Multiple Regression in SPSS

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

🔢 substan	substance_abuse_data.sav [DataSet1] - PASW Statistics Data Editor											
<u>File</u> Edit	<u>View</u> <u>D</u> ata <u>T</u>	ransform	Analyze Graphs	Utilities Add	-ons <u>W</u> indow	Help						
🔁 🖥		F 1	Reports	b b	🍇 🖬 🚦	- 4		ABG				
1 : age	68.00		Tables							Visible:	16 of 16 Variables	
	age	sex	Compare Mea	ns 🕨	educ	employ st	opcontac	dcstatus	aa	aftercar monthsfu	monthspa	
1	68.00	2	General Linea	r Model 🔹 🕨	12.00	6.00	72.00	1.00 T	T T	.0	.0 🚔	
2	42.00	2	Generalized Li	near Models 🕨	14.00	1.00	33.00	.0 F	F	.0	.0	
3	50.00	1	Mixed Models	•	14.00	1.00	66.00	1.00 F				
4	31.00	1	Correlate	•	12.00	1.00	33.00	1.00 F				
5	57.00	1	Regression	•	Linear		7.00	.0 T	F	6.00	.0	
6	29.00	1	Loglinear	۰.	Z Curve Est	imation	4.00	1.00 F				
7	40.00	1	Classify	•	- Rartial Lea	ast Squares	2.00	.0 T				
8	47.00	2	Dimension Rec	duction 🕨 🕨	Binary Lo	aistic	5.00	.0 T	F	6.00	.0	
9	47.00	1	Sc <u>a</u> le	*	Multinomia	al Logistic		-	F	6.00	.0	
10	50.00	1	Nonparametric	Tests 🕨	Crdinal		5.00	1.00 T				
11	40.00	1	Forecasting	*	B Probit		1.00	1.00 T	T	6.00	.0	
12	80.00	1	Survival	•	Nonlinear		4.00	.0 F	F	6.00	.0	
13	37.00	1	Multiple Respo	nse 🕨		timation			Т	6.00	.0	
14	36.00	1	Quality Contro	1 1		aget Caugree	2.00	1.00 T	T	6.00	.0	
15	47.00	2	ROC Curve		Z-Stage L	east Squares	6.00	1.00 F	F	6.00	.0	
16		1	<u>A</u> mos 17		Optimal S	caling (CATREG) 0.00	1.00 F	Т	6.00	.0	
17	43.00	1.0	0 5.00	2.00	12.00	1.00	9.00	.0 F	F	6.00	.0	
18	42.00	1.0	0 5.00	5.00	12.00	1.00	12.00	.0 T	F	6.00	.0	
19	29.00	1.0	0 5.00	1.00	16.00	1.00						
20	48.00	1.0	0 2.00	1.00	13.00	1.00	9.00	.0 F				
21	50.00	1.0	0 5.00	5.00	14.00	1.00	107.00	1.00 T				
22	43.00	2.0	0 5.00	2.00	19.00	.0	75.00	1.00 F	Т	6.00	.0	
23	27.00	1.0	0 5.00	5.00	12.00	1.00	51.00	.0 T	Т	6.00	.0	
24	47.00	1.0	0 5.00	5.00	13.00	3.00	33.00	.0 F	F	.0	.0	
25	55.00	1.0	0 5.00	5.00	19.00	1.00	72.00	1.00 T	Т	6.00	.0	
26	32.00	1.0	0 5.00	1.00	12.00	1.00	72.00	1.00 F	F	6.00	.0	
27	39.00	1.0	0 5.00	5.00	16.00	1.00	30.00	.0 F	F	6.00	.0	
28	33.00	1.0	0 5.00	3.00	12.00	1.00	102.00	.0 F				
29	36 00	1 0	0 5 00	5 00	16 00	1 00	58 00	1 00 F	Т	6 00	0	
	Mariable Misur					***						
Data View	variable view											
Linear									PASW Statistics	Processor is ready		

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").

🔢 substan	substance_abuse_data.sav [DataSet1] - PASW Statistics Data Editor											
File Edit	View Data T	ransform Ana	lyze Graphs	Utilities Add	ons Window	Help						
2		5 3			🍇 🔟	- 43		ABG				
1 : age	68.00										Visible:	16 of 16 Variables
<u> </u>	age	sex	ethnic	marstat	educ	employ st	opcontac	dcstatus	aa	aftercar	monthsfu	monthspa
1	68.00	2.00	5.00	5.00	12.00	6.00	72.00	1.00	Г	Т	.0	.0 🚔
2	42.00	2.00	.0	4.00	14.00	1.00	33.00	.0 I	-	F	.0	.0
3	50.00	1.00	4.00	2.00	14.00	1.00	66.00	1.00	-			
4	31.00	1.00	5.00	1.00	12.00	1.00	33.00	1.00				
5	57.00	1.00	5.00	Linear Re	ression				X	F	6.00	.0
6	29.00	1.00				Dependent				-		
7	40.00	1.00		🔗 age		monthsfu		Statistic	S		-	-
8	47.00	2.00	5.00	💑 sex	Blo	ck 1 of 1		Plots.	-	F	6.00	.0
9	47.00	1.00	5.00	🗞 ethnic	6	-	Ne	Save.		F	6.00	.0
10	50.00	1.00	5.00	adur		Tevious	<u>Ne</u>	<u>Option</u>	s			
11	40.00	1.00	5.00	equic equic		age	t(S):			Т	6.00	.0
12	80.00	1.00	5.00	I opcontac		educ				F	6.00	.0
13	37.00	1.00	5.00	💑 dcstatus		opcontac	h			Т	6.00	.0
14	36.00	1.00	5.00	💑 aa		Mett	hod: Enter	-)		Т	6.00	.0
15	47.00	2.00	5.00	monthspa	-					F	6.00	.0
16		1.00		sent absent		Selection Var	able:	ile		т	6.00	.0
17	43.00	1.00	5.00	lunderinf						F	6.00	.0
18	42.00	1.00	5.00	abstain				_		F	6.00	.0
19	29.00	1.00	5.00			WI S Moight						
20	48.00	1.00	2.00			WLS Weight.						
21	50.00	1.00	5.00									
22	43.00	2.00	5.00		OK	aste Reset	Cancel Hel	p		Т	6.00	.0
23	27.00	1.00	5.00	5.00	12.00	1.00	51.00	.0		Ţ	6.00	.0
24	47.00	1.00	5.00	5.00	13.00	3.00	33.00	.0		F	.0	.0
25	55.00	1.00	5.00	5.00	19.00	1.00	72.00	1.00	Ē.	Т	6.00	.0
26	32.00	1.00	5.00	1.00	12.00	1.00	72.00	1.00	-	F	6.00	.0
27	39.00	1.00	5.00	5.00	16.00	1.00	30.00	.0		F	6.00	.0
28	33.00	1.00	5.00	3.00	12.00	1.00	102.00	.0	-	_		
29	36 00	1 00	5 00	5 00	16.00	1 00	58.00	1 00 1	-	T	6 00	0
Data View	Variable View											
				/					PASW Sta	atistics Processor i	s ready	

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

📰 substan	ce_abuse_dat	a.sav [DataSe	et1] - PASW S	tatistics Data	a Editor							_ ð X
<u>File</u> Edit	View Data T	ransform Ana	alyze Graphs	Utilities Add	l-ons Window	Help						
		5 7		A1	🍇 🔟 🚦	- 43		ABS				
1 : age	68.00										Visible:	16 of 16 Variable
	age	sex	ethnic	marstat	educ	employst	opcontac	dcstatus	aa	aftercar	monthsfu	monthspa
1	68.00	2.00	5.00	5.00	12.00	6.00	72.00	1.00	Г	Т	.0	.0
2	42.00	2.00	.0	4.00	14.00	1.00	33.00	.0	F	F	.0	.0
3	50.00	1.00	III Linear	Regression:	Statistics	5	66.00	1.00	F			
4	31.00	1.00		0			33.00	1.00	F		-	
5	57.00	1.00	Regressi	on Coefficients	Model fit				×	F	6.00	.0
6	29.00	1.00	Estima	tes	R squared o	hange						
7	40.00	1.00		ence intervals	Descriptives	dial completions		Statistic	:S			
8	47.00	2.00	Levei(%). 95	Par and pa	rual correlations	,	Plots	-	F	6.00	.0
9	47.00	1.00	Covari	ance matrix	Commeanly	ulagilostics		Save		F	6.00	.0
10	50.00	1.00	Residuals	5			Nex	d Option	s		-	
11	40.00	1.00	🔲 D <u>u</u> rbi	n-Watson						т	6.00	.0
12	80.00	1.00		vise diagnostics	3					F	6.00	.0
13	37.00	1.00	© <u>O</u> uti	ers outside:	3 standa	ard deviations				т	6.00	.0
14	36.00	1.00	© <u>A</u> ll c	ases			: Enter	Ch	ock th	nis ^T hov t		crintive
15	47.00	2.00		Continue							gc 6.00	,ci ipugo
16	3	1.00	_	Continue	Cancel)	IIE:	stat	istics	for the	differen	<u>t</u> .0
17	43.00	1.00	5.00	winderinf		Case Labels		var	iable	s in the c	equation	.0
18	42.00	1.00	5.00	abstain		Case Labels.				F	6.00	.0
19	29.00	1.00	5.00	etinic₂		WI & Woight						
20	48.00	1.00	2.00		•	WES WEIGHT.						
21	50.00	1.00	5.00									-
22	43.00	2.00	5.00		OK Pa	iste <u>R</u> eset	Cancel Hel			Т	6.00	.0
23	27.00	1.00	5.00	5.00	12.00	<mark>1.00</mark>	51.00	.0	T,	Т	6.00	.0
24	47.00	1.00	5.00	5.00	13.00	3.00	33.00	.0	F	F	.0	.0
25	55.00	1.00	5.00	5.00	19.00	1.00	72.00	1.00	Г	Т	6.00	.0
26	32.00	1.00	5.00	1.00	12.00	1.00	72.00	1.00	F	F	6.00	.0
27	39.00	1.00	5.00	5.00	16.00	1.00	30.00	.0	F	F	6.00	.0
28	33.00	1.00	5.00	3.00	12.00	1.00	102.00	.0	F			
29	36.00	1 00	5 00	5 00	16 00	1 00	58.00	1 00	F	T	6.00	0
Data View	Variable View											
	,								PASW Sta	atistics Processor	is ready	
🏄 star	t 📄 CNR	Website	💼 *Output1	[Doc	substance_abus	🖻 Multipl	eRegressi	e .	Google	Q ▼ □ (C EN)		💧 🐙 10:15 AM

Click on the "Statistics" button to view this dialog box.

Here's the output:

	Descript	tive Statistics	
	Mean	Std. Deviation	Ν
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	1
	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		
	Ν	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 *rs* between the variables (in the top part of this table).



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			
	<u> </u>					

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			$\langle \rangle$			Standardized		
			Unstandardize	ed	Coefficients	Coefficients		
			В		Std. Error	Beta	t	Sig.
1	(Constant)		3.347		1.733		1.932	.061
	age		081		.026	467	-3.168	.003
	educ		.363		.121	.447	3.006	.005
	opcontac	1	004		.012	045	319	.752
a Dan	andant Variable: m	antk						

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.

🔢 substan	substance_abuse_data.sav [DataSet1] - PASW Statistics Data Editor 📃 🗗 🔀											
File Edit	View Data T	ransform Ana	lyze Graphs	Utilities Add	ons Window	Help						
		5 3			*	- 43		ABG				
1 : age	68.00										Visible: 1	6 of 16 Variables
	age	sex	ethnic	marstat	educ	employ st	opcontac	dcstatus	aa	aftercar	monthsfu	monthspa
1	68.00	2.00	5.00	5.00	12.00	6.00	72.00	1.00	Г	Т	.0	.0 🚔
2	42.00	2.00	.0	4.00	<mark>14.00</mark>	1.00	33.00	.0	F	F	.0	.0
3	50.00	1.00	4.00	2.00	14.00	1.00	66.00	1.00	F			
4	31.00	1.00	5.00	1.00	12.00	1.00	33.00	1.00	-			-
5	57.00	1.00	5.00	Linear Reg	ression				×	F	6.00	.0
6	29.00	1.00				Dependent:		(
7	40.00	1.00		age age	•	monthsfu	1	Statistic	:S		1	
8	47.00	2.00	5.00	💰 sex	Blog	:k 1 of 1		Plots.		F	6.00	.0
9	47.00	1.00	5.00	💑 ethnic				Save.		F	6.00	.0
10	50.00	1.00	5.00	adus adus	P	revious	<u>N</u> ex	d Option	s		1	
11	40.00	1.00	5.00	<pre>euuc</pre>		Independen	t(s):			Т	6.00	.0
12	80.00	1.00	5.00	Ø opcontac		educ	_			F	6.00	.0
13	37.00	1.00	5.00	\delta dcstatus		A oncontac				Т	6.00	.0
14	36.00	1.00	5.00	💑 aa		Met	hod: Stepwise	-		Т	6.00	.0
15	47.00	2.00	5.00	Monthspa						F	6.00	.0
16		1.00		absent 🖉		Selection Va	iable:	le		Т	6.00	.0
17	43.00	1.00	5.00	🛷 underinf				-		F	6.00	.0
18	42.00	1.00	5.00	abstain		Case Labels:		\rightarrow		F	6.00	.0
19	29.00	1.00	5.00			MI C Moight:						
20	48.00	1.00	2.00			WLS Weight		\neg				
21	50.00	1.00	5.00									
22	43.00	2.00	5.00			aste <u>R</u> eset	Cancel Hel			Т	6.00	.0
23	27.00	1.00	5.00	5.00	12.00	1.00	51.00	.0		Т	6.00	.0
24	47.00	1.00	5.00	5.00	13.00	3.00	33.00	.0	F	F	.0	.0
25	55.00	1.00	5.00	5.00	19.00	1.00	72.00	1.00	r 🔪	Т	6.00	.0
26	32.00	1.00	5.00	1.00	12.00	1.00	72.00	1.00	F 🔪	F	6.00	.0
27	39.00	1.00	5.00	5.00	16.00	1.00	30.00	.0	F	F	6.00	.0
28	33.00	1.00	5.00	3.00	12.00	1.00	102.00	.0	F	\mathbf{X}		
29	36.00	1 00	5 00	5 00	16 00	1 00	58 00	1.00	F	Т	6 00	0
Data View	Variable View					***						
	PASW Statistics Processon's ready											

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				Std. Error of the							
	R	R Square	Adjusted R Square	Estimate							
1	.321ª	.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. Prec	lictors: (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		Unstandardize	ed Coefficients	Standardized Coefficients		
		В	Std. Error	Beta	t	Sig.
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.

🔢 substan	substance_abuse_data.sav [DataSet1] - PASW Statistics Data Editor											
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help												
		5			* 2	- 43		ABC				
1 : age	age 68.00 Visible: 16 of 16 Variables											
	age	sex	ethnic	marstat	educ	employ st	opcontac	dcstatus	aa	aftercar	monthsfu	monthspa
1	68.00	2.00	5.00	5.00	12.00	6.00	72.00	1.00	Т	Т	.0	.0 🚔
2	42.00	2.00	.0	4.00	14.00	1.00	33.00	.0	F	F	.0	.0
3	50.00	1.00	4.00	2.00	14.00	1.00	66.00	1.00	F			
4	31.00	1.00	5.00	1.00	12.00	1.00	33.00	1.00	F			
5	57.00	1.00	5.00	Linear Rea	ression				×	F	6.00	.0
6	29.00	1.00				Dependent:		(
7	40.00	1.00		🔗 age	•	monthsfu	1	Statisti	cs		-	-
8	47.00	2.00	5.00	💰 sex	Blo	ck 1 of 1		Plots		F	6.00	.0
9	47.00	1.00	5.00	🗞 ethnic	6	reutieure	Ne	Save		F	6.00	.0
10	50.00	1.00	5.00	and uc	P	revious	<u>Ne</u>	<u>Option</u>	IS			
11	40.00	1.00	5.00	employst		age age	l(S).			Т	6.00	.0
12	80.00	1.00	5.00	opcontac		educ				F	6.00	.0
13	37.00	1.00	5.00	🗞 dcstatus		A opcontac				Т	6.00	.0
14	36.00	1.00	5.00	💑 aa		Met	hod: Backward	1-		Т	6.00	.0
15	47.00	2.00	5.00	Monthspa				_)		F	6.00	.0
16		1.00		🛷 absent		Selection var	able:	le		Т	6.00	.0
17	43.00	1.00	5.00	🖋 underinf		Caso Labols:				F	6.00	.0
18	42.00	1.00	5.00	abstain	•		/			F	6.00	.0
19	29.00	1.00	5.00			WLS Weight	/					
20	48.00	1.00	2.00		•							
21	50.00	1.00	5.00				Onneal Link					
22	43.00	2.00	5.00			asie <u>R</u> esei	Cancel Hei	5		Т	6.00	.0
23	27.00	1.00	5.00	5.00	12.00	1.00	51.00	.0	Т	Т	6.00	.0
24	47.00	1.00	5.00	5.00	13.00	3.00	33.00	.0	F	F	.0	.0
25	55.00	1.00	5.00	5.00	19.00	1.00	72.00	1.00	Т	Т	6.00	.0
26	32.00	1.00	5.00	1.00	12.00	1.00	72.00	1.00	F	F	6.00	.0
27	39.00	1.00	5.00	5.00	16.00	1.00	30.00	.0	F	F	6.00	.0
28	33.00	1.00	5.00	3.00	12.00	1.00	102.00	.0	F			
29	36 00	1 00	5 00	5 00	16 90	1 00	58 00	1 00	F	Τ	6 00	0
Data View	Variable View					***						
Data view												
					_/				PASW Sta	tistics Processor i	s ready	
					1							

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			Std. Error of the					
		R Square	Adjusted R Square	Estimate					
1	.525ª	.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	u and a second se					
	Total	217.619	41							
2	Regression	59.604	2	29.802	7.355	.002 ^b				
	Residual	158.015	39	4.052	u					
	Total	217.619	41							

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

b. Predictors: (Constant), age, educ

c. Dependent Variable: monthsfu

			Coefficients			
Model				Standardized		
		Unstandardize	ed Coefficients	Coefficients		
		В	Std. Error	Beta	t	Sig.
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						Collinearity Statistics
		Beta In	t	Sig.	Partial Correlation	Tolerance
2	opcontac	045 ^a	319	.752	052	.954

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

b. Dependent Variable: monthsfu

Paul F. Cook, University of Colorado Denver, Center for Nursing Research

Revised 1/10 with SPSS (PASW) version 18.0