Birla Institute of Technology & Science, Pilani

MATH F432 (Applied Statistical Methods)
First Semester 2023–2024
Comprehensive Examination (Open Book)

Time: 180 Minutes **Date: December 13, 2023 (Wednesday)** **Max. Marks: 80**

The question paper has two parts: PART A and PART B. The first part (PART A) contains 15 questions, each carrying 4 marks. The second part (PART B) has two subjective questions, each carrying 10 marks. Once you submit PART A, the question paper of PART B will be distributed. Notations and symbols have their usual meaning.

Time: 120 Minutes **PART A** **Max. Marks: 60**

For the following objective-type questions, **just write the final answer(s)** in the space provided below. Neither there is any part marking nor is there any negative marking for these questions.

1. Let X be a continuous random variable with pdf

$$f(x) = \begin{cases} x, & 0 \le x \le 1 \\ 2-x, & 1 \le x \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

Write down the pdf of Y, where $Y = 2X + 3$.
Answer:

2. If bolt thread length in a factory is normally distributed, what is the probability that the thread length of a randomly selected bolt is within the 1.5 SDs (standard deviations) of its mean value?
Answer:

3. Suppose that the lifetime of Philips brand light bulbs is modeled by an exponential distribution with rate parameter λ. We test 5 bulbs and find they have lifetimes of 2, 3, 1, 3, and 6 years, respectively. What is the MLE estimate for $(\lambda^2 + \lambda)$?
Answer:

4. To test the hypothesis $H_0 : \mu = 95$ versus $H_1 : \mu \neq 95$, a sample of size 16 resulted in $\bar{x} = 94.32$. Assume that the population is normal with $\sigma = 1.2$. For $\alpha = 0.01$, compute $\beta(94)$.
Answer:

5. For a dataset with 40 rows (observations) and 4 columns (variables), a factor analysis is carried out. How many eigen values are there for the data covariance matrix? Let the eigen value of the first factor be 1.365. If the factor loadings corresponding to the first factor are $k, 0.35, 1-k, 0.85$, then compute the value of k.
Answer:

6. In order to determine the effects of temperature on the survival of insect eggs, following observations are made. At 11^oC, 73 of 91 eggs survived to the next stage of development, whereas at 30^oC, 102 of 110 eggs survived. Do the results suggest that the proportion of survival of insect eggs differs for the two temperatures? Write down the $p-$value and mention your decision at $\alpha = 0.02$.
Answer:

7. Below are the US box office gross for the first 11 weekends of the release of “Mission Impossible III” in 2006.

Weekend (x)	1	2	3	4	5	6	7	8	9	10	11
Gross in million USD (y)	47.70	25.00	11.35	7.00	4.68	3.02	1.34	0.72	0.49	0.31	0.20

Obtain an exponential model of type $y = ab^x$ for this data.
Answer:

8. In a discriminant analysis to discriminate between "like" and "dislike", the following portion of analysis is available.

Person	Like/Dislike	Discriminant Score (Y)
1	dislike	0.148
2	dislike	0.809
3	dislike	0.735
4	dislike	1.250
5	dislike	1.176
6	like	1.691
7	like	2.353
8	like	2.279
9	like	2.721
10	like	2.721

Based on above, obtain the cut-off score that can be used to discriminate between "like" and "dislike". If the discriminant scores for three new persons are 1.581, 2.021, and 1.750, in which group would you classify them?

Answer:

9. Based on the information in question Q8, obtain between-group variability, within group variability, Wilk's λ, and eigen value of the discriminant function.

Answer:

10. When $n = 150$, what is the smallest value of $\bar{p}$ for which the LCL in a $p-$chart is positive? If $n = 75$, what would be this value?

Answer:

11. In a SLRM with $n = 14, \sum x_i = 517, \sum y_i = 346, \sum {x_i}^2 = 39095, \sum {y_i}^2 = 17454, \sum x_i y_i = 25825$, obtain the estimated simple linear regression equation. If $SSE = 395.747$, calculate the value of r^2.

Answer:

12. Assuming that two population standard deviations σ_1 and σ_2 are unknown but equal, obtain the formula for a $(1-\alpha)\%$ pooled CI for $(\mu_1 - \mu_2)$. If $n_1 = 16, n_2 = 11, \bar{x}_1 = 25, \bar{x}_2 = 30, {s_1}^2 = 2.25, {s_2}^2 = 4.00$, obtain a 90% pooled CI estimate for $(\mu_1 - \mu_2)$ when $\sigma_1 = \sigma_2$.

Answer:

13. A random sample of $n = 15$ heat pumps of a certain type yielded the following observations on lifetime (in years): 2.0, 1.3, 6.0, 1.9, 5.1, 0.4, 5.3, 1.0, 15.7, 0.7, 4.8, 0.9, 12.2, 5.3, 0.6. Assuming that the lifetime distribution is exponential, provide a 95% CI estimate for the standard deviation of the lifetime distribution.

Answer:

14. For a sample of 20 observations $\{1, 3, 5, 7, 9, 14, 16, 24, 26, 27, 32, 37, 40, 40, 41, 72, 80, 86, 87, 90\}$ on component life-time (hr), the sign-test is applied to test $H_0 : \tilde{\mu} = 25$ versus $H_1 : \tilde{\mu} > 25$. Let the test statistic be Y, the number of observations that exceed 25. (i) If we decide to reject H_0 for $Y \geq 15$, what would be the probability of type-I error for this test? (ii) For the given 20 sample observations, would you reject the null hypothesis at $\alpha = 5\%$?

Answer:

15. Compute the mean and variance for a continuous random variable X having the following pdf

$$f(x) = \begin{cases} \frac{1}{24}x\left(1 - \frac{x}{12}\right), & 0 \leq x \leq 12 \\ 0, & \text{otherwise.} \end{cases}$$

Answer:

Birla Institute of Technology & Science, Pilani

MATH F432 (Applied Statistical Methods)
First Semester 2023–2024
Comprehensive Examination (Open Book)

Time: 60 Minutes **PART B** **Max. Marks: 20**

Notations and symbols have their usual meaning. Assumptions, made if any, should be stated clearly. Start new question on fresh page. Moreover, answer each subpart of a question in continuation.

1. Police records show the following number of daily crime reports for 10 sample days during the winter months and 10 sample days during the summer months. Using the MWW test at $\alpha = 0.05$, decide whether there is any significant difference between the winter and summer months in terms of the number of crime reports? Discuss each step clearly. **[10]**

Winter	18	20	15	16	21	20	12	16	19	20
Summer	28	18	24	32	18	29	23	38	28	18

2. Time series analysis of crime counts is an inevitable tool to understand seasonal patterns of crime counts, such as in summer months and in winter months, and to know if crime count is gradually on the increasing or decreasing over time. Below is an example of crime counts in a city over four years.

Year	Jan–Mar	Apr–Jun	Jul–Sep	Oct–Dec
2010	34	51	38	26
2011	40	60	48	32
2012	42	65	51	34
2013	46	70	54	38

(i) Construct a time series plot. What type of pattern seems to exist in the data? (ii) Calculate the four-quarter and centered moving average values for the data. (iii) Compute seasonal indexes for the four quarters and (iv) deseasonalize the time series. (v) Compute the linear trend equation for the deseasonalised data and perform quarterly forecast for the next year. (vi) Finally, adjust the linear trend forecasts using the seasonality indexes computed above. **[10]**

END