STATS 500 - Homework 5

Due in class on **Wednesday**, October 18, 2017

1. Based on Problem 5 (p. 97)

Using the `cheddar` data, fit a model with taste as the response and the other three variables as predictors. Perform regression diagnostics on this model to answer the following questions. Display any plots that are relevant. **Do not provide any plots about which you have nothing to say.** Present your diagnostics in a logical order, which may not match the order of the questions below.

- Check the constant variance assumption for the errors and for evidence of non-linearity via residual plots, and adjust model as appropriate
- Check the normality assumption.
- Check for large leverage points.
- Check for outliers.
- Check for influential points.
- Check the structure of the relationship between the predictors and response.

Solutions to this problem should not exceed 5 pages.

2. Verify rigorously the equation on page 100 of textbook Faraway, that in the case of measurement error in simple linear regression, the expected value of the least-squares estimator is given by

$$E(\hat{\beta}_1) = \beta_1 \, \frac{\sigma_x^2 + \sigma_{x\delta}}{\sigma_x^2 + \sigma_\delta^2 + 2\sigma_{x\delta}}$$

where it is assumed that the only random variables are the errors ϵ_i, and that for this formula, σ_x^2, σ_δ^2, and $\sigma_{x\delta}$ are all the sampled versions, i.e.,

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^A - \overline{x^A})^2, \quad \sigma_\delta^2 = \frac{1}{n}\sum_{i=1}^{n}(\delta_i - \overline{\delta})^2, \quad \sigma_{x\delta} = \frac{1}{n}\sum_{i=1}^{n}(x_i^A - \overline{x^A})(\delta_i - \overline{\delta})$$

Hint: Utilize the standard formula for the least squares estimator of β_1 in simple linear regression being

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

where the x_i's in the above formula are x_i^O's.