1. Categorize the following variables as being qualitative or quantitative and specify the level of measurement scale.

qualitative ① quantitative ②

(a) hair colour ① N
(b) number of television sets in a private home ② R
(c) possible responses to questionnaire items: strongly agree, agree, etc. ① O
(d) Intelligence scale of 0 to 100 ② I
(e) the country where you were born in ① N

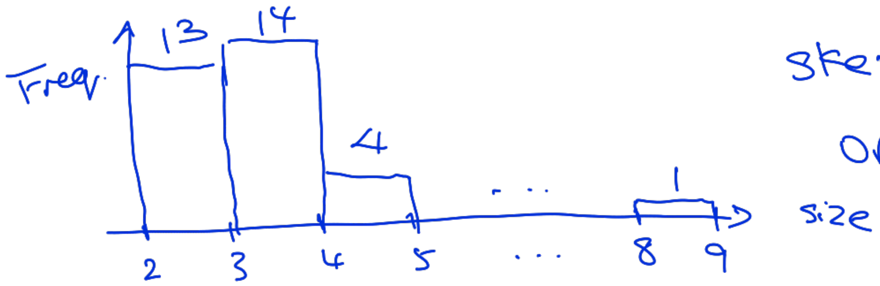
2. The following data were obtained for a measurement of certain fuel droplet size.

2.1	2.2	2.2	2.3	2.3	2.4	2.5	2.5	2.5	2.8
2.9	2.9	2.9	3.0	3.1	3.1	3.2	3.3	3.3	3.3
3.4	3.5	3.6	3.6	3.6	3.7	3.7	4.0	4.2	4.5
4.9	5.1	5.2	5.3	5.7	6.0	6.1	7.1	7.8	7.9
8.9									

(a) Group the droplet sizes and obtain a frequency table using class interval of 1 unit. Construct the frequency histogram and comment on the shape of the distribution.
(b) Construct the Stem-and-leaf diagram.
(c) Compute the mean, the 25th, 50th and 75th percentile.
(d) Construct the box plot for the droplet size data.

ⓐ

Class	[2,3)	[3,4)	[4,5)	...	[8,9)
Freq	13	14	4		1



skewed to the right

or

+ve skew

ⓑ

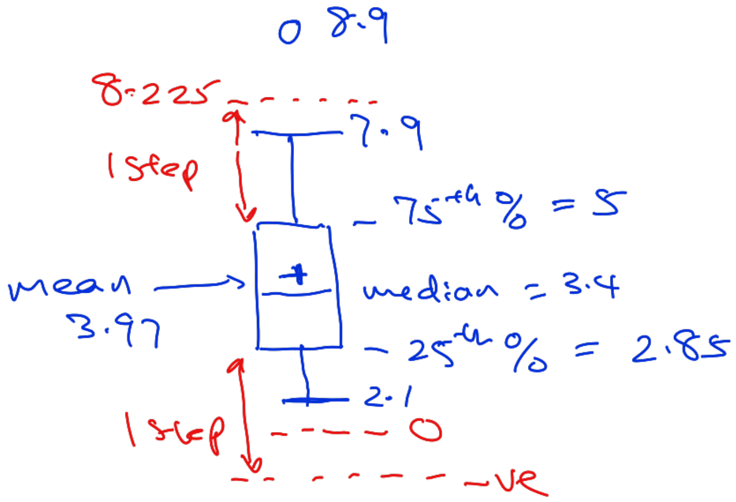
8 | 9
⋮
4 | 0 2 5 9
3 | 0 1 1 2 .... 7 7
2 | 1 2 2 3 .... 9 9

ⓒ mean $= \Sigma X / N = 3.97$

25^{th} % = data @ rank $\frac{25}{100}(41+1) = 10.5$

$= 2.85$

d



$$1 \text{ Step} = 1.5 \text{ IQR}$$
$$= 1.5(5 - 2.85)$$
$$= 3.225$$

3. A frequency distribution of the length of telephone calls monitored at the switchboard of an office is given below. Obtain the mean and the mode of the call duration.

Length of Calls (minutes)	Number of Calls
0 and under 2	10
2 and under 4	25
4 and under 6	20
6 and under 8	40
8 and under 10	5
Total	100

mid-pt x_i:
1
3
5
7
9

$$\text{mean} = \frac{10\times 1 + 25\times 3 + \ldots + 5\times 9}{100} = 5.1$$

mode = 7 — largest freq

4. Find the sample standard deviation for a set of data for which $n = 10$, $\sum x = 50$ and $\sum x^2 = 500$. Explain why it is impossible to have $n = 10$, $\sum x = 50$ and $\sum x^2 = 100$ for a given set of data.

$$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] = 27.78$$

$$s = \sqrt{27.78}$$

not possible because $s^2 \not< 0$

5. Given the following data, determine the Pearson's correlation between variable X and variable Y. Comment on your result.

X	2.5	3.4	5.6	6.7	7.9
Y	10.2	11.8	13.7	19.7	20.6

$$\rho = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}\sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}} \quad \text{or} \quad \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \, \sigma_Y}$$

get these values using cal.

$$= 0.955$$

6. In an attempt to find the mean number of hours his tutorial classmates spent per day preparing for tutorials, John collected data from 10 of his friends in the tutorial group and found that the mean is 2.4 hours with a standard deviation of 0.8 hours. However, a day later he felt that the sample size is too small. So he collected data from another 5 of his friends and found that the mean is 2.0 hours with a standard deviation of 1.2 hours. Find the mean and standard deviation when these 2 sets of data are combined.

Given $\bar{x} = \frac{\Sigma x_i}{n_x} = 2.4 \Rightarrow \Sigma x_i = 2.4 n_x = 24$

$\bar{y} = \frac{\Sigma y_i}{n_y} = 2.0 \Rightarrow \Sigma y_i = 2 n_y = 10$

$\left.\begin{matrix} y_i \end{matrix}\right\} x_i \quad z_i$

$\therefore$ mean $\bar{z} = \frac{\Sigma x_i + \Sigma y_i}{n_x + n_y} = 2.267$

$$S_x^2 = \frac{1}{n_x - 1}\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n_x}\right] = 0.8^2 \Rightarrow \Sigma x_i^2 = 63.36$$

$$S_y^2 = \frac{1}{n_y - 1}\left[\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n_y}\right] = 1.2^2 \Rightarrow \Sigma y_i^2 = 25.76$$

$$S_z^2 = \frac{1}{n_z - 1}\left[\Sigma z_i^2 - \frac{(\Sigma z_i)^2}{n_z}\right]$$

Σz_i

$$= \frac{1}{15-1}\left[\Sigma x_i^2 + \Sigma y_i^2 - \frac{(\Sigma x_i + \Sigma y_i)^2}{n_x + n_y}\right] = 0.86$$

$$\therefore S_z = \sqrt{0.86}$$