Statistical Methods - Quiz #1(70 minutes)

March 25, 2026 (Wednesday)

Section(교반): __A1__ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

- All solutions must include a detailed step-by-step explanation.
- Reference sheet is available on the last page.

1. Consider the multiple linear regression model:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \sigma^2 I_n),$$

where

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n,\ X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \in \mathbb{R}^{n\times(p+1)},\ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \in \mathbb{R}^{p+1},\ \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \in \mathbb{R}^n,$$

and $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^\top \in \mathbb{R}^{n\times(p+1)}$ is full rank ($\mathrm{rank}(X) = p+1$).

(A) Find the least squares estimate of $\boldsymbol{\beta}$, denoted $\hat{\boldsymbol{\beta}}$, that minimizes

$$S(\boldsymbol{\beta}) = (\boldsymbol{y} - X\boldsymbol{\beta})^\top(\boldsymbol{y} - X\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2.$$

<u>Solution:</u> Differentiate and set to zero:

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\sum_{i=1}^{n}\boldsymbol{x}_i\left(y_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}\right) = 0$$
$$\Rightarrow \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\top\right)\hat{\boldsymbol{\beta}} = \sum_{i=1}^{n}\boldsymbol{x}_i y_i.$$

Since $\mathrm{rank}(X) = \mathrm{rank}(X^\top X) = p+1$, $X^\top X \in \mathbb{R}^{(p+1)\times(p+1)}$ is invertible. Therefore,

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\top\right)^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i y_i = \left(X^\top X\right)^{-1}X^\top\boldsymbol{y}}.$$

(B) Find $\mathrm{E}(\hat{\boldsymbol{\beta}})$. Also, show that

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 \left(X^\top X\right)^{-1}.$$

Solution: By linearity of expectation,

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = (X^\top X)^{-1} X^\top \mathrm{E}(\boldsymbol{y}) = (X^\top X)^{-1} X^\top X \boldsymbol{\beta} = \boxed{\boldsymbol{\beta}}.$$

For the variance,

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}\big((X^\top X)^{-1} X^\top \boldsymbol{\varepsilon}\big) \\
&= (X^\top X)^{-1} X^\top \mathrm{Var}(\boldsymbol{\varepsilon}) \left((X^\top X)^{-1} X^\top\right)^\top \\
&= (X^\top X)^{-1} X^\top (\sigma^2 I_n)\, X (X^\top X)^{-1} \\
&= \sigma^2 (X^\top X)^{-1} X^\top X (X^\top X)^{-1} \\
&= \boxed{\sigma^2 (X^\top X)^{-1}}.
\end{aligned}$$

(C) Let the fitted values be $\hat{y}_i = \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$ and the hat matrix be $H = X(X^\top X)^{-1} X^\top$. Show that $(I - H)^2 = I - H$ and

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \boldsymbol{y}^\top (I - H) \boldsymbol{y}.$$

Solution: Since $H^2 = H$,

$$(I - H)^2 = (I - H)(I - H) = I - 2H + H^2 = I - H.$$

Also, since $\hat{\boldsymbol{y}} = H\boldsymbol{y}$, we have $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (I - H)\boldsymbol{y}$. Hence

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \boldsymbol{e}^\top \boldsymbol{e} = \boldsymbol{y}^\top (I - H)^\top (I - H) \boldsymbol{y} = \boldsymbol{y}^\top (I - H) \boldsymbol{y},$$

because $I - H$ is symmetric and idempotent.

(D) Show that

$$\frac{1}{\sigma^2} \boldsymbol{y}^\top (I - H) \boldsymbol{y} \sim \chi^2(n - p - 1).$$

Solution: Under the model, $\boldsymbol{y} \sim MVN(\boldsymbol{\mu}, \sigma^2 I_n)$ with $\boldsymbol{\mu} = X\boldsymbol{\beta}$.
Let $A = I - H$; then A is symmetric idempotent and
$\mathrm{rank}(A) = \mathrm{trace}(I - H) = n - \mathrm{trace}(X(X^\top X)^{-1} X^\top) = n - \mathrm{trace}((X^\top X)^{-1} X^\top X) = n - p - 1.$
Since $HX = X$, we have $AX = (I - H)X = \mathbf{0}$, so $\boldsymbol{\mu}^\top A \boldsymbol{\mu} = (X\boldsymbol{\beta})^\top A (X\boldsymbol{\beta}) = 0$.
By the hint, $\dfrac{1}{\sigma^2} \boldsymbol{y}^\top (I - H) \boldsymbol{y} = \dfrac{1}{\sigma^2} \boldsymbol{y}^\top A \boldsymbol{y} \sim \chi^2(n - p - 1)$.

$$\boxed{\frac{1}{\sigma^2} \boldsymbol{y}^\top (I - H) \boldsymbol{y} \sim \chi^2(n - p - 1).}$$

(E) Show that the estimators

$$H\boldsymbol{y} \quad \text{and} \quad (I-H)\boldsymbol{y}$$

are independent.

<u>Solution:</u> Note that $(H\boldsymbol{y}, (I-H)\boldsymbol{y})$ is jointly multivariate normal since both are linear transforms of $\boldsymbol{y}$. $\mathrm{Cov}(H\boldsymbol{y}, (I-H)\boldsymbol{y}) = H(\sigma^2 I_n)(I-H) = \sigma^2(H-H^2) = \mathbf{0}$, so $H\boldsymbol{y}$ and $(I-H)\boldsymbol{y}$ are independent.

(F) For a nonzero constant vector $\boldsymbol{a} \in \mathbb{R}^{p+1}$, find the distribution of $\boldsymbol{a}^\top \hat{\boldsymbol{\beta}}$. Also, find the distribution of the following statistic:

$$\frac{\boldsymbol{a}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{a}^\top \boldsymbol{\beta}}{\hat{\sigma}\sqrt{\boldsymbol{a}^\top (X^T X)^{-1}\boldsymbol{a}}},$$

where $\hat{\sigma}^2 = \dfrac{1}{n-p-1}\boldsymbol{y}^\top (I-H)\boldsymbol{y}$.

Note that

$$\mathrm{E}(\boldsymbol{a}^\top \hat{\boldsymbol{\beta}}) = \boldsymbol{a}^\top \mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{a}^\top \boldsymbol{\beta},$$
$$\mathrm{Var}(\boldsymbol{a}^\top \hat{\boldsymbol{\beta}}) = \boldsymbol{a}^\top \mathrm{Var}(\hat{\boldsymbol{\beta}})\boldsymbol{a} = \sigma^2 \boldsymbol{a}^\top \left(X^\top X\right)^{-1}\boldsymbol{a},$$

which gives $\boldsymbol{a}^\top \hat{\boldsymbol{\beta}} \sim N(\boldsymbol{a}^\top \boldsymbol{\beta}, \sigma^2 \boldsymbol{a}^\top \left(X^\top X\right)^{-1}\boldsymbol{a})$, or equivalently,

$$Z := \frac{\boldsymbol{a}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{a}^\top \boldsymbol{\beta}}{\sigma\sqrt{\boldsymbol{a}^\top (X^T X)^{-1}\boldsymbol{a}}} \sim N(0,1).$$

From problem (D),

$$V := \frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p-1),$$

and Z and V are independent from problem (E), since $\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{y} = (X^\top X)^{-1} X^\top (H\boldsymbol{y})$ is a function of $H\boldsymbol{y}$ (using $X^\top H = X^\top$), while $V = ((I-H)\boldsymbol{y})^\top ((I-H)\boldsymbol{y})/\sigma^2$ is a function of $(I-H)\boldsymbol{y}$. By the definition of t distribution,

$$\frac{\boldsymbol{a}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{a}^\top \boldsymbol{\beta}}{\hat{\sigma}\sqrt{\boldsymbol{a}^\top (X^T X)^{-1}\boldsymbol{a}}} = \frac{Z}{\sqrt{V/(n-p-1)}} \sim t(n-p-1).$$

Reference Sheet — Statistical Methods

PROBABILITY DISTRIBUTIONS

NORMAL DISTRIBUTION

A random variable Y follows $N(\mu, \sigma^2)$ if its pdf is

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \quad y \in \mathbb{R}.$$

If $Y \sim N(\mu, \sigma^2)$, then $Z = (Y-\mu)/\sigma \sim N(0,1)$.

CHI-SQUARE DISTRIBUTION

If $Z_1, \dots, Z_k$ are independent $N(0,1)$ random variables, then

$$V = \sum_{i=1}^{k} Z_i^2 \;\sim\; \chi^2(k).$$

$\mathrm{E}(V) = k$, $\mathrm{Var}(V) = 2k$. The sum of independent $\chi^2(k_i)$ is $\chi^2(\sum k_i)$.

t DISTRIBUTION

If $Z \sim N(0,1)$, $V \sim \chi^2(k)$, and $Z \perp V$, then

$$\frac{Z}{\sqrt{V/k}} \sim t(k).$$

F DISTRIBUTION

If $V \sim \chi^2(k_1)$, $W \sim \chi^2(k_2)$, and $V \perp W$, then

$$\frac{V/k_1}{W/k_2} \sim F(k_1, k_2).$$

MULTIVARIATE NORMAL DISTRIBUTION

A random vector $\boldsymbol{X} \in \mathbb{R}^p$ follows $\mathrm{MVN}(\boldsymbol{\mu}, \Sigma)$ if its pdf is

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left(-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p\times p}$ is positive definite.

MULTIVARIATE NORMAL: LINEAR COMBINATION

If $\boldsymbol{X} \sim \mathrm{MVN}(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{a} \in \mathbb{R}^p$ is a constant vector, then

$$\boldsymbol{a}^\top \boldsymbol{X} \;\sim\; N\big(\boldsymbol{a}^\top \boldsymbol{\mu},\; \boldsymbol{a}^\top \Sigma\, \boldsymbol{a}\big).$$

More generally, if $A \in \mathbb{R}^{k\times p}$ is a constant matrix, then $A\boldsymbol{X} \sim \mathrm{MVN}(A\boldsymbol{\mu},\, A\Sigma A^\top)$.

MULTIVARIATE NORMAL: ZERO COVARIANCE AND INDEPENDENCE

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ jointly follow multivariate normal distribution. Then

$$\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{0} \implies \boldsymbol{u} \perp \boldsymbol{v}.$$

MATRIX ALGEBRA ESSENTIALS

IDEMPOTENT MATRIX

A square matrix A is *idempotent* if $A^2 = A$.

SYMMETRIC MATRIX

A square matrix A is *symmetric* if $A = A^\top$.

RANK OF AN IDEMPOTENT MATRIX

$\mathrm{rank}(A) = \mathrm{trace}(A)$ whenever A is idempotent.

TRACE CYCLIC PROPERTY

$\mathrm{trace}(AB) = \mathrm{trace}(BA)$

EXPECTATION & VARIANCE OF LINEAR FORMS

Let A be a constant matrix and $\boldsymbol{y}$ a random vector with $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\mu}$ and $\mathrm{Var}(\boldsymbol{y}) = V$.

EXPECTATION OF A LINEAR FORM

$\mathrm{E}(A\boldsymbol{y}) = A\boldsymbol{\mu}$

VARIANCE OF A LINEAR FORM

$\mathrm{Var}(A\boldsymbol{y}) = A\,V\,A^\top$

EXPECTED VALUE OF A QUADRATIC FORM

If A is symmetric, then $\mathrm{E}(\boldsymbol{y}^\top A\boldsymbol{y}) = \mathrm{trace}(AV) + \boldsymbol{\mu}^\top A\boldsymbol{\mu}$

DISTRIBUTION THEORY FOR QUADRATIC FORMS

CHI-SQUARED FROM A QUADRATIC FORM

Let $\boldsymbol{y} \sim \mathrm{MVN}(\boldsymbol{\mu}, \sigma^2 I_n)$. If A is **symmetric idempotent** with $\mathrm{rank}(A) = \mathrm{trace}(A) = k$ and $\boldsymbol{\mu}^\top A\boldsymbol{\mu} = 0$, then

$$\frac{1}{\sigma^2}\,\boldsymbol{y}^\top A\,\boldsymbol{y} \;\sim\; \chi^2(k).$$