Statistics 206

Homework 4

Due: October 23, 2019, In Class

1. Confirm the formula for inverting a 2×2 matrix.

2. **Projection matrices**. Show the following are projection matrices, i.e., being symmetric and idempotent. Which linear subspace each of these matrices projects to? What are the ranks of these matrices? Here $\mathbf{H}$ is the hat matrix from a simple linear regression model with n cases (where the X values are not all equal).

   (a) $\mathbf{I}_n - \mathbf{H}$

   (b) $\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$

   (c) $\mathbf{H} - \frac{1}{n}\mathbf{J}_n$

3. Under the simple linear regression model, using matrix algebra, show that:

   (a) The residuals vector $\mathbf{e}$ is uncorrelated with the fitted values vector $\hat{\mathbf{Y}}$ and the LS estimator $\hat{\boldsymbol{\beta}}$.

   (b) With Normality assumption on the error terms, SSE is independent with SSR and the LS estimator $\hat{\boldsymbol{\beta}}$. (*Hint:* If $\mathbf{Z}$ is a multivariate Normal random vector, then $A\mathbf{Z}$ and $B\mathbf{Z}$ are jointly normally distributed.)

4. Derive $E(SSTO)$ and $E(SSR)$ under the simple linear regression model using matrix algebra.

5. **(Optional Problem.)** Under the simple linear regression model with Normal errors, derive the sampling distributions for SSR and $SSTO$ when $\beta_1 = 0$.

6. For each of the following regression models, answer whether it can be expressed as a general linear regression model or not. If so, indicate which transformations and/or new variables need to be introduced.

   (a) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \log X_{i2} + \beta_3 X_{i1}^2 + \epsilon_i$.

   (b) $Y_i = \epsilon_i \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2)$. $(\epsilon_i > 0)$

   (c) $Y_i = \beta_0 \exp(\beta_1 X_{i1}) + \epsilon_i$.

   (d) $Y_i = \{1 + \exp(\beta_0 + \beta_1 X_{i1} + \epsilon_i)\}^{-1}$.

7. Answer the following questions with regard to the general linear regression model and provide a brief explanation.

   (a) What is the maximum number of X variables that can be included in a general linear regression model used to fit a data set with 10 cases?

(b) With 4 predictors, how many X variables are there in the interaction model with all main effects and all interaction terms (2nd order, 3rd order, etc.)?

(c) Are the residuals uncorrelated? Do they have constant variance? How about the fitted values?