High-dimensional data analysis, Midterm exam #1: **[+15 points]**

Name [illegible].

+15

- Not only answer but also calculation

+4 1. [+4] Consider an ANOVA model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$,

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

not 4

+1 (1) [+1] $\mathbf{X}^{\mathrm{T}}\mathbf{y} =$

$$\underbrace{\begin{bmatrix} 1\cdots1 & 1\cdots1 & 1\cdots1 \\ 1\cdots1 & 0\cdots0 & 0\cdots0 \\ 0\cdots0 & 1\cdots1 & 0\cdots0 \\ 0\cdots0 & 0\cdots0 & 1\cdots1 \end{bmatrix}}_{4\times3n} \underbrace{\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \\ y_{31} \\ \vdots \\ y_{3n} \end{bmatrix}}_{3n\times1} = \begin{bmatrix} \sum_{j=1}^{3}\sum_{i=1}^{n} y_{ij} \\ \sum_{i=1}^{n} y_{1i} \\ \sum_{i=1}^{n} y_{2i} \\ \sum_{i=1}^{n} y_{3i} \end{bmatrix}_{4\times1}.$$

+1 (2) [+1] $\mathbf{X}^{\mathrm{T}}\mathbf{X} =$

$$\underbrace{\begin{bmatrix} 1\cdots1 & 1\cdots1 & 1\cdots1 \\ 1\cdots1 & 0\cdots0 & 0\cdots0 \\ 0\cdots0 & 1\cdots1 & 0\cdots0 \\ 0\cdots0 & 0\cdots0 & 1\cdots1 \end{bmatrix}}_{4\times3n} \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{bmatrix}}_{3n\times4} = \begin{bmatrix} 3n & n & n & n \\ n & n & 0 & 0 \\ n & 0 & n & 0 \\ n & 0 & 0 & n \end{bmatrix}_{4\times4} = n\begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}_{4\times4}$$

+2 (3) [+2] Prove or disprove the uniqueness of the LSE

(Proof must be clear)

$$\because X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{bmatrix} \Rightarrow X_0 = X_1 + X_2 + X_3.$$

$X_0 \quad X_1 \quad X_2 \quad X_3$

$\Rightarrow$ X is linear dependent

$\therefore$ LSE of $\hat{\beta}$ is not uniqle.

+5 2. [+5] Let $\mathbf{y} = \mathbf{X\beta} + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$, where the first column of $\mathbf{X}$ has ones.

Assume that the LSE $\hat{\boldsymbol\beta} = (\hat\beta_0, \hat\beta_1, \ldots, \hat\beta_p)$ is unique.

$\because \varepsilon \sim N(0, \sigma^2 I)$

$\therefore Y \sim N(X\beta, \sigma^2 I)$.

$\Rightarrow \hat\beta \sim$ Normal.

+1 1) [+1] Derive the distribution of $\hat{\boldsymbol\beta}$.

$$y = X\beta + \varepsilon$$

$$\Rightarrow RSS(\beta) = (y - X\beta)^T (y - X\beta) = y^T y + \beta^T X^T X \beta - y^T X \beta - \beta^T X^T y.$$

$$= y^T y - 2\beta^T X^T y + \beta^T X^T X \beta.$$

$$\frac{\partial RSS(\beta)}{\partial \beta} = -2X^T y + 2X^T X\beta \overset{set}{=} 0 \Rightarrow \hat\beta = (X^T X)^{-1} X^T y.$$

$$E(\hat\beta) = E((X^T X)^{-1} X^T y)$$
$$= (X^T X)^{-1} X^T E(y)$$
$$= (X^T X)^{-1} X^T X \beta = \beta.$$
$$V(\hat\beta) = V((X^T X)^{-1} X^T y)$$
$$= (X^T X)^{-1} X^T V(y) X (X^T X)^{-1}$$
$$= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1}$$
$$= \sigma^2 (X^T X)^{-1}!$$

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

+1 2) [+1] Derive the unbiased estimator of σ^2.

$\varepsilon \sim N(0, \sigma^2 I)$

$\frac{\varepsilon^2}{\sigma^2} \sim \chi^2_n$.

$\frac{(Y - X\beta)^T (Y - X\beta)}{\sigma^2} \sim \chi^2_n$

$\frac{(Y - X\hat\beta)^T (Y - X\hat\beta)}{\sigma^2} \sim \chi^2_{n-p-1}$

$E\left(\frac{(Y - X\hat\beta)^T (Y - X\hat\beta)}{\sigma^2}\right) = n-p-1$

$\Rightarrow \hat\sigma^2 = \frac{(Y - X\hat\beta)^T (Y - X\hat\beta)}{n-p-1}$ #.

+1 3) [+1] Derive the standard error (SE) of $\hat\beta_1$:

$V(\hat\beta) = \sigma^2 (X^T X)^{-1}$

let $(X^T X)^{-1}_{2\times 2} = V_1$ $\{(X^T X)^{-1}\}_{2,2}$

$\Rightarrow V(\hat\beta_1) = \sigma^2 V_1$

$\hat\beta_1 = [0, 1, 0 \ldots 0]\hat\beta$

$V(\hat\beta_1) = [0, 1, 0, \ldots, 0] V(\hat\beta) \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \sigma^2 (X^T X)^{-1}_{2\times 2}$.

$SE(\hat\beta_1) =$

if σ is known: $SE(\hat\beta_1) = \sigma\sqrt{V_1}$

if σ is unknown: $SE(\hat\beta_1) = \hat\sigma\sqrt{V_1}$

+1 4) [+1] Derive the distribution of $\frac{\hat\beta_1 - \beta_1}{SE(\hat\beta_1)}$.

if σ is known

$\hat\beta_1 \sim N(\beta_1, \sigma\sqrt{V_1})$

$\frac{\hat\beta_1 - \beta_1}{SE(\hat\beta_1)} \sim N(0,1)$

if σ is unknown

$\hat\beta_1 \sim N(\beta_1, \hat\sigma\sqrt{V_1})$

$$\frac{\hat\beta_1 - \beta_1}{\hat\sigma\sqrt{V_1}} = \frac{\hat\beta_1 - \beta_1}{\sigma\sqrt{V_1}} \Big/ \frac{\hat\sigma}{\sigma} \sim \frac{N(0,1)}{\sqrt{\frac{\chi_{n-p-1}}{n-p-1}}} \sim t_{n-p-1}$$

+1 5) [+1] Derive the normal approximation test for $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ at level $\alpha = 0.05$.

$$Z_1 = \frac{\hat\beta_1 - \beta_1}{\hat\sigma V_1} = \frac{\hat\beta_1 - 0}{\hat\sigma V_1} = \frac{\hat\beta_1}{\hat\sigma V_1} \sim t^2_{n-p-1}(0.95)$$

if $n \gg (p+1) \Rightarrow t^2_{n-p-1}(0.95) \approx Z(0.95)$

$\Rightarrow -1.96 \le Z_1 \le 1.96$

don't reject H_0.

+6 Q3 [+6] Consider a model: $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$, where $\sum_{i=1}^{n} x_i = 0$ ------ (1).

$$\frac{\partial RSS(\beta)}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) \overset{set}{=} 0$$

+1 (1) [+1] Write down the LSE $\begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \end{bmatrix}$.

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\beta_0 - \beta_1\sum_{i=1}^{n} x_i^2 = 0 \quad \leftarrow \because \sum_{i=1}^{n} x_i = 0$$

$$RSS(\beta) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2.$$

$$\hat\beta_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

$$\frac{\partial RSS(\beta)}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) \overset{set}{=} 0$$

$$\hat\beta_0 = \bar y - \hat\beta_1 \bar x \; (\because \sum x_i = 0) \Rightarrow \hat\beta_0 = \bar y \qquad \Rightarrow \begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \end{bmatrix} = \begin{bmatrix} \bar y \\ \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \end{bmatrix} \#$$

+1 (2) [+1] Give an example of $x_1, \ldots, x_n$ such that the LSE is not defined under (1).

$$\text{if } (x_1, \ldots, x_n) = (0, \ldots, 0) \quad \sum_{i=1}^{n} x_i = 0$$

$$\hat\beta_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{0}{0} \quad \therefore \hat\beta_1 \text{ is undefined.}$$

+2 (3) [+2] Show that F- statistics and t-statistics for $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ are equivalent.

(Proof must be clear and detailed) if $\varepsilon \sim N(0, \hat\sigma^2 I_{n\times n})$. $\quad \frac{\varepsilon^2}{\sigma^2} \sim \chi^2_n$.

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad X^TX = \begin{bmatrix} n & 0 \\ 0 & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \quad \Rightarrow \hat\beta \sim N(\beta, \hat\sigma^2 (X^TX)^{-1}) \quad \frac{(y - X\hat\beta)^T(y - X\hat\beta)}{\sigma^2} \sim \chi^2_{n-2}$$

$$(X^TX)^{-1} = \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum_{i=1}^{n} x_i^2} \end{bmatrix}. \quad \Rightarrow \hat\beta_1 \sim N\left(\beta_1, \hat\sigma^2 \frac{1}{\sum_{i=1}^{n} x_i^2}\right) \quad E\left(\frac{(y - X\hat\beta)^T(y - X\hat\beta)}{\sigma^2}\right) = n-2$$

$$V_1 = \frac{1}{\sum x_i^2} \qquad t_1 = \frac{\hat\beta_1 - 0}{\hat\sigma \frac{1}{\sum_{i=1}^{n} x_i^2}} = \frac{\hat\beta_1 \sum_{i=1}^{n} x_i^2}{\hat\sigma} \sim t_{n-2} \#$$

$$\Rightarrow \hat\sigma^2 = \frac{(y - X\hat\beta)^T(y - X\hat\beta)}{n-2} = \frac{RSS_1}{n-2}.$$

+2 (4) [+2] Draw the 95% confidence set for $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$

(Figure must be detailed)

$$F = \frac{RSS_0 - RSS_1 \,/\, (n-1)-(n-2)}{RSS_1 / n-2}.$$

next page.

$$RSS_0 = \sum_{i=1}^{n}(y_i - \hat\beta_0)^2 = \sum_{i=1}^{n}(y_i - \bar y)^2$$

$$RSS_1 = \sum_{i=1}^{n}(y_i - \bar y - x_i\hat\beta_1)^2.$$

$$= \sum_{i=1}^{n}(y_i - \bar y)^2 - 2\sum_{i=1}^{n}(y_i - \bar y)x_i\hat\beta_1 + \sum_{i=1}^{n} x_i^2\hat\beta_1^2$$

$$= RSS_0 - 2\sum_{i=1}^{n} x_i y_i \hat\beta_1 + \sum_{i=1}^{n} x_i^2 \hat\beta_1^2.$$

$$= RSS_0 - \hat\beta_1^2 \sum_{i=1}^{n} x_i^2$$

$$\Rightarrow F = \frac{(RSS_0 - RSS_1)/1}{RSS_1/n-2} = \frac{\hat\beta_1^2 \sum_{i=1}^{n} x_i^2}{\hat\sigma^2} = t_1^2.$$

④

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

$$\frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \sim N(0,1)$$

$$\frac{(\hat{\beta}-\beta)^T (X^T X) (\hat{\beta}-\beta)}{\hat{\sigma}^2} \sim \chi^2_2.$$

95% C.I. for β $\quad n(\hat{\beta}_0-\beta_0) \quad \Sigma X_i(\hat{\beta}_1-\beta_1)$

$$\Rightarrow \frac{[\hat{\beta}_0-\beta_0,\ \hat{\beta}_1-\beta_1] \begin{bmatrix} n & 0 \\ 0 & \Sigma X_i^2 \end{bmatrix}}{\hat{\sigma}^2} \begin{bmatrix} \hat{\beta}_0-\beta_0 \\ \hat{\beta}_1-\beta_1 \end{bmatrix} \le \chi^2_2(0.95).$$

$$\Rightarrow \quad n(\hat{\beta}_0-\beta_0)^2 + \Sigma X_i^2 (\hat{\beta}_1-\beta_1)^2 \le \chi^2_2(0.95)\hat{\sigma}^2.$$

$$\Rightarrow n(\beta_0-\bar{y})^2 + \Sigma X_i^2 (\beta_1-\hat{\beta}_1)^2 \le \chi^2_2(0.95)\hat{\sigma}^2.$$

if $\beta_0 = \hat{\beta}_0 = \bar{y}$.

$$\Sigma X_i^2 (\hat{\beta}_1-\beta_1)^2 \le \chi^2_2(0.95)\hat{\sigma}^2.$$

$$\left(\hat{\beta}_1 - \sqrt{\frac{\chi^2_2(0.95)\hat{\sigma}^2}{\Sigma X_i^2}} \le \beta_1 \le \hat{\beta}_1 + \sqrt{\frac{\chi^2_2(0.95)\hat{\sigma}^2}{\Sigma X_i^2}}\right).$$

if $\beta_1 = \hat{\beta}_1$

$$n(\hat{\beta}_0-\beta_0)^2 \le \chi^2_2(0.95)\hat{\sigma}^2.$$

$$\left(\hat{\beta}_0 - \sqrt{\frac{\chi^2_2(0.95)\hat{\sigma}^2}{n}} \le \beta_0 \le \hat{\beta}_0 + \sqrt{\frac{\chi^2_2(0.95)\hat{\sigma}^2}{n}}\right).$$