

Indian Statistical Institute
B S D S, First Semester, 2025-26
Mid-semester Examination
Statistics III: Multivariate Data and Regression

08.10.25

Maximum Score 60
Total score 65

Duration: 120 minutes

Name

Student ID

1. Write your name and ID on each page.
2. Numbers in brackets denote total points allotted to each question.
3. You may use calculator.
4. Laptops and phones are not allowed.
5. This is a closed book and closed notes examination.
6. Show all your work.

Qn no	1	2	3	4	Total
Marks	15	15	20	15	65
Obtained					

Name

Student ID

1. (2+8+2+3=15) Suppose you have fitted a regression line $Y = 556.26 - 4.27X$ on 38 pairs of observations where 556.26 and -4.27 are the least squares estimators. The residual standard error is 36.37. The sample mean and standard deviation of X are 169.1 and 11.17 respectively.
 - (a) For the value 170 of X , what is the estimate of the expected value of Y under the simple linear regression model?
 - (b) What is the standard error of the estimator in part (a)? Derive the expression for the standard error and report the numerical value.
 - (c) For the value 170 of X , what is the predicted value of Y ?
 - (d) What is the standard error of this prediction?

Name

Student ID

Name

Student ID

2. (5+10=15) Consider the following model:

$$y_i = \beta x_i + \epsilon_i x_i, \quad i = 1, 2, \dots, n$$

where y_i are observed, x_i are known positive constants and β is an unknown parameter. The errors ϵ_i are independently and identically distributed random variables having the probability density function

$$f(u) = \frac{1}{2\lambda} \exp\left(-\frac{|u|}{\lambda}\right), \quad -\infty < u < \infty$$

and $\lambda > 0$ is an unknown parameter.

- (a) Find the least squares estimator of β .
- (b) Find the maximum likelihood estimators of β and λ

Name

Student ID

3. (2+5+4+5+4=20) Consider the following R code and output

```
> model<-lm(total ~ Q1+Q2+Q3+Attendance+Midterm)
> summary(model)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.5175     4.8612  -0.106   0.9157
Q1             0.4799     0.2465   1.947   0.0574 .
Q2             0.3376     0.1855   1.820   0.0750 .
Q3            -0.1756     0.2454  -0.716   0.4778
Attendance     0.7309     0.3331   2.194   0.0331 *
Midterm        1.3647     0.1361  10.026 2.33e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.39 on 48 degrees of freedom
Multiple R-squared:  0.867, Adjusted R-squared:  0.8531
F-statistic: 62.57 on 5 and 48 DF,  p-value: < 2.2e-16

> best<-ols_step_best_subset(model)
> best[["metrics"]][,c("predictors","adjr","cp","aic","sbc")]
              predictors      adjr      cp      aic      sbc
1                Midterm 0.7971602 21.816282 427.4404 433.4074
2      Attendance Midterm 0.8332057  9.918558 417.8265 425.7824
3          Q1 Attendance Midterm 0.8492731  5.312991 413.2874 423.2323
4      Q1 Q2 Attendance Midterm 0.8545927  4.511955 412.2562 424.1901
5 Q1 Q2 Q3 Attendance Midterm 0.8531299  6.000000 413.6833 427.6062
> final<-lm(total ~ Q1+Attendance+Midterm-1)
> summary(final)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
Q1             0.5867     0.2057   2.853  0.00624 **
Attendance     0.8700     0.2948   2.951  0.00477 **
Midterm        1.3260     0.1189  11.148 2.79e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.42 on 51 degrees of freedom
Multiple R-squared:  0.9763, Adjusted R-squared:  0.9749
F-statistic: 701.1 on 3 and 51 DF,  p-value: < 2.2e-16
```

Answer the following question regarding the model final.

- What is the model final?
- What are the estimates of the regression coefficients and their standard errors?
- For model choice, which criterion among `adjr`, `cp`, `aic`, `bic` has been used ? Explain your answer.
- For each of the criteria, identify the optimal model.
- What is the function of the term "-1" inside the `lm` function? Why is it justified?

Name

Student ID

Name

Student ID

Name

Student ID

4. (10+2+3=15) Consider a multiple linear regression model with p predictors and an intercept. Assume that the errors are independent with mean zero and variance σ^2 . Let $\hat{y}_i$ be the fitted value of the response Y for the i -th observation with $i = 1, \dots, n$.
- (a) Find the value of $\sum_{i=1}^n \text{Var}(\hat{y}_i)$ in terms of n , p and σ^2 . Show all your derivation.
 - (b) State any additional assumptions that you may need.
 - (c) What does this result say about the impact of including more predictors in the model?

Name

Student ID
