a dependent t-test as we have two related groups. The first related group consists of the subjects at the beginning (prior to) the computerised spell training and the second related group consists of the same subjects but now at the end of the computerised training.

Does the dependent t-test test for "changes" or "differences" between related groups?

The dependent t-test can be used to test either a "change" or a "difference" in means between two related groups but not both at the same time. Whether you are measuring a "change" or "difference" between the means of the two related groups depends on your study design. The two types of study design are indicated in the following diagrams.

How do you detect differences between experimental conditions using the dependent t-test?

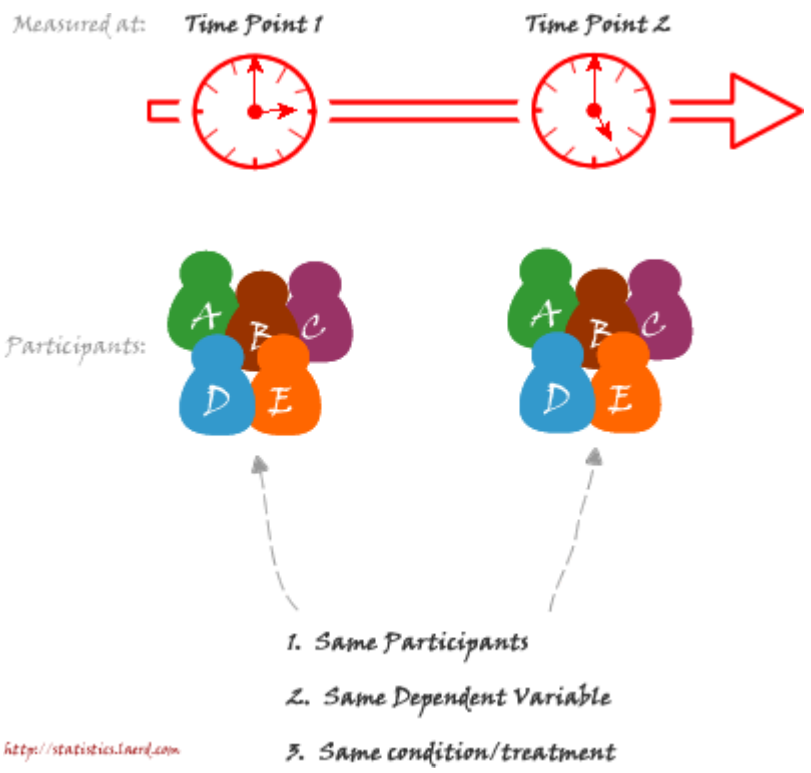
The dependent t-test can look for "differences" between means when subjects are measured on the same dependent variable under two different conditions. For example, you might have tested subjects' eyesight (dependent variable) when wearing two different types of spectacle (independent variable). See the diagram below for a general schematic of this design approach (click the image to enlarge):



Dependent T-Test for Paired Samples (cont...) 2

How do you detect changes in time using the dependent t-test?

The dependent t-test can also look for "changes" between means when the subjects are measured on the same dependent variable but at two time points. A common use of this is in a pre-post study design. In this type of experiment we measure subjects at the beginning and at the end of some intervention, e.g. an exercise-training programme or business-skills course. A general schematic is provided below (click image to enlarge):



How else might you use the dependent t-test?

You can also use the dependent t-test to study more complex study designs although it is not normally recommended. The most common, more complex study design where you might use the dependent t-test is where you have a crossover design with two different interventions that are both performed by the same subjects. One example of this design is where you have one of the interventions act as a control. For example, you might want to investigate whether a course of diet counselling can help people lose weight. To study this you could simply measure subjects' weight before and after the diet counselling course for any changes in weight using a dependent t-test. However, to improve the study design you also include want to include a control trial. During this control trial, the subjects could either receive "normal" counselling or do nothing at all or something else you deem appropriate. In order to assess this study using a dependent t-test you would use the same subjects for the control trial as the diet counselling trial. You then measure the differences between the interventions at the end, and only at the end, of the two interventions. Remember, however, that this is unlikely to be the preferred statistical analysis for this study design.

What are the assumptions of the dependent t-test?

The types of variable needed for the dependent t-test have already been discussed earlier in this guide. In addition, the distribution of the differences between the scores of the two related groups needs to be normally distributed. We do this by simply subtracting each individuals' score in one group from their score in the other related group and then testing for normality in the normal way (see our guide on how to test for normality in SPSS here). It is important to note that the two related groups do not need to be normally distributed themselves - just the differences between the groups.

Go to the last page of our guide here.

Dependent T-Test for Paired Samples (cont...)

3

What hypothesis is being tested?

The dependent t-test is testing the null hypothesis that there are no differences between the means of the two related groups. If we get a significant result then we can reject the null hypothesis that there are no significant differences between the means and accept the alternative hypothesis that there are statistically significant differences between the means. We can express this as follows:

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

What is the advantage of a dependent t-test over an independent t-test?

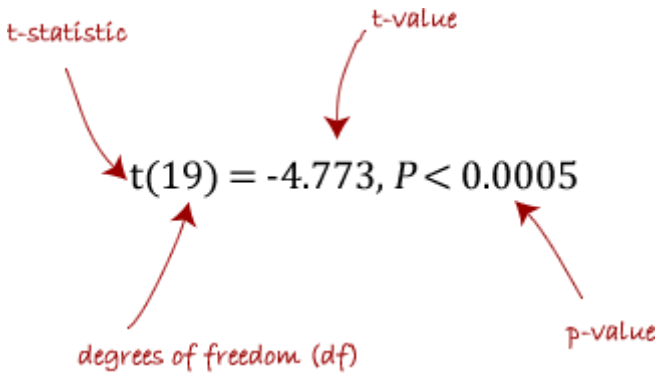
Before we answer this question, we need to point out that you cannot choose one test over the other unless your study design allows it. What we are discussing here is whether it is advantageous to design a study that uses one set of subjects whom are measured twice or two separate groups of subjects measured once each. The major advantage of choosing a repeated-measures design (and therefore running a dependent t-test) is that you get to eliminate the individual differences that occur between subjects - the concept that no two people are the same - and this increases the power of the test. What this means is that you are more likely to detect any significant differences, if they do exist, using the dependent t-test versus the independent t-test.

Can the dependent t-test be used to compare different subjects?

Yes, but this does not happen very often. You can use the dependent t-test instead of using the usual independent t-test when each subject in one of the independent groups is closely related to another subject in the other group on many individual characteristics. This approach is called a "matched-pairs" design. The reason we might want to do this is that the major advantage of running a within-subject (repeated-measures) design is that you get to eliminate between-groups variation from the equation (each individual is unique and will react slightly differently than someone else), thereby increasing the power of the test. Hence, the reason why we use the same subjects - we expect them to react in the same way as they are, after all, the same person. The most obvious case of when a "matched-pairs" design might be implemented is when using identical twins. Effectively you are choosing parameters to match your subjects on which you believe will result in each pair of subjects reacting in a similar way.

How do I report the result of a dependent t-test?

You need to report the test as follows:



where df is N - 1, where N = number of subjects.

Should I report confidence levels?

Confidence intervals (CI) are a useful statistic to include as they indicate the direction and size of a result. It is common to report 95% confidence intervals, which you will most often see reported as 95% CI. Programmes such as SPSS will automatically calculate these confidence intervals for you otherwise you need to calculate them by hand. You will want to report the mean and 95% confidence levels for the differences between the two related groups.

If you wish to run a dependent t-test in SPSS then you can find out how to do this in our guide here.

One-way ANOVA

35

What is this test for?

The one-way analysis of variance (ANOVA) is used to determine whether there are any significant differences between the means of three or more independent (unrelated) groups. This guide will provide a brief introduction to the one-way ANOVA including the assumptions of the test and when you should use this test. If you are familiar with the one-way ANOVA then you can skip this guide and go straight to how to run this test in SPSS by clicking here.

What does this test do?

The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are significantly different from each other. Specifically, it tests the null hypothesis:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

where μ = group mean and k = number of groups. If, however, the one-way ANOVA returns a significant result then we accept the alternative hypothesis (H_A), which is that there are at least 2 group means that are significantly different from each other.

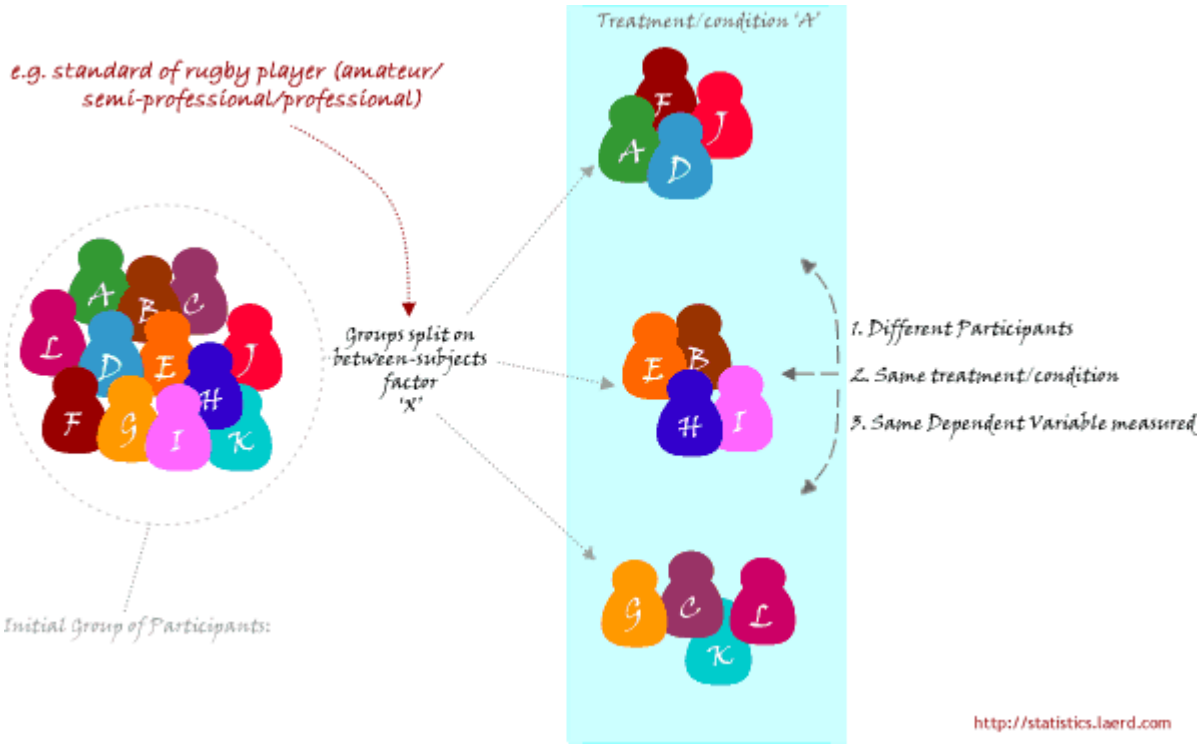
At this point, it is important to realise that the one-way ANOVA is an *omnibus* test statistic and cannot tell you which specific groups were significantly different from each other, only that at least two groups were. To determine which specific

groups differed from each other you need to use a *post-hoc* test. *Post-hoc* tests are described later in this guide.

When might you need to use this test?

If you are dealing with individuals, you are likely to encounter this situation using two different types of study design:

One study design is to recruit a group of individuals and then randomly split this group into 3 or more smaller groups, i.e. each subject is allocated to one, and only one, group. You then get each group to undertake different tasks (or put them under different conditions) and measure the outcome/response on the same dependent variable. For example, a researcher wishes to know whether different pacing strategies affect the time to complete a marathon. The researcher randomly assigns a group of volunteers to either a group that (a) starts slow and then increases their speed, (b) starts fast and slows down or (c) runs at a steady pace throughout. The time to complete the marathon is the outcome (dependent) variable. This study design is illustrated schematically in the Figure below:



When you might use this test is continued on the next page.

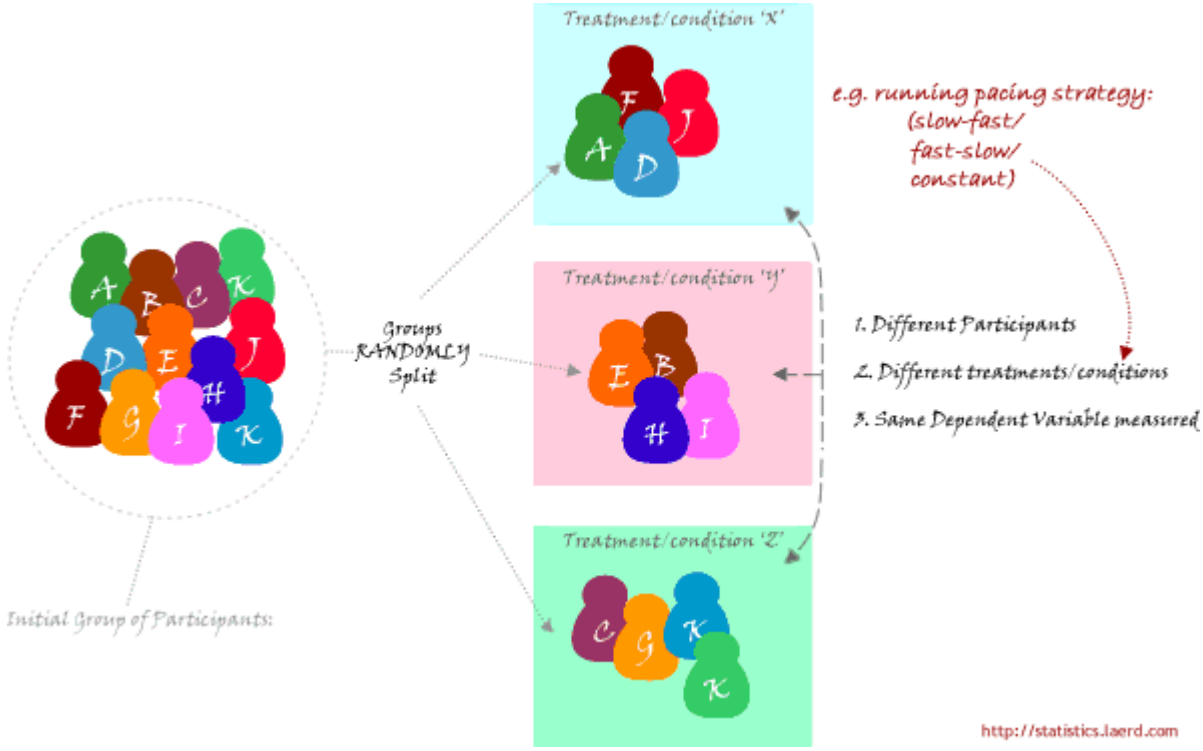
One-way ANOVA (cont...)

6

When might you need to use this test? (cont...)

A second study design is to recruit a group of individuals and then split them into groups based on some independent variable. Again, each individual will be

assigned to one group only. This independent variable is sometimes called an attribute independent variable because you are splitting the group based on some attribute that they possess, e.g. their level of education; every individual has a level of education, even if it is "none". Each group is then measured on the same dependent variable having undergone the same task or condition (or none at all). For example, a researcher is interested in determining whether there are differences in leg strength between amateur, semi-professional and professional rugby players. The force/strength measured on an isokinetic machine is the dependent variable. This type of study design is illustrated schematically in the Figure below:



Why not compare groups with multiple t-tests?

Every time you conduct a t-test there is a chance that you will make a Type 1 error. An ANOVA controls for these errors so that the Type 1 error remains at 5% and you can be more confident that any significant result you find is not just down to chance. See our guide on hypothesis testing for more information on Type I errors.

What assumptions does the test make?

There are three main assumptions, listed here:

1. The dependent variable is normally distributed in each group that is being compared in the one-way ANOVA. So, for example, if we were comparing three groups; amateur, semi-professional and professional rugby players; on their leg strength, then their leg strength values (dependent variable) would have to be normally distributed for the amateur group of players, normally distributed for the semi-professionals and normally distributed for

the professional players. You can test for normality in SPSS (see our guide here).

2. There is homogeneity of variances. This means that the population variances in each group are equal. If you use SPSS, Levene's Test for Homogeneity of Variances is included in the output when you run a one-way ANOVA in SPSS (see our One-way ANOVA using SPSS guide).
3. This is a study design issue that you will so you will need to examine your study design to determine whether this could have occurred.

One-way ANOVA (cont...)

5

What happens if my data fail these assumptions?

Firstly, don't panic! The first two of these assumptions are easily fixable, even if the last assumption is not. Lets go through the options as above:

1. The one-way ANOVA is considered a robust test against the normality assumption. This means that it tolerates violations to its normality assumption rather well. As regards the normality of group data, the one-way ANOVA can tolerate data that is non-normal (skewed or kurtotic distributions) with only a small effect on the Type I error rate. However, platykurtosis can have a profound effect when your group sizes are small. This leaves you with two options: (1) transform your data using various algorithms so that the shape of your distributions become normally distributed (see our normality guide here) or (2) choose the non-parametric Kruskal-Wallis H Test which does not require the assumption of normality (read our guide on this test here).
2. There are two tests that you can run that are applicable when the assumption of homogeneity of variances has been violated: (1) Welch or (2) Brown and Forsythe test. Alternatively, you could run a Kruskal-Wallis H Test. For most situations it has been shown that the Welsh test is best. Both the Welch and Brown and Forsythe tests are available in SPSS (see our One-way ANOVA using SPSS guide).
3. A lack of independence of cases has been stated as the most important assumptions to fail. Often, there is little you can do that offers a good solution to this problem. A full explanation of this problem and all assumptions mentioned here, including numerical explanations, are provided in Intermediate Statistics: A Modern Approach by Dr James Stevens.

How do I run a one-way ANOVA?

There are numerous ways to run a one-way ANOVA, however, we provide a comprehensive, step-by-step guide on how to do this using SPSS.

How do I report the results of a one-way ANOVA?

You will have calculated the following results or obtained them from SPSS:

Structure of results:

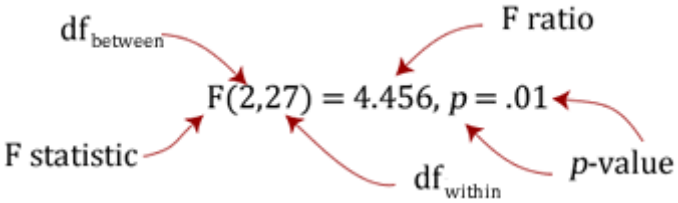
Source	*SS*	*df*	*MS*	*F*	*Sig.*
Between	SS_b	k-1	MS_b	MS_b/MS_w	*p* value
Within	SS_w	N-k	MS_w		
Total	$SS_b + SS_w$	N-1			

An example:

Source	*SS*	*df*	*MS*	*F*	*Sig.*
Between	91.476	2	45.733	4.467	.021
Within	276.400	27	10.237		
Total	367.867	29			

You will want to report this as follows:

There was a statistically significant difference between groups as determined by one-way ANOVA ($F(2,27) = 4.467$, $P = .021$). This is all you will need to write for the one-way ANOVA per se. However, in reality you will want probably also want to report means ± SD for your groups as well as follow-up a significant result with *post-hoc* tests. If you use SPSS then these descriptive statistics will be reported in the output along with the result from the one-way ANOVA. The general form of writing the result of a one-way ANOVA is as follows:



where df = degrees of freedom.

It is very important that you do not report the result as "significant difference" but that you report it as "statistically significant difference". This is because your decision as to whether the result is significant or not should not be based solely on your statistical test. Therefore, to indicate to readers that this "significance" is a statistical one, include this is your sentence.

Find out what else you have to do when you have a significant or a not-significant ANOVA result on the next page.

One-way ANOVA (cont...)

12

My p-value is greater than 0.05, what do I do now?

Report the result of the one-way ANOVA, e.g. "there were no statistically significant differences between group means as determined by one-way ANOVA

($F(2,27) = 1.397$, $P = .15$)". Not achieving a statistically significant result does not mean you should not report group means +/- SD also. However, running post-hoc tests is not warranted and should not be carried out.

My p-value is less than 0.05, what do I do now?

Firstly, you need to report your results as highlighted in the "How do I report the results?" section above. You then need to follow-up the one-way ANOVA by running *post-hoc* tests.

Homogeneity of variances was violated; how to continue?

You need to perform the same procedures as in the above three sections but add into your results section that this assumption was violated and you needed to run a Welch F test.

What are post-hoc tests?

Recall from earlier that the ANOVA test tells you whether you have an overall difference between your groups but it does not tell you which specific groups differed - post-hoc tests do. Because *post-hoc* tests are run to confirm where the differences occurred between groups, they should, therefore, only be run when you have a shown an overall significant difference in group means (i.e. a significant one-way ANOVA result). Post-hoc tests attempt to control the experimentwise error rate usually alpha = 0.05) in the same manner that the one-way ANOVA is used instead of multiple t-tests. *Post-hoc* tests are termed a posteriori tests - that is, performed after the event (the event in this case being a study).

Which *post-hoc* test should I use?

There are a great number of different *post-hoc* tests that you can use, however, you should only run one *post-hoc* test - do not run multiple *post-hoc* tests. For a one-way ANOVA, you will probably find that just one of four tests need to be considered. If your data meet the assumption of homogeneity of variances then either use the Tukey's honestly significant difference (HSD) or Scheffé *post-hoc* tests. Often, Tukey's HSD test is recommended by statisticians as it is not as conservative as the Scheffe test (which means that you are more likely to detect differences if they exist with Tukey's HSD test). Note that if you use SPSS, Tukey's HSD test is simply referred to as "Tukey" in the post-hoc multiple comparisons dialogue box). If your data did not meet the homogeneity of variances assumption then you should consider running either the Games Howell or Dunnett's C *post-hoc* test. The Games Howell test is generally recommended.

How should I graphically present my results?

First off, it is not essential that you present your results in a graphical form. However, it can add a lot of clarity to your results. There are a few key points to producing a good graph. Firstly, you need to present error bars for each group mean. It is customary to use the standard deviation of each group but standard error and confidence limits are also used in the literature. You should also make sure that the scale is appropriate for what you are measuring. These points and

more are discussed in our guide on selecting an appropriate graph (guide here). Generally, if graphically presenting data from an ANOVA, we recommend using a bar chart with standard deviation bars.

What to do now?

Now that you understand the one-way ANOVA, go to our guide on how to run the test in SPSS here.

Repeated Measures ANOVA

50

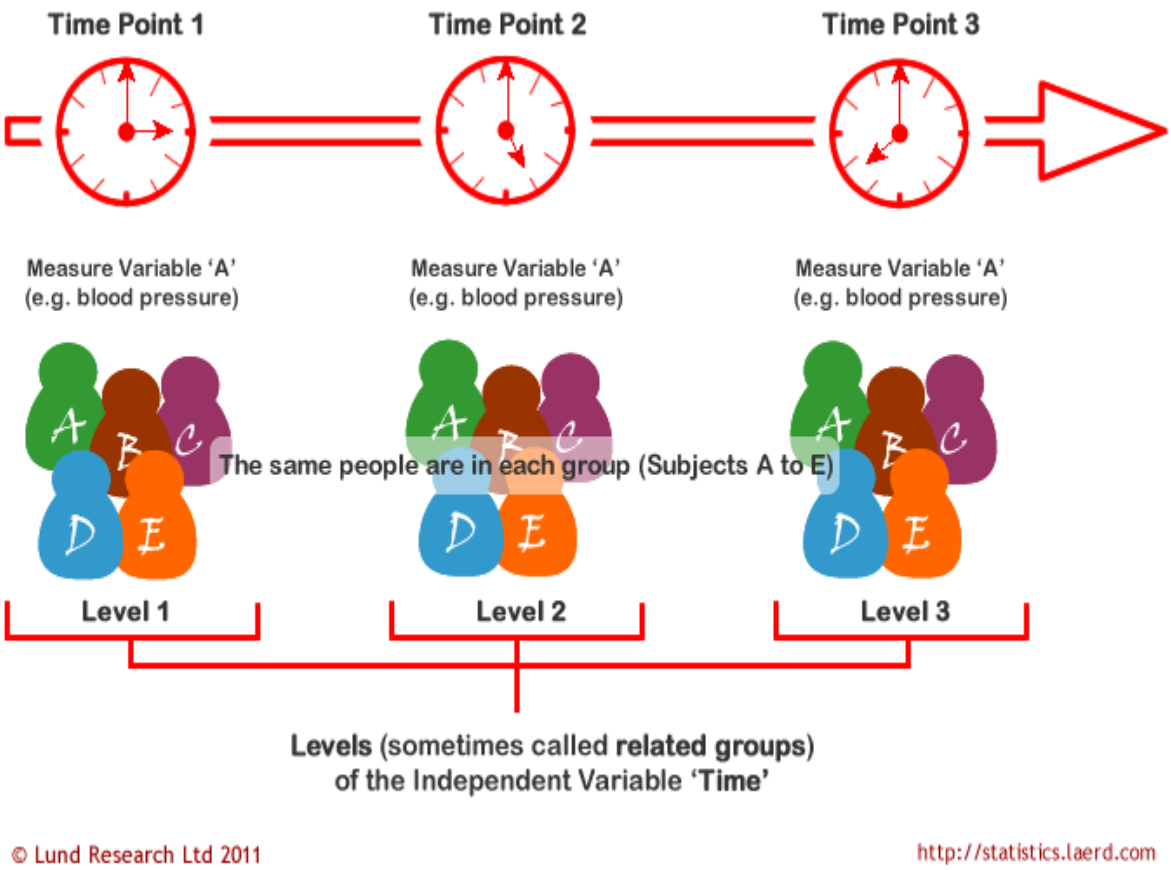
Introduction

Repeated measures ANOVA is the equivalent of the one-way ANOVA but for related not independent groups and is the extension of the dependent t-test. A Repeated Measures ANOVA is also referred to as a within-subjects ANOVA or ANOVA for correlated samples. All these names imply the nature of the Repeated Measures ANOVA, that of a test to detect any overall differences between related means. There are many complex designs that can make use of repeated measures but, throughout this guide, we will be referring to the most simple case, that of a one-way Repeated Measures ANOVA. This particular test requires one independent variable and one dependent variable. The dependent variable needs to be continuous (interval or ratio) and the independent variable categorical (either nominal or ordinal).

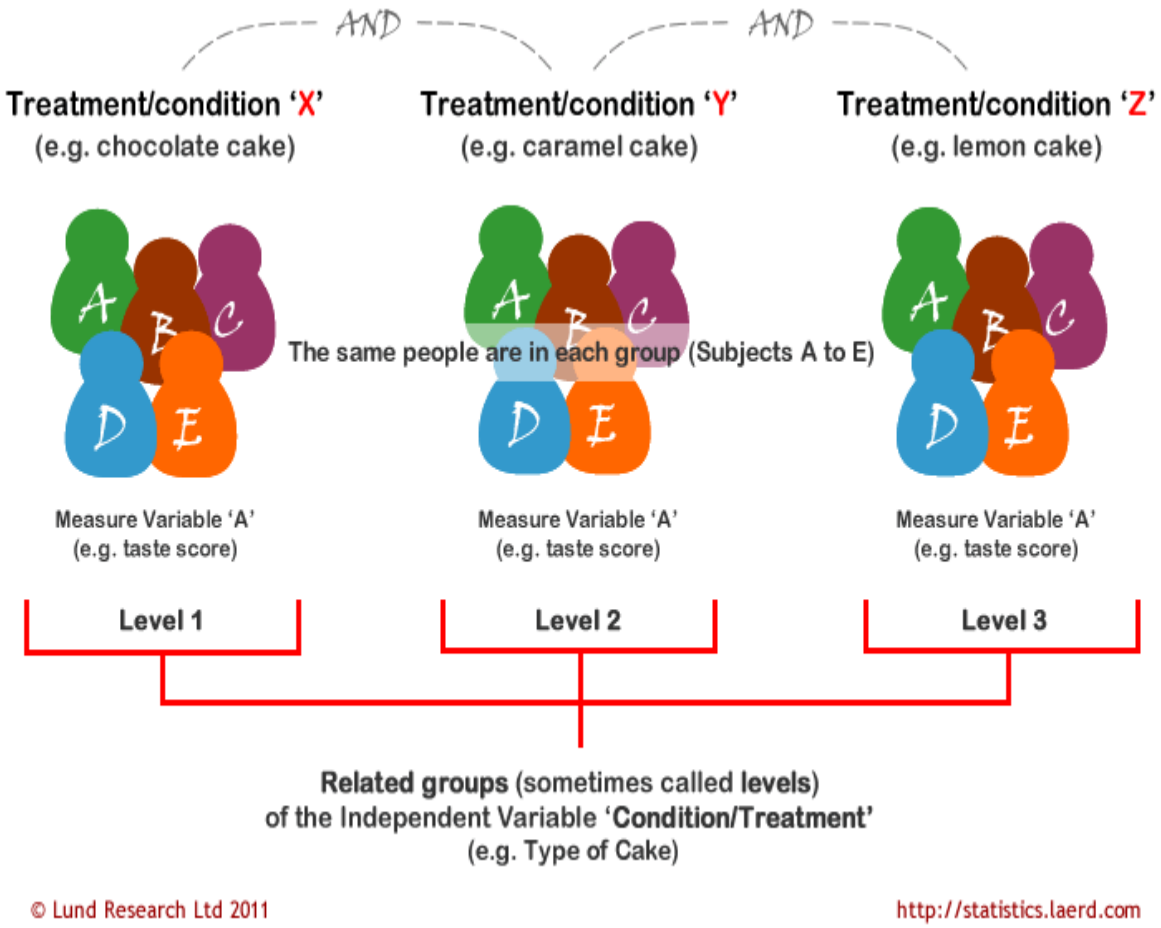
When to use a Repeated Measures ANOVA

We can analysis data using a Repeated Measures ANOVA for two types of study design. Studies that investigate either (1) changes in mean scores over three or more time points, or (2) differences in mean scores under three or more different conditions. For example, for (1), you might be investigating the effect of a 6-month exercise training programme on blood pressure and want to measure blood pressure at 3 separate time points: pre-, midway and post-exercise intervention, which would allow you to develop a time-course for any exercise effect. For (2), you might get the same subjects to eat different types of cake (chocolate, caramel, and lemon) and rate each one for taste, rather than having different people flavour each different cake. The important point with these two study designs is that the same people are being measured more than once on the same dependent variable, hence, why it is called repeated measures.

In Repeated Measures ANOVA, the independent variable has categories called **levels** or **related groups**. Where measurements are repeated over time, such as when measuring changes in blood pressure due to an exercise-training programme, the independent variable is **time**. Each **level** (or **related group**) is a specific time point. Hence, for the exercise-training study, there would be three time points and each time-point is a level of the independent variable (a schematic of a time-course repeated measures design is shown below):



Where measurements are made under different conditions, the conditions are the levels (or related groups) of the independent variable, e.g. type of cake is the independent variable with chocolate, caramel, and lemon cake as the levels of the independent variable (a schematic of a different-conditions repeated measures design is shown below). It should be noted that often the levels of the independent variable are not referred to as conditions but **treatments**. Which one you want to use is up to you, there is no right or wrong naming convention. You will also see the independent variable more commonly referred to as the **within-subjects factor**.



The above two schematics have shown an example of each type of Repeated Measures ANOVA design, but you will also often see these designs expressed in tabular form, such as shown below:

Subjects	Time/Condition		
	T_1	T_2	T_3
S_1	S_1	S_1	S_1
S_2	S_2	S_2	S_2
S_3	S_3	S_3	S_3
S_4	S_4	S_4	S_4
S_5	S_5	S_5	S_5
S_6	S_6	S_6	S_6

This particular table describes a study with six subjects (S_1 to S_6) performing under three conditions or at three time points (T_1 to T_3). As highlighted earlier, the within-subjects factor could also have been labelled "treatment" instead of "time/condition". They all relate to the same thing: subjects undergoing repeated measurements at either different time points or under different conditions/treatments.

Hypothesis for Repeated Measures ANOVA

The Repeated Measures ANOVA tests for whether there are any differences between related population means. The null hypothesis (H_0) states that the means are equal:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$$

where μ = population mean and k = number of related groups. The alternative hypothesis (HA) states that the related population means are not equal (at least one mean is different to another mean):

$$H_A: \textit{at least two means are significantly different}$$

For our exercise-training example, the null hypothesis (H0) is that mean blood pressure is the same at all time points (pre-, 3 months, and 6 months). The alternative hypothesis is that mean blood pressure is significantly different at one or more time points. A Repeated Measures ANOVA will not inform you of where the differences between groups lie as it is an omnibus statistical test. The same would be true if you were investigating different conditions or treatments rather than time points, as used in this example. If your Repeated Measures ANOVA is statistically significant then you can run *post-hoc* tests that can highlight exactly where these differences occur. How to run appropriate *post-hoc* tests for a Repeated Measures ANOVA is SPSS can be found (here).

Logic of the Repeated Measures ANOVA

The logic behind a Repeated Measures ANOVA is very similar to that of a between-subjects ANOVA. Recall that a between-subjects ANOVA partitions total variability into between-groups variability (SS_b) and within-groups variability (SS_w), as shown below:

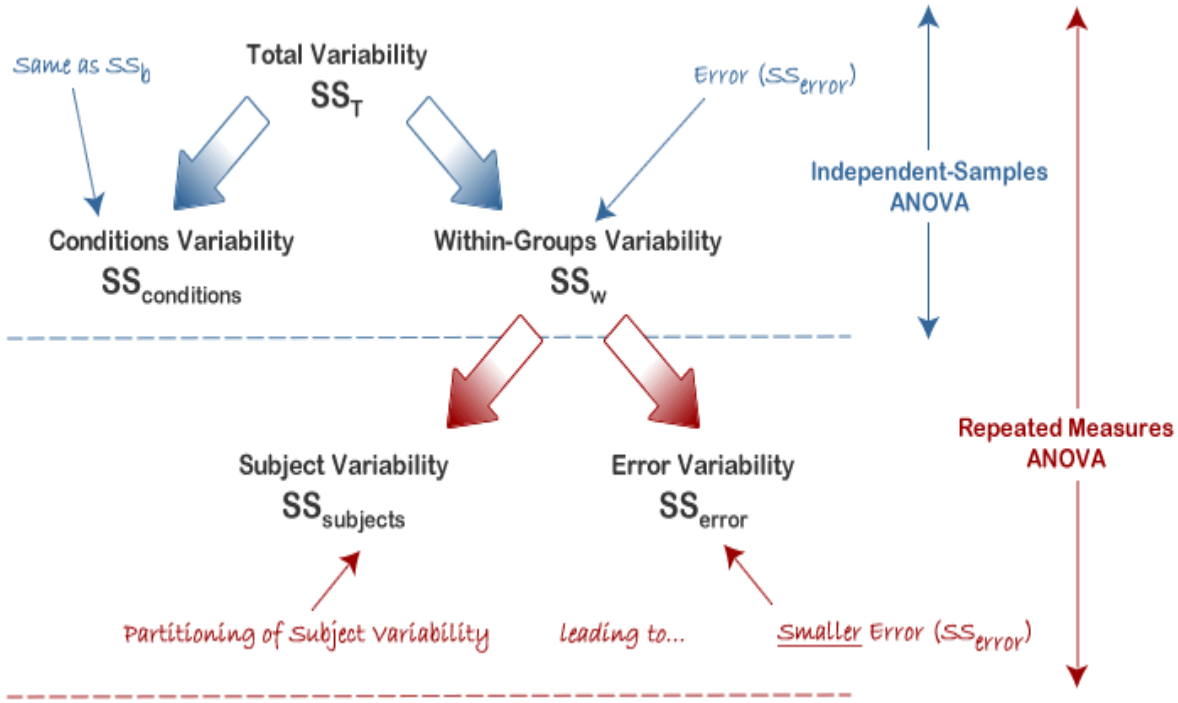
In this design, within-group variability (SS_w) is defined as the error variability (SS_{error}). Following division by the appropriate degrees of freedom, a mean sum of squares for between-groups (MS_b) and within-groups (MS_w) is determined and an *F*-statistic is calculated as the ratio of MS_b to MS_w (or MS_{error}), as shown below:

$$\textit{Independent ANOVA:} \quad F = \frac{MS_b}{MS_w} = \frac{MS_b}{MS_{error}}$$

A Repeated Measures ANOVA calculates an *F*-statistic in a similar way:

$$\textit{Repeated Measures ANOVA:} \quad F = \frac{MS_{conditions}}{MS_{error}}$$

The advantage of a Repeated Measures ANOVA is that whereas within-group variability (SS_w) expresses the error variability (SS_{error}) in an independent (between-subjects) ANOVA, a Repeated Measures ANOVA can further partition this error term, reducing its size, as is illustrated below:



This has the effect of increasing the value of the F-statistic due to the reduction of the denominator and leading to an increase in the power of the test to detect significant differences between means (this is discussed in more detail later). Mathematically, and as illustrated above, we partition the variability attributable to the differences between groups ($SS_{conditions}$) and variability within groups (SS_w) exactly as we do in a between-subjects (independent) ANOVA. However, with a Repeated Measures ANOVA, as we are using the same subjects in each group we can remove the variability due to the individual differences between subjects, referred to as $SS_{subjects}$, from the within-groups variability (SS_w). How is this achieved? Quite simply, we treat each subject as a block. That is, each subject becomes a level of a factor called subjects. We then calculate this variability as we do with any between-subjects factor. The ability to subtract $SS_{subjects}$ will leave us with a smaller SS_{error} term, as highlighted below:

$$\textit{Independent ANOVA:} \qquad SS_{error} = SS_w$$

$$\textit{Repeated Measures ANOVA:} \qquad SS_{error} = SS_w - SS_{subjects}$$

Now that we have removed the between-subjects variability, our new SS_{error} only reflects individual variability to each condition. You might recognise this as the interaction effect of subject by conditions; that is, how subjects react to the different conditions. Whether this leads to a more powerful test will depend on whether the reduction in SS_{error} more than compensates for the reduction in degrees of freedom for the error term (as degrees of freedom go from (n - k) to

(n - 1)(k - 1) (remembering that there are more subjects in the independent ANOVA design).

The next page of our guide deals with how to calculate a Repeated Measures ANOVA.

Repeated Measures ANOVA (cont...)

5

Calculating a Repeated Measures ANOVA

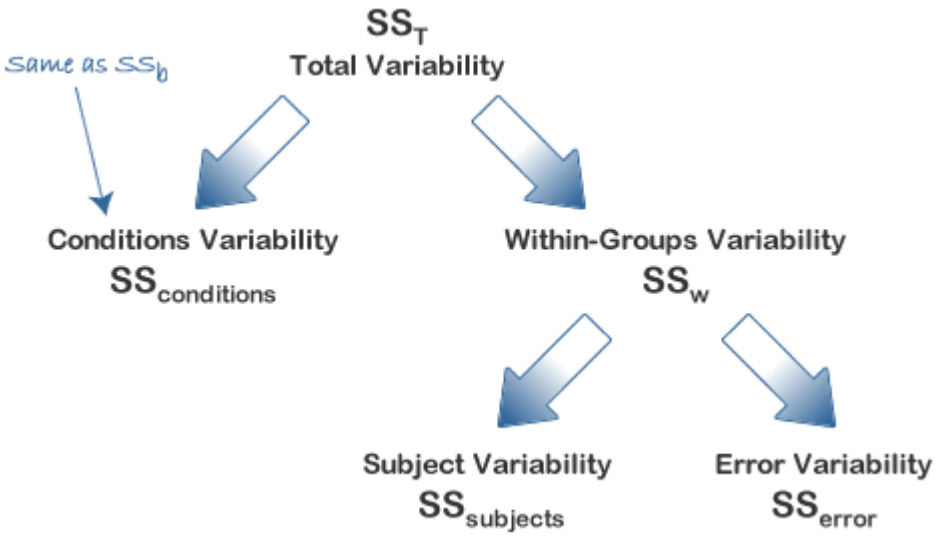
In order to provide a demonstration of how to calculate a Repeated Measures ANOVA, we shall use the example of a 6-month exercise-training intervention where six subjects had their fitness level measured on three occasions: pre-, 3 months, and post-intervention. Their data is shown below along with some initial calculations:

Exercise Intervention				
Subjects	**Pre-**	**3 Months**	**6 Months**	**Subject Means:**
1	45	50	55	**50**
2	42	42	45	**43**
3	36	41	43	**40**
4	39	35	40	**38**
5	51	55	59	**55**
6	44	49	56	**49.7**
Monthly Means:	**42.8**	**45.3**	**49.7**	
		Grand Mean:	**45.9**	

The Repeated Measures ANOVA, like other ANOVAs, generates an *F*-statistic that is used to determine statistical significance. The *F*-statistic is calculated as below:

$$F = \frac{MS_{time}}{MS_{error}} \quad or \quad F = \frac{MS_{conditions}}{MS_{error}}$$

You will already have been familiarised with $SS_{conditions}$ from earlier in this guide but in some of the calculations in the preceding sections you will see $SS_{conditions}$ sometimes referred to as SS_{time}. They both represent the sum of squares for the differences between related groups but SS_{time} is a more suitable name when dealing with time-course experiments, as we are in this example. The diagram below represents the partitioning of variance that occurs in the calculation of a Repeated Measures ANOVA.



In order to calculate an F-statistic we need to calculate $SS_{conditions}$ and SS_{error}. $SS_{conditions}$ can be calculated directly quite easily (as you will have encountered in an independent ANOVA as SS_b). Although SS_{error} can also be calculated directly it is somewhat difficult in comparison to deriving it from knowledge of other sums of squares which are easier to calculate, namely $SS_{subjects}$, and either SS_T or SS_w. SS_{error} can then be calculated in either of two ways:

$$\textit{either} \qquad SS_{error} = SS_w - SS_{subjects}$$

$$\textit{or} \qquad SS_{error} = SS_T - SS_{conditions} - SS_{subjects}$$

Both methods to calculate the *F*-statistic require the calculation of $SS_{conditions}$ and $SS_{subjects}$ but you then have the option to determine SS_{error} by first calculating either SS_T or SS_w. There is no right or wrong method, and other methods exist, it is simply personal preference as to which method you choose. For the purposes of this demonstration we shall calculate it using the first method, namely calculating SS_w.

Calculating SS_{time}

As mentioned previously, the calculation of SS_{time} is the same as for SS_b in an independent ANOVA, and can be expressed as:

$$SS_{time} = SS_b = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$$

where k = number of conditions, n_i = number of subjects under each (i^{th}) condition, $\bar{x}_i$= mean score for each (i^{th}) condition, $\bar{x}$= grand mean. So, in our example, we have:

$$SS_{time} = SS_b = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$$
$$= 6[(42.8 - 45.9)^2 + (45.3 - 45.9)^2 + (49.7 - 45.9)^2]$$
$$= 6[9.61 + 0.36 + 14.44]$$
$$= 143.44$$

Notice that because we have a repeated measures design, n_i is the same for each iteration: it is the number of subjects in our design. Hence we can simply multiple each group by this number. To better visualize the calculation above, the table below highlights the figures used in the calculation:

Exercise Intervention				
Subjects	**Pre-**	**3 Months**	**6 Months**	**Subject Means:**
1	45	50	55	**50**
2	42	42	45	**43**
3	36	41	43	**40**
4	39	35	40	**38**
5	51	55	59	**55**
6	44	49	56	**49.7**
Monthly Means:	**42.8**	**45.3**	**49.7**	
		Grand Mean:	**45.9**	

Calculating SS_w

Within-subjects variation (SS_w) is also calculated in the same way as in an independent ANOVA, expressed as follows:

$$SS_w = \sum_1 (x_{i1} - \bar{x}_1)^2 + \sum_2 (x_{i2} - \bar{x}_2)^2 + \cdots + \sum_k (x_{ik} - \bar{x}_k)^2$$

where x_{i1} is the score of the ith subject in group 1, x_{i2} is the score of the ith subject in group 2, and x_{ik} is the score of the ith subject in group k. In our case, this is:

$$SS_w = \sum_1 (x_{i1} - \bar{x}_1)^2 + \sum_2 (x_{i2} - \bar{x}_2)^2 + \sum_3 (x_{i3} - \bar{x}_3)^2$$
$$= [(45 - 42.8)^2 + (42 - 42.8)^2 + (36 - 42.8)^2 + (39 - 42.8)^2 + (51 - 42.8)^2 + (44 - 42.8)^2] +$$
$$[(50 - 45.3)^2 + (42 - 45.3)^2 + (41 - 45.3)^2 + (35 - 45.3)^2 + (55 - 45.3)^2 + (49 - 45.3)^2] +$$
$$[(55 - 49.7)^2 + (45 - 49.7)^2 + (43 - 49.7)^2 + (40 - 49.7)^2 + (59 - 49.7)^2 + (56 - 49.7)^2]$$
$$= \mathbf{715.5}$$

To better visualize the calculation above, the table below highlights the figures used in the calculation:

Exercise Intervention				
Subjects	**Pre-**	**3 Months**	**6 Months**	**Subject Means:**
1	45	50	55	**50**
2	42	42	45	**43**
3	36	41	43	**40**
4	39	35	40	**38**
5	51	55	59	**55**
6	44	49	56	**49.7**
Monthly Means:	**42.8**	**45.3**	**49.7**	
		Grand Mean:	**45.9**	

Calculating $SS_{subjects}$

As mentioned earlier, we treat each subject as its own block. In other words, we treat each subject as a level of an independent factor called subjects. We can then calculate $SS_{subjects}$ as follows:

$$SS_{subjects} = k \cdot \sum (\bar{x}_i - \bar{x})^2$$

where k = number of conditions, $\bar{x}_i$ mean of subject *i*, and $\bar{x}$ = grand mean. In our case, this is:

$$SS_{subjects} = k \cdot \sum (\bar{x}_i - \bar{x})^2$$

$$= 3[(50-45.9)^2 + (43-45.9)^2 + (40-45.9)^2 + (38-45.9)^2 + (55-45.9)^2 + (49.7-45.9)^2]$$

$$= \mathbf{658.3}$$

To better visualize the calculation above, the table below highlights the figures used in the calculation:

Exercise Intervention				
Subjects	**Pre-**	**3 Months**	**6 Months**	**Subject Means:**
1	45	50	55	**50**
2	42	42	45	**43**
3	36	41	43	**40**
4	39	35	40	**38**
5	51	55	59	**55**
6	44	49	56	**49.7**
Monthly Means:	**42.8**	**45.3**	**49.7**	
		Grand Mean:	**45.9**	

Calculating SS_{error}

We can now calculate SS_{error} by substitution:

$$SS_w = SS_{subjects} + SS_{error}$$

which, in our case, is:

$$SS_{error} = SS_w - SS_{subjects}$$
$$= 715.5 - 658.3$$
$$= \mathbf{57.2}$$

Determining MS_{time}, MS_{error} and the *F*-statistic

To determine the mean sum of squares for time (MS_{time}) we divide SS_{time} by its associated degrees of freedom (*k* - *1*), where *k* = number of time points. In our case:

$$MS_{time} = \frac{SS_{time}}{(k-1)}$$
$$= \frac{146.44}{2}$$
$$= 71.72$$

We do the same for the mean sum of squares for error (MS_{error}), this time dividing by (*n* - *1*)(*k* - *1*) degrees of freedom, where *n* = number of subjects and *k* = number of time points. In our case:

$$MS_{error} = \frac{SS_{error}}{(n-1)(k-1)}$$
$$= \frac{57.2}{(5)(2)}$$
$$= 5.72$$

Therefore, we can calculate the *F*-statistic as:

$$F = \frac{MS_{time}}{MS_{error}}$$
$$= \frac{71.72}{5.72}$$
$$= 12.53$$

We can now look up (or use a computer programme) to ascertain the critical *F*-statistic for our *F*-distribution with our degrees of freedom for time (df_{time}) and error (df_{error}) and determine whether our *F*-statistic indicates a statistically significant result.

We can now look up (or use a computer programme) to ascertain the critical F-statistic for our F-distribution with our degrees of freedom for time (df_{time}) and error (df_{error}) and determine whether our *F*-statistic indicates a statistically significant result.

How to report the result of a Repeated Measures ANOVA is shown on the next page.

Repeated Measures ANOVA (cont...)

11

Reporting the Result of a Repeated Measures ANOVA

We report the *F*-statistic from a Repeated Measures ANOVA as:

$$F(df_{time}, df_{error}) = F \text{ value}, p = p \text{ value}$$

which for our example would be:

$$F(2, 10) = 12.53, p = .002$$

This means we can reject the null hypothesis and accept the alternative hypothesis. As we will discuss later, there are assumptions and effect sizes we can calculate that can alter how we report the above result. However, we would otherwise report the above findings for this example exercise study as:

"There was a statistically significant effect of time on exercise-induced fitness, $F(2, 10) = 12.53$, $p < .002$."

or

"The six-month exercise-training programme had a statistically significant effect on fitness levels, $F(2, 10) = 12.53$, $p < .002$."

Tabular Presentation of a Repeated Measures ANOVA

Normally, the result of a Repeated Measures ANOVA is presented in the written text, as above, and not in a tabular form when writing a report. However, most statistical programmes, such as SPSS, will report the result of a Repeated Measures ANOVA in tabular form. Doing so allows the user to gain a fuller understanding of all the calculations that were made by the programme. The table below represents the type of table that you will be presented with and what the different sections mean.

Source	SS	df	MS	F
Conditions	$SS_{conditions}$	$(k-1)$	$MS_{conditions}$	$\frac{MS_{conditions}}{MS_{error}}$
Subjects	$SS_{subjects}$	$(n-1)$	$MS_{subjects}$	$\frac{MS_{subjects}}{MS_{error}}$
Error	SS_{error}	$(k-1)(n-1)$	MS_{error}	
Total	SS_T	$(N-1)$		

Most often, the Subjects row is not presented and sometimes the Total row is also omitted. The *F*-statistic found on the first row (time/conditions row) is the *F*-statistic that will determine whether there was a significant difference between at least two means or not. For our results, omitting the Subjects and Total rows, we have:

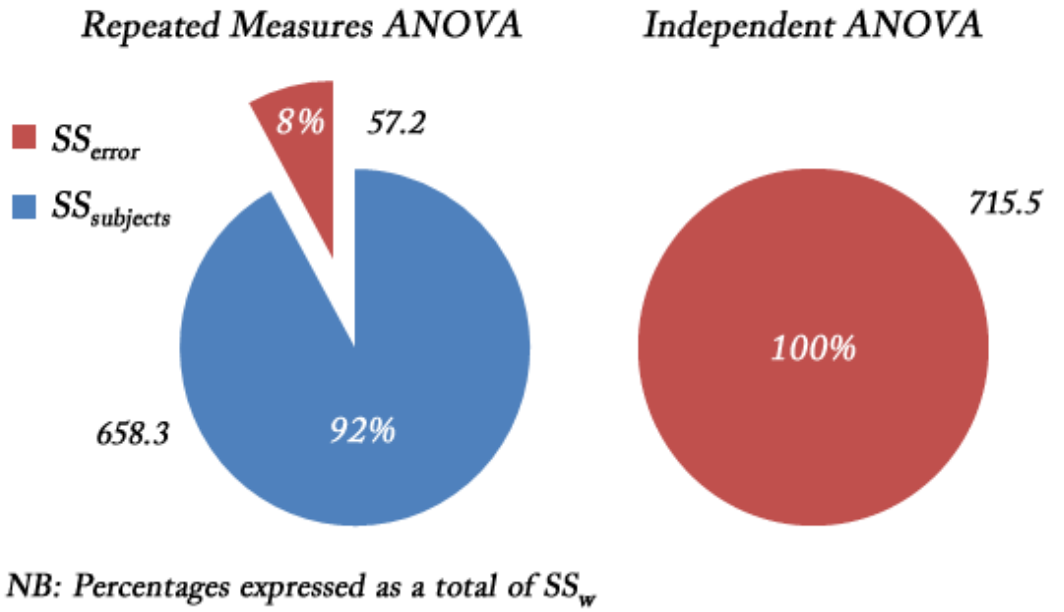
Source	SS	df	MS	F
Time	143.44	2	71.72	12.53
Error	57.2	10	5.72	

which is similar to the output produced by SPSS.

Increased Power in a Repeated Measures ANOVA

The major advantage with running a Repeated Measures ANOVA over an independent ANOVA is that the test is generally much more powerful. This particular advantage is achieved by the reduction in MSerror (the denominator of the *F*-statistic) that comes from the partitioning of variability due to differences between subjects ($SS_{subjects}$) from the original error term in an independent ANOVA (SS_w): i.e. $SS_{error} = SS_w - SS_{subjects}$. We achieved a result of $F(2, 10) = 12.53$, $p < .002$, for our example Repeated Measures ANOVA. How does this compare to if we had run an independent ANOVA instead? Well, if we ran through the calculations we would have ended up with a result of $F(2, 15) = 1.504$, $p = .254$, for the independent ANOVA. We can clearly see the advantage of using the same subjects in a Repeated Measures ANOVA as opposed to different subjects. For our exercise-training example, the illustration below shows that after taking away $SS_{subjects}$ from SS_w we are left with an error term (SS_{error}) that is only 8% as large as the independent ANOVA error term.

Comparison of Error Terms in Independent vs. Repeated Measures ANOVA



This does not lead to an automatic increase in the F-statistic as there are a greater number of degrees of freedom for SS_w than SS_{error}. However, it is usual for $SS_{subjects}$ to account for such a large percentage of the within-groups variability

that the reduction in the error term is large enough to more than compensate for the loss in the degrees of freedom (as used in selecting an *F*-distribution).

Effect Size for Repeated Measures ANOVA

It is becoming more common to report effect sizes in Partial eta-squared is where the the $SS_{subjects}$ has been removed from the denominator (and is what is produced by SPSS):

$$\eta^2_{partial} = \frac{SS_{conditions}}{(SS_{conditions} + SS_{error})} \quad or \quad \eta^2_{partial} = \frac{SS_{time}}{(SS_{time} + SS_{error})}$$

So, for our example, this would lead to a partial eta-squared of:

$$\eta^2_{partial} = \frac{SS_{time}}{(SS_{time} + SS_{error})}$$
$$= \frac{143.44}{(143.44 + 57.2)}$$
$$= 0.715$$

Underlying Assumptions: Normality

Similar to the other ANOVA tests, each level of the independent variable needs to be approximately normally distributed. How to check for this is provided in our Testing for Normality in SPSS guide.

Underlying Assumptions: Sphericity

The concept of sphericity, for all intents and purposes, is the repeated measures equivalent of homogeneity of variances. An explanation of sphericity is provided in our Sphericity guide. Testing for sphericity is an option in SPSS using Mauchly's Test for Sphericity as part of the GLM Repeated Measures procedure. A guide on running a Repeated Measures ANOVA in SPSS can be found here. We can write up our results (not the exercise example), where we have included Mauchly's Test for Sphericity as:

"Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, $\chi2(2) = 22.115$, $p < .0005$, and, therefore, a Greenhouse-Geisser correction was used. There was a significant effect of time on cholesterol concentration, $F(1.171, 38) = 21.032$, $p < .0005$."

Pearson Product-Moment Correlation

47

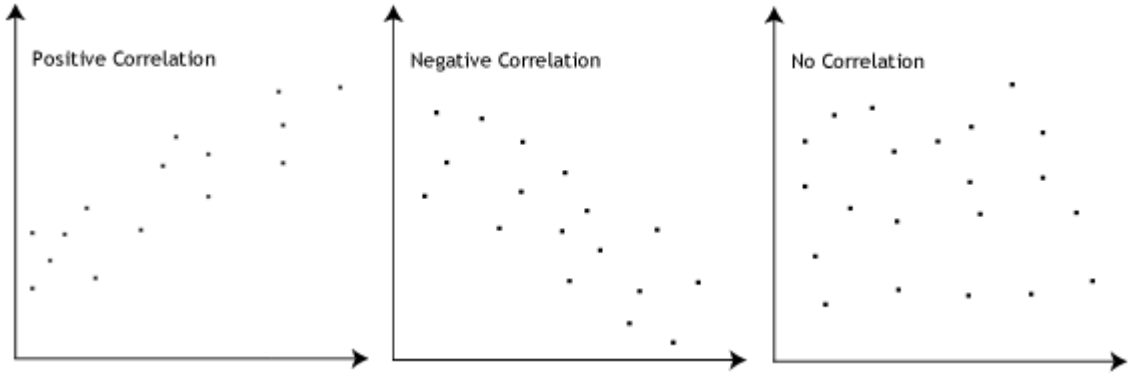
What does this test do?

The Pearson product-moment correlation coefficient (or Pearson correlation coefficient for short) is a measure of the strength of a linear association between two variables and is denoted by *r*. Basically, a Pearson product-moment correlation attempts to draw a line of best fit through the data of two variables, and the Pearson correlation coefficient, *r*, indicates how far away all these data

points are to this line of best fit (how well the data points fit this new model/line of best fit).

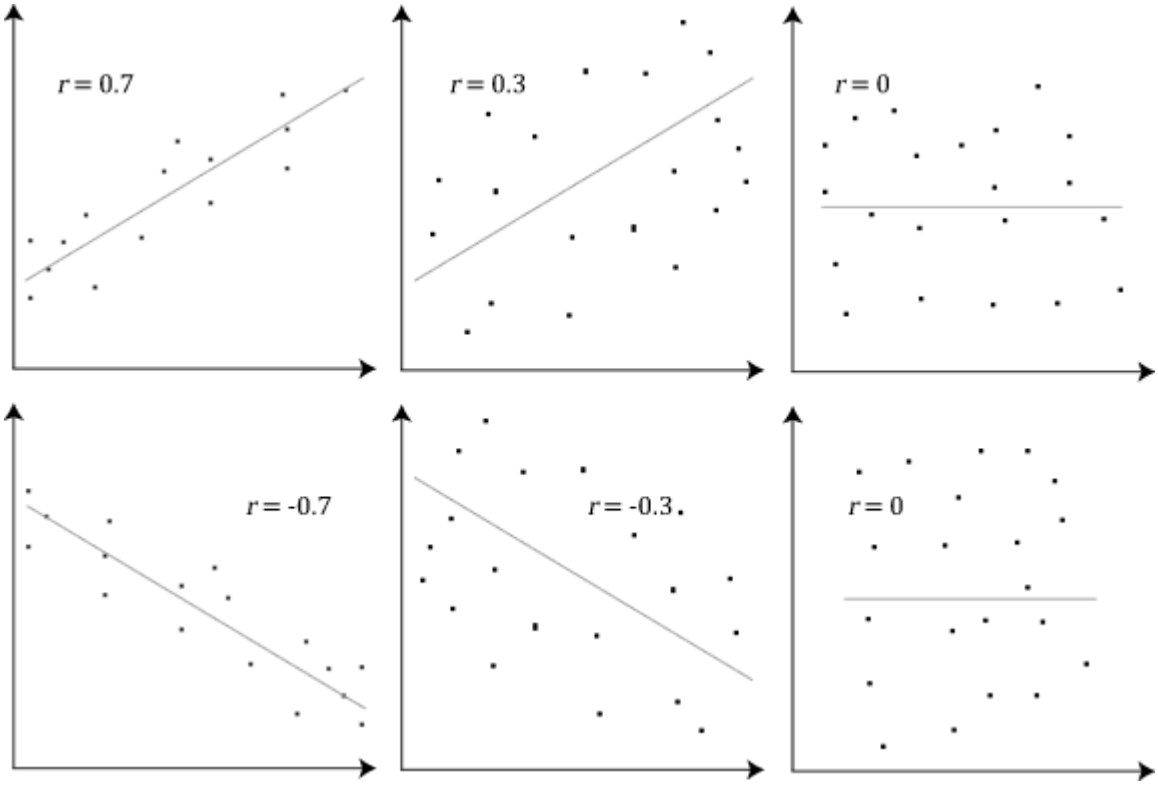
What values can the Pearson correlation coefficient take?

The Pearson correlation coefficient, *r*, can take a range of values from +1 to -1. A value of 0 indicates that there is no association between the two variables. A value greater than 0 indicates a positive association, that is, as the value of one variable increases so does the value of the other variable. A value less than 0 indicates a negative association, that is, as the value of one variable increases the value of the other variable decreases. This is shown in the diagram below (click the image to enlarge):



How can we determine the strength of association based on the Pearson correlation coefficient?

The stronger the association of the two variables the closer the Pearson correlation coefficient, *r*, will be to either +1 or -1 depending on whether the relationship is positive or negative, respectively. Achieving a value of +1 or -1 means that all your data points are included on the line of best fit - there are no data points that show any variation away from this line. Values for *r* between +1 and -1 (for example, *r* = 0.8 or -0.4) indicate that there is variation around the line of best fit. The closer the value of *r* to 0 the greater the variation around the line of best fit. Different relationships and their correlation coefficients are shown in the diagram below (click the image to enlarge):



Are there guidelines to interpreting Pearson's correlation coefficient?

Yes, the following guidelines have been proposed:

	Coefficient, *r*	
Strength of Association	Positive	Negative
Small	.1 to .3	-0.1 to -0.3
Medium	.3 to .5	-0.3 to -0.5
Large	.5 to 1.0	-0.5 to -1.0

Remember that these values are guidelines and whether an association is strong or not will also depend on what you are measuring.

Can you use any type of variable for Pearson's correlation coefficient?

No, the two variables have to be measured on either an interval or ratio scale. However, both variables do not need to be measured on the same scale, e.g. one variable can be ratio and one can be interval. Further information about types of variable can be found in our Types of Variable guide. If you have ordinal data then you will want to use Spearman's Rank Order Correlation or a Kendall's Tau Correlation instead of the Pearson product-moment correlation.

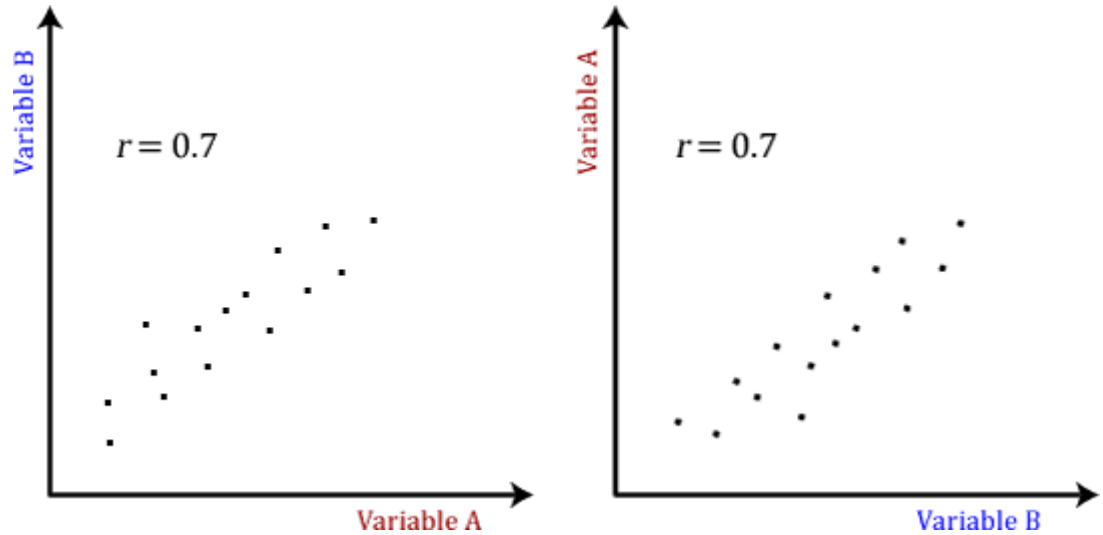
Do the two variables have to be measured in the same units?

No, the two variables can be measured in entirely different units. For example, you could correlate a person's age with their blood sugar levels. Here, the units are completely different; age is measured in years and blood sugar level measured in mmol/L (a measure of concentration). Indeed, the calculations for Pearson's correlation coefficient were designed such that the units of

measurement do not affect the calculation - this allows the correlation coefficient to be comparable and not influenced by the units of the variables used.

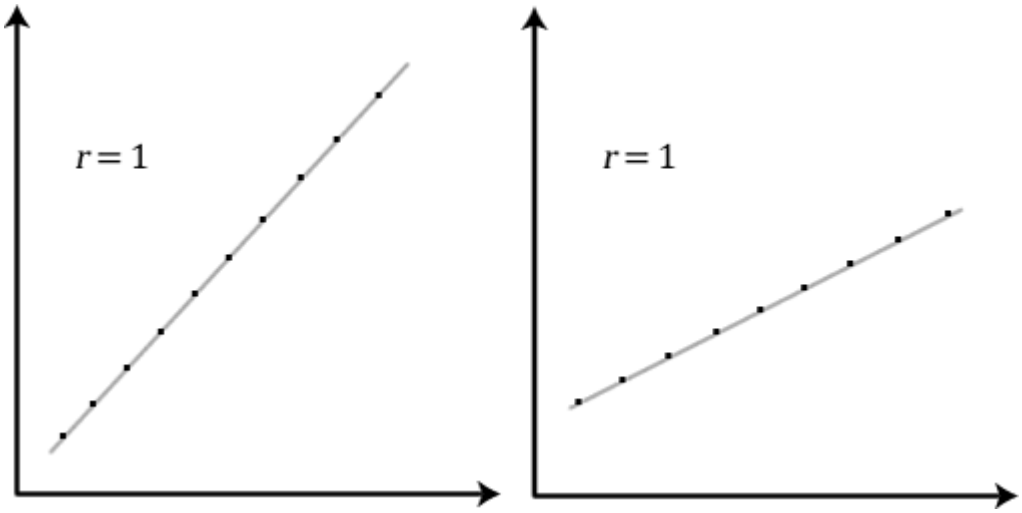
What about dependent and independent variables?

The Pearson product-moment correlation does not take into consideration whether a variable has been classified as a dependent or independent variable. It treats all variables equally. For example, you might want to find out whether basketball performance is correlated to a person's height. You might, therefore, plot a graph of performance against height and calculate the Pearson correlation coefficient. Lets say, for example, that r = .67. That is, as height increases so does basketball performance. This makes sense. However, if we plotted the variables the other way around and wanted to determine whether a person's height was determined by their basketball performance (which makes no sense) we would still get r = .67. This is because the Pearson correlation coefficient makes no account of any theory behind why you chose the two variables to compare. This is illustrated below:



Does the Pearson correlation coefficient indicate the slope of the line?

It is important to realise that the Pearson correlation coefficient, r, does not represent the slope of the line of best fit. Therefore, if you get a Pearson correlation coefficient of +1 this does not mean that for every unit increase in one variable there is a unit increase in another. It simply means that there is no variation between the data points and the line of best fit. This is illustrated below:



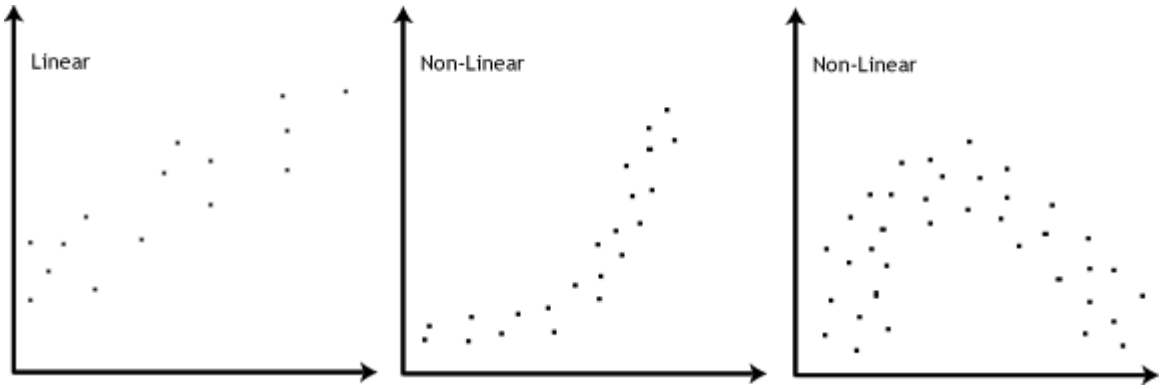
What assumptions does Pearson's correlation make?

There are four assumptions that are made with respect to Pearson's correlation:

1. The variables must be either interval or ratio measurements (see our Types of Variable guide for further details).
2. The variables must be approximately normally distributed (see our Testing for Normality guide for further details).
3. There is a linear relationship between the two variables. We discuss this later in this guide (jump to this section here).
4. Outliers are either kept to a minimum or are removed entirely. We also discuss this later in this guide (jump to this section here).
5. There is homoscedasticity of the data. This is discussed later in this guide (jump to this section here).

How can you detect a linear relationship?

To test to see whether your two variables form a linear relationship you simply need to plot them on a graph (a scatterplot, for example) and visually inspect the graph's shape. In the diagram below (click image to enlarge) you will find a few different examples of a linear relationship and some non-linear relationships. It is not appropriate to analyse a non-linear relationship using a Pearson product-moment correlation.



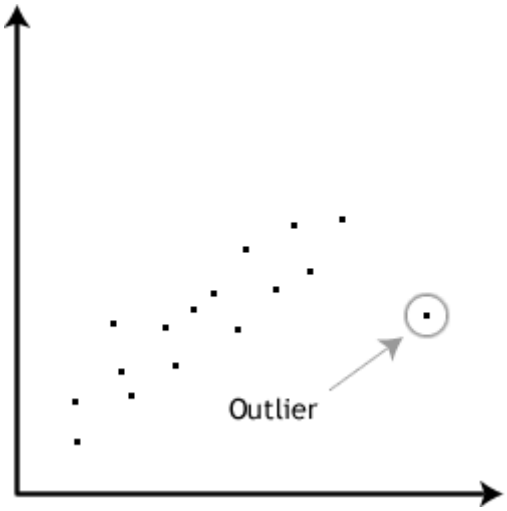
Go to the next page for more.

Pearson Product-Moment Correlation

15

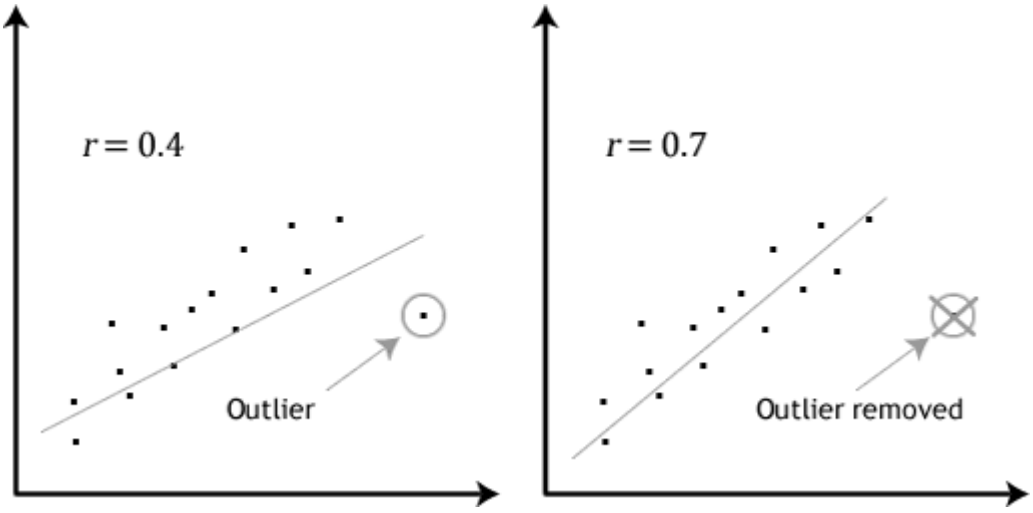
How can you detect outliers?

An outlier (in correlation analysis) is a data point that does not fit the general trend of your data but would appear to be a wayward (extreme) value and not what you would expect compared to the rest of your data points. You can detect outliers in a similar way to how you detect a linear relationship, by simply plotting the two variables against each other on a graph and visually inspecting the graph for wayward (extreme) points. You can then either remove or manipulate that particular point as long as you can justify why you did so (there are far more robust methods for detecting outliers in regression analysis). Alternatively, if you cannot justify removing the data point(s) then you can run a non-parametric test such as Spearman's Rank Order Correlation or Kendall's Tau Correlation instead, which are much less sensitive to outliers. This might be your best approach if you cannot justify removing the outlier. The diagram below indicates what a potential outlier might look like:



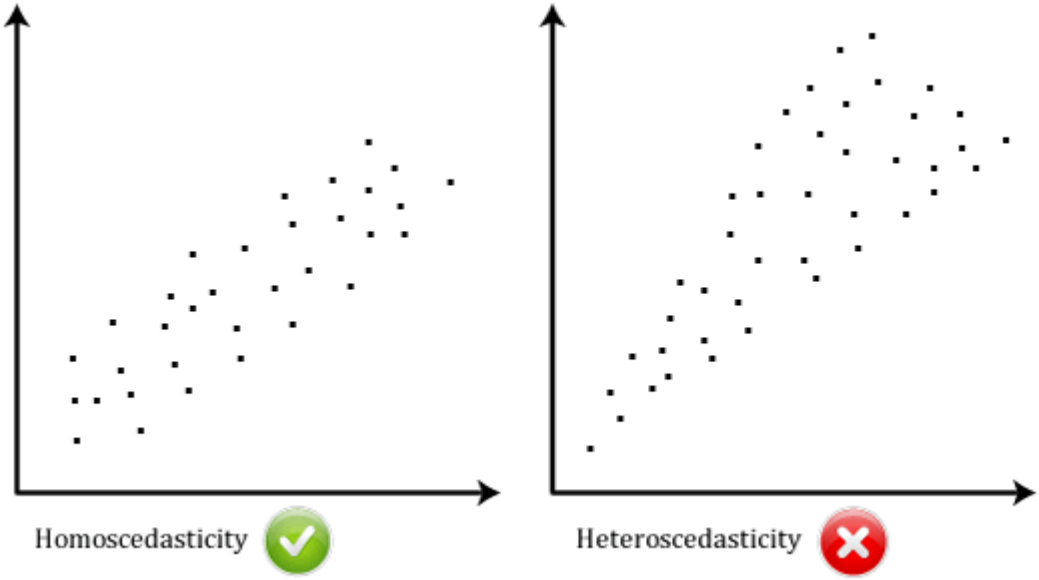
Why is testing for outliers so important?

Outliers can have a very large effect on the line of best fit and the Pearson correlation coefficient, which can lead to very different conclusions regarding your data. This point is most easily illustrated by studying scatterplots of a linear relationship with an outlier included and after its removal, with respect to both the line of best fit and the correlation coefficient. This is illustrated in the diagram below:



What is homoscedasticity?

Homoscedasticity basically means that the variances along the line of best fit remain similar as you move along the line. It is required that your data show homoscedasticity for you to run a Pearson product-moment correlation. Homoscedasticity is most easily demonstrated diagrammatically as below:

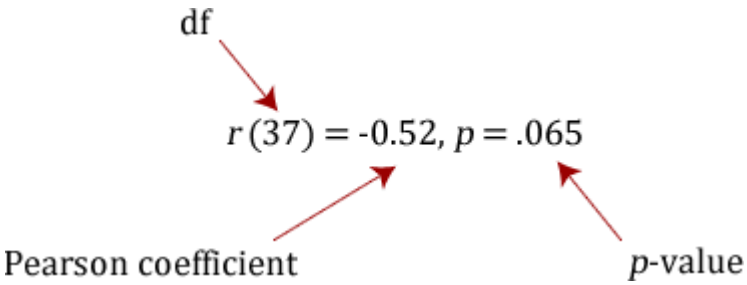


Can you establish cause-and-effect?

No, the Pearson correlation cannot determine a cause-and-effect relationship. It can only establish the strength of the association between two variables. As stated earlier, it does not even distinguish between independent and dependent variables.

How do I report the output of a Pearson product-moment correlation?

You need to state that you used the Pearson product-moment correlation and report the value of the correlation coefficient, *r*, as well as the degrees of freedom (df). You should express the result as follows:



where the degrees of freedom (df) is the number of data points minus 2 (*N* - 2). If you have not tested the significance of the correlation then leave that section out of the results.

Can I determine whether the association is statistically significant?

Yes, the easy way to do this is through a statistical programme, such as SPSS. We provide a guide on how to do this, which you can find here. You need to be careful how you interpret the statistical significance of a correlation. If your correlation coefficient has been determined to be statistically significant this does not mean that you have a strong association. It simply tests the null hypothesis that there is no relationship. By rejecting the null hypothesis you accept the alternative hypothesis that states that there is a relationship but with no information about the strength of the relationship or its importance.

What is the Coefficient of Determination?

The coefficient of determination, r^2, is the square of the Pearson correlation coefficient *r* (i.e. r^2). So, for example, a Pearson correlation coefficient of 0.6 would result in a coefficient of determination of 0.62, which is 0.36. Therefore, r^2 = 0.36. The coefficient of determination, with respect to correlation, is the proportion of the variance that is shared by both variables. It gives a measure of the amount of variation that can be explained by the model (the correlation is the model). It is sometimes expressed as a percentage (for example, 36% instead of 0.36) when we discuss the proportion of variance explained by the correlation. However, we must never write r^2 = 36%, or any other percentage. We must always write it as a proportion, e.g. r^2 = 0.36.

To run a Pearson correlation in SPSS, go to our guide here.

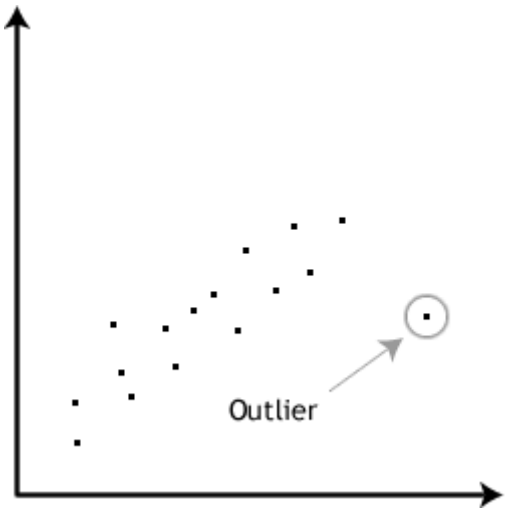
Pearson Product-Moment Correlation

15

How can you detect outliers?

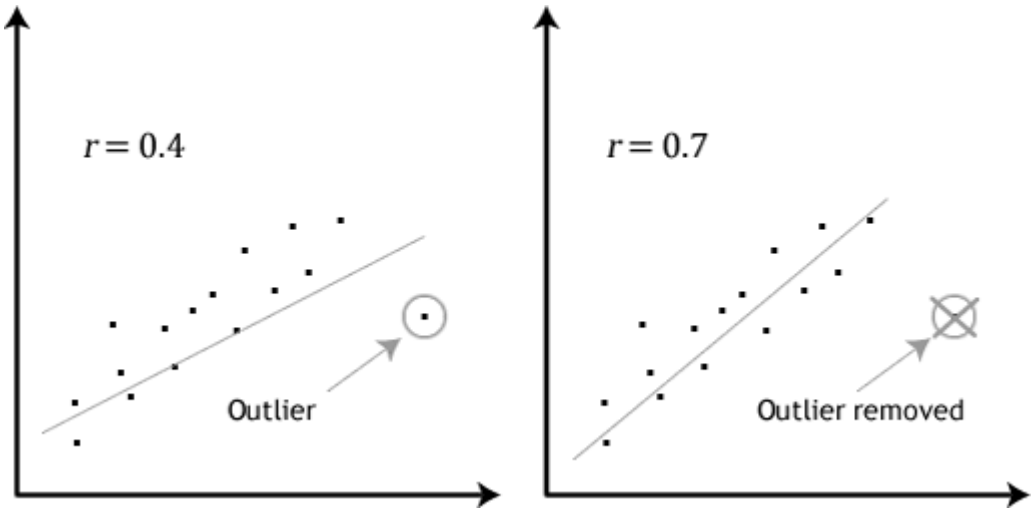
An outlier (in correlation analysis) is a data point that does not fit the general trend of your data but would appear to be a wayward (extreme) value and not what you would expect compared to the rest of your data points. You can detect outliers in a similar way to how you detect a linear relationship, by simply plotting the two variables against each other on a graph and visually inspecting the graph for wayward (extreme) points. You can then either remove or manipulate that particular point as long as you can justify why you did so (there are far more

robust methods for detecting outliers in regression analysis). Alternatively, if you cannot justify removing the data point(s) then you can run a non-parametric test such as Spearman's Rank Order Correlation or Kendall's Tau Correlation instead, which are much less sensitive to outliers. This might be your best approach if you cannot justify removing the outlier. The diagram below indicates what a potential outlier might look like:



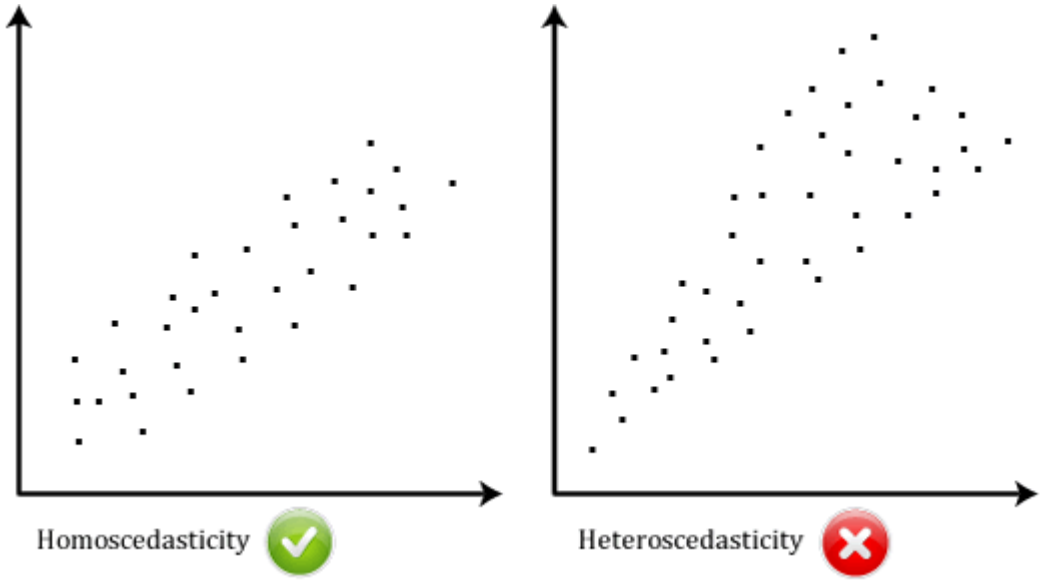
Why is testing for outliers so important?

Outliers can have a very large effect on the line of best fit and the Pearson correlation coefficient, which can lead to very different conclusions regarding your data. This point is most easily illustrated by studying scatterplots of a linear relationship with an outlier included and after its removal, with respect to both the line of best fit and the correlation coefficient. This is illustrated in the diagram below:



What is homoscedasticity?

Homoscedasticity basically means that the variances along the line of best fit remain similar as you move along the line. It is required that your data show homoscedasticity for you to run a Pearson product-moment correlation. Homoscedasticity is most easily demonstrated diagrammatically as below:

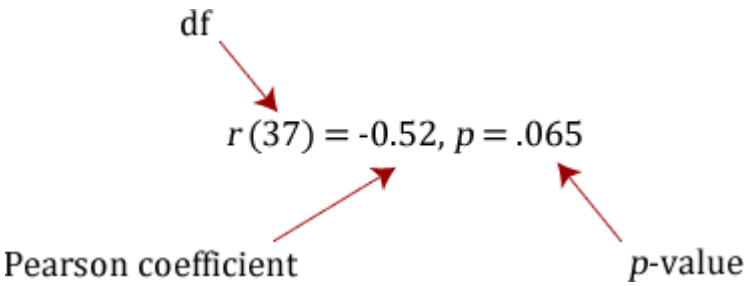


Can you establish cause-and-effect?

No, the Pearson correlation cannot determine a cause-and-effect relationship. It can only establish the strength of the association between two variables. As stated earlier, it does not even distinguish between independent and dependent variables.

How do I report the output of a Pearson product-moment correlation?

You need to state that you used the Pearson product-moment correlation and report the value of the correlation coefficient, *r*, as well as the degrees of freedom (df). You should express the result as follows:



where the degrees of freedom (df) is the number of data points minus 2 (*N* - 2). If you have not tested the significance of the correlation then leave that section out of the results.

Can I determine whether the association is statistically significant?

Yes, the easy way to do this is through a statistical programme, such as SPSS. We provide a guide on how to do this, which you can find here. You need to be careful how you interpret the statistical significance of a correlation. If your correlation coefficient has been determined to be statistically significant this does not mean that you have a strong association. It simply tests the null hypothesis that there is no relationship. By rejecting the null hypothesis you accept the alternative hypothesis that states that there is a relationship but with no information about the strength of the relationship or its importance.

What is the Coefficient of Determination?

The coefficient of determination, r^2, is the square of the Pearson correlation coefficient r (i.e. r^2). So, for example, a Pearson correlation coefficient of 0.6 would result in a coefficient of determination of 0.62, which is 0.36. Therefore, $r^2 = 0.36$. The coefficient of determination, with respect to correlation, is the proportion of the variance that is shared by both variables. It gives a measure of the amount of variation that can be explained by the model (the correlation is the model). It is sometimes expressed as a percentage (for example, 36% instead of 0.36) when we discuss the proportion of variance explained by the correlation. However, we must never write r^2 = 36%, or any other percentage. We must always write it as a proportion, e.g. $r^2 = 0.36$.

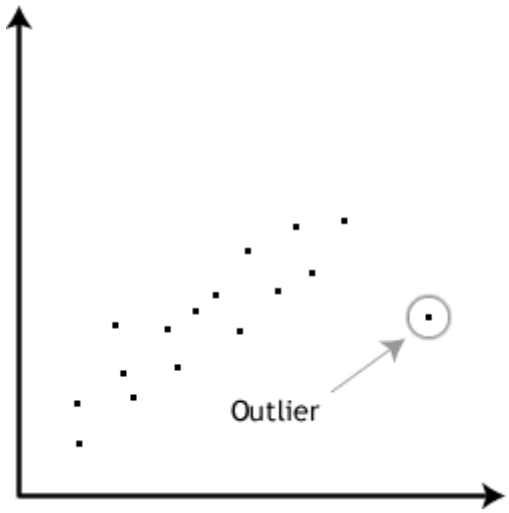
To run a Pearson correlation in SPSS, go to our guide here.

Pearson Product-Moment Correlation

15

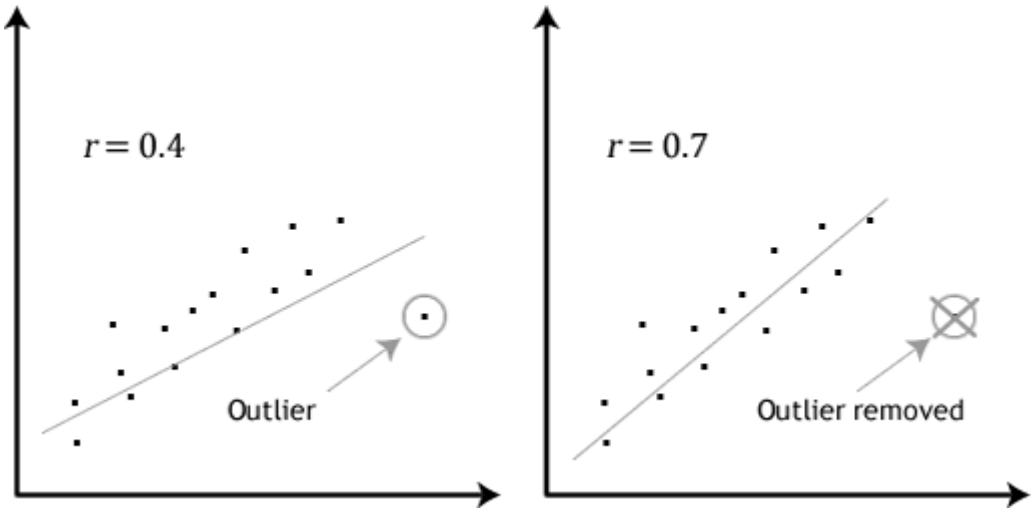
How can you detect outliers?

An outlier (in correlation analysis) is a data point that does not fit the general trend of your data but would appear to be a wayward (extreme) value and not what you would expect compared to the rest of your data points. You can detect outliers in a similar way to how you detect a linear relationship, by simply plotting the two variables against each other on a graph and visually inspecting the graph for wayward (extreme) points. You can then either remove or manipulate that particular point as long as you can justify why you did so (there are far more robust methods for detecting outliers in regression analysis). Alternatively, if you cannot justify removing the data point(s) then you can run a non-parametric test such as Spearman's Rank Order Correlation or Kendall's Tau Correlation instead, which are much less sensitive to outliers. This might be your best approach if you cannot justify removing the outlier. The diagram below indicates what a potential outlier might look like:



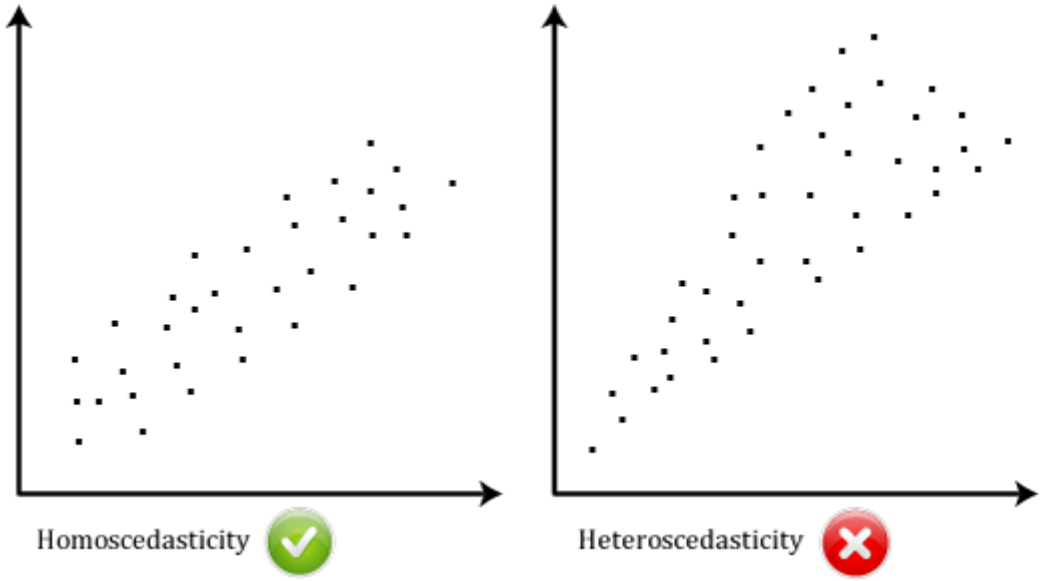
Why is testing for outliers so important?

Outliers can have a very large effect on the line of best fit and the Pearson correlation coefficient, which can lead to very different conclusions regarding your data. This point is most easily illustrated by studying scatterplots of a linear relationship with an outlier included and after its removal, with respect to both the line of best fit and the correlation coefficient. This is illustrated in the diagram below:



What is homoscedasticity?

Homoscedasticity basically means that the variances along the line of best fit remain similar as you move along the line. It is required that your data show homoscedasticity for you to run a Pearson product-moment correlation. Homoscedasticity is most easily demonstrated diagrammatically as below:

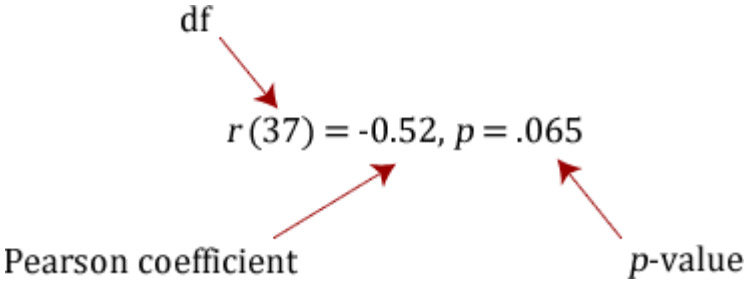


Can you establish cause-and-effect?

No, the Pearson correlation cannot determine a cause-and-effect relationship. It can only establish the strength of the association between two variables. As stated earlier, it does not even distinguish between independent and dependent variables.

How do I report the output of a Pearson product-moment correlation?

You need to state that you used the Pearson product-moment correlation and report the value of the correlation coefficient, *r*, as well as the degrees of freedom (df). You should express the result as follows:



where the degrees of freedom (df) is the number of data points minus 2 (*N* - 2). If you have not tested the significance of the correlation then leave that section out of the results.

Can I determine whether the association is statistically significant?

Yes, the easy way to do this is through a statistical programme, such as SPSS. We provide a guide on how to do this, which you can find here. You need to be careful how you interpret the statistical significance of a correlation. If your correlation coefficient has been determined to be statistically significant this does not mean that you have a strong association. It simply tests the null hypothesis that there is no relationship. By rejecting the null hypothesis you accept the alternative hypothesis that states that there is a relationship but with no information about the strength of the relationship or its importance.

What is the Coefficient of Determination?

The coefficient of determination, r^2, is the square of the Pearson correlation coefficient *r* (i.e. r^2). So, for example, a Pearson correlation coefficient of 0.6 would result in a coefficient of determination of 0.62, which is 0.36. Therefore, r^2 = 0.36. The coefficient of determination, with respect to correlation, is the proportion of the variance that is shared by both variables. It gives a measure of the amount of variation that can be explained by the model (the correlation is the model). It is sometimes expressed as a percentage (for example, 36% instead of 0.36) when we discuss the proportion of variance explained by the correlation. However, we must never write r^2 = 36%, or any other percentage. We must always write it as a proportion, e.g. r^2 = 0.36.

Spearman's Rank-Order Correlation

32

This guide will tell you when you should use Spearman's rank-order correlation to analyse your data, what assumptions you have to satisfy, how to calculate it, and how to report it. If you want to know how to run a Spearman correlation in SPSS then go to our guide here. If you want to calculate the correlation coefficient manually then we have a calculator you can use that also shows all the working out. This can be found at our Laerd Mathematics site here.

When should you use the Spearman's Rank-Order Correlation?

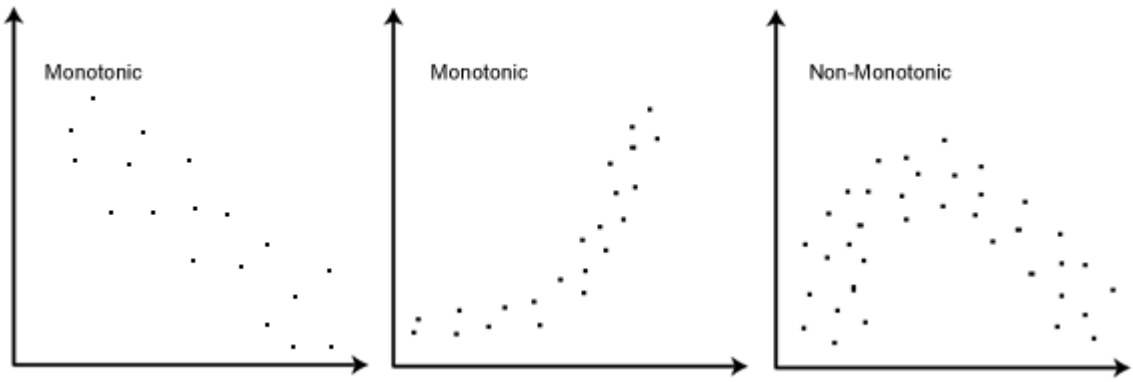
The Spearman's rank-order correlation is the nonparametric version of the Pearson product-moment correlation. Spearman's correlation coefficient, (ρ, also signified by r_s) measures the strength of association between two ranked variables.

What are the assumptions of the test?

You need two variables that are either ordinal, interval or ratio (see our Types of Variable guide if you need clarification). Although you would normally hope to use a Pearson product-moment correlation on interval or ratio data, the Spearman correlation can be used when the assumptions of the Pearson correlation are markedly violated. A second assumption is that there is a monotonic relationship between your variables.

What is a monotonic relationship?

A monotonic relationship is a relationship that does one of the following: (1) as the value of one variable increases so does the value of the other variable or (2) as the value of one variable increases the other variable value decreases. Examples of monotonic and non-monotonic relationships are presented in the diagram below (click image to enlarge):



Why is a monotonic relationship important to Spearman's correlation?

A monotonic relationship is an important underlying assumption of the Spearman rank-order correlation. It is also important to recognize the assumption of a monotonic relationship is less restrictive than a linear relationship (an assumption that has to be met by the Pearson product-moment correlation). The middle image above illustrates this point well: A non-linear relationship exists but the relationship is monotonic and is suitable for analysis by Spearman's correlation but not by Pearson's correlation.

How to rank data?

In some cases your data might already be ranked but often you will find that you need to rank the data yourself (or use SPSS to do it for you). Thankfully, ranking data is not a difficult task and is easily achieved by working through your data in a table. Let us consider the following example data regarding the marks achieved in a maths and English exam:

English	56	75	45	71	61	64	58	80	76	61
Maths	66	70	40	60	65	56	59	77	67	63

The procedure for ranking these scores is as follows:

First, create a table with four columns and label them as below:

English (mark)	Maths (mark)	Rank (English)	Rank (maths)
56	66	9	4
75	70	3	2
45	40	10	10
71	60	4	7
61	65	6.5	5
64	56	5	9
58	59	8	8
80	77	1	1
76	67	2	3
61	63	6.5	6

You need to rank the scores for maths and English separately. The score with the highest value should be labelled "1" and the lowest score should be labelled "10" (if your data set has more than 10 cases then the lowest score will be how many cases you have). Look carefully at the two individuals that scored 61 in the English exam (highlighted in bold). Notice their joint rank of 6.5. This is because when you have two identical values in the data (called a "tie") you need to take the average of the ranks that they would have otherwise occupied. We do this as, in this example, we have no way of knowing which score should be put in rank 6 and which score should be ranked 7. Therefore, you will notice that the ranks of 6 and 7 do not exist for English. These two ranks have been averaged ((6 + 7)/2 = 6.5) and assigned to each of these "tied" scores.

What is the definition of Spearman's rank-order correlation?

There are two methods to calculate Spearman's rank-order correlation depending on whether: (1) your data does not have tied ranks or (2) your data has tied ranks. The formula for when there are no tied ranks is:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where d_i = difference in paired ranks and *n* = number of cases. The formula to use when there are tied ranks is:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

where *i* = paired score.

Spearman's Rank-Order Correlation (cont...)

21

What values can the Spearman correlation coefficient, r_s, take?

The Spearman correlation coefficient, r_s, can take values from +1 to -1. A r_s of +1 indicates a perfect association of ranks, a r_s of zero indicates no association between ranks and a r_s of -1 indicates a perfect negative association of ranks. The closer r_s is to zero, the weaker the association between the ranks.

An example of calculating Spearman's correlation

To calculate a Spearman rank-order correlation on data without any ties we will use the following data:

English	56	75	45	71	62	64	58	80	76	61
Maths	66	70	40	60	65	56	59	77	67	63

We then complete the following table:

English (mark)	Maths (mark)	Rank (English)	Rank (maths)	d	d^2
56	66	9	4	5	25
75	70	3	2	1	1
45	40	10	10	0	0
71	60	4	7	3	9
62	65	6.5	5	1	1
64	56	5	9	4	16
58	59	8	8	0	0
80	77	1	1	0	0
76	67	2	3	1	1
61	63	6.5	6	1	1

Where d = difference between ranks and d^2 = difference squared.

We then calculate the following:

$$\sum d_i^2 = 25 + 1 + 9 + 1 + 16 + 1 + 1 = 54$$

We then substitute this into the main equation with the other information as follows:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$\rho = 1 - \frac{6 \times 54}{10(10^2-1)}$$

$$\rho = 1 - \frac{324}{990}$$

$$\rho = 1 - 0.33$$

$$\rho = 0.67$$

as n = 10. Hence, we have a ρ of 0.67. This indicates a strong positive relationship between the ranks individuals obtained in the maths and English exam. That is, the higher you ranked in maths, the higher you ranked in English also, and vice versa.

How do you report a Spearman's correlation?

How you report a Spearman's correlation coefficient depends on whether or not you have determined the statistical significance of the coefficient. If you have simply run the Spearman correlation without any statistical significance tests then you are able to simple state the value of the coefficient as shown below:

$$\rho = coefficient \quad OR \quad r_s = coefficient$$

e.g.

$$\rho = 0.67 \quad OR \quad r_s = 0.67$$

However, if you have also run statistical significance tests then you need to include some more information as shown below:

$$\rho(df) = \rho\ coefficient, P = P\ value$$

$$\rho(8) = 0.67, P = 0.033$$

where df = N - 2, where N = number of pairwise cases.

How do you express the hypothesis for this test?

The general form of a null hypothesis for a Spearman correlation is:

H_0: There is no association between the two variables [in the population].

Remember, you are making an inference from your sample to the population that the sample is supposed to represent. However, as this a general understanding of an inferential statistical test, it is often not included. A null hypothesis statement for the example used earlier in this guide would be:

H_0: There is no association between maths and English marks.

How do I interpret a statistically significant Spearman correlation?

It is important to realize that statistical significance does not indicate the strength of the Spearman rank-order correlation. In fact, the statistical significance testing of the Spearman correlation does not provide you with *any* information about the strength of the relationship. Thus, achieving a value of $P = 0.001$, for example, does not mean that the relationship is stronger than if you achieved a value of $P = 0.04$. This is because the significance test is investigating whether you can accept or reject the null hypothesis. If you set $\alpha = 0.05$ then achieving a statistically significant Spearman rank-order correlation means that you can be sure that there is less than a 5% chance that the strength of the relationship you found (your rho coefficient) happened by chance if the null hypothesis were true.

Testing Assumptions

Sphericity

13

Introduction

ANOVAs with repeated measures (within-subject factors) are particularly susceptible to the violation of the assumption of sphericity. Sphericity is the condition where the variances of the differences between all combinations of related groups (levels) are equal. Violation of sphericity is when the variances of the *differences* between all combinations of related groups are not equal. Sphericity can be likened to homogeneity of variances in a between-subjects ANOVA.

The violation of sphericity is serious for the Repeated Measures ANOVA, with violation causing the test to become too liberal (i.e. an increase in the Type I error rate). Determining whether sphericity has been violated is, therefore, very important. Luckily, if violations of sphericity do occur then corrections have been developed to produce a more valid critical *F*-value (i.e. reduce the increase in Type I error rate). This is achieved by estimating the degree to which sphericity has been violated and applying a correction factor to the degrees of freedom of the *F*-distribution. We will discuss this in more detail later in this guide. Firstly, we will illustrate what sphericity is by way of an example.

An Example of Sphericity

To illustrate the concept of sphericity as equality of variance of the *differences* between each pair of values, we will analyse the fictitious data in the Table 1 below. This data is from a fictitious study that measured aerobic capacity (units: ml/min/kg) at three time points (Time 1, Time 2, Time 3) for six subjects.

Subject	Time 1	Time 2	Time 3	Time 1 – Time 2	Time 1 – Time 3	Time 2 – Time 3
1	45	50	55	-5	-10	-5
2	42	42	45	0	-3	-3
3	36	41	43	-5	-7	-2
4	39	35	40	4	-1	-5
5	51	55	59	-4	-8	-4
6	44	49	56	-5	-12	-7
			Variance:	**13.9**	**17.4**	**3.1**

Firstly, as we are interested in the differences between related groups (time points) we must calculate the differences between each combination of related group (time point) (the last three columns in the table above). The more time points (or conditions) the greater the number of possible combinations. For three time points, we have three different combinations. We then need to calculate the variance of each group difference, again presented in the table above. Looking at our results, at first glance, it would appear that the variances between the paired differences are not equal (13.9 *vs.* 17.4 *vs.* 3.1); the variance of the difference between Time 2 and Time 3 is much less than the other two combinations. This might lead us to conclude that our data violates the assumption of sphericity. We can, however, test our data for sphericity using a formal test called Mauchly's Test of Sphericity.

Testing for Sphericity: Mauchly's Test of Sphericity

As just mentioned, Mauchly's Test of Sphericity is a formal way of testing the assumption of sphericity. Although this test has been heavily criticised, often failing to detect departures from sphericity in small samples and over-detecting them in large samples, it is nonetheless a commonly used test. This is probably due to its automatic print out in SPSS for repeated measures ANOVAs and the lack of an otherwise readily available test. However, despite these shortcomings, because it is widely used, we will explain the test in this section and how to interpret it.

Mauchly's Test of Sphericity tests the null hypothesis that the variances of the differences are equal. Thus, if Mauchly's Test of Sphericity is statistically significant ($p < .05$), we can reject the null hypothesis and accept the alternative hypothesis that the variances of the differences are not equal (i.e. sphericity has been violated). Results from Mauchly's Test of Sphericity are shown below for our example data (see the red section below):

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	.434	3.343	2	.188	.638	.760	.500

The results of this test show that sphericity has not been violated ($p = .188$) (you need to look under the "Sig." column). We can thus report the result of Mauchly's Test of Sphericity as follows:

"Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 3.343$, $p = .188$."

You might have noticed the discrepancy between the result of Mauchly's Test of Sphericity, which indicates that the assumption of sphericity is not violated, and the large differences in the variances calculated earlier (13.9 *vs.* 17.4 *vs.* 3.1), suggesting violation of the assumption of sphericity. Unfortunately, this is one of the problems of Mauchly's test when dealing with small sample sizes, which was mentioned earlier.

If your data does not violate the assumption of sphericity then you do not need to modify your degrees of freedom. [If you are using SPSS, then your results will be presented in the "sphericity assumed" row(s).] Not violating this assumption means that the *F*-statistic that you have calculated is valid and can be used to determine statistical significance. If, however, the assumption of sphericity is violated, the *F*-statistic is positively biased rendering it invalid and increasing the risk of a Type I error. To overcome this problem corrections must be applied to the degrees of freedom *(df)*, such that a valid critical *F*-value can be obtained. It should be noted that it is not uncommon to find that sphericity has been violated.

The corrections that you will encounter to combat the violation of the assumption of sphericity are the **lower-bound estimate**, **Greenhouse-Geisser correction** and the **Huynh-Feldt correction**. These corrections rely on estimating sphericity.

Estimating Sphericity (ε) and How Corrections Work

The degree to which sphericity is present, or not, is represented by a statistic called epsilon (ε). An epsilon of 1 (i.e. $\varepsilon = 1$) indicates that the condition of sphericity is exactly met. The further epsilon decreases below 1 (i.e. $\varepsilon < 1$), the greater the violation of sphericity. Therefore, you can think of epsilon as a statistic that describes the degree to which sphericity has been violated. The lowest value that epsilon (ε) can take is called the lower-bound estimate whilst both the Greenhouse-Geisser and the Huynd-Feldt procedures attempt to estimate epsilon (ε) albeit in different ways (it is an estimate as we are dealing with samples not populations). For this reason, the estimates of sphericity (ε) tend to always be different depending on which procedure is used. By estimating epsilon (ε), all these procedures then use their sphericity estimate (ε) to correct the degrees of freedom for the *F*-distribution. As you will see later on in this guide, the actual value of the *F*-statistic does not change as a result of applying the corrections.

So what effect are the corrections on the degrees of freedom having? The answer to this lies in how the critical values for the *F*-statistic are calculated. The corrections affect the degrees of freedom of the *F*-distribution such that larger critical values are used (i.e. the *p*-value increases). This is to counteract the fact that when the assumption of sphericity is violated there is an increase in Type I errors due to the critical values in the *F*-table being too small. These corrections attempt to correct this bias.

Recall that the degrees of freedom used in the calculation of the *F*-statistic in a Repeated Measures ANOVA are:

$$df_{time/condition} = (k-1)$$

$$df_{error} = (k-1)(n-1)$$

where *k* = number of repeated measures and *n* = number of subjects. The three corrections (lower-bound estimate, Greenhouse-Geisser and Huynh-Feldt correction) all alter the degrees of freedom by multiplying these degrees of freedom by their estimated epsilon (ε) as below:

$$df_{time/condition} = \hat{\varepsilon}(k-1)$$

$$df_{error} = \hat{\varepsilon}(k-1)(n-1)$$

Please note that the different corrections use different mathematical symbols for estimated epsilon (ε), which will be shown later on.

Also recall that the *F*-statistic is calculated as:

$$F = \frac{MS_{time/conditions}}{MS_{error}} = \frac{\frac{SS_{time/conditions}}{df_{time/conditions}}}{\frac{SS_{error}}{df_{error}}} = \frac{\frac{SS_{time/conditions}}{(k-1)}}{\frac{SS_{error}}{(n-1)(k-1)}}$$

As stated earlier, these corrections do not lead to a different *F*-statistic. But how does the *F*-statistic remain unaffected when the degrees of freedom are being altered? This is because the estimated epsilon is added as a multiplier to the degrees of freedom for both the numerator and denominator, and thus they cancel each other out, as shown below:

$$F = \frac{\frac{SS_{time/conditions}}{\hat{\varepsilon}(k-1)}}{\frac{SS_{error}}{\hat{\varepsilon}(n-1)(k-1)}} = \frac{\frac{SS_{time/conditions}}{\cancel{\hat{\varepsilon}}(k-1)}}{\frac{SS_{error}}{\cancel{\hat{\varepsilon}}(n-1)(k-1)}} = \frac{\frac{SS_{time/conditions}}{(k-1)}}{\frac{SS_{error}}{(n-1)(k-1)}} = \frac{MS_{time/conditions}}{MS_{error}}$$

For our example, we have the three estimates of epsilon (ε) calculated as follows (using SPSS):

Epsilon[a]		
Greenhouse-Geisser	Huynh-Feldt	Lower-bound
.638	.760	.500

Lower-Bound Estimate

The lowest value that epsilon (ε) can take is called the lower-bound estimate (or the lower-bound adjustment) and is calculated as:

$$\textbf{lower bound}\ estimate\ (\varepsilon) = \frac{1}{(1-k)}$$

where k = number of repeated measures. As you can see from the above equation, the greater the number of repeated measures, the greater the potential for the violation of sphericity. So, for our example which has three repeated measures, the lowest value epsilon (ε) can take would be:

$$\textbf{lower bound } estimate(\varepsilon) = \frac{1}{(1-3)}$$
$$= 0.5$$

This represents the greatest possible violation of sphericity and, therefore, using the lower-bound estimate means that you are correcting your degrees of freedom for the "worst case scenario". This provides a correction that is far too conservative (incorrectly rejecting the null hypothesis). This correction has been superseded by the **Greenhouse-Geisser** and **Huynd-Feldt** corrections, and the lower-bound estimate is no longer a recommended correction, i.e. *do not use the lower-bound estimate.*

The other types of correction and interpreting statistical printouts of sphericity are to be found on the next page.

Sphericity (cont...)

4

Greenhouse-Geisser Correction

The Greenhouse-Geisser procedure estimates epsilon (referred to as $\hat{\varepsilon}$) in order to correct the degrees of freedom of the *F*-distribution as has been mentioned previously, and shown below:

$$df_{time/condition} = \hat{\varepsilon}(k-1)$$
$$df_{error} = \hat{\varepsilon}(k-1)(n-1)$$

Using our prior example, and if sphericity had been violated, we would have:

$$df_{time/condition} = 0.638(3-1) \qquad = 1.276$$
$$df_{error} = 0.638(3-1)(6-1) = 6.380 \qquad where\ \hat{\varepsilon} = 0.638$$

So our *F*-test result is corrected from F (2,10) = 12.534, p = .002 to F (1.277,6.384) = 12.534, p = .009 (degrees of freedom are slightly different due to rounding). The correction has elicited a more accurate significance value; it has increased the p-value to compensate for the fact that the test is too liberal when sphericity is violated.

Huynd-Feldt Correction

As with the Greenhouse-Geisser correction, the Huynd-Feldt correction estimates epsilon, (represented as $\tilde{\varepsilon}$) in order to correct the degrees of freedom of the *F*-distribution as shown below:

$$df_{time/condition} = \tilde{\varepsilon}(k-1)$$
$$df_{error} = \tilde{\varepsilon}(k-1)(n-1)$$

Using our prior example, and if sphericity had been violated, we would have:

$$df_{time/condition} = 0.760(3-1) = 1.520$$

$$df_{error} = 0.760(3-1)(6-1) = 7.600 \qquad where\ \tilde{\varepsilon} = 0.760$$

So our F test result is corrected from F (2,10) = 12.534, p = .002 to F (1.520,7.602) = 12.534, p = .005 (degrees of freedom are slightly different due to rounding). As with the Greenhouse-Geisser correction, this correction has elicited a more accurate significance value; it has increased the p-value to compensate for the fact that the test is too liberal when sphericity is violated.

Greenhouse-Geisser vs. Huynd-Feldt Correction

The Greenhouse-Geisser correction tends to underestimate epsilon (ε) when epsilon (ε) is close to 1 (i.e. it is a conservative correction) whilst the Huynd-Feldt correction tends to overestimate epsilon (ε) (i.e. it is a more liberal correction). Generally, the recommendation is to use the Greenhouse-Geisser correction, especially if estimated epsilon (ε) is less than 0.75. However, some statisticians recommend using the Huynd-Feldt correction if estimated epsilon (ε) is greater than 0.75. In practice, both corrections produce very similar corrections, so if estimated epsilon (ε) is greater than 0.75, you can equally justify using either.

Interpreting Statistical Printouts

To see all the above in action, consider the data set we have been using for this article. We can see from our earlier table that, for our data set, the estimated epsilon (ε) using the Greenhouse-Geisser method is 0.638 (i.e. $\hat{\varepsilon}$= 0.638). The following table shows the output of our Repeated Measures ANOVA (in SPSS):

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	143.444	2	71.722	12.534	.002
	Greenhouse-Geisser	143.444	1.277	112.350	12.534	.009
	Huynh-Feldt	143.444	1.520	94.351	12.534	.005
	Lower-bound	143.444	1.000	143.444	12.534	.017
Error(TIME)	Sphericity Assumed	57.222	10	5.722		
	Greenhouse-Geisser	57.222	6.384	8.964		
	Huynh-Feldt	57.222	7.602	7.528		
	Lower-bound	57.222	5.000	11.444		

In SPSS, the "Sphericity Assumed" row(s) are where sphericity has not been violated and, therefore, represents the normal calculations we would make to calculate a significance value for a Repeated Measures ANOVA. Notice how the sum of squares and F-statistic are identical regardless of whether or which correction is applied (shown below in blue). This further highlights that the corrections are not being applied to the partitioning of sum of squares but to the degrees of freedom.

Corrections affect degrees of freedom (df), Mean Sum of Squares (MS) and *p*-value

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	143.444	2	71.722	12.534	.002
	Greenhouse-Geisser	143.444	1.277	112.350	12.534	.009
	Huynh-Feldt	143.444	1.520	94.351	12.534	.005
	Lower-bound	143.444	1.000	143.444	12.534	.017
Error(TIME)	Sphericity Assumed	57.222	10	5.722		
	Greenhouse-Geisser	57.222	6.384	8.964		
	Huynh-Feldt	57.222	7.602	7.528		
	Lower-bound	57.222	5.000	11.444		

Corrections have **NO** effect on Sum of Squares (SS) or the *F*-statistic

We can see in the diagram above that the corrections have altered the degrees of freedom (df), which in turn have altered the Mean Sum of Squares (MS) for both the TIME factor and its error, and have altered the level of significance of the *F*-statistic.

Univariate *vs.* Multivariate Analysis

An alternative method is to use a MANOVA instead of a Repeated Measures ANOVA. The reason for doing this is that the MANOVA does not require the assumption of sphericity. There are a number of reasons for choosing a MANOVA over a Repeated Measures ANOVA and *vice-versa* and we will be adding this information to this guide soon.

Population Standard Deviation

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$$

$$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$$

Sample Standard Deviation