Chapter 11-Multiple Regression

11.1 Predicting quality of life:

a) All other variables held constant, a difference of +1 degree in Temperature is associated with a difference of -.01 in perceived Quality of Life. A difference of $1000 in median income, again with all other variables held constant, is associated with a +.05 difference in perceived Quality of Life. A similar interpretation applies of b_3 and b_4. Since values of 0 cannot reasonably occur for all predictors, the intercept has no meaningful interpretation.

b) $\hat{Y}$ = 5.37 - .01(55) + .05(12) + .003(500) - .01(200) = 4.92

c) $\hat{Y}$ = 5.37 - .01(55) + .05(12) + .003(100) - .01(200) = 3.72

11.3 Religious Influence and religious Hope contribute significantly to the prediction, but not religious Involvement.

> It is worth pointing out here that even though religion Involvement does not contribute significantly to the multiple regression, it does have a significant simple correlation with Optimism. The matrix of correlations (where N = 600) is

	OPTIMISM	RELINVOL	RELINF	RELHOPE
OPTIMISM	1.0000 P= .	.1667 P= .000	.2725 P= .000	.2663 P= .000
RELINVOL	.1667 P= .000	1.0000 P= .	.4487 P= .000	.5439 P= .000
RELINF	.2725 P= .000	.4487 P= .000	1.0000 P= .	.4187 P= .000
RELHOPE	.2663 P= .000	.5439 P= .000	.4187 P= .000	1.0000 P= .

11.5 I would have speculated that religious Involvement was not a significant predictor because of its overlap with the other predictors, but the tolerances kick a hole in that theory to some extent.

> That's what happens when you ask a question before you are sure of the answer. ☹

11.7 Adjusted R^2 for 15 cases in Exercise 11.6:

$$R^2_{0.1234} = .173$$

$$\text{est } R^{*2} = 1 - \frac{(1-R^2)(N-1)}{(N-p-1)} = 1 - \frac{(1-.173)(14)}{(15-4-1)} = -.158$$

Since a squared value cannot be negative, we will declare it undefined. This is all the more reasonable in light of the fact that we cannot reject $H_0{:}R^* = 0$.

11.9 The multiple correlation between the predictors and the percentage of births under 2500 grams is .855. The incidence of low birthweight increases when there are more mothers under 17, when mothers have fewer than 12 years of education, and when mothers are unmarried. All of the predictors are associated with young mothers. (As the question noted, there are too few observations for a meaningful analysis of the variables in question.)

11.11 The multiple correlation between Depression and the three predictor variables was significant, with R = .49 [F(3,131) = 14.11, p = .0000]. Thus approximately 25% of the variability in Depression can be accounted for by variability in these predictors. The results show us that depression among students who have lost a parent through death is positively associated with an elevated level of perceived vulnerability to future loss and negatively associated with the level of social support. The age at which the student lost his or her parent does not appear to play a role.

11.13 The fact that the frequency of the behavior was not a factor in reporting in an interesting finding. My first thought would be that it is highly correlated with the Offensiveness, and that Offensiveness is carrying the burden. But a look at the simple correlation shows that the two variables are correlated at less than r = .20.

11.15 Using random variables as predictors:

I drew the following data directly from the random number tables in the appendix (and I didn't cheat).

Y	X_1	X_2	X_3	X_4	X_5
5	3	7	2	7	5
2	1	6	0	9	5
3	5	2	9	1	2
6	4	1	8	7	9
9	1	0	2	9	4
2	7	6	7	1	7
6	9	2	8	8	1
3	7	3	0	4	9
9	3	3	7	9	4
8	5	6	5	6	4

The multiple correlation for these data is .739, which is astonishingly high. Fortunately, the F test on the regression is not significant. Notice that we have only twice as many subjects as predictors.

This question is bound to lead to the question of how many cases we need per variable. There is no good answer to this question. Some will tell you that there should be at least 10 cases per predictor. I know of no argument in defense of such a rule. Harris (1985) has suggested a rule that says that *N* should exceed the number of predictors by at least 50. Cohen (1988) has argued from the point of view of power, and gives the example that a population correlation coefficient of .30 would require a sample size of 187 to have power = .80. This latter is sobering, but it is not a good argument here because we have not yet discussed power in any meaningful way.

11.17 Predicting weight:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-204.741	29.160		-7.021	.000
	height	5.092	.424	.785	12.016	.000
2	(Constant)	-88.199	43.777		-2.015	.047
	height	3.691	.572	.569	6.450	.000
	sex	-14.700	4.290	-.302	-3.426	.001

a. Dependent Variable: weight

11.19 The weighted average is 3.68, which is very close to the regression coefficient for Height when we control for Gender.

11.21 Sex is important to include in this relationship because women tend to be smaller than men, and thus probably have smaller, though not less effective, brains, but we probably don't want that contamination in our data. However, note that Sex was not significant in the previous answer, though the sample size (and hence power) is low.

11.23 I could argue that PctSAT is a nuisance variable because we are not particularly interested in the variable itself, but only in controlling it to allow us to have a clearer view of Expend, which is the variable in which we are interested. At the same time, it is an important contributor to the prediction of Combined, but we are led away from noticing that because of our predominant interest in Expend.

11.25 The scatterplot follows and shows that the squared correlation is .434, which is just what we found from the regression solution.