UNIVERSITY OF JAFFNA

FACULTY OF ENGINEERING

END SEMESTER EXAMINATION– AUGUST, 2023

MC3020- PROBABILITY AND STATISTICS

Date: 02 - 08 -2023 Duration: TWO Hours

Instructions

1. This paper contains **FIVE (5)** questions:
2. Answer **<u>all</u>** questions in the answer book provided.
3. If you have any doubt as to the interpretation of the wording of a question, make your own decision, but clearly state in the script.
4. This examination accounts for **50%** of module assessment. Total maximum mark attainable is **100**.

Question 1[20 marks]

1. The Engineering Students' Union (ESU) at University of Jaffna has Seven members: four women and three men. Three are selected to attend a workshop seminar related ragging awareness program, which is held in Colombo. Estimate the following probabilities:
   (a) All three selected will be women.
   (b) At least one woman will be selected.
2. In a group of engineering students at the University of Jaffna, 60% of the students are proficient in coding, and 40% are proficient in robotics. If 30% of the students are proficient in both coding and robotics, what is the probability that a randomly chosen engineering student from the University of Jaffna is proficient in either coding or robotics (or both)?

Question 2[20 marks]

1. If you put an ordinary incandescent bulb into a light fixture designed for a high efficiency (low heat) bulb, it is possible–but not certain–that the contact point will melt and the light bulb will become stuck in the socket. Extra labor is then required to remove it and clean the fixture's contact point. An electrician estimates that these extra tasks must be performed 0.5370 of the time when replacing an incorrect bulb. The electrician has been called to a house on another matter, and notices they have used the wrong light bulb type in five of their downlights.

(a) If the electrician points out the hazard and is employed to replace the bulbs, what distribution would best be used to model the number of bulbs that will be stuck in the socket?

(b) If the electrician is working his way through a large apartment block, and each apartment has eight incorrect bulbs that must be replaced, what is the long run proportion of times at least five of the bulbs will be stuck and require the extra steps?

(c) What is the mean and variance of the distribution of the number of stuck bulbs?

2. The traffic Police officer's radar speed gun causes grief for many drivers who use the Jaffna-Kandy (A9) road. A particular location between Ariviyal Nagar and Iranaimadu junction in Kilinochchi catches 12 speeding motorists per hour. This figure is an average of all hours between 9am and 5pm.

    (a) Which distribution would best be used to model the number of speeding motorists in a 10 minutes period? and what proportion of 10 minutes period would you expect the Police officer's radar speed gun to remain unused to deliver a fine?

    (b) What is the probability that the Police officer catches at-least three motorists in 15 minutes period?

    (c) Which distribution would best be used to model the amount of time between speeding motorists passing the Police officer?

    (d) What proportion of speeding motorists will get their fine taken within 30 minutes of the previous speeding motorist?

Question 3[20 marks]

1. As a part of the evaluation of Colombo municipal council employees, the city manager audits the parking tickets issued by city parking officers to determine the number of tickets that were contested by the car owner and found to be improperly issued. In past years, the number of improperly issued tickets per officer had a normal distribution with mean $\mu = 380$. Because there has recently been a change in the city's parking regulations, the city manager suspects that the mean number of improperly issued tickets has increased. An audit of 12 randomly selected officers is conducted to test whether there has been an increase in improper tickets. Use the sample data given here which is audit collected from each officers:

$$390, 380, 369, 392, 398, 393, 392, 396, 399, 391, 387, 393$$

    (a) Give a point estimate of the mean number of improperly issued tickets. Construct a 95% confidence interval on the mean number of improperly issued tickets. Interpret this interval.

    (b) Is there sufficient evidence that the mean number of improperly issued tickets is greater than 380? Use a $\alpha = 0.01$.

(c) Is there sufficient evidence that the variance number of improperly issued tickets is greater than 35? Use a $\alpha = 0.05$.

2. An engineering team is developing a new coating material for metal surfaces used in a specific industrial application. They are testing two different formulations to determine if there is a significant difference in the wear resistance between the two coatings. To evaluate this, the team performs wear tests on metal samples coated with each formulation. A total of 24 metal samples are tested, with 12 samples coated using Formulation A (treated group) and 12 samples coated using Formulation B (untreated group). After conducting the wear tests, the wear depth in micrometers for each sample is measured, and the data are summarized in the table below:

   Formulation A (Treated Group - Wear Depth in Micrometers):

   $$18, 43, 28, 50, 16, 32, 13, 35, 38, 33, 6, 7$$

   Formulation B (Untreated Group - Wear Depth in Micrometers):

   $$40, 54, 26, 63, 21, 37, 39, 23, 48, 58, 28, 39$$

   The engineering team wants to determine if Formulation A provides a significant improvement in wear resistance compared to Formulation B. Assume that the population distributions of the measurements are normal with equal variances.

   (a) Construct a 95% confidence interval for the difference in mean wear depth $(\mu_1 - \mu_2)$ to assess the size of the difference between the two means.

   (b) Test whether the mean wear depth of metal samples coated with Formulation A is significantly less than the mean wear depth of samples coated with Formulation B using a significance level (α) of 0.05.

Question 4[20 marks]

1. Dylan is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year, so the prior probability of rain is just 5/365. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 5% of the time. What is the probability that it will rain on the day of Dylan's wedding?

2. In the dynamic city of Ariviyal Nagar, renowned for its scientific and technological advancements, engineering companies constantly seek talented individuals to join their workforce. One such company is TechSolutions, a leading tech firm with a reputation for hiring exceptional engineering professionals. As part of their recruitment process, TechSolutions considers various factors to ensure they hire the best candidates. One crucial aspect is evaluating the IQ scores of engineering job applicants. These IQ scores are known to follow a normal distribution with a mean of 125 and a standard deviation of 10.

(a) What is the probability that an engineering job applicant has an IQ score less than 135?

(b) TechSolutions aims to identify top-performing candidates. Among engineering job applicants, what IQ score corresponds to the 95th percentile?

(c) TechSolutions has a policy of rejecting engineering job applicants with an IQ score below 110. In their recent recruitment drive, there were 800 engineering job applicants. How many applicants were rejected based on this IQ score threshold?

(d) TechSolutions seeks to hire candidates with above-average intelligence. Between which two IQ scores, symmetrically located around the mean, do 94% of the engineering job applicants' IQ scores fall?

Question 5[20 marks]

1. Suppose you are an electrical engineer tasked with predicting the power consumption of a household appliance based on its operating voltage. You collected data from various appliances and measured their operating voltage (in volts) and corresponding power consumption (in watts). The dataset is as follows:

   Operating Voltage (V): 110, 125, 130, 141, 145, 160
   Power Consumption (W): 75, 68, 70, 80, 90, 100

   (a) Find the equation of the least squares regression line for predicting power consumption based on operating voltage.

   (b) Interpret the slope coefficient of the regression line in the context of the problem.

   (c) Use the regression line to predict the power consumption for an appliance with an operating voltage of 135V.

   (d) Calculate the coefficient of determination (R-squared) for the regression line and interpret its meaning.

2. Suppose you are a civil engineer tasked with predicting the flexural strength of concrete beams based on their dimensions. You collected data from various concrete beams and measured their width (in inches), depth (in inches), and length (in feet), as well as their corresponding flexural strength (in pounds per square inch, psi). The dataset is as follows:

Width (inches)	Depth (inches)	Length (feet)	Flexural Strength (psi)
8	6	2	3000
10	10	5	3500
12	14	6	4000
14	16	7	4200
16	18	10	4500
18	20	18	5000
20	22	20	5500
22	26	22	6000

Answer the following questions using the R programs output given Figure 1 (without necessarily employing any formulas):

(a) Find the equation of the multiple linear regression line for predicting flexural strength based on the dimensions of the concrete beams.

(b) Interpret the coefficients of the regression line in the context of the problem.

(c) Use the regression line to predict the flexural strength of a concrete beam with width 15 inches, depth 17 inches, and length 16 feet.

(d) Perform an individual coefficient test to determine the significance of the width variable in predicting flexural strength. State your hypotheses, the test statistic, and the conclusion clearly.

```
> summary(Model_2)

Call:
lm(formula = Flexural_Strength ~ Width + Depth + Length)

Residuals:
      1        2        3        4        5        6        7        8 
 37.113   -6.443   41.753  -12.629  -58.763 -134.278   65.464   67.784 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) 2234.536    394.297   5.667  0.00478 **
Width          5.541     89.517   0.062  0.95361   
Depth         98.711     48.191   2.048  0.10990   
Length        45.876     23.201   1.977  0.11916   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 91.77 on 4 degrees of freedom
Multiple R-squared:  0.9952,	Adjusted R-squared:  0.9917 
F-statistic: 278.9 on 3 and 4 DF,  p-value: 4.239e-05

>
```

Figure 1: R output for part (2) of question 5

Important useful formulas

1. If X follows binomial with parameters n, p. Then, the probability mass function is given by

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}; \ x = 0, 1, 2, \cdots n$$

2. If X follows Poisson with parameters λ. Then, the probability mass function is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}; \ x = 0, 1, 2, \cdots\cdots$$

3. If X follows exponential distribution with parameter λ. Then, the probability density function is given by

$$f(x) = \lambda e^{-\lambda x}; \quad x > 0$$

4. $\mu = E(X) = \sum_{\text{for all x}} xP(X = x)$ and $\sigma^2 = Var(X) = E(X^2) - (E(X))^2$
or $Var(X) = E(X - \mu)^2 = \sum_{\text{for all x}} (x - \mu)^2 P(X = x)$

5. Let us consider that a sample space S is divided into two mutually exclusive partitions S_1 and S_2. An event H has occurred, and $P(S_1|H)$ can be written as

$$P(S_1|H) = \frac{P(S_1)P(H|S_1)}{P(S_1)P(H|S_1) + P(S_2)P(H|S_2)}$$

6. If X follows normal distribution with parameter μ and σ then $\left(\frac{X-\mu}{\sigma}\right) = Z$ follows standard normal distribution with $\mu = 0$ and $\sigma = 1$.

7. Sample mean ($\bar{X}$) and standard deviation (s) can be estimated

$$\bar{X} = \frac{\sum X}{n}, \quad s = \sqrt{\frac{\sum \left(X - \bar{X}\right)^2}{n-1}} \quad = \sqrt{\frac{n\sum X^2 \; - \; \left(\sum X\right)^2}{n(n-1)}}$$

8.

Parameter	$(1-\alpha) * 100\%$ confidence interval	Test statistic value
Mean (μ)	Case 1: $\left(\bar{X} \mp Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right)$	Case 1: $Z = \frac{\bar{X}-\mu_0}{(\frac{\sigma}{\sqrt{n}})}$
	Case 2: $\left(\bar{X} \mp t_{\frac{\alpha}{2},df} * \frac{S}{\sqrt{n}}\right)$	Case 2: $T = \frac{\bar{X}-\mu_0}{(\frac{S}{\sqrt{n}})}, df = n-1$
	Case 3: $\left(\bar{X} \mp Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right)$	Case 3: $Z = \frac{\bar{X}-\mu_0}{(\frac{\sigma}{\sqrt{n}})}$
	Case 4: $\left(\bar{X} \mp Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}\right)$	Case 4: $Z = \frac{\bar{X}-\mu_0}{(\frac{S}{\sqrt{n}})}$
Variance (σ)	$\left(\frac{(n-1)*S^2}{\chi^2_U}, \frac{(n-1)*S^2}{\chi^2_L}\right)$	$\chi = \frac{(n-1)s^2}{\sigma_0^2}, df = n-1$

Case 1: when population is normal and σ is known, case 2: when population is normal and σ is unknown, case 3: when population is not normal and σ is known, sample size n is large and case 3: when population is not normal and σ is known, sample size n is large.

9.

Parameter	$(1-\alpha) * 100\%$ confidence interval	Test statistic value
Dependent populations	Case 1: $\left(\bar{D} \mp Z_{\frac{\alpha}{2}} * \frac{\sigma_D}{\sqrt{n}}\right)$	Case 1: $Z = \frac{\bar{D}-\mu_0}{(\frac{\sigma_D}{\sqrt{n}})}$
Mean difference ($\mu_1 - \mu_2 = \mu_D$)	Case 2: $\left(\bar{D} \mp t_{\frac{\alpha}{2},df} * \frac{s_D}{\sqrt{n}}\right)$	Case 2: $T = \frac{\bar{D}-\mu_0}{(\frac{s_D}{\sqrt{n}})}, df = n-1$
	Case 3: $\left(\bar{D} \mp Z_{\frac{\alpha}{2}} * \frac{\sigma_D}{\sqrt{n}}\right)$	Case 3: $Z = \frac{\bar{D}-\mu_0}{(\frac{\sigma_D}{\sqrt{n}})}$
	Case 4: $\left(\bar{D} \mp Z_{\frac{\alpha}{2}} * \frac{s_D}{\sqrt{n}}\right)$	Case 4: $Z = \frac{\bar{D}-\mu_0}{(\frac{s_D}{\sqrt{n}})}$

Case 1: when population distribution of the differences is normal and σ_D is known, case 2: when population distribution of the differences is normal and σ_D is unknown, case 3: when population distribution of the differences is not normal and σ_D is known, sample size n is large and case 4: when population distribution of the differences is not normal and σ_D is known, sample size n is large.

10. Sample mean ($\bar{D}$) and standard deviation (s_D) for the differences can be estimated

$$\bar{D} = \frac{\sum D}{n}, \quad s_D = \sqrt{\frac{\sum \left(D - \bar{D}\right)^2}{n-1}} \quad = \sqrt{\frac{n\sum D^2 \; - \; \left(\sum D\right)^2}{n(n-1)}}$$

11.

Parameter	$(1-\alpha)*100\%$ **confidence interval**	**Test statistic value**
Independent populations	Case 1: $\left(\bar{X}-\bar{Y} \mp Z_{\frac{\alpha}{2}} * \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$	Case 1: $Z=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$
Mean difference $(\mu_1-\mu_2)$	Case 2: $\left(\bar{X}-\bar{Y} \mp t_{\frac{\alpha}{2}} * \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)$	Case 2: $T=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ $df=\frac{(A+B)^2}{\frac{A^2}{n_1-1}+\frac{B^2}{n_2-1}}; A=\frac{s_1^2}{n_1},\ B=\frac{s_2^2}{n_2},$
	Case 3: $\left(\bar{X}-\bar{Y} \mp t_{\frac{\alpha}{2}} * s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$ where $s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$	Case 3: $T=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $df=n_1+n_2-2$
	Case 4: $\left(\bar{X}-\bar{Y} \mp Z_{\frac{\alpha}{2}} * \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$	Case 4: $Z=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$
	Case 5: $\left(\bar{X}-\bar{Y} \mp Z_{\frac{\alpha}{2}} * \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)$	Case 5: $Z=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$
Proportions difference (p_1-p_2)	$\left(\hat{p_1}-\hat{p_2} \mp Z_{\frac{\alpha}{2}} * \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1}+\frac{\hat{p_2}(1-\hat{p_2})}{n_2}}\right)$	$Z=\frac{\hat{p_1}-\hat{p_2}-(p_1-p_2)}{\bar{p}(1-\bar{p})(\frac{1}{n_1}+\frac{1}{n_2})}$ $\bar{p}=\frac{X_1+X_2}{n_1+n_2}$

Case 1: When the two independent population distributions are normal and the population variances σ_1^2 and σ_2^2 are known, case 2: When the two independent population distributions are normal and the population variances σ_1^2 and σ_2^2 are unknown and unequal, case 3: When the two independent population distributions are normal and the population variances σ_1^2 and σ_2^2 are unknown but equal, case 4: When the two independent population distributions are not normal and the population variances σ_1^2 and σ_2^2 are known, and the sample size n_1 and n_2 are large and case 5: When the two independent population distributions are not normal and the population variances σ_1^2 and σ_2^2 are known, and the sample size n_1 and n_2 are large.

12. Linear regression coefficient estimation formulas,

$$\hat{\beta_1}=\frac{\sum_{i=1}^{n} X_iY_i-n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2-n\bar{X}^2} \quad \hat{\beta_0}=\bar{Y}-\hat{\beta_1}\bar{X}$$

13. Standard error (S_e) is given by

$$S_e=\sqrt{\frac{\sum_{i=1}^{n}\left(Y_i-\hat{Y_i}\right)^2}{n-2}}$$

14. Correlation (r) is given by

$$r=\frac{\sum_{i=1}^{n} X_iY_i-n\bar{X}\bar{Y}}{\sqrt{\left(\sum_{i=1}^{n} X_i^2-n\bar{X}^2\right)\left(\sum_{i=1}^{n} Y_i^2-n\bar{Y}^2\right)}}$$

15. Test statistic value when testing the hypothesis for correlation is given by,

$$T=\frac{r}{\sqrt{\frac{1-r^2}{n-2}}}, \quad \text{d.f}=n-2$$

16. Test statistic value when testing the hypothesis for slope coefficient is given by,

$$T=\frac{\hat{\beta_1}-0}{\frac{Se}{\sqrt{\sum_{i=1}^{n} X_i^2-n\bar{X}^2}}}, \quad \text{d.f}=n-2$$

17. Test statistic value when testing the hypothesis for intercept coefficient is given by,

$$T = \frac{\hat{\beta}_0 - 0}{Se\sqrt{\dfrac{1}{n} + \dfrac{\bar{X}^2}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}}}, \quad \text{d.f} = n - 2$$

——————— *End of Examination* ———————