|  | **PES University, Bengaluru**  (Established under Karnataka Act No. 16 of 2013) | | **UE20CS902** |
| --- | --- | --- | --- |
| **October 2024: END SEMESTER ASSESSMENT (ESA)**  **M TECH DATA SCIENCE AND ARTIFICIAL INTELLIGENCE\_SEMESTER I**  **UE20CS902 – Statistical Methods for Decision Making** | | | |
| Time: 3 Hrs | | * All questions are compulsory. * Section A should be handwritten in the answer script provided * Section B and C are coding questions which have to be answered in the system. | Max Marks: 100 |

**Section A : 20 Marks**

| 1 | a) | Define skewness and kurtosis. Explain the difference between them in relation to the shape of a distribution. | 2 |
| --- | --- | --- | --- |
| b) | What is the Central Limit Theorem (CLT)? Why is it important in inferential statistics? | 2 |
| c) | Describe the difference between Type I and Type II errors in hypothesis testing. Provide an example for each. | 2 |
| d) | Consider the following dataset:  10, 12, 14, 16, 18, 20  Calculate the variance and standard deviation for the dataset. | 2 |
| e) | Explain the difference between correlation and causation. Provide an example where a strong correlation does not imply causation. | 2 |
|  | | | |
| 2 | a) | A contingency table for the relationship between gender and voting preference is given below:     | Gender | Voted | Did Not Note | | --- | --- | --- | | Male | 100 | 50 | | Female | 120 | 30 |   Calculate the marginal and conditional probabilities. | 2 |
| b) | A company claims that the average time taken to resolve customer complaints is 48 hours. To verify this claim, a random sample of 35 complaints is selected, and the sample mean resolution time is 50 hours with a standard deviation of 8 hours.  Tasks:   1. Construct a 95% confidence interval for the mean resolution time. 2. Based on the confidence interval, can we conclude that the company's claim is incorrect? Explain why or why not. | 4 |
|  | c) | If the probability of event A is 0.3, event B is 0.6, and the probability of both A and B occurring is 0.2, are the events A and B independent? Explain your reasoning with calculations. | 2 |
|  | d) | A factory manufactures light bulbs, and it is known that the proportion of defective bulbs is 0.05. If a sample of 200 bulbs is selected, use the binomial distribution to calculate the probability of finding exactly 10 defective bulbs. | 2 |
| **SECTION B – 40 MARKS** | | | |
| 3 | a) | A company produces light bulbs, and it is claimed that the mean life of the bulbs is 1000 hours. A sample of 10 bulbs gave the following lifetimes (in hours):  940, 955, 1020, 1005, 980, 970, 995, 1015, 990, 1010.  (i) Perform a hypothesis test at the 0.05 significance level to determine if the mean lifetime of the bulbs differs from 1000 hours. (5 marks)  (ii) Calculate the 95% confidence interval for the mean lifetime. (3 marks) | 8 |
| b) | A manufacturer of metal rods wants to ensure that the diameter of the rods meets a specified standard. A random sample of 25 rods is taken, and their diameters (in cm) are measured. The sample has a mean diameter of 5.05 cm and a standard deviation of 0.08 cm.  (i) Conduct a hypothesis test to check if the mean diameter of the rods is different from the specified standard of 5.00 cm, using a 0.01 significance level. (5 marks)  (ii) Construct the 99% confidence interval for the mean diameter. (3 marks) | 8 |
| 3 | c | A health researcher is studying the effect of a new diet on body weight. She records the weights of 15 participants before and after 6 months on the diet:  Before Diet: [82, 85, 90, 88, 84, 87, 89, 92, 93, 88, 90, 91, 86, 89, 90]  After Diet: [78, 83, 87, 85, 80, 83, 86, 88, 89, 85, 86, 87, 82, 85, 86]  (i) Test whether the diet has significantly reduced the participants' weight at a 0.05 significance level. (4 marks)  (ii) Calculate the mean difference in weight and construct a 95% confidence interval for the mean difference. (4 marks) | 8 |
| 3 | d | A study investigates whether there is a relationship between hours of study and exam performance. The following data shows the number of hours studied and the corresponding exam scores for 10 students:  Hours Studied: [2, 4, 6, 8, 10, 12, 14, 16, 18, 20]  Exam Scores: [55, 60, 65, 70, 73, 78, 82, 85, 90, 92]  (i) Perform a linear regression analysis to determine the relationship between hours studied and exam scores. (4 marks)  (ii) Predict the exam score for a student who studied for 15 hours. (2 marks)  (iii) Test the significance of the regression model using a 0.05 significance level. (2 marks) | 8 |
| 3 | e | A clinical trial compares two treatments for lowering blood pressure. The blood pressure levels of 20 patients before and after each treatment are recorded.  Treatment A:  Before = [142, 150, 148, 152, 145, 149, 153, 146, 151, 148]  After = [135, 140, 138, 142, 137, 139, 143, 135, 141, 140]  Treatment B:  Before = [148, 155, 150, 153, 147, 152, 158, 149, 156, 154]  After = [140, 145, 143, 147, 140, 144, 149, 141, 148, 146]  (i) Perform a paired t-test to compare the effect of Treatment A and Treatment B in reducing blood pressure. (5 marks)  (ii) Which treatment is more effective, based on the test results? (1 mark)  (iii) Construct a 95% confidence interval for the difference in blood pressure reduction between the two treatments. (2 marks) | 8 |
| **Section C- 40 Marks** | | | |
| 4 | a | Consider the water\_potability.csv file and answer the following questions   1. What are the summary statistics (mean, median, min, max) for Potability and Hardness? (2 marks) 2. Calculate the 95% confidence intervals for the mean Hardness across potable and non-potable water. Interpret the intervals and discuss any overlaps. (5 marks) 3. Identify and analyze outliers in the Turbidity data using the IQR (Interquartile Range) method. How many outliers are there, and what impact might they have on the overall analysis? (5 marks) 4. Compute the Z-scores for Solids to standardize the data. Identify any values that are considered outliers (Z-score > 3 or < -3). (5 marks) 5. Is there a correlation between Turbidity and Conductivity? If so, how strong is it? (3 marks) (3 marks) | 20 |
|  | b | Consider the car\_data.csv file and answer the following questions   * + 1. Conduct hypothesis testing to see whether the average ph level in potable water is higher than that in non-potable water (state hypothesis, conduct test, and conclude). (4 marks)     2. Conduct hypothesis testing to see whether the Organic\_carbon levels in potable water are higher than in non-potable water (state hypothesis, conduct test, and conclude). (4 marks)     3. Conduct hypothesis testing to determine if the Hardness levels differ between potable and non-potable water (state hypothesis, conduct test, and conclude). (4 marks)     4. Conduct a statistical analysis to find whether there is a statistical difference in Hardness among water samples across different pH ranges (e.g., low pH < 6.5, normal pH 6.5–8.5, high pH > 8.5). (8 marks) | 20 |