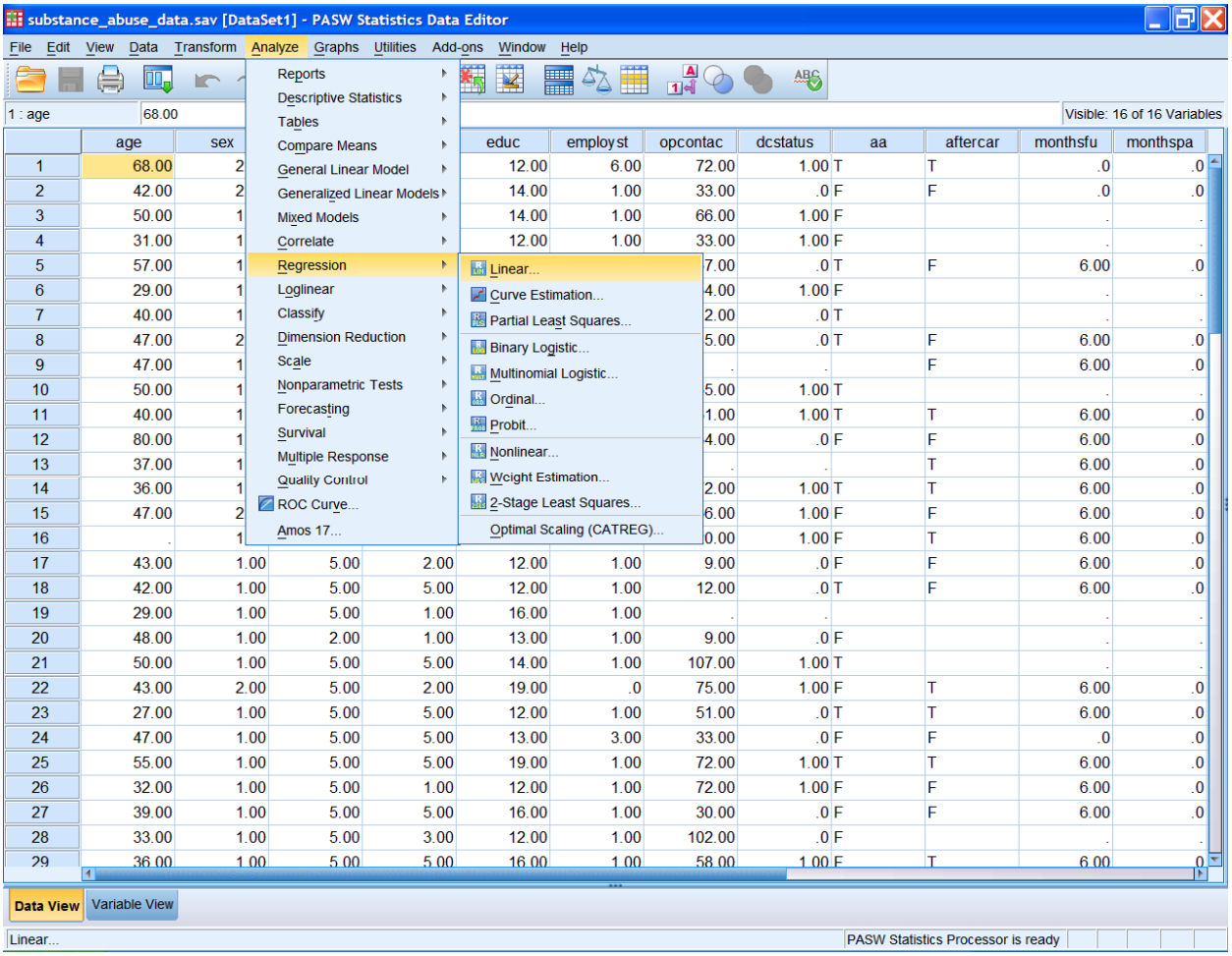
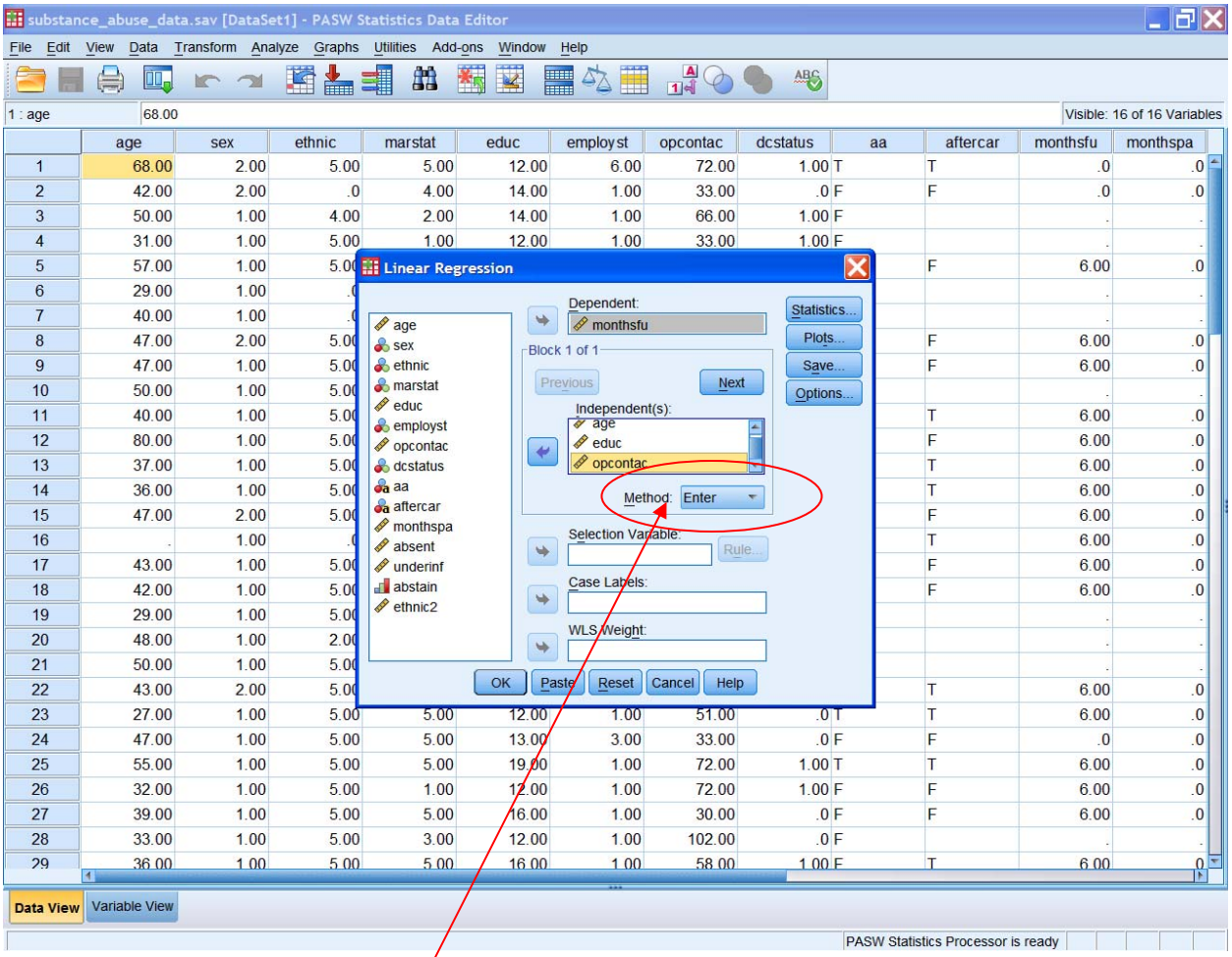
Multiple Regression in SPSS

This example shows you how to perform **multiple regression**. The basic command is "regression": "linear."

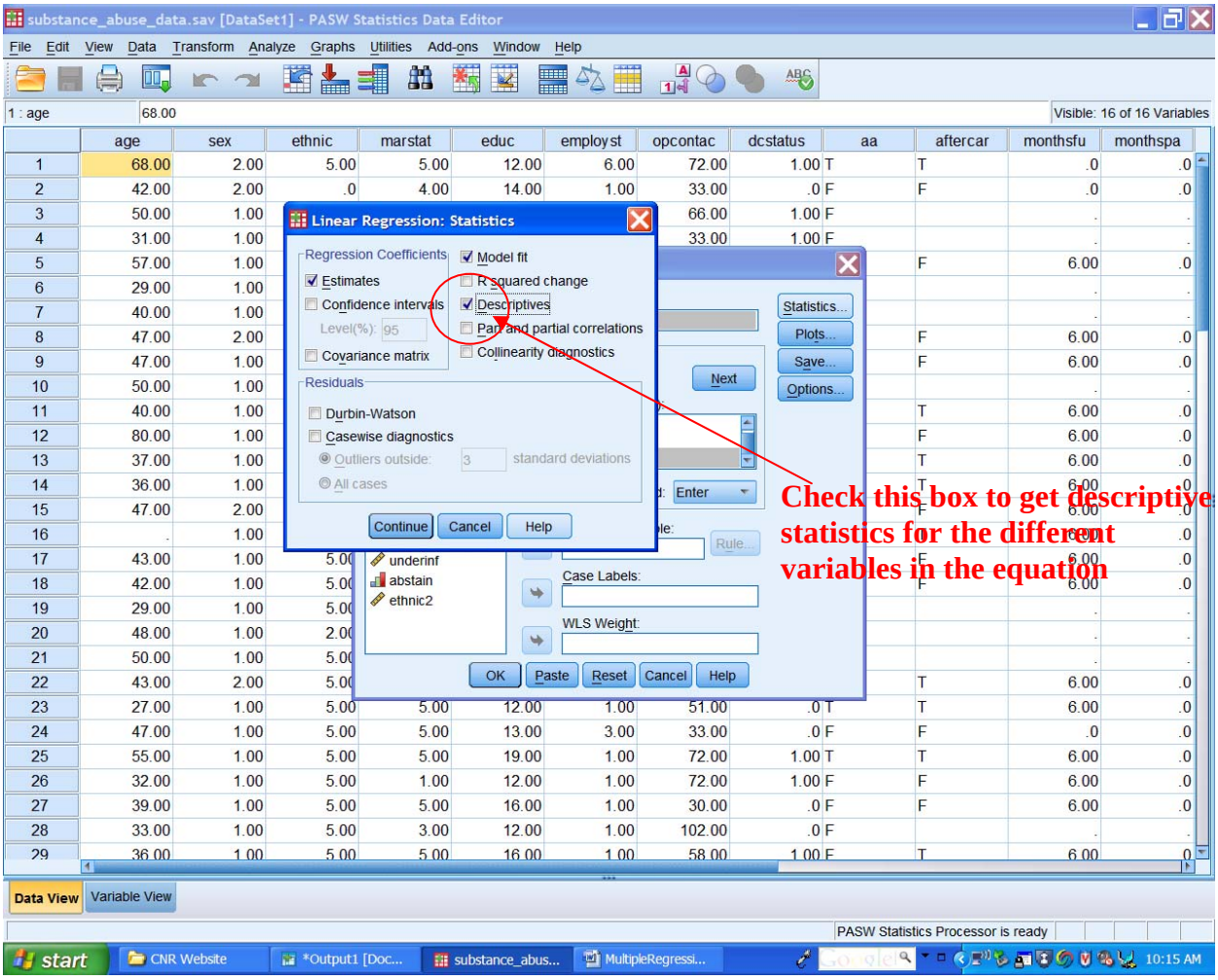


In the main dialog box, input the dependent variable and several predictors. In this case, we want to predict "months of full-time employment" ("monthsfu") among participants in a substance abuse treatment program. We're going to use three predictors: participants' age, number of years of education, and number of outpatient sessions completed ("opcontact").



Leave this drop-down menu set to the default value ("Enter"), for now.

Click on the “Statistics” button to view this dialog box.



Here’s the output:

Descriptive Statistics

	Mean	Std. Deviation	N
monthsfu	4.9048	2.30386	42
age	42.8810	13.22821	42
educ	14.4048	2.83755	42
opcontac	46.6905	26.96474	42

These are the descriptive statistics, based on the option that we selected.

Correlations

		monthsfu	age	educ	opcontac
Pearson Correlation	monthsfu	1.000	-.321	.278	-.028
	age	-.321	1.000	.342	.151
	educ	.278	.342	1.000	.195
	opcontac	-.028	.151	.195	1.000
Sig. (1-tailed)	monthsfu	.	.019	.037	.429
	age	.019	.	.013	.170
	educ	.037	.013	.	.108
	opcontac	.429	.170	.108	.
N	monthsfu	42	42	42	42
	age	42	42	42	42
	educ	42	42	42	42
	opcontac	42	42	42	42

The "descriptives" command also gives you a correlation matrix, showing you the Pearson *r*s between the variables (in the top part of this table).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.525[a]	.276	.219	2.03647

a. Predictors: (Constant), opcontac, age, educ

This table tells you what % of variability in the DV is accounted for by all of the IVs together (it's a *multiple* R-square). The footnote on this table tells you which variables were included in this equation (in this case, all three of the ones that we put in).

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	60.025	3	20.008	4.825	.006[a]
	Residual	157.594	38	4.147		
	Total	217.619	41			

a. Predictors: (Constant), opcontac, age, educ

b. Dependent Variable: monthsfu

This table gives you an *F*-test to determine whether the model is a good fit for the data. According to this *p*-value, it is.

Coefficients[a]

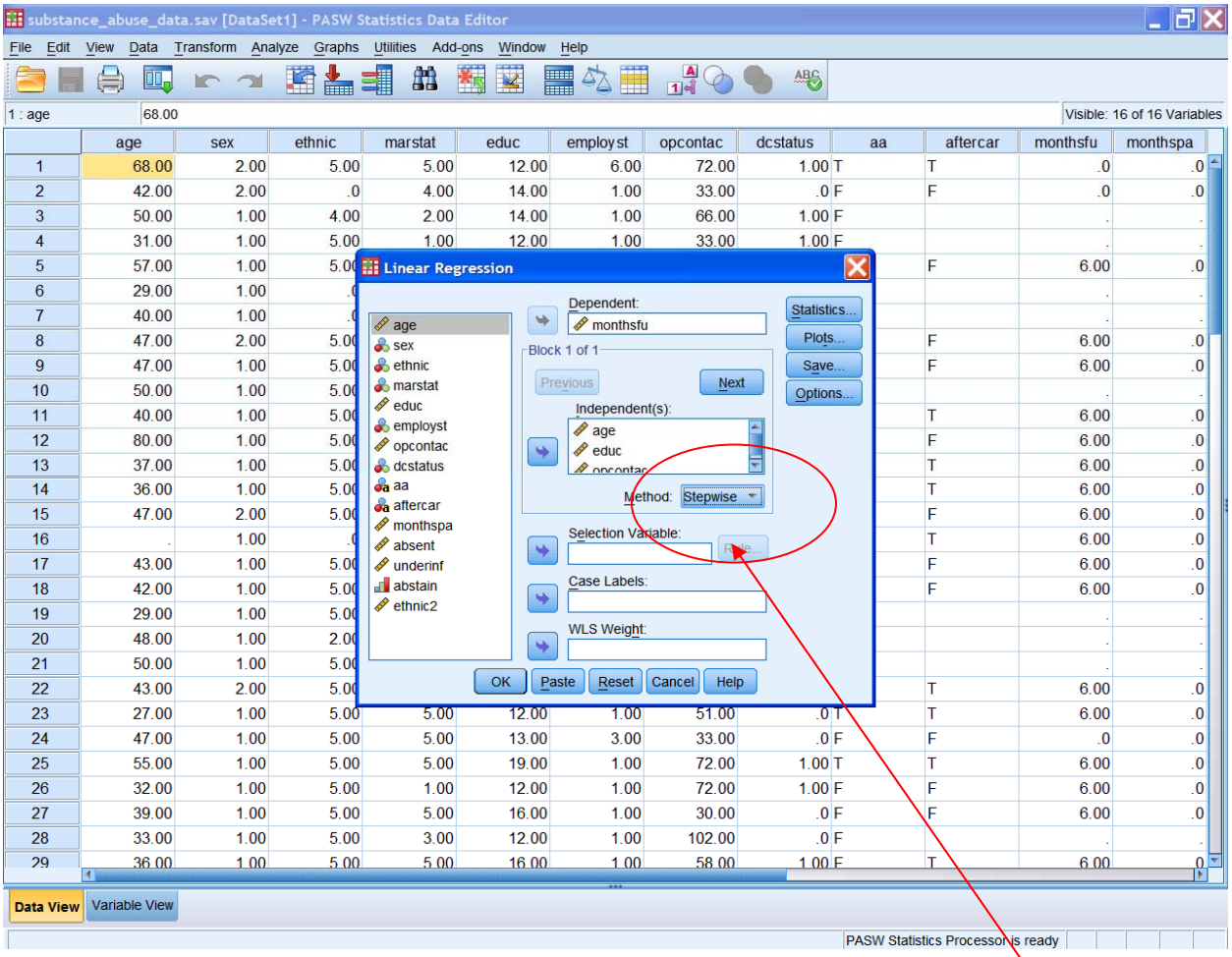
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.347	1.733		1.932	.061
	age	-.081	.026	-.467	-3.168	.003
	educ	.363	.121	.447	3.006	.005
	opcontac	-.004	.012	-.045	-.319	.752

a. Dependent Variable: monthsfu

Finally, here are the beta coefficients—one to go with each predictor. (Use the "unstandardized coefficients," because the constant [beta zero] is included). Based on this table, the equation for the regression line is:

$$y = 3.347 - .081(age) + .363(educ) - .004(opcontact)$$

Using this equation, given values for "age," "educ," and "opcontact," you can come up with a prediction for the "months of full-time work" variable.



Now go back to the original dialog box, and change this drop-down menu to use the “stepwise” method instead.

[For the sake of simplicity, I also went under “statistics” and turned *off* the “descriptives” option for the following tests]

Here's the revised output:

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	age	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	educ	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: monthsfu

This first table tells you which variables were included in the model at each step: "Age" was the single best predictor (step 1), and "Educ" was the next best predictor (added the most), after "Age" was included in the model (step 2).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.321[a]	.103	.081	2.20880
2	.523[b]	.274	.237	2.01288

a. Predictors: (Constant), age

b. Predictors: (Constant), age, educ

Again, here are the R-squares. With "Age" alone (step 1), 10.3% of the variance was accounted for. With both "Age" and "Educ" (step 2), 27.4% of the variance was accounted for.

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.466	1	22.466	4.605	.038[a]
	Residual	195.153	40	4.879		
	Total	217.619	41			
2	Regression	59.604	2	29.802	7.355	.002[b]
	Residual	158.015	39	4.052		
	Total	217.619	41			

a. Predictors: (Constant), age

b. Predictors: (Constant), age, educ

c. Dependent Variable: monthsfu

This table now gives two *F*-tests, one for each step of the procedure. Both steps had overall significant results (p = .038 for Age alone, and p = .002 for Age and Educ).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.304	1.169		6.248	.000
	age	-.056	.026	-.321	-2.146	.038
2	(Constant)	3.285	1.702		1.930	.061
	age	-.082	.025	-.472	-3.248	.002
	educ	.357	.118	.440	3.028	.004

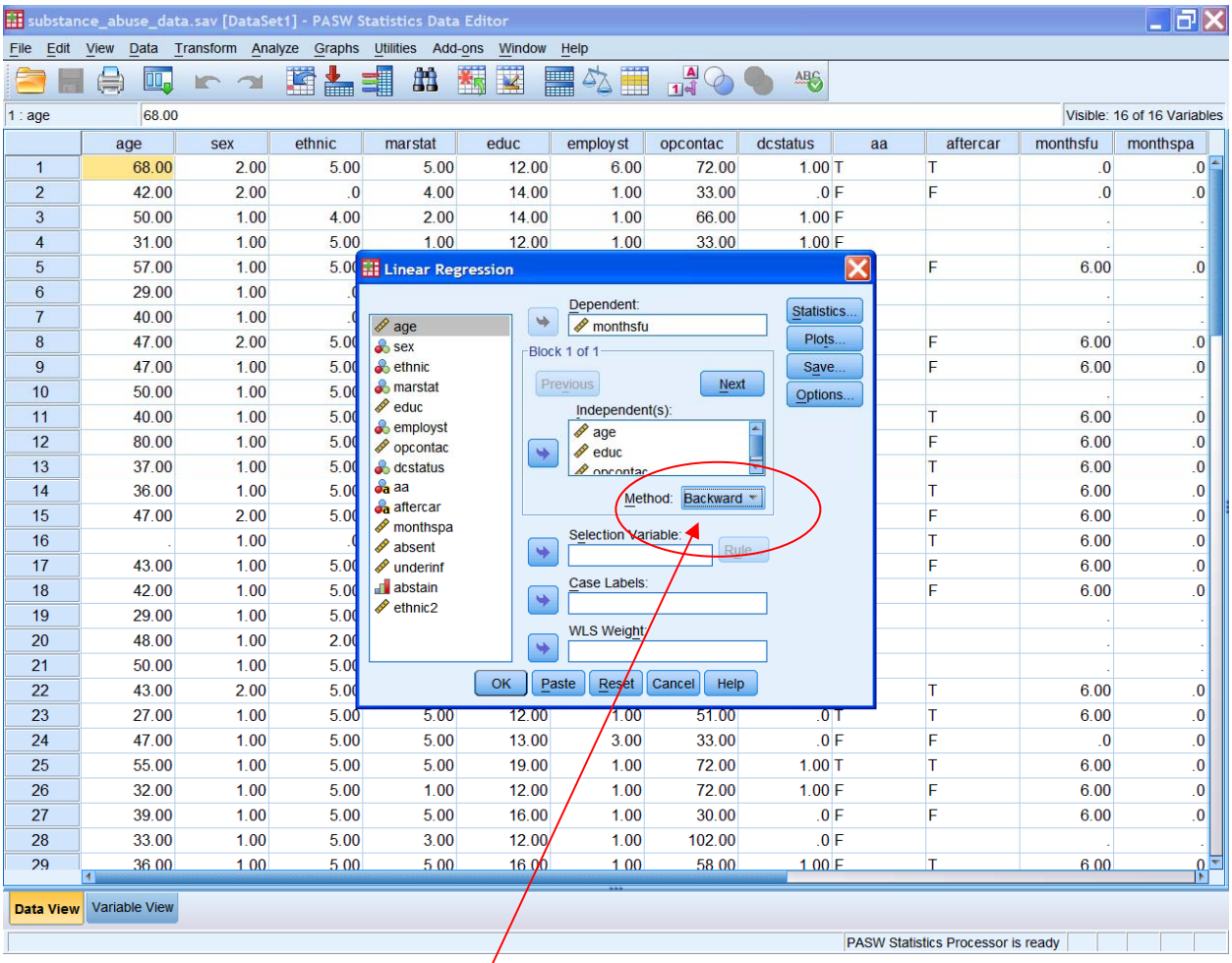
a. Dependent Variable: monthsfu

Again, this table gives beta coefficients so that you can construct the regression equation. Notice that the betas *change*, depending on which predictors are included in the model.

These are the weights that you want, for an equation that includes just Age and Education (the two best predictors). The equation would be:

Predicted Months Full-Time Work = 3.285 - .082(Age) + .357(Educ)

The last table ("Variables Excluded from the Equation") just lists the variables that *weren't* included in the model at each step.



Finally, try the “backward” stepwise regression procedure.

Here's the output:

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	opcontac, age, educ[a]	.	Enter
2	age, educ	opcontac	Backward (criterion: Probability of F-to-remove >= .100).

a. All requested variables entered.

b. Dependent Variable: monthsfu

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.525[a]	.276	.219	2.03647
2	.523[b]	.274	.237	2.01288

a. Predictors: (Constant), opcontac, age, educ

b. Predictors: (Constant), age, educ

From this model summary, you can see that step 2 gets down to the same two predictors that we wound up with in the "forward stepwise" procedure (Age and Educ). You can see the small difference in the R-square between step 1 and step 2—that's why the model discarded the third predictor as not being particularly useful.

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	60.025	3	20.008	4.825	.006[a]
	Residual	157.594	38	4.147		
	Total	217.619	41			
2	Regression	59.604	2	29.802	7.355	.002[b]
	Residual	158.015	39	4.052		
	Total	217.619	41			

a. Predictors: (Constant), opcontac, age, educ

b. Predictors: (Constant), age, educ

c. Dependent Variable: monthsfu

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.347	1.733		1.932	.061
	age	-.081	.026	-.467	-3.168	.003
	educ	.363	.121	.447	3.006	.005
	opcontac	-.004	.012	-.045	-.319	.752
2	(Constant)	3.285	1.702		1.930	.061
	age	-.082	.025	-.472	-3.248	.002
	educ	.357	.118	.440	3.028	.004

a. Dependent Variable: monthsfu

Excluded Variables[b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
2	opcontac	-.045[a]	-.319	.752	-.052	.954

a. Predictors in the Model: (Constant), age, educ

b. Dependent Variable: monthsfu