



SECI2143 – PROBABILITY AND STATISTICAL DATA ANALYSIS

ASSIGNMENT 1

SECTION 02

DR CHAN WENG HOWE

MEMBERS :

NAME	MATRIC NO
LEE CAI XUAN	A20EC0062
CHLOE RACQUELMAE KENNEDY	A20EC0026
MYZA NAZIFA BINTI NAZRY	A20EC0219
AMIRAH RAIHANAH BINTI ABDUL RAHIM	A20EC0182

Question 1

- 1a) Football clubs in Malaysiana
- b) Malaysiana Super League matches
- c) Total number of goals scored
- d) Amount of time played before a goal is scored
- e) Mean number of goals per match on the first saturday of the season
- f) Mean number of goals per game in the previous season

- 2a) All students at a school in JayBie
- b) Second year pupils
- c) Place of birth, gender, age (in years)
- d) Height and weight
- e) Number of pupils weighed and the mean of the weights recorded
- f) Data on first year pupils

Question 2.

1. Summarize Table 1 in a frequency table as below.

Factors	Scales	Frequency				Total
		1	2	3	4	
Price		14	6	0	0	20
Condition of the car		3	6	11	0	20
Fuel Efficiency		1	11	7	1	20
Car Depreciation		6	11	2	1	20

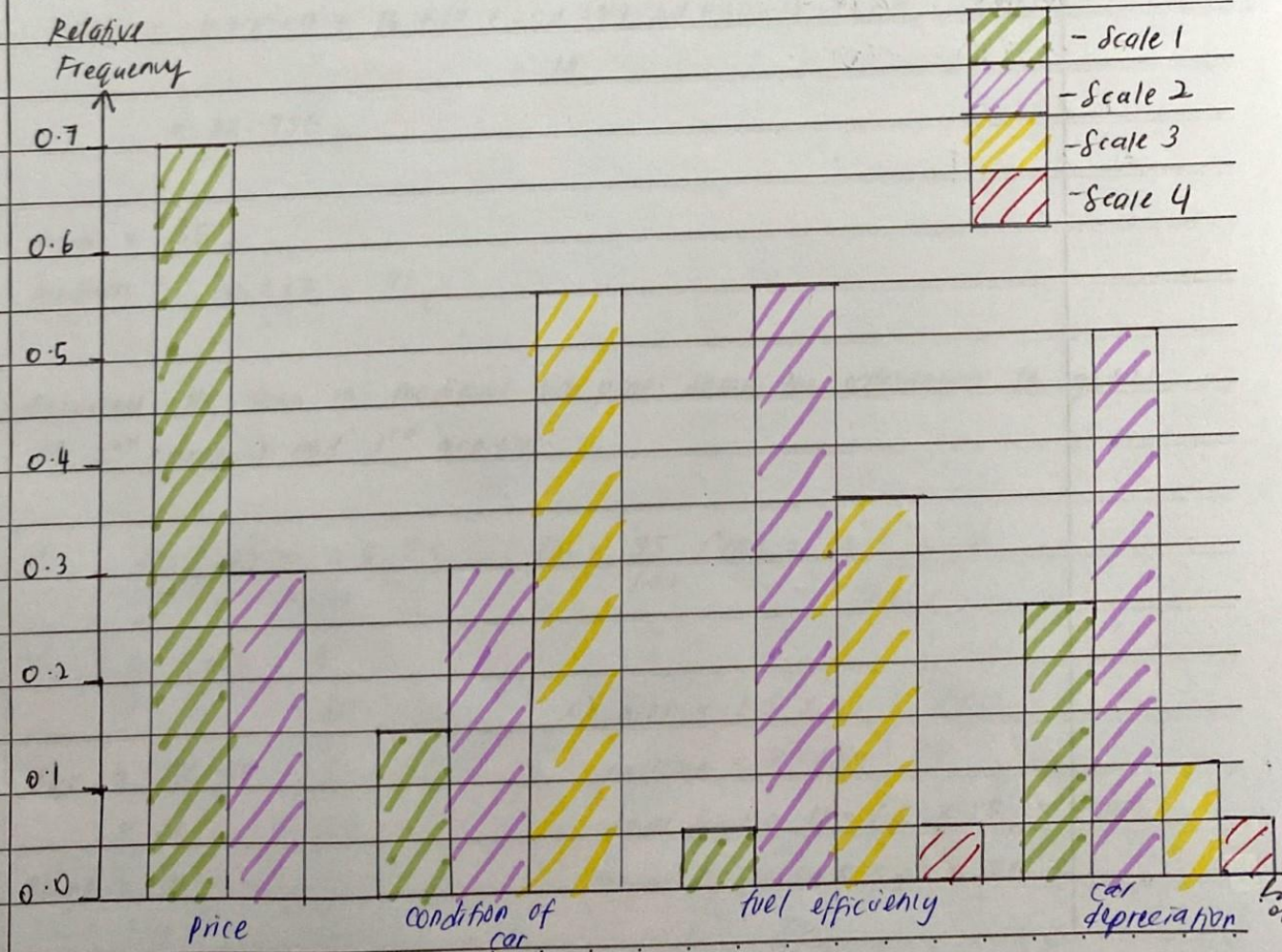
2. Summarize all the factors (Price, condition of car, fuel efficiency and car depreciation) from Table 1 in a frequency distribution. For each of the factor, use the table as below.

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	14	0.7	14	0.7
2	6	0.3	20	1.0
3	0	0.0	20	1.0
4	0	0.0	20	1.0
Total	20	1.0		

Scale	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
1	3	0.15	3	0.15
2	6	0.3	9	0.45
3	11	0.55	20	1.0
4	0	0.0	20	1.0
Total	20	1.0		

Scale	Frequency	Relative Frequency	Cumulative Frequency	Relative CF
1	1	0.05	1	0.05
2	11	0.55	12	0.60
3	7	0.35	19	0.95
4	1	0.05	20	1.0
Total	20	1.0		

Scale	Frequency	Relative Frequency	CF	Relative CF
1	6	0.3	6	0.3
2	11	0.55	17	0.85
3	2	0.1	19	0.95
4	1	0.05	20	1.0
Total	20	1.0		



Question 3

1) 4, 8, 9, 16, 18, 21, 23, 24, 30, 32, 33, 38, 42, 43, 44, 55, 65, 81

Number of days taken to finish developing mobile app for a competition

Stem | Leaf

0 | 4 8 9

1 | 6 8

2 | 1 3 4

3 | 0 2 3 8

4 | 2 3 4

5 | 5

6 | 5

8 | 1

Keys: 1|6 means 16

$$2a) \bar{x} = \frac{4 + 8 + 9 + 16 + 18 + 21 + 23 + 24 + 30 + 32 + 33 + 38 + 42 + 43 + 44 + 55 + 65 + 81}{18}$$

$$= \frac{586}{18}$$

$$= 32.5556$$

b) No mode. This is because there is no number repeats in the data.

$$c) i = \frac{50}{100} (18) \quad (Y[9] + Y[10]) \div 2 = (30 + 32) \div 2$$

$$= 9 \quad = 62 \div 2$$

$$= 31$$

Question 3

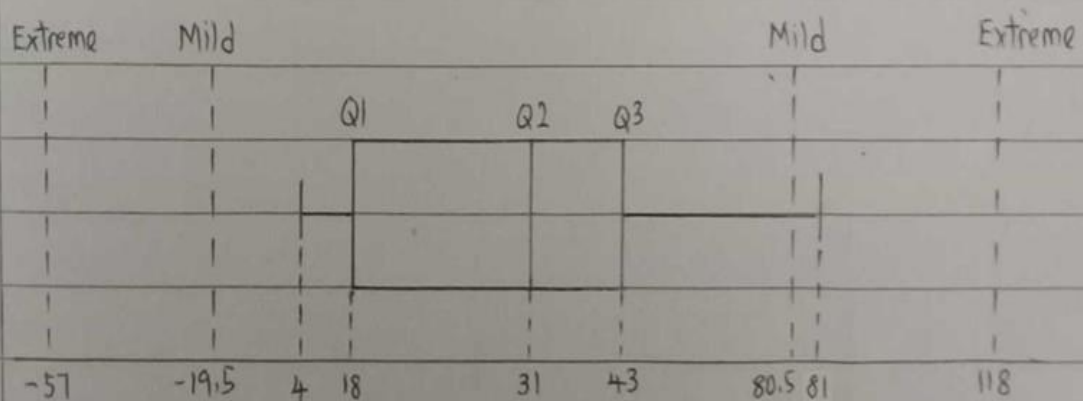
3a) $i = \frac{25}{100} (18)$ First quartile = 18 Median = 31
 $= 4.5$
 $= 5$

$i = \frac{75}{100} (18)$ Third quartile = 43
 $= 13.5$
 $= 14$

b) Interquartile range = $43 - 18$
 $= 25$

c) Mild = $Q3 + 1.5 \times IQR$ Extreme = $80.5 + 1.5 \times IQR$
 $= 43 + 1.5 \times 25$ $= 80.5 + 1.5 \times 25$
 $= 80.5$ $= 118$

Mild = $Q1 - 1.5 \times IQR$ Extreme = $-19.5 - 1.5 \times IQR$
 $= 18 - 1.5 \times 25$ $= -19.5 - 1.5 \times 25$
 $= -19.5$ $= -57$



QUESTION 4

The value of 11 houses in a new development area in Secudai are shown in Table 2.

Value per house	Number of houses
RM 175 000	1
RM 250 000	5
RM 600 000	4
RM 700 000	1

Table 2

Based on the table :

1. Find the mean value of these houses in RM.

$$\begin{aligned} &\text{Mean value of houses} \\ &= \frac{(RM175\,000 \times 1) + (RM250\,000 \times 5) + (RM600\,000 \times 4) + (RM700\,000 \times 1)}{11} \end{aligned}$$

$$= RM\,375\,000$$

2. Find the median value of these houses in RM.

$$\begin{aligned} \frac{n}{2} &= 5.5 \\ &\approx 6 \end{aligned}$$

RM175 000, RM 250 000, RM 250 000, RM 250 000, RM 250 000, RM 250 000, RM 600 000, RM 600 000, RM 600 000, RM 600 000, RM 700 000

RM 600 000, RM 600 000, RM 700 000

$$\therefore \text{Median} = RM250\,000$$

3. State which measures the central tendency, the mean or the median, "best" represents the values of the 11 houses. Justify your answer.

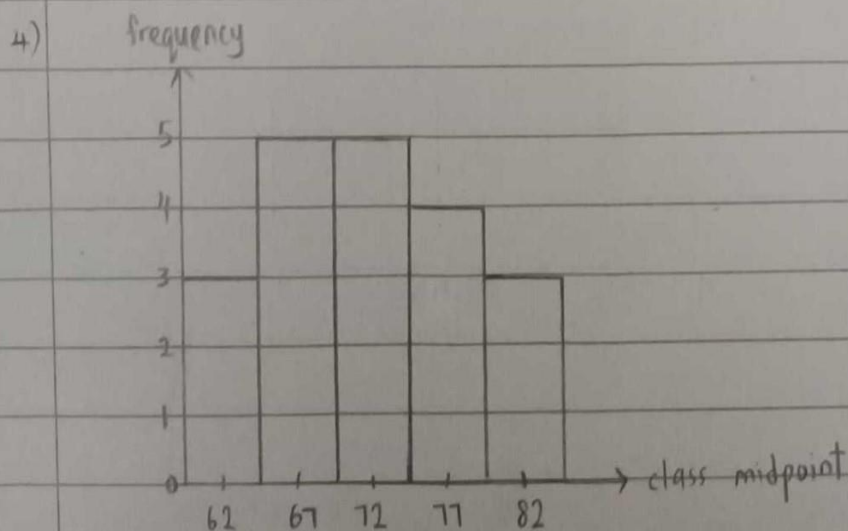
Mean because it uses all the values in the data/table.

Question 5

1)	Age	Ratio
	Current smoker	Nominal
	Body Mass Index (BMI)	Interval
	Hypertension	Nominal

2)	Class Interval	Category	Frequency
	$0 < \text{BMI} < 18.5$	Under weight	0
	$18.5 \leq \text{BMI} < 25.0$	Normal	12
	$25.0 \leq \text{BMI} < 30.0$	Overweight	7
	$30.0 \leq \text{BMI}$	obesity	1

3)	Class Interval	class Boundaries	Class midpoint	Frequency	Cumulative frequency
	60-64	59.5 - 64.5	62	3	3
	65-69	64.5 - 69.5	67	5	8
	70-74	69.5 - 74.5	72	5	13
	75-79	74.5 - 79.5	77	4	17
	80-84	79.5 - 84.5	82	3	20
	Total			20	



QUESTION 6

A flight delay is when an airline flight takes off and/or lands later than its scheduled time. The Federal Aviation Administration (FAA) considers a flight to be delayed when it is 15 minutes later than its scheduled time. The following distribution shows a sample of 100 flights with delay times (rounded to the nearest minute) recorded in ABC Airport.

Time (minutes)	Frequencies
16 - 30	3
31 - 45	13
46 - 60	30
61 - 75	25
76 - 90	14
91 - 105	8
106 - 120	4
121 - 135	2
136 - 150	1

Table 4 : Flight Delay Time

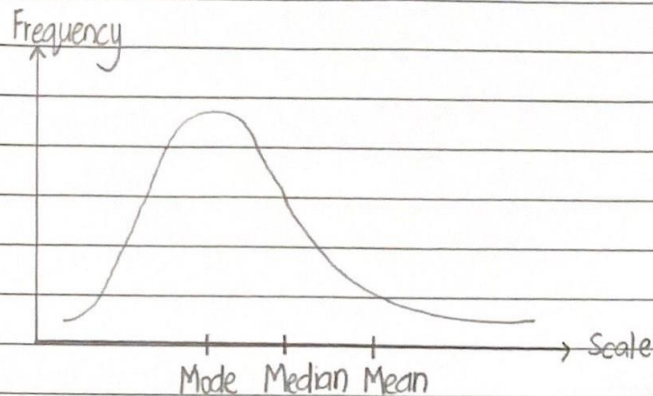
1. Find the median, mean and mode for the values above. Show the calculation to get each answer.

Time (minutes)	Frequency	Midpoint (x)	Cumulative frequency	$f(x)$	
16 - 30	3	23	3	69	
31 - 45	13	38	16	494	
46 - 60	30	53	46	1590	= mode
61 - 75	25	68	71	1700	= median
76 - 90	14	83	85	1162	
91 - 105	8	98	93	784	
106 - 120	4	113	97	452	
121 - 135	2	128	99	256	
136 - 150	1	143	100	143	
Total	100			6650	

Mean	$\frac{100}{2}$	Median		Mode
$= \frac{6650}{100}$	$= 50$	$= 60.5 + \frac{100/2 - 46}{25} (15)$	$= 45.5 + 15 \times \frac{30 - 13}{2(30) - 13 - 25}$	
$= 66.5$		$= 62.9$	$= 57.09$	

2. Based on your answer in (1):

(a) Draw the distribution graph.



(b) What can you tell on the shape of the distribution (skewness / kurtosis)?
It is right-skewed and positively-skewed.

(c) What does this tell you about the data?
From the data, it tells us that most flights take longer delay time.