

Exercises

Practicing the Basics

13.1. The regression equation relating y = education (number of years completed) to race ($z = 1$ for whites, $z = 0$ for nonwhites) in a certain country is $E(y) = 11 + 2z$. The regression equation relating education to race and to father's education (x) is $E(y) = 3 + 0.8x - 0.6z$.

(a) Ignoring father's education, find the mean education for whites, the mean education for nonwhites, and the difference between them.

(b) Plot the relationship between x and the mean of y for whites and for nonwhites.

(c) Controlling for father's education, find the difference between the mean education of whites and nonwhites. Illustrate by finding the mean education for each group when father's education equals 12 years.

13.2. Table 3.9 on page 65 showed data for several nations on y = CO2 emissions (in metric tons per capita) and x = per capita GDP (in thousands of dollars). Let z = whether the nation is in Europe ($1 = \text{yes}$, $0 = \text{no}$).

(a) The prediction equation for the effect of z is $\hat{y} = 10.61 - 2.48z$. Interpret the coefficients.

(b) The prediction equation for the effects of x and z is $\hat{y} = 2.10 + 0.22x - 3.58z$. Interpret the coefficients.

13.3. A regression analysis for the 100th Congress predicted the proportion of each representative's votes on abortion issues that took the "pro-choice" position.⁶ The prediction equation was

$$\hat{y} = 0.350 + 0.011id + 0.094r + 0.005nw + 0.005inc + 0.063s - 0.167p,$$

where r = religion (1 for non-Catholics), s = sex (1 for women), p = political party (1 for Democrats), id = ideology is the member's ADA score (ranging from 0 at most conservative to 100 at most liberal), nw = nonwhite is the percentage nonwhite of the member's district, and inc = income is the median family income of the member's district.

(a) Interpret the coefficient for political party.

(b) Using standardized variables, the prediction equation is

$$\hat{z}_y = 0.83z_{id} + 0.21z_r + 0.18z_{nw} + 0.05z_{inc} + 0.03z_s - 0.18z_p.$$

Comment on the relative sizes of the partial effects. Interpret the coefficient of ideology.

13.4. For 2014 data, the GSS website yields the prediction equation $\hat{y} = 9.59 + 0.166x_1 + 0.347x_2$ for y = highest year of school completed, x_1 = sex (1 = male, 2 = female), and x_2 = highest year of mother's education completed.

(a) Interpret the estimated partial effects.

(b) A more usual dummy coding for sex would be 0 = male and 1 = female. Would the estimated effects of x_1 and x_2 then change? Explain.

13.5. Based on a national survey, Table 13.16 shows results of a prediction equation for y = alcohol consumption, measured as the number of alcoholic drinks the subject drank during the past month.

(a) For x = alcohol consumption three years ago and dummy variables f for whether father died in the past three years, s for sex, and (m_1, m_2, m_3) for the four categories of marital status, report the prediction equation.

(b) Find the predicted alcohol consumption for a divorced male whose father died in the previous three years and whose consumption three years previously was (i) 0 drinks per month, (ii) 10 drinks per month.

TABLE 13.16

Explanatory Variable	Estimate	Std. Error
Intercept	8.3	
Death of father in past three years (0 = no)	9.8	2.9
Sex (0 = male)	-5.3	1.6
Marital status (0 = married)		
Divorced, separated	7.0	2.0
Widowed	2.0	3.6
Never married	1.2	2.4
Alcohol consumption three years ago	0.501	0.023

13.6. Consider the results in the previous exercise.

(a) Marital status has three estimates. Dividing the coefficient of the divorced dummy variable by its standard error yields a t statistic. What hypothesis does it test?

(b) What would you need to do to test the effect of marital status (all categories at once), controlling for the other variables?

13.7. For the Houses data file at the text website, Table 13.17 shows results of modeling y = selling price (in dollars) in terms of size of home (in square feet) and whether the home is new ($1 = \text{yes}$; $0 = \text{no}$).

TABLE 13.17

Parameter	Coef.	Std. Error	t	Sig
Intercept	-40230.867	14696.140	-2.738	.007
size	116.132	8.795	13.204	.000
new	57736.283	18653.041	3.095	.003

(a) Report and interpret the prediction equation, and form separate equations relating selling price to size for new and for not new homes.

⁶R. Tatalovich and D. Schier, *American Politics Quarterly*, vol. 21 (1993).

(b) Find the predicted selling price for a home of 3000 square feet that is (i) new, (ii) not new.

13.8. For the previous exercise, Table 13.18 shows results of fitting the model allowing interaction.

(a) Report the lines relating the predicted selling price to the size for homes that are (i) new, (ii) not new.

(b) Find the predicted selling price for a home of 3000 square feet that is (i) new, (ii) not new.

(c) Find the predicted selling price for a home of 1500 square feet that is (i) new, (ii) not new. Comparing to (b), explain how the difference in predicted selling prices changes as size of home increases.

TABLE 13.18

Parameter	Coef.	Std. Error	t	Sig
Intercept	-22227.808	15521.110	-1.432	.155
size	104.438	9.424	11.082	.000
new	-78527.502	51007.642	-1.540	.127
new#size	61.916	21.686	2.855	.005

13.9. Using software, replicate all the analyses shown in Sections 13.1 and 13.2 using the Income data file at the text website.

13.10. The software outputs in Table 13.19 show results of fitting two models to data from a study of the relationship

TABLE 13.19

NO INTERACTION MODEL						
Source	Partial SS	df	Mean Square	F	Sig	
race	40.08	2	20.04	1.07	.354	
register	2317.43	1	2317.43	123.93	.000	
Source	Sum of Squares	df	Mean Square	Parameter	Estimate	
Model	7936.734	3	2645.578	INTERCEPT	-2.7786	
Residual	673.166	36	18.699	REGISTER	0.7400	
Total	8609.900	39		RACE	a	-1.3106
					b	-2.8522
					ma	0.0000
INTERACTION MODEL						
Source	Partial SS	df	Mean Square	F	Sig	
race*register	53.79	2	27.89	1.47	.243	
Source	Sum of Squares	df	Mean Square	Parameter	Estimate	
Model	7990.523	5	1598.105	INTERCEPT	-8.245	
Residual	619.377	34	18.217	REGISTER	0.878	
Total	8609.900	39		RACE	a	6.974
					b	9.804
					ma	0.000
				REGISTER*RACE	a	-0.175
					b	-0.283
					ma	0.000

between y = percentage of adults voting, percentage of adults registered to vote, and racial-ethnic representation, for a random sample of 40 precincts in the state of Texas for a gubernatorial election. Racial-ethnic representation of a precinct is the group (Anglo, black, or Mexican American) having the strongest representation in a precinct.

(a) State a research question that could be addressed using these data.

(b) Report the prediction equation for the model assuming no interaction. Interpret the parameter estimates.

(c) Report the prediction equation for the model allowing interaction. Interpret the parameter estimates and describe the nature of the estimated interaction.

(d) Test whether the regression lines for the three categories have the same slope. Report the test statistic and P -value, and interpret.

(e) For the model assuming no interaction, test whether the mean voting percentages are equal for the three categories of racial-ethnic representation, controlling for percentage registered. Report the test statistic and P -value, and interpret.

(f) Report the test statistic and P -value for testing the null hypothesis that percentage voting and percentage registered are independent, controlling for racial-ethnic representation. Interpret.

13.20. Table 13.21 shows output for GSS data with y = index of attitudes toward premarital, extramarital, and homosexual sex, for which higher scores represent more permissive attitudes. The categorical explanatory variables are race (0 for whites, 1 for blacks), gender (0 for males, 1 for females), region (0 for South, 1 for non-South), and religion ($r_1 = 1$ for liberal Protestant sect, $r_2 = 1$ for conservative Protestant, $r_3 = 1$ for fundamentalist Protestant sect, $r_4 = 1$ for Catholic, $r_5 = 1$ for Jewish; no religious affiliation when $r_1 = \dots = r_5 = 0$). The quantitative explanatory variables are age, education (number of years), attendance at church (higher values represent more frequent attendance), and a variable for which higher values represent greater intolerance of freedom of speech for atheists and communists.

- (a) Based on the parameter estimates, give a profile of a person you would expect to be (i) least permissive, (ii) most permissive, with respect to sexual attitudes.
- (b) Summarize your main conclusions from studying the output.

TABLE 13.21

Analysis of Variance					
	Sum of Squares	df	Mean Square	F	
Regression	2583.326	12	215.277	54.098	
Residual	4345.534	1092	3.979		
R Square		0.373			
Variable	Coef.	Std. Error	Beta	t	Sig
(Constant)	9.373				
RACE	0.993	0.2040	0.125	4.869	.000
AGE	-0.029	0.0042	-0.189	-6.957	.000
SEX	-0.289	0.1230	-0.058	-2.353	.019
EDUC	0.073	0.0223	0.092	3.281	.001
REGION	0.617	0.1401	0.115	4.403	.000
ATTEND	-0.286	0.0255	-0.304	-11.217	.000
R1	-0.296	0.2826	-0.049	-1.048	.295
R2	-0.605	0.2782	-0.113	-2.174	.030
R3	-1.187	0.3438	-0.128	-3.454	.001
R4	-0.127	0.2856	0.023	0.446	.656
R5	0.521	0.4417	0.034	1.179	.238
FREESPCH	-0.465	0.0581	-0.227	-8.011	.000

13.21. You plan a study of factors associated with fertility (a woman's number of children) in a Latin American city. Of particular interest is whether migrants from other cities or migrants from rural areas differ from natives of the city in their family sizes. The groups to be compared are urban natives, urban migrants, and rural migrants. Since fertility is negatively related to educational level, and since edu-

cation might differ among the three groups, you control that variable. Table 13.22 shows some of the data for a random sample of married women above age 45. Analyze the complete data, which are the Fertility data file at the text website. In your report, provide graphical presentations as well as interpretations for all your analyses, and summarize the main results.

TABLE 13.22

Urban Natives		Urban Migrants		Rural Migrants	
Education	Fertility	Education	Fertility	Education	Fertility
0	7	0	7	0	4
0	5	0	6	0	6
1	5	0	7	0	10
1	4	1	5	0	8

13.22. Analyze the Houses2 data file at the text website by modeling selling price in terms of size of house and whether it is new.

- (a) Fit the model allowing interaction, and test whether the interaction term is needed in the model.
- (b) Construct a scatterplot, identifying the points by whether the home is new or not. The observation with the highest selling price is a new home that is somewhat removed from the general trend of points. Fit the interaction model after removing this single observation. Again,