Week 6 homework problems

1. Consider the National Football League data set named `table.b1` in the `MPV` package. (Make sure you install `MPV` before trying to load it.) Use `?table.b1` to learn about its contents. Then do the following:

   a. Fit a multiple linear regression model relating the number of games won to the team's passing yardage (x_2), the percentage of rushing plays (x_7), and the opponents' yards rushing (x_8). Calculate t statistics for testing the hypotheses $H_0 : \beta_2 = 0$, $H_0 : \beta_7 = 0$, and $H_0 : \beta_8 = 0$. What conclusions can you draw about the role the variables x_2, x_7, and x_8 play in the model?

   b. Find a 95% confidence interval on β_7 and provide an interpretation of it.

   c. Find a 95% confidence interval on the mean number of games won by a team when $x_2 = 2300$, $x_7 = 56.0$, and $x_8 = 2100$.

   d. Find a 95% prediction interval on the number of games won by a new team when $x_2 = 2300$, $x_7 = 56.0$, and $x_8 = 2100$.

2. Data on last year's sales (y, in 100,000s of dollars) in 15 sales districts are given in the file "sales" posted on Canvas. This file also contains promotional expenditures (x_1, in thousands of dollars), the number of active accounts (x_2), the number of competing brands (x_3), and the district potential (x_4) for each of the districts.

   A model with all four regressors is proposed:

   $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + e, \quad e \sim N(0, \sigma^2)$$

   Test the following hypotheses:

   a. $\beta_4 = 0$

   b. $\beta_2 = \beta_3 = 0$

   c. $\beta_2 = \beta_3$

   d. $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

3. The variable Y is believed to be associated with the variables x_1, x_2, x_3, and x_4. All possible subsets of these variables are used in fitting a multiple linear regression model and the RSS (and its df) of the model is recorded below.

Variables included	RSS	df	Variables included	RSS	df
–	1300.6	57	x_2, x_3	376.75	55
x_1	1297.0	56	x_2, x_4	253.45	55
x_2	843.83	56	x_3, x_4	717.11	55
x_3	936.97	56	x_1, x_2, x_3	376.18	54
x_4	726.59	56	x_1, x_2, x_4	228.19	54
x_1, x_2	843.76	55	x_1, x_3, x_4	698.46	54
x_1, x_3	935.62	55	x_2, x_3, x_4	252.06	54
x_1, x_4	716.07	55	x_1, x_2, x_3, x_4	228.14	53

   a. Create an ANOVA table for the linear model $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$ using Type I sums of squares. Include F statistics and p-values for testing individual predictors.

b. Create an ANOVA table for the linear model $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$ using Type II sums of squares. Include F statistics and p-values for testing individual predictors.

c. What is $SSreg$ in both of the previous ANOVA tables, and in which table do the predictors' sums of squares add up to it?

d. What is R^2 in the full model containing x_1, x_2, x_3, and x_4?