Data Analysis & Interpretation

End-term Examination

BBA 2020

Set- 1

Duration: 12 Hours

Maximum Marks: 50

There are a total of 11 questions. Please answer 10 questions. The maximum marks for the paper is 50.

Q1 A car manufacturer is concerned about poor customer satisfaction at one of its dealerships. The management decides to evaluate the satisfaction surveys of its next 40 customers. The dealer will be fined if the number of customers who report favourably is between 22 and 26. The dealership will be dissolved if fewer than 22 customers report favourably. It is known that 70% of the dealer’s customers report favourably on satisfaction surveys.

a. What is the probability that the dealer will be fined?

b. What is the probability that the dealership will be dissolved?

2.5+2.5

Q2 Using data from 50 workers, a researcher estimates

Wage = , where Wage is the hourly wage rate and Education, Experience, and Age are the years of higher education, the years of experience, and the age of the worker, respectively. A portion of the regression results is shown in the following table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Standard Error | t Stat | p-value |
| Intercept | 7.87 | 4.09 | 1.93 | 0.0603 |
| Education | 1.44 | 0.34 | 4.24 | 0.0001 |
| Experience | 0.45 | 0.14 | 3.16 | 0.0028 |
| Age | −0.01 | 0.08 | - 0.14 | 0.8920 |

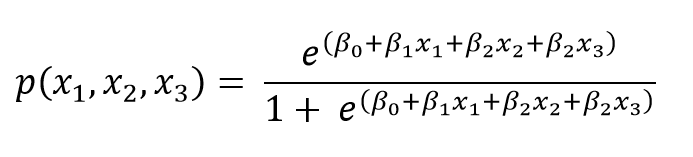
a. Interpret the estimates for β1 and β2.

b. What is the sample regression equation?

c. Predict the hourly wage rate for a 30-year-old worker with 4 years of higher education and 3 years of experience.

1.5+1.5+2

Q3 A research analyst present data that can be used to investigate allegations of gender discrimination in the hiring practices of a particular firm. These data are given in the table below. In this table, y is a dummy variable that equals 1 if a potential employee was hired and 0 otherwise; is the number of years of education of the potential employee; is the number of years of experience of the potential employee; and is a dummy variable that equals 1 if the potential employee was a male and 0 if the potential employee was a female. If we use the logistic regression model



to analyse these data, we find that the point estimates of the model parameters and their associated p-values (given in parentheses) are = -14.2483 (.0191), =1.1549 (.0552), = 0.9098 (.0341), and = 5.6037 (.0313).

a) Consider a potential employee having 4 years of education and 5 years of experience. Find (1) a point estimate of the probability that the potential employee will be hired if the potential employee is a male, and (2) a point estimate of the probability that the potential employee will be hired if the potential employee is a female.

B)

(1) Using = 5.6037, find a point estimate of the odds ratio for .

(2) Interpret this odds ratio.

(3) Using the p-value describing the importance of , can we conclude that there is strong evidence that gender is related to the probability that a potential employee will be hired?

(2.5+2.5)

Q4 You have got a new job as a regional head of a drug company. Your company has recently launched a drug to reduce cholesterol level of human blood. However, you are given a task to statistically establish the fact that drug has some effects. As a first step, you identify 10 people with v high cholesterol and run a mini-clinical trial. The sample data here is the cholesterol levels for these 10 men diagnosed with high cholesterol (also in sheet Q4) before and after the drug. The first row represents levels before the drug and the second row, after the drug. (1+4)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cholesterol levels in mg/DL | | | | | | | | |  |
| Before | 237 | 289 | 257 | 228 | 303 | 275 | 262 | 304 | 244 | 233 |
| After | 194 | 240 | 230 | 186 | 265 | 222 | 242 | 281 | 240 | 212 |

1. State your null and alternate hypothesis.
2. At 0.05% level of significance, can you claim that, on average, the drug has some effects on cholesterol levels?

Q5 Your university is planning to introduce special menus for Italian cuisine in their canteen. However, they would like to understand the interest of people in different categories (stakeholders) who would prefer Italian cuisine, especially if the interest is similar across all groups. They would like to understand if the action will be equally encouraging for different types of users. To test this assumption, a random sample of 200 students, 100 faculty, 50 administrative staffs and 50 support staff were randomly asked if they would prefer Italian cuisine. The response of different categories is given in Q5 sheet. Based on this survey outcome, can you say if the interest is equal in all groups of university members at alpha 0.05? What is the conclusion at 1% level of confidence? (2.5+2.5)

Q6 Calculate cost of living using Fisher method, Laspeyre’s Price and Paache method for the data share in Q6 worksheet. ( 2+2+1 marks)

Q7 The manager of a warehouse for a telecommunications company is involved in a process that receives expensive circuit boards and returns them to central stock so that they can be reused at a later date. Speedy processing of these circuit boards is critical in providing good service to customers and reducing capital expenditures.

The data in Q7 excel sheet represent the number of circuit boards processed per day by a subgroup of five employees over a 30-day period.

a. Construct a control chart for the range. (2 marks)

b. Construct a control chart for the mean. (2 marks)

c. Is the process in control? Explain. (1 mark)

Q8 A coffee shop must decide how many newspapers it should order. Each newspaper costs $2 and is sold for $4.25. Any unsold newspapers at the end of the day are returned for a refund of $0.50.

Assume that the daily demand for newspapers varies according to the following rules:

Case a - The demand follows a triangular distribution, with minimum, maximum & most likely values of 90, 140 & 100.

Case b - The demand follows a triangular distribution, with minimum, maximum & most likely values of 140, 160 & 150.

Case c - The demand follows a triangular distribution, with minimum, maximum & most likely values of 160, 210 & 200.

Case d - The demand follows a uniform distribution, between 210 and 250.

Additionally, there is a 35% chance that case A occurs, a 20% chance that case B occurs, a 30% chance that case C occurs and a 15% chance that case D occurs.

Data is shared in Q8 worksheet.

Run Monte Carlo simulation for the profit distribution for this business problem. Determine number of newspaper to be ordered. Show all relevant graphs and figures. (5 marks)

Q9 Suppose that a survey conducted in the 2020 indicated that 7% of healthcare users said that N95 masks would not be enough to protect from COVID. Is there evidence that the proportion of healthcare who said that has changed from the previous year? To test this hypothesis, a sample of 500 healthcare workers has been taken and 25 has reported that they felt N95 masks would not be enough to protect from COVID. (5 marks)

Q10 The average cost of a hotel room in Chicago is said to be $ 170 per night. To determine if this is true, a random sample of 25 hotels is taken and resulted in a sample mean of $ 174 and an S of $ 16.1 Test the appropriate hypotheses at a = 0.05. (Assume the population distribution is normal). (5 marks)

Q11 You want to see if three different cricket clubs yield different scores in T20 match tournaments. You randomly select five matches for each club. At the 0.05 significance level, is there a difference in mean distance? (5 marks)

|  |  |  |
| --- | --- | --- |
| **Cricket Club 1** | **Cricket Club 2** | **Cricket Club 3** |
| 255 | 235 | 201 |
| 264 | 219 | 223 |
| 242 | 236 | 198 |
| 238 | 228 | 207 |
| 252 | 217 | 205 |