

MIDTERM EXAMINATION (UTS)

Statistics and Probability

Weeks 1–7

Lecturer:

Alfrian Carmen Talakua, S.SI., M.Kom

Duration:

90 Minutes

Department of Information Systems

Instructions:

- Answer all questions carefully.
 - Multiple choice questions: choose the best answer.
 - Show all calculation steps clearly for problem-solving questions.
-

PART A — MULTIPLE CHOICE (20 QUESTIONS)

1. Statistics is the science of ...
- A. Computer programming
 - B. Data analysis and interpretation
 - C. Operating systems
 - D. Network design

Answer: B

2. Data in numerical form is called ...
- A. Qualitative data
 - B. Nominal data
 - C. Quantitative data
 - D. Ordinal data

Answer: C

3. Descriptive statistics is used to ...
- A. Test hypotheses
 - B. Summarize data
 - C. Build AI systems
 - D. Delete data

Answer: B

4. The average value of a dataset is called ...
- A. Median
 - B. Mode
 - C. Mean
 - D. Range

Answer: C

5. The middle value in an ordered dataset is called ...

- A. Mean
- B. Variance
- C. Median
- D. Mode

Answer: C

6. The value that appears most frequently in a dataset is called ...

- A. Mean
- B. Variance
- C. Mode
- D. Standard deviation

Answer: C

7. Standard deviation measures ...

- A. The center of data
- B. Data variability
- C. Total observations
- D. Data percentage

Answer: B

8. Probability refers to ...

- A. Total amount of data
- B. Chance of an event occurring
- C. A database system
- D. A graph format

Answer: B

9. Probability values always range between ...

- A. -1 and 1
- B. 0 and 100
- C. 0 and 1
- D. 1 and 10

Answer: C

10. The sample space of rolling one die is ...

- A. {1,2,3}
- B. {2,4,6}
- C. {1,2,3,4,5,6}
- D. {0,1}

Answer: C

11. The probability of rolling an even number on a die is ...

- A. $1/6$
- B. $2/6$
- C. $3/6$
- D. $6/6$

Answer: C

12. A random variable is ...

- A. A text variable
- B. A numerical outcome of a random experiment
- C. A chart type
- D. A linear equation

Answer: B

13. Which of the following is a discrete random variable?

- A. Weight
- B. Height
- C. Number of students
- D. Temperature

Answer: C

14. Which of the following is a continuous random variable?

- A. Number of cars
- B. Number of students
- C. Height
- D. Number of patients

Answer: C

15. A normal distribution has the shape of a ...

- A. Triangle
- B. Rectangle
- C. Bell curve
- D. Straight line

Answer: C

16. In a normal distribution ...

- A. Mean \neq Median
- B. Mean = Median = Mode
- C. Median $>$ Mean
- D. Mode = 0

Answer: B

17. A Z-score is used to measure ...

- A. The position of data relative to the mean
- B. Data deletion
- C. Graph creation
- D. Mode calculation

Answer: A

18. The Central Limit Theorem states that ...

- A. All data are discrete
- B. All data are continuous
- C. The sampling distribution approaches normality as sample size increases
- D. All distributions have mean zero

Answer: C

19. The formula for standard error is:

- A. $\sigma \times n$
- B. σ / n
- C. σ / \sqrt{n}
- D. n / σ

Answer: C

20. As sample size increases ...

- A. Standard error decreases
- B. Standard error increases
- C. Mean becomes zero
- D. Distribution disappears

Answer: A

PART B — CALCULATION & CASE STUDY QUESTIONS (30 QUESTIONS)

Central Tendency and Dispersion

21. Calculate the mean of the following dataset:

60, 70, 80, 90, 100

Answer:

$$\text{Mean} = (60 + 70 + 80 + 90 + 100) / 5$$

$$\text{Mean} = 400 / 5$$

$$\text{Mean} = 80$$

22. Calculate the median of the following dataset:

45, 50, 55, 60, 65

Answer:

$$\text{Median} = 55$$

23. Calculate the mode of the following dataset:

70, 75, 75, 80, 90

Answer:

$$\text{Mode} = 75$$

24. Calculate the range of the following dataset:

30, 40, 55, 70, 90

Answer:

$$\text{Range} = 90 - 30 = 60$$

25. Calculate the mean of the following dataset:

65, 70, 75, 80, 85, 90

Answer:

Mean = 77.5

26. Calculate the median of the following dataset:

55, 60, 65, 70

Answer:

Median = 62.5

27. Calculate the mode of the following dataset:

5, 7, 7, 8, 9, 9, 9

Answer:

Mode = 9

28. Calculate the variance of the following dataset:

60, 70, 80

Answer:

Variance = 66.67

29. Calculate the standard deviation of the following dataset:

60, 70, 80

Answer:

Standard Deviation \approx 8.16

30. Interpret the following result:

Mean = 70

Standard Deviation = 5

Explain what this says about the data spread.

Answer:

The data values are relatively close to the mean, indicating low variability.

Probability

31. A die is rolled once. Calculate the probability of getting the number 4.

$$P(A) = \frac{n(A)}{n(S)}$$

Answer:

$$P(4) = 1/6$$

32. A die is rolled once. Calculate the probability of getting an odd number.

Answer:

$$P(\text{odd}) = 3/6 = 1/2 = 0.5$$

33. Two dice are rolled simultaneously. Calculate the probability that the sum equals 7.

Answer:

$$P(\text{sum} = 7) = 6/36 = 1/6$$

34. A bag contains:

- 5 red balls
- 3 blue balls
- 2 green balls

Calculate the probability of selecting a blue ball.

Answer:

$$P(\text{blue}) = 3/10 = 0.3$$

35. During the last 30 days, it rained on 18 days. Calculate the empirical probability of rain.

Answer:

$$P(\text{rain}) = 18/30 = 0.6$$

36. If $P(A)=0.7$, calculate the complement probability.

$$P(A') = 1 - P(A)$$

Answer:

$$P(A') = 1 - 0.7 = 0.3$$

37. A number is randomly selected from 1–20. Calculate the probability of selecting a number greater than 15.

Answer:

$$P(x > 15) = 5/20 = 1/4 = 0.25$$

38. One card is drawn from a standard deck of 52 cards. Calculate the probability of drawing a heart card.

Answer:

$$P(\text{heart}) = 13/52 = 1/4 = 0.25$$

39. Calculate the probability of rolling an even number on a die.

Answer:

$$P(\text{even}) = 3/6 = 1/2 = 0.5$$

40. In a survey of 200 children, 45 were identified as stunted. Calculate the probability that a randomly selected child is stunted.

Answer:

$$P(\text{stunted}) = 45/200 = 0.225$$

Random Variables and Probability Distribution

41. Determine whether the following variables are discrete or continuous:

- Number of vehicles
- Weight
- Number of patients
- Travel time

Answer:

- Discrete
 - Continuous
 - Discrete
 - Continuous
-

42. Given the following probability distribution:

X	P(X)
0	0.2
1	0.5
2	0.3

Calculate the expected value.

$$E(X) = \sum x \cdot P(x)$$

Answer:

$$\begin{aligned} E(X) &= (0 \times 0.2) + (1 \times 0.5) + (2 \times 0.3) \\ &= 0 + 0.5 + 0.6 \\ &= 1.1 \end{aligned}$$

43. Using the distribution in Question 42, calculate the variance.

Answer:

$$\text{Variance} = 0.49$$

44. Using the distribution in Question 42, calculate the standard deviation.

Answer:

$$\text{Standard Deviation} = 0.7$$

45. Explain the meaning of expected value in a probability distribution.

Answer:

Expected value represents the long-run average outcome of a random variable.

Normal Distribution and Z-Score

46. Given:

- Student score = 85
- Mean = 75
- Standard deviation = 10

Calculate the Z-score.

$$z = \frac{x - \mu}{\sigma}$$

x 1.2

μ 0.0

σ 1.0

$z = \frac{x - \mu}{\sigma} \approx 1.2$

$\Phi(z) \approx 88.5\%$



Answer:

$$z = (85 - 75) / 10$$

$$z = 1$$

47. Interpret the meaning of a Z-score equal to 1.5.

Answer:

The value is 1.5 standard deviations above the mean.

48. Explain the 68–95–99.7 empirical rule in normal distribution.

Answer:

- 68% of data lies within 1 standard deviation
 - 95% lies within 2 standard deviations
 - 99.7% lies within 3 standard deviations
-

Sampling Distribution

49. Given:

- $\sigma = 12$
- $n = 36$

Calculate the standard error.

$$SE = \frac{\sigma}{\sqrt{n}}$$

Answer:

$$SE = 12 / \sqrt{36}$$

$$SE = 12 / 6$$

$$SE = 2$$

50. A company wants to estimate the average package delivery time. They collect a sample of 50 transactions from a population of 10,000 transactions.

Explain:

- a. Which group represents the population
- b. Which group represents the sample
- c. Why sampling is important in data science

Answer:

a. Population = 10,000 transactions

b. Sample = 50 transactions

c. Sampling helps estimate population characteristics efficiently while reducing time and