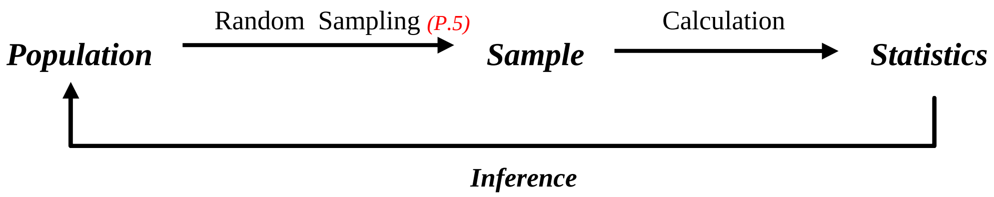
A. Definition

Statistics : It is the procedures for *obtaining*, *processing* and *analysing* data in an orderly manner, and for making logical *decision* based on the data.

B. Populations & Samples *(P.2)*



Notation / parameters

Characteristics	Population	Sample
size	N	n
mean	μ	$\bar{x}$
standard deviation	σ	s

C. Discrete & Continuous Variables *(P.76)*

1. Ungrouped discrete data
   { 0, 2, 5, 8, 10, 14,. . . }

2. Grouped data
   For a sample containing a large no. of observations, treatment can be simplified by grouping the individual observations into classes or class intervals. In this case, the data are referred to as grouped data.

   (a) discrete data

x	f
0	1
1	2
2	1
3	4
4	3

   (b) Continuous data

class interval	f
151 - 155	5
156 - 160	10
161 - 165	20
166 - 170	5

D. Frequency Tables (P.79)

1. Relative frequencies

$$r_i = \frac{f_i}{N} \quad where \ \ N = \Sigma f$$

2. Percentage frequencies

$$r_i = \frac{f_i}{N} \times 100\%$$

3. Cumulative frequencies
   Example 11.9

E. Graphical Representations

1. Bar Charts *Fig.11.2, 11.3* (P.90)
   Example 11.2, 11.3

2. Histograms *Fig.11.4, 11.5* (P.91)
   If the class width is unequal, apply,

$$\text{freq. density of a class} = \frac{\text{frequency of the class}}{\text{class width}} \quad (P.92)$$

   Example 11.5, 11.6

3. Frequency polygons / curves (*class freq. Against class mark*) (P.93)
   Example 11.6, 11.7

4. Cumulative frequency polygons / curves *Fig. 11.10, 11.11* (P.97)

5. Stem - and - leaf diagrams (P.101)
   Example 11.10 Self - study(P.101 - 103)

F. Measures of Central tendency

The tendency of the statistical data to concentrate at certain values.

1. Arithmetic Mean / Mean
   (a) Ungrouped discrete data (P.13)

Population	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$
Sample	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

   Example 10.6, 10.7, 10.8

   (b) Grouped data (P.15)
   (i) discrete

Population	$\mu = \frac{1}{N}\sum_{i=1}^{k} f_i x_i$ where $N = \sum_{i=1}^{k} f_i$
Sample	$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i$ where $n = \sum_{i=1}^{k} f_i$

   Example 10.9, 10.10

1. (b) (ii) Continuous *(P.111)*
same as 1(b)(i) for x_i = class mark

class interval	f	class mark
151 - 155	5	
156 - 160	10	
161 - 165	20	
166 - 170	5	

2. Weighted Mean = $\dfrac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i}$ *(P.20)*

Application：Construction of price index（物價指數）
Example 10.12

Example

The mean annual salary paid to all employees in a company was $10000. The mean annual salary paid to male and female employees of the company were $12000 and $9000 respectively. Determine the percentages of the males and females employed by the company.

<u>Combined mean</u> (P.28)

If a population consist of 2 sets of data with respective mean $\overline{x_1}$ and $\overline{x_2}$, then the mean of the population is

$$\overline{x} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2},$$

Where n_1 and n_2 are the % of male and female workers.

3. Median *(P.22)*
Data should be arranged either in ascending or descending order first.

(a) Ungrouped discrete data

$$\text{median} = \begin{cases} x_{\frac{N+1}{2}} & \text{if } N \text{ is odd} \\ \frac{1}{2}\left(x_{\frac{N}{2}} + x_{\frac{N}{2}+1}\right) & \text{if } N \text{ is even} \end{cases}$$

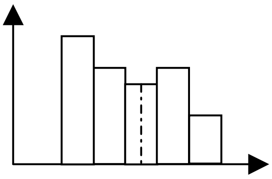
Example 10.13, 10.14

(b) Grouped data *(P.23)*

(i) discrete
Imagine the data in the groups as being spread out, and use 3(a).
Example 10.15

(ii) Continuous
(1) use Cumulative frequency polygon / curve

3. (b) (ii) (2) use Histogram *(P.115)*



$$\text{median} = L + C\frac{\frac{N}{2} - Fc}{fm}$$

where N = total freq.
C = class width
L = lower boundary of median class
fm = freq. of median class
Fc = cumulative freq. for the class next lower than the median class

Example 11.12

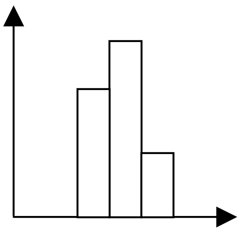
Example Find the median and median class of the data which represent the grades of an examination.

Class interval	Freq.	less than	c.f.
30 - 39	6		
40 - 49	12		
50 - 59	15		
60 - 69	25		
70 - 79	18		
80 - 89	6		
90 - 99	4		

4. Mode *(Unimodal, bimodal & multimodal set)*

(a) Discrete (ungrouped and grouped) data *(P.25)*
The value occurs most frequently in the set.

(b) Continuous data (Modal class) *(P.114)*



$$\text{Mode} = L + C\left(\frac{\Delta_L}{\Delta_L + \Delta_R}\right)$$

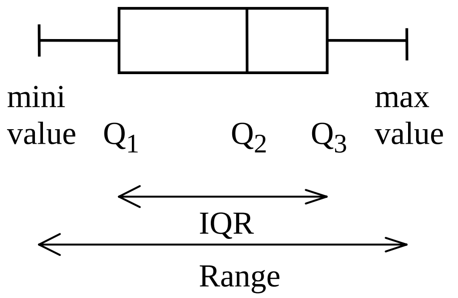
where L = lower boundary of modal class

G. Measures of Dispersion

1. Range = largest value of the set − smallest value of the set *(P.35)*
Example 10.18

2. Interquartile range (IQR) = Upper quartile (Q_3) − Lower quartile (Q_1) *(P.36)*
Example 10.19
Example Find the range and IQR of the sequence of number
{ 3, 7, 8, 12, 16, 24, 31, 59, 63, 72, 91}

3. Box and whiskers diagram / Box plots *(P.37)*



Example 10.21

Example Draw the box plot of the sequence of number
{ 3, 7, 8, 12, 16, 24, 31, 59, 63, 72, 91}

4. Quantiles

(a) Discrete (ungrouped and grouped) data *(P.39)*

(i) deciles (divide the set of data into 10 equal parts)
(ii) percentiles (divide the set of data into 100 equal parts)

General rule for locating quantiles:

$$r = \frac{p}{q}(n+1)$$

where p = pth data
q = (10 for decile, 100 for percentile)
n = total no. of data
r = the required observation, *round to nearest integer*

If r is obtained mid-way between integers

(1) quartiles — following rounding off rule. 四捨五入
(2) quantiles — round towards the median.

Example 10.22, 10.23, 10.24, 10.25

Exercise 10.3 Q13

(b) Continuous data *(P.113)*

(i) Use cumulative frequency polygons / curve

(ii) General formula

$$\text{pth q-tile} = L + \frac{1}{fm}\left(\frac{pN}{q} - Fc\right)C$$

where
N = total freq.
C = class width
L = lower boundary of pth q-tile class
fm = freq. of pth q-tile class
Fc = cumulative freq. for the class next lower than the pth q-tile class

(c) Use of the stem - and - leaf diagram *(P.118)*

To calculate median, quartile & quantile, use

$$r = \frac{p}{q}(n+1)$$

Example 11.13

Exercise 11.3 Q17, 18, 19; Exercise 11.4 Q4; Revision Exercise 11 Q14

5. Variance

	Population $(\sigma_n)^2$ *(P.43)*	Sample $(\sigma_{n-1})^2$ *(P.50)*
Ungrouped data	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$
Grouped data	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k}f_i(x_i-\mu)^2$ where $N=\sum_{i=1}^{k}f_i$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}f_i(x_i-\bar{x})^2$ where $n=\sum_{i=1}^{k}f_i$

6. Standard deviation

$\sigma, s = \sqrt{\text{variance}}$

Example 10.26

7. ‘ Short - cut ’ formula

	Population *.(P 45, 47)*	Sample *(P.52)*
Ungrouped data	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}x_i^2-\mu^2$ where $\mu=\frac{1}{N}\sum_{i=1}^{N}x_i$	$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)$ where $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$
Grouped data	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k}f_ix_i-\mu^2$ where $\mu=\frac{1}{N}\sum_{i=1}^{k}f_ix_i$	$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k}f_ix_i^2-n\bar{x}^2\right)$ where $\bar{x}=\frac{1}{n}\sum_{i=1}^{k}f_ix_i$

Example 10.27, 10.28, 10.29, 10.30, 11.11, 11.12

Exercise 10.3 Q20, 27, 29; Exercise 11.4 Q12

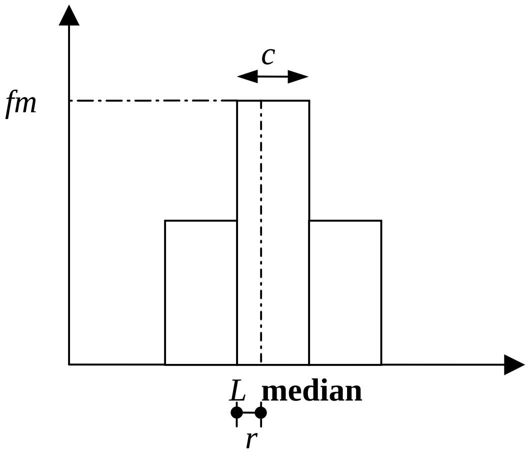
Proof :

$$\begin{aligned}\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 &= \frac{1}{N}\sum_{i=1}^{N}(x_i^2-2\mu x_i+\mu^2) \\ &= \frac{1}{N}(\sum_{i=1}^{N}x_i^2-2\mu\sum_{i=1}^{N}x_i+N\mu^2) \\ &= \frac{1}{N}(\sum_{i=1}^{N}x_i^2-2N\mu^2+N\mu^2) \\ &= \frac{1}{N}\sum_{i=1}^{N}x_i^2-\mu^2 .\end{aligned}$$

H. Self Studying

P.24 Comparison of the median with the mean,
P.37 Interquartile range compared with range
P.83 Procedure in constructing frequency tables
P.120 Symmetry, skewness & relative positions of measures of central tendency

(I) Median



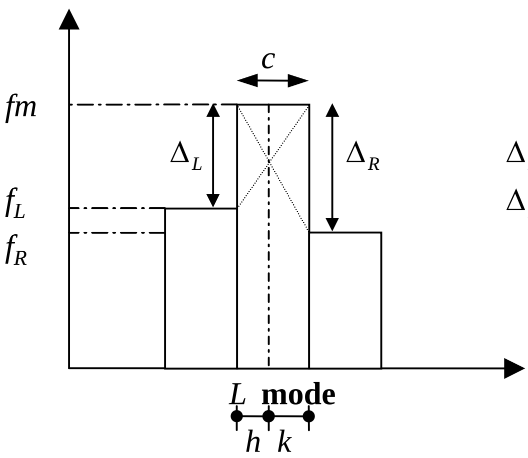
$$\frac{N}{2} - F_C : fm = r : c$$

$$\therefore \quad r = c\left[\frac{\frac{N}{2} - F_c}{fm}\right]$$

$$\therefore \quad \text{median} = L + r$$

$$\boxed{\therefore \quad \text{median} = L + c\left[\frac{\frac{N}{2} - F_c}{fm}\right]}$$

(II) Modal Class



$$\Delta_L = fm - f_L$$
$$\Delta_R = fm - f_R$$

By similar triangle

$$\frac{h}{k} = \frac{\Delta_L}{\Delta_R}$$

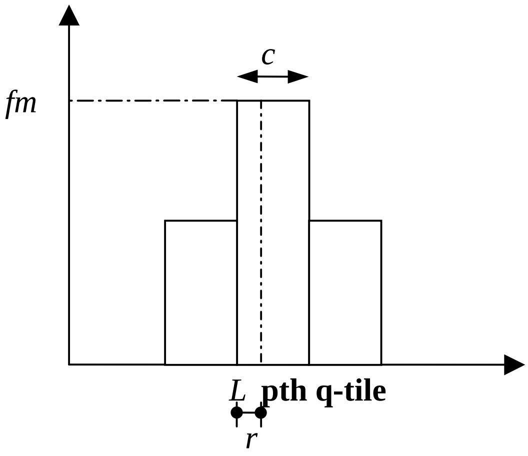
$$\therefore \quad h = \frac{\Delta_L}{\Delta_R}(c - h)$$

$$\therefore \quad h = \left(\frac{\Delta_L}{\Delta_L + \Delta_R}\right) c$$

$$\therefore \quad \text{mode} = L + h$$

$$\boxed{\therefore \quad \text{mode} = L + \left(\frac{\Delta_L}{\Delta_L + \Delta_R}\right) c}$$

(III) Quantiles



$$\frac{p}{q}N - F_c : fm = r : c$$

$$\therefore \quad r = c\left[\frac{\frac{p}{q}N - F_c}{fm}\right]$$

$$\therefore \quad \text{pth q-tile} = L + r$$

$$\boxed{\therefore \quad \text{pth q-tile} = L + c\left[\frac{\frac{p}{q}N - F_c}{fm}\right]}$$

The Consumer Price Index

Index Series	Expenditure range (monthly household expenditure during Oct 94 - Sep 95)	% of households covered
CPI(A)	$4000 - $15999	50
CPI(B)	$16000 - $29999	30
Hang Seng CPI	$30000 - $59999	10
Composite CPI	$4000 - $59999	90

Expenditure Weight (%) of the 1994/95-based CPIs