

What is the role of the Remove method in Multiple Regression?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the role of the Remove method in Multiple Regression?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=149754>

The Remove method in Multiple Regression is a statistical technique used to determine the most influential independent variables in a regression model. It is a stepwise process that involves removing one independent variable at a time and observing the effect on the model's overall performance. This method helps researchers identify the most significant predictors and eliminates irrelevant variables, thus improving the accuracy and reliability of the regression model. The Remove method is an essential tool for simplifying and interpreting complex regression models and is widely used in various fields such as economics, psychology, and social sciences.

Remove method of Multiple Regression

In this method, we will learn the Remove method, and it is the last method. In the case of remove method, we can specify a variable that needs to be removed from our side. Keep this in mind that this remove method cannot be used alone. We use the remove method of regression in conjunction with or in association with any other method, typically the Enter method. First, we enter all the variables in the model, and as a second step, we try to eliminate some variables based on their less contribution to the total variance. For example, in the following case, we will first enter all the variables in the model following the Enter method.

The screenshot shows the IBM SPSS Statistics Data Editor interface. A dialog box titled "Linear Regression" is open, allowing the user to configure a regression model. The dependent variable is "Current Salary [salary]". The independent variables are "Educational Level [educ]", "Employment Category [jobcat]", "Beginning Salary [salbegin]", and "Previous Experience [prevexp]". The method is set to "Enter". The dialog box also includes options for "Selection Variable", "Case Labels", and "WLS Weight".

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role	
1	id	Numeric	4	0	Employee Code	None	None	8	Right	Scale	Input
2	gender	String								Input	Input
3	educ	Numeric								Input	Input
4	jobcat	Numeric								Input	Input
5	salary	Dollar								Input	Input
6	salbegin	Dollar								Input	Input
7	jobtime	Numeric								Input	Input
8	prevexp	Numeric								Input	Input
9	minority	Numeric								Input	Input
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											

Now we will find out which variable is least important for explaining the total variance. It seems, in the coefficients table, all the variables are good predictors. T values are quite high.

Output2 [Document2] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help

Log
Regression
Title
Notes
Active Dataset
Variables Entered/Removed
Model Summary
ANOVA
Coefficients

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.955 ^a	.912	.904	\$10,214.743

a. Predictors: (Constant), Previous Experience, Beginning Salary, Educational Level, Employment Category

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.868E+10	4	1.217E+10	116.637	.000 ^b
	Residual	4695343461	45	104340965.8		
	Total	5.338E+10	49			

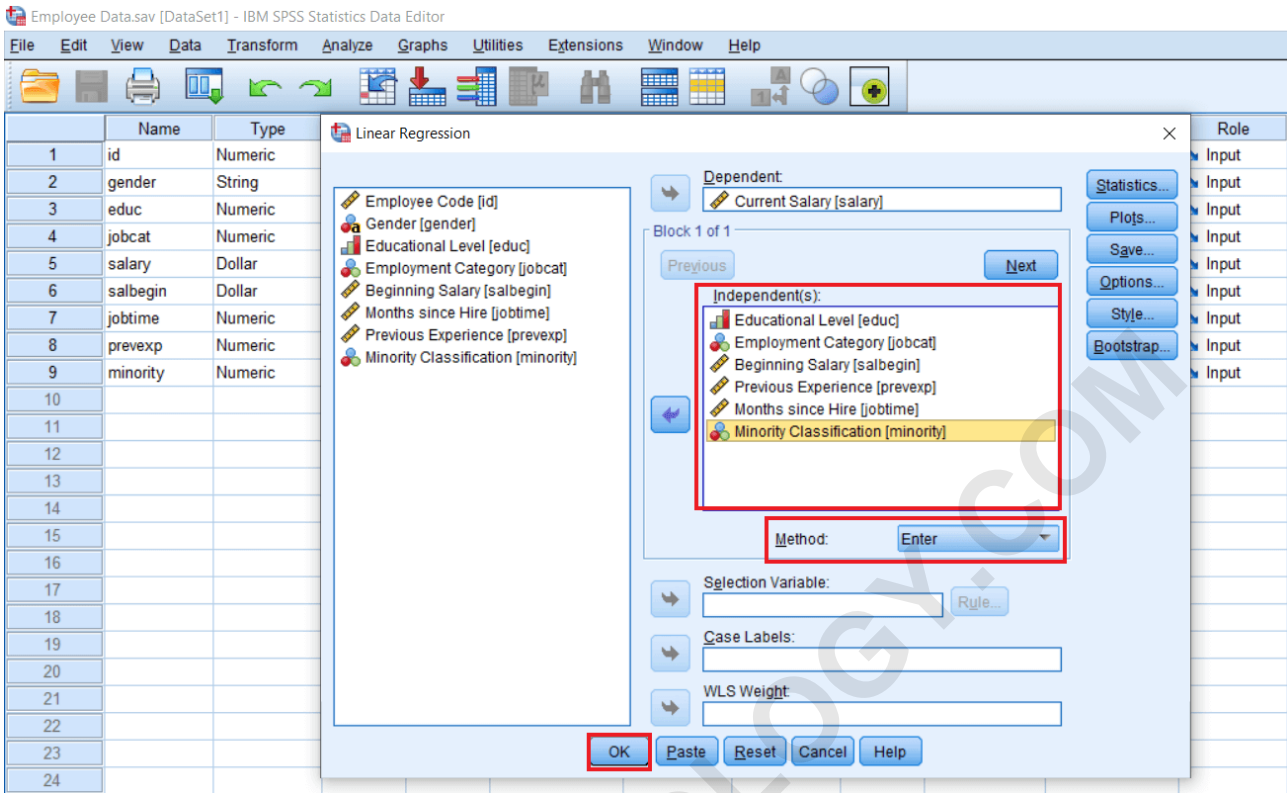
a. Dependent Variable: Current Salary
b. Predictors: (Constant), Previous Experience, Beginning Salary, Educational Level, Employment Category

Coefficients^a

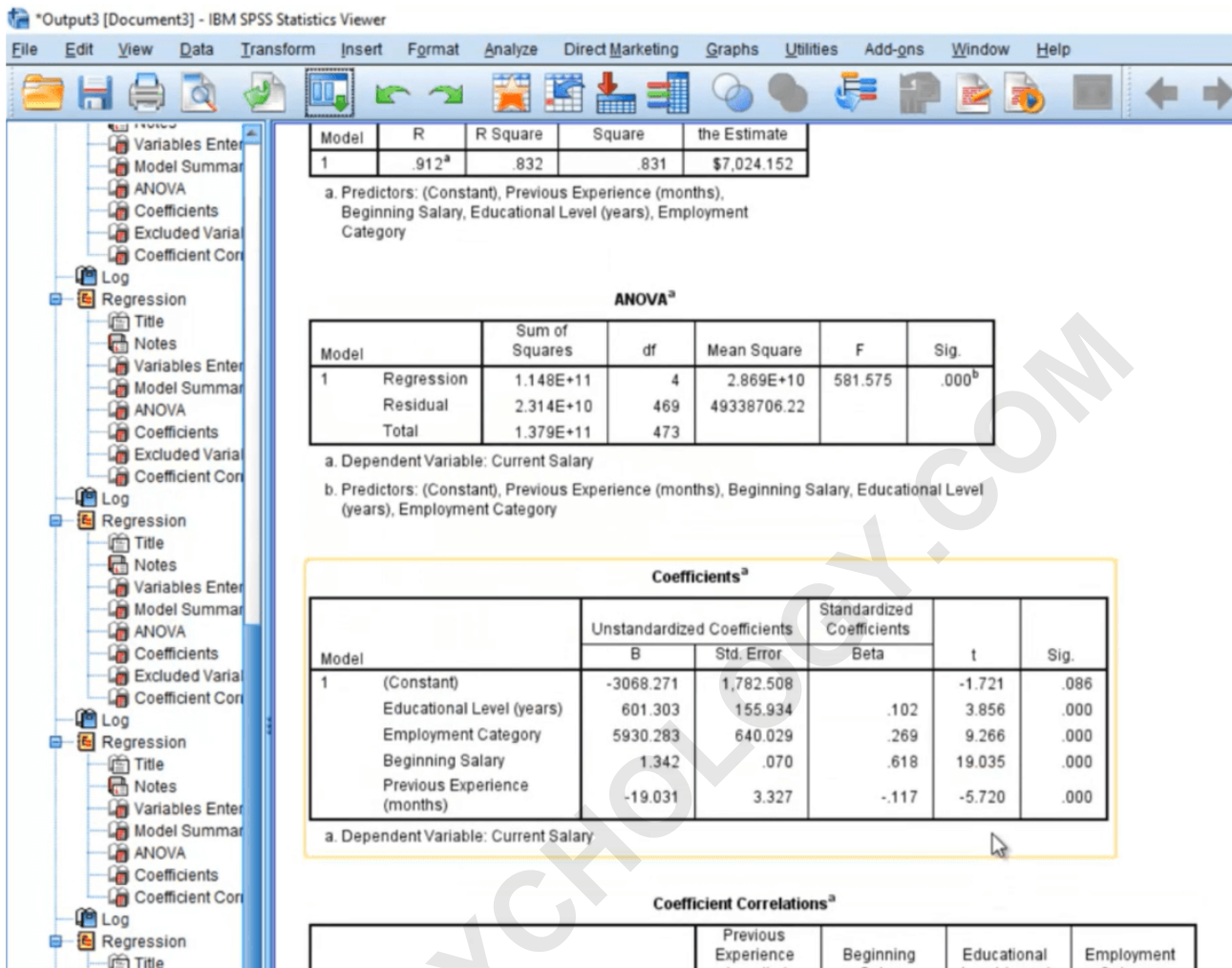
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4974.500	6423.437		.774	.443
	Educational Level	-173.681	438.974	-.021	-.396	.694
	Employment Category	18978.073	2649.744	.523	7.162	.000
	Beginning Salary	.631	.099	.503	6.403	.000
	Previous Experience	-39.238	14.910	-.122	-2.632	.012

a. Dependent Variable: Current Salary

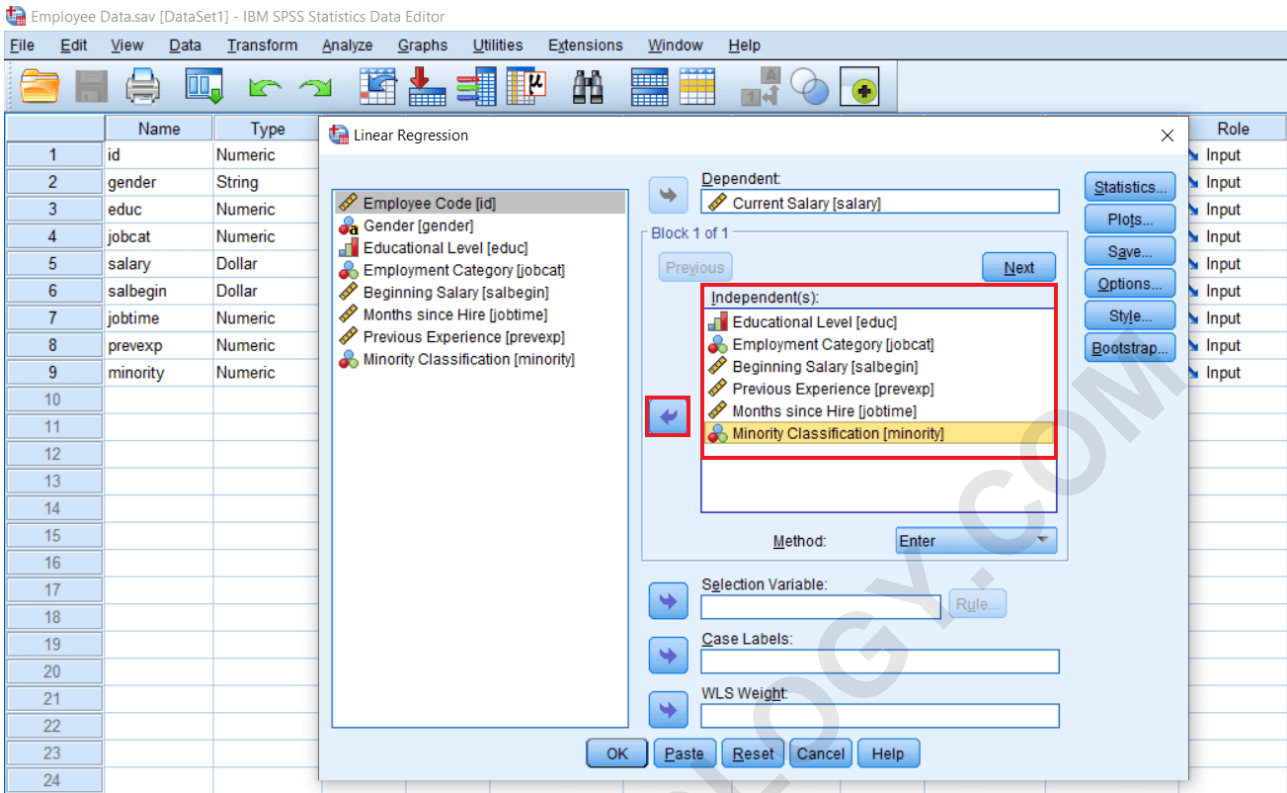
So let's take some other variable, which might not be related that well. Let's take months since hire and minority classification. Now out of these variables, we are guessing minority classification to be less important for predicting the current salary and click Ok like this:



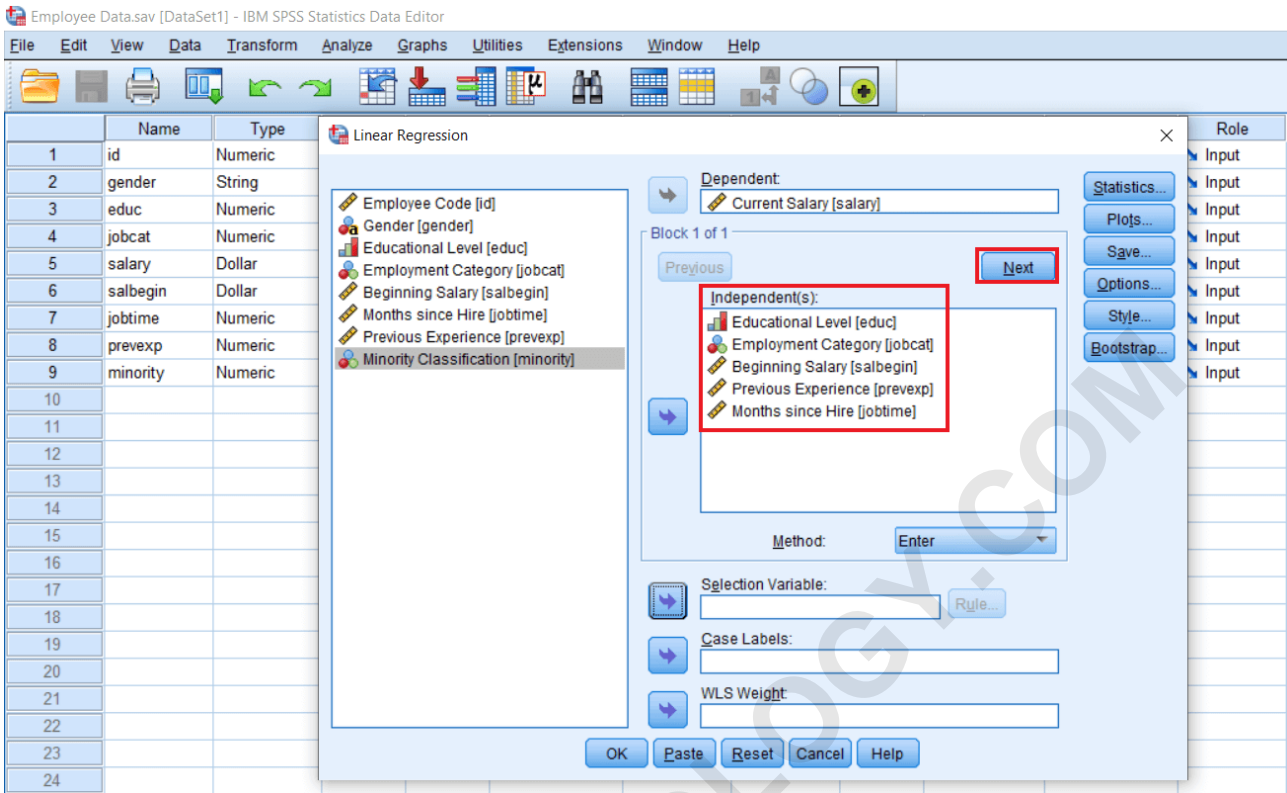
All the variables have been entered. Now we can see, in the coefficient table, minority classification is getting a non-significant value.



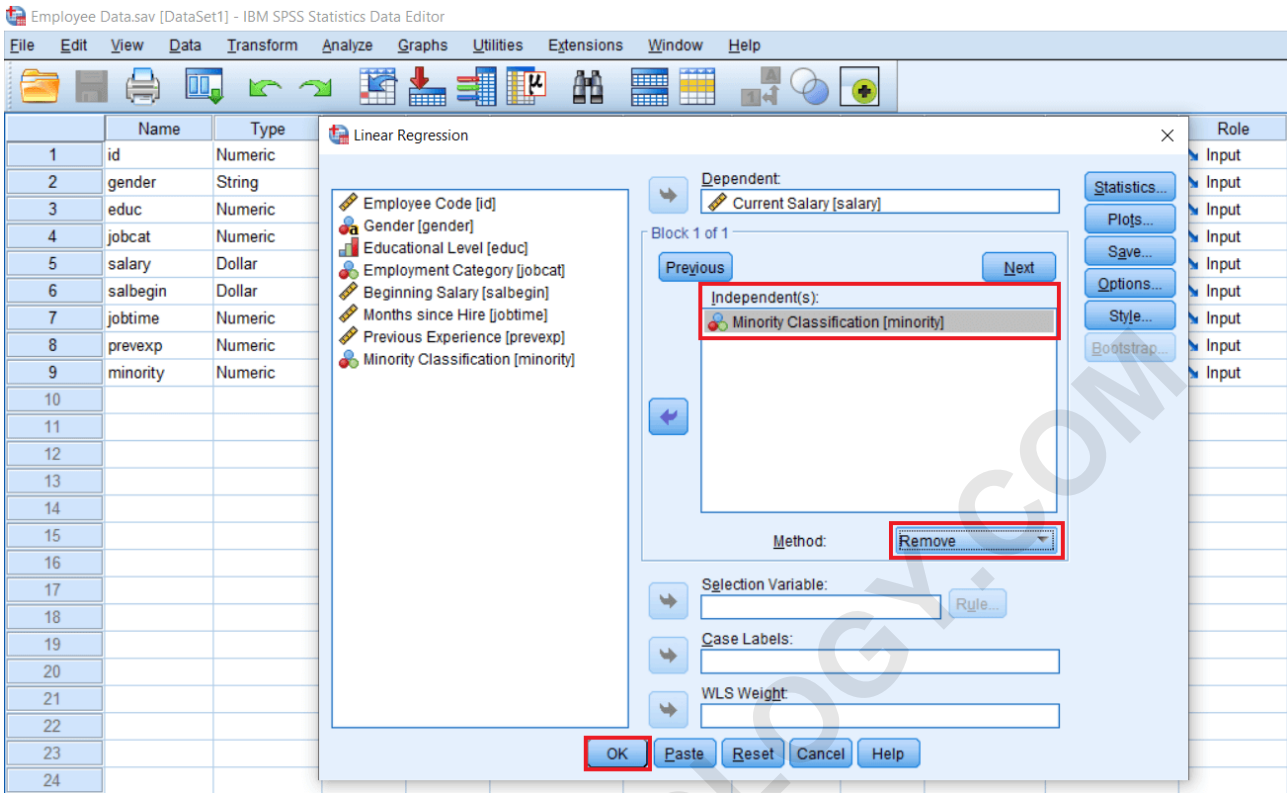
For this, we can carry out an Enter method followed by a Remove method. For that, we will go again on Linear regression and define our variables. So in the first case, we will select Minority Classification under Independent and click on the following arrow.



Now we have five variables. Now we will click on Next to create a second block.



In the second block, we are going to use the Remove method. We will also use Minority classification as a remove variable. After that, click on Ok.



After the Coefficients table, we have Excluded variables from the analysis, which is the minority classification.

The screenshot displays the IBM SPSS Statistics Viewer interface with the following components:

- Model Summary:**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.159E+11	5	2.318E+10	493.084	.000 ^b
	Residual	2.200E+10	468	47015635.57		
	Total	1.379E+11	473			

a. Dependent Variable: Current Salary
 b. Predictors: (Constant), Months since Hire, Previous Experience (months), Beginning Salary, Educational Level (years), Employment Category
- Coefficients:**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-15038.574	2,992.525		-5.025	.000
	Educational Level (years)	539.642	152.735	.091	3.533	.000
	Employment Category	5859.585	624.945	.265	9.376	.000
	Beginning Salary	1.365	.069	.629	19.796	.000
	Previous Experience (months)	-19.553	3.250	-.120	-6.017	.000
	Months since Hire	154.698	31.464	.091	4.917	.000

a. Dependent Variable: Current Salary
- Excluded Variables:**

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Minority Classification	-.016 ^b	-.838	.403	-.039	.948

a. Dependent Variable: Current Salary
 b. Predictors in the Model: (Constant), Months since Hire, Previous Experience (months), Beginning Salary, Educational Level (years), Employment Category
- Coefficient Correlations:**

Double-click to activate

Model		Months since Hire	Previous Experience (months)	Beginning Salary	Educational Level (years)	Employment Category	
1	Correlations	Months since Hire	1.000	-.033	.070	-.082	
		Previous Experience		1.000	-.023	-.023	
		Beginning Salary			1.000	-.023	
		Educational Level				1.000	
		Employment Category					1.000

In the Second model, we can see that we did not enter our previous 5 variables, but they have been kept in this model because we followed the remove method.

The screenshot displays the IBM SPSS Statistics Viewer interface. The left sidebar shows a tree view of the analysis steps: Regression > Coefficients. The main window shows the following tables:

Model Summary

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.159E+11	5	2.318E+10	493.084	.000 ^b
	Residual	2.200E+10	468	47015635.57		
	Total	1.379E+11	473			

a. Dependent Variable: Current Salary
 b. Predictors: (Constant), Months since Hire, Previous Experience (months), Beginning Salary, Educational Level (years), Employment Category

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-15038.574	2,992.525		-5.025	.000
	Educational Level (years)	539.642	152.735	.091	3.533	.000
	Employment Category	5859.585	624.945	.265	9.376	.000
	Beginning Salary	1.365	.069	.629	19.796	.000
	Previous Experience (months)	-19.553	3.250	-.120	-6.017	.000
	Months since Hire	154.698	31.464	.091	4.917	.000

a. Dependent Variable: Current Salary

Excluded Variables^a

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	Minority Classification	-.016 ^b	-.838	.403	-.039	.948

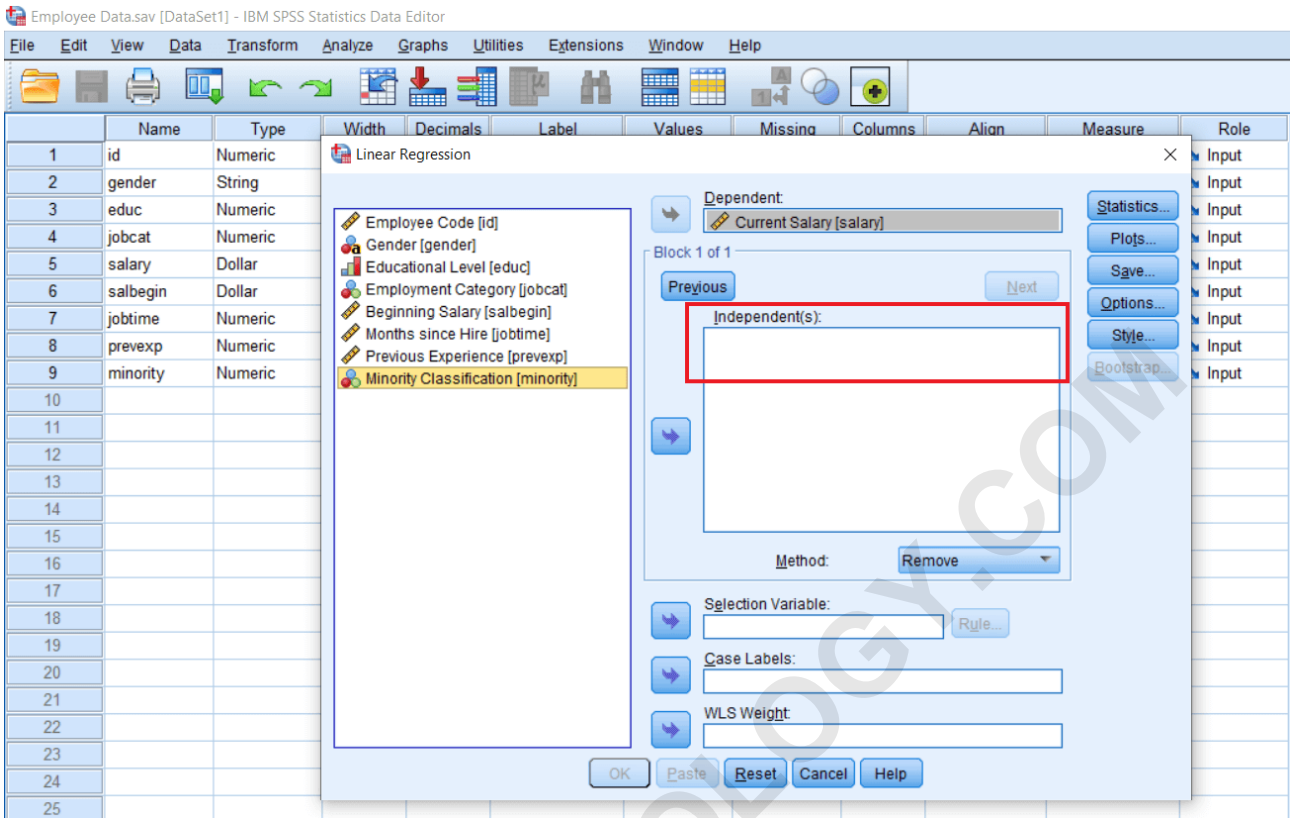
a. Dependent Variable: Current Salary
 b. Predictors in the Model: (Constant), Months since Hire, Previous Experience (months), Beginning Salary, Educational Level (years), Employment Category

Coefficient Correlations^a

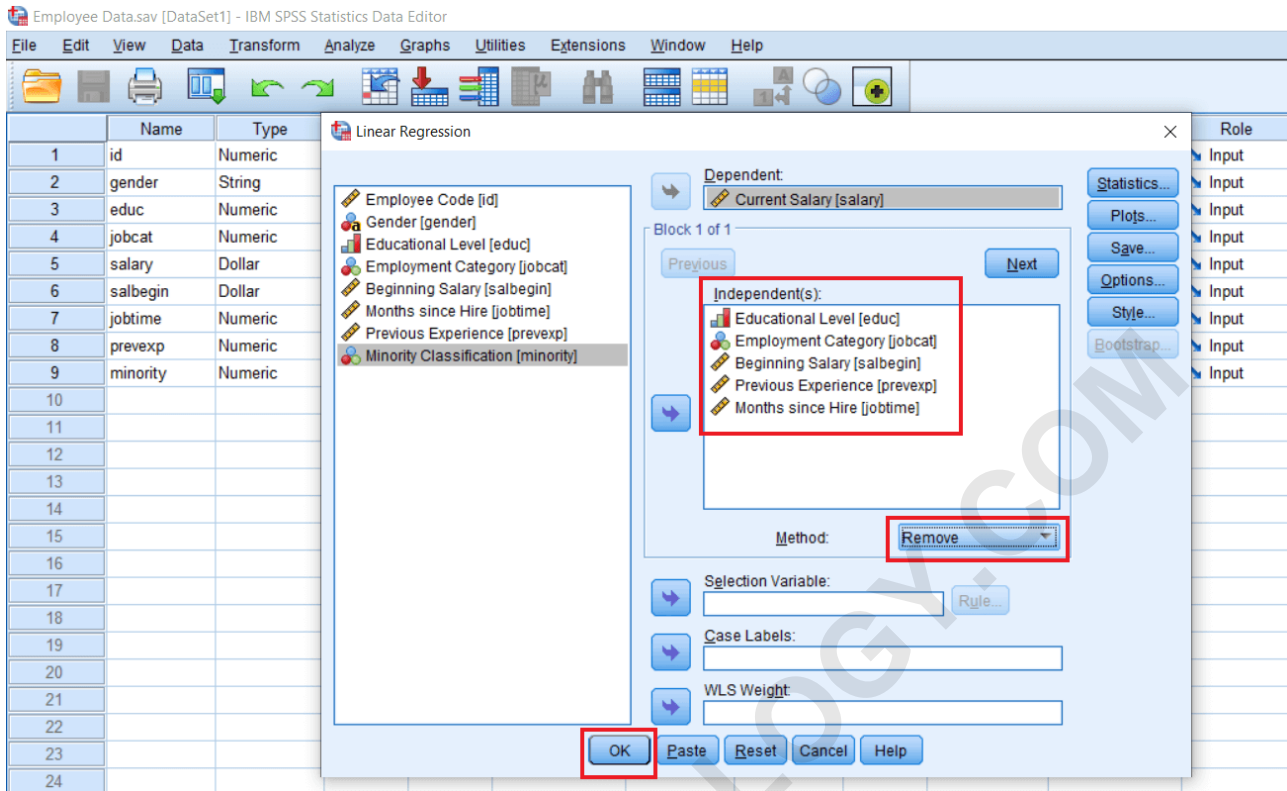
Model		Months since Hire	Previous Experience (months)	Beginning Salary	Educational Level (years)	Employment Category	
1	Correlations	Months since Hire	1.000	-.033	.070	-.082	
		Previous Experience		1.000	-.001	-.001	
		Beginning Salary			1.000	-.001	
		Educational Level				1.000	
		Employment Category					1.000

Double-click to activate

So if we specify a Remove method in step 2, it means it already contains all the variables that have been defined in step 1. So that is the beauty of the remove method. Keep this in mind that we cannot use the remove method in isolation. For example, suppose in step 2, we remove the minority classification variable like this:



Suppose we start with the remove method without carrying out the enter method. So select the Remove method as our first method and click Ok.



Now the software will give us an error warning, which says **Invalid REGRESSION METHOD subcommand specification--REMOVE cannot be used as the first method when building an equation. REGRESSION has inserted ENTER as the first method; REMOVE is now the second method.**

The screenshot shows the IBM SPSS Statistics Viewer interface. The main window displays the 'Regression' output for the dataset 'Employee Data.sav'. A red box highlights a warning message: 'Invalid REGRESSION METHOD subcommand specification-- REMOVE cannot be used as the first method when building an equation. REGRESSION has inserted ENTER as the first method; REMOVE is now the second method.' Below the warning is a table titled 'Variables Entered/Removed^a'.

Model	Variables Entered	Variables Removed	Method
1	Months since Hire, Employment Category, Previous Experience, Educational Level, Beginning Salary ^b		Enter
2	^b	Beginning Salary, Educational Level, Months since Hire, Employment Category, Previous Experience ^c	Remove

a. Dependent Variable: Current Salary
 b. All requested variables entered.
 c. All requested variables removed.

So it has carried out an enter method and followed by a remove method. But in the remove method, it has removed all the variables that we entered in the first case. So that's not going to be any logical value for

ourselves, but that instructs us about the procedure that we should follow while carrying out the remove method.

*Output2 [Document2] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window

Output
 Log
 Regression
 Title
 Notes
 Active Dataset
 Warnings
 Variables Entered
 Model Summary
 ANOVA
 Coefficients
 Excluded Variable

Regression

[DataSet1] C:\Users\ASUS\Desktop\Employee Data.sav

Warnings

Invalid REGRESSION METHOD subcommand specification-- REMOVE cannot be used as the first method when building an equation. REGRESSION has inserted ENTER as the first method; REMOVE is now the second method.

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Months since Hire, Employment Category, Previous Experience, Educational Level, Beginning Salary ^b	.	Enter
2	^b	Beginning Salary, Educational Level, Months since Hire, Employment Category, Previous Experience ^c	Remove

a. Dependent Variable: Current Salary
 b. All requested variables entered.
 c. All requested variables removed.