B. Tech. Third Semester (Artificial Intelligence and Data Science) / 21-22_SOE_ADS_203.1 Examination

Course Code : AIDS 2202 **Course Name : Statistics for Data Science**

Time : 3 Hours] [Max. Marks : 50

Instructions to Candidates :—

(1) Do not write anything on question paper except your exam seat number.

(2) Write the accurate question number in left margin of answer book along with answers.
(3) All questions are compulsory.
(4) Assume suitable data wherever necessary and mention at the beginning of answer.
(5) Use of Logarithmic tables, non – programmable calculator, Steam tables, Mollier's chart, Drawing instruments, Thermodynamic tables for moist air, Psychrometric charts and Refrigeration charts is permitted.
(6) **No additional answer book will be supplied.**

1. (A) Virginia Suboleski is an aircraft maintenance supervisor. A recent delivery of bolts from a new supplier caught the eye of a clerk. Suboleski sent 25 of the bolts to a testing lab to determine the force necessary to break each of the bolts. In thousands of pounds of force, the results are as follows :

147·8	137·4	125·2	141·1	145·7
119·9	133·3	142·3	138·7	125·7
142·0	130·8	129·8	141·2	134·9
125·0	128·9	142·0	118·6	133·0
151·1	125·7	126·3	140·9	138·2

(a) Arrange the data into an array from highest to lowest.

(b) What proportion of the bolts withstood at least 120,000 pounds of force ? What proportion withstood at least 150,000 pounds ?

(c) If Suboleski knows that these bolts when installed on aircraft are subjected to up to 140,000 pounds of force, what proportion of the sample

bolts would have failed in use ? What should Suboleski recommend the company do about continuing to order from the new supplier ? 3(CO1)

(B) On June 30, 1992, the capitalizations of nine Asian/Pacific stock markets were :

Country	Capitalization (Billions of U.S. $)
Philippines	17
Indonesia	21
Thailand	44
Singapore	50
Malaysia	79
South Korea	86
Taiwan	140
Hong Kong	178
Australia	203

(a) Find the arithmetic mean of the data.

(b) Find the median of the data.

(c) Find the mode of the data.

(d) Which is the best measure of central tendency of the data ?

(e) Find the standard deviation of the data. (The entire population is included in the data.) 5(CO1)

2. (A) In a staff meeting called to address the problem of returned checks at the supermarket where you are interning as a financial analyst, the bank reports that 12 percent of all checks are returned for insufficient funds, and of those, in 50 percent of cases, there was cash given back to the customer. Overall, 10 percent of customers ask for cash back at the end of their transaction with the store. For 1,000 customer visits, how many transactions will involve :

(a) Insufficient funds ?

(b) Cash back to the customer ?

(c) Both insufficient funds and cash back ?

(d) Either insufficient funds or cash back ? 4(CO2)

(B) Draw Venn diagrams to represent the following situations involving three events, A, B and C, which are part of a sample space of events but do not include the whole sample space :

(a) Each pair of events (A and B; A and C, and B and C) may occur together, but all three may not occur together.

(b) A and B are mutually exclusive, but not A and C nor B and C.

(c) A, B and C are all mutually exclusive of one another.

(d) A and B are mutually exclusive, B and C are mutually exclusive, but A and C are not mutually exclusive. 4(CO2)

3. (A) Explain probability distribution. Also explain Binomial Distribution, Poisson Distribution and Normal Distribution. 5(CO2)

(B) Glenn Howell, VP of personnel for the Standard Insurance Company, has developed a new training program that is entirely self-paced. New employees work various stages at their own pace ; completion occurs when the material is learned. Howell's program has been especially effective in speeding up the training process, as an employee's salary during training is only 67 percent of that earned upon completion of the program. In the last several years, average completion time of the program was 44 days, and the standard deviation was 12 days :

(a) Find the probability an employee will finish the program in 33 to 42 days.

(b) What is the probability of finishing the program in fewer than 30 days ?

(c) Fewer than 25 or more than 60 days ? 3(CO2)

4. (A) (1) Explain Sampling and different sampling types.

(2) Tread-On-Us has designed a new tire, and they don't know what the average amount of tread life is going to be. They do know that tread life is normally distributed with a standard deviation of 216·4 miles :

(a) If the company samples 800 tires and records their tread life, what is the probability the sample mean is between the true mean and 300 miles over the true mean ?

(b) How large a sample must be taken to be 95 percent sure the sample mean will be within 100 miles of the true mean ?

2+2(CO4)

(B) Farmer Braun, who sells grain to Germany, owns 60 acres of wheat fields. Based on past experience, he knows that the yield from each individual acre is normally distributed with mean 120 bushels and standard deviation 12 bushels. Help Farmer Braun plan for his next year's crop by finding :

(a) The expected mean of the yields from Farmer Braun's 60 acres of wheat.

(b) The standard deviation of the sample mean of the yields from Farmer Braun's 60 acres.

(c) The probability that the mean yield per acre will exceed 123.8 bushels.

(d) The probability that the mean yield per acre will fall between 117 and 122 bushels.

4(CO4)

5. (A) Gwen Taylor, apartment manager for Willow Wood Apartments, wants to inform potential renters about how much electricity they can expect to use during August. She randomly selects 61 residents and discovers their average electricity usage in August to be 894 kilowatt hours (kwh). Gwen believes the variance in usage is about 131 $(kwh)^2$.

(a) Establish an interval estimate for the average August electricity usage so Gwen can be 68.3 percent certain the true population mean lies within this interval.

(b) Repeat part (a) with a 99.7 percent certainty.

(c) If the price per kwh is \$0.12, within what interval can Gwen be 68.3 percent certain that the average August cost for electricity will lie ?

3(CO3)

(B) Suppose a sample of 50 is taken from a population with standard deviation 27 and that the sample mean is 86.

(a) Establish an interval estimate for the population mean that is 95·5 percent certain to include the true population mean.

(b) Suppose, instead, that the sample size was 5,000. Establish an interval for the population mean that is 95·5 percent certain to include the true population mean.

(c) Why might estimate (a) be preferred to estimate (b) ? Why might (b) be preferred to (a) ? 3(CO3)

(C) The following sample of eight observations is from an infinite population with a normal distribution :

75·3, 76·4, 83·2, 91·0, 80·1, 77·5, 84·8, 81·0

(a) Find the sample mean.

(b) Estimate the population standard deviation.

(c) Construct a 98 percent confidence interval for the population mean. 3(CO3)

6. (A) Maxwell's Hot Chocolate is concerned about the effect of the recent year-long coffee advertising campaign on hot chocolate sales. The average weekly hot chocolate sales two years ago was 984·7 pounds and the standard deviation was 72·6 pounds. Maxwell's has randomly selected 30 weeks from the past year and found average sales of 912·1 pounds.

(a) State appropriate hypotheses for testing whether hot chocolate sales have decreased.

(b) At the 2 percent significance level, test these hypotheses. 3(CO1)

(B) The average commission charged by full-service brokerage firms on a sale of common stock is \$144, and the standard deviation is \$52. Joel Freelander has taken a random sample of 121 trades by his clients and determined that they paid an average commission of \$151. At a 0·10 significance level, can Joel conclude that his clients' commissions are higher than the industry average ? What happens to the power of the test for μ. = \$140, \$160 and \$175 if the significance level is changed to 0·05 ? 6(CO1)