Quiz 1: MTH441A

Question 1

Consider the following simple linear regression model without the intercept

$$y_i = \beta z_i + \epsilon_i \quad \text{for } i = 1, \ldots, n,$$

where $\{\epsilon_1, \ldots, \epsilon_n\}$ are independently distributed with mean 0 and variance σ^2. Find

1. the least square estimate $\hat{\beta}$ of β. (2 marks)
2. the expected value of $\hat{\beta}$. (1 mark)
3. the variance of $\hat{\beta}$. (2 marks)

Question 2

Consider the following multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \quad \text{for } i = 1, \ldots, n,$$

where $\{\epsilon_1, \ldots, \epsilon_n\}$ are independently distributed with mean 0 and variance σ^2. Show that

$$\sum_{i=1}^{n} \mathrm{Var}(\hat{y}_i) = (p+1)\sigma^2.$$

(5 marks)

Question 3

The adjusted R^2 for a linear regression model is defined as

$$R^2_{\text{adj}} = 1 - \frac{\mathrm{SS}_{\text{res}}/(n-p)}{\mathrm{SS}_{\mathrm{T}}/(n-1)}$$

Find $E[R^2_{\text{adj}}]$. (5 marks)

Mid Sem: MTH441A

Question 1:

Consider the model

$$y_i = x_i \beta + \epsilon_i$$

where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independently distributed with mean 0 and variance $\mathrm{Var}(\epsilon) = V = \mathrm{Diag}(1/\omega_1, \ldots, 1/\omega_n)$.

(i)

Find a general formula for the weighted least-squares estimator of β.

(ii)

What is the variance of the weighted least-squares estimator?

(iii)

Suppose that $\mathrm{Var}(\epsilon_i) = cx_i$. Using the results of parts (i) and (ii), find the weighted least-squares estimator of β and the variance of this estimator.

(iv)

Suppose that $\mathrm{Var}(\epsilon_i) = cx_i^2$. Using the results of parts (i) and (ii), find the weighted least-squares estimator of β and the variance of this estimator.

Question 2:

Consider the following multiple linear regression model

$$y = X\beta + \epsilon$$

where $E[\epsilon] = 0$ and $\mathrm{Var}(\epsilon) = \sigma^2 V$. Verify that

$$[y'V^{-1}y - y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y]/(n-p)$$

is an unbiased estimator of σ^2.

Hint: $E[\mathbf{A}^T\mathbf{A}] = \mathrm{trace}(A'V) + \mu' A\mu$ where $E[\epsilon] = \mu$ and $\mathrm{Var}(\epsilon) = V$.

Question 3

If $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

prove or disprove that

$$\left(Y'\Sigma^{-1}Y - \frac{Y_2^2}{\sigma_{22}}\right) \sim \chi_1^2.$$

Question 4:

Consider a multiple linear regression model with usual assumptions and n observations and p regressors. Derive the distribution of Residual Sum of Squares divided by σ^2.

Quiz 2: MTH441A

Name: ____________________________ **Roll Number:** ____________________________

MCQ: Choose the correct options (more than one correct option might be possible) in questions 1 to 5. Each question carries 2 marks.

Question 1:

Consider the following simple regression model with usual assumptions:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n.$$

Let $\bar{x} = \sum_{i=1}^n x_i/n$, $\hat{\beta}_1$ is the least square estimate of β_1 and $\mathrm{Var}(\epsilon_i) = \sigma^2$. The covariance between $\bar{y}$ and $\hat{\beta}_1$ is:

(a) $\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$

(b) 0

(c) $\frac{\sigma^2}{n}$

(d) $\sum_{i=1}^n (x_i - \bar{x})^2$

Answer:

Question 2:

Consider the model (1). Define $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$. When σ^2 is known, an appropriate test statistic for testing the hypothesis $H_0 : \beta_1 = \beta_0$ versus $H_1 : \beta_1 \neq \beta_0$ is:

(a) $Z_0 = (\hat{\beta}_1 - \beta_0)/\sqrt{\sigma^2/S_{xx}}$ and $Z_0 \sim N(0,1)$

(b) $t_0 = (\hat{\beta}_1 - \beta_0)/\sqrt{MS_{\text{res}}/S_{xx}}$ and $t_0 \sim t_{n-2}$

(c) $Z_0 = (\hat{\beta}_1 - \beta_0)/\sqrt{\sigma^2/S_{xx}}$ and $Z_0 \sim t_{n-2}$

(d) $W_0 = (\hat{\beta}_1 - \beta_0)^2/(\sigma^2/S_{xx})$ and $W_0 \sim \chi_1^2$

Answer:

Question 3:

Consider the following regression model with n observations and p parameters in matrix form:

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \epsilon_{n\times 1}$$

(a) Let $H = X(X'X)^{-1}X'$ and h_{ii} is the i-th diagonal element of H. Then the i-th observation is a leverage point if $h_{ii} > 2p/n$.

(b) The i-th observation is influential/outlier if the associated `COOKS` distance is greater than 1.

(c) The i-th observation is influential/outlier if the associated `DFFITS` is greater than $2\sqrt{p/n}$.

(d) The i-th observation is influential/outlier if the associated `DFBETAS` is greater than $2/\sqrt{n}$.

Answer:

Question 4

(a) Subset selection criterion based on residual mean square is equivalent to the criterion based on adjusted R^2.

(b) For a subset model if Mallows's C_p Statistic is closed to p, then that model is not recommended.

(c) Forward selection and Backward elimination methods lead to the same subset model.

(d) Forward selection and Stepwise methods lead to the same subset model.

Question 5:

Consider the model

(a) Ridge estimator is an unbiased estimator of the regression parameters.

(b) Principal-Component Regression can be used to address the issue of multicollinearity.

(c) VIF is a measure of multicollinearity.

(d) The data collection method can not be a problem of multicollinearity.

Answer:

Question 6

Consider the following regression model with usual assumptions:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i, \quad i = 1, \ldots, n.$$

(i)

Write an appropriate test statistic for testing $H_0 : T\beta = 0$ versus $H_1 : T\beta \neq 0$, where

$$T = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

(ii)

What is the distribution of the test statistic you proposed?

Question 7

Let $Y \sim \text{Gamma}(\alpha, \beta)$ with α known. The density of Y is given by

$$f(y|\alpha,\beta) = \frac{\beta^\alpha y^{\alpha-1} e^{-\beta y}}{\Gamma(\alpha)}, \quad 0 < y < \infty.$$

(i) Is Y a member of the exponential family distribution? (1 mark)

(ii) Find the mean and variance of Y (2 marks)

Endsem

Question 1: Consider the following arrangement of complete randomized block design (CRBD) with blocks $i = a, b, c, d$ and treatments $j = 1, 2, 3, 4$.

Table 1: CRBD

Block	Treatment	.	.	.
a	1	2	3	4
b	2	1	4	3
c	3	4	2	1
d	4	3	1	2

The following ANOVA model is used:

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = a, b, c, d; \quad j = 1, 2, 3, 4,$$

where μ is the overall mean parameter, α_i and β_j are parameters associated with ith block and jth treatment effects respectively. Following constraints are imposed for the identifiability:

$$\sum_{i=a}^{d} \alpha_i = 0 \quad \text{and} \quad \sum_{j=1}^{4} \beta_j = 0.$$

Table 2: Data

Block	Treatment	.	.	.
a	3	2	5	6
b	2	4	7	8
c	9	2	1	5
d	5	6	4	3

The CRBD given in Table 1 is used for the experiment and the responses are given in Table 2. Find the least square estimates of μ, α_i and β_j for $i = a, b, c, d$ and $j = 1, 2, 3, 4$. (7 Marks)

Question 2: Consider a Poisson regression model with the following linear predictor

$$\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j, \quad i = 1, \ldots, n; \quad j = 1, \ldots, p$$

and probability mass function:

$$f(y_i; \mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}$$

(a) Find the canonical link function relating mean to the linear predictor. (2 Marks)

(b) Find the mean and variance of y_i using the form of exponential family distribution. (3 Marks)

(c) Find the asymptotic variance-covariance matrix of the least square estimator of $\beta = (\beta_1, \ldots, \beta_p)^T$. (3 Marks)

Question 3:

(i) Consider the following regression model:

$$y_i = \beta_0 + \beta_1(x_{i1} - \bar{x}_1) + \cdots + \beta_{p-1}(x_{i,p-1} - \bar{x}_{p-1}) + \epsilon_i, \quad i = 1, \ldots, n$$

where $\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}$. Find the LSE of β_0 and $\beta = (\beta_1, \ldots, \beta_{p-1})^T$. Express your solution in a compact form. (4 Marks)

(ii) Consider a particular case of above regression model when $n = p = 3$, i.e.

$$y_i = \beta_0 + \beta_1(x_{i1} - \bar{x}_1) + \beta_2(x_{i2} - \bar{x}_2) + \epsilon_i, \quad i = 1, 2, 3$$

where:

$$x_1 = (1, 0, 0)^T, \quad x_2 = (0, 1, 0)^T, \quad y = (1, 2, 3)^T$$

Find the least square estimators of $(\beta_0, \beta_1, \beta_2)$. (4 Marks)

Question 4: Define

$$X_A = \begin{pmatrix} X_{n\times p} \\ \sqrt{k} I_p \end{pmatrix}, \quad y_A = \begin{pmatrix} y \\ 0_p \end{pmatrix}$$

where I_p is a square matrix of order $p \times p$ and 0_p is a $p \times 1$ vector of zeros. Moreover, $X^T X$ is a diagonal matrix with jth diagonal element as $j - k$ for $j = 1, \ldots, p$. Now consider the following model:

$$y_A = X_A \beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n)$$

Find $\sum_{j=1}^{p} \mathrm{Var}(\hat{\beta}_j)$ where $\hat{\beta}_j$ is the least square estimator of β. (7 Marks)