DESCRIPTIVE STATISTICS

➤ **<u>The weighted arithmetic mean</u>:**

Consider k data sets. The mean of data set i is denoted by $\bar{x}_i$ and the size of data set i by n_i, $i = 1, \ldots, k$. The joint arithmetic mean of the k data sets can then be expressed in terms of the means of the individual data sets as follows:

$$\bar{x}_\omega = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i}$$

<u>Example</u>:

Suppose Statistics students are divided into three groups of sizes 100, 160, and 140 respectively. These students wrote a test and the averages (means) of the respective groups were 66%, 52% and 70%. Calculate the mean score obtained by the whole group of 400 students.

$$\bar{x}_\omega = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$$

$$= \frac{n_1}{n_1 + n_2 + n_3}\bar{x}_1 + \frac{n_2}{n_1 + n_2 + n_3}\bar{x}_2 + \frac{n_3}{n_1 + n_2 + n_3}\bar{x}_3$$

$$= 0{,}25\bar{x}_1 + 0{,}4\bar{x}_2 + 0{,}35\bar{x}_3$$

$$= (0{,}25)(66) + (0{,}4)(52) + (0{,}35)(70)$$

$$= 61{,}8$$

Let $Y_1, Y_2, \ldots, Y_n$ denote n numbers with the relative importance of $w_1, w_2, \ldots, w_n$. The weighted mean of $Y_1, Y_2, \ldots, Y_n$ is given by $\bar{y}_\omega = \dfrac{\sum_{i=1}^{n} w_i\, y_i}{\sum_{i=1}^{n} w_i}$

<u>Example</u>:

The three most important dams in the Vaal River catchment area are the Vaal Dam, Bloemhof Dam and the Sterkfontein Dam. At the end of 1992 the net full supply capacity (FSC) in millions of cubic metres and the percentage content of these dams were as follows:

DAM	Net FSC	%-CONTENT
Vaal Dam	2529	20
Bloemhof Dam	1269	20
Sterkfontein Dam	2617	99

Example:

Consider the average number of rainy days for January to December in Bloemfontein:

15; 13; 11; 5; 4; 5; 4; 3; 7; 9; 16; 17

Calculate i) the first quartile;
ii) the eighth decile.

- 3; 4; 4; 5; 5; 7; 9; 11; 13; 15; 16; 17

i)

- $q_1 = p_{25}$

- $\dfrac{i(n+1)}{100} = \dfrac{25(13)}{100} = 3.25$

$$
\begin{aligned}
p_{25} &= obs(3) + 0.25(obs(4) - obs(3)) \\
&= 4 + 0.25(5 - 4) \\
&= 4 + 0.25 \\
&= 4.25
\end{aligned}
$$

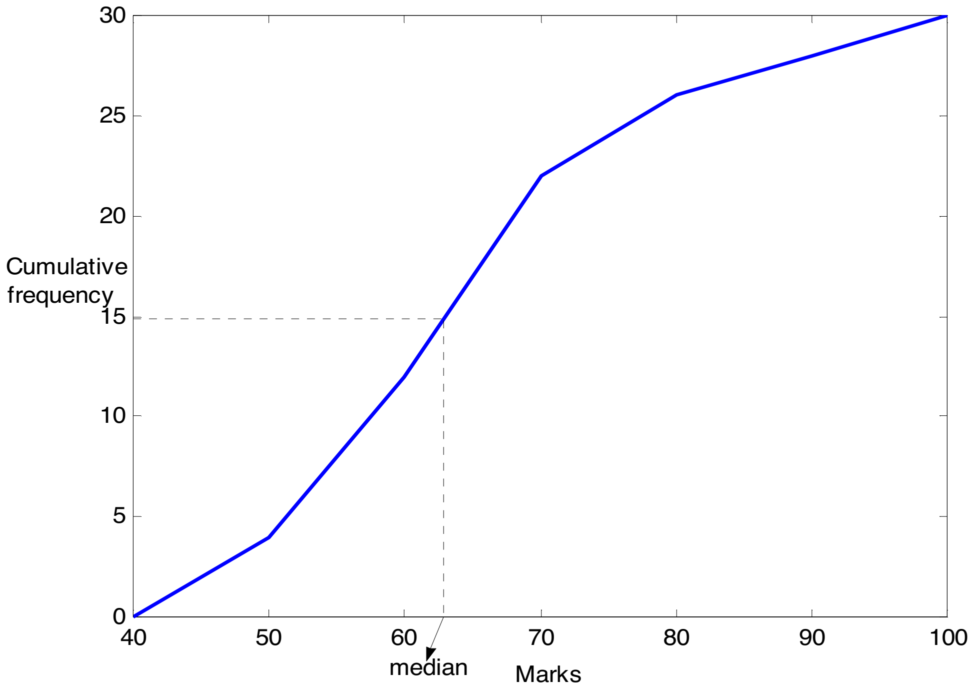
ii)

- $d_8 = p_{80}$

- $\dfrac{i(n+1)}{100} = \dfrac{80(13)}{100} = 10.4$

$$
\begin{aligned}
p_{80} &= obs(10) + 0.4(obs(11) - obs(10)) \\
&= 15 + 0.4(16 - 15) \\
&= 15 + 0.4 \\
&= 15.4
\end{aligned}
$$

Cumulative frequency polygon of the marks obtained by 30 students.



1.5.2 Measures of spread:

If only the mean of a data set is known, we do not know how the observations are spread about the mean.

Measures of spread are the range and the standard deviation.

Consider rainfall figures for Pretoria and Durban:

Month	Nov	Dec	Jan	Feb	March
Pretoria	119.6	119.9	134.4	113.0	94.2
Durban	110.9	106.6	110.1	124.7	105.7

According to the means there is apparently a strong similarity:

Pretoria: $\bar{x}_P = 116.22$ Durban: $\bar{x}_D = 111.6$

The next graph, however, illustrates the difference between the rainfall figures by taking the spread into account.

$$= \frac{\sum_{i=1}^{n} x_i^2 - \frac{2}{n}\left(\sum_{i=1}^{n} x_i\right)^2 + \frac{n}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} x_i^2 - \frac{2}{n}\left(\sum_{i=1}^{n} x_i\right)^2 + \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$$

<u>Properties of the standard deviation</u>:

- The standard deviation is based on all values in the data set and is the most important measure of spread.
- s fluctuates very little from one sample to the next taken from the same population.
- The larger the value of s, the further the observations are from $\bar{x}$.
- A variance and standard deviation is nonnegative (positive or zero).

<u>Example</u>:

Consider the scores obtained by seven gymnasts:

8.10 7.10 6.65 8.60 6.20 6.55 7.75

Calculate the standard deviation and the variance.

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

$$= \frac{375.6075 - \frac{(50.95)^2}{7}}{6}$$

$$= 0.7941$$

$$s = 0.8911$$