Statistics 206

Homework 4

Due: October 23, 2019, In Class

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

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

   Check if the following equality holds.

$$\frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

   $=$

$$\frac{1}{ad - bc} \begin{bmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{bmatrix}$$

   $=$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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? You can take $\mathbf{H}$ as the hat matrix from a simple linear regression model with n cases (where the X values are not all equal).

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

$$(\mathbf{I}_n - \mathbf{H})' = \mathbf{I}_n' - \mathbf{H}' = \mathbf{I}_n - \mathbf{H}$$

$$(\mathbf{I}_n - \mathbf{H})^2 = \mathbf{I}_n^2 - \mathbf{I}_n\mathbf{H} - \mathbf{H}\mathbf{I}_n + \mathbf{H}^2 = \mathbf{I}_n - \mathbf{H}$$

   It projects a vector onto the linear subspace of $\mathbf{R}^n$ that is orthogonal to the column space of X. Its rank is $n - p = n - 2$.

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

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

$$(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n)^2 = \mathbf{I}_n^2 - \mathbf{I}_n\frac{1}{n}\mathbf{J}_n - \frac{1}{n}\mathbf{J}_n\mathbf{I}_n + \frac{1}{n^2}\mathbf{J}_n^2 = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$$

   It projects a vector onto the linear subspace of $\mathbf{R}^n$ that is orthogonal to the subspace spanned by $\mathbf{1}_n$. Its rank is $n - 1$.

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

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

$$(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)^2 = \mathbf{H} - \frac{1}{n}\mathbf{J}_n\mathbf{H} - \mathbf{H}\frac{1}{n}\mathbf{J}_n + \frac{1}{n^2}\mathbf{J}_n^2 = \mathbf{H} - \frac{1}{n}\mathbf{J}_n\mathbf{H} - \mathbf{H}\frac{1}{n}\mathbf{J}_n + \frac{1}{n}\mathbf{J}_n = \mathbf{H} - \frac{1}{n}\mathbf{J}_n$$

since $\mathbf{J}_n\mathbf{H} = \mathbf{J}_n$

$\mathbf{J}_n\mathbf{H} = \mathbf{J}_n$ because $\mathbf{H}$ is the projection matrix onto the column space of X and every column of $\mathbf{J}_n$, namely $\mathbf{1}_n$, is in the column space of X.

It projects a vector onto the linear subspace of column space of X that is orthogonal to the subspace spanned by $\mathbf{1}_n$. Its rank is $p - 1 = 1$.

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 $\hat{\boldsymbol{\beta}}$.

Proof.

$$e = (I - H)Y, \quad \hat{\beta} = (X'X)^{-1}X'Y$$

$$\mathrm{Cov}(e, \hat{\beta}) = (I - H)\mathrm{Cov}(Y)((X'X)^{-1}X')' = \sigma^2(I - H)X(X'X)^{-1} = 0$$

since $(I - H)X = X - X = 0$. Therefore $\hat{\beta}$ and the residuals e are uncorrelated.
Also, $\hat{Y} = X\hat{\beta}$.
Hence, $\mathrm{Cov}(\hat{Y}, e) = \mathrm{Cov}(X\hat{\beta}, e) = X\mathrm{Cov}(\hat{\beta}, e) = 0$.
Therefore $\hat{Y}$ and the residuals e are uncorrelated. □

(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.)

Proof. Clearly, $e = (I_n - H)Y$ and $d = (H - \frac{1}{n}J_n)Y$ are jointly normally distributed from Hint. Also $\mathrm{Cov}(e, d) = (I_n - H)\mathrm{Var}(Y)(H - \frac{1}{n}J_n) = \sigma^2(H - H^2 - \frac{1}{n}J_n + H\frac{1}{n}J_n) = 0$ as $H^2 = H$ and $HJ_n = J_n$ as they are projection matrices.
Since e and d are jointly normally distributed and uncorrelated, they are independent. Hence, $SSE = e^T e$ and $SSR = d^T d$ being functions of e and d are also independent. From part (a), e and $\hat{\beta}$ are uncorrelated and using Hint they are jointly normal. Hence e and $\hat{\beta}$ are independent and so is $SSE = e^T e$ and $\hat{\beta}$, SSE being a function of e. □

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

algebra.

$$
\begin{aligned}
E(SSTO) &= E\{Y'(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n)Y\} \\
&= E\{Tr((\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n)YY')\} \\
&= Tr\{(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n)E(YY')\} \\
&= Tr\{(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n)(\sigma^2\mathbf{I}_n + X\beta\beta'X')\} \\
&= (n-1)\sigma^2 + Tr(\beta'X'(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n)X\beta) \\
&= (n-1)\sigma^2 + Tr(\beta'X'(\mathbf{I}_n - \mathbf{H})X\beta) + Tr(\beta'X'(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)X\beta) \\
&= (n-1)\sigma^2 + 0 + \beta_1^2\sum(X_i - \overline{X})^2 \quad \text{by } (\mathbf{I}_n - \mathbf{H})X = 0 \text{ and next part} \\
&= (n-1)\sigma^2 + \beta_1^2\sum(X_i - \overline{X})^2
\end{aligned}
$$

$$
\begin{aligned}
E(SSR) &= E\{Y'(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)Y\} \\
&= E\{Tr((\mathbf{H} - \frac{1}{n}\mathbf{J}_n)YY')\} \\
&= Tr\{(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)E(YY')\} \\
&= Tr\{(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)(\sigma^2\mathbf{I}_n + X\beta\beta'X')\} \\
&= (2-1)\sigma^2 + Tr(\beta'X'(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)X\beta) \\
&= \sigma^2 + Tr(\beta'X'X\beta - \beta'X'\frac{1}{n}\mathbf{J}_nX\beta) \quad \text{since } \mathbf{H}X = X \\
&= \sigma^2 + \beta'X'X\beta - \beta'X'\frac{1}{n}\mathbf{J}_nX\beta \\
&= \sigma^2 + (n\beta_0^2 - 2\beta_1\sum X_i + \beta_1^2\sum X_i^2) - (n\beta_0^2 - 2\beta_1\sum X_i + n\beta_1^2(\overline{X}_i)^2) \\
&= \sigma^2 + \beta_1^2\sum(X_i - \overline{X})^2
\end{aligned}
$$

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

When $\beta_1 = 0$, $\boldsymbol{X\beta} = \beta_0\mathbf{1}_n$.

$$
\begin{aligned}
SSR &= Y^{'}(H - \frac{1}{n}J_n)Y \\
&= d'd
\end{aligned}
$$

Where $d = (H - \frac{1}{n}J_n)Y = (H - \frac{1}{n}J_n)(\beta_0 1_n + \epsilon) = (H - \frac{1}{n}J_n)\epsilon$, since $(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J}_n)\mathbf{1}_n = \mathbf{0}_n$. Thus,

$$SSR = \epsilon'(H - \frac{1}{n}J_n)\epsilon$$

Let $z = Q\epsilon$, then

$$SSR = \epsilon'\mathbf{Q}'\boldsymbol{\Lambda}\mathbf{Q}\epsilon = z'\boldsymbol{\Lambda}z = \sum_{1}^{p-1} z_i^2$$

$$E(z) = E(Q\epsilon) = 0_n, Var(z) = var(Q\epsilon) = Q'var(\epsilon)Q = \sigma^2 I_n$$

Under normal error model, z_i are *iid* $N(0, \sigma^2)$. Thus, $SSR \sim \sigma^2\chi_1^2$

Similar for SSTO.

$$\begin{aligned}
SSTO &= \boldsymbol{Y}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{Y} \\
&= (\beta_0 1_n + \epsilon)'(I_n - \frac{1}{n}J_n)(\beta_0 1_n + \epsilon) \\
&= \epsilon'(I_n - \frac{1}{n}J_n)\epsilon \\
&= \epsilon'\mathbf{Q}'\boldsymbol{\Lambda}\mathbf{Q}\epsilon \\
&= (Q\epsilon)'\boldsymbol{\Lambda}(Q\epsilon) \\
&= \sum_{1}^{n-1} z_i^2
\end{aligned}$$

Under normal error model, z_i are *iid* $N(0, \sigma^2)$. Thus, $SSTO \sim \sigma^2\chi_{n-1}^2$

6. For each of the following regression models, indicate whether it can be expressed as a general linear regression model. If so, indicate which transformations and/or new variables need to be introduced.

   (a) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \log X_{i2} + \beta_3 X_{i1}^2 + \epsilon_i$.
   Yes. Define $\tilde{X}_{i2} = \log X_{i2}, X_{i3} = X_{i1}^2$, then

   $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \tilde{X}_{i2} + \beta_3 X_{i3} + \epsilon_i$$

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

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

   define $\tilde{Y}_i = \log(Y_i), \tilde{X}_{i2} = X_{i2}^2$, and $\tilde{\epsilon}_i = \log(\epsilon_i)$,

   $$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \tilde{X}_{i2} + \tilde{\epsilon}_i$$

   (c) $Y_i = \beta_0 \exp(\beta_1 X_{i1}) + \epsilon_i$.
   No.

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

$$\log(1/Y_i - 1) = \beta_0 + \beta_1 X_{i1} + \epsilon_i$$

define $\tilde{Y}_i = \log(1/Y_i - 1)$,

$$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$$

7. Answer the following questions with regard to the general linear regression model and explain your answer.

(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?
Here $n = 10$ and we know $p \leq n - 1 = 9$. So maximum value of p is 9. The maximum number of X variables is $p - 1 = 8$.

(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.)?

$$2^4 - 1 = 15$$

(c) Are the residuals uncorrelated? Do they have constant variance? How about the fitted values?
The residuals have variance covariance matrix $\sigma^2(\mathbf{I}_n - \mathbf{H})$. They are correlated unless $\mathbf{H}$ is diagonal. They do not have constant variance unless the diagonal terms of $\mathbf{H}$ are constant. The fitted values have variance covariance matrix $\sigma^2\mathbf{H}$. They are correlated unless $\mathbf{H}$ is diagonal. They do not have constant variance unless the diagonal terms of $\mathbf{H}$ are constant.

If $\mathbf{H}$ is diagonal, then it must be $\mathbf{I_n}$ due to the fact that $\mathbf{H}$ is a projection matrix and $\mathbf{1}$ is in the space it projects to. However, $\mathbf{H} = \mathbf{I_n}$ could only possibly happen when $p = n$.