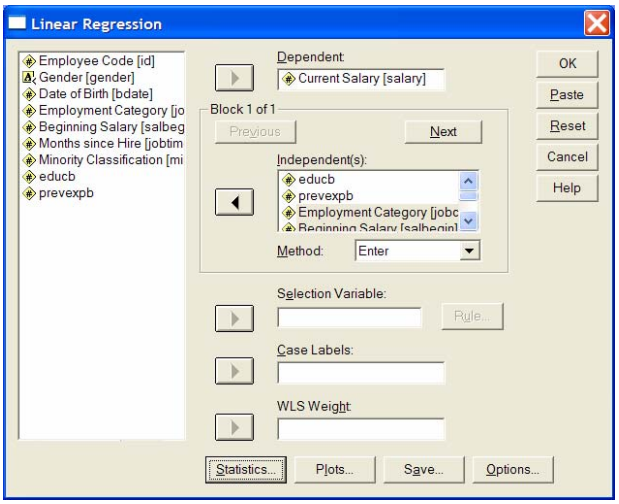
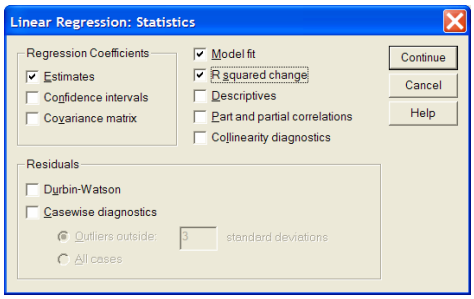
Something to Remember about Model Comparison and Sample Sizes…

The criterion variable is salary. The question is whether education level (years of education) and previous experience with the job are important contributors to the model. So we'll compare the full model with a reduced model dropping these variables (formed by removing these two variables on the second step of the regression analysis).

Analysis → Regression → Linear



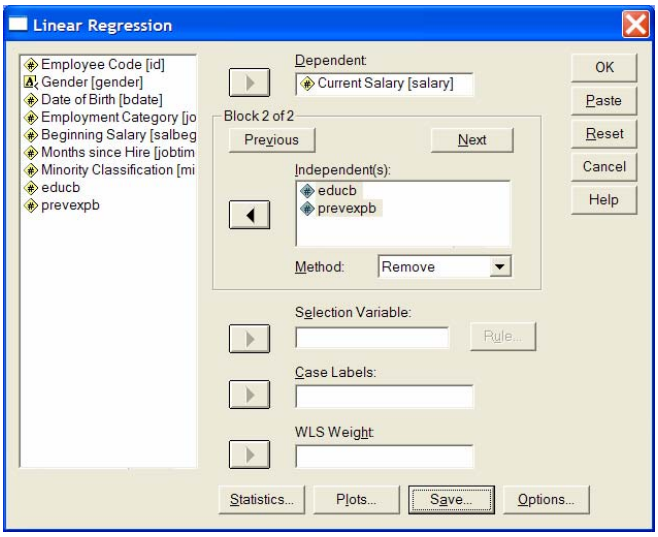


Select the criterion variable

Select the predictors for the full model

Be sure "Method" shows "Enter"

Click Statistics and check "R-squared change"



Click "Next"

Notice it now says "Block 2 of 2"

Select the predictors to remove to create the reduced model from the full model.

Be sure "Method" shows "Remove"

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.913[a]	.833	.831	$7,113.898	.833	361.613	6	434	.000
2	.900[b]	.810	.809	$7,570.344	-.023	29.872	2	434	.000

a. Predictors: (Constant), prevexpb, Months since Hire, Beginning Salary, Minority Classification, educb, Employment Category

b. Predictors: (Constant), Months since Hire, Beginning Salary, Minority Classification, Employment Category

As expected, the R^2 for the reduced model is smaller than for the full model – a mathematical necessity!

With this large sample, even a small $R^2\Delta$ based on 2 predictors is significant.

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	109801984020.633	6	18300330670.106	361.613	.000[a]
	Residual	21963673523.698	434	50607542.681		
	Total	131765657544.331	440			
2	Regression	106778449392.845	4	26694612348.211	465.792	.000[b]
	Residual	24987208151.487	436	57310110.439		
	Total	131765657544.331	440			

a. Predictors: (Constant), prevexpb, Months since Hire, Beginning Salary, Minority Classification, educb, Employment Category

b. Predictors: (Constant), Months since Hire, Beginning Salary, Minority Classification, Employment Category

c. Dependent Variable: Current Salary

ANOVA tests of each model's R^2

This table is the only place that SPSS shows you the sample size

$N = df_{residual} + df_{regression} + 1$

$N = df_{total} + 1$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-15068.1	3239.307		-4.652	.000
	Employment Category	5829.823	661.484	.264	8.813	.000
	Beginning Salary	1.361	.073	.628	18.637	.000
	Months since Hire	156.844	34.140	.091	4.594	.000
	Minority Classification	-667.259	842.024	-.016	-.792	.429
	educb	544.773	167.479	.090	3.253	.001
	prevexpb	-19.154	3.697	-.112	-5.181	.000
2	(Constant)	-11366.0	3063.569		-3.710	.000
	Employment Category	5770.390	698.833	.261	8.257	.000
	Beginning Salary	1.464	.069	.676	21.365	.000
	Months since Hire	162.237	36.191	.094	4.483	.000
	Minority Classification	-1600.819	883.609	-.038	-1.812	.071

a. Dependent Variable: Current Salary

Regression weights for each model.

The

Here's the same hypothesis testing, but this time getting each model separately from SPSS

The Full model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.913[a]	.833	.831	$7,113.898

a. Predictors: (Constant), prevexpb, Months since Hire, Beginning Salary, Minority Classification, educb, Employment Category

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	109801984021	6	18300330670.1	361.613	.000[a]
	Residual	21963673523.7	434	50607542.681		
	Total	131765657544	440			

a. Predictors: (Constant), prevexpb, Months since Hire, Beginning Salary, Minority Classification, educb, Employment Category

b. Dependent Variable: Current Salary

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-15068.1	3239.307		-4.652	.000
	Employment Category	5829.823	661.484	.264	8.813	.000
	Beginning Salary	1.361	.073	.628	18.637	.000
	Months since Hire	156.844	34.140	.091	4.594	.000
	Minority Classification	-667.259	842.024	-.016	-.792	.429
	educb	544.773	167.479	.090	3.253	.001
	prevexpb	-19.154	3.697	-.112	-5.181	.000

a. Dependent Variable: Current Salary

The same result as the first step of the previous procedure!

The Reduced model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.918[a]	.842	.815	$7,347.221

a. Predictors: (Constant), Minority Classification, Months since Hire, Employment Category, Beginning Salary

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	112599096777.9	4	28149774194.475	521.469	.000[a]
	Residual	25317398658.44	469	53981660.253		
	Total	137916495436.3	473			

a. Predictors: (Constant), Minority Classification, Months since Hire, Employment Category, Beginning Salary

b. Dependent Variable: Current Salary

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-11937.0	2859.302		-4.175	.000
	Employment Category	5797.404	666.766	.263	8.695	.000
	Beginning Salary	1.471	.066	.678	22.409	.000
	Months since Hire	166.238	33.640	.098	4.942	.000
	Minority Classification	-1549.233	827.335	-.038	-1.873	.062

a. Dependent Variable: Current Salary

But look what happens when we get the reduced model as a separate model (rather than getting it from the same run as the full model, as we did above)

The R^2 for the reduced model is larger than for the full model?????

Check out the sample size for this model – N = 472 + 1 = 473 → different from the full model above, or from the reduced mode obtained previously.

SPSS uses "**listwise deletion**" which means that for any analysis, it uses only the cases that have data for all the variables involved in the analysis.

It appears that there were several cases that didn't have data for education level and/or previous experience.

"Odd" things can happen when using different samples for different models – be careful!

Always check your sample sizes?