8. Statistics and Probability

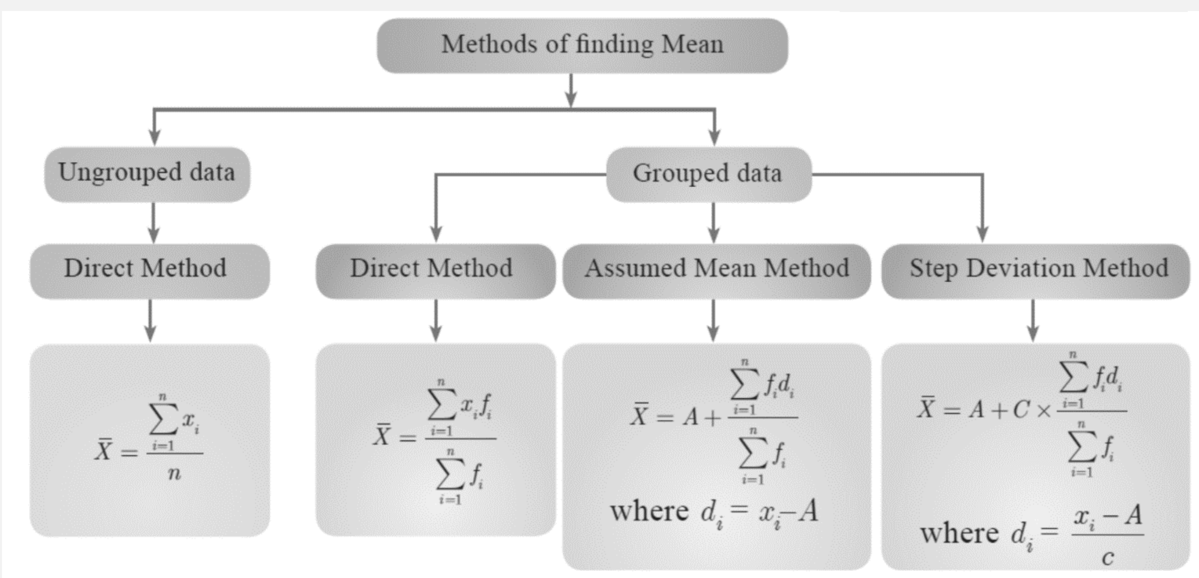
Introduction for Exercise 8.1

Concept corner

Note: Measures of Central Tendency: It is often convenient to have one number that represent the whole data. Such a number is called a Measures of Central Tendency.

The most common among them are Arithmetic Mean, Median, Mode.

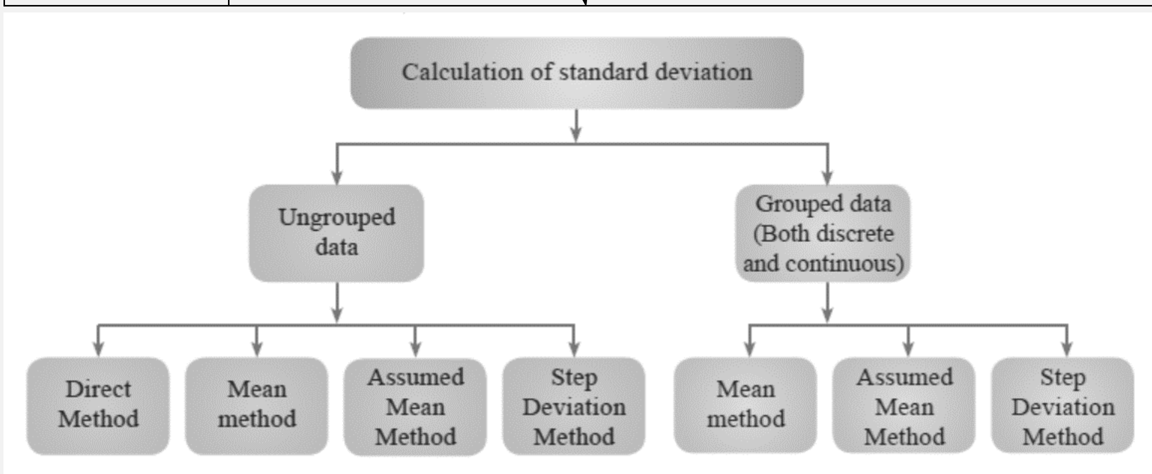
Data	The numerical representation of facts is called data.
Observation	Each entry in the data is called an observation.
Variable	The quantities which are being considered in a survey are called variables. Variables are generally denoted by x_i, where $i = 1,2,3,\dots,n$.
Frequencies	The number of times, a variable occurs in a given data is called the frequency of that variable. Frequencies are generally denoted as f_i, where $i = 1,2,3,\dots,n$.
Arithmetic Mean	The Arithmetic Mean or Mean of the given values is sum of all the observations divided by the total number of observations. It is denoted by $\bar{x}$ (pronounced as x bar) $\bar{x} = \frac{\text{Sum of all the observations}}{\text{Number of observations}}$



- Measures of Variation (or) Dispersion of a data provide an idea of how observations spread out (or) scattered throughout the data.
- Different Measures of Dispersion are

1. Range
2. Mean deviation
3. Quartile deviation
4. Standard deviation
5. Variance
6. Coefficient of Variation

Range	The difference between the largest value and the smallest value is called Range. Range $R = L - S$ Coefficient of range $= \frac{L-S}{L+S}$ (L – Largest value, S – Smallest value)
Deviation from the mean	For a given data with n observations $x_1, x_2, x_3 \ldots . x_n$ the deviations from the mean $\bar{x}$ are $x_1 - \bar{x}\,, x_2 - \bar{x}\,, \ldots\, x_n - \bar{x}$
Squares of deviations from the mean	The squares of deviation from the mean $\bar{x}$ of the observations $x_1, x_2 . . x_n$ are $(x_1 - \bar{x})^2\,, (x_2 - \bar{x})^2\,, \ldots\, (x_n - \bar{x})^2$ or $\sum_{i=1}^{n}(x_i - \bar{x})^2$ $(x_i - \bar{x})^2 \geq 0$ for all observations $x_i, i = 1,2,3, \ldots n$. If the deviations from the mean $(x_i - \bar{x})$ are small, then the squares of the deviations will be very small.
Variance	The mean of the squares of the deviations from the mean is called Variance. It is denoted by σ^2 (read as sigma square) Variance $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$
Standard Deviation	The positive square root of Variance is called Standard deviation. That is, standard deviation is the positive square root of the mean of the squares of deviations of the given values from their mean. It is denoted by σ Standard deviation $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$



Calculation of Standard Deviation for ungrouped data

(i)	Direct Method	$\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$
(ii)	Mean Method	If $d_i = x_i - \bar{x}$ are the deviations, then $\sigma = \sqrt{\frac{\Sigma d_i^2}{n}}$
(iii)	Assumed Mean Method	$\sigma = \sqrt{\frac{\Sigma d_i^2}{n} - \left(\frac{\Sigma d_i}{n}\right)^2}$
(iv)	Step deviation Method	$\sigma = c \times \sqrt{\frac{\Sigma d_i^2}{n} - \left(\frac{\Sigma d_i}{n}\right)^2}$

Calculation of Standard Deviation for grouped data

(i)	Mean Method	$\sigma = \sqrt{\frac{\Sigma f_i d_i^2}{N}}$, where $N = \sum_{i=1}^{n} f_i$
(ii)	Assumed Mean Method	$d_i = x - A\,,\quad \sigma = \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2}$

Calculation of Standard deviation for continuous frequency distribution

(i)	Mean Method	$\sigma = \sqrt{\frac{\Sigma f_i(x_i-\bar{x})^2}{N}}$, where x_i = Middle value of the i^{th} class f_i =Frequency of the i^{th} class
(ii)	Shortcut Method (or) Step deviation method	$d_i = \frac{x_i - A}{c}$, $\sigma = c \times \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2}$

- The SD of first 'n' natural numbers, $\sigma = \sqrt{\frac{n^2-1}{12}}$
- The SD **will not change** when we add or subtract some fixed constant to all the values
- SD of a collection of data gets multiplied or divided by the quantity k, if each item is multiplied or divided by k
- If the frequency of initial class is zero, then the next class will be considered for the calculation of range.
- The range of a set of data does not give the clear idea about the dispersion of the data from measures of Central Tendency. For this, we need a measure which depend upon the deviation from the measures of Central Tendency.
- $(x_i - \bar{x}) \geq 0$ for all observations $x_i, i = 1,2,3, \dots n$. If the deviations from the mean $(x_i - \bar{x})$ are small, then the squares of the deviations will be very small.

Note: While computing standard deviation, arranging data in ascending order is not mandatory.

- If the data values are given directly then to find standard deviation we can use the formula $\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$
- If the data values are not given directly but the squares of the deviations from the mean of each observation is given then to find standard deviation we can use the formula $\sigma = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}$

Introduction for Exercise 8.2

Concept corner

Definition: For comparing two or more data for corresponding changes the relative measure of standard deviation called **Coefficient of variation.**

Coefficient of variation of first data (C.V1)= $\frac{\sigma_1}{\bar{x}_1} \times 100\%$

Coefficient of variation of second data (C.V2)= $\frac{\sigma_2}{\bar{x}_2} \times 100\%$

i) The data with **lesser** coefficient of variation is more consistent or stable than the other data.

ii) The data with **greater** coefficient of variation is inconsistent.

iii) The data have equal coefficient of variation values one data depends on the other.

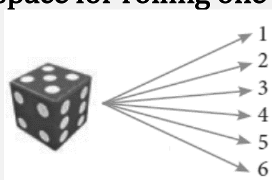
To Find the Square root:	**Example: To find the square root of 75**
$\sqrt{X} = \sqrt{S} + \frac{(X-S)}{2\sqrt{S}}$ X –the number you want the square root S – the closet square number you know to X	$X = 75, S = 81$ (nearest square) $\sqrt{S} = 9$ $\sqrt{75} = \sqrt{81} + \frac{(75-81)}{2(\sqrt{81})} = 9 + \frac{-6}{2(9)} = 9 - \frac{6}{18}$ $= 9 - 0.333 = 8.667$

Introduction for Exercise 8.3

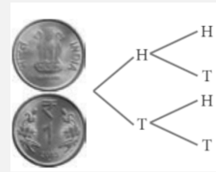
Concept corner

- A **random experiment** is an experiment in which
  (i) The set of all possible outcomes are known (ii) Exact outcome is not known
- **Sample space:** The set of all possible outcomes in a random experiment is called a sample space. It is generally denoted by S
- **Sample point :** Each element of a sample space is called a sample point.
- **Tree diagram:** Tree diagram allow us to see visually all possible outcomes of an random experiment. Each branch in a tree diagram represent a possible outcome.

Sample space for rolling one die



Sample space for toss two coins



Events	Explanation	Example
Equally likely events	Two or more events are said to be **equally likely** if each one of them has an equal chance of occurring.	Head and tail are equally likely events in **tossing a coin.**
Certain events	In an experiment, the event which surely occur is called **certain event.**	When we **roll a die,** the event of getting any natural number from 1 to 6 is a certain event.
Impossible events	In an experiment if an event has no scope to occur then it is called an **impossible event.**	When we **toss two coins**, the event of getting three heads is an impossible event.
Mutually exclusive events	Two or more events are said to be **mutually exclusive** if they don't have common sample points. i.e., events A, B are said to be mutually exclusive if, $A \cap B = \emptyset$.	When we **roll a die** the events of getting odd numbers and even numbers are mutually exclusive events.
Exhaustive events	The collection of events whose union is the whole sample space are called **exhaustive events.**	When we **toss a coin twice**, the collection of events of getting two heads, exactly one head, no head are exhaustive events.
Complementary events	The **complement of an event** A is the event representing collection of sample points not in A. It is denoted A' or A^c or $\bar{A}$ The event A and its complement A' are mutually exclusive and exhaustive.	When we **roll a die**, the event 'rolling 5 or 6' and the event of rolling 1, 2, 3 or 4 are complementary events.

Elementary Event: If an event E consists of only one outcome then it is called an elementary event.

Probability of an event:

In a random experiment, let S be the sample space and $E \subseteq S$. Then if E is an event, the probability of occurrence of E is defined as

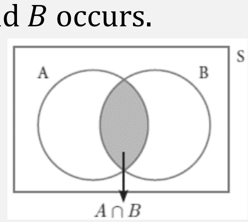
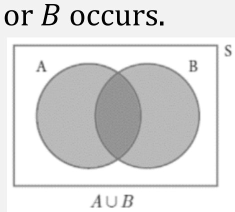
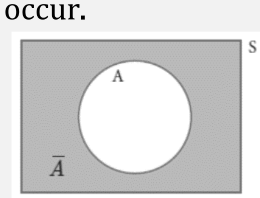
$$P(E) = \frac{\text{Number of outcomes favourable to occurence of } E}{\text{Number of all possible outcomes}} = \frac{n(E)}{n(S)}$$

- $P(E) = \frac{n(E)}{n(S)}$
- $P(S) = \frac{n(S)}{n(S)} = 1.$ The probability of sure event is 1.
- $P(\emptyset) = \frac{n(\emptyset)}{n(S)} = \frac{0}{n(s)} = 0.$ The probability of impossible event is 0.
- Since E is a subset of S and $\emptyset$ is a subset of any set,
  $\emptyset \subseteq E \subseteq S$
  $P(\emptyset) \leq P(E) \leq P(S)$
  $0 \leq P(E) \leq 1$
  Therefore, the probability value always lies from 0 to 1.
- The complement event of E is $\bar{E}$.
  Let $P(E) = \frac{m}{n}$
  (Where m is the number of favorable outcomes of E and
  n is the total number of possible outcomes).
  $$P(\bar{E}) = \frac{\text{Number of outcomes unfavourable to occurance of } E}{\text{Number of all possible outcomes}}$$
  $$P(\bar{E}) = \frac{n-m}{n} = 1 - \frac{m}{n}$$
  $P(\bar{E}) = 1 - P(E)$
- $P(E) + P(\bar{E}) = 1$

Introduction for Exercise 8.4

Concept corner

Algebra of events In a random experiment, let S be the sample space. Let $A \subseteq S$ and $B \subseteq S$ be the events in S. We say that

(i) $(A \cap B)$ is an event that occurs only when both A and B occurs. 	(ii) $(A \cup B)$ is an event that occurs when either one of A or B occurs. 	(iii) $\bar{A}$ is an event that occurs only when A doesn't occur. 

- $A \cap \bar{A} = \emptyset$, $A \cup \bar{A} = S$
- If A, B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$

- P (Union of mutually exclusive events)= $\sum(Probability\ of\ events)$

Verbal description of the event	**Equivalent set theoretic notation**
Not A	$\bar{A}$
A or B (at least one of Aor B)	$A \cup B$
A and B	$A \cap B$
A but not B	$A \cap \bar{B}$
Neither A nor B	$\bar{A} \cap \bar{B}$
At least one of A, B or C	$A \cup B \cup C$
Exactly one of A and B	$(A \cap \bar{B}) \cup (\bar{A} \cap B)$
All three of A, B and C	$A \cap B \cap C$
Exactly two of A, B and C	$(A \cap B \cap \bar{C}) \cup (A \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap C)$

Theorem 1: If A and B are two events associated with a random experiment, then prove that (i) $P(A \cap \bar{B}) = P$ (only A)= $P(A) - P(A \cap B)$(ii) $P(\bar{A} \cap B) = P$(only B)= $P(B) - P(A \cap B)$

Addition Theorem of Probability:

(i) If A and B are any two events then, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(ii) If A, B and C are any three events then,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$$