Stat 6021: Guided Question Set 8

In this guided question set, we will use the "nfl.txt" data set that we used in the last module. As a reminder, the data are on NFL team performance from the 1976 season. The variables are:

- y: Games won (14-game season)
- x_1: Rushing yards (season)
- x_2: Passing yards (season)
- x_3: Punting average (yards/punt)
- x_4: Field goal percentage (FGs made/FGs attempted)
- x_5: Turnover differential (turnovers acquired minus turnovers lost)
- x_6: Penalty yards (season)
- x_7: Percent rushing (rushing plays/total plays)
- x_8: Opponents' rushing yards (season)
- x_9: Opponents' passing yards (season)

We will continue to regress the number of games won against three predictors: passing yards, x_2, percent rushing, x_7, and opponents' rushing yards in the season, x_8.

1. For this first question, you will generate partial regression plots for each of the predictors. As a reminder, a partial regression plot for predictor x_k is obtained by:
    - Regressing y against the other predictors, $x_1, \cdots, x_{k-1}$,that are already in the model and obtaining the residuals, $e(y|x_1, \cdots, x_{k-1})$.
    - Regressing the predictor in question, x_k, against the predictors that are already in the model and obtaining the residuals, $e(x_k|x_1, \cdots, x_{k-1})$.
    - Plotting the residuals against each other, $e(y|x_1, \cdots, x_{k-1})$ against $e(x_k|x_1, \cdots, x_{k-1})$.

(a) Produce the partial regression plot for x_2. Interpret what this partial regression is telling us.

(b) Fit a linear regression for the partial residual plot for x_2. Report the estimated coefficients.

(c) Fit a linear regression for the response against the three predictors. Report the estimated coefficient for x_2 and compare the value with the estimated slope from the previous part. What do you notice?

(d) Before producing the partial regression plots for x_7 and x_8, what do you think will be the values of the estimated coefficients for the linear regression for each of these plots?

(e) Produce the partial regression plots for x_7 and x_8. Interpret what both of these plots are telling us.

2. Produce plots of the residuals, studentized residuals, and externally studentized residuals (each against the fitted values for the multiple linear regression). Based on these, do we have any outliers?

3. Do we have any high leverage data points for this multiple linear regression? What teams are these?

4. Use $DFFITS_i$, $DFBETAS_{j,i}$, and Cook's distance to check for influential observations. What teams are influential?