STA 6127: Solutions of Exercises for Chap. 13

1a. Use first equation (ignoring father's education) to get 11 + 2(1) = 13 for whites and 11 for nonwhites, for which the difference = 2. b. See p. 102a of solutions manual.
c. The coefficient of Z in the second equation, which is $-.6$.
d. Substitute $X = 12$ in second equation, yielding 12.0 for whites ($Z = 1$) and 12.6 for nonwhites ($Z = 0$).

2a. The predicted proportion of pro-choice votes increased by .005 for each 1% increase in the percentage nonwhite of the district, controlling for the other predictors in the model.
b. The predicted proportion of pro-choice votes was .063 higher for women members, controlling for the other predictors.
c. The predicted proportion of pro-choice votes was .167 lower for Democrats, controlling for the other predictors. No, the relationship could be very different when we ignore rather than control the other factors.
d. Ideology seems to be, by far, the most important predictor of proportion of pro-choice votes. A standard deviation increase in ideology corresponds to a .83 standard deviation predicted increase in the response, controlling for the other variables in the model.

3a. $\hat{Y} = 8.3 + 9.8F - 5.3G + 7.0M_1 + 2.0M_2 + 1.2M_3 + .501X$.
b. $\hat{Y} = 8.3 + 9.8(1) + 7.0(1) + .501(10) = 30.1$.
c. Predicted number of alcoholic drinks was 5.3 lower for females than for males, controlling for father's death, marital status, and alcohol consumption three years ago.
d. $t = -5.3/1.6 = -3.3$, $P = .001$; the mean alcohol consumption is lower for females than males, controlling for other variables.
e. $-5.3 \pm 1.96(1.6) = (-8.4, -2.2)$; the mean alcohol consumption is between 2.2 and 8.4 drinks per month lower for females than males, controlling for other variables.
f. It compares the mean for the divorced group to the baseline group (married), controlling for other predictors. To compare all four levels of marital status, we need to conduct the F test comparing this model to the simpler model without these three dummy variables.

4a. The regression model is $E(Y) = \alpha + \beta(X) + \beta_1 Z$. Here β denotes the partial effect of X on the mean of Y i.e it is the amount by which the price of a house would change for 1 unit change in its size controlling for the type (old or new). On the other hand, β_1 denotes the differences in the means of Y between the two categories of houses (old and new), controlling for size of home.
b. $\hat{Y} = -26.1 + 72.6X + 19.6Z$. The predicted selling price is 19.6 thousand dollars higher for new homes, controlling for size. The separate equations are $\hat{Y} = -26.1 + 72.6X$ for older homes and $\hat{Y} = -6.5 + 72.6X$ for new homes.
c. (i) $\hat{Y} = -6.5 + 72.6(3) = 211.2$, (ii) 191.6.
d. Adjusted mean for older homes would be larger, since mean selling price would increase if the average size were larger.

5a. (i) $\hat{Y} = -48.4 + 96.0X$, (ii) $\hat{Y} = -16.6 + 66.6X$; the coefficient 29.4 of the cross-product term is the difference between the two slopes.
b. (i) 239.6 thousand dollars, (ii) 183.2 thousand dollars; the predicted difference is larger than with the no interaction model.
c. (i) 95.6, (ii) 83.3; for the model allowing interaction, the difference in predicted prices increases as the size of home increases.
d. $t = 29.4/8.2 = 3.6$, $P = .0005$, which provides very strong evidence of interaction. There seems to be a greater slope for new homes; that is, the effect of size of home on price is greater for new homes.

6a. $\hat{Y} = -16.6 + 66.6X$ for older homes, and $\hat{Y} = -10.6 + 71.6X$ for new homes; now the lines are quite similar.
b. (i) 207.13 thousand dollars, (ii) 183.21 thousand dollars. On comparing the results with that of problem 13.4 (c) we conclude that the predicted difference is larger than with the no interaction model.

c. (i) 99.76 thousand dollars, (ii) 83.3 thousand dollars; for the model allowing interaction, the difference in predicted prices increases as the size of home increases.
d. $t = .497$, $P = .62$. Now, the no interaction model is adequate. The one outlying point is highly influential, the decision about whether the interaction term is needed depending on whether we include it in the analysis.

16a. The regression model for this analysis is $E(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 Z_1 + \beta_6 Z_2 + \beta_7 Z_3 + \beta_8 R_1 + \beta_9 R_2 + \beta_{10} R_3 + \beta_{11} R_4 + \beta_{12} R_5$. Here X_1 denotes age, X_2 denotes education, X_3 denotes church attendance, X_4 denotes tolerance level, Z_1 denotes race, Z_2 denotes gender, Z_3 denotes region and $R_1 - R_3$ have their usual meanings. The prediction equation is given by: $\hat{y} = 9.373 - 0.029X_1 + 0.073X_2 - 0.286X_3 - 0.465X_4 + 0.993Z_1$ - $0.289Z_2 + 0.617Z_3 - 0.296R_1 - 0.605R_2 - 1.187R_3 - 0.127R_4 + 0.521R_5$.
b. (i) Least permissive in sexual attitudes are old white females with low levels of education from the South who are fundamentalist Protestants and who are intolerant of freedom of speech. (ii) Most permissive are young black males with high levels of education from the non-South who are Jewish and are tolerant of freedom of speech.
c. The main conclusions from the given output are as follows: (i) Since the R^2 value is quite low, the given regression model does not provide good predictive power. (ii) Since the F-statistic value is quite large, I conclude that at least one of the above explanatory variables has a significant impact on the attitude towards sex. (iii) Since the p-values corresponding to all the variables except R1, R4 and R5 are very small, we conclude that all the other variables (except R1, R4 and R5) have a significant impact on the attitude towards sex. The standardized coefficients (in the column labelled misleadingly as "beta" suggest that church attendance has the largest effect on the response, controlling for the other variables.

22. See the figure on p. 109a. in the solutions manual

23. See Figure 13.3b in the textbook.

26. a