SRN

| | | |
|---|---|---|
| PES UNIVERSITY | **PES University, Bengaluru**<br>(Established under Karnataka Act No. 16 of 2013) | **UX21MB143A** |

**March 2024: END SEMESTER ASSESSMENT (ESA)**
**EXECUTIVE MBA- SEMESTER I**
**Statistical Methods**
**UX21MB143A**

| Time: 3 Hrs | Answer All Questions | Max Marks: 100 |
|---|---|---|

**Instructions**

- Q1 and Q2 are to be answered in the answer script provided.
- Q3, Q4 and Q5 are coding questions which have to be answered in the system.

| | | | |
|---|---|---|---|
| 1 | a) | Explain the difference between a population and a sample. Why are samples used in statistics? | 5 |
| | b) | What are the 4 sampling techniques used in statistical research? Describe any one of them in detail with an example. | 5 |
| | c) | Why do most of the sample means differ from the population mean? What is this difference called? Define the Central limit Theorem. | 5 |
| | d) | Give the definition of a 95% confidence interval for estimation of a population mean. Also give its formula along with the meanings of the symbols. | 5 |
| | | | |
| 2 | a) | What is meant by a Type I error and a Type II error? How are they related? What symbols are used to represent the probabilities of Type I and Type II errors.Which is more serious with respect to its consequences and how can we safeguard against that error ? | 6 |
| | b) | Two manufacturers supply food to a large cafeteria. Manufacturer A supplies 2400 cans of soup, and 3% are found to be dented. Manufacturer B supplies 3600 cans, and 1% are found to be dented. Given that a can of soup is dented, find the probability that it came from manufacturer B. | 7 |
| | c) | Fastra Auto has the following cars in stock<br><br>**SUV** **Compact** **Mid-sized**<br>Foreign 20 50 20<br>Domestic 65 100 45<br><br>If a car is selected at random find the probability that it is :<br>1. Foreign<br>2. Foreign and mid-sized<br>3. Domestic or an SUV<br>4. SUV given it is domestic<br>5. Foreign given it is mid-sized<br>6. Are the events ‘compact’ and ‘domestic’ independent? | 7 |

SRN

| | | | |
|---|---|---|---|
| 3 | | Assume the below for part a & b:<br>The mean systolic blood pressure of normal adults is 120 millimeters of mercury (mm Hg) and the standard deviation is 5.6. Assume the variable is normally distributed. | |
| | a) | If an individual is selected, find the probability that the individual's pressure will be between 120 and 121.8 mm Hg. | 5 |
| | b) | For another study a researcher wants to select the middle 60% of the population based on blood pressure. What are the lower and upper limits of blood pressure that would qualify a person to participate in the study? | 5 |
| | c) | A random sample of state gasoline taxes (in cents) is shown here for 12 states. Use the data to estimate the true population mean gasoline tax with 90% confidence. Does your interval contain the national average of 44.7 cents? Does it suggest that the population average is different from 44.7. Assume normal distribution.<br>38.4 40.9 67 32.5 51.5 43.4 38 43.4 50.7 35.4 39.3 41.4 | 10 |
| | | | |
| 4 | | Consider the healthinsurance.csv file and answer the following questions | |
| | a) | Provide a summary statistics of all the numerical variables, based on the summary which variable do you think has more variability. | 5 |
| | b) | Provide a histogram for the variable bmi, based on histogram and calculation of mean, median and mode what would be the closest distribution shape you would suggest (use bin_size=20). | 5 |
| | c) | Plot the boxplots for variable claim based on smoker, diabetes, sex. Based on these plots what do you observe and what do you conclude. | 5 |
| | d) | Draw a scatter plot for all of the variables, what is your observation and conclusion for the relationship between age and claim charges, bmi and claim charges, would you prefer to further sub classify and develop a scatter plot, if yes or no why | 5 |
| | | | |
| 5 | | Consider the healthinsurance.csv file and answer the following questions. Assume normality of the variable Claim, conduct the following at a significance level $\alpha = 0.01$ | |
| | a) | Perform a T test to compare whether smokers claim more than nonsmokers (use the complete data rather than random sampling of data, state the hypothesis and conclude accordingly) | 5 |
| | b) | Perform a T test to compare whether diabetics claim more than non diabetics (use the complete data rather than random sampling of data, state the hypothesis and conclude accordingly) | 5 |
| | c) | Conduct an ANOVA to compare between insurance charges of various age category and conclude accordingly (Assume of Equality of variance) | 5 |
| | d) | Conduct an ANOVA to compare between insurance charges of various BMI_Levels and conclude accordingly (Assume of Equality of variance) | 5 |