Practice Problems

Stat 151A, Spring 2014

1. For each statement below, indicate whether it is TRUE or FALSE and give your reasoning

   (a) The OLS estimator of $\boldsymbol{\beta}$ in a linear regression model is unbiased even if the errors in the model are dependent.

   (b) The sum of the residuals is zero only when there is an intercept term in the linear model.

   (c) We are interested in finding high leverage points because such points are changing the fit of the model independent of the rest of the data.

2. Let h_i be the leverage point for an observation i.

   (a) Show that $\sum h_i = tr(H) = p + 1$.
   Hint: Recall that $tr(ABC) = tr(BCA) = tr(CAB)$ when the dimension of the matrices are such that all of those operations are well defined.

   (b) What implications does this have for interpreting large leverage values?

3. Consider the following model,

$$y = \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 \epsilon$$

   with the following data for each variable

$$\begin{aligned} \boldsymbol{y}^T &= \begin{pmatrix} 3 & 2 & 3 & 5 \end{pmatrix} \\ \boldsymbol{Z}_1^T &= \begin{pmatrix} 2 & 0 & 0 & 1 \end{pmatrix} \\ \boldsymbol{Z}_2^T &= \begin{pmatrix} 0 & 2 & 4 & 0 \end{pmatrix} \\ \boldsymbol{Z}_3^T &= \begin{pmatrix} 1 & 4 & 2 & -2 \end{pmatrix} \end{aligned}$$

   (a) Using the observed data, write down the model in terms of $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta}$

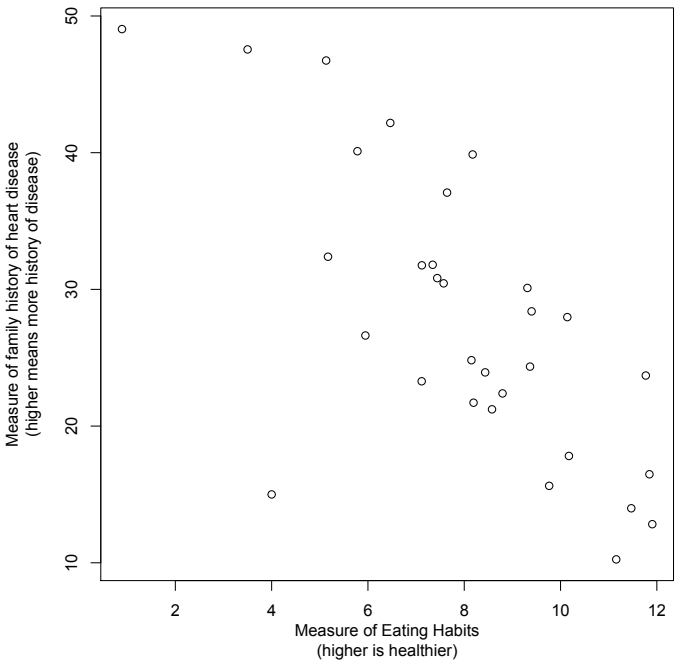
   (b) Derive the estimate $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ using our standard assumptions on the error terms.

   (c) Derive the t-statistic for testing the significance of each of the coefficients

4. A researcher wants to evaluate the effect of eating habits and family history on patients with heart conditions that agree to undergo a certain type of preventative treatment. She has started enrolling patients and has recorded information

on their current eating habits and family history. She plans to followup with these patients in a year and measure how successful the preventative treatment was and evaluate how their eating habits and family history influence the effectiveness of the treatment (assume the treatment does not require a change in their eating habits).

She shows you the data she has collected from the patients she has enlisted so far, in the form of a scatter plot of a measure of eating habits and family history.



(a) What problems might she have in drawing conclusions about how eating habits and family history influence the treatment?

(b) What would you recommend she do as she continues to enroll patients in her study?