Statistics 206

Homework 4 (Solution)

Due : Oct. 27, 2022, 11:59PM

Instructions:

- You should upload homeworkX files on canvas (under "Assignments/hwX") before its due date.
- Your homework may be prepared by a word processor (e.g., Latex) or through handwriting.
- For handwritten homework, you should either scan or take photos of your homework: Please make sure the pages are clearly numbered and are in order and the scans/photos are complete and clear; Check before submitting.
- Please name the files following the format: "FirstName-LastName-HwX". If there are several files, you can use "-Questions1-5", "-Questions6", etc., to distinguish them. E.g., "Jie-Peng-Hw1-Questions1-5.pdf", "Jie-Peng-Hw1-Questions6.rmd".
- Your name should be clearly shown on the submitted files: By putting on your name, you also acknowledge that you are the person who did and prepared the submitted homework.
- **Optional Problems** are more advanced and are not counted towards the grade.
- Showing/sharing/uploading homework or solutions outside of this class is prohibited.

1. Derive $E[SSTO]$ and $E[SSR]$ under the simple linear regression model using matrix algebra.

ANS. Denote $\boldsymbol{\mu} = E[\boldsymbol{Y}] = \boldsymbol{X\beta}$. First, note that for any $n \times n$ matrix $\boldsymbol{A}$,

$$E[(\boldsymbol{Y}-\boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{Y}-\boldsymbol{\mu})] = E[\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}] - \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}.$$

Using the commutative property of the trace of a matrix,

$$\begin{aligned} E[(\boldsymbol{Y}-\boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{Y}-\boldsymbol{\mu})] &= E[\mathrm{tr}(\boldsymbol{Y}-\boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{Y}-\boldsymbol{\mu})] = E[\mathrm{tr}\boldsymbol{A}(\boldsymbol{Y}-\boldsymbol{\mu})(\boldsymbol{Y}-\boldsymbol{\mu})'] \\ &= \mathrm{tr}(\boldsymbol{A}E[(\boldsymbol{Y}-\boldsymbol{\mu})(\boldsymbol{Y}-\boldsymbol{\mu})']) = \mathrm{tr}(\boldsymbol{A}\mathrm{Var}(\boldsymbol{Y})). \end{aligned}$$

It follows that

$$E[\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}] = \mathrm{tr}(\boldsymbol{A}\mathrm{Var}(\boldsymbol{Y})) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}.$$

Denote $\boldsymbol{P}_0 = \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n' = \frac{1}{n}\boldsymbol{J}_n$.

$$\begin{aligned} E[SSTO] &= E[\boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{Y}] \\ &= \mathrm{tr}((\boldsymbol{I}_n - \boldsymbol{P}_0)\sigma^2\boldsymbol{I}_n) + \boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{X}\boldsymbol{\beta} \end{aligned}$$

Note that $\boldsymbol{X} = (\mathbf{1}_n\ \boldsymbol{x})$ and hence $\boldsymbol{X}\boldsymbol{\beta} = \beta_0\mathbf{1}_n + \beta_1\boldsymbol{x}$. Also, $\boldsymbol{P}_0\mathbf{1}_n = \mathbf{1}_n$ and $(\boldsymbol{I}_n - \boldsymbol{P}_0)\mathbf{1}_n = \mathbf{0}$,

$$\begin{aligned} E[SSTO] &= \sigma^2\mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{P}_0) + \beta_1^2\boldsymbol{x}'(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{x} \\ &= (n-1)\sigma^2 + \beta_1^2\sum(x_i - \overline{x})^2. \end{aligned}$$

Similarly, using the fact that $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \mathbf{0}$,

$$\begin{aligned} E[SSE] &= E[\boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y}] \\ &= \sigma^2\mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{H}) + \boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} \\ &= (n-2)\sigma^2. \end{aligned}$$

$E[SSR] = E[SSTO] - E[SSE] = \sigma^2 + \beta_1^2\sum(x_i - \overline{x})^2$, or

$$\begin{aligned} E[SSR] &= E[\boldsymbol{Y}'(\boldsymbol{H} - \boldsymbol{P}_0)\boldsymbol{Y}] \\ &= \sigma^2\mathrm{tr}(\boldsymbol{H} - \boldsymbol{P}_0) + \boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{H} - \boldsymbol{P}_0)\boldsymbol{X}\boldsymbol{\beta} \\ &= \sigma^2 + \beta_1^2\sum(x_i - \overline{x})^2. \end{aligned}$$

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

ANS. Since $\beta_1 = 0$, $\boldsymbol{X}\boldsymbol{\beta} = \beta_0\mathbf{1}_n$. Using the fact that $(\boldsymbol{H} - \boldsymbol{P}_0)\mathbf{1}_n$

$$SSR = \boldsymbol{Y}'(\boldsymbol{H} - \boldsymbol{P}_0)\boldsymbol{Y} = (\boldsymbol{Y} - \beta_0\mathbf{1}_n)'(\boldsymbol{H} - \boldsymbol{P}_0)(\boldsymbol{Y} - \beta_0\mathbf{1}_n) = \boldsymbol{\epsilon}'(\boldsymbol{H} - \boldsymbol{P}_0)\boldsymbol{\epsilon}.$$

As we have shown in HW3, $\boldsymbol{H} - \boldsymbol{P}_0$ is a projection matrix with rank $p - 1 = 1$. Therefore it can be diagonalized as $\boldsymbol{H} - \boldsymbol{P}_0 = \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}$, where $\boldsymbol{Q}$ is an orthogonal matrix, Λ is a $n \times n$ diagonal matrix with $(p-1)$ entries being 1 and $(n-p+1)$ entries being 0. Therefore,

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

where $\boldsymbol{z} \sim N(0, \sigma^2\boldsymbol{I}_n)$, namely z_i are i.i.d. $N(0, \sigma^2)$. Therefore $SSR \sim \sigma^2\chi_1^2$. (This can also be obtained directly from the Cochran's theorem.)

Similarly,

$$SSTO = \boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{Y} = (\boldsymbol{Y} - \beta_0\mathbf{1}_n)'(\boldsymbol{I}_n - \boldsymbol{P}_0)(\boldsymbol{Y} - \beta_0\mathbf{1}_n) = \boldsymbol{\epsilon}'(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{\epsilon}.$$

Here $\boldsymbol{I}_n - \boldsymbol{P}_0$ is a projection matrix with rank $n - 1$. Using the same argument above, $SSTO \sim \sigma^2\chi_{n-1}^2$.

Alternatively, from the fact that $SSE \sim \sigma^2\chi_{n-2}^2$ (This holds regardless of whether $\beta_1 = 0$), and that SSE and SSR are independent, and that $SSTO = SSE + SSR$, we also have $SSTO \sim \sigma^2\chi_{n-1}^2$.

3. For each of the following models, answer whether it can be expressed as a multiple regression model or not. If so, indicate which transformations and/or new variables need to be introduced. (ϵ_i's denote the error terms)

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

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

   ANS. Yes.

   $$\log(Y_i) = \beta_0 + E[\log(\epsilon_i)] + \beta_1 X_{i1} + \beta_2 X_{i2}^2 + \log(\epsilon_i) - E[\log(\epsilon_i)].$$

   Define $\tilde{Y}_i = \log(Y_i), \tilde{X}_{i2} = X_{i2}^2$, $\tilde{\epsilon}_i = \log(\epsilon_i) - E[\log(\epsilon_i)]$, and $\tilde{\beta}_0 = \beta_0 + E[\log(\epsilon_i)]$; then

   $$\tilde{Y}_i = \tilde{\beta}_0 + \beta_1 X_{i1} + \beta_2 \tilde{X}_{i2} + \tilde{\epsilon}_i$$

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

   ANS. No.

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

   ANS. Yes.

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

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

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

4. Answer the following questions with regard to multiple regression models and provide a brief explanation.

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

   ANS. Here $n = 10$ and in order to the fit the model, the design matrix $X_{n\times p}$ must be of full column rank p. Since the rank of a matrix can not exceed its number of rows (which is n here), so $p \le n$ and so the number of X variables $p - 1 \le n - 1 = 9$. If one considers to estimate σ^2, the number of X variables $p - 1 < n - 1 = 9$.

   (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.)?

   ANS. A straightforward way is simply listing all of them and count the number. Alternatively, by the binomial theorem,

   $$\sum_{k=1}^{4} \binom{4}{k} = 15.$$

   Or, consider four variables they are either power 0 or power 1,

   $$2^4 - 1 = 15$$

5. Tell true or false of the following statements with regard to multiple regression models.

   (a) The multiple coefficient of determination R^2 is always larger/not-smaller for models with more X variables.

ANS. **False / True [obscure question]**. When the X variables in a smaller model are not entirely nested within the larger model, then it is possible that the R^2 of the smaller model is larger than the R^2 of the larger model.

   (b) If all the regression coefficients associated with the X variables are estimated to be zero, then $R^2 = 0$.

ANS. **True**. The fitted regression surface is horizontal so $\widehat{Y}_i = \overline{Y}$ for all i and thus $SSR = 0$, then $R^2 = \frac{SSR}{SSTO} = 0$.

   (c) The adjusted multiple coefficient of determination R_a^2 may decrease when adding additional X variables into the model.

ANS. **True**. $R_a^2 = 1 - \frac{n-1}{n-p}\frac{SSE}{SSTO}$, the decrease in SSE may be more than offset by the loss of degrees of freedom in the denominator $n - p$.

   (d) Models with larger R^2 is always preferred.

ANS. **False**. Models with larger R^2 may have a lot of X variables that are unrelated to the response variable or are highly correlated with each other which just over-fit the data and make prediction and interpretation difficult.

   (e) If the response vector is a linear combination of the columns of the design matrix $\mathbf{X}$, then the coefficient of multiple determination $R^2 = 1$.

ANS. **TRUE**. Since now $Y_i = \widehat{Y}_i$ for all $i = 1, \cdots, n$. Thus all residuals $e_i \equiv 0$ and then $SSE = 0$.

6. **Multiple regression by matrix algebra in R**.

   See solution to Question 6.

7. Under the multiple regression model, show that the residuals are uncorrelated with the fitted values and the estimated regression coefficients.

Proof.

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

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

since $(I - \mathbf{H})X = X - X = 0$. Therefore $\hat{\beta}$ and the residuals e are uncorrelated.

- $\hat{Y} = X\hat{\beta}$. Hence, $\text{Cov}(\hat{Y}, e) = \text{Cov}(X\hat{\beta}, e) = X\text{Cov}(\hat{\beta}, e) = 0$. Therefore $\hat{Y}$ and the residuals e are uncorrelated.
- (Alternative)

$$\hat{Y} = \mathbf{H}Y, \quad e = (I - \mathbf{H})Y$$

$$\text{Cov}(\hat{Y}, e) = \text{Cov}(\mathbf{H}Y, (I - \mathbf{H})Y) = H\text{Cov}(Y)(I - \mathbf{H})^t = \sigma^2\mathbf{H}(I - \mathbf{H}) = 0$$

  since $\mathbf{H}(I - \mathbf{H}) = \mathbf{H} - \mathbf{H} = 0$. Therefore $\hat{Y}$ and the residuals e are uncorrelated.

$\square$