STA 302f15 Assignment Twelve[1]

This assignment is preparation for the final exam; the homework questions are not to be handed in. The final exam may or may not have material from this assignment.

1. Regression diagnostics are mostly based on the residuals. This question compares the error terms ϵ_i to the residuals $\widehat{\epsilon}_i$. Answer True or False to each statement. For statements about the residuals, show a calculation that proves your answer. You may use anything on the formula sheet.

   (a) $E(\epsilon_i) = 0$

   (b) $E(\widehat{\epsilon}_i) = 0$

   (c) $Var(\epsilon_i) = 0$

   (d) $Var(\widehat{\epsilon}_i) = 0$

   (e) ϵ_i has a normal distribution.

   (f) $\widehat{\epsilon}_i$ has a normal distribution.

   (g) $\epsilon_1, \ldots, \epsilon_n$ are independent.

   (h) $\widehat{\epsilon}_1, \ldots, \widehat{\epsilon}_n$ are independent.

2. One of these statements is true, and the other is false. Pick one, and show it is true with a quick calculation. Start with something from the formula sheet.

   - $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\epsilon}}$
   - $\mathbf{y} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\epsilon}}$
   - $\widehat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \widehat{\boldsymbol{\epsilon}}$

   As the saying goes, "Data equals fit plus residual."

3. The *deleted residual* is $\widehat{\epsilon}_{(i)} = y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{(i)}$, where $\widehat{\boldsymbol{\beta}}_{(i)}$ is defined as usual, but based on the $n-1$ observations with observation i deleted.

   (a) Guided by an expression on the formula sheet, write the formula for the Studentized deleted residual. You don't have to prove anything. You will need the symbols $\mathbf{X}_{(i)}$ and $MSE_{(i)}$, which are defined in the natural way.

   (b) If the model is correct, what is the distribution of the Studentized deleted residual? Make sure you have the degrees of freedom right.

   (c) Why are numerator and denominator independent?

[1] Copyright information is at the end of the last page.

4. For the general linear regression model, are $\widehat{\mathbf{y}}$ and $\widehat{\boldsymbol{\epsilon}}$ independent?
   (a) Answer Yes or No and prove your answer.
   (b) What does this imply about the plot of predicted values against residuals?
5. For the general linear regression model, are $\mathbf{y}$ and $\widehat{\mathbf{y}}$ independent? Answer Yes or No and prove your answer.
6. For the general linear regression model, are $\mathbf{y}$ and $\widehat{\boldsymbol{\epsilon}}$ independent? Answer Yes or No and prove your answer.
7. For the general linear regression model, calculate $\mathbf{X}'\widehat{\boldsymbol{\epsilon}}$ one more time. This will help with the next question.
8. For the general linear regression model in which $\mathbf{X}$ is a matrix of constants,
   (a) Why does it not make sense to ask about independence of the independent variable values and the residuals?
   (b) Prove that the sample correlation between residuals and independent variable values must equal exactly zero.
   (c) Does this result depend on the correctness of the model?
   (d) What does the correlation between residuals and independent variable values imply about the corresponding plots?
