UTM UNIVERSITI TEKNOLOGI MALAYSIA

FACULTY OF COMPUTING
UTM Johor Bahru

Semester II 2024/2025

Subject	:	PROBABILITY & STATISTICAL DATA ANALYSIS (SECI 2143)
Task	:	Chapter 3 & Chapter 4
Due Date	:	**Week 9 (12 – 16 May 2025)**

INSTRUCTION:

1. This is a **GROUP** assignment. Please clearly write the group members **NAME & MATRIC NUMBER** in the front page of the submission.
2. This assignment contributes to 5% of overall course marks.
3. Only **HANDWRITTEN** submission is accepted:
    a. Submissions using any reporting or statistical tools (e.g.: MS Word, MS Excel, etc.,) will be **REJECTED**.
    b. Make sure the submission is neatly written. Any submission with handwriting that is unreadable, will be **REJECTED**.
    c. For answer that need to draw graphs, using graph paper is optional. You can use plain paper.
    d. Round your answers to **THREE** decimal places.
    e. Please scan/snapshot your work and save as a PDF file.
4. Submission via eLearning – only **ONE** group member needs to submit on behalf of the group.

Group 5

	Team Members	**Matrix Number**
1	ASWINI A/P CHANDRASAGARAN	SX242452ECRHF01
2	IZWAN AZIZ BIN ISMAIL @ ABD MALEK	SX241894ECJHF01
3	FATIN SYAHIRAH BINTI NOR RASHID	SX241920ECRHF01
4	SITI NURNAJIHAH BINTI MOHAMAD ANUAR	SX232351ECRHF04
5	YUARAJ A/L PARTHIPAN	SX241919ECRHF01

No:
Subject:
Date:

QUESTION 1

Based on the value, calculate the :

a) Mean

Class Interval	Mid Point , x_i	frequency , f_i	$f_i x_i$
150 - 160	$(150+160) \div 2 = 155$	12	$(155 \times 12) = 1860$
160 - 170	$(160+170) \div 2 = 165$	20	$(165 \times 20) = 3300$
170 - 180	$(170+180) \div 2 = 175$	5	$(175 \times 5) = 875$
180 - 190	$(180+190) \div 2 = 185$	3	$(185 \times 3) = 1555$
Total		40	6590

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{n} = \frac{6590}{40}$$

$$= 164.75 \quad \#$$

Subject:

No:
Date:

b) Median

class interval	class boundary	frequency	cumulative frequency (CF)	
150 - 160	149.5 - 159.5	12	12	
160 - 170	159.5 - 169.5	20	32	← median class
170 - 180	169.5 - 179.5	5	37	
180 - 190	179.5 - 189.5	3	40	
		40		

$L = 159.5$

$N = 40$, $\frac{N}{2} = 20$

$Cf_p = 12$

$W = (159.5 - 169.5) = 10$

$f_{med} = 20$

$$\text{median} = L + \frac{\frac{N}{2} - Cf_p}{f_{med}} (W)$$

$$= 159.5 + \left(\frac{20 - 12}{20} \right) \times 10$$

$$= 163.5 \; \#$$

Subject:

No:

Date:

c. Mode

Class Interval	frequency	
150 - 160	12	
160 - 170	20	← modal class
170 - 180	5	
180 - 190	3	

$L = 160$

$h = (160 - 170) = 10$

$f_0 = 12$

$f_1 = 20$

$f_2 = 5$

$$\text{mode} = 160 + \left(\frac{20 - 12}{(2 \times 20) - 12 - 5} \right) \times 10$$

$$= 160 + \left(\frac{8}{23} \right) \times 10$$

$$= 160 + 3.478$$

$$= 163.478 \quad \#$$

d) Modal class

$= 160 \leq x < 170$ #

Assignment 2 Aswhi No. Date

Question 2

a) mean = $\frac{85+90+75+88+92+80+85+82+90+85}{10} = \frac{852}{10}$

$= 85.2$

median = 75, 80, 82, 85, 85, 85, 88, 90, 90, 92.

$= \frac{85+85}{2}$

$= 85$.

Mode = 85 (frequency 3).

b) The mean, median and mode are values which nearly same. Average students scores 85.2. According to median most of the students scored above 85 and another half below 85. Based on the mode 85, most numbers of students scores 85.

In this case, the mean 85.2 would best describe the measure to use when reporting the overall summary of the participation scores since it takes all the points into account and provides an average value.

c) 1. mean = $\frac{55+65+65+70+85+95+95+95+100+100}{10} = \frac{825}{10}$

$= 82.5$

median = 55, 65, 65, 70, 85, 95, 95, 95, 100, 100.

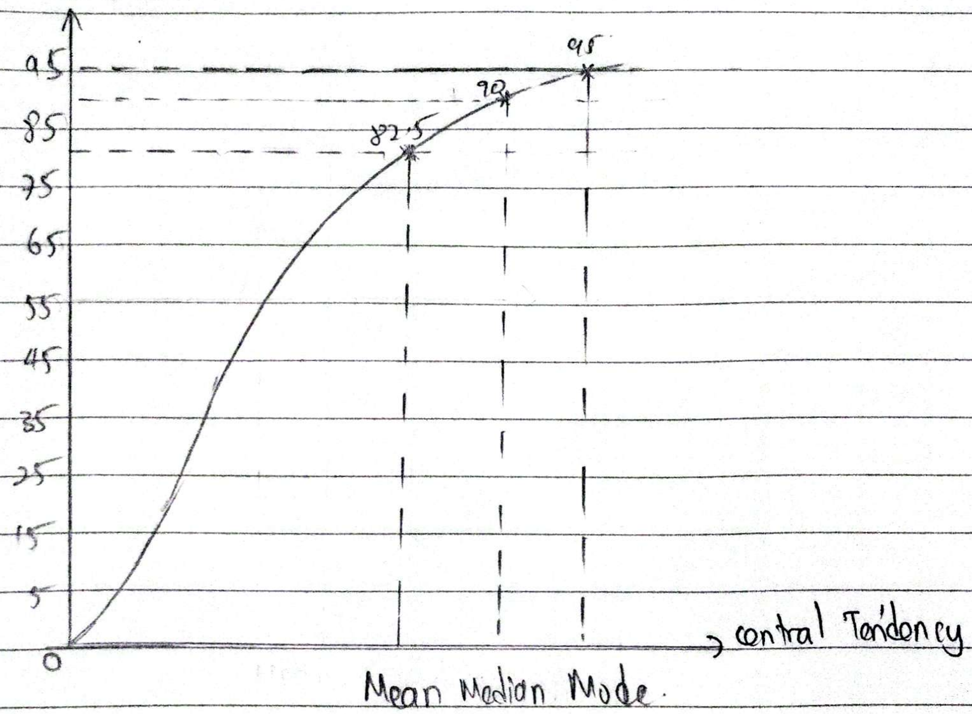
$= \frac{85+95}{2} = 90$.

Mode = 95 (frequency 3).

Assignment 2 Asuini No. Date

Question 2

c) ii. Scores value.



iii.

	Original data	Corrected scores.
Mean	85.2	82.5
Median	85	90
Mode.	85.	95

- Mean dropped from 85.2 to 82.5 indicating students participation was lower.
- median increased from 85 to 90 indicating half of the students scores above 90.
- Mode increased from 85 to 95 indicating that most numbers of students scored 95.

Assignment 2

Question 3

a) i. Range. Highest sales − lowes sales.

= 40,000 − 25,000

= 15,000.

ii. Variance.

$$\frac{(\text{value} - \text{Mean})^2}{12 - 1 \rightarrow (n-1)}$$

$$= \frac{25+30+28+35+32+27+40+38+33+36+31+29}{12} = \frac{384}{12}$$

$$= \text{Thousands} \rightarrow \frac{384{,}000}{12}$$

= 32,000 Mean.

$(x - 32{,}000)^2$

$(25{,}000 - 32{,}000)^2 = 49{,}000{,}000$

$(30{,}000 - 32{,}000)^2 = 4{,}000{,}000$

$(28{,}000 - 32{,}000)^2 = 16{,}000{,}000$

$(35{,}000 - 32{,}000)^2 = 9{,}000{,}000$

$(32{,}000 - 32{,}000)^2 = 0$

$(27{,}000 - 32{,}000)^2 = 25{,}000{,}000$

$(40{,}000 - 32{,}000)^2 = 64{,}000{,}000$

$(38{,}000 - 32{,}000)^2 = 36{,}000{,}000$

$(33{,}000 - 32{,}000)^2 = 1{,}000{,}000$

$(36{,}000 - 32{,}000)^2 = 16{,}000{,}000$

$(31{,}000 - 32{,}000)^2 = 1{,}000{,}000$

$(29{,}000 - 32{,}000)^2 = 9{,}000{,}000$.

~~49,000 + 4,000 + 16,000 + 9000 + 0 + 25,000 + 64,000 + 36,000 + 1,000 + 16,000 + 1,00~~
~~9,000~~

Assignmen 2

Question 3

a ii. Variance.

= 49,000,000 + 4,000,000 + 16,000,000 + 9,000,000 + 0 + 25,000,000 + 64,000,000 + 36,000,000 + 1,000,000 + 16,000,000 + 1,000,000 + 9,000,000

= 230,000,000.

$$\frac{230{,}000{,}000}{11} = 20909090.91$$

iii. standard deviation.

$= \sqrt{20909090.91}$

$= 4572.646.$

b) - Range indicates the monthly sales diffrent between the highest and lowest sales of the months.

- variance (20909090.91) and the standard deviation (4572.646) indiates that sales figures fluctuate around average of ~~32~~ RM32,000, with most months falling within approximately RM 4572.646. above or below average.

c) Financial Management:

- Budgeting: Get realistic forecasts from sales to more effectively manage cashflow.
- Contigency Planning: set aside money or make short term financing arrangments for slow months.

Operations Management:

- Inventory: Balance inventory levels with sales trends, reducing costs and stockouts.
- Staffing: Staff according to predicted sales to control costs without compromising quality of service.

No. Date

Marketing and sales:

- Promotions: Run promotions during slow sales months to create added revenue.
- Resource Allocation: Concentrate marketing effort in high sales months for greater returns.

Question 4

- mean = 50
- standard deviation = 10
- productivity scores are normally distributed

a) what percentage of employees in the sample showed an increase in productivity after the training program?

= 50% of employees showed an increase (assuming increase or means above the mean

b) what is the probability of employees having a productivity score between 37 and 65 after the training?

formula : $Z = \frac{x - \mu}{\sigma}$

37 : $Z = \frac{37 - 50}{10} = -1.3$

65 : $Z = \frac{60 - 50}{10} = 1.5$

standard normal distribution

- $P(Z < 1.5) = 0.9332$
- $P(Z < -1.3) = 0.0968$

probility 37 and 65

- $P(37 < x < 65) = P(Z < 1.5) - P(Z < -1.3)$

$= 0.9332 - 0.0968 = 0.8364$

= probabitiy is 0.8364 / 83.64%

c) company have 1000 employees and want to send for team building if score < 20 unit. cost is RM 200 per person. estimate total cost.

Z for 20 : $Z = \frac{20 - 50}{10} = -3$

standard normal table :

- $P(Z < -3) = 0.0013$

calculate employees :

employees = 0.0013 x 1000 = 1.3 = 1 employee

calculate cost =

cost = 1 x 200 = RM200

= estimate total budget = RM200 x

d) if the company want to identity the top 5% of employees with the highest productivity scores after the training, what minimum score would an employee need achieve?

top 5% means the cutoff is at 95% cumulative from the left

find z-table, z = 1.645

$x = \mu + z\sigma$

$= 50 + (1.645)(10)$

$= 50 + 16.45$

$= 66.45$

= minimum score = 66.45 units to be in the top 5%.

Assignment 2 Asuini No. Date

Question 5

a) The random variable x represents the number of correct answers a student gets on the exam.

b) $P = \frac{1}{4}, q = 1 - \frac{1}{4} - \frac{3}{4}, n = 6, x = 0,1,2,3,4,5,6.$

$P(x=n)$: $P(x=0)\ {}^6C_0 \times \left(\frac{1}{4}\right)^0 \times \left(\frac{3}{4}\right)^6 = 0.178$

$P(x=1)\ {}^6C_1 \times \left(\frac{1}{4}\right)^1 \times \left(\frac{3}{4}\right)^5 = 0.356$

$P(x=2)\ {}^6C_2 \times \left(\frac{1}{4}\right)^2 \times \left(\frac{3}{4}\right)^4 = 0.297$

$P(x=3)\ {}^6C_3 \times \left(\frac{1}{4}\right)^3 \times \left(\frac{3}{4}\right)^3 = 0.132$

$P(x=4)\ {}^6C_4 \times \left(\frac{1}{4}\right)^4 \times \left(\frac{3}{4}\right)^2 = 0.033$

$P(x=5)\ {}^6C_5 \times \left(\frac{1}{4}\right)^5 \times \left(\frac{3}{4}\right)^1 = 0.004$

$P(x=6)\ 6C_6 \times \left(\frac{1}{4}\right)^6 \times \left(\frac{3}{4}\right)^0 = 0.0002 = 0$

Table

x	0	1	2	3	4	5	6
P(x=n)	0.178	0.356	0.297	0.132	0.033	0.004	0

c) $\mu = np \quad 6\left(\frac{1}{4}\right) = 1.5$

d) $P(x \geq 3)$

$= 0.132 + 0.033 + 0.004 + 0$

$= 0.169$

e) $P = \frac{3}{4} \quad q = \frac{1}{4} \quad n = 4$

$= {}^{4-1}C_{1-1} \times \left(\frac{3}{4}\right)^1 \times \left(\frac{1}{4}\right)^3 = 0.01172$

$= 0.012$

Assignmont 2 Asuni. No. Date

Question 6.

a) The probability distribution of x is a negative binomial distribution.
$r = 4, p = 0.7$
r (cappucino orders), p (probability of order cappucino)

b) $\frac{r(1-p)}{p^2} = \frac{4(1-0.7)}{(0.7)^2}$

$= \frac{1.2}{0.49} = 2.449.$

standard deviation $= \sqrt{2.449} = 1.564.$

c) $x = 6, r = 4, p = 0.7$
$P(x=6)$
$= \left(\frac{6-1}{4-1}\right)(0.7)^4 (1-0.7)^2$
$= {}^5C_2 \times (0.7)^4 \times (0.3)^2$
$= 0.216.$

d) $n = 12, k = 7, p = 0.7.$ $\left[\binom{n}{x} p^x (1-p)^{n-x}\right]$

$P(k=7) = {}^{12}C_7 \times (0.7)^7 \times (1-0.7)^{12-7}.$
$= 0.159$