

Ph. D. IN STATISTICS
(PHDSTAT)
Term-End Examination
December, 2025

RST-004 : ADVANCES IN STATISTICS

Time : 3 Hours

Maximum Marks : 100

Note : (i) *Question No. 1 is compulsory.*

(ii) *Attempt any **four** questions from
Q. No. 2 to 7.*

(iii) *Scientific (non-programmable)
calculator is allowed.*

(iv) *Symbols have their usual meanings.*

1. (a) State whether the following statements are True or False. Give reasons in support of your answers : $5 \times 2 = 10$

(i) If the coefficient of determination $R^2 = 0.90$, the number of

observations $n = 20$, and the number of explanatory variables $k = 3$, then the adjusted R^2 will be approximately 0.88.

- (ii) For a logistic regression model :

$$\hat{\pi} = \frac{\exp(1.5 + 2x)}{1 + \exp(1.5 + 2x)},$$

the odds ratio corresponding to a one-unit increase in X is 5.44.

- (iii) A residual plot showing a linear pattern indicates that the assumptions of linearity and homoscedasticity are reasonably satisfied.
- (iv) In Bayesian estimation, a conjugate prior is chosen such that the posterior distribution is of a different family than the prior.
- (v) The absolute error loss function leads to the posterior mode as the Bayes estimator.

(b) Describe the following : 5+5

(i) Multicollinearity

(ii) Symmetric loss functions

2. The data on the annual yield (in kg/hectare) of a crop and annual rainfall (in cm) are given as follows : 20

Yield	Rainfall
21.4	134
23.1	121
20.4	142
24.0	120
22.4	131
23.3	125
22.1	131
23.4	119
22.8	128
21.6	134
23.3	124
21.8	139

(i) Fit a linear regression model.

(ii) Obtain the prediction interval.

- (iii) Plot the prediction interval along with the given data.

[Given that $t_{10, 0.025} = 2.228$]

3. (a) A company monitors the lifespan of rechargeable batteries used in electric vehicles. The time (in hours) until a battery fails is believed to follow an exponential distribution, with the probability density function given as : 8

$$f(x / \theta) = \theta e^{-\theta x}; x > 0, \theta > 0$$

where θ represents failure rate of batteries.

Find the Jeffrey's prior for θ .

- (b) Define loss function and its role in Bayesian estimation. 6
- (c) Mention the underlying assumptions of multiple regression model. 6
4. (a) The preparation times (in minutes) required to complete customers' orders in a restaurant are assumed to follow a normal distribution with mean θ and known standard deviation $\sigma = 1$ minute.

Based on prior experience, the chef believes that the average preparation time is around 5 minutes, but with some uncertainty. This prior belief about θ is modeled by a normal distribution with mean 3 minutes and standard deviation 1 minute : 13

- (i) Find Bayes estimate of θ under the squared error loss function.
- (ii) If the preparation times of 5 orders are observed as 2, 1, 3, 2, 3 minutes, then find the Bayes estimate of average preparation time.

- (b) Differentiate between informative and non-informative priors. 7

5. (a) The following data were collected to study the effect of an exposure on a disease :

	Disease	No disease
Exposure	152	48
No exposure	28	172

Fit a logistic regression model to examine the effect of exposure on disease. Also, obtain the log-odds and odds ratio. 8

- (b) Differentiate between the forward selection and backward elimination approaches of variables selection methods. 12
6. (a) What is Poisson regression model ? Explain with an example. 6
- (b) Differentiate between P-P and Q-Q plots. 14
7. (a) Define a regression model with one quantitative and one qualitative (with three categories) explanatory variables. Also, state the regression models corresponding to each categories. 16
- (b) Write down the like functions of the profit and complementary log-log models. 4

× × × × ×