Regression

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)

بەکاردێ بۆ زانینی کاریگەری گۆڕاوی سەربەخۆ (X) لەسەر گۆڕاوی پاشکۆ (y)

مەرجە سەرەکیەکانی ئەنجامدانی REGRESSION

Assumption 1: Your two variables should be measured at the **continuous** level (they are either **interval** or **ratio** variables)

Assumption 2 There needs to be a **linear relationship** between the two variables

Assumption 3: There should be **no significant outliers**

Assumption4 : You should have **independence of observations**

Assumption 5: Your data needs to show **homoscedasticity**

Assumption 6: you need to check that the **residuals (errors)** of the regression line are **approximately normally distributed**

Simple Linear Regression

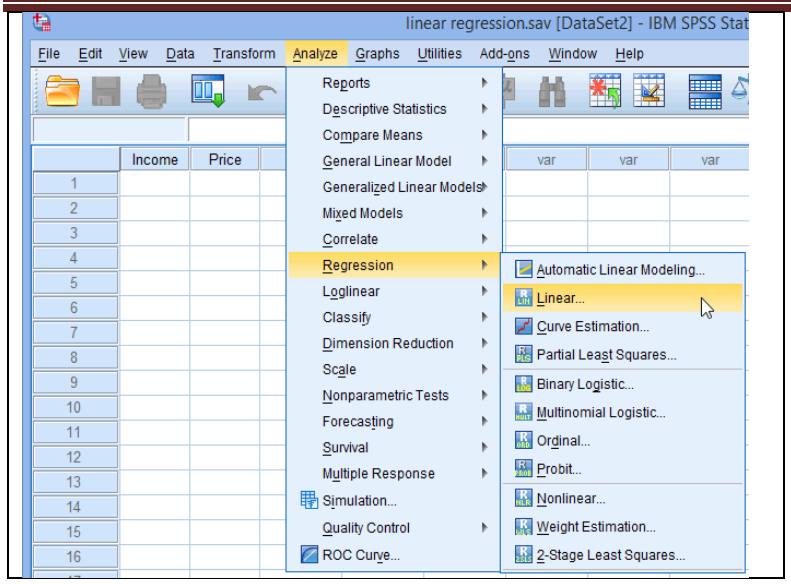
$$Yi = \beta 0 + \beta\ 1Xi + \varepsilon i,$$

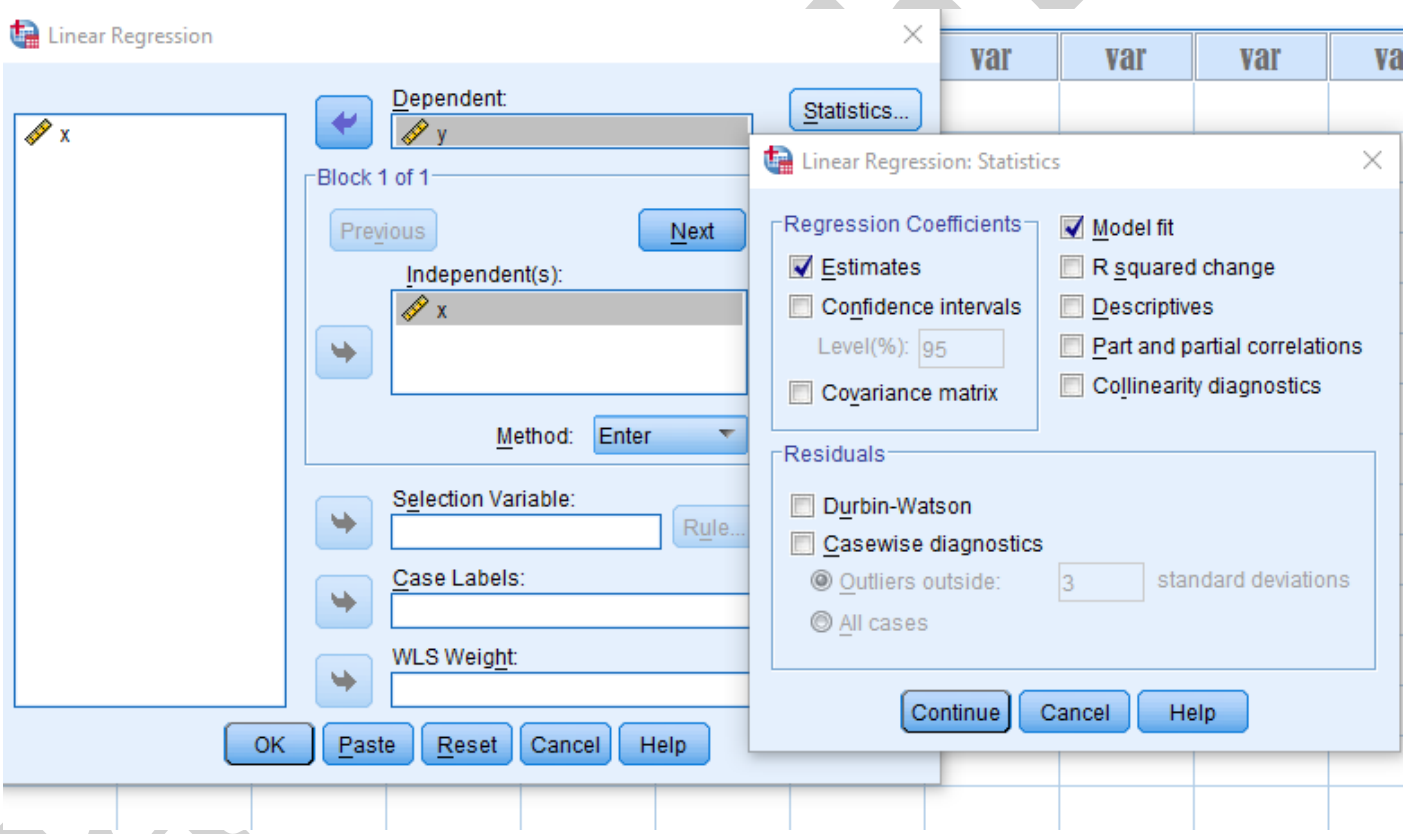
To perform a simple linear regression in SPSS

- Choose **Analyze → Regression→ Linear**.
- Move the DV to the **Dependent** box.
- Move the IV to the **Independent(s)** box.
- Click the **Continue** button.
- Click the **OK** button.

Example\\ find regression between x , y

X	Y
6	82
2	86
15	43
9	74
12	58
5	90
8	78
6	80
12	50
9	54
9	56
10	52
2	90
1	97
3	88





CONTINUE ---------OK ------RESULT

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

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.923[a]	.852	.840	7.104

a. Predictors: (Constant), x

The next table is the **ANOVA** table, which reports how well the regression equation fits the data

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3773.638	1	3773.638	74.772	.000[b]
	Residual	656.095	13	50.469		
	Total	4429.733	14			

a. Dependent Variable: y

b. Predictors: (Constant), x

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)	100.507	3.786		26.546	.000
	X	-3.941	.456	-.923	-8.647	.000

a. Dependent Variable: y

to present the regression equation as:

$$y = 100.507 - 3.941(x)$$

Multiple Regression

زیاتر لەیەك گۆڕاوی سەربەخۆمان هەبوو

The general form of the multiple regression model is

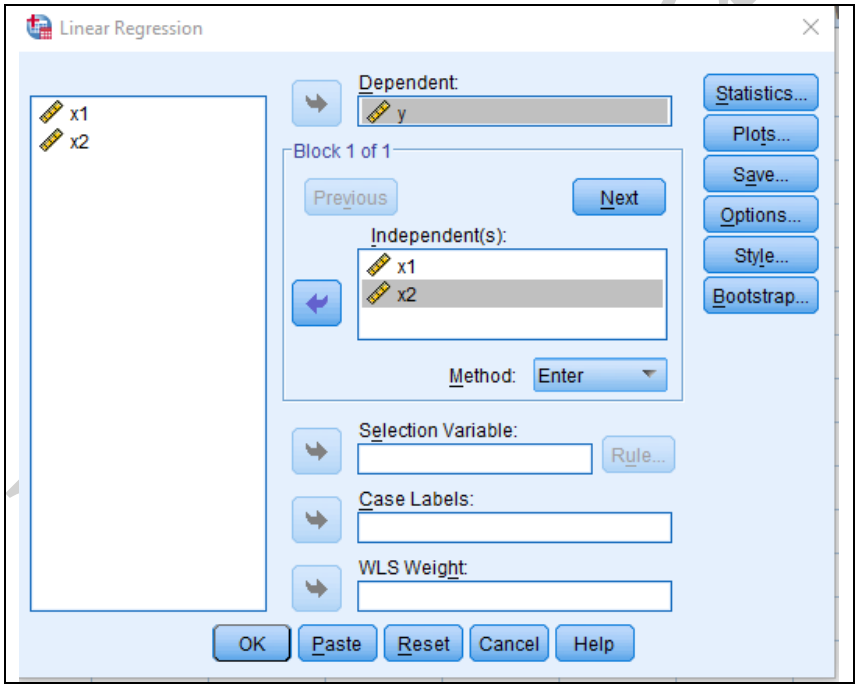
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ + \beta_k X_{ik} + \varepsilon_i,.$$

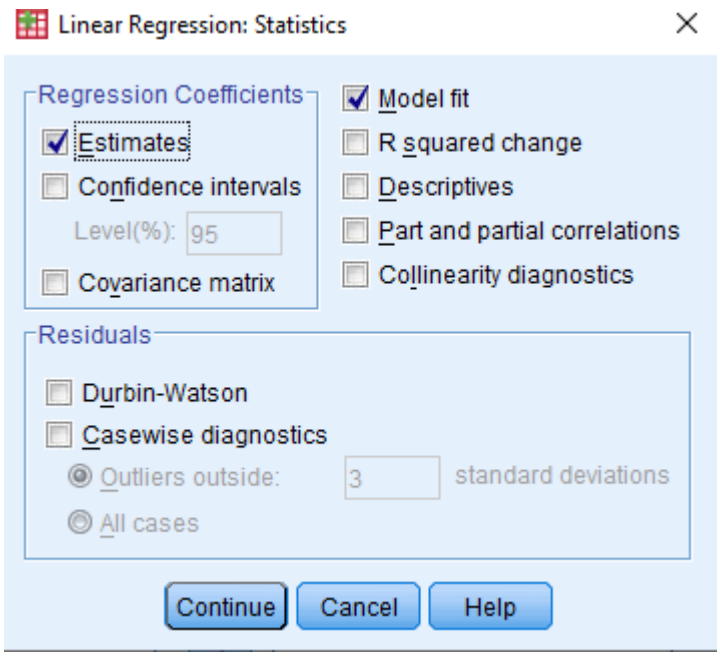
To perform a multiple regression in SPSS

- Choose **Analyze → Regression → Linear**.
- Move the DV to the **Dependent** box.
- Move all of the IVs to the **Independent(s)** box.
- Click the **Continue** button.
- Click the **OK** button.

Example\\ find multiple regression between y and x1, x2

	y	x1	x2
1	58	89	58
2	25	52	59
3	48	47	57
4	79	58	58
5	96	93	95
6	85	95	96
7	89	56	93
8	95	95	94
9	65	65	45
10			





CONTINUE ------OK – RESULT

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.765[a]	.585	.447	17.901

a. Predictors: (Constant), x2, x1

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2712.235	2	1356.117	4.232	.071[b]
	Residual	1922.654	6	320.442		
	Total	4634.889	8			

a. Dependent Variable: y

b. Predictors: (Constant), x2, x1

Coefficients[a]

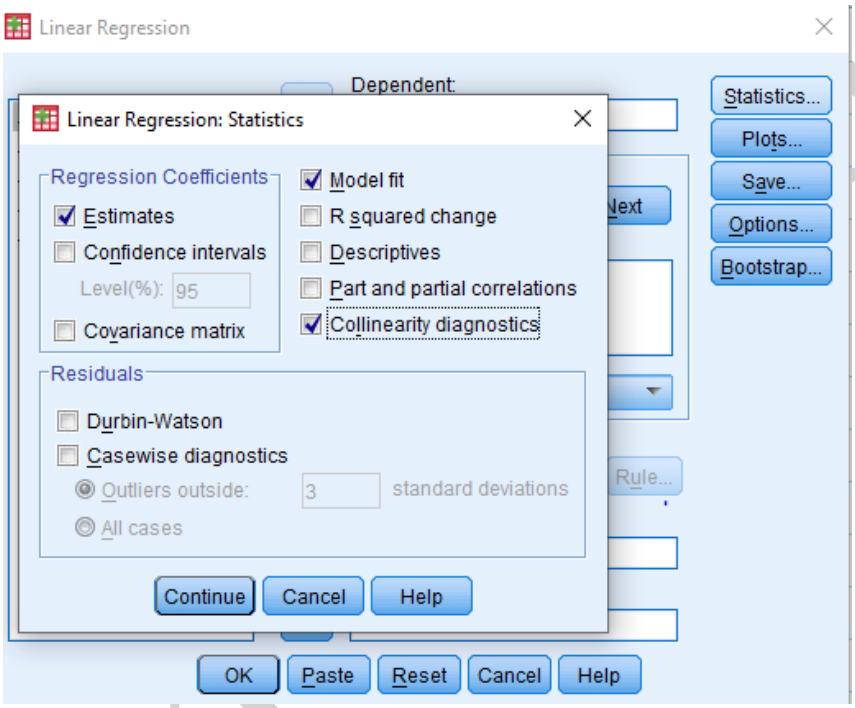
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.517	25.666		.020	.985
	x1	.290	.382	.245	.759	.477
	x2	.683	.369	.597	1.849	.114

a. Dependent Variable: y

$$Y= 0.517 +0.290x_1+0.683x_2$$

Multi collinearity

Multi collinearity is a problem when for any predictor the R^2 between that predictor and the remaining predictors is very high. Upon request, SPSS will give you two transformations of the squared multiple correlation coefficients. One is tolerance, which is simply 1 minus that R^2. The second is VIF, the **variance inflation factor**, which is simply the reciprocal of the tolerance. Very low values of tolerance (.1 or less) indicate a problem. Very high values of VIF (10 or more, although some would say 5 or even 4) indicate a problem. As you can see in the table below, we have no multi collinearity problem here.



Continue ---------OK--------RESULT

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Age	.980	1.021
	Conscientiousness	.980	1.021

تیبینی " ئەگەر هاتوو بەهای VIFلە 10کەمتر بوو واتە کیشەی Collinearityمان نیە بەلامگەورەتر بوو لە 10 ئەو کیشەیەمان هەیە پیویستە چارەسەری بکەین