

PSDA Assignment 1Question 1

- (a) ratio (b) ordinal (c) interval (d) nominal
 (e) ordinal (f) ratio (g) ratio (h) interval
 (i) ratio (j) ratio

Question 2

- (a) i) Population: freshmen class at Lincoln High School (b) i) nominal
 ii) Sample: 100 students from the freshmen class ii) ratio
 iii) nominal

(c) i) $n=20$

Arranging data: 280, 340, 440, 490, 520, 540, 560, 560,
 580, 580, 600, 610, 630, 650, 660, 680,
 710, 730, 740, 740

minimum = 280, maximum = 740,

$Q_1 = 530, Q_2 = 590, Q_3 = 670$

ii) $Q_1: 20 \times \frac{1}{4} = 5$

$$Q_1 = \frac{Y[5] + Y[6]}{2} = \frac{520 + 540}{2} = 530 = Q_3 - Q_1$$

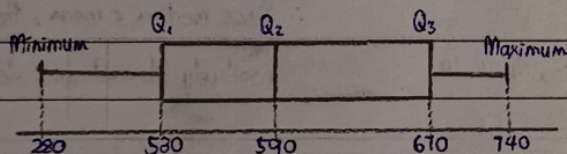
$$Q_2: 20 \times \frac{1}{2} = 10 = 670 - 530$$

$$Q_2 = \frac{Y[10] + Y[11]}{2} = \frac{580 + 600}{2} = 590 = 140$$

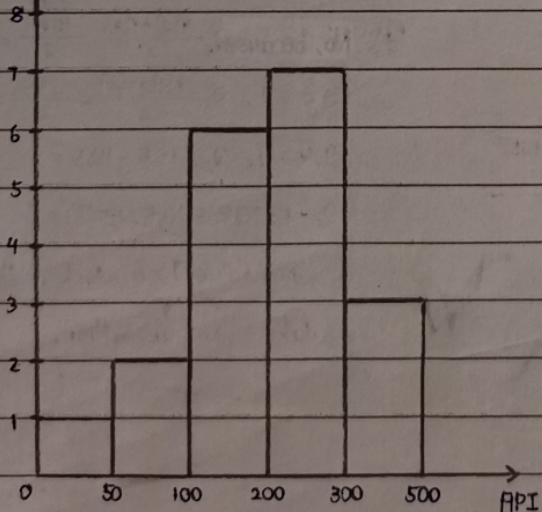
$$Q_3: 20 \times \frac{3}{4} = 15$$

$$Q_3 = \frac{Y[15] + Y[16]}{2} = \frac{660 + 680}{2} = 670$$

$\therefore$ Interquartile range, IQR

Question 3

(a) Frequency



(b) Areas affected with $API \geq 100 = 6 + 7 + 3 = 16$

API	Frequency	Relative frequency
0-50	1	0.05
50-100	2	0.10
100-200	6	0.32
200-300	7	0.37
300-500	3	0.16

$$\text{Percentage of area with } API \geq 100 = \frac{16}{19} \times 100\% = 84.21\%$$

- (c) Yes, mean value is affected. Because mean is unique for every set of data. The value 950 added is an outlier that is different from the other data, hence affecting the mean.

Question 4

(a) Mean, $\bar{x}$

$$= \frac{\sum x}{n}$$

$$= \frac{92+98+112+88+134+36+35+92+215+336+237}{11}$$

$$= 134.09$$

≈ 135 deaths

Mode = 92 (Year 2006, 2013)

Arranging data: 35, 36, 88, 92, 92, 98, 112, 134, 215, 237, 336

$$n=11, 11 \times \frac{1}{2} = 5.5 \therefore \text{Median} = Y[6] = 98 \text{ deaths}$$

$$\text{Standard deviation, } s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{(92-135)^2 + (98-135)^2 + (112-135)^2 + (88-135)^2 + (134-135)^2 + (36-135)^2 + (35-135)^2 + (92-135)^2 + (215-135)^2 + (336-135)^2 + (237-135)^2}{11-1}}$$

$$\therefore s = \sqrt{\frac{84812}{10}} = 92.09$$

(b) Percentile of value 215

$$= \frac{\text{no. of values less than 215}}{\text{total no. of values}} (100)$$

$$= \frac{8}{11} (100)$$

$$= 72.73$$

$\approx 73^{\text{th}}$ percentile

(c) Mean, because every data is counted even when there are outliers.

Question 5

(a) Mean, $\bar{x} = \frac{\sum x}{n} = \frac{14+14+10+12+11+13+11+11+14+10+13+8}{12} = 11.75$

Mode = 11, 14

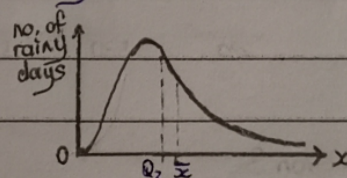
Arranging data: 8, 10, 10, 11, 11, 11, 12, 13, 13, 14, 14, 14

$$n=12, 12 \times \frac{1}{2} = 6 \therefore \text{Median} = \frac{Y[6] + Y[7]}{2} = \frac{11+12}{2} = 11.5$$

$$\text{Range} = x_h - x_l = 14 - 8 = 6$$

(b) Median = 11.5, Mean = 11.75

$\therefore$ Since median < mean, the distribution is positively skewed and skewed to the right.



(c) Standard deviation, s

$$= \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{(8-11.75)^2 + (10-11.75)^2 + (10-11.75)^2 + 3(11-11.75)^2 + (12-11.75)^2 + 2(13-11.75)^2 + 3(14-11.75)^2}{12-1}}$$

$$s = \sqrt{\frac{40.25}{11}} = 1.91$$

(d) $12 \times \frac{1}{4} = 3, Q_1 = \frac{Y[3] + Y[4]}{2} = \frac{10+11}{2} = 10.5$

$$12 \times \frac{3}{4} = 9, Q_3 = \frac{Y[9] + Y[10]}{2} = \frac{13+14}{2} = 13.5$$

$$\text{IQR} = Q_3 - Q_1 = 13.5 - 10.5 = 3$$

$$Q_1 - 1.5 \text{IQR} = 10.5 - 1.5(3) = 6$$

$\therefore$ Since value 7 > 6 which is the lower limit, the data is not an outlier.

Question 6

(a) Arranged data: 12, 14, 21, 28, 30

$$\text{Mean, } \bar{x} = \frac{\sum x}{n} = \frac{12+14+21+28+30}{5} = 21$$

$$n=5, 5 \times \frac{1}{2} = 2.5, \text{ Median} = Y[3] = 21$$

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{(12-21)^2 + (14-21)^2 + (21-21)^2 + (28-21)^2 + (30-21)^2}{5-1}}$$

$$s = \sqrt{\frac{260}{4}} = 8.06$$

$\therefore$ Skewness value

$$= \frac{\sum (x - \bar{x})^3}{(n-1)s^3}$$

$$= \frac{(12-21)^3 + (14-21)^3 + (21-21)^3 + (28-21)^3 + (30-21)^3}{(5-1)(8.06)^3}$$

$$= 0$$

(b) Skewness = 0

$\therefore$ Since skewness = 0, the sample is symmetrically distributed.