



---

**SECI2143**

**PROBABILITY AND STATISTICAL DATA ANALYSIS**

---

**SEMESTER 2 2020/2021**

**ASSIGNMENT 1 :**

**LECTURER :**

**DR SHARIN HAZLIN HUSPI**

**SUBMITTED BY:**

**INDRADEVI A/P VIKNESHWARAN (A20EC0050)**

**KESHINIY A/P MOGAN (A20EC0057)**

**ROSHINI A/P THAVAN NAIR (A20EC0139)**

## Question 1

1. a. Football clubs in Malaysia
- b. Malaysia Super League matches in the season
- c. Total number of goals scored
- d. Amount of time played before a goal is scored
- e. Number of goals scored in each match on the first Saturday of season

Mean number of goals per match as 2.97

- f. Mean number of goals per game in last season was 3.08

2. a. All students at a school in Jaybir.
- b. Second year pupils
- c. Number of people weighed and age.
- d. Height and weight
- e. Weight of second year pupils
- f. Summary data of first year pupils.

summary data: place of birth, gender, age (in years), height and weight.

## Question 2

1.

Factors	Scales				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. Price

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	-	-

Condition of the Car

scale	frequency	relative frequency	cumulative frequency	cumulative relative frequency
1	3	0.15	3	0.15
2	6	0.30	9	0.45
3	11	0.55	20	1.00
4	0	0.00	20	1.00
total	20	1.00	-	-

Fuel Efficiency

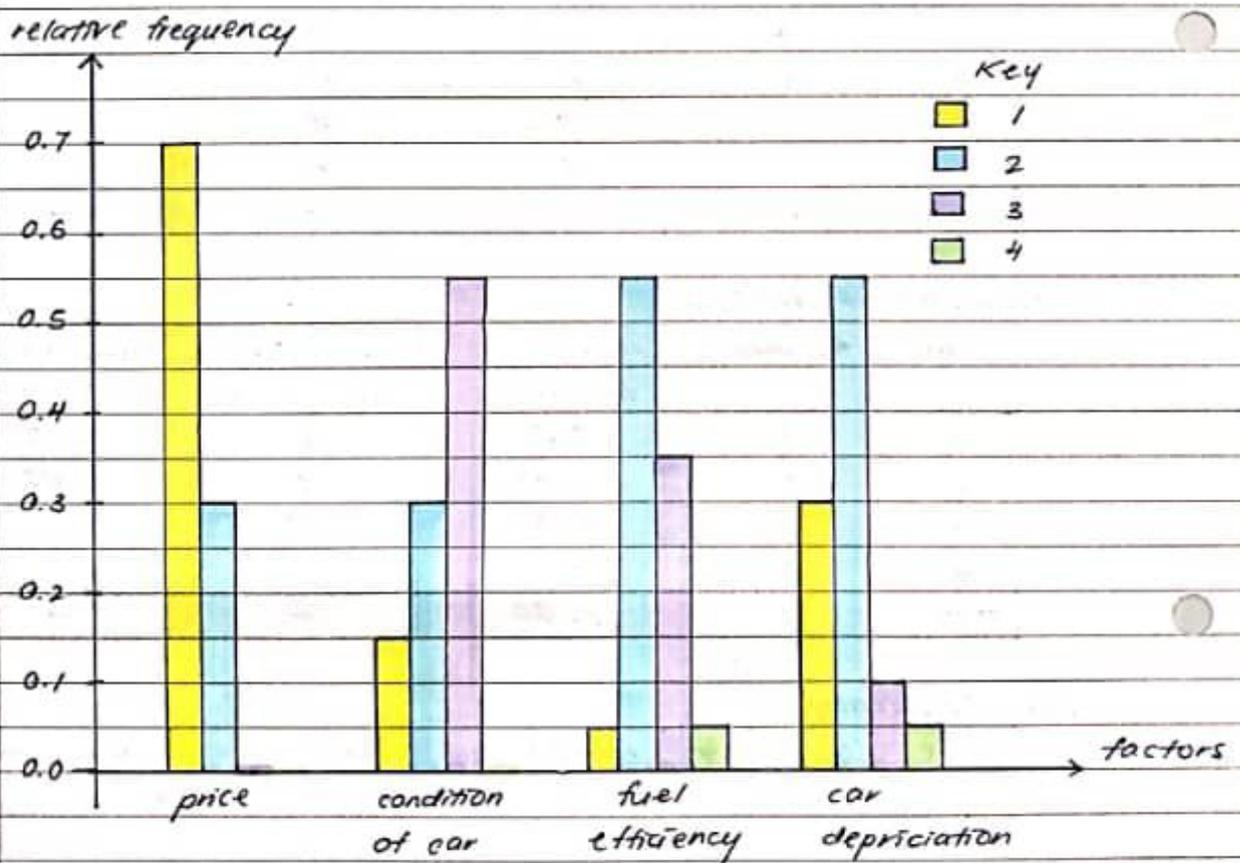
scale	frequency	relative frequency	cumulative frequency	cumulative relative frequency
1	1	0.05	6	0.30
2	11	0.55	17	0.85
3	7	0.35	19	0.95
4	1	0.05	20	1.00
total	20	1.00	-	-

## Car Depreciation

scale	frequency	relative frequency	cumulative frequency	cumulative relative frequency
1	6	0.30	6	0.30
2	11	0.55	17	0.85
3	2	0.10	19	0.95
4	1	0.05	20	1.00
total	20	1.00	-	-

3.

## Graph of Survey on Car Buying Factors



3) 30, 43, 32, 21, 65, 8, 4, 18, 16, 38, 9, 44, 33, 23, 24, 81, 42, 55  
 → 4, 8, 9, 16, 18, 21, 23, 24, 30, 32, 33, 38, 42, 43, 44, 55, 65, 81

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
7	
8	1

Stem = tens Leaf = Ones, Key: 0|4 = 04

2)

$$a) \text{Mean} = \frac{\sum_{i=1}^n x_i}{n}$$

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

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

$$= 32.56 *$$

b) Mode = No mode \*

$$c) \text{Median} = \frac{18}{2} = 9^{\text{th}}$$

$$= \frac{30+32}{2}$$

$$= 31 *$$

3

$$\begin{aligned} a) 1st &= \frac{1}{4} \times 18 \\ &= 4.5 \approx 5^{th} \\ Q_1 &= 18 \end{aligned}$$

$$\begin{aligned} 2nd &= \frac{1}{2} \times 18 \\ &= 9^{th} \end{aligned}$$

$$Q_2 = \frac{30 + 32}{2}$$

$$Q_2 = 31$$

$$\begin{aligned} 3rd &= \frac{3}{4} \times 18 \\ &= 13.5 \approx 14^{th} \end{aligned}$$

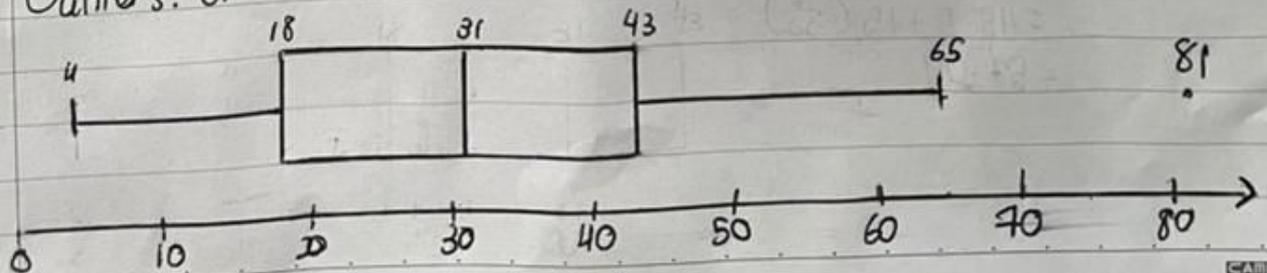
$$Q_3 = 43$$

$$\begin{aligned} b) IQR &= Q_3 - Q_1 \\ &= 43 - 18 \\ &= 25 \end{aligned}$$

$$\begin{aligned} c) \text{Lower Limit} &= Q_1 - 1.5 IQR \\ &= 18 - (1.5)(25) \\ &= -19.5 \end{aligned}$$

$$\begin{aligned} \text{Upper Limit} &= Q_3 + 1.5 IQR \\ &= 43 + (1.5)(25) \\ &= 80.5 \end{aligned}$$

Outliers: 81



Question 4

$$1) \text{ mean, } \bar{x} = \frac{(1 \times 175\,000) + (5 \times 250\,000) + (4 \times 500\,000) + (1 \times 700\,000)}{11}$$

$$= \frac{4\,125\,000}{11}$$

$$= \text{Rm } 375\,000$$

$$2) \text{ median} = \frac{50}{100} \times 11$$

$$= 5.5 \approx 6$$

$$Y[6] = \text{Rm } 250\,000$$

3) mean measures the central tendency and best represents the values of 11 houses because it uses all values in the data set and give an average value of these houses.

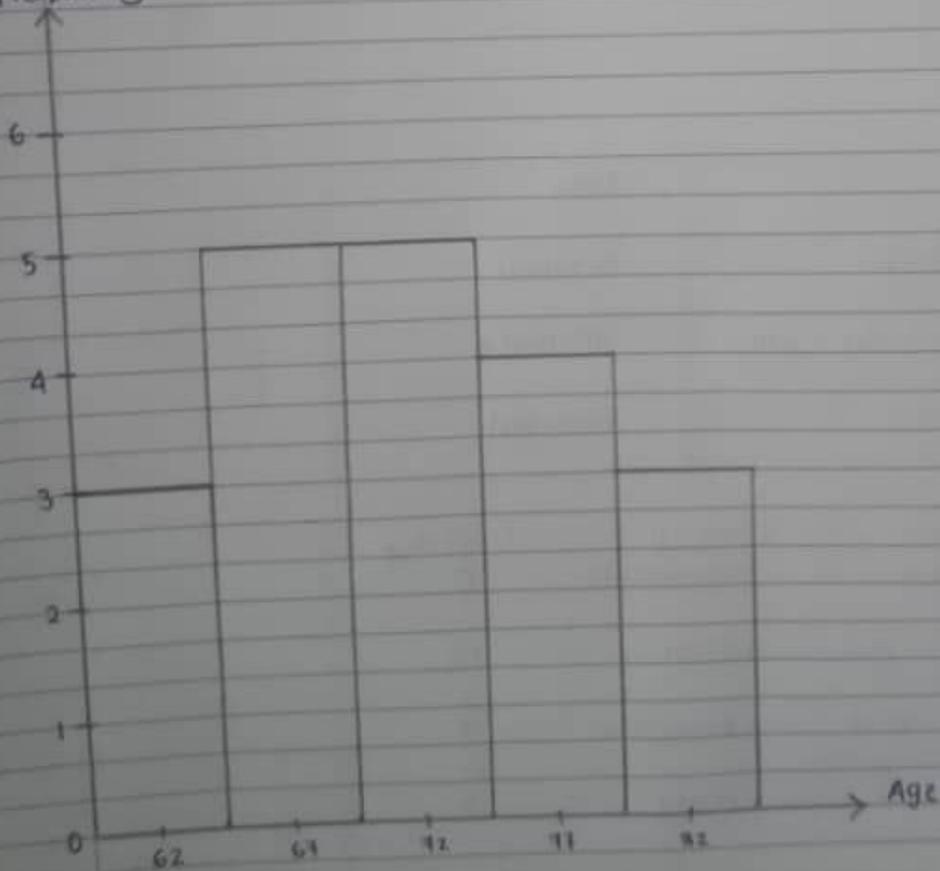
Question 5

1)	Age	ratio
	Current smoker	nominal
	Body mass index (BMI)	interval
	Hypertension	nominal

2)	Class Interval	Category	Frequency
	$0 < \text{BMI} < 18.5$	Underweight	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 $f$	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	

Frequency



6)

	Class Boundaries	Midpoint, $m$	Frequency, $f$	Cumulative Frequency	Relative Frequency	$f_m$
	15.5 - 30.5	23	3	3	0.03	69
	30.5 - 45.5	38	13	16	0.13	494
Mode $\rightarrow$	45.5 - 60.5	53	30	46	0.30	1590
Median $\rightarrow$	60.5 - 75.5	68	25	71	0.25	1700
	75.5 - 90.5	83	14	85	0.14	1162
	90.5 - 105.5	98	8	93	0.08	784
	105.5 - 120.5	113	4	97	0.04	452
	120.5 - 135.5	128	2	99	0.02	256
	135.5 - 150.5	143	1	100	0.01	143
	Total		100			6650

$$\text{Median} = \frac{100}{2} = 50^{\text{th}}$$

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

$$= 60.5 + \frac{100}{2} - 46}{25} (15)$$

$$= 62.9$$

$$\text{Mean} = \frac{\sum f_m}{\sum f}$$

$$= \frac{6650}{100}$$

$$= 66.5$$

$$\text{Mode} = l + h \times \left[ \frac{(f_1 - f_0) \div (2f_1 - f_0 - f_2)}{1} \right]$$

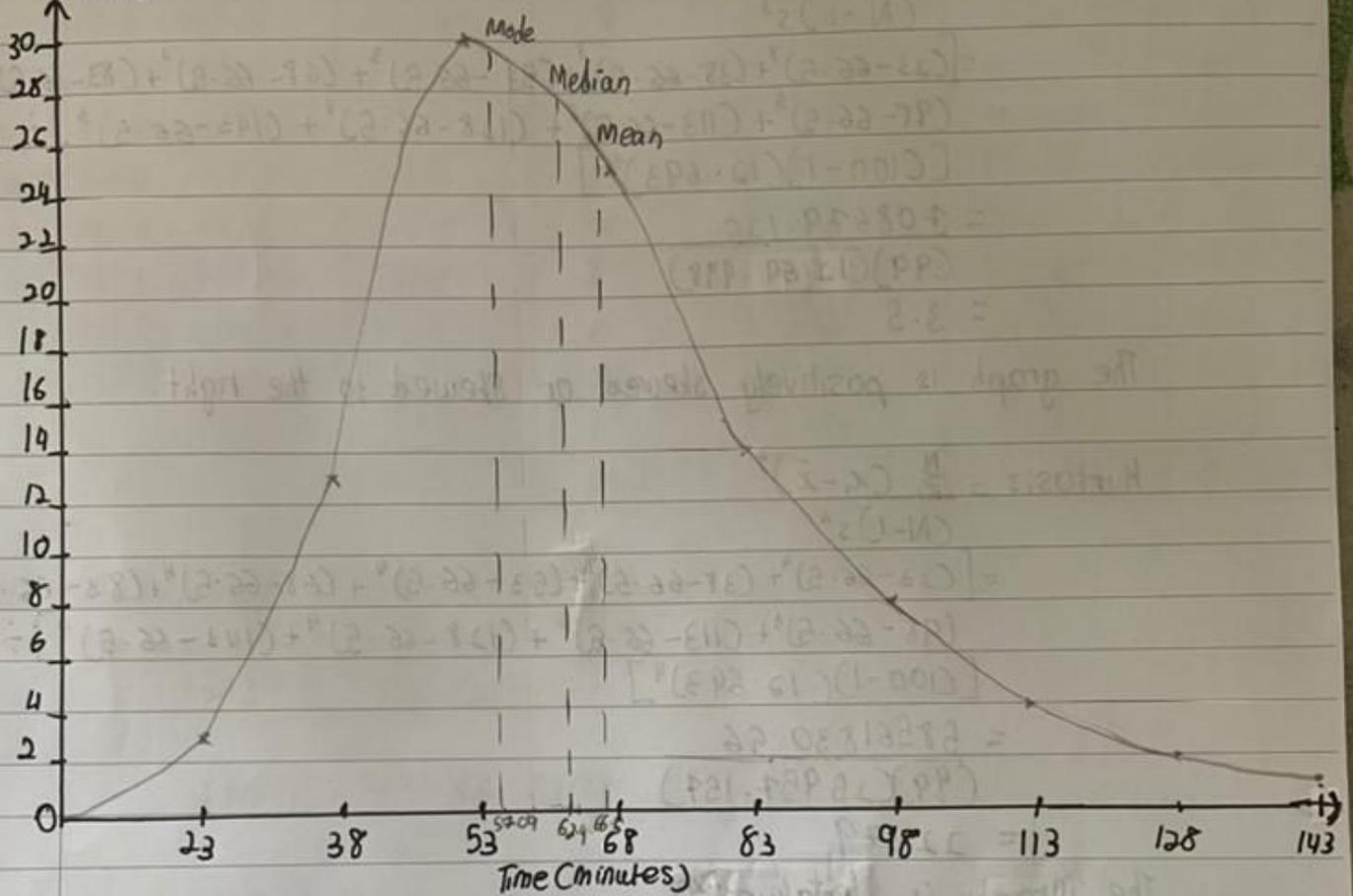
$$= 45.5 + 15 \left[ \frac{(30 - 13) \div (2(30) - 13 - 25)}{1} \right]$$

$$= 45.5 + 15 \left( \frac{17}{22} \right)$$

$$= 57.09$$

a)

Frequency



$$b) \text{ Variance, } s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

$$= \frac{[(23-66.5)^2 + (38-66.5)^2 + (53-66.5)^2 + (68-66.5)^2 + (83-66.5)^2 + (98-66.5)^2 + (113-66.5)^2 + (128-66.5)^2 + (143-66.5)^2]}{(100-1)}$$

$$= \frac{15950.25}{99}$$

$$99$$

$$= 161.114$$

$$\text{Std Deviation, } s = \sqrt{161.114}$$

$$= 12.693$$

$$\text{Skewness} = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{(N-1)s^3}$$

$$= \frac{[(23-66.5)^3 + (38-66.5)^3 + (53-66.5)^3 + (68-66.5)^3 + (83-66.5)^3 + (98-66.5)^3 + (113-66.5)^3 + (128-66.5)^3 + (143-66.5)^3]}{[(100-1)(12.693)^3]}$$

$$= \frac{708679.125}{(99)(2044.998)}$$

$$= 3.5$$

The graph is positively skewed or skewed to the right.

$$\text{Kurtosis} = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{(N-1)s^4}$$

$$= \frac{[(23-66.5)^4 + (38-66.5)^4 + (53-66.5)^4 + (68-66.5)^4 + (83-66.5)^4 + (98-66.5)^4 + (113-66.5)^4 + (128-66.5)^4 + (143-66.5)^4]}{[(100-1)(12.693)^4]}$$

$$= \frac{58561830.56}{(99)(25957.157)}$$

$$= 22.79$$

The Kurtosis is  $> 3$ :

$\therefore$  The graph is leptokurtic

c) Mode  $<$  Median  $<$  Mean

Most of the flight did not delayed for longer time