

M. Sc. (APPLIED STATISTICS)
(MSCAST)

Term-End Examination

December, 2024

MST-017 : APPLIED REGRESSION ANALYSIS

Time : 3 Hours

Maximum Marks : 50

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

(ii) *Attempt any **four** questions from the remaining question nos. 2 to 6.*

(iii) *Use of scientific calculator (non-programmable) is allowed.*

(iv) *Symbols have their usual meanings.*

1. State whether the following statements are true or false. Give reasons in support of your answers : 5×2=10

(a) The model $Y = (\beta_0 + \beta_1 X_1 + \beta_2 X_2) \varepsilon$ is a linear model.

(b) For a regression model with 4 explanatory variables fitted on 20 observations, if the value of DFFITS for a observation is 1.5,

then that observation is considered as an influential point.

- (c) We use Durbin-Watson test for ensuring the presence of heteroscedasticity of error terms in the regression model.
- (d) In case of Poisson regression model, we use the identity link function.
- (e) If any *one* of the explanatory variables is highly correlated with other explanatory variables, the issue of autocorrelation occurs.

2. To study the relationship between age (in years) and cholesterol level (in mg/dL) of two groups of patients, the following data were obtained :

10

Group 1								
Age	46	52	39	65	54	49	76	71
Cholesterol	186	218	182	241	224	193	262	240
Group 2								
Age	20	33	78	51	43	44	63	
Cholesterol	208	248	335	264	242	268	315	

Fit the multiple linear regression model. Also obtain the regression model for each group separately.

3. The following table provides data on the average price of a product (in ₹) and total monthly sales (in ₹ lakhs) of 12 departmental stores to check the effect of price of the product on the monthly sales :

Store	Total Sales	Price of Product
1	2.4	34.0
2	2.3	34.7
3	2.0	29.8
4	2.9	38.2
5	2.7	36.1
6	3.2	42.8
7	3.4	40.8
8	2.8	37.8
9	3.2	38.4
10	3.7	41.3
11	4.0	42.0
12	3.4	40.1

If the fitted regression model for the given data is $Y = -2.4 + 0.14X$, then construct the residual and normal probability plots. Also interpret both plots.

6+4

4. (a) Explain the concept of heteroscedasticity in regression analysis with example. 5
- (b) Mention the underlying assumptions of the multiple linear regression model. 5
5. To study the effect of smoking on a disease, the following data were collected : 4+2+2+2

	Disease	No disease
Smoker	452	215
Non-Smoker	548	785
Total	1000	1000

- (i) Fit a logistic model.
- (ii) Determine the probability of having disease in both groups.
- (iii) Compute log-odds for smokers and non-smokers.
- (iv) Obtain the odds ratio for disease.
6. Write short notes on the following topics : 5+5
- (a) Polynomial Regression Model
- (b) Polytomous Regression Model

× × × × × × ×