BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
Work Integrated Learning Programmes Division
First Semester 2025-2026
Mid-Semester Test
(EC-2 Regular)

Course No.	:	AIMLCZC418
Course Title	:	Introduction to Statistical Methods
Nature of Exam	:	Closed Book
Weightage	:	30%
Duration	:	2 Hours
Date of Exam	:	20-12-2025_FN

No. of Pages = 4
No. of Questions = 6

Note to Students:
1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

Answer all the questions:

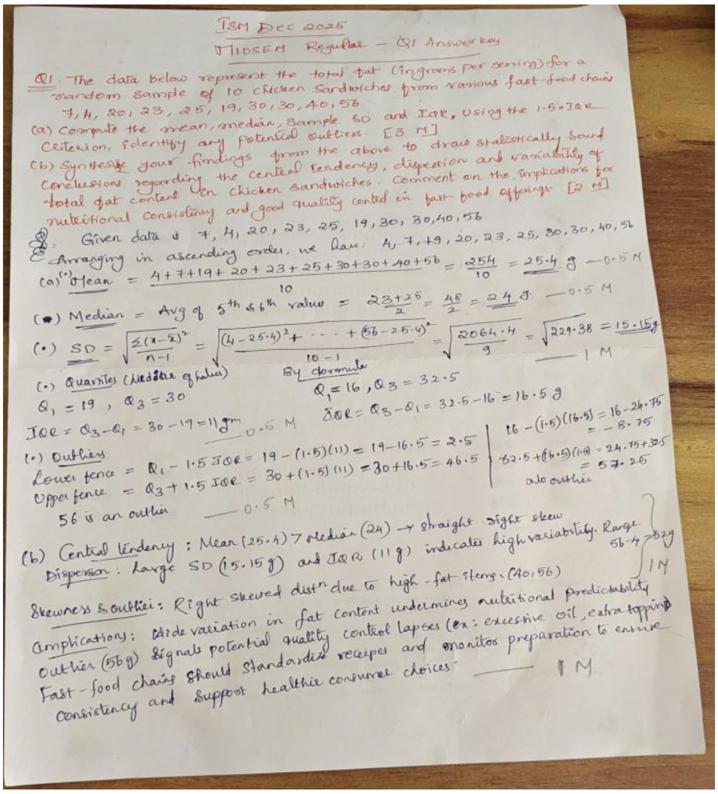
1. The data below represent the total fat (in grams per serving) for a random sample of 10 chicken sandwiches from various fast-food chains:

   7, 4, 20, 23, 25, 19, 30, 30, 40, 56

a. Compute the mean, median, sample standard deviation and interquartile range (IQR). Using the 1.5 × IQR criterion, identify any potential outliers. [3 marks]

b. Synthesize your findings from the above to draw statistically sound conclusions regarding the central tendency, dispersion, and variability of total fat content in chicken sandwiches. Comment on the implications for nutritional consistency and quality control in fast-food offerings. [2 marks]

Solution:



ISM Dec 2025

MIDSEM Regular – Q1 Answer key

Q1. The data below represent the total fat (in grams per serving) for a random sample of 10 chicken sandwiches from various fast-food chains
7, 4, 20, 23, 25, 19, 30, 30, 40, 56

(a) Compute the mean, median, Sample SD and IQR, using the 1.5×IQR criterion, identify any potential outliers. [3 M]

(b) Synthesize your findings from the above to draw statistically sound conclusions regarding the central tendency, dispersion and variability of total fat content in chicken sandwiches. Comment on the implications for nutritional consistency and good quality control in fast food offerings. [2 M]

Sol: Given data is 7, 4, 20, 23, 25, 19, 30, 30, 40, 56

Arranging in ascending order, we have: 4, 7, 19, 20, 23, 25, 30, 30, 40, 56

(a) (•) Mean $= \frac{4+7+19+20+23+25+30+30+40+56}{10} = \frac{254}{10} = 25.4$ g — 0.5 M

(•) Median = Avg of 5^{th} & 6^{th} value $= \frac{23+25}{2} = \frac{48}{2} = 24$ g — 0.5 M

(•) $SD = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{(4-25.4)^2 + \cdots + (56-25.4)^2}{10-1}} = \sqrt{\frac{2064.4}{9}} = \sqrt{229.38} = 15.15$ g — 1 M

(•) Quartiles (Median of halves)
$Q_1 = 19$, $Q_3 = 30$
$IQR = Q_3 - Q_1 = 30 - 19 = 11$ gm — 0.5 M

By formula
$Q_1 = 16$, $Q_3 = 32.5$
$IQR = Q_3 - Q_1 = 32.5 - 16 = 16.5$ g

(•) Outliers
Lower fence $= Q_1 - 1.5\,IQR = 19 - (1.5)(11) = 19 - 16.5 = 2.5$
Upper fence $= Q_3 + 1.5\,IQR = 30 + (1.5)(11) = 30 + 16.5 = 46.5$
56 is an outlier — 0.5 M

$16 - (1.5)(16.5) = 16 - 24.75 = -8.75$
$32.5 + (16.5)(1.5) = 24.75 + 32.5 = 57.25$
no outlier

(b) Central tendency: Mean (25.4) > Median (24) → slight right skew
Dispersion: Large SD (15.15 g) and IQR (11 g) indicates high variability. Range 56 − 4 = 52 g
Skewness & outlier: Right skewed $dist^n$ due to high-fat items. (40, 56) — 1 M

Implications: Wide variation in fat content undermines nutritional predictability. Outlier (56 g) signals potential quality control lapses (ex: excessive oil, extra toppings). Fast-food chains should standardise recipes and monitor preparation to ensure consistency and support healthier consumer choices — 1 M

2. A school wants to build a simple system to classify student messages as "Urgent" or "Not Urgent" using the Naïve Bayes classifier. Here is a small sample dataset: [5 marks]

Message ID	Words in Message	Label
1	assignment due tomorrow please help	Urgent
2	lunch plans today?	Not Urgent
3	exam rescheduled urgent attention	Urgent
4	thanks for the notes	Not Urgent

New Message to Classify: "assignment urgent" Using Naïve Bayes with Laplace smoothing ($\alpha = 1$), classify this message as Urgent or Not Urgent. Show brief working.

Solution:

Step 1: Vocabulary & Counts

Unique words (V): {assignment, due, tomorrow, please, help, lunch, plans, today, exam, rescheduled, urgent, attention, thanks, for, the, notes} → $|V| = 16$

Class counts:
Urgent (U): 2 messages
Not Urgent (N): 2 messages
$P(U) = P(N) = 2/4 = 0.5$

Word counts per class (with Laplace $\alpha = 1$):

Word	Count in U	Count in N
assignment	1	0
urgent	1	0
others	—	—

Total words in U: 9 → after smoothing: 9 + 16 = 25
Total words in N: 7 → after smoothing: 7 + 16 = 23

Step 2: Likelihoods for "assignment urgent"

$P(\text{assignment} \mid U) = (1 + 1)/(9 + 16) = 2/25$
$P(\text{urgent} \mid U) = (1 + 1)/25 = 2/25$
$\rightarrow P(U \mid \text{msg}) \propto P(U) \times 2/25 \times 2/25 = 0.5 \times 4/625 = 0.0032$

$P(\text{assignment} \mid N) = (0 + 1)/23 = 1/23$
$P(\text{urgent} \mid N) = (0 + 1)/23 = 1/23$
$\rightarrow P(N \mid \text{msg}) \propto 0.5 \times 1/23 \times 1/23 = 0.5 / 529 \approx 0.000945$

Step 3: Compare

$0.0032 > 0.000945 \Rightarrow$ Classify as Urgent

Answer: Urgent

3. a) A random variable X has the following probability function:

x	0	1	3	4	5	6	7
p (x)	0	k	2k	2k	3k	k^2	$7k^2 + k$

i) Find the value of k [1 mark]
ii) Evaluate P (X < 6), P (X ≥ 6) [1 mark]
iii) P (0 < X < 5) [1 mark]

Solution:

i) $k = \frac{1}{10}$
ii) $P(X < 6) = \frac{8}{10} = 0.8$, $P(X \geq 6) = \frac{2}{10} = 0.2$
iii) $P(0 < X < 5) = P(X = 1) + P(X = 3) + P(X = 4) = k + 2k + 2k = 5k = \frac{5}{10} = 0.5$

b) A factory produces high-precision surgical blades, and quality control data shows that 0.2% (i.e., one-fifth of a percent) of blades are defective. The blades are packed in sterilized packets of 10 each. In a consignment of 10,000 packets, use the Poisson approximation to estimate how many packets contain: (Assume defects occur independently.) [2 marks]

i. No defective blade
ii. Exactly one defective blade

Solution:

Poisson parameter

For one packet, the expected number of defective blades is

$\lambda = np = 10 \times 0.002 = 0.02$

So, the number of defectives per packet is approximated by $X \sim \text{Poisson}(\lambda = 0.02)$

i) Packet contains no defective blade

P(X=0)= e^{-0.02}≈0.9802

In 10,000 packets: 10,000×0.9802≈9,802 packets

(ii) Packet contains exactly one defective blade

$P(X=1) = \lambda e^{-\lambda} = 0.02 \times e^{-0.02} \approx 0.0196$

In 10,000 packets: 10,000×0.0196≈196 packets

4. Let X and Y be continuous random variables with joint probability density function:

$$f(x,y) = \begin{cases} 6x^2 y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$ [5marks]

i. Verify that f(x,y) is a valid joint PDF.
ii. Find the marginal densities $f_X(x)$ and $f_Y(y)$
iii. Compute P(X<0.5, Y>0.5)
iv. Are X and Y independent? Justify.
v. Compute E[X], E[Y] and E[XY].

Solution: **(i) Verification of a valid joint PDF**

A joint PDF must satisfy:

1. **Non-negativity:** $6x^2y \geq 0$ for $x, y \geq 0$
2. **Total probability = 1:**

$$\int_0^1 \int_0^1 6x^2 y\, dy\, dx = \int_0^1 6x^2 \left[\frac{y^2}{2}\right]_0^1 dx = \int_0^1 3x^2 dx$$

$$= 3\left[\frac{x^3}{3}\right]_0^1 = 1$$

Hence, $f(x, y)$ is a valid joint PDF.

(ii) Marginal densities

Marginal of X:

$$f_X(x) = \int_0^1 6x^2 y\, dy = 6x^2\left[\frac{y^2}{2}\right]_0^1 = 3x^2, \quad 0 \le x \le 1$$

Marginal of Y:

$$f_Y(y) = \int_0^1 6x^2 y\, dx = 6y\left[\frac{x^3}{3}\right]_0^1 = 2y, \quad 0 \le y \le 1$$

(iii) Probability P(X<0.5, Y>0.5)

$$P = \int_0^{0.5} \int_{0.5}^1 6x^2 y \, dy \, dx$$

First integrate w.r.t. y:

$$\int_{0.5}^1 y \, dy = \left[\frac{y^2}{2} \right]_{0.5}^1 = \frac{1}{2} - \frac{1}{8} = \frac{3}{8}$$

So,

$$P = \int_0^{0.5} 6x^2 \cdot \frac{3}{8} \, dx = \frac{9}{4} \int_0^{0.5} x^2 dx$$

$$= \frac{9}{4} \left[\frac{x^3}{3} \right]_0^{0.5} = \frac{9}{4} \cdot \frac{1}{24} = \boxed{\frac{3}{32}}$$

(iv) Independence of X and Y

Check whether

$$f(x, y) = f_X(x) f_Y(y)$$

$$f_X(x) f_Y(y) = (3x^2)(2y) = 6x^2 y = f(x, y)$$

Hence, X and Y are independent.

(v) Expectations

$E[X]$

$$E[X] = \int_0^1 x f_X(x) \, dx = \int_0^1 x(3x^2) \, dx = 3 \int_0^1 x^3 dx = \frac{3}{4}$$

$E[Y]$

$$E[Y] = \int_0^1 y f_Y(y) \, dy = \int_0^1 y(2y) \, dy = 2 \int_0^1 y^2 dy = \frac{2}{3}$$

$E[XY]$

Since X and Y are independent:

$$E[XY] = E[X]E[Y] = \frac{3}{4} \times \frac{2}{3} = \frac{1}{2}$$

5. The inside diameters of washers produced by an automated manufacturing machine are normally distributed. A random sample of 200 washers yields a mean diameter of 0.500 cm and a standard deviation of 0.005 cm. For proper function in an automotive assembly, each washer must have an inside diameter between 0.492 cm and 0.506 cm. Washers outside this range are classified as defective. Assuming the sample estimates accurately reflect the population parameters, [5 marks]

   (a) Compute the proportion of washers expected to be defective.
   (b) Interpret the result in the context of process capability (no calculation required).

Solution:(a) Proportion of washers expected to be defective (3.5Marks)

Step 1: Convert limits to z-scores

Lower limit:

$$z_1 = \frac{0.492 - 0.500}{0.005} = \frac{-0.008}{0.005} = -1.6$$

Upper limit:

$$z_2 = \frac{0.506 - 0.500}{0.005} = \frac{0.006}{0.005} = 1.2$$

Step 2: Find probability within specifications

From standard normal tables: $P(Z<1.2)=0.8849$, $P(Z<-1.6)=0.0548$

$P(0.492 \le X \le 0.506) = P(-1.6 \le Z \le 1.2) = 0.8849 - 0.0548 = 0.8301$

Step 3: Probability of defective washers

$P(\text{defective}) = 1 - 0.8301 = 0.1699$

About 16.99% (≈ 17%) of washers are expected to be defective.

(b) Interpretation in terms of process capability (1.5Marks)

A defect rate of nearly **17%** is **unacceptably high** for an automated manufacturing process. This indicates that:

- The process variation is **too large** relative to the specification limits.
- The process is **not capable** of consistently meeting design requirements.
- Improvements such as **reducing variability**, **better machine calibration**, or **process control adjustments** are needed.

6. A meteorological department is analyzing the probabilities of three independent weather alerts being issued on a given day in a region: [5 marks]

A: Flood Alert **B**: Landslide Alert **C**: Storm Alert

Based on historical data, they observe:

Each alert (A, B, or C) is **equally likely** to be issued: P(A)=P(B)=P(C)
The probability that **any two alerts occur together** (e.g., flood and landslide, landslide and storm, or storm and flood) is: $P(A \cap B) = P(B \cap C) = P(C \cap A) = \frac{1}{8}$
The probability that **all three alerts occur together** is: $P(A \cap B \cap C) = \frac{1}{16}$
The probability of **no alerts at all** (completely calm weather) is: $P(\bar{A} \cap \bar{B} \cap \bar{C}) = \frac{1}{4}$
Now, help the department to compute:

a) The probability that a **Flood Alert** is issued
b) The probability that **exactly two** of the alerts are issued
c) The probability that **at least one** alert is issued
d) The probability that **exactly one** alert is issued

Solution:

(a) Probability that a Flood Alert is issued, P(A)

Use inclusion–exclusion for "no alert":

$P(\bar{A} \cap \bar{B} \cap \bar{C}) = 1 - [P(A)+P(B)+P(C)] + [P(A \cap B)+P(B \cap C)+P(C \cap A)] - P(A \cap B \cap C)$

Substitute values:

$$\frac{1}{4} = 1 - 3p + 3\left(\frac{1}{8}\right) - \frac{1}{16}$$

$$\frac{1}{4} = 1 - 3p + \frac{3}{8} - \frac{1}{16} = \frac{21}{16} - 3p$$

$$3p = \frac{17}{16} \quad \Rightarrow \quad p = \frac{17}{48}$$

$$\boxed{P(A) = \frac{17}{48}}$$

(b) Probability that exactly two alerts are issued

Exactly two alerts occur when a pair occurs but not all three.

$P(\text{exactly two}) = [P(A \cap B)+P(B \cap C)+P(C \cap A)] - 3P(A \cap B \cap C)$

$$= 3\left(\frac{1}{8}\right) - 3\left(\frac{1}{16}\right) = \frac{3}{16}$$

$$\boxed{P(\text{exactly two alerts}) = \frac{3}{16}}$$

(c) Probability that at least one alert is issued

$$P(\text{at least one}) = 1 - P(\text{no alert}) = 1 - \frac{1}{4} = \frac{3}{4}$$

d) Probability that exactly one alert is issued

P(exactly one)=P(at least one)−P(exactly two)−P(all three)

$$= \frac{3}{4} - \frac{3}{16} - \frac{1}{16} = \frac{12}{16} - \frac{4}{16} = \frac{8}{16} = \frac{1}{2}$$

$$\boxed{P(\text{exactly one alert}) = \frac{1}{2}}$$

****