# Probability :-

Galileo was the 1st mathematician who attempted to discover probability theory while he was dealing with some problems in gambling, but failed.

During 17 the century, chavalier de-mere a great gambler was found, some difficulties in gambling. He turned to pascal and requested him to solve the game. Pascal in consulation with Fermat (Pronounciation - Ferma) looked into problems. They laid the foundation probability theory. Then it was developed by Bernoullie. The other pioneers are Laplace, Bayes etc.

The modern theory of probability using sets was discovered by Russian mathematician A.N. Kolmogorov in 1930-33.

The probability theory is used in budget planning, defence tactis of govt & in gambeling, betting etc.

## Scientific experiment & statistical experiment

**experiment**:

An experiment or trial is called scientific if the outcome is sure & is called statistical if the outcome is random or uncertain.

E.g:- If we take up an experiment

Of mixing 2 parts of hydrogen and 1 part of oxygen then the outcome is surely water. It is scientific expt.

Tossing coin is statistical because the outcome is not sure, tail or head may occurs.

## Outcome or sample point, sample space & event

Imagine an expt or trial ~~treat~~ in which there are several logical possibilities. Each logical possibility is called outcome or sample point.

The set of all logical possibilities is called sample space which is written as S or $\Omega$ (Capital Omega)

An event is a set of some ~~of~~ all possible outcomes & thus a subset of sample space. It is denoted by any capital letter.

E.g :- Consider the expt of rolling of a die. There are 6 outcomes 1, 2, 3, 4, 5, 6

S = Sample space = $\{1, 2, 3, 4, 5, 6\}$

Here 1, 2 etc are outcomes or sample points

Let A be the event of getting an odd no

$A = \{1, 3, 5\}$

## Note :

Sample ~~and pace~~ space is like universal set of set theory. Sample points or outcome is like an element of set theory.

Event is any subset of sample space.

## Equally likely outcomes :

Two outcomes are said to be equally likely if they have equal chance of happening.

Eg: In tossing of coin, if the coin is fair or unbiased then the outcomes head and tail have equal chance of happening & ~~then~~ thus they are equally likely.

## Probability :

Let S be non empty finite sample space & let A be an event. Then the probability of event A is written as P(A) and is defined as

$$P(A) = \frac{\text{No of cases favourable to A}}{\text{Total no. of possible outcomes}}$$

$$= \frac{|A|}{|S|}$$ provided that the outcomes ~~are~~ equally likely.

**Ex:** In the rolling die, what is the probability of getting an odd number?

**Sol^n :** S = Sample space = $\{1,2,3,4,5,6\}$

A be the event of getting an odd number.

$\therefore A = \{1,3,5\}$

$|A| = 3$, $|S| = 6$

$\therefore P(A) = \frac{|A|}{|S|} = \frac{3}{6} = \frac{1}{2}$

**Theorem**: Prove that (i) $P(\phi) = 0$ (ii) $P(S) = 1$ (iii) $0 \leq P(A) \leq 1$ where $S$ is non-empty finite sample space & $A$ is any event.

**Proof**: Let $S$ be non-empty finite sample space

(i) $P(\phi) = \frac{|\phi|}{|S|} = \frac{0}{|S|} = 0$

(ii) $P(S) = \frac{|S|}{|S|} = 1$

(iii) Let $A$ be any ~~proper~~ event i.e. $A \subset S$

~~or $A \subseteq S$~~

$\Rightarrow |A| \leq |S|$

$\therefore P(A) = \frac{|A|}{|S|} \leq \frac{|S|}{|S|} = 1$

Also $|A| \geq 0$ (Equality holds for $A = \phi$)

& $|S| > 0$ $(\because S \neq \phi)$

$\therefore \frac{|A|}{|S|} \geq 0 \Rightarrow P(A) \geq 0$

$\therefore 0 \leq P(A) \leq 1$ $\square$

**Certain Event or Sure event**

An event $A$ is called certain if $P(A) = 1$. Ex: Sample space '$S$' is a certain event.

**Impossible event**:

An event $A$ is called impossible if $P(A) = 0$

Ex: $\phi$ is impossible event.

## Mutually exclusive event

Two events A & B are called mutually exclusive or disjoint if they cannot occur simultaneously or if they have no common element i.e $A \cap B = \phi$

Ex :- Getting an odd no. & getting an even number on a throw of die are mutually exclusive events.

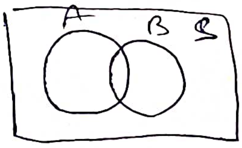
**Thm** (General Addition Theorem)

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

i.e $P(A \text{ or } B) = P(A) + P(B) - P(A \,\&\, B)$

where A, B are any 2 events,

Proof → Let 'S' be the non-empty finite sample space. Let A & B be any two events.




$A \subset S$, $B \subset S$

We know $|A \cup B| = |A| + |B| - |A \cap B|$

$$\therefore \frac{|A \cup B|}{|S|} = \frac{|A|}{|S|} + \frac{|B|}{|S|} - \frac{|A \cap B|}{|S|}$$

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

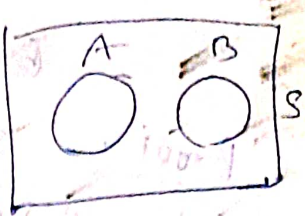
$$\Rightarrow P(A \text{ or } B) = P(A) + P(B) - P(A \,\&\, B)$$

□.

**Thm** ~~Special~~ Special addition theorem

If A & B are mutually exclusive events then $P(A \text{ or } B) = P(A) + P(B)$

**Proof →** Let $S$ be the non empty finite sample space. Let $A$ & $B$ be elements. Also mutually exclusive




$$\therefore A \cap B = \phi$$

Now $|A \cup B| = |A| + |B|$

$$\Rightarrow \frac{|A \cup B|}{|S|} = \frac{|A|}{|S|} + \frac{|B|}{|S|}$$

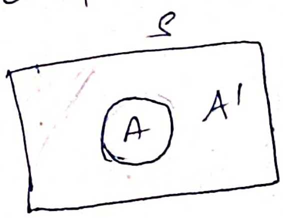
$$\Rightarrow P(A \cup B) = P(A) + P(B) \qquad \square$$

$$\Rightarrow P(A \text{ or } B) = P(A) + P(B)$$

**Corollary - 1**

$$P(\text{not } A) = 1 - P(A)$$

**Proof :** Let $S$ be the non-empty finite sample space. Let $A$ be any event.




$$\therefore A \subset S$$

$$\& \; A \cap A' = \phi$$

Also $A \cup A' = S$

i.e $A$ & $A'$ are mutually exclusive.

$$P(A \cup A') = P(A) + P(A')$$

$$\therefore P(S) = P(A) + P(A')$$

$$\cancel{\Rightarrow P(S)} = P(A) + P(A')$$

$$\Rightarrow 1 = P(A) + P(A')$$

$$\Rightarrow P(A') = 1 - P(A)$$

$$\Rightarrow P(\text{not } A) = 1 - P(A) \qquad \square$$

**Corollary : 2.**

Let $A$ & $B$ be two events

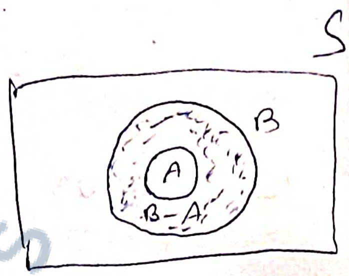
$A \subset B$.

Then (a) $P(A) \leq P(B)$

(b) $P(B-A) = P(B) - P(A)$

**Proof:**

Let $A$ & $B$ be two events in a non-empty finite sample space 'S'.




Given that $A \subset B$

$\therefore$ $(B-A) \cup A = B$ & $(B-A) \cap A = \phi$

i.e. $(B-A)$ & $A$ are mutually exclusive.

$\therefore$ $P(B) = P((B-A) \cup A)$

$= P(B-A) + P(A)$ $\square$

$\Rightarrow$ $P(B-A) = P(B) - P(A)$

But $P(B-A) \geq 0$

$\therefore$ $P(B) - P(A) \geq 0$

$\Rightarrow$ $P(B) \geq P(A)$

$\Rightarrow$ $P(A) \leq P(B)$ $\square$.

## Independent events

Two events $A$ & $B$ are said to be independent if the occurrence or non-occurrence of ~~the other~~. one event does not affect the probability of the occurrence & non-occurrence of the other. Here the word probability is important. Or we omit the defn this word then the defn becomes the defn of mutually exclusive events.

**Ex:** The event of drawing a king & an ace in two success(ive) event in drawing from a pack of 52 cards with replacement is independent (not affected) but without replacement is dependent. (affected)

**Theorem:** If $A$ & $B$ are independent events then $P(A \& B) = P(A) \cdot P(B)$

i.e. $P(A \cap B) = P(A) \cdot P(B)$

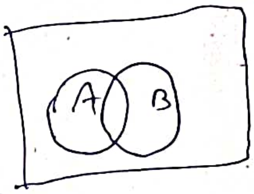
**Q:** If $A$ & $B$ be two independent events then prove that $A'$ & $B$ are independent & $A$ & $B'$ are independent.

Imp.

**Proof:** Given that $A$ & $B$ are independent

$\therefore P(A \cap B) = P(A) \cdot P(B)$

Now $A' \cap B = B - (A \cap B)$




$= (A \cup B)'$ (By De morgan's law)

$\Rightarrow P(A' \cap B') = P((A \cup B)') = 1 - P(A \cup B)$

$= 1 - \{P(A) + P(B) - P(A \cap B)\}$

$= 1 - P(A) - P(B) + P(A \cap B)$

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

$\Rightarrow P(A' \cap B) = P(B - A \cap B)$

$= P(B) - P(A \cap B)$ $\quad (\because A \cap B \subset B$ corollary$)$

$= P(B) - P(A) \cdot P(B)$

$= P(B)\,(1 - P(A)) = P(B) \cdot P(A')$

$\therefore = P(A') \cdot P(B)$

$\therefore A'$ & $B$ are independent.

Q2) If $A$ & $B$ be two indipendent events then prove that $A'$ & $B'$ are independent.

Proof: Given that $A$ & $B$ are indipendent

$\therefore P(A \cap B) = P(A) \cdot P(B)$

Now $A' \cap B' = (A \cup B)'$ (By Demorgan's law)

$\Rightarrow P(A' \cap B') = P((A \cup B)') = 1 - P(A \cup B)$

$= 1 - \{P(A) + P(B) - P(A \cap B)\}$

$= 1 - P(A) - P(B) + P(A \cap B)$

~~Q -) Prove that A & B' are Independent provided A & B are indepent~~

$= (1 - P(A)) - P(B) + P(A) P(B)$

$= (1 - P(A)) - P(B)\,(1 - P(A))$

$= (1 - P(A)) \cdot (1 - P(B))$

$= P(A') \cdot P(B')$

$\therefore A'$ and $B'$ are independent. (Proved)

**Conditional Probability:**

Let $B$ be an event on a sample space. The Conditional probability of $B$ subjected to $A$ denoted by $P(B/A)$ is defined as $P(B/A) = \dfrac{P(B \cap A)}{P(A)}$

**Multiplication theorem**

$P(A) \cdot P(B/A) = P(B \cap A)$

Imp Note :-

P(B|A) Means we have to find the probability of B such that probability of A given ( | stands for such that)

P(A|B) Means we have to find the probability of A such that Probability of B given.

# Statistics

The word statistics seems to have been derived from Latin word status or Italian word statistia or German word statistik each of which means a political state. In ancient ~~farms~~ times the govt. used to collect the information regarding the population & property of the wealth of the country. This helps the govt to have an idea of the man power of the country & the financial pos<sup></sup>

General theory of statistics was discovered after discovery of probability theory. Any how R.A. fisher is known as "Father of statistics"

Before studing details we define some term which will be used in future.

**Variable** : A quantity which has measurable ie characteristic is called a variable or vareate. These are of 2 types.

(a) **Discrete variable** (Discontinous variable) :- A variable which assumes integral value is called a discrete or discontinuous variable.

Ex :- No of students in a class.

Marks secured in Maths in Test exam.

(b) Continuous Variable :- A variable which takes any real value is called a continuous variable.

Ex :- The weight of students in a class, The height of " " " etc.

Sample :- A selected portion of a collection of measurements of a given variable is called a sample. The process of selection is called sampling.

Ex :- A handful of rice from a rice bag constitutes a sample.

Frequency :- The no of times a value of variable occurs is called the frequency of that variable.

Raw data :- If the data is not in an order then it is called raw data or ungrouped data.

Simple frequency distribution (S.f.d)

The arrangement of raw data with corresponding frequencies in an order is called a simple frequency distribution.

Grouped frequency distribution (G.f.d)

The arrangement of raw data showing the class interval & the corresponding frequency in a tabular form is called grouped frequency distribution.

Ex: The marks of 15 students in math are

20, 25, 32, 82, 72, 45, 20, 32, 25, 25, 26, 25, 26, 82, 72

Here the data is Raw data.

It can be arranged & tabulated as follows

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x | 20 | 25 | 26 | 32 | 45 | 72 | 82 |
| f | 2 | 4 | 2 | 2 | 1 | 2 | 2 |

Here x stands for value of variable & f stands for the frequency of corresponding value of a varible.

It is called g.f.d

The above data can also be arranged showing C.I as follows.

| Class-interval C.I | frequency (f) | ~~Cumulative~~ cumulative frequency C.f |
|---|---|---|
| 20-30 | 8 | 8 |
| 30-40 | 2 | 10 |
| 40-50 | 1 | 11 |
| 50-60 | 0 | 11 |
| 60-70 | 0 | 11 |
| 70-80 | 2 | 13 |
| 80-90 | 2 | 15 |

It is called G.f.d

## ~~Chumulative~~ Commulative frequency.

C.f is obtained by adding each frequency to the sum of the previous ones.

## Class-intervals

These are of two types : exclusive class interval & inclusive class interval

10-20, 20-30, - - - - - - - etc are exclusive classes.

The classes 10-20 means [10, 20)

20-30 means [20, 30) etc.

Here in the class 10-20, 10 is called lower limit & 20 is called upper limit of the class 10-20.

The difference between lower & upper limits is called class size or class width

Here it is 10.

Class mid mark is the value mid way between upper & lower limit. Mid mark is here 15 for class 10-20,

$$\text{Mid mark} = \frac{\text{Lower limit} + \text{upper limit}}{2}$$

The frequency of any class is centered at its class mid-mark.

The class interval 10-19, 20-29 are called inclusive class intervals.

The class 10-19 means [10, 19]. Here the class size is 10. Lower limit is 10 & upper limit is 19

Class mid mark $= \frac{10+19}{2} = 14.5$

We can make inclusive class to exclusive. Suppose inclusive classes are 10-19, 20-29 etc. Here the difference of upper limit of one class & lower limit of next class is 1. Thus consider $1/2 = .5$.

Now the corresponding exclusive classes are 9.5 - 19.5, 19.5 - 29.5 etc

After conversion to exclusive classes the mid mark is not changed.

## Measures of central tendency or Measures of location

~~Poon~~ Plainly speaking a measure of central tendency of a statistical data is the value of the variable which is the general representative of entire distribution.

There are 6 measures of central tendency.

1 - Arithmetic mean or Mean
2 - Mode
3 - Median
4. Geometric mean 5. Harmonic mean 6. Weighted mean

## Arithmetic Mean or Mean of raw data

If $x_1, x_2 - - - x_n$ be $n$ variables then their mean is written as $\bar{x}$ & is defined by

$$\bar{x} = \frac{x_1 + x_2 + - - - - + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

## Mean of S.f.d

If $x_1, x_2 - - - x_n$ are $n$ variables & $f_1, f_2 - - - f_n$ are their respective frequencies then their mean $\bar{x}$ is given by

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + - - - + x_n f_n}{f_1 + f_2 + - - - + f_n}$$

$$= \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

## Mean of G.f.d

If $m_1, m_2, - - - m_n$ are the midvalues of C.I and $f_1, f_2 - - f_n$ are their respective frequencies then $\bar{x}$ is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} m_i f_i}{\sum_{i=1}^{n} f_i}$$

Sometimes the calculation of mean by the above methods becomes tedious so a suitable number is chosen from set of variables. It is called '*working mean* or *Assumed Mean*'. (~~Deviat~~

Them-1 (Deviation method)

Let the working mean be A. Let $d_i = x_i - A$ be the deviation, then formula

$$\bar{x} = A + \frac{\sum_i f_i d_i}{\sum_i f_i}$$

(Deviation Method)

* Proof: Given that $d_i = x_i - A$

$$\therefore \quad \bar{x} = \frac{\sum_i f_i x_i}{\sum_i f_i} = \frac{\sum_i f_i (d_i + A)}{\sum_i f_i}$$

$$= \frac{\sum_i f_i d_i}{\sum_i f_i} + \frac{\sum_i f_i A}{\sum f_i} = \frac{\sum f_i d_i}{\sum f_i} + \frac{A \sum_i f_i}{\sum_i f_i}$$

$$= A + \frac{\sum_i f_i d_i}{\sum f_i} \quad \square.$$

Them 2 (Step deviation method)

If $x_1, x_2, \ldots\ldots x_n$ be the mid value of which $f_1, f_2, \ldots\ldots f_n$ are frequencies. Suppose A is working mean, C is the width of C.I then

$$\bar{x} = A + \frac{C \cdot \sum f_i U_i}{\sum f_i} \quad \text{where } U_i = \frac{x_i - A}{C}$$

**Proof** :- $U_i = \dfrac{x_i - A}{C}$

$\Rightarrow C U_i = x_i - A$

$\Rightarrow C U_i + A = x_i$

$$\therefore \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i (C U_i + A)}{\sum f_i}$$

$$= \frac{C \sum f_i U_i}{\sum f_i} + \frac{A \sum f_i}{\sum f_i}$$

$$= A + C \frac{\sum f_i U_i}{\sum f_i} \quad \square$$

Q → Find the mean of first n natural numbers.

**Ans** :- The 1st n natural nos are 1, 2, 3, 4 - - - - n. It is raw data.

$$\text{Mean} = \bar{x} = \frac{1+2+3 - - - + n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

Q → Find the mean of the following data

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Varianu | 5 | 4.5 | 4 | 3.5 | 3 | 2.5 | 2 | 1 | 0 |
| frequency | 1 | 3 | 4 | 1 | 5 | 2 | 4 | 3 | 2 |

Soln  A = Working mean = 3

| $x_i$ | $f_i$ | $d_i = x_i - 3$ | $f_i d_i$ |
|---|---|---|---|
| 5 | 1 | 2 | 2 |
| 4.5 | 3 | 1.5 | 4.5 |
| 4 | 4 | 1 | 4 |
| 3.5 | 1 | .5 | .5 |
| 3 | 5 | 0 | 0 |
| 2.5 | 2 | -.5 | -1 |
| 2 | 4 | -1 | -4 |
| 1 | 3 | -2 | -6 |
| 0 | 2 | -3 | -6 |
| | $\sum f_i = 25$ | | $\sum f_i d_i = -6$ |

$$\therefore \bar{x} = A + \frac{\sum f_i d_i}{\sum f_i} = 3 - \frac{6}{25} = \frac{69}{25}$$

$$= 2.76$$

Q→ Find the mean of the following data.

| 20-40 | 40-60 | 60-80 | 80-100 | 100-120 | 120-140 | 140-160 | 160-180 | 180-200 |
|---|---|---|---|---|---|---|---|---|
| 6 | 9 | 12 | 14 | 20 | 15 | 10 | 7 | 7 |

Soln  A = working Mean = 110

c = class width = 20

| C.I | Frequency ($f_i$) | M.e $y_i$ | $U_i = \frac{y_i - 110}{20}$ | $f_iU_i$ |
|---|---|---|---|---|
| 20-40 | 6 | 30 | -4 | -24 |
| 40-60 | 9 | 50 | -3 | -27 |
| 60-80 | 12 | 70 | -2 | -24 |
| 80-100 | 14 | 90 | -1 | -14 |
| 100-120 | 20 | 110 | 0 | 0 |
| 120-140 | 15 | 130 | 1 | 15 |
| 140-160 | 10 | 150 | 2 | 20 |
| 160-180 | 7 | 170 | 3 | 21 |
| 180-200 | 7 | 190 | 4 | 28 |
| | $\sum f_i = 100$ | | | $\sum f_iU_i = -5$ |

$$\bar{x} = A + c\frac{\sum f_iU_i}{\sum f_i} = 110 + \frac{20 \times -5}{100}$$

$$= 110 - 1 = 109 \text{ (Ans)}$$

## Median

Def: Median is the middle most value of a given set of values.

Median of Raw-data

It can be found by the following steps.

Step 1: Arrange the data in increasing or decreasing order.

Step 2: of $n$ is the no of values

"n" is odd then median is the $\left(\frac{n+1}{2}\right)^{th}$ value.

If is n is even, median is the mean of $\left(\frac{n}{2}\right)^{th}$ & $\left(\frac{n}{2}+1\right)^{th}$ value.

## Median of simple frequency distribution

Step-1 → Write the distribution in a tabular form with frequency & cumulative frequency column.

Step-2 : If the total no of frequency $N = \sum f_i$ is odd then taken $\frac{N+1}{2}$

If N is even take $N/2$

Step-3

Find the corresponding value from the variable column.

Then it is median.

## Median of Grouped frequency distribution

Step-1 → Construct the table for distribution showing frequency & C.F column.

Step 2 : Find the total no of frequency $= \sum f_i = N$ & find $N/2$ (If N is even or odd)

**Step-3** Find the Corresponding median class.

**Step-4** Calculate $j = N/2 - C.F$ of previous class

**Step-5** $$\text{Median} = L_{med} + \frac{j \times c}{f_{med}}$$

Where $L_{med}$ = Lower limit of median class.

$f_{med}$ = Frequency of median.

$c$ = class size.

**Example** :- Find the median of 15, 7, 22, 13, 27, 17, 13

**Soln** :- First we arrange in ascending order
7, 13, 13, 15, 17, 22, 27

No of obsⁿs is ~~even~~ 7 i.e odd

Median is $\left(\frac{7+1}{2}\right)^{th}$ i.e $4^{th}$ observation which is 15.

**Q** :- Find the median of 17, 11, 13, 15, 7, 9

**Soln** First we arrange ~~a~~ in decreasing order
17, 15, 13, 11, 9, 7

No of obsⁿs is 6 i.e even.

Median is mean of $\frac{6}{2}$ = 3rd and 4th

Obs ~~n~~ & mean of 13 & mean ~~of~~

$= \frac{13+11}{2} = 12$

Q → Find the median of the following data

| | | | | | |
|---|---|---|---|---|---|
| 180 | 150 | 120 | 100 | 80 | 200 |
| 10 | 20 | 30 | 16 | 26 | 6 |

Sol. Arrange the data in tabular form

| $x_i$ | $f_i$ | c.f |
|---|---|---|
| 180 → | 10 | 10 |
| 150 → | 20 | 30 |
| 120 → | 30 | 60 |
| 100 → | 16 | 76 |
| 80 → | 26 | 102 |
| 200 → | 6 | 108 |
| | $\Sigma f_i = 108$ | |

Here $N = 108$ which is even. Take $N/2 = \frac{108}{2} = 54$. which is written in c.f 60. The corresponding value from variable column 120

∴ Median = 120

Q → Find the median of the following data

| | | | | | | |
|---|---|---|---|---|---|---|
| 40 | 45 | 50 | 55 | 60 | 65 | 70 |
| 5 | 7 | 8 | 15 | 4 | 3 | 1 |

| $x_i$ | $f_i$ | c.f |
|---|---|---|
| 40 → | 5 → | 5 |
| 45 → | 7 → | 12 |
| 50 → | 8 → | 20 |
| 55 → | 15 → | 35 |
| 60 → | 4 → | 39 |
| 65 → | 3 → | 42 |
| 70 → | 1 → | 43 |
| | $\Sigma f_i = 43$ | |

Here $N = 43$ which is odd.

∴ $\frac{N+1}{2} = 22$ which is written in c.f 35 & corresponding value from variable column = 55

∴ Median = 55

Q→ Find the median of following data

| | | | | |
|---|---|---|---|---|
| 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
| 8 | 12 | 16 | 35 | 29 |

Soln

| C.I | | $f_i$ | | C.f |
|---|---|---|---|---|
| 0-10 | → | 8 | → | 8 |
| 10-20 | → | 12 | → | 20 |
| 20-30 | → | 16 | → | 36 |
| $\boxed{30-40}$ | → | 35 | → | $\boxed{71}$ |
| 40-50 | → | 29 | → | 100 |

Now $N/2 = 50$ which is written CF 71

∴ The corresponding median Class = 30-40

∴ $L_{med} = 30$, $f_{med} = 35$, C = Class width = 10.

∴ $f = N/2 -$ C.f of previous class

$= 50 - 36 = 14$

∴ Median $= L_{med} + \frac{f \times C}{f_{med}} = 30 + \frac{14 \times 10}{35} = 34$

# Mode

Defn Mode or modal value is that value which occures most often (i.e. which has the greatest frequency)

If there are 2 modes then the distribution is called 'Bimodal'. If each observation occures only once or equally often then the distribution has no mode.

For grouped distribution, find the Cla[ss] corresponding to max$^m$ frequency. It is called modal class. The mark of the modal class as the mode.

There is an emperical relation between mean, mode and median

$$\boxed{\text{Mode} = \text{Mean} - 3(\text{Mean} - \text{Median})}$$

$$\boxed{\text{i.e. mode} = 3\ \text{median} - 2\ \text{mean}.}$$

Mode of a g.f.d can be also calculated by

$$\text{Mode} = l_1 + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2}(l_2 - l_1)$$

where $l_1$ is the lower limit of modal form

$l_2$ " " upper limit " " "

$F_1$ " " frequency " " "

$F_0$ " " " of the class preceding (or) modal form class

$f_2$ " " " " (or) following " "

Provided the classes are exclusive.

Note →

In symmetrical (bell shaped) mean, median, mode coincide

Mode = 3 median − 2 mean

Graph of bell shaped distribution.

Q→ Find the mode of the following data

| | | | | | |
|---|---|---|---|---|---|
| 40 | 43 | 52 | 55 | 58 | 59 |
| 2 | 2 | 3 | 4 | 2 | 2 |

Soln

| $x_i$ | | $f_i$ |
|---|---|---|
| 40 | → | 2 |
| 43 | → | 2 |
| 52 | → | 3 |
| 55 | → | 4 |
| 58 | → | 2 |
| 59 | → | 2 |

Here 55 has the greatest frequency.

∴ Mode = 55

Q→ Find the mode of following data

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 30-32 | 33-35 | 36-38 | 39-41 | 42-44 | 45-47 | 48-50 | 51-53 |
| 3 | 4 | 10 | 16 | 12 | 8 | 4 | 2 |

Soln

| Class limit | | freq |
|---|---|---|
| 30 - 32 | → | 3 |
| 33 - 35 | → | 4 |
| 36 - 38 | → | 10 |
| 39 - 41 | → | 16 |
| 42 - 44 | → | 12 |
| 45 - 47 | → | 8 |
| 48 - 50 | → | 4 |
| 51 - 53 | → | 2 |

Here the max frequencies is 16 which occurs in the class 39-41.

∴ Modal class is 39-41.

Making it exclusive we have the class as 38.5 - 41.5

$L_1$ = Lower limit of modal class = 38.5.

$L_2$ = Upper " " " " = 41.5.

$f_1$ = Frequency of " " " = 16.

$f_0$ = Frequency of previous class = 10.

$f_2$ = " " following class = 12

$$\therefore \text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1)$$

$$= 38.5 + \frac{16-10}{32-10-12}(41.5-38.5)$$

$$= 38.5 + \frac{6}{10} \times 3$$

$$= 40.3$$

Sat - 4.30 PM

## Measures of dispersion

Suppose marks secured in 5 subjects by 2 students are given below

A : 38, 36, 37, 35, 39

B : 02, 70, 03, 60, 50

Mean of each is 37.

But A is more consistent than B.

The dispersion is defined as the scattering or variate values around their central value.

More is the dispersion less is the consistent.

There are 3 measures of dispersion,

① Range ② Mean deviation (MD)
③ Standard deviation (S.d)

① **Range** :- It is a distribution which is defined as the difference between the highest & lowest values of the variable.

More range implies more dispersion.
Less " " less dispersion.

In ex :- Range of marks secured by B is $70-2=68$ but range of marks secured by A is $39-35=4$

$\therefore$ A is more consistent than B.

2. **Mean deviation from mean**

**M.D for Raw data**

$$M.D = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$ where $\bar{x}$ is mean.

**M.D for S.f.d**

$$M_D = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

M.D for g.f.d

$$M.D = \frac{\sum_{i=1}^{n} f_i \left| y_i - \bar{x} \right|}{\sum_{i=1}^{n} f_i}$$

where $y_i \{i=1,2,\ldots n\}$ are class mid marks with respective frequency $f_i$.

Q → Find M.D of $a, a+d, a+2d$

Card - - - - - $a+2nd$

Sol $a, a+d, a+2d$ - - - $a+2nd$

are in AP.

Mean $\bar{x} = \frac{(\text{first}+\text{last}) \text{ term}}{2} = \frac{a+a+2nd}{2}$

$= a+nd$

| $x_i$ | $\lvert x_i - \bar{x} \rvert = \lvert x_i - a - nd \rvert$ |
|---|---|
| $a$ | $nd$ |
| $a+d$ | $(n-1)d$ |
| $a+2d$ | $(n-2)d$ |
| $\vdots$ | $\vdots$ |
| | $0$ |
| $a+nd$ | $\vdots$ |
| $a+2nd$ | $nd$ |

$$\therefore M.D = \frac{\sum_{i=1}^{2n+1} \left| x_i - \bar{x} \right|}{2n+1}$$

$$= \frac{2\{nd + (n-1)d + \dots + 2d + d\}}{2n+1}$$

$$= \frac{2d\{1+2+\dots+n\}}{2n+1} = \frac{2dn(n+1)}{2(2n+1)}$$

$$= \frac{n(n+1)}{(2n+1)} d$$

**Q4** Find the M.D from mean of the following data.

| | | | | | |
|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 |
| 1 | 3 | 7 | 6 | 2 | 1 |

**Sol**

| $x_i$ | $f_i$ | $f_i x_i$ | $\lvert x_i - \bar{x} \rvert$ | $\lvert x_i - 2.4 \rvert . f_i$ |
|---|---|---|---|---|
| 0 | 1 | 0 | 2.4 | 2.4 |
| 1 | 3 | 3 | 1.4 | 4.2 |
| 2 | 7 | 14 | .4 | 2.8 |
| 3 | 6 | 18 | .6 | 3.6 |
| 4 | 2 | 8 | 1.6 | 3.2 |
| 5 | 1 | 5 | 2.6 | 2.6 |
| $\Sigma f_i = 20$ | | $\Sigma f_i x_i = 48$ | | $\Sigma f_i \lvert x_i - \bar{x} \rvert = 18.8$ |

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{48}{20} = 2.4$$

$$\text{M.e.d from mean} = \frac{\sum_i f_i \, |x_i - \bar{x}|}{\sum f_i}$$

$$= \frac{18.8}{20} = 0.94$$

Q-) Find the M.D from mean, of the following data.

| 0-20 | 20-40 | 40-60 | 60-80 | 80-100 |
|---|---|---|---|---|
| 5 | 5 | 20 | 15 | 5 |

Sol^n

| C.I | Class mark $y_i$ | frequency $f_i$ | $f_i y_i$ | $\lvert y_i - \bar{x}\rvert = \lvert y_i - 54\rvert$ | $f_i \lvert y_i - \bar{x}\rvert$ |
|---|---|---|---|---|---|
| 0-20 | 10 | 5 | 50 | 44 | 220 |
| 20-40 | 30 | 5 | 150 | 24 | 120 |
| 40-60 | 50 | 20 | 1000 | 4 | 80 |
| 60-80 | 70 | 15 | 1050 | 16 | 240 |
| 80-100 | 90 | 5 | 450 | 36 | 180 |
| | | $\sum f_i = 50$ | $\sum f_i y_i = 2700$ | | $\sum f_i (y_i - \bar{x}) = 840$ |

$$\bar{x} = \frac{\sum f_i y_i}{\sum f_i} = \frac{2700}{50} = 54$$

$$\text{M.D from mean} = \frac{\sum f_i \, |y_i - \bar{x}|}{\sum f_i}$$

$$= \frac{840}{50} = 16.8$$

## Mean deviation from Median

**M.D for raw data**

$$M.D = \frac{\sum_{i=1}^{n} |x_i - a|}{n}$$

where a is median.

**M.D for sfd**

$$M.D = \frac{\sum_{i=1}^{n} f_i |x_i - a|}{\sum_{i=1}^{n} f_i}$$

**M.D for Gfd**

$$M.D = \frac{\sum_{i=1}^{n} f_i |y_i - a|}{\sum_{i=1}^{n} f_i}$$

where $y_i = (i = 1, 2, 3 \dots n)$ are class mid marks with respective frequency $f_i$

Q → Find M.D from median of the following data

| | | | | | |
|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 |
| 1 | 3 | 7 | 6 | 2 | 1 |

**Sol^n**

| $x_i$ | $f_i$ | C.F | $\lvert x_i - a\rvert = \lvert x_i - 2\rvert$ | $f_i \lvert x_i - a\rvert = f_i \lvert x - 2\rvert$ |
|---|---|---|---|---|
| | | | 2 | 2 |
| 0 | 1 | 1 | 1 | 3 |
| 1 | 3 | 4 | 0 | 0 |
| 2 | 7 | 11 | 1 | 6 |
| 3 | 6 | 17 | 2 | 4 |
| 4 | 2 | 19 | 3 | 3 |
| 5 | 1 | 20 | | |

$N = \sum f_i = 20$

$\sum f_i |x_i - a| = 18$

$N/2 = 10$ which is within CF 11.

The corresponding value from variable column $= 2$

$\therefore$ Median $= a = 2$

$$\text{M.D from median} = \frac{\sum_{i}^{n} f_i |x_i - a|}{\sum f_i} = \frac{18}{20} = 0.9$$

## Standard deviation

It is denoted by $\sigma$. Its square denotes $(\sigma)^2$ is called variance. The +ve sq root of variance is S.d

$\Sigma \frac{(x_i - \bar{x})^2}{n}$

S.d for Raw data

$\frac{\Sigma x_i^2}{n} - \bar{x}^2$

$$\sigma = + \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

S.d for S.f.d

$$\sigma = + \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i x_i}{\Sigma f_i}\right)^2}$$

S.d for G.F.D

$$\sigma = C \sqrt{\frac{\Sigma f_i U_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i U_i}{\Sigma f_i}\right)^2}$$

where $C$ = Class size

$$U_i = \frac{y_i - A}{C}$$

$y_i$ is class mid mark, $A$ is assumed mean.

# Co-efficient of variation – C.V

$C.V = \cancel{100 \times \frac{\sigma}{\bar{x}}}$

$C.V\% = 100 \times \frac{\sigma}{\bar{x}}$

C. dispersion $= \frac{\sigma}{\bar{x}}$

Q) Find the S.d & variance of the following 7, 10, 12, 13, 15, 20, 21, 28, 29, 35

Soln

| $x_i$ | $x_i^2$ |
|---|---|
| 7 | 49 |
| 10 | 100 |
| 12 | 144 |
| 13 | 169 |
| 15 | 225 |
| 20 | 400 |
| 21 | 441 |
| 28 | 784 |
| 29 | 841 |
| 35 | 1225 |

$\sum x_i = 190$, $\sum x_i^2 = 4378$

$n = 10$, $\therefore \sigma = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$

$= \sqrt{\frac{4378}{10} - \left(\frac{190}{10}\right)^2}$

$= \sqrt{437.8 - 361} = \sqrt{76.8}$

$= 8.76$.

VarEance $= \sigma^2 = 76.8$

Q) Find the S.d & Vareence of

1st n natural numbers.

$x_i = 1, 2, 3 \dots\dots n$

$x_i^2 = 1^2 \quad 2^2 \quad 3^2 \dots\dots n^2$

Varience $= \sigma^2 = \dfrac{\sum x_i^2}{n} - \left(\dfrac{\sum x_i}{n}\right)^2$

$$= \frac{1^2+2^2+\dots\dots+n^2}{n} - \left(\frac{1+2+3\dots\dots+n}{n}\right)^2$$

$$= \frac{\cancel{n}(n+1)(2n+1)}{6\cancel{n}} - \left[\frac{\cancel{n}(n+1)}{2\cancel{n}}\right]^2$$

$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$= \frac{n+1}{12}\left[4n+2-3n-3\right]$$

$$= \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$$

S.d $= \sqrt{\dfrac{n^2-1}{12}}$

Q-) Find the S.d of Varience of the following data

| | | | | | |
|---|---|---|---|---|---|
| 3 | 5 | 7 | 9 | 11 | 13 |
| 2 | 7 | 10 | 9 | 5 | 1 |

P.T.O

| $x_i$ | $f_i$ | $f_i x_i$ | $x_i^2$ | $f_i x_i^2$ |
|---|---|---|---|---|
| 3 | 2 | 6 | 9 | 18 |
| 5 | 7 | 35 | 25 | 175 |
| 7 | 10 | 70 | 49 | 490 |
| 9 | 9 | 81 | 81 | 729 |
| 11 | 5 | 55 | 121 | 605 |
| 13 | 1 | 13 | 169 | 169 |
| | $\Sigma f_i = 34$ | $\Sigma f_i x_i = 260$ | | $\Sigma f_i x_i^2 = 2186$ |

$$\sigma = \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i x_i}{\Sigma f_i}\right)^2}$$

$$= \sqrt{\frac{2186}{34} - \left(\frac{260}{34}\right)^2} = \sqrt{5.93} \approx 2.43$$

Varience $\sigma^2 = 5.93$

Q → Find varience & S.d of the following

| 1-5 | 6-10 | 11-15 | 16-20 | 21-25 |
|---|---|---|---|---|
| 4 | 7 | 10 | 5 | 4 |

Sum $C$ = Class size = 5

$a$ = assumed mean = 13

| C.I | $f_i$ | Class mid mark $y_i$ | $U_i = \frac{y_i - a}{c}$ | $f_i U_i$ | $f_i U_i^2$ |
|---|---|---|---|---|---|
| 1-5 | 4 | 3 | -2 | -8 | 16 |
| 6-10 | 7 | 8 | -1 | -7 | 7 |
| 11-15 | 10 | 13 | 0 | 0 | 0 |
| 16-20 | 5 | 18 | 1 | 5 | 5 |
| 21-25 | 4 | 23 | 2 | 8 | 16 |
| | $\Sigma f_i = 30$ | | | $\Sigma f_i U_i = -2$ | $\Sigma f_i U_i^2 = 44$ |

$$\sigma = c\sqrt{\frac{\sum f_i u_i^2}{\sum f_i} - \left(\frac{\sum f_i u_i}{\sum f_i}\right)^2}$$

$$= 5\sqrt{\frac{44}{30} - \left(\frac{-2}{30}\right)^2} = 5\sqrt{\left(\frac{44}{30} - \frac{4}{900}\right)}$$

$$= 5 \cdot \sqrt{\frac{329}{225}} = 5 \times \sqrt{1.4622}$$

$$= 5 \times 1.209 = 6.045$$

$$\text{Varianc} = \sigma^2 = (5)^2 \times (1.4622) = 25 \times 1.4622$$

$$=$$

Q. Find the Co-efficient of Variation of following data.

| | | | | | |
|---|---|---|---|---|---|
| 3 | 5 | 7 | 9 | 11 | 13 |
| 2 | 7 | 10 | 9 | 5 | 1 |

Soln It is found that $\sigma = 2.43$

$$\bar{x} = \frac{\sum_i f_i x_i}{\sum f_i} = \frac{260}{34} = 7.64$$

$$C.V = \frac{100 \times \sigma}{\bar{x}} = \frac{100 \times 2.43}{7.64} = \frac{243}{7.64}$$

$$= \frac{24300}{764}$$

Ex-9-e

5.

| $x_i$ | $f_i$ | $f_i x_i$ | $x_i^2$ | $f_i x_i^2$ |
|---|---|---|---|---|
| 0 | $^nC_0$ | 0 | 0 | 0 |
| 1 | $^nC_1$ | $^nC_1$ | 1 | $^nC_1$ |
| 2 | $^nC_2$ | $2\,^nC_2$ | $2^2$ | $2^2\,^nC_2$ |
| ⋮ | | | | |
| n | $^nC_n$ | $n\,^nC_n$ | $n^2$ | $n^2\,^nC_n$ |

$B = \{bgg, ggb, gbg, ggg\}$

$A \cap B = \{bgg, ggb, gbg\}$

$|A \cap B| = 3$, $P(A \cap B) = 3/8$, $P(A) = 7/8$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{3/8}{7/8} = 3/7$$

22. (i) $P(A) = 1/2$, $P(B) = \frac{1}{3}$, $P(C) = 1/4$

$P(A^c) = 1/2$, $P(B^c) = \frac{2}{3}$, $P(C^c) = 3/4$

Let D be the event of exactly one hits & ~~two~~ other ~~intersect~~. missed

$$P(D) = P(A \cap B^c \cap C^c) + P(A^c \cap B \cap C^c) + P(A^c \cap B^c \cap C)$$

$$= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} = \frac{11}{24}$$

(ii) Prob of A hits when 2 other missed

~~$= P(A/B)$~~ $$P(A/D) = \frac{P(A \cap D)}{P(D)} = \frac{P(A \cap B^c \cap C^c)}{P(D)}$$

$$= \frac{\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4}}{\frac{11}{24}} = \frac{6/24}{11/24} = \frac{6}{11}$$

## Exercise – 9(d)

2. Hints (I)

$P(1b)$ = $P(1) \cdot P(b/1) = \frac{1}{3} \times \frac{3}{8} = \frac{1}{8}$

$P(2b) = \frac{1}{3} \cdot \frac{5}{9} = 5/27$

$P(3b) = \frac{1}{3} \times \frac{6}{9} = 6/27$

$Prob = \frac{3}{24} + \frac{5}{27} + \frac{6}{27} = \frac{115}{216}$

(ii) Acc. to question

Prob of 3rd bag $= P$

" " 2nd " $= 2P$

" " 1st " $= 4P$

$\therefore P + 2P + 4P = 1$

$\Rightarrow P = 1/7$

$\therefore$ Prob of 1st bag $= \frac{4}{7}$

" " 2nd bag $= \frac{2}{7}$

" " 3rd " $= \frac{1}{7}$

Prob of ball is black from 1st bag $= P(1b)$

" " " " " " 2nd " $= P(2b)$

" " " " " " 3rd " $= P(3b)$

Now $P(1b) = \frac{4}{7} \times \frac{3}{8}$, $P(2b) = \frac{2}{7} \times \frac{5}{9}$

$P(3b) = \frac{1}{7} \times \frac{6}{9}$

$\therefore$ Prob of black ball $= \frac{4}{7} \times \frac{3}{8} + \frac{2}{7} \times \frac{5}{9} + \frac{1}{7} \times \frac{6}{9}$

V.Imp Q → What is the prob of a leap year containing 53 Sundays?

Soln : In leap year (366 days) there are 52 complete weeks & 2 days over.

The following possible combinations for these 2 over days

(i) Sun & Mon (ii) Mon & Tue (iii) Tue & Wed

(iv) Wed & Thu (v) Thu & Fri (vi) Fri & Sat.

(vii) Sat & Sun,

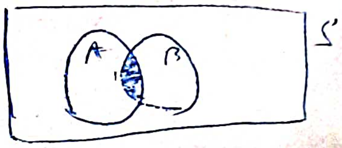
In order that a leap year is selected with 53 Sundays one of the two over days must be a Sunday. Thus are 2 possibilities

$\therefore$ Probability $= \frac{2}{7}$

---

Q → If A & B are independent events then prove that A & B' are independent events.

**Proof :**




Given that A & B are independent events.

~~ie P(A~~ ie $P(A \cap B) = P(A) \cdot P(B)$

Now, $A \cap B' = A - (A \cap B)$

$P(A \cap B') = P(A - A \cap B)$

$= P(A) - P(A \cap B)$ $(\because A \cap B \subset A)$

$= P(A) - P(A) \cdot P(B)$ ($\because$ A & B are independent events)

$= P(A)(1 - P(B))$

$= P(A) \cdot P(B')$

$\therefore$ A & B' are independent events

(Proved.

# Playing Cards.

There $\underline{52 \text{ cards}}$ in a packet.

Red 26 cards:

- 13 → ♥ → Heart
- 13 → ♦ → Diamond.

Black 26 Cards:

- 13 → ♠ → Spades
- 13 → ♣ → Club

16 face cards:

- 4 → A → Ace
- 4 → ~~King~~ K → King
- 4 → Q → Queen
- 4 → J → Jack

Honour Card → King, Queen, Jack, Ace.

Face/Court Card → King, Queen, Jack

Suits → 4

denominants → Same lable like 4 King, 4 Queen, 4 Ace
4 → 2H, 2D, 2S, 2C

## Notes

Probability of number divivible by 2 = $\frac{1}{2}$

" " " 3 = $\frac{1}{3}$

" " " 2 and 3 = ~~both~~ $\frac{1}{\text{L.C.M of 2 \& 3}}$