STAT 203. HW 2. Yifan Zheng 3035272865.

1. Lec 3. P3.

3 Invariance of OLS Assume that $X^\top X$ is non-degenerate and Γ is a $p \times p$ orthogonal matrix. Define $\tilde{X} = X\Gamma$. Give the formulas of the coefficient of $\tilde{X}$ and the fitted values in the OLS fit of Y on $\tilde{X}$. How do they depend on Γ?

suppose OLS of Y on $\tilde{X}$:

$$\hat{\beta}' = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T Y = [(X\Gamma)^T(X\Gamma)]^{-1}(X\Gamma)^T Y.$$

$$= (\Gamma^T X^T X \Gamma)^{-1}(\Gamma^T X^T) Y.$$

Because

$$= \Gamma^{-1}(X^TX)^{-1}\Gamma\,\Gamma^T X^T Y$$

$(\Gamma^T = \Gamma^{-1})$

$$= \Gamma^{-1}(X^TX)^{-1}X^T Y. = \Gamma^{-1}\hat{\beta},$$

where $\hat{\beta}$ is coefficient of X in the OLS of Y on X.

Hence $\hat{Y}' = \tilde{X}\cdot\hat{\beta}' = X\Gamma\cdot\Gamma^{-1}\hat{\beta} = X\hat{\beta} = \hat{Y}.$

Hence the coefficient is scaled by Γ^{-1}, while the fitted values doesn't change.

2. Lec3. P5

5 OLS with multiple responses For each unit $i = 1, \dots, n$, we have multiple responses $y_i = (y_{i1}, \dots, y_{iq})^{\mathrm{T}}$ and multiple covariates $x_i = (x_{i1}, \dots, x_{ip})^{\mathrm{T}}$. Define

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1q} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{nq} \end{pmatrix} = \begin{pmatrix} y_1^{\mathrm{T}} \\ \vdots \\ y_n^{\mathrm{T}} \end{pmatrix} = (Y_1, \dots, Y_q), \quad X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} x_1^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{pmatrix} = (X_1, \dots, X_p)$$

as the $n \times q$ response matrix and $n \times p$ covariate matrix. Define the multiple OLS matrix as

$$\boxed{\hat{B} = \arg\min_{B \in \mathbb{R}^{p \times q}} \sum_{i=1}^{n} \|y_i - B^{\mathrm{T}} x_i\|^2}$$

Show that $\hat{B} = (\hat{B}_1, \dots, \hat{B}_q)$, where

$$\hat{B}_1 = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y_1, \dots, \hat{B}_q = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y_q.$$

This result tells us that if the multiple OLS for a vector outcomes reduces to multiple independent OLS fits.

Suppose $B_k = \begin{bmatrix} B_{k1} \\ \vdots \\ B_{kp} \end{bmatrix}$ Hence $B = \begin{bmatrix} B_{11} & B_{21} & \cdots & B_{q1} \\ B_{12} & B_{22} & \cdots & B_{q2} \\ \vdots & \vdots & & \vdots \\ B_{1p} & B_{2p} & & B_{qp} \end{bmatrix}$

$$y_i - B^T \cdot x_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iq} \end{bmatrix} - \begin{bmatrix} B_{11} & \cdots & B_{1p} \\ \vdots & & \vdots \\ B_{q1} & \cdots & B_{qp} \end{bmatrix} \cdot \begin{bmatrix} X_{i1} \\ \vdots \\ X_{ip} \end{bmatrix} = \begin{bmatrix} y_{i1} - \sum_{j=1}^{p} B_{1j} X_{ij} \\ \vdots \\ y_{iq} - \sum_{j=1}^{p} B_{qj} X_{ij} \end{bmatrix}$$

$$\hat{B} = \arg\min_{B} \sum_{i=1}^{n} \sum_{k=1}^{q} \left[y_{ik} - \sum_{j=1}^{p} B_{kj} X_{ij} \right]^2$$

To get $\hat{B}_k$ need to $\frac{\partial}{\partial B_{kj}} \sum_{i=1}^{n} \left[y_{ik} - \sum_{j=1}^{p} B_{kj} X_{ij} \right]^2 = 0$

where $j = 1, \cdots, p$. That is: $2 \cdot \sum_{i=1}^{n} \left(y_{ik} - \sum_{j=1}^{p} B_{kj} X_{ij} \right) \cdot X_j = 0$

Hence $\hat{B}_k = (X^T X)^{-1} X^T Y_k$. $k = 1 \cdots q$

3. Lec 4. P1. Prove Lemma 1.

Lemma 1. Both H and $I_n - H$ are projection matrices onto the column space of X and its complement, respectively. In particular, $HX = X$, $(I_n - H)X = 0$, and they are orthogonal:

$$H(I_n - H) = (I_n - H)H = 0.$$

① $H = X(X^TX)^{-1}X^T$.

Because $H^2 = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H$

And $H^T = X(X^TX)^{-1}X^T = H$

$\Rightarrow$ H is projection matrix

$HX = X(X^TX)^{-1}X^TX = X$.

$\Rightarrow$ H is projection matrix onto the column space of X

② $(I-H)^2 = I - 2H + H^2 = I - H$

$(I-H)^T = I - H^T = I - H$

$\Rightarrow$ $I-H$ is projection matrix

$(I-H)X = X - HX = X - X = 0$.

$\Rightarrow$ H is projection matrix onto the complement of coloum space of X.

③. $H(I_n - H) = H - H^2 = H - H = 0$

$(I_n - H)H = H - H^2 = H - H = 0$

Hence they are orthogonal

4. Lec 4. P3.

3 Gauss–Markov Theorem for prediction Under the Gauss–Markov model, the OLS predictor $\hat{Y} = X\hat{\beta}$ for the mean $X\beta$ is the best linear unbiased predictor in the sense that $\text{cov}(\hat{Y}) \preceq \text{cov}(\tilde{Y})$ for any predictor $\tilde{Y}$ satisfying

(C1) $\tilde{Y} = \tilde{H}Y$ for some $\tilde{H} \in \mathbb{R}^{n\times n}$ not depending on Y;

(C2) $\tilde{Y}$ is unbiased for $X\beta$.

To prove $\hat{Y}$ is BLP

$\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY = HY$. H is projection matrix

(C1) + (C2) $\Rightarrow E(\tilde{Y}) = X\beta \Rightarrow E(\tilde{H}Y) = \tilde{H}E(Y) = X\beta$

$\Rightarrow \tilde{H}X\beta = X\beta. \Rightarrow \tilde{H}X = X.$

$$\text{Cov}(\tilde{Y}) = \text{Cov}(\tilde{H}Y) = \text{Cov}(\tilde{H}Y - HY + HY)$$
$$= \text{Cov}[(\tilde{H}-H)Y + HY]$$

$$\text{Cov}((\tilde{H}-H)Y, HY) = (\tilde{H}-H)\text{Cov}(Y)H = \sigma^2(\tilde{H}\cdot H - H)$$
$$= \sigma^2(\underline{\tilde{H}\cdot X}(X^TX)^{-1}X^T - H) = \sigma^2(H-H) = 0$$

Hence $\text{Cov}(\tilde{Y}) = (\tilde{H}-H)\text{Cov}(Y)(\tilde{H}-H)^T + H\,\text{Cov}Y\,H$
$$= \sigma^2(\tilde{H}-H)(\tilde{H}-H)^T + \sigma^2 H$$
$$\text{Cov}(\hat{Y}) = \sigma^2 H.$$

Hence $\text{Cov}(\hat{Y}) \leq \text{Cov}(\tilde{Y})$

5. Lec5. P1.

1 MLE Under the Gaussian linear model, ① show that the maximum likelihood estimator (MLE) for β is the OLS estimator, but the MLE for σ^2 is $\tilde{\sigma}^2 = \text{RSS}/n$. ② Compare the mean squared errors of $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ for estimating σ^2.

① $Y = X\beta + \varepsilon \sim N(X\beta, \sigma^2 I_n)$

Hence $\log L(Y) = \text{Constant} - n\log\sigma + \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i^T\beta)^2$

$$\frac{\partial \log L(Y)}{\partial \beta} = \sum_{i=1}^n x_i^T (y_i - x_i^T\beta) = 0$$

Hence $\hat{\beta}_{MLE} = (X^TX)^{-1}X^TY$, exactly OLS estimator

$$\frac{\partial \log L(Y)}{\partial \sigma} = -\frac{n}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^n (y_i - x_i^T\beta)^2 = 0$$

Hence $\hat{\sigma}^2_{MLE} = \frac{\sum_{i=1}^n (y_i - x_i^T\beta)^2}{n} = \frac{RSS}{n} = \tilde{\sigma}^2$

② $\hat{\sigma}^2 = \frac{RSS}{n-p}$, $\tilde{\sigma}^2 = \frac{RSS}{n}$, $\hat{\sigma}^2 \sim \frac{\sigma^2}{n-p}\chi^2_{n-p}$

- Bias. $E(\hat{\sigma}^2) = \sigma^2$

  $E(\tilde{\sigma}^2) = E(\hat{\sigma}^2 \cdot \frac{n-p}{n}) = \frac{n-p}{n}\sigma^2$

- Variance. $Var(\hat{\sigma}^2) = \frac{\sigma^4}{(n-p)^2} \cdot 2(n-p) = \frac{2\sigma^4}{n-p}$

  $Var(\tilde{\sigma}^2) = Var(\hat{\sigma}^2 \cdot \frac{n-p}{n}) = \frac{n-p}{n^2} \cdot 2\sigma^4$

- MSE

$$MSE(\hat{\sigma}^2) = [E(\hat{\sigma}^2) - \sigma^2]^2 + Var(\hat{\sigma}^2)$$
$$= \frac{2\sigma^4}{n-p}$$

$$MSE(\tilde{\sigma}^2) = [E(\tilde{\sigma}^2) - \sigma^2]^2 + Var(\tilde{\sigma}^2)$$
$$= \frac{(n-p)^2}{n^2}\sigma^4 - \frac{2(n-p)}{n^2}\sigma^4$$
$$= \frac{(n-p)^2 - 2(n-p)}{n^2}\sigma^4.$$

$$MSE(\hat{\sigma}^2) - MSE(\tilde{\sigma}^2) = \frac{\sigma^4}{n^2(n-p)}\left(2n^2 - (n-p)^3 + 2(n-p)^2\right)$$

Compare $2(n^2+(n-p)^2)$ and $(n-p)^3$.

If $2(n^2+(n-p)^2) < (n-p)^3$, $MSE(\hat{\sigma}^2) < MSE(\tilde{\sigma}^2)$

If $2(n^2+(n-p)^2) \geqslant (n-p)^3$, $MSE(\hat{\sigma}^2) \geqslant MSE(\tilde{\sigma}^2)$

6. Lec 5 P4.

4 **Analysis of Variance (ANOVA) with multi-level treatment** Let x_i be the indicator vector for J treatment levels in a completely randomized experiment, for example, $x_i = e_j = (0, \dots, 1, \dots, 0)^T$ with the jth element being one if unit i receives treatment level j ($j = 1, \dots, J$). Let y_i be the outcome of unit i ($i = 1, \dots, n$). Let $\mathcal{T}_j$ be the indices of units receiving treatment j, and let $n_j = |\mathcal{T}_j|$ be the sample size and $\bar{y}_j = n_j^{-1} \sum_{i \in \mathcal{T}_j} y_i$ be the sample mean of the outcomes under treatment j. Define $\bar{y} = n^{-1} \sum_{i=1}^n y_i$ as the grand mean. We can test whether the treatment has any effect on the outcome by testing the null hypothesis

$$H_0 : \beta_1 = \cdots = \beta_J$$

in the Gaussian linear model $Y = X\beta + \varepsilon$ assuming $\varepsilon \sim \mathrm{N}(0, \sigma^2 I_n)$. This is a special case of testing $C\beta = 0$. Find C and show that the F statistic is identical to

$$F = \frac{\sum_{j=1}^J n_j (\bar{y}_j - \bar{y})^2 / (J-1)}{\sum_{j=1}^J \sum_{i \in \mathcal{T}_j} (y_i - \bar{y}_j)^2 / (n-J)} \sim F_{J-1, n-J}.$$

Remarks: (1) This is Fisher's F statistic. (2) In this linear model formulation, X does not contain a column of 1's. (3) The choice of C is not unique, but the final formula for F is. (4) You may use the Sherman–Morrison formula in the proof.

(1). $C \cdot \beta = 0 \Leftrightarrow \beta_1 = \cdots = \beta_J$

Hence one solution for C could be $C \cdot \beta = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_1 - \beta_3 \\ \vdots \\ \beta_1 - \beta_J \end{bmatrix}$

$$C = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & 0 \cdots & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & \cdots & 0 & -1 \end{bmatrix}_{(J-1) \times (J)}$$

(2). $C\hat{\beta} \sim N(C\beta, \sigma^2 C (X^T X)^{-1} C^T)$

$$\Rightarrow (C\hat{\beta} - C\beta)^T \cdot (\sigma^2 C (X^T X)^{-1} C^T)^{-1} \cdot (C\hat{\beta} - C\beta) \sim \chi^2_{J-1}$$

Hence $F_c = \dfrac{(C\hat{\beta} - C\beta)^T \cdot (\sigma^2 C (X^T X)^{-1} C^T)^{-1} (C\hat{\beta} - C\beta) / (J-1)}{\hat{\sigma}^2 / \sigma^2}$

$\sim F_{J-1, n-J}$.

Next. prove $F_c = F$

$$X = \left[\begin{array}{cccccc} 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\ \vdots\\ 1\end{array}}\right\} n_1 \\ \left.\vphantom{\begin{array}{c}1\\ \vdots\\ 1\end{array}}\right\} n_2 \\ \vdots \\ \left.\vphantom{\begin{array}{c}1\\ \vdots\\ 1\end{array}}\right\} n_J \end{array} \quad (n \times J).$$

① $$X^T X = \begin{bmatrix} n_1 & & \\ & \ddots & \\ & & n_J \end{bmatrix}_{J \times J} \Rightarrow (X^T X)^{-1} = \begin{bmatrix} 1/n_1 & & \\ & \ddots & \\ & & 1/n_J \end{bmatrix}_{J \times J}$$

② $$X^T Y = \begin{bmatrix} n_1 \bar{y}_1 \\ \vdots \\ n_J \bar{y}_J \end{bmatrix}_{J \times 1}$$

③ $$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_J \end{bmatrix}_{J \times 1} \Rightarrow C \cdot \hat{\beta} = \begin{bmatrix} \bar{y}_1 - \bar{y}_2 \\ \vdots \\ \bar{y}_1 - \bar{y}_J \end{bmatrix}_{(J-1) \times 1}$$

④ $$C (X^T X)^{-1} C^T = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}_{(J-1) \times J} \times \begin{bmatrix} \frac{1}{n_1} & & \\ & \ddots & \\ & & \frac{1}{n_J} \end{bmatrix} \times \begin{bmatrix} 1 & 1 & \cdots & 1 \\ -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & -1 \end{bmatrix}_{J \times (J-1)}$$

$$= \begin{bmatrix} \frac{1}{n_1} + \frac{1}{n_2} & \frac{1}{n_1} & \frac{1}{n_1} & \cdots & \frac{1}{n_1} \\ \frac{1}{n_1} & \frac{1}{n_1 + n_3} & \frac{1}{n_1} & \cdots & \frac{1}{n_1} \\ \vdots & & & & \\ \frac{1}{n_1} & \frac{1}{n_1} & \frac{1}{n_1} & \cdots & \frac{1}{n_1 + n_J} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n_1} & \cdots & \frac{1}{n_1} \\ \vdots & & \vdots \\ \frac{1}{n_1} & \cdots & \frac{1}{n_1} \end{bmatrix}_{(J-1) \cdot (J-1)} + \begin{bmatrix} \frac{1}{n_2} & & \\ & \ddots & \\ & & \frac{1}{n_J} \end{bmatrix}$$

Let $u = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ $V = \begin{bmatrix} \frac{1}{n_1} \\ \vdots \\ \frac{1}{n_1} \end{bmatrix}$ $A = \begin{bmatrix} \frac{1}{n_2} & & \\ & \ddots & \\ & & \frac{1}{n_J} \end{bmatrix}$

By Sherman-Morrison Formula.

$$(A+uv^T)^{-1} = A^{-1} - (1+v^TA^{-1}u)^{-1} A^{-1} u v^T A^{-1}$$

$$= \begin{bmatrix} n_2 & & \\ & \ddots & \\ & & n_J \end{bmatrix} - \frac{1}{n_1 + \sum_{j=2}^{J} n_j} \begin{bmatrix} n_2 & & \\ & \ddots & \\ & & n_J \end{bmatrix} \cdot \begin{bmatrix} 1 & \cdots & 1 \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} n_2 & & \\ & \ddots & \\ & & n_J \end{bmatrix}$$

$$= \begin{bmatrix} n_2 & & \\ & \ddots & \\ & & n_J \end{bmatrix} - \frac{1}{\sum_{j=1}^{J} n_j} \cdot \begin{bmatrix} n_2^2 & n_2 n_3 & \cdots & n_2 n_J \\ n_2 n_3 & n_3^2 & \cdots & n_3 n_J \\ \vdots & & & \vdots \\ n_2 n_J & n_3 n_J & \cdots & n_J^2 \end{bmatrix}$$

Hence $[C\hat{\beta}]^T (C(X^TX)^{-1}C^T)^{-1} C\hat{\beta}$

$$= \begin{bmatrix} \bar{y}_1 - \bar{y}_2 \\ \vdots \\ \bar{y}_1 - \bar{y}_J \end{bmatrix}^T \cdot \begin{bmatrix} n_2 & & \\ & \ddots & \\ & & n_J \end{bmatrix} \cdot \begin{bmatrix} \bar{y}_1 - \bar{y}_2 \\ \vdots \\ \bar{y}_1 - \bar{y}_J \end{bmatrix} - \frac{1}{\sum_{j=1}^{n} n_j} \begin{bmatrix} \bar{y}_1 - \bar{y}_2 \\ \vdots \\ \bar{y}_1 - \bar{y}_J \end{bmatrix}^T \begin{bmatrix} n_2^2 & \cdots & n_2 n_J \\ \vdots & & \\ n_2 n_J & \cdots & n_J^2 \end{bmatrix} \begin{bmatrix} \bar{y}_1 - \bar{y}_2 \\ \vdots \\ \bar{y}_1 - \bar{y}_J \end{bmatrix}$$

$$= \sum_{j=2}^{J} (\bar{y}_1 - \bar{y}_j)^2 n_j - \frac{1}{n} \cdot \left[\sum_{j=2}^{J} (\bar{y}_1 - \bar{y}_j) n_j\right]^2$$

$$= n\bar{y}_1^2 - 2n\bar{y}_1\bar{y} + \sum_{j=1}^{J} \bar{y}_j^2 n_j - \frac{1}{n}[n\bar{y}_1 - n\bar{y}]^2$$

$$= \sum_{j=1}^{J} \bar{y}_j^2 n_j - n\bar{y}^2 = \sum_{j=1}^{J} n_j (\bar{y}_j - \bar{y})^2$$

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-J} = \frac{1}{n-J}(Y - X\hat{\beta})^T(Y - X\hat{\beta}) = \frac{1}{n-J}\sum_{j=1}^{J}\sum_{i\in T_j}(y_i - \bar{y}_j)^2$$

Hence $F_c = \dfrac{\sum_{j=1}^{J} n_j (\bar{y}_j - \bar{y})^2 / (J-1)}{\sum_{j=1}^{J} \sum_{i\in T_j} (y_i - \bar{y})^2 / (n-J)} = F$