# MTH 207A: Assignment 4

**P 1.** Consider the one–way ANOVA model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ i = 1, \ldots, K,\ j = 1, \ldots, n_i \tag{1}$$

where $\epsilon_{ij}$ are i.i.d. $N(0, \sigma^2)$.

(a) Find the residual sum of squares ($RSS$) for the model (1).

(b) Find the distribution of $RSS/\sigma^2$.

**Solution:** For the model (1), the design matrix X is

$$\mathrm{X} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{1}_{n_2} & \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_K} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_K} \end{bmatrix}_{n \times (K+1)}.$$

where $n = \sum_{i=1}^{K} n_i$.

Since X is not a full rank matrix, the least square estimator of $\boldsymbol{\beta} = (\mu, \alpha_1, \ldots, \alpha_K)'$ is given by

$$\widehat{\boldsymbol{\beta}} = (\mathrm{X}'\mathrm{X})^{-}\mathrm{X}\mathbf{y}.$$

First we shall compute $\widehat{\boldsymbol{\beta}}$.

$$\mathrm{X}'\mathrm{X} = \begin{bmatrix} n & n_1 & n_2 & \cdots & n_K \\ n_1 & n_1 & 0 & \cdots & 0 \\ n_2 & 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_K & 0 & 0 & \cdots & n_K \end{bmatrix}.$$

It can be shown that

$$(\mathrm{X}'\mathrm{X})^{-} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1/n_1 & 0 & \cdots & 0 \\ 0 & 0 & 1/n_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/n_K \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0}' & \mathrm{D}^{-1} \end{bmatrix}.$$

where

$$D = \begin{bmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_K \end{bmatrix}.$$

Also,

$$\mathrm{X}'\mathbf{y} = \begin{bmatrix} \sum_{i=1}^{K} \sum_{j=1}^{n_i} y_{ij} \\ \sum_{j=1}^{n_1} y_{1j} \\ \vdots \\ \sum_{j=1}^{n_K} y_{Kj} \end{bmatrix}.$$

Therefore,

$$\widehat{\boldsymbol{\beta}} = (\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathbf{y} = \begin{bmatrix} 0 \\ \sum_{j=1}^{n_1} y_{1j}/n_1 \\ \vdots \\ \sum_{j=1}^{n_K} y_{Kj}/n_K \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{y}_1 \\ \vdots \\ \bar{y}_K \end{bmatrix}.$$

Thus $\widehat{\mu} = 0$ and $\widehat{\alpha} = \bar{y}_i$ for $i = 1, \ldots, K$. Now

$$\begin{aligned} RSS &= \sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \widehat{\alpha})^2 = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 \\ &= \sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij}^2 + (\bar{y}_i)^2 - 2y_{ij}\bar{y}_i) = \sum_{i=1}^{K}(\sum_{j=1}^{n_i} y_{ij}^2 - n_i(\bar{y}_i)^2) \end{aligned}$$

Now observe that

$$\begin{aligned} n_i(\bar{y}_i)^2 &= n_i(\bar{y}_i)'(\bar{y}_i) \\ &= \left[(1/n_i)\left[\mathbf{0}'_{n_1} \quad \cdots \quad \mathbf{0}'_{n_{i-1}} \quad \mathbf{1}'_{n_i} \quad \mathbf{0}'_{n_{i+1}} \quad \cdots \quad \mathbf{0}'_{n_K}\right]\mathbf{y}\right]' \\ &\quad \times\left[(1/n_i)\left[\mathbf{0}'_{n_1} \quad \cdots \quad \mathbf{0}'_{n_{i-1}} \quad \mathbf{1}'_{n_i} \quad \mathbf{0}'_{n_{i+1}} \quad \cdots \quad \mathbf{0}'_{n_K}\right]\mathbf{y}\right] \\ &= \frac{1}{n_i}\mathbf{y}'\begin{bmatrix} \mathbf{0}_{n_1} \\ \cdots \\ \mathbf{0}_{n_{i-1}} \\ \mathbf{1}_{n_i} \\ \mathbf{0}_{n_{i+1}} \\ \cdots \\ \mathbf{0}_{n_K} \end{bmatrix}\left[\mathbf{0}'_{n_1} \quad \cdots \quad \mathbf{0}'_{n_{i-1}} \quad \mathbf{1}'_{n_i} \quad \mathbf{0}'_{n_{i+1}} \quad \cdots \quad \mathbf{0}'_{n_K}\right]\mathbf{y} \\ &= \mathbf{y}'\mathrm{E}_{n_i}\mathbf{y} \end{aligned}$$

where

$$\mathrm{E}_{n_i} = \begin{bmatrix} \mathrm{Z}_{n_1\times n_1} & \mathrm{Z}_{n_1\times n_2} & \cdots & \mathrm{Z}_{n_1\times n_i} & \cdots & \mathrm{Z}_{n_1\times n_K} \\ \mathrm{Z}_{n_2\times n_1} & \mathrm{Z}_{n_2\times n_2} & \cdots & \mathrm{Z}_{n_2\times n_i} & \cdots & \mathrm{Z}_{n_2\times n_K} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathrm{Z}_{n_{i-1}\times n_1} & \mathrm{Z}_{n_{i-1}\times n_2} & \cdots & \mathrm{Z}_{n_{i-1}\times n_i} & \cdots & \mathrm{Z}_{n_{i-1}\times n_K} \\ \mathrm{Z}_{n_i\times n_1} & \mathrm{Z}_{n_i\times n_2} & \cdots & (1/n_i)\mathrm{J}_{n_i\times n_i} & \cdots & \mathrm{Z}_{n_i\times n_K} \\ \mathrm{Z}_{n_{i+1}\times n_1} & \mathrm{Z}_{n_{i+1}\times n_2} & \cdots & \mathrm{Z}_{n_{i+1}\times n_i} & \cdots & \mathrm{Z}_{n_{i+1}\times n_K} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \mathrm{Z}_{n_K\times n_1} & \mathrm{Z}_{n_K\times n_2} & \cdots & \mathrm{Z}_{n_K\times n_i} & \cdots & \mathrm{Z}_{n_K\times n_K} \end{bmatrix}$$

$\mathrm{Z}_{n_i\times n_j}$ is $n_i \times n_j$ matrix of all zeros and $\mathrm{J}_{n_i\times n_j}$ is $n_i \times n_j$ matrix of all ones.

Thus

$$RSS = \mathbf{y}'\mathbf{y} - \sum_{i=1}^{K}\mathbf{y}'\mathrm{E}_{n_i}\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathrm{J}_n\mathbf{y} = \mathbf{y}'[\mathrm{I} - \mathrm{J}_n]\mathbf{y},$$

where

$$\mathrm{J}_n = \begin{bmatrix} (1/n_1)\mathrm{J}_{n_1\times n_1} & \mathrm{Z}_{n_1\times n_2} & \cdots & \mathrm{Z}_{n_1\times n_i} & \cdots & \mathrm{Z}_{n_1\times n_K} \\ \mathrm{Z}_{n_2\times n_1} & (1/n_2)\mathrm{J}_{n_2\times n_2} & \cdots & \mathrm{Z}_{n_2\times n_i} & \cdots & \mathrm{Z}_{n_2\times n_K} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathrm{Z}_{n_{i-1}\times n_1} & \mathrm{Z}_{n_{i-1}\times n_2} & \cdots & \mathrm{Z}_{n_{i-1}\times n_i} & \cdots & \mathrm{Z}_{n_{i-1}\times n_K} \\ \mathrm{Z}_{n_i\times n_1} & \mathrm{Z}_{n_i\times n_2} & \cdots & (1/n_i)\mathrm{J}_{n_i\times n_i} & \cdots & \mathrm{Z}_{n_i\times n_K} \\ \mathrm{Z}_{n_{i+1}\times n_1} & \mathrm{Z}_{n_{i+1}\times n_2} & \cdots & \mathrm{Z}_{n_{i+1}\times n_i} & \cdots & \mathrm{Z}_{n_{i+1}\times n_K} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \mathrm{Z}_{n_K\times n_1} & \mathrm{Z}_{n_K\times n_2} & \cdots & \mathrm{Z}_{n_K\times n_i} & \cdots & (1/n_K)\mathrm{J}_{n_K\times n_K}. \end{bmatrix}$$

Next, it can be shown that $[\mathrm{I} - \mathrm{J}_n]\mathrm{X} = 0$. Hence

$$RSS = \mathbf{y}'\mathbf{y} - \sum_{i=1}^{K}\mathbf{y}'\mathrm{E}_{n_i}\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathrm{J}_n\mathbf{y} = \mathbf{y}'[\mathrm{I} - \mathrm{J}_n]\mathbf{y} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'[\mathrm{I} - \mathrm{J}_n](\mathbf{y} - \mathrm{X}\boldsymbol{\beta}),$$

Also, $[\mathrm{I} - \mathrm{J}_n]$ is symmetric and idempotent. $\mathrm{Rank}[\mathrm{I} - \mathrm{J}_n] = \mathrm{trace}[\mathrm{I} - \mathrm{J}_n] = n - K$. Therefore by Cochran Theorem

$$\frac{RSS}{\sigma^2} = \frac{(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'[\mathrm{I} - \mathrm{J}_n](\mathbf{y} - \mathrm{X}\boldsymbol{\beta})}{\sigma^2} \sim \chi^2_{n-K}.$$

**P 2.** Consider the ANOVA model (1). The regression sum of squares $SS_{\mathrm{Reg}}$ is defined as

$$SS_{\mathrm{Reg}} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (\widehat{y}_{ij} - \bar{y})^2, \tag{2}$$

where $\widehat{y}_{ij} = \widehat{\mu} + \widehat{\alpha}_i$, $\widehat{\mu}$ and $\widehat{\alpha}_i$ are least squares estimators of $\mu$ and $\alpha_i$, $\bar{y} = \sum_{i=1}^{K} n_i \bar{y}_i / \sum_{i=1}^{K} n_i$ and $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}$ for $i = 1, \ldots, K$. Find the distribution of $SS_{\mathrm{Reg}}/\sigma^2$.

**Solution:** From **P 1**,

$$\begin{aligned}
SS_{\mathrm{Reg}} &= \sum_{i=1}^{K} \sum_{j=1}^{n_i} (\widehat{y}_{ij} - \bar{y})^2 = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{K} n_i (\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{K} n_i (\bar{y}_i)^2 - n(\bar{y})^2 \\
&= \mathbf{y}'\mathrm{J}_n\mathbf{y} - (1/n)\mathbf{y}'\mathrm{W}_n\mathbf{y} = \mathbf{y}'[\mathrm{J}_n - (1/n)\mathrm{W}_n]\mathbf{y}
\end{aligned}$$

Where $\mathrm{W}_n$ is $n \times n$ matrix of all ones.

Next, it can be shown that $[\mathrm{J}_n - (1/n)\mathrm{W}_n]\mathrm{X} = 0$. Thus,

$$SS_{\mathrm{Reg}} = \mathbf{y}'[\mathrm{J}_n - (1/n)\mathrm{W}_n]\mathbf{y} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'[\mathrm{J}_n - (1/n)\mathrm{W}_n](\mathbf{y} - \mathrm{X}\boldsymbol{\beta}).$$

Also, $[\mathrm{J}_n - (1/n)\mathrm{W}_n]$ is symmetric and idempotent. $\mathrm{Rank}[\mathrm{J}_n - (1/n)\mathrm{W}_n] = \mathrm{trace}[\mathrm{J}_n - (1/n)\mathrm{W}] = K - 1$. Therefore by Cochran Theorem

$$\frac{SS_{\mathrm{Reg}}}{\sigma^2} = \frac{(\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'[\mathrm{J}_n - (1/n)\mathrm{W}](\mathbf{y} - \mathrm{X}\boldsymbol{\beta})}{\sigma^2} \sim \chi^2_{K-1}.$$

**Remark:** The hypothesis

$$H_0 : \alpha_1 = \cdots = \alpha_K \text{ versus } H_1 : \alpha_i \neq \alpha_j \text{ for at least one pair of } (i, j)$$

can tested using the statistics

$$F = \frac{SS_{\mathrm{Reg}}/df_1}{RSS/df_2} \sim F_{df_1, df_2},$$

where $df_1$ and $df_2$ are degrees of freedoms associated to the distributions of $RSS$ and $SS_{\mathrm{Reg}}$ respectively.

Note that it can be shown that $[\mathrm{J}_n - (1/n)\mathrm{W}_n][\mathrm{I} - \mathrm{J}_n] = 0$. Thus, $RSS$ and $SS_{\mathrm{Reg}}$ both are independent. Thus,

$$F = \frac{SS_{\mathrm{Reg}}/(K-1)}{RSS/(n-K)} \sim F_{K-1, n-K}.$$

**P 3.** Consider the model $y_1 = \theta_1 + \theta_2 + \epsilon_1$, $y_2 = 2\theta_1 + \epsilon_2$, $y_3 = \theta_1 - \theta_2 + \epsilon_3$, where $\epsilon_i$, $i = 1, 2, 3$, are i.i.d. $N(0, \sigma^2)$. Derive the $F$–statistics for testing $\theta_1 = \theta_2$.

**R: Assignment**

**P 4.** Three teaching methods, A, B, C, are to be compared. Each method was ad- ministered to a group of 4 students, and the scores obtained by the students on a test are given below. Carry out an F -test at level of significance 0.01 to decide whether the mean scores under the three methods are significantly different.

Table 1: Data.

| Method | | |
|---|---|---|
| A | B | C |
| 75 | 82 | 78 |
| 79 | 93 | 81 |
| 71 | 86 | 76 |
| 69 | 88 | 81 |

**Solution**
Comment: Data Import from a CSV file named *Score_Data.csv*

Comment: ANOVA Method
$Score.Data < -read.csv("/Users/satya/Desktop/MTH207A/Score\_Data.csv")$

$one.way < -aov(Score \sim Method, data = Score.Data)$

summary(one.way)

Comment: Regression method

$y < -c(75, 79, 71, 69, 82, 93, 86, 88, 78, 81, 76, 81)$
$x1 < -c(rep(1, 4), rep(0, 4), rep(0, 4))$
$x2 < -c(rep(0, 4), rep(1, 4), rep(0, 4))$
$x3 < -c(rep(0, 4), rep(0, 4), rep(1, 4))$

$model.lm.P1 < -lm(y\ x1 + x2 + x3)$

summary(model.lm.P1)

**P 5.** Consider the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ with the usual assumptions. Test the following hypotheses: **(a)** $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$, **(b)** $H_0 : \beta_3 = 0$ versus $H_1 : \beta_3 \neq 0$ using the data given below.

Table 2: Data.

| $y$ | $x_1$ | $x_2$ | $y$ | $x_1$ | $x_2$ |
|---|---|---|---|---|---|
| 10 | 21 | 2.67 | 13 | 48 | 7.20 |
| 12 | 32 | 3.12 | 21 | 81 | 9.12 |
| 6 | 46 | 2.11 | 14 | 93 | 3.21 |
| 14 | 91 | 4.21 | 11 | 88 | 4.87 |
| 20 | 20 | 6.43 | 18 | 46 | 5.38 |
| 5 | 65 | 1.76 | 17 | 24 | 8.71 |
| 8 | 26 | 2.88 | 27 | 11 | 8.11 |
| 15 | 74 | 6.15 | | | |

**Solution:**
Comment: Problem P5
$P5.Data < -read.csv("/Users/satya/Desktop/MTH207A/P5.Data.csv")$

$model.lm.P2 < -lm(y \sim x1 + x2, data = P5.Data)$
$summary(model.lm.P2)$
$linearHypothesis(model.lm.P2, c("x1 = 0"), test = "F")$
$linearHypothesis(model.lm.P2, c("x1 - x2 = 0"), test = "F")$