FACULTY OF COMPUTING

SECI1143 PROBABILITY & STATISTICAL DATA ANALYSIS

SEMESTER 2024/2025

LECTURER'S NAME

DR. SEAH CHOON SEN

ASSIGNMENT 2

NAME	MATRIC ID
NUR AMIRAH HUDA BINTI AMIRRUDDIN	SX242451ECRHF01
SYAHIDATUL NAJWA BINTI ZAMRI	SX241889ECJHF01
CHE KU MARDIAH BINTI CHE KU MAHHADI	SX241899ECJHF01
SABARIAH BINTI ALIAS	SX230889ECJHS01
DANISH NURHAKIMI BIN ZULKEFLY	SX241897ECJHF01
MUHAMMAD NUR AIMAN BIN AZLEY	SX241892ECJHF01

No.: Date:

Assignment 2 : Chapter 3 and Chapter 4

QUESTION 1

Table 1 shows data collected on the heights (in cm) of 60 participants from a Youth Basketball Training Camp.

Table 1

Class Interval	Frequency
$150 \leq x < 160$	12
$160 \leq x < 170$	20
$170 \leq x < 180$	5
$180 \leq x < 190$	3

Based on the above value, calculate the :

a) Mean

Class Interval	Midpoint, x_i	Frequency, f_i	$f_i x_i$
150 - 159	$(150+159)\div 2 = 154.5$	12	$154.5 \times 12 = 1854$
160 - 169	$(160+169)\div 2 = 164.5$	20	$164.5 \times 20 = 3290$
170 - 179	$(170+179)\div 2 = 174.5$	5	$174.5 \times 5 = 872.5$
180 - 189	$(180+189)\div 2 = 184.5$	3	$184.5 \times 3 = 553.5$
Total		40 (n)	6570

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{n} = \frac{6570}{40} = 164.25$$

No.: .... Date: ....

b) Median

Class Interval	Frequency	Cumulative frequency
150 - 159	12	12
160 - 169	20	32
170 - 179	5	37
180 - 189	3	40
Total	40 (N)	

$N \div 2 = 40 \div 2 = 20$

$\therefore$ median class = 160 - 169

$L = 159.5 \quad N = 40 \quad cf_p = 12$

$W = (169.5 - 159.5) = 10 \quad f_{med} = 20$

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

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

$$= 163.5$$

c) Mode

$\therefore$ modal class = 160 - 169

$l = 160 - 0.5 = 159.5 \quad , \quad h = 169 - 160 = 9$

$f_1 = 20 \quad , \quad f_0 = 12 \quad , \quad f_2 = 5$

$$\text{Mode} = l + \left[h \times \frac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)}\right]$$

$$= 159.5 + \left[9 \times \frac{(20-12)}{(2 \times 20 - 12 - 5)}\right]$$

$$= 159.5 + \left[9 \times \frac{12}{23}\right]$$

$$= 159.5 + 4\frac{16}{23}$$

$$= 164.20$$

d) Modal class = 160 - 169

No.: Date:

Question 2

a) 10 club members

85, 90, 75, 88, 92, 80, 85, 82, 90, 85

75, 80, 82, 85, 85 | 85, 88, 90, 90, 92

$$\text{Mean} = \frac{(85+90+75+88+92+80+85+82+90+85)}{10}$$

$$= \frac{852}{10}$$

$$= 85.2$$

$$\text{Median} = (85+85) \div 2$$

$$= 170 \div 2$$

$$= 85$$

Mode = 85

The mean is 85.2, median is 85 and mode is 85 of the participation scores.

b)
- Mean is 85.2 where it show the average student performance.
- Median is 85 where it is indicates the middle score of students. This is meaning that half of the students scored above and half below of 85.
- Mode is 85 where it shows that the most common participation score.
- Statistic that more appropriate represent the overall summary of the participation scores is mean, Mean is 85.2.

c) 55, 65, 65, 70, 85, 95, 95, 95, 100, 100

i) $$\text{Mean} = \frac{(55+65+65+70+85+95+95+95+100+100)}{10}$$

$$= \frac{825}{10}$$

$$= 82.5$$

No.: Date:

c) i) Median = $(85 + 95) \div 2$

$= 180 \div 2$

$= 90$

Mode = 95

iii) The corrected scores :

Mode = 95 , Median = 90 , Mean = 82.5

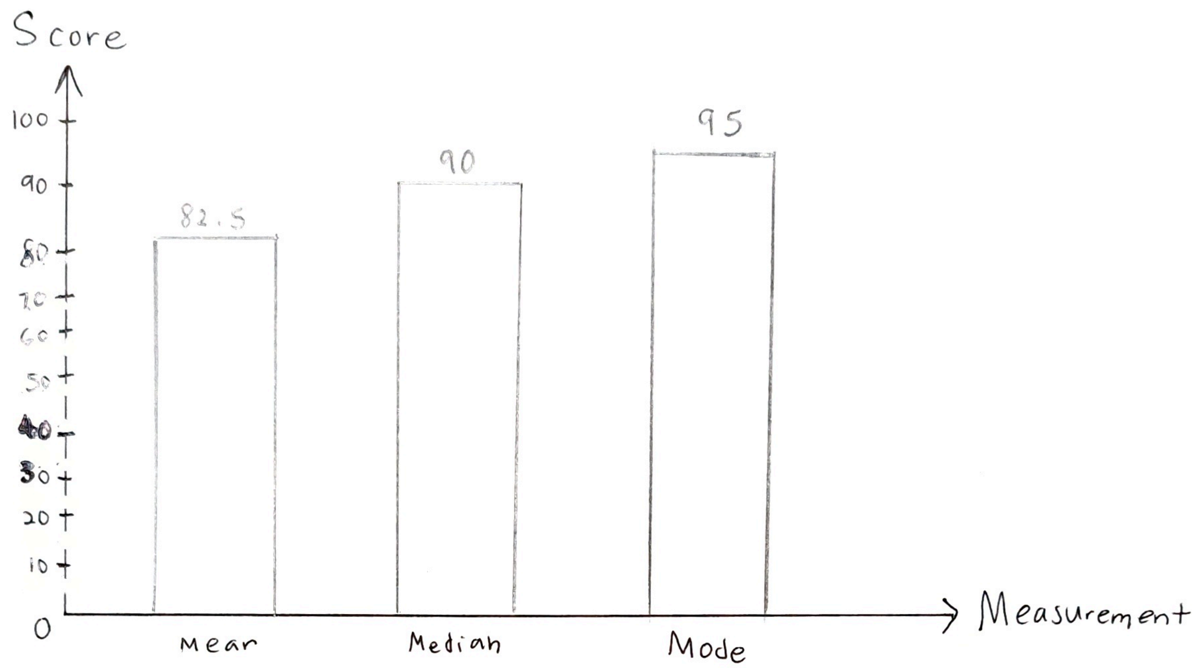
The original data :

Mode = 85 , Median = 85, Mean = 85.2

- The corrected data shows more variation of participation score which is from 55 to 100.
- The original data only from 75 to 92.
- The mean is drop in corrected score compared to original data.
- The mode and median is increase in corrected score compared to orignal data.

c) ii) Plot the Graph of Central Tendency

Central Tendency of Corrected Participation Scores



No.: Date:

Question 3

a) i) Range $= \text{Max} - \text{Min} = 40 - 25 = 15$

ii) Variance

$$\text{Mean} = \frac{\Sigma x}{n}$$

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

$$= 32$$

Compute squared deviations from the mean:

$$\sum (x_i - \bar{x})^2$$

$$= (25-32)^2 + (30-32)^2 + (28-32)^2 + (35-32)^2 + (32-32)^2 + (27-32)^2 + (40-32)^2 + (38-32)^2 + (33-32)^2 + (36-32)^2 + (31-32)^2 + (29-32)^2$$

$$= 49 + 4 + 16 + 9 + 0 + 25 + 64 + 36 + 1 + 16 + 1 + 9$$

$$= 230$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

$$= \frac{230}{12}$$

$$= 19.17$$

iii) Standard deviation

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

$$= \sqrt{19.17}$$

$$= 4.38$$

No.: .... Date: ....

b) Throughout the year, the average monthly sales were RM32,000. The range of sales was RM15,000 with a standard deviation of RM4380. This indicates that monthly sales typically deviate around RM 4380 from the mean. Although the variation is not extreme, it reflects moderate fluctuations from month to month. The highest sales were recorded in July at RM40,000 while the lowest occured in January at RM25,000. This incosistency suggests that sales performance was not stable throughout the year. The monthly variations highlight the need to anticipate and adapt to both peak and off-peak periods in sales performance.

c) ① Marketing and Promotion Timing

→ Special offers and advertisement can be used in low sales months such as in January or June to help boost sales and keep income more steady.

② Cash Flow Planning

→ By knowing which months have lower sales, the business can save enough money to pay for regular expenses during slower times and avoid running out of cash.

③ Inventory and Staffing Adjustments

→ In months with higher sales, the business can stock up on more products and add staff hours. In slower months, it can reduce stock and staff hours to save on costs.

No:

Date: ....

Question 4.

a. mean = 50

standard deviation = 50

= Percentage of employees with increased productivity after training is 50 %

b. i) $z = \frac{x - \mu}{\sigma}$

$x = 37$ $z = \frac{37-50}{10} = -1.3$ $P(z < -1.3) = 0.0968$

$x = 65$ $z = \frac{65-50}{10} = 1.5$ $P(z < 1.5) = 0.9332$

ii. $P(37 < x\ 65)$ $0.9332 - 0.0968 = 0.8364$

c. mean = 50

Standard deviation = 10

Score = 20

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

$z = \frac{20-50}{10} = -3.0 \approx 0.0013$

0.13% of employees scored below 20

$0.0013 \times 1000 = 1.3$ employees ≈ 1 employee

$1 \times RM200 = RM200$

= 0.13% of employees scored below 20 units, which ~~quater~~ is 1 employee out of 1000. The total cost for the them building in program is RM200.

No:

Date: ....

d. mean $= 50$ unit.

standard deviation $= 10$ unit

minimun score $=$ Top 5%.

95% lower

$P(Z < z) = 0.9500 \rightarrow z = 1.645$

$x = \mu + z\sigma = 50 + (1.645)(10) = 50 + 16.45 = 66.45$

= ~~Minimum score untuk berada dalam top~~

= Minimun score for stay in top 5% is 66.45 unit.

No:

Date: ....

Question 5.

a. Identify the random variable X.

= x is a 6 multiple - choice questions

b. Construct up the table for the probability of x

$p = \frac{1}{4} = 0.25$ (true)

$1 - p = 1 - 0.25 = 0.75$ (false)

Formula = $b(x; n, p) = \binom{n}{x} p^x q^{n-x}$, $x = 0, 1, 2, \dots n$

$P(X = x) = \binom{6}{x} (0.25)^x (0.75)^{6-x}$

x	Formula	$P(X = x)$
0	$\binom{6}{0}(0.25)^0(0.75)^6$	0.1780
1	$\binom{6}{1}(0.25)^1(0.75)^5$	0.3560
2	$\binom{6}{2}(0.25)^2(0.75)^4$	0.2966
3	$\binom{6}{3}(0.25)^3(0.75)^3$	0.1318
4	$\binom{6}{4}(0.25)^4(0.75)^2$	0.0329
5	$\binom{6}{5}(0.25)^5(0.75)^1$	0.0044
6	$\binom{6}{6}(0.25)^6(0.75)^0$	0.0002

c. Mean of the distribution.

= Mean of binominal distribution

$\mu = n \times p$

$= 6 \times 0.25$

$= 1.5$

d. Probability of ~~getting~~ a student getting at least 3 correct answers.

= At least 3 correct answers means $x = 3, 4, 5, 6$

$P(x \geq 3) = P(3) + P(4) + P(5) + P(6)$

$= 0.1318 + 0.0329 + 0.0044 + 0.0002 = 0.1693$

e. Probability that the first incorrect answer occurs in the fourth question is: (geometric distribution)

$\cancel{1,2,3 \; (\overset{0.25}{0.75})}$ $\cancel{P = (0.25)^3 \times (0.75) = 0.0117}$

$\cancel{4 \; (\overset{0.75}{0.25})}$

$\cancel{(0.73)^3 \times (0.25)}$

$\cancel{=}$

$$b^*\left(4; 1, \frac{1}{4}\right) = \binom{4-1}{1-1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^3$$

$$= \binom{3}{0}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^3$$

$$= 0.1054$$

No:

Date: ....

Question 6 -

a. What is the probability distribution of X ?

$r = 4, p = 0.70$ (negoti binominal)

b. What the standard deviation of this distribution ?

$$\sigma_x = \sqrt{\frac{(1-p)}{p^2}}$$

$$= \sqrt{\frac{(1-0.70)}{0.70^2}}$$

$$= \sqrt{\frac{0.3}{0.49}} = 0.7825$$

c. Determine the probability that exactly 6 customers will arive until the 4th customer who orders a cappuccino.

$$g(x;p) = [(1-p)^{x-1}](p)$$

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{n-k}$$

$$b^*(6;4,0.7) = \binom{6-1}{4-1} (0.7)^4 (0.3)^3$$

$$= \binom{5}{3} (0.7)^4 (0.3)^2$$

$$= 0.2160$$

No:

Date: ....

a. The probability that exactly 7 cappuccinos will be ordered within the next 12 customer arrivals.

$$\cancel{b^*(x;k,p) = \binom{x-1}{k-1}(p^k)(q^{x-k})}$$

$$\cancel{b^*(12;7,0.7)}$$

$$\cancel{= \binom{11}{6}(0.7)^7(0.3)^5}$$

$$b(x;n,p) = \binom{n}{x} p^x q^{n-x}$$

$$b(12;7,0.7) = \binom{12}{7}(0.7)^7(0.3)^5$$

$$= 0.1584$$