Review for Midterm Exam 1

Chapter 1: Linear Regression with One Predictor Variable

1. Understand the difference between a functional relationship and a statistical relationship among variables. Regression utilizes a statistical relationship to predict one variable from the other(s).

2. Understand the simple linear regression model.

   - $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, for $i = 1, 2, \ldots, n$, where $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.
   - You should know in the above model which components are parameters, which components are known, and which components are random variables.
   - You should know the three basic assumptions on the error terms of a linear regression model (independent, normal errors with constant variance).
   - You should know how to interpret the regression coefficients β_0 and β_1. In particular you should be aware of when β_0 actually has meaning (i.e. you should understand what is meant by scope of the model).

3. Understand the difference between unknowns and their estimates.

   - (X_i, Y_i) are both known (Y_i is a random variable)
   - $E(Y)$ is the mean response – essentially it is the unknown regression line which we estimate by the fitted value $\hat{Y}_i = b_0 + b_1 X_i$. The mean response is the mean (center) of the probability distribution of Y corresponding to a specific value of X.
   - Errors ε_i's are unknown; Residuals e_i's are the difference between the observed Y_i and the fitted value $\hat{Y}_i$. Residuals are used to study whether the assumptions of the model are satisfied.
   - The parameters β_k's are unknown and estimated by the b_k's.
   - The population variance σ^2 of the error terms is unknown and is estimated by the sample variance $s^2 = MSE$.

Chapter 2: Inferences in Regression Analysis

1. We may estimate parameters by either point estimates or interval estimates (confidence intervals).

2. The general form of a confidence interval is: Point Estimate ± Critical Value * SE(Point Estimate).

3. The general form of a test statistic is: [Point Estimate E(Point Est. under the null)]/ SE(Point Estimate).

4. Know how to compute the standard errors for the parameter estimates $b_0, b_1, etc.$, fitted values $\hat{Y}_h$, and predictions for new observations $\hat{Y}_{h(new)}$. Again, you will not be asked to calculate SS_X, but would need to know how to use it to calculate any of these standard errors, if it was given to you.

5. You should be able to discuss the differences between the estimated variance for the mean response and that for the prediction of a new observation.

6. Know how to form confidence intervals and hypothesis tests for the parameters, the mean response, and for prediction of new observations. Know how to state your hypotheses, calculate test statistics, degrees of freedom, implement a decision rule, and conclusions for a hypothesis test.

7. Understand the ANOVA table – degrees of freedom, SS, MS, and overall F-test for model significance.

8. Understand the General Linear Test approach. Know what is meant by Full and Reduced models. Know how to get degrees of freedom for this test.

9. Understand the coefficient of determination (R^2) and how to calculate it. It is a measure of the proportion of variation in Y that has been explained by our model. Also understand its limitations.

10. Understand the coefficient of correlation (r) and how to calculate it. It is a measure of the strength of linear association between Y and X. It can be positive or negative depending the nature of the relationship between Y and X.