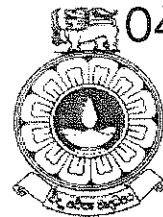


J

0433

The Open University of Sri Lanka
Faculty of Engineering Technology
Department of Mathematics and Philosophy of Engineering



Study Programme : Bachelor of Software Engineering Honours
Name of the Examination : Final Examination
Course Code and Title : MHZ4377 Applied Statistics
Academic Year : 2024/25
Date : 21st February 2026
Time : 0930-1230h
Duration : 3 hours

General instructions

REGISTRATION NO:

- **Part A is Compulsory**

- Provide short answers in the given space.
- Do not need to show any workings.
- Answer **five (05)** questions only out of seven (07) questions from **Part B**.
 - Provide answers in separate sheets (answer booklet) which will be provided in the examination.
 - Show all your workings.
- The number of pages in the paper is Eight(08).
- All the symbols are in standard notation unless they are defined.
- All the relevant statistical tables and equations are provided.
- You are allowed to use a non-programmable calculator during the examination.
- **Attach Part A to the Answer script of Part B.**
- Answer for each question should commence from a new page.
- This is a Closed Book Test (CBT).

Provide short answers in the given space. Do not need to show any workings or work. Please attach this part to your answer script.

1. State whether the following variables are discrete or continuous:

(i) Number of machine breakdowns per month:

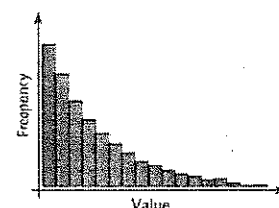
(ii) Time taken to complete a production cycle :

2. What happens to the standard deviation if all observations in a dataset are multiplied by 5?

Answer:

3. State the shape of the distribution represented by the following histogram.

Answer:



4. If $X \sim \text{Bin}(50, 0.3)$, find mean of X .

Answer:

5. If A and B are mutually exclusive events with $P(A)=0.4$ and $P(B)=0.5$, find $P(A \cup B)$.

Answer:

6. If the probability of selecting a defective electric component is 0.2, the probability of selecting a non-defective electric component is,

Answer:

7. What is the meaning of the level of significance in hypothesis testing?

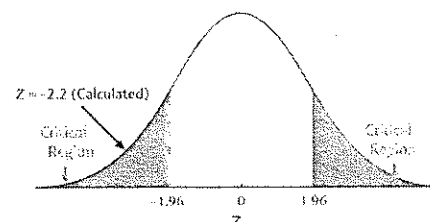
Answer:

8. Find the missing value k : Mean of 5, 7, k , 9, 11 is 8.

Answer:

9. State the decision that can be made according to the position of calculated and critical Z values as displayed in the graph when doing hypothesis testing.

Answer:



10. In hypothesis testing, the probability of failing to reject H_0 when it is false is called,

Answer:

Part B

Q1.

- (a) Consider the following extract from a Software Bug Tracking Log maintained at TechNova Solutions (Pvt) Ltd.

Software Bug Tracking Log

Bug ID	Application Module	Platform / Version	Number of Errors	Error Severity
B100	User Authentication	Web v1.2	3	Minor
B101	Payment Processing	Android v2.0	2	Minor
B102	Inventory Management	Web v1.3	6	Major
B103	Reporting Dashboard	iOS v2.1	0	None
B104	Order Management	Web v1.2	4	Moderate

- (i) Determine the type of data and scale of the measurement associated with each of the following variables in the above Software Bug Tracking log.

(12 marks)

	<i>Type of data</i>	<i>Scale of measurement</i>
Application Module		
Platform / Version		
Number of Errors		

- (ii) Suppose you are the QA team member at TechNova Solutions (Pvt) Ltd. Based on the defect severity levels recorded, propose a simple coding scheme to quantify defect severity for further statistical analysis.

(05 marks)

<i>Defect Severity</i>	<i>Suggested Code</i>
None	
Minor	
Moderate	
Major	

- 0433
- (iii) Then identify the scale of measurement associated with the coded variable in part (ii) and suggest the best statistical summary for central tendency to use on this variable.

(08 marks)

- (b) A large-scale software development company aims to analyze software resource waste generated in three development teams (Team A, Team B, and Team C) during one sprint week.

The company uses two types of development approaches:

- Agile Methodology (Iterative Development Process)
- Traditional Waterfall Methodology (Sequential Development Process)

At the end of the sprint week, project supervisors recorded the resource waste (measured in hours of rework due to bugs, failed builds, or redundant coding).

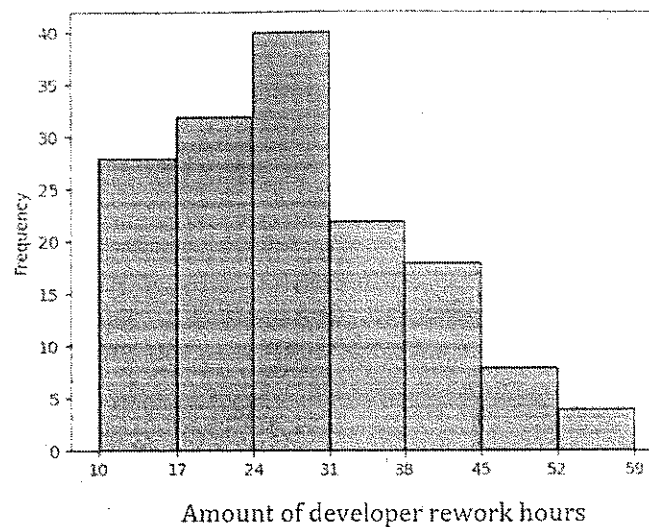
Team	Agile Methodology (Hours)	Waterfall Methodology (Hours)
A	12	24
B	15	18
C	10	30

The company management plans to reduce resource waste under the Traditional Waterfall methodology by 25% in to minimize development inefficiencies, reduce rework hours, and improve overall software quality.

- (i) Draw a suitable graph to compare resource waste across development teams under Agile and Waterfall methodologies.
(15 marks)
- (ii) Explain the differences in resource waste among Team A, Team B, and Team C under both Agile and Waterfall methodologies.
(10 marks)
- (iii) Identify which team must reduce the most resource waste under the Waterfall methodology to achieve the company's 25% waste reduction target.
(15 marks)

- (c) A large-scale software development company monitored the amount of developer rework hours (due to bugs, failed builds, and redundant coding) per development module during one sprint week. A sample of 152 development modules was recorded and summarized in a histogram.

Histogram of amount of developer rework hours



- (i) Comment about the shape of the distribution of the data. (10 marks)
 - (ii) Draw the frequency polygon corresponding to the histogram. (15 marks)
 - (iii) Determine the total number of development modules that are required between 24 and 45 rework hours during the sprint week. (10 marks)
- Q2.** At a large-scale software development company, quality assurance inspectors recorded the number of software bugs detected per 1000 lines of code (LOC) from 21 different software modules during a sprint review. Each value represents the number of detected bugs per module. The stem represents the tens digit, and the leaf represents the units digit.

Stem	Leaf
0	1 1 2 2 3 4
1	0 1 1 1 2 5 5
2	3 3 4
3	5 6
4	8
5	9
6	9

- (a) Find the Central Tendency measures (Mean, Median, Mode) of the given data. (25 marks)
- (b) The software development company wants to divide the 21 software modules into four equal performance groups based on the number of bugs detected per 1000 lines of code: Best performing modules (lowest bug counts), Lower-middle group, Upper-middle group, Worst performing modules (highest bug counts). Using the previously given data, find:

- i. First Quartile (Q_1)
- ii. Second Quartile (Median, Q_2)
- iii. Third Quartile (Q_3)

(30 marks)

- (c) Obtain the inter quartile range of the number of bugs detected by using the results obtained in part (b). Interpret what this tells about the performance levels of the software development. **(10 marks)**
- (d) Construct a box and whisker plot to represent the given data, clearly indicating the presence of any outliers. **(25 marks)**
- (e) State, giving a reason for which measure of central tendency, you would recommend summarizing this data. **(10 marks)**

Q3. A quality assurance analyst studying software defects defines the random variable X = "number of critical bugs detected in a selected software module"

The probability distribution of X is given by:

$$P(X = x) = c(x + 2), \quad x = 0, 1, 2, 3$$

where c is a constant.

- (a) Find the value of c such that the function becomes a valid probability distribution. **(15 marks)**
- (b) Find the probability that:
 - (i) At most 2 critical bugs were detected
 - (ii) Exactly 3 critical bugs were detected**(20 marks)**
- (c) Find the expected number of critical bugs per software module. **(25 marks)**
- (d) Find the variance of the number of critical bugs per software module. **(25 marks)**
- (e) Find the probability that the number of critical bugs lies between 1 and 3.

(15 marks)**Q4.**

- (a) Discuss the characteristics of Geometric distribution and the characteristics of Binomial distribution. **(20 marks)**
- (b) In a software company, a quality assurance team inspects software modules to identify defective modules before release. Previous studies show that the probability that a module is defective is 0.03.
 - (i) Find the probability that the first defective is found at the 4th module inspected. **(20 marks)**

- (ii) Find the probability that more than 2 modules must be inspected before finding the first defective module. (20 marks)
- (iii) Suppose the same inspection process is applied to a batch of 25 modules. Find the probability that 12% of the modules are defective. (40 marks)

Q5.

- (a) A software engineering research team studies the percentage performance degradation of a web application after the first optimization cycle to evaluate system efficiency. A random sample of 10 software modules was tested, and the performance degradation percentages were recorded as follows:

performance degradation (%)	2.1	1.8	2.4	2.0	1.9	2.3	2.2	1.7	2.0	2.1
--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- (i) Determine the best estimate for the population mean performance degradation. (25 marks)
- (ii) Obtain the best estimate for population variance. (25 marks)
- (b) A software engineering research team evaluates the average execution time per software module (measured in milliseconds) after implementing a new performance optimization technique. A random sample of 120 software modules was selected from the system. The findings revealed: that the sample mean execution time is $\bar{x} = 85\text{ms}$ and the sample standard deviation is $s = 4\text{ms}$.
- (i) Find the standard error of the estimation of population mean. (15 marks)
- (ii) Obtain the interval estimate for the population mean at 5% significant level. (30 marks)
- (iii) Interpret your result. (05 marks)

Q6. A software company claims that a new code optimization framework reduces the average execution time per batch process without affecting system functionality, thereby improving deployment efficiency. The traditional system produces an average execution time of $\mu_0 = 1200$ milliseconds with known population standard deviation of $\sigma = 60$ milliseconds. To test the claim, a random sample of $n = 150$ batch processes using the new framework is examined. The sample mean execution time is found to be $\bar{x} = 1185$ milliseconds.

- (a) State the null hypotheses and alternative hypotheses for testing the claim. (20 marks)
- (b) Find the test statistics associated with this hypothesis testing.

- (c) At the 1% significance level, determine whether there is sufficient evidence to support the claim that the new code optimization framework reduces the average execution time. (30 marks)
- (d) State your conclusion regarding the hypothesis and comment on the claim. (20 marks)

Q7.

A software engineering researcher investigates the relationship between development effort measured in person-days and Module development cost (thousand USD) during system testing for a set of completed software modules. The following summary statistics were obtained from the collected data:

Development effort(person days)	120	130	140	150	160	170
Module development cost (Thousand USD)	68	72	75	78	80	83

You may use the following summary statistics:

$$\sum X = 870, \sum Y = 456, \sum X^2 = 126,700, \sum Y^2 = 34,722, \sum XY = 66,840$$

- (a) Draw a scatter diagram and comment on the form and strength of the relationship between development effort and Module development cost. (10 marks)
- (b) Calculate the correlation coefficient for the above data and interpret it. (20 marks)
- (c) Determine the least squares estimates of the regression coefficients β_0 and β_1 for the model: $Y = \beta_0 + \beta_1 X$ (30 marks)
- (d) Write down the estimated regression equation and interpret the slope coefficient in a software cost context. (10 marks)
- (e)
- (i) Predict the development cost when effort is 155 person-days. (10 marks)
- (ii) If the observed cost at 155 person-days is 77 thousand USD, calculate the residual and interpret its sign. (10 marks)
- (f) Explain why prediction at 220 person-days may not be reliable. (10 marks)

All Rights Reserved

Important Formulas:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Binomial distribution

$$Pr(X = r) = \frac{n!}{(r!(n-r)!)} p^r q^{(n-r)}$$

Geometric distribution

$$Pr(X = r) = pq^{(r-1)}$$

Poisson Distribution

$$Pr(X = r) = e^{(-\mu)} \frac{\mu^r}{r!}$$

If random variable X , follows normal distribution ($X \sim N(\mu, \sigma^2)$) Then $Z = \frac{(X-\mu)}{\sigma}$ follows standard normal distribution where $Z \sim (1, 0)$.

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$

$$\beta_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

