

Multivariate OLS Regression

Example 1:

I hypothesize that children who grow up in families with lots of rules will smoke fewer cigarettes per day than children who grow up in families that do not have lots of rules. I am going to control for a set of demographic variables that I think might also be associated with smoking behavior. These include race/ethnicity, sex, child's age, birth weight, and father involvement.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.213 ^a	.045	.036	8.108

This first box gives me the R-Square and adjusted R-Square values. Unless told otherwise, use the adjusted R-Square. The adjusted value takes into account the number of independent variables and the use of variables that do not add much to the analysis.

The R^2 tells me that the set of independent/control variables explains 3.6% of the variance in number of cigarettes smoked per day (so 96.4% of the variance is left unexplained).

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3149.862	10	314.986	4.792	.000 ^a
	Residual	66288.026	1008	65.737		
	Total	69437.887	1018			

This second box gives me the F-statistic for the analysis *as a whole*. We are not going to work much with this F-statistic, but you should be able to determine from the Sig. column whether the analysis is statistically significant. In this case, the analysis is significant at $p < .001$ ($.000 < .001$). So, we can say that we are 99.99% confident that there is a relationship between all my independent/control variables together, and the number of cigarettes smoked per day.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.697	.747		8.971	.000
	child is African American	-2.453	.754	-.106	-3.256	.001
	child is Hispanic	-5.071	4.353	-.036	-1.165	.244
	child is Asian American	-3.355	5.980	-.017	-.561	.575
	child is of some other race	-1.650	1.045	-.049	-1.579	.115
	child is female	-.864	.510	-.052	-1.693	.091
	Avg fam incm/needs thru the childs life	-.581	.097	-.194	-6.017	.000
	child low birth weight (under 51/2 lbs)	.936	1.047	.028	.893	.372
	father involv taking care of child	1.014	.546	.057	1.858	.064
	family has lots of rules	-.407	.511	-.025	-.795	.427
	age of child at 1997 PSID core interview	-.009	.071	-.004	-.128	.898

a. Dependent Variable: cigspcrdayall07

This third and final box gives me information about the relationship between each of the IV/control variables and the outcome. Unless told otherwise, you may interpret only the significant variables.

1. African American children smoke 2.5 fewer cigarettes daily than white children. This relationship is significant at $p = .001$. I am 99.99% confident that the relationship exists in the population.
2. Girls smoke .86 fewer cigarettes daily than boys. This relationship is significant at $p < .10$. I am 90% confident that the relationship exists in the population.
3. For every 1 unit increase in family income-to-needs, children are predicted to smoke .58 fewer cigarettes per day, 10 years later. This relationship is significant at $p < .001$. I am 99.99% confident that the relationship exists in the population.
4. Children whose fathers are involved in care-taking are predicted to smoke 1.04 more cigarettes per day 10 years later, than children whose fathers are not involved in care-taking. This relationship is significant at $p < .10$. I am 90% confident that the relationship exists in the population.

I can also use the data to make predictions by using the \hat{y} formula and plugging in values for each IV/control. Notice that you need to use all model variables, both significant and non-significant, to make predictions.

$$\hat{y} = 6.697 + (-2.453)(AA) + (-5.071)(Hisp) + (-3.355)(Asian) + (-1.650)(OthRace) + (-.864)(female) + (-.581)(inc) + (.936)(lbweight) + (1.014)(fathinv) + (-.407)(rules) + (-.009)(age)$$

So, a 15-year-old white male, whose family is at poverty level ($inc = 1$), who was not born low birthweight, had an uninvolved father, and whose family did not have many rules, is predicted to smoke 5.981 cigarettes per day.

If we looked at a white male who was the same as above in every respect except that his father was involved in caretaking when he was younger, we would predict smoking 6.995 cigarettes per day ($5.981 + 1.014$).

Example 2: Using a log dependent variable**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.229 ^a	.053	.037	.27697

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.863	7	.266	3.470	.001 ^a
	Residual	33.606	438	.077		
	Total	35.469	445			

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.155	.101		11.462	.000
	child is female	.023	.027	.041	.864	.388
	age of child at 1997 PSID core interview	.023	.010	.115	2.452	.015
	% of childhd in poverty	.032	.058	.031	.548	.584
	prnts use time out/tlk w/kid when prob	.062	.043	.069	1.460	.145
	finhdhd	.015	.006	.121	2.424	.016
	head of household is a female	-.019	.038	-.027	-.500	.617
	child have any phys/mntl cond limit	-.177	.081	-.103	-2.184	.030

a. Dependent Variable: logpsy07

The model above is statistically significant at $p = .001$; the full set of IV/controls explains 5.3% of the variance in the log of psychological well-being.

Because we are using a logged dependent variable, we interpret the IV/control variables as *percent of change*.

1. For each additional year of age for the child, we predict a 2.3% increase in psychological well-being.
2. For each additional year of education achieved by the head of household, we predict a 1.5% increase in the child's psychological well-being.
3. If the child grows up with physical or mental conditions that limit activity, we predict a 17.7% decline in psychological well-being.

Example 3:

Outcome = Log of Aggravation in Parenting Scale (2007)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.210 ^a	.044	.035	.33417

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.600	7	.514	4.606	.000 ^a
	Residual	77.992	698	.112		
	Total	81.592	705			

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.905	.104		8.697	.000
	child is female	.008	.026	.012	.323	.746
	femalehead07	.095	.031	.115	3.063	.002
	edhd07	-.014	.004	-.121	-3.238	.001
	bigcity07	.062	.037	.065	1.686	.092
	kidage	.001	.006	.007	.189	.850
	total number of probs doc mentioned	.024	.012	.077	2.039	.042
	urban07	.014	.030	.018	.469	.639

a. Dependent Variable: logpar07

So...if you want to lower your parenting aggravation, stay away from big cities (of 500,000 or more), get more education, live in a dual parent family, and have a child with few problems. Oh, and lest anyone tells you otherwise, girls are just as aggravating as boys (and vice versa).