[This question paper contains ____ printed pages.]

Your Roll No……………………..

Sr. No. of Question Paper	:	To be filled by Examination Branch
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।

SET-1

Q1.

A retail company wants to study the factors affecting monthly sales of its stores. The management believes that advertising expenditure, store size, and number of employees influence sales 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	12.50	4.20	2.98	0.005
Advertising (X1)	3.40	0.80	4.25	0.000
Store Size (X2)	0.015	0.006	2.50	0.018
Employees (X3)	-0.20	0.10	-2.00	0.054

Where:

- Y= Monthly Sales (in ₹ lakhs)
- X_1= Advertising Expenditure (₹ lakhs)
- X_2= Store Size (in sq. ft.)
- X_3= Number of Employees

Additional Information:

- $R^2 = 0.78$
- Adjusted $R^2 = 0.74$
- F-statistic = 19.60
- Prob(F-statistic) = 0.0001
- Number of observations = 30

Questions

a) Write the estimated regression equation.
b) Interpret the coefficient of Advertising Expenditure and Store Size.
c) Test whether Advertising Expenditure significantly affects sales at 5% level of significance.
d) Interpret the meaning of $R^2$and Adjusted R^2.
e) Explain the significance of the F-statistic in this regression model.

f) Is the variable "Number of Employees" statistically significant at 5% level? Justify your answer.

(18)

(OR)

Q1.

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

I. Define Multiple Regression Analysis. Explain its objectives and importance in business and economic research.

II. Explain the following assumptions of the Classical Linear Regression Model (CLRM):
a) Linearity
b) No Multicollinearity
c) Homoskedasticity
d) No Autocorrelation

III. What is Multicollinearity? Explain its causes, consequences, and methods to detect it in a regression model.

IV. Differentiate between $R^2$and Adjusted R^2.

V. Explain the importance of the F-test and t-test in multiple regression analysis.

(18)

प्रश्न 1

एक रिटेल कंपनी अपने स्टोर की मासिक बिक्री को प्रभावित करने वाले कारकों का अध्ययन करना चाहती है। मैनेजमेंट का मानना है कि विज्ञापन पर खर्च, स्टोर का आकार और कर्मचारियों की संख्या बिक्री के प्रदर्शन पर असर डालते हैं।

निम्नलिखित मल्टीपल रिग्रेशन मॉडल का अनुमान लगाया गया है:

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

वेरिएबल	**कोएफ़िशिएंट**	**स्टैंडर्ड एरर**	**t-सांख्यिकी**	**p-मान**
इंटरसेप्ट	12.50	4.20	2.98	0.005
विज्ञापन (X1)	3.40	0.80	4.25	0.000
स्टोर का आकार (X2)	0.015	0.006	2.50	0.018
कर्मचारी (X3)	-0.20	0.10	-2.00	0.054

जहाँ:

Y= मासिक बिक्री (₹ लाख में)

X_1= विज्ञापन पर खर्च (₹ लाख)

X_2= स्टोर का आकार (वर्ग फ़ीट में)

X_3= कर्मचारियों की संख्या

अतिरिक्त जानकारी:

R^2=0.78

एडजस्टेड R^2=0.74

F-सांख्यिकी = 19.60

Prob(F-सांख्यिकी) = 0.0001

अवलोकनों की संख्या = 30

प्रश्न

a) अनुमानित रिग्रेशन समीकरण लिखें।
b) विज्ञापन पर खर्च और स्टोर के आकार के कोएफ़िशिएंट की व्याख्या करें।
c) जाँच करें कि क्या विज्ञापन पर खर्च 5% सार्थकता स्तर पर बिक्री को महत्वपूर्ण रूप से प्रभावित करता है।
d) R^2 और एडजस्टेड R^2 के अर्थ की व्याख्या करें।
e) इस रिग्रेशन मॉडल में F-सांख्यिकी के महत्व को समझाएँ।
f) क्या वेरिएबल "कर्मचारियों की संख्या" 5% स्तर पर सांख्यिकीय रूप से महत्वपूर्ण है? अपने उत्तर का औचित्य सिद्ध करें।

(18)

(या)

प्रश्न 1

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

१. Multiple Regression Analysis को परिभाषित कीजिए। इसके उद्देश्यों तथा व्यवसाय और आर्थिक अनुसंधान में इसके महत्व को समझाइए।

२. Classical Linear Regression Model (CLRM) की निम्नलिखित मान्यताओं को समझाइए:

(क) रेखीयता (Linearity)

(ख) बहुसहसंबंध का अभाव (No Multicollinearity)

(ग) समान विचरणता (Homoskedasticity)

(घ) स्व-सहसंबंध का अभाव (No Autocorrelation)

३. Multicollinearity क्या है? इसके कारण, परिणाम तथा Regression Model में इसे ज्ञात करने की विधियों को समझाइए।

४. R^2 तथा Adjusted R^2 में अंतर स्पष्ट कीजिए।

५. Multiple Regression Analysis में F-test तथा t-test के महत्व को समझाइए।

(18)

Q2.

A bank wants to study the factors affecting whether a customer will default on a loan or not. The management believes that monthly income, credit score, and existing debt influence the probability of loan default.

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 the customer defaults on the loan
- X_1= Monthly Income (₹ thousand)
- X_2= Credit Score
- X_3= Existing Debt (₹ thousand)

The regression output obtained from statistical software is given below:

Variable	Coefficient (B)	Standard Error	Wald Statistic	p-value
Constant	3.20	1.10	8.46	0.004
Monthly Income (X1)	-0.045	0.015	9.00	0.003
Credit Score (X2)	-0.008	0.003	7.11	0.008
Existing Debt (X3)	0.060	0.020	9.00	0.003

Model Summary:

- -2 Log Likelihood = 42.80
- Cox & Snell $R^2 = 0.41$
- Nagelkerke $R^2 = 0.56$

- Classification Accuracy = 82%

Questions

a) Write the estimated binary logistic regression equation using the above output.

b) Interpret the coefficients of Monthly Income and Existing Debt.

c) Explain the meaning of the negative coefficient of Credit Score.

d) Test whether Existing Debt significantly affects the probability of loan default at 5% level of significance.

e) Explain the meaning of Cox & Snell R^2 and Nagelkerke R^2 in logistic regression.

f) What does the Classification Accuracy of 82% indicate about the model?

g) Differentiate between Multiple Linear Regression and Binary Logistic Regression.

(18)

(OR)

Q2. Attempt the following question:

I. Define Binary Logistic Regression. Explain its objectives and importance in statistical and business research. How is it different from Multiple Linear Regression?

II. Explain the following concepts related to Binary Logistic Regression:
a) Logit Function
b) Odds and Odds Ratio

III. Discuss the assumptions of Binary Logistic Regression. Why are these assumptions important for obtaining reliable results?

IV. Explain the role and interpretation of the following in Binary Logistic Regression:
a) Wald Statistic
b) p-value
c) Confusion Matrix

V. Differentiate between Dependent Variable and Independent Variable in Logistic Regression.

VI. Explain any two practical applications of Binary Logistic Regression in real-life business or research situations.

(18)

प्रश्न २

एक बैंक यह अध्ययन करना चाहता है कि कौन-से कारक किसी ग्राहक द्वारा ऋण (Loan) चूक (Default) करने की संभावना को प्रभावित करते हैं। प्रबंधन का मानना है कि मासिक आय, क्रेडिट स्कोर तथा वर्तमान ऋण ग्राहक के loan default 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= ग्राहक द्वारा ऋण चूक (Loan Default) करने की संभावना

X_1= मासिक आय (₹ हजार में)

X_2= क्रेडिट स्कोर

X_3= वर्तमान ऋण (₹ हजार में)

Statistical Software से प्राप्त Regression Output निम्नलिखित है:

Variable	Coefficient (B)	Standard Error	Wald Statistic	p-value
Constant	3.20	1.10	8.46	0.004
Monthly Income (X_1)	-0.045	0.015	9.00	0.003
Credit Score (X_2)	-0.008	0.003	7.11	0.008
Existing Debt (X_3)	0.060	0.020	9.00	0.003

Model Summary:

१. -2Log Likelihood = 42.80

२. Cox & Snell $R^2 = 0.41$

३. Nagelkerke $R^2 = 0.56$

४. Classification Accuracy = 82%

प्रश्न

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

२. Monthly Income तथा Existing Debt के coefficients की व्याख्या कीजिए।

३. Credit Score के negative coefficient का अर्थ समझाइए।

४. 5% significance level पर जाँच कीजिए कि Existing Debt, loan default की संभावना को significantly प्रभावित करता है या नहीं।

५. Logistic Regression में Cox & Snell R^2तथा Nagelkerke R^2का अर्थ समझाइए।

६. Model की 82% Classification Accuracy क्या दर्शाती है?

७. Multiple Linear Regression तथा Binary Logistic Regression में अंतर स्पष्ट कीजिए।

(18)

(या)

प्रश्न २

निम्नलिखित प्रश्न का उत्तर दीजिए:

I. Binary Logistic Regression को परिभाषित कीजिए। इसके उद्देश्यों तथा सांख्यिकीय एवं व्यावसायिक अनुसंधान में इसके महत्व को समझाइए। यह Multiple Linear Regression से किस प्रकार भिन्न है?

II. Binary Logistic Regression से संबंधित निम्नलिखित अवधारणाओं को समझाइए:

(क) Logit Function

(ख) Odds तथा Odds Ratio

III. Binary Logistic Regression की मान्यताओं (Assumptions) पर चर्चा कीजिए। विश्वसनीय परिणाम प्राप्त करने के लिए ये मान्यताएँ क्यों महत्वपूर्ण हैं?

IV. Binary Logistic Regression में निम्नलिखित की भूमिका तथा व्याख्या समझाइए:

(क) Wald Statistic

(ख) p-value

(ग) Confusion Matrix

V. Logistic Regression में Dependent Variable तथा Independent Variable के बीच अंतर स्पष्ट कीजिए।

VI. वास्तविक व्यावसायिक अथवा अनुसंधान परिस्थितियों में Binary Logistic Regression के किसी दो व्यावहारिक अनुप्रयोगों को समझाइए।

(18)

Q3.

A researcher wants to examine whether the average monthly sales differ among three different retail store locations: Urban, Semi-Urban, and Rural.

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

Group Statistics

Store Location	Sample Size (n)	Mean Monthly Sales (₹ lakhs)	Standard Deviation
Urban	10	82.5	6.2

Store Location	Sample Size (n)	Mean Monthly Sales (₹ lakhs)	Standard Deviation
Semi-Urban	10	75.4	5.8
Rural	10	68.1	7.0

ANOVA Table

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F-value	p-value
Between Groups	1052.4	2	526.2	11.48	0.0003
Within Groups	1237.5	27	45.83		
Total	2289.9	29			

Assumption Testing

Test	Result
Levene's Test p-value	0.42
Shapiro-Wilk Test p-value	0.18

Questions

1. State the null hypothesis and alternative hypothesis for the One-Way ANOVA test.
2. Interpret the mean monthly sales for the three store locations.
3. Calculate and interpret the F-statistic using the ANOVA table.
4. At 5% level of significance, test whether there is a significant difference in average monthly sales among the store locations.
5. Interpret the results of Levene's Test and comment on the assumption of homogeneity of variance.
6. Interpret the Shapiro-Wilk Test result and comment on the normality assumption.

(18)

(OR)

Q3. Answer the following questions:

I. Define One-Way Analysis of Variance (ANOVA). Explain its objectives and importance in statistical analysis and research studies.

II. Explain the assumptions of One-Way ANOVA in detail. Why is it important to verify these assumptions before conducting ANOVA?

III. Explain the following terms used in One-Way ANOVA:
a) Between Group Variation
b) Within Group Variation
d) F-statistic

IV. State and explain the null hypothesis and alternative hypothesis used in One-Way ANOVA.

V. Differentiate between One-Way ANOVA and t-test

VI. Explain any two practical applications of One-Way ANOVA in business, healthcare, education, or social science research.

(18)

प्रश्न ३

एक शोधकर्ता यह अध्ययन करना चाहता है कि क्या तीन विभिन्न रिटेल स्टोर स्थानों — Urban, Semi-Urban तथा Rural — के बीच औसत मासिक बिक्री में अंतर है।

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

समूह सांख्यिकी

स्टोर लोकेशन	नमूना आकार (n)	मासिक बिक्री (₹ लाख)	मानक विचलन
शहरी	10	82.5	6.2
अर्ध-शहरी	10	75.4	5.8
ग्रामीण	10	68.1	7.0

ANOVA तालिका

Variation का स्रोत	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F-value	p-value
समूहों के बीच	1052.4	2	526.2	11.48	0.0003
समूहों के भीतर	1237.5	27	45.83		
कुल	2289.9	29			

मान्यता परीक्षण

परीक्षण	परिणाम
Levene’s Test p-value	0.42
Shapiro-Wilk Test p-value	0.18

प्रश्न

१. One-Way ANOVA Test के लिए Null Hypothesis और Alternative Hypothesis लिखिए।

२. तीनों स्टोर लोकेशन के Mean Monthly Sales की व्याख्या कीजिए।

३. ANOVA Table का उपयोग करके F-statistic की गणना कीजिए और उसकी व्याख्या कीजिए।

४. ५% significance level पर जाँच कीजिए कि स्टोर लोकेशन के बीच औसत मासिक बिक्री में significant difference है या नहीं।

५. Levene’s Test के परिणाम की व्याख्या कीजिए और homogeneity of variance की assumption पर टिप्पणी कीजिए।

६. Shapiro-Wilk Test के परिणाम की व्याख्या कीजिए और normality assumption पर टिप्पणी कीजिए।

(18)

(या)

प्रश्न ३

निम्नलिखित प्रश्नों के उत्तर दीजिए:

I. One-Way Analysis of Variance (ANOVA) को परिभाषित कीजिए। सांख्यिकीय विश्लेषण तथा अनुसंधान अध्ययनों में इसके उद्देश्यों और महत्व को समझाइए।

II. One-Way ANOVA की मान्यताओं (Assumptions) को विस्तार से समझाइए। ANOVA करने से पहले इन मान्यताओं की जाँच करना क्यों आवश्यक है?

III. One-Way ANOVA में प्रयुक्त निम्नलिखित शब्दों को समझाइए:

(क) Between Group Variation / समूहों के बीच विचरण

(ख) Within Group Variation / समूहों के भीतर विचरण

(ग) F-statistic

IV. One-Way ANOVA में प्रयुक्त Null Hypothesis तथा Alternative Hypothesis को लिखिए और समझाइए।

V. One-Way ANOVA तथा t-test में अंतर स्पष्ट कीजिए।

VI. व्यवसाय, स्वास्थ्य सेवा (Healthcare), शिक्षा अथवा सामाजिक विज्ञान अनुसंधान में One-Way ANOVA के किसी दो व्यावहारिक अनुप्रयोगों को समझाइए।

(18)

Q4.

A banking researcher conducted a longitudinal study to analyze the impact of digital banking usage, customer trust, and service quality on customer retention over a period of 5 years.

The same customers were observed repeatedly every year to study changes in retention behavior over time.

The following statistical output was obtained.

Model Summary

Statistic	Value
Number of Customers	110
Total Observations	550
R^2	0.78
Adjusted R^2	0.75
F-value	44.90
Significance (p-value)	0.000

Coefficients Table

Variable	Coefficient (β)	Standard Error	t-value	p-value
Constant	38.60	5.75	6.71	0.000
Digital Banking Usage Score	3.85	0.66	5.83	0.000
Customer Trust Score	2.40	0.58	4.14	0.001
Service Quality Rating	2.95	0.72	4.09	0.001
Complaint Frequency	-1.90	0.64	-2.97	0.004

Variable	Coefficient (β)	Standard Error	t-value	p-value
Year Number	2.25	0.48	4.69	0.000

Average Customer Retention Scores Over Time

Year	Average Retention Score
Year 1	58
Year 2	64
Year 3	69
Year 4	75
Year 5	81

Correlation Between Repeated Measurements

Year Pair	Correlation
Year 1 & Year 2	0.66
Year 2 & Year 3	0.71
Year 3 & Year 4	0.78
Year 4 & Year 5	0.84

Questions

(a) What is a longitudinal study? Why is the given study considered longitudinal in nature?

(b) Interpret the following:

1. R^2 value
2. F-value and significance level

(c) Interpret the coefficients of:

1. Digital Banking Usage Score
2. Customer Trust Score
3. Service Quality Rating

(d) Identify the statistically significant variables at the 5% significance level using p-values.

(e) Analyze the trend in customer retention scores over the five years. What conclusion can be drawn from the trend?

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

(18)

(OR)

Q4.

Explain the concept of Longitudinal Studies in research methodology. Discuss its types, features, advantages, limitations, and applications in business and social science research. Also differentiate between longitudinal and cross-sectional studies with suitable examples.

(18)

प्रश्न 4

एक बैंकिंग शोधकर्ता ने 5 वर्षों की अवधि में डिजिटल बैंकिंग उपयोग, ग्राहक विश्वास तथा सेवा गुणवत्ता का ग्राहक प्रतिधारण (Customer Retention) पर प्रभाव जानने के लिए एक Longitudinal Study संचालित की।

समय के साथ ग्राहक प्रतिधारण व्यवहार में होने वाले परिवर्तनों का अध्ययन करने हेतु उन्हीं ग्राहकों का प्रत्येक वर्ष बार-बार अवलोकन किया गया।

मॉडल सारांश

Statistic	Value
ग्राहकों की संख्या	110
कुल अवलोकन	550
R^2	0.78
Adjusted R^2	0.75
F-value	44.90
Significance (p-value)	0.000

गुणांक तालिका

Variable	Coefficient (β)	Standard Error	t-value	p-value
Constant	38.60	5.75	6.71	0.000

Variable	Coefficient (β)	Standard Error	t-value	p-value
Digital Banking Usage Score	3.85	0.66	5.83	0.000
Customer Trust Score	2.40	0.58	4.14	0.001
Service Quality Rating	2.95	0.72	4.09	0.001
Complaint Frequency	-1.90	0.64	-2.97	0.004
Year Number	2.25	0.48	4.69	0.000

समय के साथ औसत ग्राहक प्रतिधारण स्कोर

वर्ष	औसत प्रतिधारण स्कोर
वर्ष 1	58
वर्ष 2	64
वर्ष 3	69
वर्ष 4	75
वर्ष 5	81

पुनरावृत्त मापों के बीच सहसंबंध

वर्ष युग्म	सहसंबंध
वर्ष 1 & वर्ष 2	0.66
वर्ष 2 & वर्ष 3	0.71
वर्ष 3 & वर्ष 4	0.78
वर्ष 4 & वर्ष 5	0.84

प्रश्न

(क) Longitudinal Study क्या है? दी गई study को longitudinal nature की क्यों माना जाता है?

(ख) निम्नलिखित की व्याख्या कीजिए:

१. R^2value

२. F-value तथा significance level

(ग) निम्नलिखित coefficients की व्याख्या कीजिए:

१. Digital Banking Usage Score

२. Customer Trust Score

३. Service Quality Rating

(घ) p-values का उपयोग करते हुए 5% significance level पर statistically significant variables की पहचान कीजिए।

(ङ) पाँच वर्षों में customer retention scores के trend का विश्लेषण कीजिए। इस trend से क्या निष्कर्ष निकाला जा सकता है?

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

(18)

(या)

प्रश्न 4

Research Methodology में Longitudinal Studies की अवधारणा को समझाइए। इसके प्रकार, विशेषताएँ, लाभ, सीमाएँ तथा व्यवसाय एवं सामाजिक विज्ञान अनुसंधान में इसके अनुप्रयोगों पर चर्चा कीजिए। साथ ही, उपयुक्त उदाहरणों सहित Longitudinal Studies तथा Cross-Sectional Studies में अंतर स्पष्ट कीजिए।

(18)

Q5.

Do any three questions of the following carrying 6 marks each:

a) Write a short note on multicollinearity, its causes, effects, detection methods, and remedies in Multiple Regression Analysis.
b) Write a short note on the concept of Odds Ratio in Logistic Regression with suitable examples and interpretation.
c) Write a short note on One-Way ANOVA and Two-Way ANOVA, highlighting their differences and applications.
d) Write a short note on the importance and applications of Longitudinal Studies in business and social science research.
e) Write a short note explaining the major advantages and limitations 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) Multiple Regression Analysis में Multicollinearity, उसके कारण, प्रभाव, ज्ञात करने की विधियाँ तथा उसके उपचारों पर संक्षिप्त टिप्पणी लिखिए।

(ख) Logistic Regression में Odds Ratio की अवधारणा को उपयुक्त उदाहरण एवं व्याख्या सहित संक्षेप में समझाइए।

(ग) One-Way ANOVA तथा Two-Way ANOVA पर संक्षिप्त टिप्पणी लिखिए तथा उनके अंतर एवं अनुप्रयोगों को स्पष्ट कीजिए।

(घ) व्यवसाय एवं सामाजिक विज्ञान अनुसंधान में Longitudinal Studies के महत्व एवं अनुप्रयोगों पर संक्षिप्त टिप्पणी लिखिए।

(ङ) Research Methodology में Experimental Design के प्रमुख लाभ एवं सीमाओं को समझाते हुए संक्षिप्त टिप्पणी लिखिए।

(18)