High-dimensional data analysis HW#1

102225014 Ai-Chun Chen

1. In one-way ANOVA:

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $i = 1, 2, 3;\ j = 1, \ldots, n$ where $\alpha_1 + \alpha_2 + \alpha_3 = 0$.

Matrix form: $\mathbf{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \\ y_{31} \\ \vdots \\ y_{3n} \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$, $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n} \\ \varepsilon_{31} \\ \vdots \\ \varepsilon_{3n} \end{bmatrix}$.

Compute rank($\mathbf{X}$), $\mathbf{X}^{\mathrm{T}}\mathbf{y}$, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and find a LSE $\hat{\boldsymbol{\beta}}$.

<u>Solution</u>:

(1)

rank($\mathbf{X}$) = 3 ($x_0 = x_1 + x_2 + x_3$)

(2)

$$\mathbf{X}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} y_{\cdot\cdot} \\ y_{1\cdot} \\ y_{2\cdot} \\ y_{3\cdot} \end{bmatrix}$$

(3)

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = n\begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

(4)

LSE => satisfy $\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{X}^{\mathrm{T}}\mathbf{X}\,\hat{\boldsymbol{\beta}}$

Consider the estimator of $\boldsymbol{\beta}$ by method of moments:

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $i = 1, 2, 3;\ j = 1, \ldots, n$

$$\Rightarrow \sum_{i=1}^{3}\sum_{j=1}^{n} y_{ij} = \sum_{i=1}^{3}\sum_{j=1}^{n}(\mu + \alpha_i + \varepsilon_{ij})$$

$$\Rightarrow y_{\cdot\cdot} = 3n\mu + 0 + \sum_{i=1}^{3}\sum_{j=1}^{n}\varepsilon_{ij} \quad (\because \sum_{i=1}^{3}\alpha_i = 0)$$

$$\Rightarrow E(y_{\cdot\cdot}) = 3n\mu$$

$$\Rightarrow E(\bar{y}_{\cdot\cdot}) = \mu$$

$$\Rightarrow \hat{\mu} = \bar{y}_{\cdot\cdot}$$

Similarly, $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, i = 1, 2, 3;\ j = 1, \ldots, n$

$$\Rightarrow \sum_{j=1}^{n} y_{ij} = \sum_{j=1}^{n}(\mu + \alpha_i + \varepsilon_{ij})$$

$$\Rightarrow y_{i\cdot} = n\mu + n\alpha_i + \sum_{j=1}^{n}\varepsilon_{ij}$$

$$\Rightarrow y_{i\cdot} = n\mu + n\alpha_i + \sum_{j=1}^{n}\varepsilon_{ij}$$

$$\Rightarrow E(\bar{y}_{i\cdot}) = \mu + \alpha_i$$

$$\Rightarrow \hat{\alpha} = \bar{y}_{i\cdot} - \hat{\mu} = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$$

Thus, by method of moments, estimators $\begin{cases} \hat{\mu} = \bar{y}_{\cdot\cdot} \\ \hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot} \end{cases}$, take $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix} = \begin{bmatrix} \bar{y}_{\cdot\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot} \\ \bar{y}_{3\cdot} - \bar{y}_{\cdot\cdot} \end{bmatrix}$,

which satisfy the equal $\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{X}^{\mathrm{T}}\mathbf{X}\,\hat{\boldsymbol{\beta}}$.

2. Show that the F statistic (3.13) for dropping a single coefficient from a model is equal to the square of the corresponding z-score (3.12).

<u>Solution</u>:

Assume that RSS_1 is from full model and RSS_j is from model dropping β_j.

Our goal is to show that $\mathrm{F} = \dfrac{(\mathrm{RSS}_j - \mathrm{RSS}_1)/(p - p_j)}{\mathrm{RSS}_1/(N-p-1)}$ and $z_j^2 = \dfrac{\hat{\beta}_j^2}{\hat{\sigma}^2 v_j}$ have same distribution.

The numerator of F is $\chi^2_{df=1}$ because $p - p_j = 1$ in our setting. The denominator of F is $\dfrac{\chi^2_{df=N-p-1}}{N-p-1}$.

The component $z_j = \dfrac{\hat{\beta}_j}{\hat{\sigma}\sqrt{v_j}}$ can be written as

$\dfrac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{v_j}} \Bigg/ \sqrt{\dfrac{\hat{\sigma}^2(N-p-1)}{\sigma^2} \Big/ (N-p-1)}$ where the numerator is standard normal and the denominator is $\sqrt{\dfrac{\chi^2_{df=N-p-1}}{N-p-1}}$. Thus, $z_j^2 \sim \dfrac{\chi^2_{df=1}}{\chi^2_{df=N-p-1}}$ where is the same as F.