STATISTICS AND PROBABILITY

(i) Arithmetic mean

If $x_1, x_2, \ldots, x_n$ are n values of variable x, then the arithmetic mean of these values is
given by

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right).$$

Frequency distribution: In case of a frequency distribution $x_i(f_i)$ [where f_i is the frequency of the variable x_i or we can say number of times the variable x_i appears]

Arithmetic Mean $(\bar{x}) = \dfrac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n}$

$$= \frac{1}{N}\left(\sum_{i=1}^{n} f_i x_i\right) \qquad \text{(i)}$$

where N = $f_1 + f_2 + f_3 + \ldots + f_n$

(ii) Geometric Mean

Geometric mean of a set of n observations is the nth root of their product. Thus the geometric mean G of n non–zero observations
$x_1, x_2, \ldots, x_n$ is
G = $(x_1, x_2, x_3 \ldots x_n)^{1/n}$

(iii) Harmonic Mean

Harmonic mean of a number of non–zero observations is the reciprocal of the A.M. of the reciprocal of the given values. Thus harmonic mean H of n non–zero observations $x_1\, x_2, x_3 \ldots x_n$ is

$$H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)} \qquad \text{(i)}$$

In case of a frequency distribution
$x_i/f_i : i = 1, 2, \ldots, n.$

$$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}\left(\frac{f_i}{x_i}\right)}$$

(iv)Mode

$$\text{Mode} \quad = l + \frac{(f - f_1)}{(2f - f_1 - f_2)} \times h$$

Where, l = lower limit of the modal class
h = width of the modal class
f_1 = frequency of the class preceding the modal class
f_2 = frequency of the class following the modal class
f = frequency of the modal class

Median The median of a distribution is defined as the middle or central value in the distribution when the series of data are arranged in the ascending or descending order. It divides the series of data in two equal parts.

Calculation of median for ungrouped Data

(a) Arranged the data in ascending or descending order of magnitude.

(b) Then if the number of data n is odd then $\frac{n+1}{2}$th observation will be the middle value and hence this observation will be the median.

But if the number of data n is even then find the $\frac{n}{2}$th and $\left(\frac{n}{2}+1\right)$the observations. The mean of these two observation is the median.

Calculation of Median for grouped data

Case 1: When the series of data is discrete

(a) Arrange the grouped data in ascending or descending order of magnitude together with their frequencies.

(b) Prepare the cumulative frequency table.

(c) If the total frequency, n is odd then $\frac{n+1}{2}$th observation is the median.

But if the total frequency, n is even, then find the $\frac{n}{2}$the and $\left(\frac{n}{2}+1\right)$th observations. Then the median is the mean of these two observations.

Empirical relation between Mean , Mode and Median

Mean – Mode = 3(Mean – Median)

$\Rightarrow$ Mode = 3Median – 2Mean

(iii) Mean Deviation

If $x_i/f_i i = 1, 2\ldots., n$ is the frequency distribution, then mean deviation from an average A(median, mode or Arithmetic mean) is given by

M.D. = $\frac{1}{N}\sum_{i=1}^{n} f_i\,|x_i - A|$, where $\sum_{i=1}^{n} f_i = N$

(iv) Standard Deviation (S.d) and Variance

The standard deviation of variate X is the square root of the AM of the squares of all deviations of X from the A.M. of the observations and it is denoted by (σ)

Thus, if $x_i(f_i)$, $i = 1, 2, 3, \ldots, n$ is a frequency distribution, then

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{X}\right)^2} \qquad \text{(i)}$$

–efficient of variations for each series. The series having greater CV is said to be more variable than the other. where $\bar{X}$ is the A.M. of the distribution and N = $\sum_{i=1}^{n} f_i$

The square of the standard deviation is called the variance and is given by

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{X})^2$$

Probability

Probability, in the conventional sense, is a ratio between what may be called as the 'number of favourable' to the 'total number'.

RANDOM EXPERIMENT

It is an operation which can result in any one of its well defined outcomes and the outcome cannot be predicted with certainity.

SAMPLE SPACE AND SAMPLE POINTS

The set of all possible outcomes of a random experiment is called sample space and is denoted by *S*. Every possible outcome i.e. every element of this set is a sample point.

EVENT

A subset of sample space, i.e. a set of some of possible outcomes of a random experiment is called as event.

ALGEBRA OF EVENTS

In connection with basic probability laws we shall need the following concepts and facts about events (subsets) *A*, *B*, *C*, …… of a given sample space *S*

Verbal description of the event	Equivalent set theoretic notation
Not *A*	$\bar{A}$
A or *B* (at least one of *A* or *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*	$(A \cap B \cap \bar{C}) \cup (A \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap C)$

DEFINITION OF PROBABILITY WITH DISCRETE SAMPEL SPACE

If the sample space S of an experiment consists of finitely many outcomes (points) that are equally likely, then the probability of occurrence of an event A is

$$P(A)=\frac{\text{Number of sample points in A}}{\text{Number of sample points in S}}$$

$$P(A)=\frac{n(A)}{n(S)}$$

In particular $P(S) = 1$ and $0 \leq P(A) \leq 1$.

AXIOMATIC DEFINITION OF PROBABILITY

Given a sample space S, with each event A of S, there is associated a number P(A), called the probability of A, such that the following axioms of probability are satisfied.

- For every A in S, $0 \leq P(\mathrm{A}) \leq 1$
- The entire sample space has the probability $P(S) = 1$
- For mutually exclusive events A and B $(A \cap B = \phi)$, $P(A \cup B) = P(A) + P(B)$.

BASIC THEORIES OF PROBABILITY

1. For an event A and its complement A^c in sample space S_1**$P(\mathrm{A}^c) = 1 - P(\mathrm{A})$**

$\because \quad A \cap A^c = \phi$ and $A \cup A^c = S$

and $P(A \cup \mathrm{A}^c) = P(A) + P(A^c)$

$\Rightarrow \quad P(S) = P(A) + P(A^c)$

$\Rightarrow \quad 1 = P(A) + P(A^c)$

2. THEOREM OF TOTAL PROBABILITY OR ADDITION THEOREM OF PROBABILITY

If *A* and *B* be any two events in a sample space *S*, then the probability of occurrence of at least one of the events *A* and *B* is given by

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

or $P(A') = 1 - P(A)$

Theorem: If *A*, *B* and *C* are any three events in a sample space *S*, then

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

3. For exhaustive events $A_1, A_2, \ldots. A_n$ in a sample space.

$$P(A_1 \cup A_2 \cup \ldots\ldots.. \cup A_n) = 1$$

4. For mutually exclusive events $A_1A_2 \ldots\ldots.. A_n$ in a sample space S

$$P(A_1 \cup A_2 \cup A_3 \ldots.. \cup A_n) = P(A_1) + P(A_2) + \ldots\ldots.. + P(A_n)$$

5. For mutually exclusive and exhaustive events $A_1, A_2, \ldots\ldots\ldots., A_n$ in a sample space.

$$P(A_1 \cup A_2 \cup A_3 \ldots.. \cup A_n) = P(A_1) + P(A_2) + \ldots.. + P(A_n) = 1$$

6. For an event A in sample space S, **'the odds in favour of *A*'** are $\frac{P(A)}{P(A^c)}$ where A^c is the complement of the event A.

Also **'the odds against *A*'** are $\frac{P(A^c)}{P(A)}$.