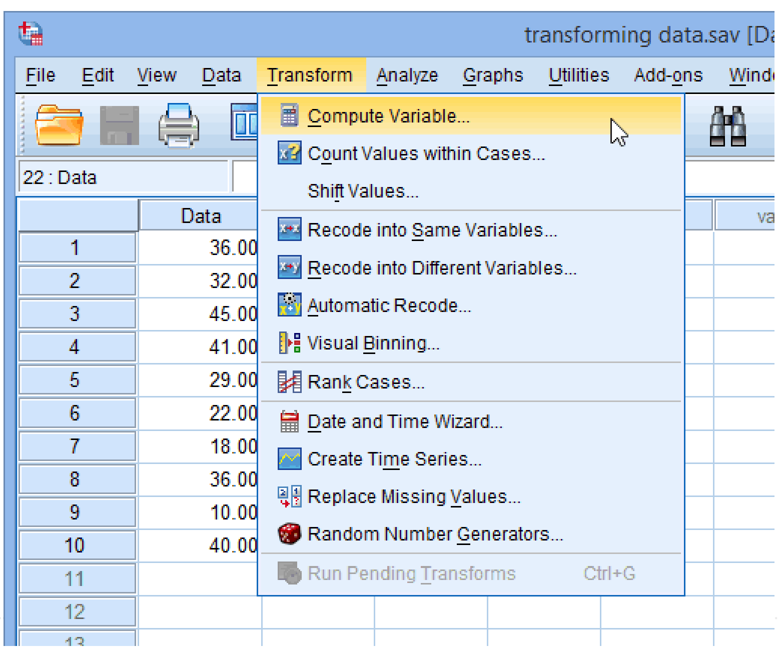
Transforming Data by Compute Action

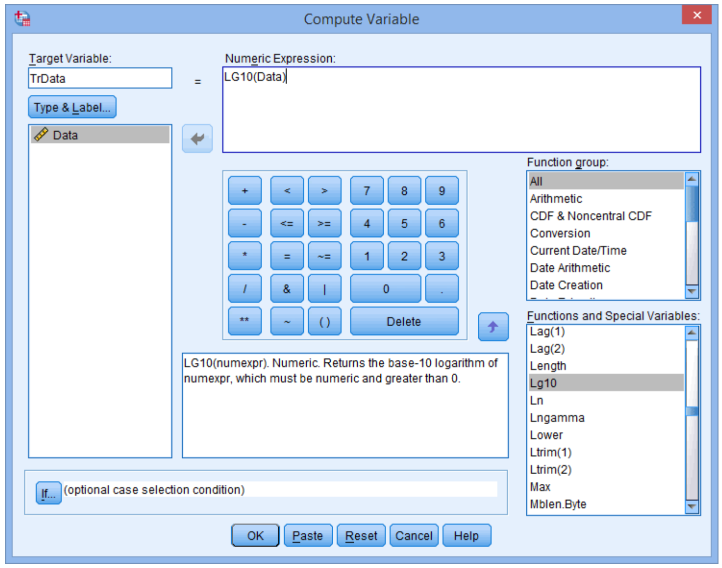
In this "quick start" guide, we will enter some data and then perform a transformation of the data. Transforming data is performed for a whole host of different reasons. We show you how to transform your data using SPSS Statistics for "square", "square root", "reflect and square root", "reflect and log", "reciprocal", "reflect and inverse" and "log" transformations. Also, we 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

Transforming Data- Compute in SPSS Statistics

Click on **Transform > Compute Variable...** in the top menu, as shown below:



You need to first select the function you would like to use. To do this, click "**All**" in the Function group: box or you can write a mathematics expression to transform the given data into the required form. In the Target Variable, write the name of the new variable and also you can make a new label and type.



The result will be new variable in new column as follows

	Data	TrData
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

Recoding Variables

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

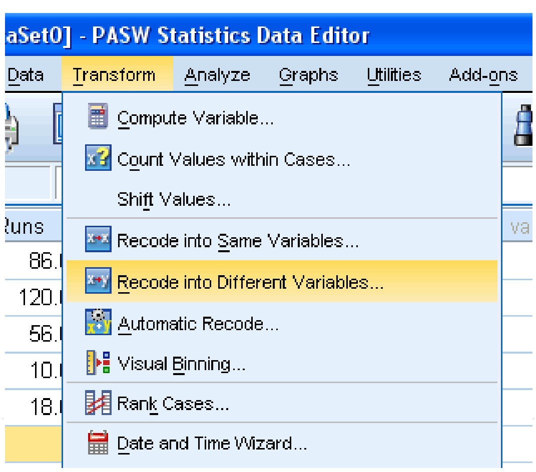
- Recoding into the same variable
- Recoding into a different variable

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

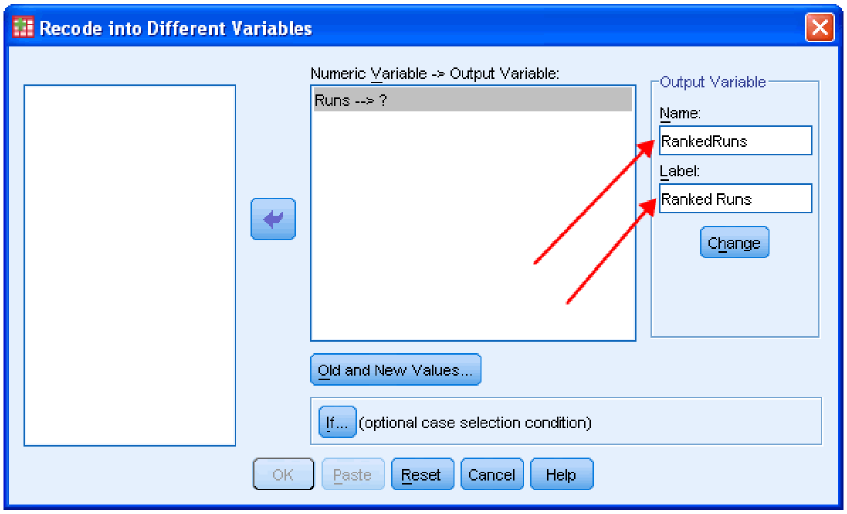
- Recode single values
- Recode a given range of values
- Recode data into two categories

Recoding Variables in SPSS Statistics

Click on **Transform > Recode Into Different Variables**

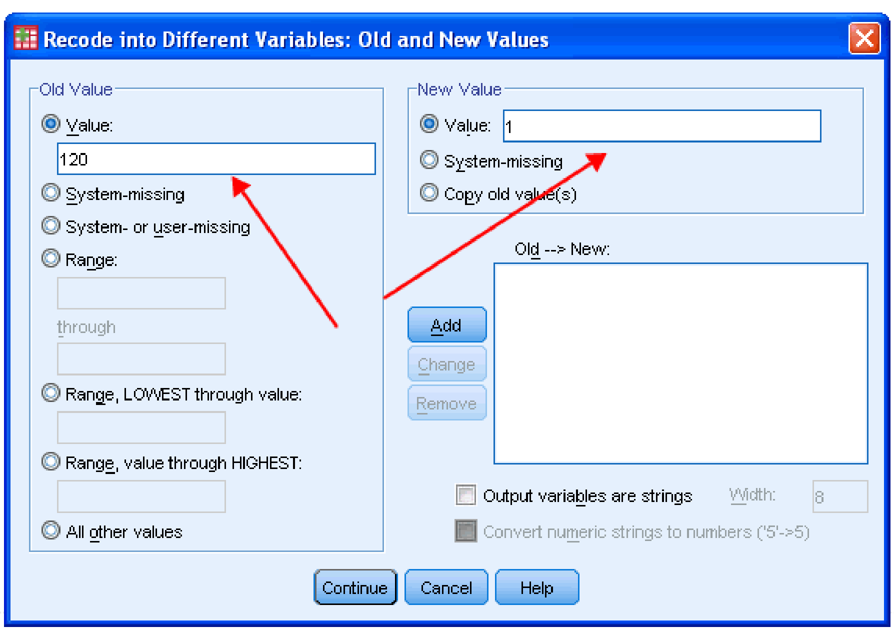


We have to write down the new name and label for the new variable as below, then click change to get it.

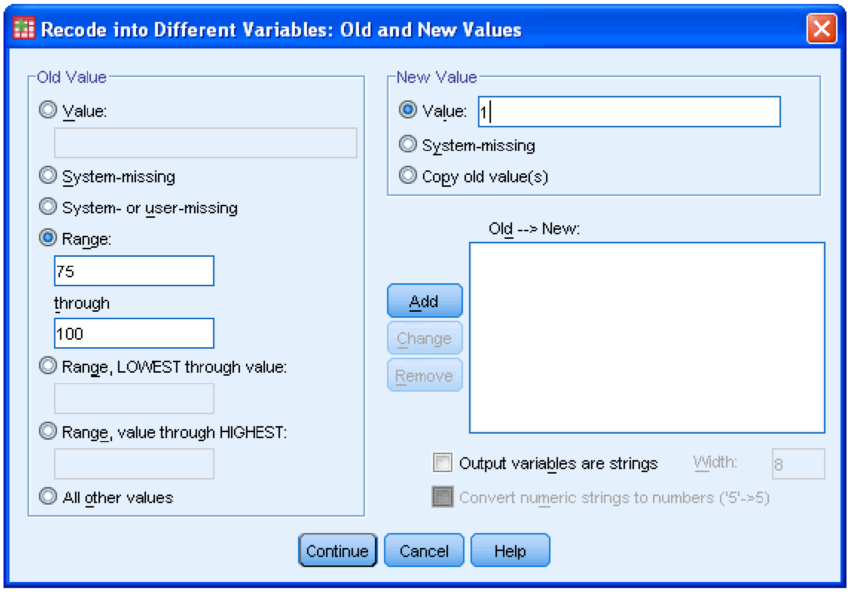


And, in **Old and New values,** we get a box to change old values into single value or we can use a range of data and transform them into new categorical data. A good example about this, if we have continues variable like AGE and we want to conduct a categorical analysis; we can recode it into many categories using ranges.

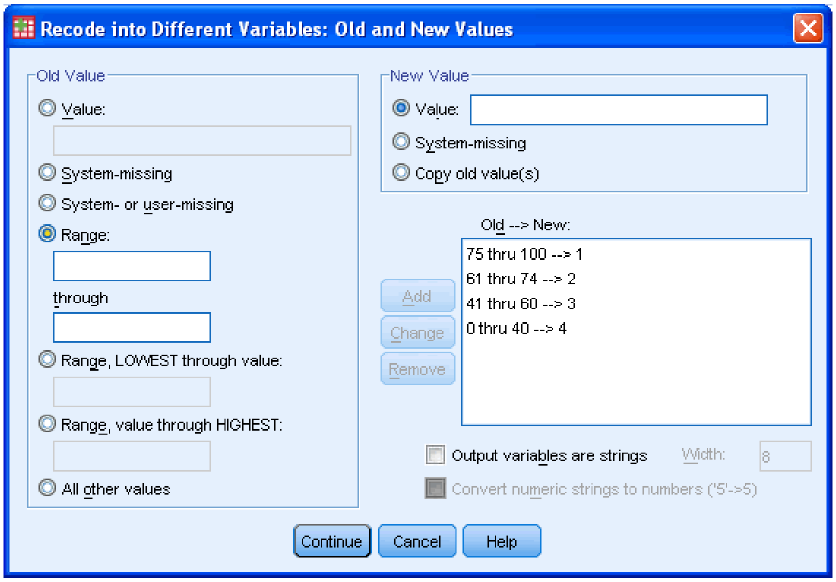
This is a single recode.



The range recodes,



Continue adding and making categories



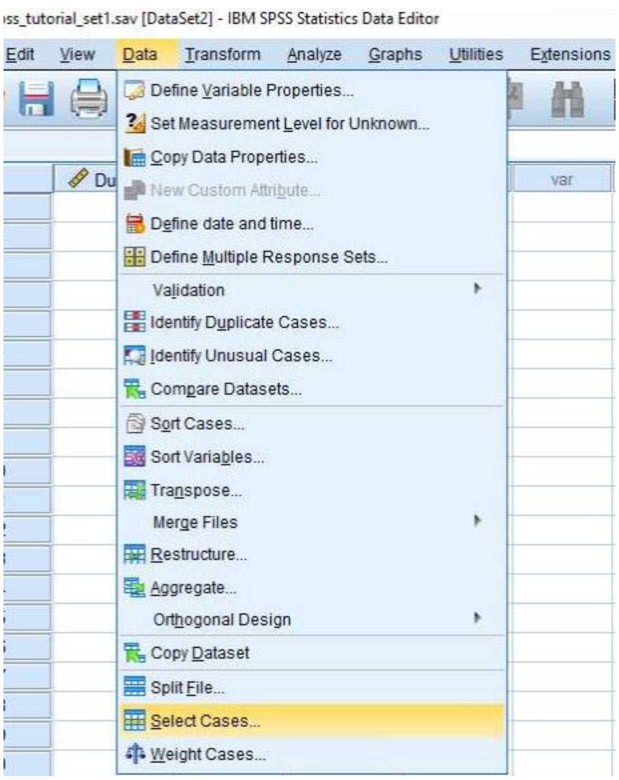
Then click, continue to get the result.

Select Cases form Data Set

You rarely want to remove/delete data from your SPSS Statistics file, but there are a number of occasions when you need to filter that data before analysing it (e.g., to take into account missing data or outliers in your data set that could be negatively affecting your results). We show you, step-by-step, how to select different types of data (i.e., specific cases) and then filter your data set.

Select Cases function in SPSS Statistics.

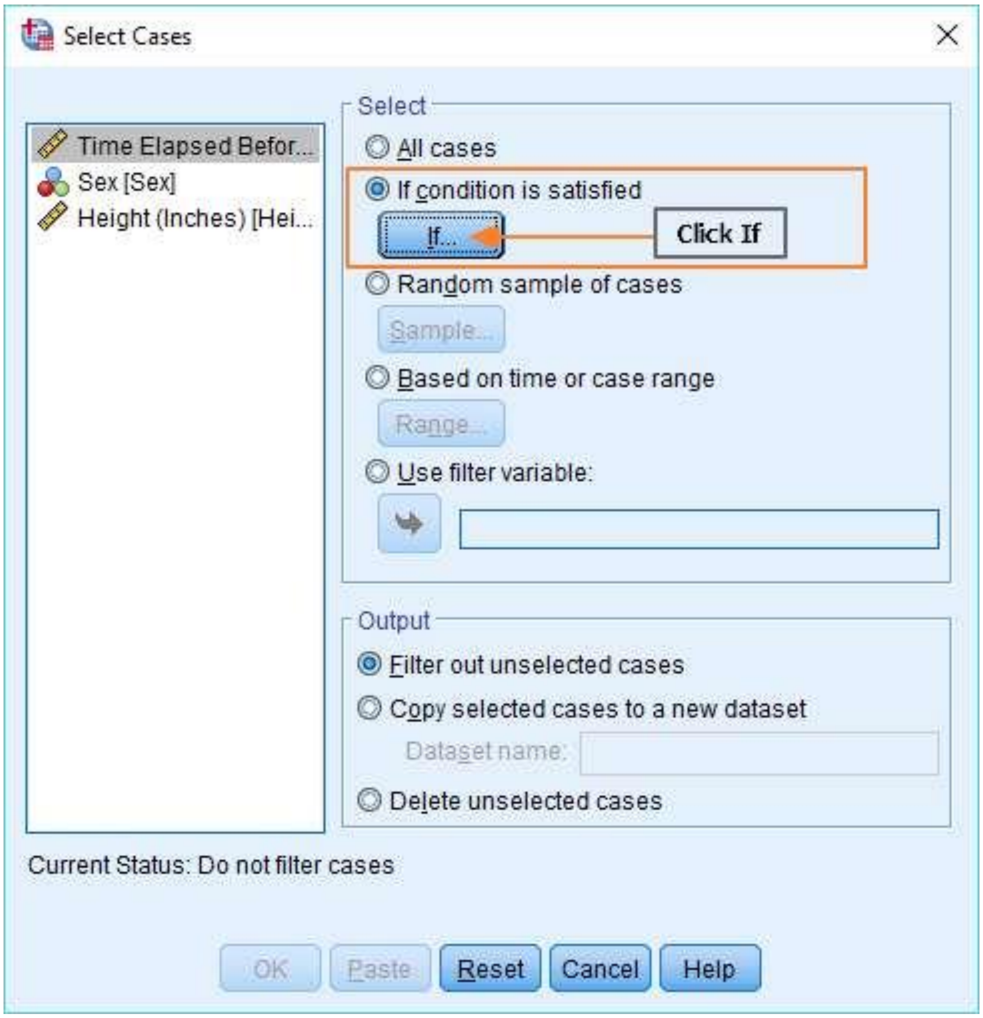
Click **Data** > **Select Cases ...** on the main menu as shown below



The Functions for expression dialog allows you to select subsets of cases using conditional expressions. A conditional expression returns a value of true, false, or missing for each case.

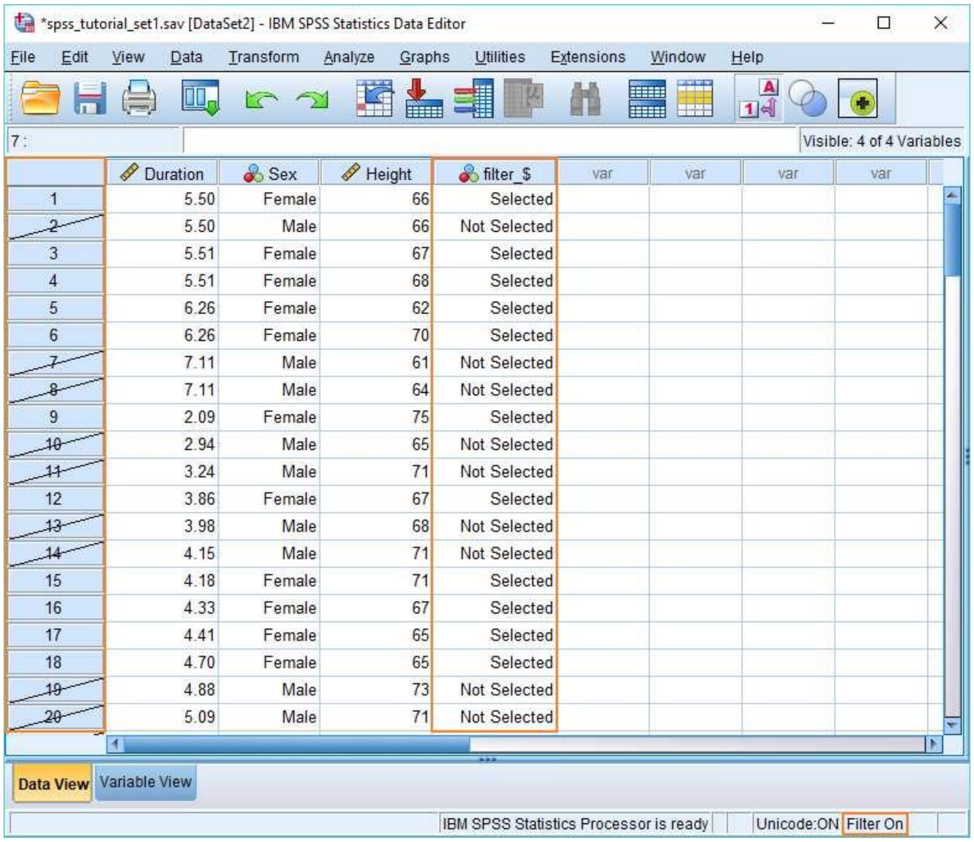
- If the result of a conditional expression is true, the case is included in the selected subset.

- If the result of a conditional expression is false or missing, the case is not included in the selected subset.



Most conditional expressions use one or more of the six relational operators (<, >, <=, >=, =, and ~=) on the calculator pad.

Conditional expressions can include variable names, constants, arithmetic operators, numeric (and other) functions, logical variables, and relational operators.



	Duration	Sex	Height	filter_$
1	5.50	Female	66	Selected
2	5.50	Male	66	Not Selected
3	5.51	Female	67	Selected
4	5.51	Female	68	Selected
5	6.26	Female	62	Selected
6	6.26	Female	70	Selected
7	7.11	Male	61	Not Selected
8	7.11	Male	64	Not Selected
9	2.09	Female	75	Selected
10	2.94	Male	65	Not Selected
11	3.24	Male	71	Not Selected
12	3.86	Female	67	Selected
13	3.98	Male	68	Not Selected
14	4.15	Male	71	Not Selected
15	4.18	Female	71	Selected
16	4.33	Female	67	Selected
17	4.41	Female	65	Selected
18	4.70	Female	65	Selected
19	4.88	Male	73	Not Selected
20	5.09	Male	71	Not Selected

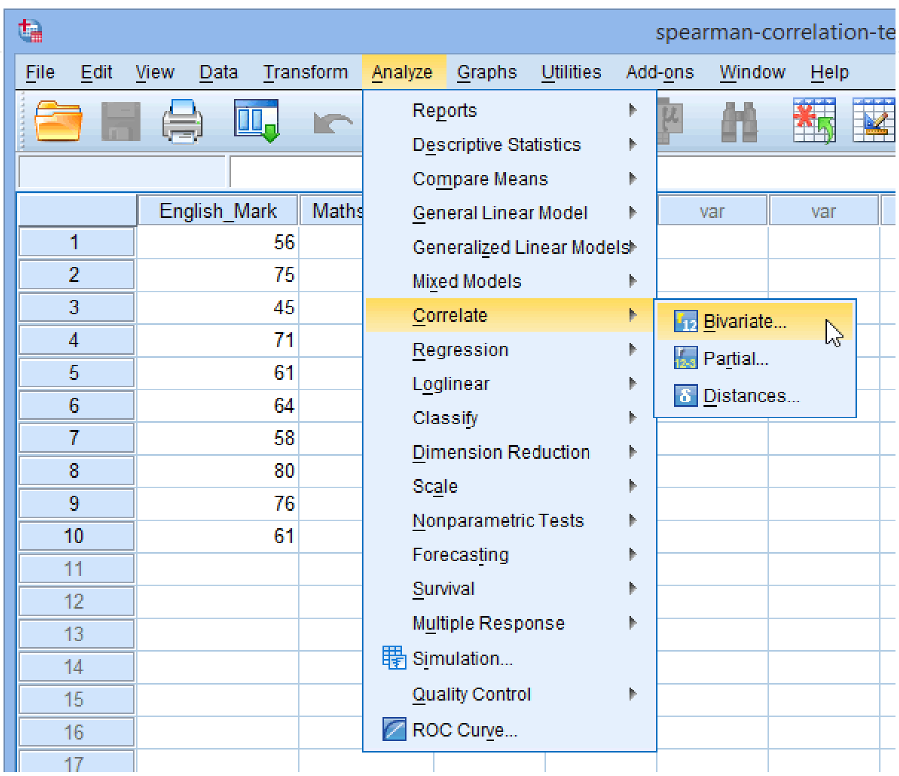
Spearman's Rank-Order Correlation

The Spearman rank-order correlation coefficient (Spearman's correlation, for short) is a nonparametric measure of the strength and direction of association that exists between two variables measured on at least an ordinal scale.

For example, you could use a Spearman's correlation to understand whether there is an association between exam performance and time spent revising; whether there is an association between depression and length of unemployment; and so forth.

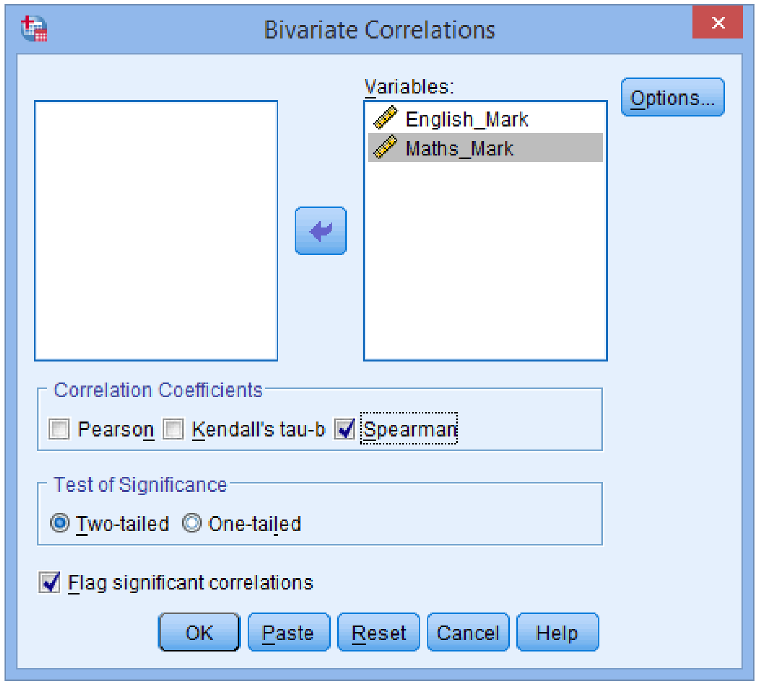
Spearman's correlation analysis in SPSS Statistics

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



You will be presented with the **Bivariate Correlations** dialogue box:

Select the Spearman checkbox in the –Correlation Coefficients– area. You will end up with a screen similar to below:



Therefore, after running the Spearman's correlation procedure, you will be presented with the **Correlations** table, as shown 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).

The **Correlations** table presents **Spearman's correlation**, its **significance value** (i.e., ***p*-value**) and the **sample size** that the calculation was based on.

In this example, the sample size, N, is **10**, Spearman's correlation coefficient, r_s, is **0.669**, which **is** statistically significant (***p* = .035**).

Reporting the results of a Spearman's correlation analysis

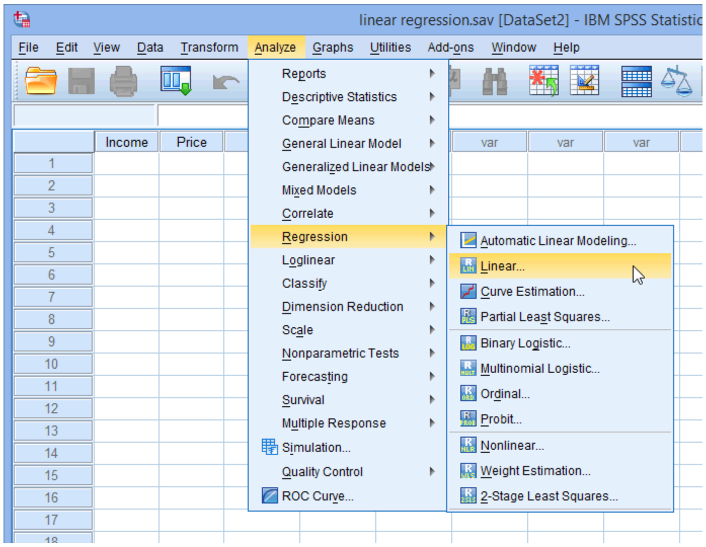
A Spearman's rank-order correlation was run to determine the relationship between 10 students' English and maths exam marks. There was a strong, positive correlation between English and maths marks, which was statistically significant ($r_s(8) = .669$, $p = .035$).

Linear Regression Analysis using SPSS Statistics

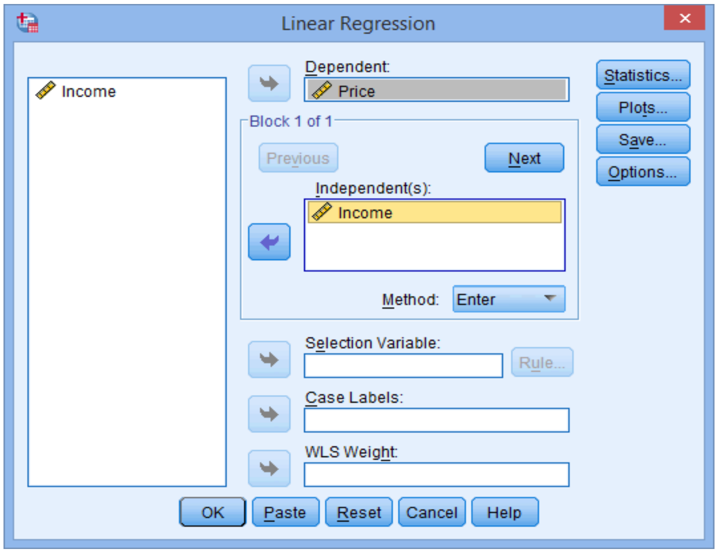
Linear regression is the next step up after correlation. It is used when we want to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable (or sometimes, the outcome variable). The variable we are using to predict the other variable's value is called the independent variable (or sometimes, the predictor variable). For example, you could use linear regression to understand whether exam performance can be predicted based on revision time; whether cigarette consumption can be predicted based on smoking duration; and so forth.

Linear Regression in SPSS Statistics

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



You will be presented with the **Linear Regression** dialogue box, then Transfer the independent variable into the Independent(s): box and the dependent variable into the Dependent: box.



Output of Linear Regression Analysis

The first table is the **Model Summary** table, as shown below:

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

This table provides the ***R*** and $\boldsymbol{R^2}$ values. The *R* value represents the simple correlation and is 0.873 (the "**R**" Column), which indicates a high degree of correlation. The *R*2 value (the "**R Square**" column) indicates how much of the total variation in the dependent variable, Price, can be explained by the independent variable, Income. In this case, 76.2% can be explained, which is very large.

The next table is the **ANOVA** table, which reports how well the regression equation fits the data (i.e., predicts the dependent variable) and is shown below:

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	44182633.37	1	44182633.37	57.737	.000[b]
	Residual	13774291.07	18	765238.393		
	Total	57956924.44	19			

a. Dependent Variable: Price

b. Predictors: (Constant), Income

This table indicates that the regression model predicts the dependent 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 run. Here, $p < 0.0005$, which is less than 0.05, and indicates that, overall, the regression model statistically significantly predicts the outcome variable (i.e., it is a good fit for the data).

The **Coefficients** table provides us with the necessary information to predict price from income, as well as determine whether income contributes statistically significantly to the model (by looking at the "**Sig.**" column). Furthermore, we can use the values in the "**B**" column under the "**Unstandardized Coefficients**" column, as shown below:

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

To present the regression equation as:

Price = 8287 + 0.564(Income)