[This question paper contains **16** printed pages.]

Your Roll No.......................

Sr. No. of Question Paper	:	**7951**
Unique Paper Code	:	2413080039
Name of the Paper	:	**Advanced Business Research**
Name of the Course	:	**B.Com. (Hons)**
Type of the Paper	:	**DSE-8.9**
Semester	:	VIII

Duration: 3 Hours Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Attempt **all** question, **all** questions are compulsory.
3. All questions carry equal marks.
4. Answers may be written either in English or Hindi; but the same medium should be used throughout the paper.
5. Use of simple calculator allowed.

परीक्षार्थियों के लिए निर्देश

१. इस प्रश्न-पत्र को प्राप्त होते ही अपना रोल नंबर ऊपर लिखिए।

२. सभी प्रश्नों का उत्तर दीजिए, सभी प्रश्न अनिवार्य हैं।

३. सभी प्रश्न समान अंक के हैं।

४. उत्तर अंग्रेज़ी अथवा हिन्दी में लिखे जा सकते हैं, किन्तु पूरे प्रश्न-पत्र में एक ही माध्यम का प्रयोग किया जाना चाहिए।

५. साधारण कैलकुलेटर के उपयोग की अनुमति है।

Q1.

A university is conducting a study to examine the factors affecting students' final examination scores. The researcher believes that study hours, attendance percentage, and internet usage for academic purposes influence student performance.

The following multiple regression model is estimated:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

Variable	Coefficient	Standard Error	t-Statistic	p-value
Intercept	28.40	6.10	4.66	0.000
Study Hours (X1)	5.20	1.10	4.73	0.000
Attendance (X2)	0.45	0.18	2.50	0.019
Internet Usage (X3)	0.90	0.50	1.80	0.082

Where:

- Y= Final Exam Score (out of 100)
- X_1= Average Study Hours per Day
- X_2= Attendance Percentage
- X_3= Internet Usage for Academic Work (hours per week)

Additional Information:

- $R^2 = 0.81$
- Adjusted $R^2 = 0.78$
- F-statistic = 27.45
- Prob(F-statistic) = 0.0000
- Number of observations = 40

Questions

a) Write the estimated regression equation from the above output.
b) Interpret the coefficient of Study Hours and Attendance Percentage..
c) est whether Attendance Percentage significantly affects exam scores at 5% level of significance.
d) Explain the meaning of R^2 and Adjusted R^2 in the context of this model.

e) What does the F-statistic indicate about the overall significance of the regression model?
f) Is Internet Usage statistically significant at 5% level? Explain with reasons.

(18)

(OR)

Q1.

Answer the following questions related to Multiple Regression Analysis and its assumptions:

a) Explain the concept of Multiple Regression Analysis. How is it different from Simple Linear Regression? Discuss its applications in economics, management, and business research.
b) Discuss the assumptions of the Classical Linear Regression Model (CLRM) in detail. Why are these assumptions important for obtaining reliable regression results?
c) Explain the problem of Autocorrelation in regression analysis. Discuss its causes, consequences, and methods of detection.
d) What is the difference between Homoskedasticity and Heteroskedasticity? Explain the impact of heteroskedasticity on Ordinary Least Squares (OLS) estimates.
e) Explain the significance of the following in regression analysis:
a) F-test
c) p-value
f) What is the Variance Inflation Factor (VIF)? How does it help in detecting multicollinearity in a regression model?

(18)

प्रश्न 1

एक विश्वविद्यालय छात्रों के अंतिम परीक्षा अंकों को प्रभावित करने वाले कारकों का अध्ययन कर रहा है। शोधकर्ता का मानना है कि अध्ययन के घंटे, उपस्थिति प्रतिशत तथा शैक्षणिक उद्देश्यों के लिए इंटरनेट उपयोग, छात्रों के प्रदर्शन को प्रभावित करते हैं।

निम्नलिखित Multiple Regression Model का अनुमान लगाया गया है:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

Variable / चर	Coefficient	Standard Error	t-Statistic	p-value
Intercept	28.40	6.10	4.66	0.000
Study Hours (X1)	5.20	1.10	4.73	0.000
Attendance (X2)	0.45	0.18	2.50	0.019
Internet Usage (X3)	0.90	0.50	1.80	0.082

जहाँ:

Y= अंतिम परीक्षा अंक (100 में से)

X_1= प्रतिदिन औसत अध्ययन घंटे

X_2= उपस्थिति प्रतिशत

X_3= शैक्षणिक कार्य हेतु इंटरनेट उपयोग (प्रति सप्ताह घंटे)

अतिरिक्त जानकारी:

१. $R^2 = 0.81$

२. Adjusted $R^2 = 0.78$

३. F-statistic = 27.45

४. Prob(F-statistic) = 0.0000

५. Number of observations = 40

प्रश्न

१. उपरोक्त Output से अनुमानित Regression Equation लिखिए।

२. Study Hours तथा Attendance Percentage के coefficients की व्याख्या कीजिए।

३. 5% significance level पर जाँच कीजिए कि Attendance Percentage परीक्षा अंकों को significantly प्रभावित करता है या नहीं।

४. इस model के संदर्भ में R^2 तथा Adjusted R^2 का अर्थ समझाइए।

५. F-statistic regression model की overall significance के बारे में क्या दर्शाता है?

६. क्या Internet Usage, 5% significance level पर statistically significant है? कारण सहित समझाइए।

(18)

(या)

प्रश्न 1

Multiple Regression Analysis तथा उसकी मान्यताओं (Assumptions) से संबंधित निम्नलिखित प्रश्नों के उत्तर दीजिए:

(क) Multiple Regression Analysis की अवधारणा को समझाइए। यह Simple Linear Regression से किस प्रकार भिन्न है? अर्थशास्त्र, प्रबंधन तथा व्यावसायिक अनुसंधान में इसके अनुप्रयोगों पर चर्चा कीजिए।

(ख) Classical Linear Regression Model (CLRM) की मान्यताओं (Assumptions) को विस्तार से समझाइए। विश्वसनीय Regression Results प्राप्त करने के लिए ये मान्यताएँ क्यों महत्वपूर्ण हैं?

(ग) Regression Analysis में Autocorrelation की समस्या को समझाइए। इसके कारण, परिणाम तथा ज्ञात करने की विधियों पर चर्चा कीजिए।

(घ) Homoskedasticity तथा Heteroskedasticity में क्या अंतर है? Ordinary Least Squares (OLS) Estimates पर Heteroskedasticity के प्रभाव को समझाइए।

(ङ) Regression Analysis में निम्नलिखित के महत्व को समझाइए:

(१) F-test

(२) p-value

(च) Variance Inflation Factor (VIF) क्या है? यह Regression Model में Multicollinearity को ज्ञात करने में किस प्रकार सहायक होता है? (18)

Q2.

A hospital wants to study the factors affecting whether a patient is diagnosed with diabetes or not. The doctors believe that age, body mass index (BMI), and daily physical activity influence the probability of diabetes.

The following binary logistic regression model is estimated:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Where:

- p= Probability that a patient has diabetes
- X_1= Age (years)
- X_2= Body Mass Index (BMI)
- X_3= Daily Physical Activity (hours)

The regression output obtained from statistical software is given below:

Variable	Coefficient (B)	Standard Error	Wald Statistic	p-value
Constant	-5.40	1.80	9.00	0.003
Age (X1)	0.055	0.018	9.34	0.002
BMI (X2)	0.120	0.040	9.00	0.003
Physical Activity (X3)	-0.75	0.28	7.17	0.007

Model Summary:

- -2 Log Likelihood = 38.60
- Cox & Snell $R^2 = 0.45$
- Nagelkerke $R^2 = 0.61$
- Classification Accuracy = 85%

a) Write the estimated binary logistic regression equation using the above output.
b) Interpret the coefficients of Age and BMI.
c) Explain the meaning of the negative coefficient of Physical Activity.
d) Test whether BMI significantly affects the probability of diabetes at 5% level of significance.
e) Interpret the Classification Accuracy of 85% for this model.
f) Explain why Binary Logistic Regression is preferred over Multiple Linear Regression when the dependent variable is categorical. B

(18)

(OR)

Q2. Attempt the following question:

I. What is Binary Logistic Regression? Explain the situations in which Binary Logistic Regression is preferred over Multiple Linear Regression. Discuss its advantages in predictive analysis.
II. Explain the following terms used in Binary Logistic Regression:
a) Logistic Function
b) Odds Ratio
III. Discuss the major assumptions of Binary Logistic Regression. Explain the consequences of violating these assumptions.
IV. Explain the significance of the following measures in evaluating a Binary Logistic Regression model:
a) -2 Log Likelihood
b) Cox & Snell R^2
V. Differentiate between Classification and Prediction
VI. Explain any one real-life application of Binary Logistic Regression in the fields of healthcare, banking, marketing, or education.

(18)

प्रश्न 2

एक अस्पताल यह अध्ययन करना चाहता है कि कौन-से कारक किसी मरीज में diabetes होने या न होने को प्रभावित करते हैं। डॉक्टरों का मानना है कि age, body mass index (BMI) और daily physical activity diabetes की संभावना को प्रभावित करते हैं।

निम्नलिखित Binary Logistic Regression Model का अनुमान लगाया गया है:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

जहाँ:

p= मरीज को diabetes होने की संभावना

X_1= Age / आयु (वर्षों में)

X_2= Body Mass Index (BMI)

X_3= Daily Physical Activity / दैनिक शारीरिक गतिविधि (घंटों में)

Statistical Software से प्राप्त Regression Output निम्नलिखित है:

Variable	Coefficient (B)	Standard Error	Wald Statistic	p-value
स्थिरांक	-5.40	1.80	9.00	0.003
आयु (X1)	0.055	0.018	9.34	0.002
BMI (X2)	0.120	0.040	9.00	0.003
शारीरिक गतिविधि (X3)	-0.75	0.28	7.17	0.007

मॉडल सारांश:

१. −2Log Likelihood = 38.60

२. Cox & Snell $R^2 = 0.45$

३. Nagelkerke $R^2 = 0.61$

४. Classification Accuracy = 85%

प्रश्न

१. उपरोक्त Output का उपयोग करते हुए अनुमानित Binary Logistic Regression Equation लिखिए।

२. Age और BMI के coefficients की व्याख्या कीजिए।

३. Physical Activity के negative coefficient का अर्थ समझाइए।

४. 5% significance level पर जाँच कीजिए कि BMI diabetes की संभावना को significantly प्रभावित करता है या नहीं।

५. इस model के लिए 85% Classification Accuracy की व्याख्या कीजिए।

६. जब dependent variable categorical हो, तब Multiple Linear Regression की तुलना में Binary Logistic Regression को क्यों प्राथमिकता दी जाती है?

(18)

(या)

प्रश्न 2

निम्नलिखित प्रश्न का उत्तर दीजिए:

(क) Binary Logistic Regression क्या है? उन परिस्थितियों को समझाइए जिनमें Multiple Linear Regression की अपेक्षा Binary Logistic Regression को प्राथमिकता दी जाती है। Predictive Analysis में इसके लाभों पर चर्चा कीजिए।

(ख) Binary Logistic Regression में प्रयुक्त निम्नलिखित शब्दों को समझाइए:

(१) Logistic Function

(२) Odds Ratio

(ग) Binary Logistic Regression की प्रमुख मान्यताओं (Assumptions) पर चर्चा कीजिए। इन मान्यताओं के उल्लंघन के परिणामों को समझाइए।

(घ) Binary Logistic Regression Model के मूल्यांकन में निम्नलिखित मापों के महत्व को समझाइए:

(१) −2Log Likelihood

(२) Cox & Snell R^2

(ङ) Classification तथा Prediction में अंतर स्पष्ट कीजिए।

(च) Healthcare, Banking, Marketing अथवा Education के क्षेत्र में Binary Logistic Regression के किसी एक वास्तविक जीवन अनुप्रयोग को समझाइए।

(18)

Q3.

A university wants to examine whether the average examination scores of students differ across three teaching methods: Traditional Classroom Teaching, Online Teaching, and Blended Learning.

The following One-Way ANOVA test is conducted using statistical software.

Group Statistics

Teaching Method	Sample Size (n)	Mean Exam Score	Standard Deviation
Traditional Classroom	12	68.4	5.6
Online Teaching	12	74.2	6.1
Blended Learning	12	81.5	5.3

ANOVA Table

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F-value	p-value
Between Groups	1089.6	2	544.8	16.72	0.0000
Within Groups	1074.9	33	32.57		
Total	2164.5	35			

Assumption Testing

Test	Result
Levene's Test p-value	0.36
Shapiro-Wilk Test p-value	0.21

Questions

a) State the null hypothesis and alternative hypothesis for the One-Way ANOVA test.
b) Interpret the mean examination scores for the three teaching methods.
c) Explain how the F-statistic is calculated in One-Way ANOVA and interpret the F-value obtained above.
d) At 5% level of significance, test whether there is a significant difference in average examination scores among the teaching methods.
e) Explain the meaning of Between Groups variation and Within Groups variation in the context of this problem.
f) Interpret the result of Levene's Test and comment on whether the assumption of equal variances is satisfied.

(18)

(OR)

Q3. Answer the following questions:

a) What is One-Way ANOVA? Explain the situations in which One-Way ANOVA is preferred over multiple t-tests. Discuss its advantages in statistical analysis.
b) Discuss the assumptions of One-Way ANOVA. Explain the consequences of violating the assumptions of normality and homogeneity of variance.
c) Explain the following concepts related to One-Way ANOVA:
 a) Sum of Squares Between (SSB)
 b) Sum of Squares Within (SSW)
 c) Degrees of Freedom
d) Explain the role and interpretation of the F-statistic in One-Way ANOVA. How is the decision regarding hypothesis testing taken using the p-value?
e) Differentiate between One-Way ANOVA and Two-Way ANOVA
f) Explain any two real-life applications of One-Way ANOVA in marketing, education, healthcare, or industrial research.

(18)

प्रश्न 3

एक विश्वविद्यालय यह जाँच करना चाहता है कि तीन शिक्षण विधियों — Traditional Classroom Teaching, Online Teaching और Blended Learning — के बीच छात्रों के औसत परीक्षा अंकों में अंतर है या नहीं।

इसके लिए Statistical Software का उपयोग करके निम्नलिखित One-Way ANOVA Test किया गया है।

समूह सांख्यिकी

शिक्षण विधि	नमूना आकार	औसत परीक्षा अंक	मानक विचलन
पारंपरिक कक्षा शिक्षण	12	68.4	5.6
ऑनलाइन शिक्षण	12	74.2	6.1
मिश्रित शिक्षण	12	81.5	5.3

ANOVA तालिका

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F-value	p-value
समूहों के बीच	1089.6	2	544.8	16.72	0.0000
समूहों के भीतर	1074.9	33	32.57		
कुल	2164.5	35			

मान्यता परीक्षण

परीक्षण	परिणाम
Levene's Test p-value	0.36
Shapiro-Wilk Test p-value	0.21

प्रश्न

(क) One-Way ANOVA Test के लिए Null Hypothesis और Alternative Hypothesis लिखिए।

(ख) तीनों teaching methods के mean examination scores की व्याख्या कीजिए।

(ग) One-Way ANOVA में F-statistic की गणना कैसे की जाती है? ऊपर प्राप्त F-value की व्याख्या कीजिए।

(घ) ५% significance level पर जाँच कीजिए कि teaching methods के बीच average examination scores में significant difference है या नहीं।

(ङ) इस समस्या के संदर्भ में Between Groups variation और Within Groups variation का अर्थ समझाइए।

(च) Levene's Test के परिणाम की व्याख्या कीजिए और बताइए कि equal variances की assumption पूरी होती है या नहीं।

(18)

(या)

प्रश्न 3

निम्नलिखित प्रश्नों के उत्तर दीजिए:

(क) One-Way ANOVA क्या है? उन परिस्थितियों को समझाइए जिनमें अनेक t-tests की तुलना में One-Way ANOVA को प्राथमिकता दी जाती है। Statistical Analysis में इसके लाभों पर चर्चा कीजिए।

(ख) One-Way ANOVA की मान्यताओं (Assumptions) पर चर्चा कीजिए। Normality तथा Homogeneity of Variance की मान्यताओं के उल्लंघन के परिणामों को समझाइए।

(ग) One-Way ANOVA से संबंधित निम्नलिखित अवधारणाओं को समझाइए:

(१) Sum of Squares Between (SSB)

(२) Sum of Squares Within (SSW)

(३) Degrees of Freedom

(घ) One-Way ANOVA में F-statistic की भूमिका तथा उसकी व्याख्या कीजिए। p-value का उपयोग करके hypothesis testing में निर्णय कैसे लिया जाता है?

(ङ) One-Way ANOVA तथा Two-Way ANOVA में अंतर स्पष्ट कीजिए।

(च) Marketing, Education, Healthcare अथवा Industrial Research में One-Way ANOVA के किसी दो वास्तविक जीवन अनुप्रयोगों को समझाइए।

(18)

Q4.

A human resource researcher conducted a longitudinal study to examine the effect of training hours, employee motivation, and work-life balance on employee performance over a period of 6 years.
The same group of employees was observed repeatedly each year.
The following output was generated from the longitudinal regression analysis.

Model Summary

Statistic	Value
Number of Employees	90
Total Observations	540
R^2	0.72
Adjusted R^2	0.70
F-value	36.48
Significance (p-value)	0.000

Coefficients Table

Variable	Coefficient (β)	Standard Error	t-value	p-value
Constant	48.20	5.40	8.93	0.000
Training Hours	3.40	0.62	5.48	0.000
Employee Motivation Score	2.15	0.58	3.71	0.001
Work-Life Balance Score	1.85	0.70	2.64	0.010
Absenteeism Days	-2.25	0.75	-3.00	0.003
Year Number	2.80	0.52	5.38	0.000

Average Employee Performance Scores Over Time

Year	Average Performance Score
Year 1	60
Year 2	66
Year 3	71
Year 4	76
Year 5	82
Year 6	87

Correlation Between Repeated Measurements

Year Pair	Correlation
Year 1 & Year 2	0.67
Year 2 & Year 3	0.72
Year 3 & Year 4	0.77
Year 4 & Year 5	0.82
Year 5 & Year 6	0.86

Questions

(a) What is meant by a longitudinal study? Explain why the given research is longitudinal in nature.

(b) Interpret the following values from the model summary:
1. R^2 value
2. F-value and significance level

(c) Interpret the coefficients of:
1. Training Hours
2. Employee Motivation Score
3. Work-Life Balance Score
4. Absenteeism Days
5. Year Number

(d) Which variables are statistically significant at the 5% level? Justify using p-values.

(e) Analyze the trend in employee performance scores over the six years. What conclusion can be drawn?

(f) Interpret the increasing correlation between repeated yearly observations in the study.

(18)

(OR)

Q4.

What is a Longitudinal Research Design? Explain the process of conducting a longitudinal study. Discuss the major challenges faced by researchers in longitudinal studies and suggest methods to overcome them. Illustrate your answer with suitable business or social research examples.

(18)

प्रश्न 4

एक मानव संसाधन शोधकर्ता ने ६ वर्षों की अवधि में प्रशिक्षण घंटों, कर्मचारी प्रेरणा तथा कार्य-जीवन संतुलन का कर्मचारी प्रदर्शन पर प्रभाव जानने के लिए एक Longitudinal Study संचालित की।
हर वर्ष कर्मचारियों के उसी समूह का बार-बार अवलोकन किया गया।
Longitudinal Regression Analysis से निम्नलिखित Output प्राप्त हुआ:

मॉडल सारांश

Statistic	Value
कर्मचारियों की संख्या	90
कुल अवलोकन	540
R^2	0.72
Adjusted R^2	0.70
F-value	36.48
Significance (p-value)	0.000

गुणांक तालिका

Variable	Coefficient (β)	Standard Error	t-value	p-value
स्थिरांक	48.20	5.40	8.93	0.000

Variable	Coefficient (β)	Standard Error	t-value	p-value
प्रशिक्षण घंटे	3.40	0.62	5.48	0.000
कर्मचारी प्रेरणा स्कोर	2.15	0.58	3.71	0.001
कार्य-जीवन संतुलन स्कोर	1.85	0.70	2.64	0.010
अनुपस्थिति के दिन	-2.25	0.75	-3.00	0.003
वर्ष संख्या	2.80	0.52	5.38	0.000

समय के साथ औसत कर्मचारी प्रदर्शन स्कोर

Year	Average Performance Score
वर्ष 1	60
वर्ष 2	66
वर्ष 3	71
वर्ष 4	76
वर्ष 5	82
वर्ष 6	87

पुनरावृत्त मापों के बीच सहसंबंध

वर्ष युग्म	Correlation / सहसंबंध
वर्ष 1 & वर्ष 2	0.67
वर्ष 2 & वर्ष 3	0.72
वर्ष 3 & वर्ष 4	0.77
वर्ष 4 & वर्ष 5	0.82
वर्ष 5 & वर्ष 6	0.86

प्रश्न

(क) Longitudinal Study से क्या अभिप्राय है? समझाइए कि दिया गया research longitudinal nature का क्यों है।

(ख) Model Summary से निम्नलिखित मानों की व्याख्या कीजिए:

१. R^2value

२. F-value तथा significance level

(ग) निम्नलिखित coefficients की व्याख्या कीजिए:

१. Training Hours

२. Employee Motivation Score

३. Work-Life Balance Score

४. Absenteeism Days

५. Year Number

(घ) ५% significance level पर कौन-से variables statistically significant हैं? p-values का उपयोग करते हुए उचित कारण दीजिए।

(ङ) ६ वर्षों में employee performance scores के trend का विश्लेषण कीजिए। इससे क्या निष्कर्ष निकाला जा सकता है?

(च) study में repeated yearly observations के बीच बढ़ते हुए correlation की व्याख्या कीजिए।

(18)

(या)

प्रश्न 4

Longitudinal Research Design की अवधारणा को समझाइए। Longitudinal Study संचालित करने की प्रक्रिया का वर्णन कीजिए। Longitudinal Studies में शोधकर्ताओं द्वारा सामना की जाने वाली प्रमुख चुनौतियों पर चर्चा कीजिए तथा उन्हें दूर करने के उपाय सुझाइए। अपने उत्तर को उपयुक्त व्यवसायिक अथवा सामाजिक अनुसंधान के उदाहरणों सहित स्पष्ट कीजिए।

(18)

Q5.

Do any three questions carrying 6 marks each

a. Explain the concept, assumptions, advantages, and applications of Multiple Regression Analysis in research.
b. Discuss the meaning, types, interpretation, and applications of Logistic Regression.
c. Explain the concept of ANOVA, its assumptions, types, and importance in hypothesis testing.
d. Explain the meaning, characteristics, types, advantages, and limitations of Longitudinal Studies.
e. Discuss the concept, elements, types, and importance of Experimental Design in research methodology.

(18)

प्रश्न 5

निम्नलिखित में से किन्हीं तीन प्रश्नों के उत्तर दीजिए। प्रत्येक प्रश्न ६ अंक का है:

(क) Research में Multiple Regression Analysis की अवधारणा, मान्यताओं (Assumptions), लाभों तथा अनुप्रयोगों को समझाइए।

(ख) Logistic Regression के अर्थ, प्रकार, व्याख्या तथा अनुप्रयोगों पर चर्चा कीजिए।

(ग) ANOVA की अवधारणा, उसकी मान्यताओं, प्रकारों तथा Hypothesis Testing में उसके महत्व को समझाइए।

(घ) Longitudinal Studies के अर्थ, विशेषताओं, प्रकारों, लाभों तथा सीमाओं को समझाइए।

(ङ) Research Methodology में Experimental Design की अवधारणा, तत्वों, प्रकारों तथा महत्व पर चर्चा कीजिए।

(18)