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:

ta 🛛						t	ransform	ing dat	a.sav [Da
<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> n	is <u>W</u> ind
		Ę		Compu Count	ute Variabl Values wit	e hin Cases		3	212
22:0	ata			Shi <u>f</u> t Va	alues				
<u> </u>		1 [)ata	🔤 Recod	e into <u>S</u> am	e Variable	S		va
	1		36.00	Recod	e into Diffe	rent Variat	oles		
	2		32.00	Autom:	ntic Docod				
	3		45.00			e		_	
	4		41.00	Visual	<u>B</u> inning				
	5		29.00	🛃 Ran <u>k</u> C	Cases				
	6		22.00	🗎 Date a	nd Time W	/izard			
	7		18.00	Create	Time Seri	es			
	8		36.00	Bil Boolog	Missing	Voluee			
	9		10.00	G Replac	e missing	values			
1	10		40.00	👹 Rando	m Numbei	r <u>G</u> enerato	rs		
1	11			📕 🐻 Run Pe	ending <u>T</u> ra	nsforms	Ctrl+	G	
1	2								
1	3								

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.

ta	Compute Variable	×
Target Variable: TrData Type & LabeL Data	Compute Variable Numeric Expression: = LG10(Data) + > + > + > + > + > + > + > + > + > + = + > + = + > + = + = + + + + + + + + + + - 1 - 1 - 1 - 1 - 1 - 1 - - - - - - - - - - - - - - - - - -	×
(optional case select	ion condition) Max Mblen.Byte	•
	OK Paste Cancel Help	

The result will be new variable in new column as follows

	Data	TrData	V
	Data	nData	v
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.

🔢 Recode into Different Va	iables 🛛
	Numeric Variable -> Output Variable: Output Variable: Runs> ? Name: RankedRuns Label: Ranked Runs Change Old and New Values If (optional case selection condition)
	OK Paste Reset Cancel Help

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.

III Recode into Different Variables: Ol	d and New Values 🛛 🔀							
Recode into Different Variables: Old Old Value Value: 120 System-missing Range: through Range, LOWEST through value: Range, value through HIGHEST:	d and New Values							
◯ All <u>o</u> ther values	Output variables are strings Width: B Convert numeric strings to numbers ('5'->5)							
Continue Cancel Help								

This is a single recode.

The range recodes,

Recode into Different Variables: OI	ld and New Values 🛛 🔀
Recode into Different Variables: 01 Old Value Yalue: System-missing System- or user-missing Range: 75 through 100 Range, LOWEST through value: Range, value through HIGHEST: All other values	d and New Values Image: New Value Image: Value Image: Old> New: Image: Remove Image: Remove Image: Output variables are strings Image: Output variables are strings to numbers ('5'->5)
Continu	e Cancel Help

Continue adding and making categories

Recode into Different Variables: Ol	d and New Values 🛛 🛛 👌	K					
Recode into Different Variables: OI Old Value Value: System-missing System-or user-missing Range: through Range, LOWEST through value: Range, value through HICHEST: All other values	d and New Values Image: System-missing Imag						
Continue Cancel Help							

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 **<u>Data</u>** > <u>Select Cases</u> ... 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.

 Time Elapsed Befor Sex [Sex] Height (Inches) [Hei 	Select O All cases If condition is satisfied If condition is satisfied
	Output

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

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

Edi	t <u>V</u> iew	Data	Transform	<u>Analyze G</u> i	raphs	Utilities	Extensions	s <u>W</u> in	dow	Help			
			5 2				H				0	۲	
			1000							14	Visib	le: 4 of 4 \	/ariat
	🖉 Di	iration	💑 Sex	🥜 Height		💦 filter_\$	var		var	V	ar	var	
1		5.50	Female		66	Selecte	d						1.5
2-	_	5.50	Male)	66 1	Not Selecte	d						
3		5.51	Female		67	Selecte	d						
4		5.51	Female	- (68	Selecte	d						
5		6.26	Female		62	Selecte	d						
6		6.26	Female		70	Selecte	d						
7-	-	7.11	Male		61	Not Selecte	d						
-8		7.11	Male)	64	Not Selecte	d						
9		2.09	Female		75	Selecte	d						
_10	-	2.94	Male)	65 1	Not Selecte	d						
	-	3.24	Male		71 1	Not Selecte	d						
12		3.86	Female	- (67	Selecte	d						
_13	-	3.98	Male	1	68	Not Selecte	d						
_14	-	4.15	Male		71 1	Not Selecte	d						
15		4.18	Female		71	Selecte	d						
16		4.33	Female)	67	Selecte	d						
17		4.41	Female		65	Selecte	d						
18		4.70	Female		65	Selecte	d						
19-	-	4.88	Male	1	73	Not Selecte	d						
_20	-	5.09	Male		71	Not Selecte	d						
	<u> </u>	_											•
ta Vicu	Variable	liew											

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 <u>Analyze > Correlate > Bivariate...</u> on the main menu as shown below:

ta 🛛							spea	rman-co	rrelati	on-te
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata <u>T</u> rar	nsform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Ado	l- <u>o</u> ns	<u>W</u> indow	<u>H</u> elp)
		IC.	Rep	orts		•	ų	14	*	4
			D <u>e</u> s	criptive Sta	atistics					
	English Mark	Mathe	Con	npare mea	ns r Model	- P-		/ar	var	
1	56	Matria	Gen	eralized Li	i Model	l ob	`	/ai	vai	
2	75		Mive		inear would	h l				
3	45		Corr	relate				Diveriat		
4	71		Reg	ression			12	Bivariate	3	
5	61			linear			12-3	Pa <u>r</u> tial		
6	64		Clas	ssifv			δ	<u>D</u> istances		
7	58		Dim	ension Re	duction					
8	80		Scal	le						
9	76		Non	parametri	c Tests	•				
10	61		 Fore	ecasting						
11			Surv	/ival		•				
12			Mult	iple Resp	onse	•				
13			🐺 Sim	ulation						
15			Qua	lity Contro	I					
16			ROC	C Curve						
17				-						

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

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

Variables:							
Pearson Kendall's tau-b Spearman							
r Test of Significance							
Test of Significance							

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

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

Correlations

*. 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 <u>Analyze > Regression > Linear...</u> on the top menu,



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

Linear Regression				
Income	Dependent: Price Price Previous Next Independent(s): Method: Enter Selection Variable: Case Labels: WLS Weight: WLS Weight: Paste Reset Cancel Help	Statistics Plots Save Options		

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 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 R2 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:

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			

ANOVA^a

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:

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

Coefficients^a

a. Dependent Variable: Price

To present the regression equation as:

Price = 8287 + 0.564(Income)