

# Assignment MST-017

for

**M.Sc. (Applied Statistics)  
(MSCAST)**

**Valid from January 2025 to December 2025**

**SCHOOL OF SCIENCES**

Indira Gandhi National Open University  
New Delhi - 110068

**Dear Learner,**

Welcome to the M.Sc. (Applied Statistics) Programme.

As per the university guidelines, you need to complete the assignment for each theory course. Note that there are no assignments for lab courses in the MSCAST programme, namely, MSTL-011, MSTL-012, MSTL-013, MSTL-014, and MSTL-015. You should remember that writing answers to an assignment's questions will improve your writing skills and prepare you for the term-end examination.

**It is compulsory to submit the assignments within the stipulated time to be eligible to appear in the term-end examination.** You will not be allowed to appear for the term-end examination for a course if you do not submit the assignment for that course by the due date. As per the University guidelines, if you appear in the term-end examination of a course without submitting its assignment, the result of the term-end examination is liable to be cancelled/ withheld.

**The assignments constitute the continuous component of the evaluation process and have 30% weightage in the final grading.**

Before you write the assignments, you are advised to first go through the self-learning material for that course and then prepare the assignments carefully by following the instructions pertaining to the assignments. Your responses should not be a verbatim reproduction of the textual materials provided for self-learning purposes, but it should be in your own words.

If you have any doubts or problems pertaining to the course material and assignments, contact the programme in charge or the academic counsellor at your study centre. If you still have problems related to this assignment, feel free to contact the course coordinator.

Wishing you all the best in successfully completing the programme.

**(Dr. Neha Garg)**  
**Course Coordinator, MST-017**  
**Email: [nehagarg@ignou.ac.in](mailto:nehagarg@ignou.ac.in)**

## Instructions:

- Submit the assignments within the stipulated time. Otherwise, you will not be permitted to appear for the term-end examination.
- Solve the latest assignments uploaded for the current year/session.
- Read the instructions related to the assignments mentioned in the Programme Guide.
- Use only A-4 size paper to write your responses. It is mandatory to write all assignments neatly in your own handwriting. Typed or printed copies of the assignments will not be accepted. Note that you may use the printout only if a question specifically asks for the output of a program in MST-015 and MST-024.
- All questions given in the assignments are compulsory for each course.
- Express your response in your own words. You are advised to restrict your response based on the marks assigned to it. This will also help you to distribute your time in writing or completing your assignments on time.
- Securely fasten multiple pages together (you can staple or tie them) and number them carefully for each assignment separately.
- Do not forget to enclose the assignment question sheet of that course after the cover page of the assignment response (answer sheets). It is not compulsory to write each question separately before answering the question. Mention the question number for each answer.
- The solved assignment must be submitted at the Study Centre allotted to you before the due date set by the University. Please check the IGNOU website for updated information regarding the due date of assignment submission.
- You are advised to mention all information on the first page of the assignment response sheet, given on the next page.
- **Keep a copy of the assignment answer sheets with you before submission for future reference.**

**ASSIGNMENT CODE: MST-017/TMA/2025**

NAME: \_\_\_\_\_

ENROLLMENT NO: \_\_\_\_\_

ADMISSION CYCLE: \_\_\_\_\_

PROGRAMME CODE: MSCAST

COURSE CODE: MST-017

COURSE TITLE: APPLIED REGRESSION ANALYSIS

REGIONAL CENTRE CODE: \_\_\_\_\_

STUDY CENTRE CODE: \_\_\_\_\_

ADDRESS: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

CONTACT NUMBER: \_\_\_\_\_

EMAIL ID: \_\_\_\_\_

DATE OF SUBMISSION: \_\_\_\_\_



School of Sciences

**Indira Gandhi National Open University**

Maidan Garhi, New Delhi-110068 (INDIA)

# MST-017: Applied Regression Analysis

Course Code: MST-017

Assignment Code: MST-017/TMA/2025

Maximum Marks: 100

**Note: All questions are compulsory. Answer in your own words.**

**1(a)** State whether the following statements are true or false and also give the reason in support of your answer. **(5×2=10)**

- (i) We define three indicator variables for an explanatory variable with three categories.
- (ii) If the coefficient of determination is 0.833, the number of observations and explanatory variables are 12 and 3, respectively, then the Adjusted  $R^2$  will be 0.84.
- (iii) For a simple regression model fitted on 15 observations, if we have  $h_{ii} = 0.37$ , then it is an indication to trace the leverage point in the regression model.
- (iv) In a regression model  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ , if  $H_0: \beta_1 = 0$  is not rejected, then the variable  $X_1$  will remain in the model.
- (v) The logit link function is  $\log[-\log(1 - \pi)]$ .

**(b)** Write a short note on the problem of multicollinearity and autocorrelation. **(10)**

**2** Suppose a researcher wants to evaluate the effect of cholesterol on the blood pressure. The following data on serum cholesterol (in mg/dL) and systolic blood pressure (in mm/Hg) were obtained for 15 patients to explore the relationship between cholesterol and blood pressure:

S. No.	Cholesterol (mg/dL)	SBP (mm/Hg)
1	300	150
2	410	270
3	380	210
4	530	310
5	570	350
6	490	310
7	340	210
8	320	150
9	280	110
10	550	320
11	340	220
12	350	170
13	410	260
14	390	230
15	450	270

- (i) Fit a linear regression model using the method of least squares.
- (ii) Construct the normal probability plot for the regression model fitted on serum cholesterol and systolic blood pressure.

(iii) Test the significance of the fitted regression model.

(20)

3(a) Explain the assumptions underlying the multiple linear regression model.

(b) For the data given in **Question 2(b)**, obtain the followings:

(i) Diagonal of the hat matrix. Also, check the leverage points, if any.

(ii) Cook's Distances. Also, verify the influence points, if any.

(6+14)

4 A company conducted a study on its employees to see the relationship of several variables with an employee's IQ. For this purpose, fifteen employees were selected, and an IQ test and five different personality tests were given to them. Each employee's IQ was recorded along with scores on five tests. The data are shown in the following table:

Employee	Test 1	Test 2	Test 3	Test 4	Test 5	IQ
1	83	80	78	77	67	99
2	73	85	67	80	63	92
3	81	80	71	81	68	94
4	96	86	82	83	56	99
5	84	73	75	75	68	94
6	72	74	71	67	59	79
7	84	79	84	84	69	97
8	54	86	61	69	53	92
9	86	85	79	78	76	94
10	42	71	60	80	56	86
11	83	72	72	78	74	98
12	63	86	65	85	56	83
13	69	76	64	85	61	98
14	81	84	65	79	64	96
15	50	85	71	65	75	76

Determine the most appropriate regression model for the employee's IQ using a stepwise approach at a 5 % level of significance and interpret the results. Does the final regression model satisfy the linearity and normality assumptions?

(20)

5. The following data on the diagnosis of coronary heart disease (where 0 indicates absence and 1 indicates presence), serum cholesterol (in mg/dl), and weight (in kg) were obtained for 80 patients to explore the relationship of coronary heart disease with cholesterol and weight:

S. No.	Serum Cholesterol (mg/dl)	Weight (kg)	Number of Patients having CHD	Total Number of Patients
1	420	60	10	20
2	450	68	15	30
3	400	54	4	15
4	510	74	2	10
5	480	62	1	5

- (i) Fit a multiple logistic model for the dependence of coronary heart disease on the average serum cholesterol and weight considering  $\hat{\beta}_0^0 = 4.279$ ,  $\hat{\beta}_1^0 = -0.035$  and  $\hat{\beta}_2^0 = 0.172$ , as the initial values of the parameters (solve only for one iteration).
- (ii) Test the significance of the fitted model using the Hosmer-Lemeshow test at 5% level of significance.

**(12+8)**