Patriot’s Day – Multiple Linear Regression, Inferences

1. Setting:
   (a) $k$ independent variables $x_1, \ldots, x_k$. One dependent variable $y$.
   (b) $n$ data points: $(x_{11}, \ldots, x_{k1}, y_1), \ldots, (x_{1i}, \ldots, x_{ki}, y_i), \ldots, (x_{1n}, \ldots, x_{kn}, y_n)$.

2. The standard linear statistical model ($\text{lm}$ in R)
   (a) $y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + e_i$ (So $\beta_0 + \beta_1 x^*_1 + \cdots + \beta_k x^*_k$ is the mean value of $y$ for a fixed tuple $(x^*_1, \ldots, x^*_k)$ of independent variables.)
   (b) The errors, $e_i$ have mean 0, variance $\sigma^2$, and are independent.
   (c) The random variables $e_i$ have normal distributions.

3. Confidence intervals for the parameters $\beta_j$.
   (a) Statistical software gives estimates $b_j$ for $\beta_j$ and estimates $s_{b_j}$ of the standard deviation of $b_j$.
   (b) Then $\frac{b_j - \beta_j}{s_{b_j}}$ has a $t$-distribution with $n - (k + 1)$ degrees of freedom.
   (c) A confidence interval for any $\beta$ is $b + t^* s_b$ where $t^*$ is the critical value for a $t$-distribution with $n - (k + 1)$ degrees of freedom.
   (d) Similarly, we can get confidence intervals and prediction intervals for fixed $(x^*_1, \ldots, x^*_k)$.
   (e) Cautions: the declining confidence in multiple confidence intervals; the interpretation of a confidence interval in the presence of other coefficients.

4. An alternate way to interpret the output - Testing the “hypothesis” that $\beta = 0$.
   (a) If $\beta = 0$, $t = b/s_b$ has a $t$-distribution with $n - (k + 1)$ degrees of freedom.
   (b) The probability that a random variable with a $t$-distribution is at least as great as $t$ is called the $p$-value of the statistic.
   (c) The $p$-value is a measure of how “surprising” that a $t$-value this extreme occurred. If the $p$-value is very small, we doubt that $\beta = 0$. Otherwise, we think that $\beta = 0$ is consistent with the data.
   (d) Computing $p$-values is directly related to computing confidence intervals and a $p$-value doesn’t provide as much information as a confidence interval.

5. Adjusted $R^2$. $R^2$ will increase as we add more variables to the model. To correct for this, there is a statistic called adjusted $R^2$:
   \[
   \text{adj}-R^2 = 1 - \frac{\text{SSResid}/(n - (k + 1))}{\text{SSTotal}/(n - 1)}
   \]

Homework

Read Devore and Farnum Section 11.5, pages 525–527, 530–532.

1. In the data section of the course website is a datafile consisting of the scores of 32 students of Mathematics 222 on three tests and a final exam. In this question we investigate using the test scores to predict the final exam score. (After all, if the test scores do a good enough job, I wouldn’t have to grade the final exam!)
   (a) Write a linear function $\text{Exam} = b_0 + b_1 \text{Test1} + b_2 \text{Test2} + b_3 \text{Test3}$ that can be used to predict the final exam score from the three test scores.
   (b) Write a 95% confidence interval for the parameter $\beta_1$ in the model corresponding to our estimate $b_1$.
   (c) Do the $p$-values for the coefficients lead you to suspect that one or more of the $\beta_i$ are not very useful in the model? Explain.
   (d) For each of the three independent variables, fit a linear function that does not include that variable. Compare the values of adj-$R^2$ for each those models to each other and to the full model. Which model would you use to predict exam scores and why?
   (e) If a student scores 85 on each test, what is the predicted exam score? What is a 90% confidence interval for that prediction?