Annamacharya Institute of Technology & Sciences: Tirupati

(Autonomous)

PROBABILITY AND STATISTICS

Subject Code: 20ABS9911

(AK20 Regulation)
(Common to CSE, AI&DS)

Annamacharya Institute of Technology & Sciences: Tirupati

(Autonomous)

PROBABILITY AND STATISTICS

Subject Code: 20ABS9911

(AK20 Regulation)
(Common to CSE, AI&DS)

MEASURES OF DISPERSION

Introduction:

Dispersion is defined as deviation or scattering of values from their central values i.e, average (Mean, Median or Mode but preferably Mean or Median). In other words, dispersion measures the degree or extent to which the values of a variable deviate from its average.

Two distributions may have:

i. Same central tendency and same dispersion
ii. Different central tendency but same dispersion
iii. Same central tendency but different dispersion
iv. Different central tendency and different dispersion

Definition: The degree to which numerical data tend to spread about an average value is called variation or dispersion or spread of the data.

The measures of dispersion in common use are:

(i) Range
(ii) Mean Deviation
(iii) Standard Deviation

(I) RANGE: Calculation of Range:

For ungrouped data:

Range = Highest Value – Lowest Value. i.e, (H – L).

For grouped frequency distribution:

Range = Upper boundary of last class – Lower boundary of 1stclass

Problem1: Compute the range for the following observation 15, 20, 25, 25, 30, 35.

Solution: Range = Largest value – Smallest value

i.e., 35-15=20

Problem 2: The following table gives the daily sales (Rs.) of two firms A and B for five days.

Firm A	**5050**	**5025**	**4950**	**4835**	**5140**
Firm B	**4900**	**3100**	**2200**	**1800**	**13000**

Solution: The sales of both the firms in average are same but distribution pattern is not similar. There is a great amount of variation in the daily sales of the firm B than that of the firm A

Range of sales of firm A = Greatest value – Smallest value = 5140-4835=305

Range of sales of firm B = Greatest value – Smallest value = 13000-1800=11200

MEAN DEVIATION:

Mean deviation is defined as arithmetic average of absolute values of the deviations of the variates measured from an average (median, mode or mean).

The absolute value of the deviation denoted by | deviation | is the numerical value of the deviation with positive sign.

Note: Mean deviation can be similarly calculated by taking deviations from the median or mode.

Mean Deviation from Mean of an Ungrouped Data:

Let $x_1, x_2, \ldots, x_n$ be the values of n variates and $\bar{x}$ be their arithmetic mean. Let $|x_i - \bar{x}|$ be the absolute value of the deviation of the variate x_i from $\bar{x}$.

$$\therefore \text{Mean deviation} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Problem 1: Calculate the mean deviation of the variates 40, 62, 54, 68, 76 from A.M

Solution: A.M = $\bar{x} = \frac{40+62+54+68+76}{5} = \frac{300}{5} = 60$

$$\therefore \text{Mean deviation} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{\sum_{i=1}^{5} |x_i - 60|}{5} = \frac{|40-60|+|62-60|+|54-60|+|68-60|+|76-60|}{5}$$

$= 52/5 = 10.4$

Problem 2: Find the mean deviation from the mean for the following data: 38, 70, 48, 40, 42, 55, 63, 46, 54, 44.

Solution: Mean $\bar{x} = \frac{38+70+48+40+42+55+63+46+54+44}{10} = \frac{500}{10} = 50$

$$\text{Mean deviation from the mean} = \frac{\sum_{i=1}^{10} |x_i - \bar{x}|}{10} =$$

$$\frac{|38-50|+|70-50|+|48-50|+|40-50|+|42-50|+|55-50|+|63-50|+|46-50|+|54-50|+|44-50|}{10}$$

$= 84/10 = 8.4$

Problem 3: Find the mean deviation from the median for the data 34, 66, 30, 38, 44, 50, 40, 60, 42, 51.

Solution : Arranging the data in ascending order, we have :

30, 34, 38, 40, 42, 44, 50, 51, 60, 66 (n=10 terms)

Now Median $(M_d) = \dfrac{\left(\frac{n}{2}\right)^{th} term + \left(\frac{n}{2}+1\right)^{th} term}{2}$

= (42+44)/2 = 43.

∴ Mean deviation from the median = $\dfrac{\sum_{i=1}^{10}|x_i - Median|}{10} = \dfrac{\sum_{i=1}^{10}|x_i - 43|}{10}$

$$\frac{|30-43|+|34-43|+|38-43|+|40-43|+|42-43|+|44-43|+|50-43|+|51-43|+|60-43|+|66-43|}{10}$$

= 87/10 = 8.7

Mean Deviation for a Grouped Data:

We know that data can be arranged as a frequency distribution in two ways

(i) Discrete Frequency Distribution and

(ii) Continuous Frequency Distribution

Mean Deviation from mean for Discrete Frequency Distribution:

Let $x_1, x_2, \ldots\ldots, x_n$ be the midvalues of n class intervals with frequencies $f_1, f_2, \ldots\ldots, f_n$ of a frequency distribution. Let $\bar{x}$ be the arithmetic mean of the distribution. Let $|x_i - \bar{x}|$ be the absolute value of the deviation of the midvalue x_i from the arithmetic mean $\bar{x}$

Then the mean deviation about the arithmetic mean

$$= \frac{|x_i - \bar{x}|f_1 + |x_i - \bar{x}|f_2 + |x_i - \bar{x}|f_3 + \ldots\ldots + |x_i - \bar{x}|f_n}{f_1 + f_2 + f_3 + \ldots\ldots\ldots + f_n} = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

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

Problem 1: Find the mean deviation about the mean for the following data

x_i	2	5	7	8	10	35
f_i	6	8	10	6	8	2

Solution : we will tabulate the values as follows:

x_i	f_i	$x_i f_i$	$\lvert x_i - \bar{x}\rvert = \lvert x_i - 8\rvert$	$\lvert x_i - \bar{x}\rvert f_i$
2	6	12	6	36
5	8	40	3	24
7	10	70	1	10
8	6	48	0	0
10	8	80	2	16

35	2	70	27	54
	$\sum f_i$=N=40	$\sum x_i f_i$=320		140

Thus A.M= $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{320}{40} = 8$

$\therefore$ Mean deviation = $\frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$ =140/40 = 3.5

Problem 2: Find the mean deviation about the median for the following data

x_i	6	9	3	12	15	13	21	22
f_i	4	5	3	2	5	4	4	3

Solution : Given observations in ascending order to get the table as follows:

x_i	3	6	9	12	13	15	21	22
f_i	3	4	5	2	4	5	4	3

Here N= 30

$\therefore$ Median is the mean of 15th and 16th observation which is equal to 13.

Now we tabulate the absolute values of the deviations.

$\lvert x_i - med\rvert = \lvert x_i - 13\rvert$	10	7	4	1	0	2	8	9
f_i	3	4	5	2	4	5	4	3
$f_i \lvert x_i - med\rvert$	30	28	20	2	0	10	32	27

Thus $\sum f_i |x_i - \text{median}| = 149$

$\therefore$ Mean deviation from the median = $\frac{\sum_{i=1}^{8} f_i |x_i - Median|}{\sum f_i} = \frac{149}{30} = 4.97$

Problem 3: Find the mean deviation about the mean for the following data

x_i	5	10	15	20	25
f_i	7	4	6	3	5

Solution:

Mean=14;

Mean deviation about the mean= 6.32

Problem 4: Find the mean deviation from median for the following data

x_i	6	7	8	9	10	11	12
f_i	3	6	9	13	8	5	4

Solution : Median = 9 ;

Mean deviation about the median = 1.25

Mean Deviation from mean for Continuous Frequency Distribution: A continuous frequency distribution is a series in which the data is classified into different class intervals along with respective frequency. We calculate the A.M. of a continuous frequency distribute, we take x_i as the mid value of the class interval.

Problem 1: The following table gives the sales of 100 companies. Find the mean deviation from the mean.

Sales in thousands	40-50	50-60	60-70	70-80	80-90	90-100
Number of companies	5	15	25	30	20	5

Solution: we construct the following table for the given data

Sales	Number of companies f_i	Midpoint of the class x_i	$x_i f_i$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert f_i$
40-50	5	45	225	26	130
50-60	15	55	825	16	240
60-70	25	65	1625	6	150
70-80	30	75	2250	4	120
80-90	20	85	1700	14	280
90-100	5	95	475	24	120
	$\sum f_i = N = 100$		$\sum x_i f_i = 7100$		$\sum \lvert x_i - \bar{x} \rvert f_i = 1040$

Now $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{7100}{100} = 71$

Mean Deviation from mean= $\frac{\sum_{i=1}^{n} f_i \lvert x_i - \bar{x} \rvert}{\sum_{i=1}^{n} f_i}$ =1040/40 = 10.4

Problem 2: Find the mean deviation of the following frequency distribution:

Class Interval	0-4	4-8	8-12	12-16	16-20	20-40
Frequency	8	12	35	25	13	7

Solution : Mean = 11.76; mean deviation = 4.176

Step Deviation Method (Short Cut method) :

Suppose in the given data the midpoints of the class intervals x_i and their associated frequencies are numerically large. Then the computations become tedious (too large).

To avoid large calculations, we take an assumed mean ***a*** which lies in the middle or close to it in the data and take the deviations of the mid points x_i from this assumed mean. This is equal to shifting the origin from 0 to assumed mean on the number line.

Again, if there is a common factor of all the deviations, we divide them by their common factor (h) to further simplify the deviations. These are known as Step Deviations.

With the assumed mean **'*a*'** and a common factor h we define a new variable,

$d_i = \frac{x_i - a}{h}$. Then A.M. = $\bar{x} = \left(\frac{\sum f_i d_i}{N} \right) h$

Problem 1: Find the mean deviation about the mean for the following data

Classes	0-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800
Freq.	4	8	9	10	7	5	4	3

Solution : Assumed mean $a = 350$

Classes	Mid values(x_i)	Frequency(f_i)	d_i	$f_i d_i$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert f_i$
0-100	50	4	-3	-12	308	1232
100-200	150	8	-2	-16	208	1664
200-300	250	9	-1	-9	108	972
300-400	350	10	0	0	8	80
400-500	450	7	1	7	92	644
500-600	550	5	2	10	192	960
600-700	650	4	3	12	290	1168
700-800	750	3	4	12	392	1176
		50		4		7896

$$d_i = \frac{x_i - assumed\ mean}{class\ size} = \frac{x_i - a}{h} = \frac{x_i - 350}{100}$$

Now $\bar{x} = a + \frac{\sum f_i d_i}{\sum f_i} \times class\ size = 350 + \frac{4}{50} x100 = 358$

Mean Deviation from mean= $\frac{\sum_{i=1}^{n} f_i \lvert x_i - \bar{x} \rvert}{\sum_{i=1}^{n} f_i}$ = 7896 / 50 = 157.92

Problem 2: Find the mean deviation about the mean for the following data

Classes	0-10	10-20	20-30	30-40	40-50
Freq.	5	8	15	16	6

Solution :

Classes	Mid values(x_i)	Frequency(f_i)	d_i	$f_i d_i$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert f_i$
0-10	5	5	-2	-10	22	110
10-20	15	8	-1	-8	12	96
20-30	25	15	0	0	2	30
30-40	35	16	1	16	8	128
40-50	45	6	2	12	18	108
		50		10		472

Now $\bar{x} = a + \frac{\sum f_i d_i}{\sum f_i} \times h = 25 + \frac{10}{50} \times 10 = 27$ and $\frac{x_i - \bar{x}}{h} = \frac{x_i - 27}{10}$

Mean Deviation from mean= $\dfrac{\sum_{i=1}^{n} f_i |x_i - \overline{x}|}{\sum_{i=1}^{n} f_i}$ = *472 / 50 = 9.44*

Problem 3: Find the mean deviation from median for the following data

Age of workers	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No. of workers	120	125	175	160	150	140	100	30

Solution : we form the following table for the given data

Classes	Mid points(x_i)	*Frequency(f_i)*	*Cumulative frequency (c.f)*	$\lvert x_i - \overline{x} \rvert = \lvert x_i - 37.5 \rvert$	$\lvert x_i - \overline{x} \rvert f_i$
20-25	22.5	120	120	15	1800
25-30	27.5	125	245	10	1250
30-35	32.5	175	420	5	875
35-40	37.5	160	580	0	0
40-45	42.5	150	730	5	750
45-50	47.5	140	870	10	1400
50-55	52.5	100	970	15	1500
55-60	57.5	30	1000	20	600
		N=1000			8175

Here N / 2 = 1000 / 2 = 500.

The C.f. just greater than N / 2 is 580. i = 5 (length of class interval)

The corresponding class interval is 35 – 40. This is the median class

$$\therefore Median\ (M_d) = l + \frac{\frac{N}{2} - cf}{f} X\ i = 35 + \frac{500 - 420}{160} X\ 5 = 35 + 2.5 = 37.5$$

$$\therefore \text{Mean deviation from the median} = \frac{\sum_{i=1}^{8} f_i |x_i - Median|}{\sum f_i} = \frac{8175}{1000} = 8.175$$

Problem 4: Find the mean deviation from median for the following data

Wages/week (Rs.)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of workers	120	125	175	160	150	140	100

Solution : Here N/2=30 ; Median = 45 ;

Mean deviation from median = 11.33

UNIT - II

Probability

1. Probability
2. probability axioms
3. Addition & multiplicative law
4. Conditional probability
5. Baye's theorem
6. Random variables. (Discrete & Continuous)
7. probability Density functions
8. properties
9. Mathematical expectation.

(Pg.no: 1 – 20)

Probability and Statistics

Statistics: Statistics is a branch of Mathematics. The word statistics is derived from the Latin word Status, Greek word Statista and Italian word statistik, each of which means political state. In ancient days the goverment used this word collect the information regarding the population of the country and wealth of the country.

The theoritical development of modern statistics came in mid of 17^{th} Centuries with introduction of "Theory of probability". (prof R.A. fisher is known as the father of statistics).

statistics has been defined differently by different authors from time to time. Mainly, statistics is defined as:

Statistics is Science which deals with the collection, classification, Analysation and interpretation of data. This is the older definition. The modern definition of Statistics is the decision making science in the face of uncertainity.

Probability theory: The word probability theory means possibility or chance. The element of chance plays a vital role in our daily life many important business decisions are made on this basis.

when we toss a coin, the result of the top face may be either head or tail. when we throw (or) role a six face die we may get either $\{1, 2, \ldots, 6\}$.

TERMINOLOGY:

Random experiment: An experiment whose result cannot be predicted with certainity before conducting the actual experiment, even though the outcomes are known is called Random experiment.

In a random experiment every result will have equal chance to happen.

EX: Tossing a coin, rolling a die.

Trial and Event: Trial is a single performance of an experiment and the outcome of such an experiment are known as events.
(or) Event is a subspace of sample space $E \subseteq S$
Ex: In tossing a coin, tossing a coin is trial and the outcomes.

Experiment: An experiment is any physical action or process i.e., observed and the result is noted.

Deterministic experiment: An experiment is called Deterministic experiment, if the result can be predicted with certainty prior to the performance of the experiment.

Ex: Throwing a stone of upwards where it is known that the stone will fall definitely to the ground, due to force of gravitation.

Finite Sample Space: A sample space is finite sample space if its sample points are finite numbers.

Infinite Sample Space: A sample space is called an infinite sample space if its sample points are infinite in numbers.

Theorem: If E_1, E_2 are any two events then $P(E_1 \cap E_2) \le P(E_1) \le P(E_1 \cup E_2) \le P(E_1) + P(E_2)$.

Proof: Clearly, $E_1 \cap E_2 \subseteq E_1 \subseteq E_1 \cup E_2$

$$P(E_1 \cap E_2) \le P(E_1) \le P(E_1 \cup E_2) \quad \text{--- (1)}$$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$\Rightarrow P(E_1 \cup E_2) \le P(E_1) + P(E_2) \quad \text{--- (2)}$$

From (1) & (2), we get $P(E_1 \cap E_2) \le P(E_1) \le P(E_1 \cup E_2) \le P(E_1) + P(E_2)$

Notations: (a) A, B are any events of sample space 'S'.
(i) Neither 'A' nor 'B' occurs : $A^c \cap B^c$ (or) $(A \cup B)^c$.
(ii) 'A' occurs and 'B' doesnot occurs : $A \cap B^c$.
(iii) Exactly one of 'A' and 'B' occurs : $(A \cap B^c) \cup (A^c \cap B)$.
(b) If A, B, C are events in a sample space 'S':
(i) Among them only 'A' occurs : $A \cap B^c \cap C^c$
(ii) Both 'A' and 'B' occur and 'C' not occur : $A \cap B \cap C^c$
(iii) All 3 of A, B, C occur : $A \cap B \cap C$.
(iv) Atleast one of A, B, C occur : $A \cup B \cup C$.

Random Experiment:

- An experiment which can be repeated any number of times under identical conditions.
- All the results of the experiment are known in advance.
- But actual result in a particular case is not known in advance.

Ex: 1. Tossing a coin
2. Rolling a die
3. Drawing a card from pack of cards.

Deterministic Experiment:

An experiment is called Deterministic experiment, if the result can be predicted with certainty prior to the performance of the experiment.

Ex: Throwing a stone of upwards where it is known that the stone will fall definitely to the ground, due to force of gravitation.

Sample Space:

A set of all possible outcomes of an experiment is called sample space and it is denoted by 'S'.

Ex: 1. When we toss a coin the possible cases are head or tail.
So, sample space $S = \{H, T\} = 2$.
2. When we roll a die then $S = \{1, 2, 3, 4, 5, 6\}$.
3. When 'n' dice are thrown $S = 6^n$.

Trial and Event:

Trial is a single performance of an experiment and the outcome of such an experiment are known as events.
(or) Event is a subspace of sample space $E \subseteq S$.

Experiment:

An experiment is any physical action or process i.e, observed and the result is noted.

Exhaustive events:

The total no. of all possible outcomes of a trial are known as exhaustive events.

(or)

Events of a random experiment are said to be exhaustive if atleast one of them necessarily occurs.

Ex: 1) Rolling a die : Sample space = {1,2,3,4,5,6}

A : Getting even number = {2,4,6}

B : Getting odd number = {1,3,5}

C : Getting prime number = {2,3,5}

The event A,B (or A∪B) are exhaustive events.

2) When we toss a coin there are two exhaustive events {H,T}

Favourable events:

The number of outcomes favourable to an event in an experiment is the number of outcomes its entail the happening of an event.

(or)

The outcomes which make necessary the happening of an event in a trial are called favourable events.

Example: 1. If two dice are thrown, the number of favourable events of getting a sum 5 is four.

i.e., (1,4), (2,3), (3,2) and (4,1).

2. In tossing two coins at a time, no. of cases favourable to the event of getting tail on both the coins are (T,T).

- No of favourable cases for getting two heads are (H,H)
- No. of favourable cases for getting atleast one head is 3 {HH, HT, TH}.

Equally likely events:

Events are said to be equally likely if each event has equal chance to happen.

Ex: When we toss an unbiased coin Head & Tail has equal chance to happen. Head may happen first or tail may happen.

Mutually Exclusive events:

Events are said to be mutually exclusive if the following happening of one event excludes the happening of all the remaining events.

Mutually exclusive events does not happen jointly. So, mutually exclusive events are disjoint events.

Ex: When we toss a coin either head alone happens or tail alone happens but not both, at a time happening of head excludes the happening of tail.

Hence head & tail are mutually exclusive events.

i.e., the two events A and B are said to be M.E.E if $A \cap B = \phi$.

Dependent event (or) Conditional event:

Events are said to be dependent if the happening of one event depends on the happening of other events.

Ex: Two coins are tossed. The event of getting two tails given that there is atleast one tail is a conditional event.

Independent events:

Events are said to be independent if the happening of one event does not depends on the happening or non-happening of other events.

Ex: When two coins are tossed independently at a time the result of the first coin doesnot depends on the result on the second coin.

Compound events (Joint events)

Events which happen jointly at a time are called compound events.

Ex: When two coins are tossed at a time then the possible joint events are (HH, HT, TH, TT).

Compliment events:

All outcomes that are not the event (A / A^c / $\bar{A}$).

Ex: 1. When the event is Heads, the complement is Tails.

2. When the event is {Monday, Wednesday}, the complement is {Tuesday, Thursday, Friday, Saturday, Sunday}.

* Definition of probability:

If an experiment is performed 'n' is the no of exhaustive cases (total no of cases) and 'm' is the no of a favourable cases of an event 'A' then the probability of an event 'A' is denoted by p(A) and it is defined as

$$p(A) = \frac{\text{Number of favourable Cases}}{\text{Number of Exhaustive Cases (total no of Cases)}} = \frac{m}{n}.$$

i.e, $p = \frac{m}{n}$

Note: $p(A) + p(\bar{A}) = 1$

i.e, The probability of success + the probability of failure = 1

The axioms of a probability:

1. If A is any event, then the probability of 'A' is lies between 0&1.
   i.e, $0 \leq p(A) \leq 1$.
2. The probability of a sample space is equal to one.
   i.e, $p(S) = 1$
3. If A and B are said to be mutually exclusive events, then
   $p(A \cup B) = p(A) + p(B)$.

Problems:

* In a sample of 446 cars, stopped at a road block, only 67 of the drivers had their seat belts fastened. Estimate the probability that a driver stopped on that road, will have his or her seat belt fastened.

Sol: Number of favourable outcomes = 67

Exhaustive events = 446

∴ Required probability = $\frac{67}{446}$

* Out of 15 items 4 are not in good condition 4 are selected at random. Find the probability that (i) All are not good (ii) Two are not good.

Sol: Total number of items = 15

Number of ways of picking 4 items = ${}^{15}C_4$

i) Suppose 4 items are chosen which are not good.

Number of ways of selecting = ${}^{4}C_4$

∴ The probability all are not good = $\frac{{}^{4}C_4}{{}^{15}C_4} = \frac{1}{1365}$

ii) Suppose two items are not good

Number of ways of selecting of 2 bad items = ${}^{4}C_2$

∴ probability of getting two items which are not good = $\frac{{}^{4}C_2}{{}^{15}C_4} = \frac{2}{455}$

* Find the probability of getting one red king if we selected a card from a pack of 52 cards.

52 Cards (A 2 3 4 5 6 7 8 9 10 J Q K)
13 13 13 13
26 - Red Cards < 13, 13
26 - Black Cards < 13, 13

Sol: There are 2 red kings.

∴ The number of possible cases = 2

Number of exhaustive cases = 52

∴ probability = $\frac{2}{52} = \frac{1}{26}$.

* Find the probability of getting a head in tossing a coin.

Sol: Number of exhaustive cases = 2

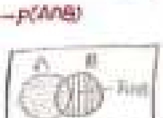
Number of possible cases = 1

∴ probability = $\frac{1}{2}$.

Addition Theorem on probability:

For any two events A and B; $p(A\cup B) = p(A)+p(B)-p(A\cap B)$

Proof: Set A is shown by horizontal lines,



$\bar{A}\cap B$ by vertical lines.

A and $\bar{A}\cap B$ are disjoint sets.

A and $\bar{A}\cap B$ are mutually exclusive events

$A\cup(\bar{A}\cap B) = A\cup B$

$p[A\cup(\bar{A}\cap B)] = p[A\cup B]$

$p(A)+p(\bar{A}\cap B) = p(A\cup B)$

$p(A)+p(B)-p(A\cap B) = p(A\cup B)$ $\quad [\because p(\bar{A}\cap B) = p(B)-p(A\cap B)]$

$\therefore p(A\cup B) = p(A)+p(B)-p(A\cap B)$

Note: For any three events A, B, C; $p(A\cup B\cup C) = p(A)+p(B)+p(C)-p(A\cap B)-p(B\cap C)-p(C\cap A)+p(A\cap B\cap C)$

Conditional Probability:

If A & B are two events then the probability of happening of event B given that A has already happend. It is denoted by

$$p(B/A) = \frac{p(A\cap B)}{p(A)} \text{ if } p(A)\neq 0 \text{ (or) } >0.$$

Ily

$$p(A/B) = \frac{p(A\cap B)}{p(B)} \text{ if } p(B)>0.$$

Multiplication Theorem on probability:

If A and B are any two events then $p(A\cap B) = p(A)\cdot p(B/A)$
$= p(B)\cdot p(A/B)$.

Proof:- From the definition of conditional probability

$$p(A/B) = \frac{p(A\cap B)}{p(B)}, \text{ if } p(B)\neq 0.$$

$\Rightarrow p(A\cap B) = p(A/B)\cdot p(B)$.

Ily $\quad p(B/A) = \frac{p(A\cap B)}{p(A)}$, if $p(A)\neq 0$.

$\Rightarrow p(A\cap B) = p(A)\cdot p(B/A)$

Note: Suppose for three events A, B & C then $p(A\cap B\cap C) = p(D\cap C)$

$p(D\cap C) = p(D)\cdot p(C/D)$

$= p(A\cap B)\cdot p(C/A\cap B)$

$= p(A)\cdot p(B/A)\cdot p(C/A\cap B)$.

For 'n' elements $A_1, A_2, A_3, \ldots A_n$ then $p(A_1\cap A_2\cap\cdots\cap A_n) = p(A_1)\cdot p(A_2/A_1)\cdot p(A_3/A_1\cap A_2)\cdots p(A_n/A_1\cap A_2\cdots A_{n-1})$

Problems:

* A Card is drawn from a well shuffled pack of Cards. what is the probability that it is either a spade or an ace.

Sol: Let 'S' be a sample space of the sample events $n(S) = 52$.

Let 'A' denote the event of getting a spade and B denote event of getting an ace.

Then $A\cup B$ = The event of getting a spade or an ace.

$A\cap B$ = The event of getting a spade and an ace.

$p(A) = \frac{13}{52}$, $p(B) = \frac{4}{52}$, $p(A\cap B) = \frac{1}{52}$

By addition theorem.

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

$= \frac{13}{52}+\frac{4}{52}-\frac{1}{52}$

$= \frac{16}{52}$

$p(A\cup B) = \frac{4}{13}$.

* A bag contains 4 green, 6 black and 7 white balls. A ball is drawn at random. what is the probability that it is either a green or black ball.

Sol: Let A be the event of getting a green ball.

B be the event of getting a black ball.

total sample space (S) = 4+6+7 = 17

$\therefore n(S) = {}^{17}C_1 = 17$

$\Rightarrow p(A) = \frac{4}{17}$, $p(B) = \frac{6}{17}$

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

$= \frac{4}{17}+\frac{6}{17}-0 = \frac{10}{17}$.

* Determine (i) $p(B/A)$ (ii) $p(A/B^c)$ if A and B are events with $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{4}$, $p(A\cup B) = \frac{1}{2}$.

Sol: Given $p(A) = \frac{1}{3}$, $p(B) = \frac{1}{4}$, $p(A\cup B) = \frac{1}{2}$

Now, $p(A\cup B) = p(A)+p(B)-p(A\cap B)$

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

$p(A\cap B) = \frac{1}{3} + \frac{1}{4} - \frac{1}{2} = \frac{1}{12}$

i) $p(B/A) = \frac{p(A\cap B)}{p(A)} = \frac{1/12}{1/3} = 1/4$

(2) $p(\bar{B}) = 1 - p(B) = 1 - \frac{1}{4} = \frac{3}{4}$

$p(A/\bar{B}) = \frac{p(A\cap \bar{B})}{p(\bar{B})} = \frac{p(A) - p(A\cap B)}{1 - p(B)} = \frac{\frac{1}{3} - \frac{1}{12}}{(3/4)} = \frac{3/12}{3/4} = 1/3$

* If $p(A) = 3/8$, $p(B) = 1/3$, $p(A\cap B) = 1/4$. Then find

(i) $p(A^c)$ (ii) $p(B^c)$ (iii) $p(A\cup B)$ (iv) $p(A^c\cup B^c)$ (v) $p(A^c\cap B^c)$

(vi) $p(A^c\cap B)$ (vii) $p(A\cap B^c)$

Sol: (i) $p(A^c) = 1 - p(A) = 1 - \frac{3}{8} = \frac{5}{8}$

(ii) $p(B^c) = 1 - p(B) = 1 - \frac{1}{3} = \frac{2}{3}$.

(iii) $p(A\cup B) = p(A) + p(B) - p(A\cap B) = \frac{3}{8} + \frac{1}{3} - \frac{1}{4} = \frac{11}{24}$ $[\because (A\cup B)^c = A^c\cap B^c$, $(A\cap B)^c = A^c\cup B^c]$

(iv) $p(A^c\cup B^c) = p[(A\cap B)^c] = 1 - p(A\cap B) = 1 - \frac{1}{4} = \frac{3}{4}$

(v) $p(A^c\cap B^c) = p[(A\cup B)^c] = 1 - p(A\cup B) = 1 - \frac{11}{24} = \frac{13}{24}$

(vi) $p(A^c\cap B) = p(B) - p(A\cap B) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

(vii) $p(A\cap B^c) = p(A) - p(A\cap B) = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$.

* Among 150 students 80 are studying Mathematics, 40 are studying physics and 30 are studying mathematics and physics. If a student is chosen at random. Find the probability that the student
(i) studying maths or physics (ii) students studying neither maths (or) physics

Sol: Let A be the event that the student studying maths.

$\therefore p(A) = \frac{80}{150}$

'B' be the event that the student studying physics.

$\therefore p(B) = \frac{40}{150}$

Let $A\cap B$ be the event that the student studying both maths and physics.

$\therefore p(A\cap B) = \frac{30}{150}$

(i) The probability that the students studying maths or physics

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

$= \frac{80}{150} + \frac{40}{150} - \frac{30}{150} = \frac{90}{150} = \frac{3}{5}$

(ii) The probability that the students studying neither maths nor physics.

$p(A^c\cap B^c) = p[(A\cup B)^c] = 1 - p(A\cup B) = 1 - \frac{3}{5} = \frac{2}{5}$.

* One card is drawn from a pack of 52 cards each of the 52 cards being "Equally likely to be drawn". Find the probability of
(i) The card is either red or king
(ii) The card is drawn neither red nor a king.

Sol: The card is drawn neither red n
(i) Let 'A' be the event that a king is drawn from pack of cards.

$p(A) = \frac{4}{52}$

Let 'B' be the event that a red is drawn from pack of cards.

$p(B) = \frac{26}{52}$

$\therefore$ The probability that the card is either red or a king

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

$= \frac{4}{52} + \frac{26}{52} - 0$

$= \frac{30}{52} = \frac{5}{7}$

(ii) The probability that the card is neither red nor a king

$p(A^c\cap B^c) = p((A\cup B)^c) = 1 - p(A\cup B) = 1 - \frac{30}{52} = \frac{22}{52} = \frac{11}{26}$

* Three students A, B, C are in running race, A and B have the same probability of winning and each is twice as likely to win as 'C'. Find the probability that B or C wins.

Given that, $P(A) = P(B)$,

$$2P(A) = P(C) \quad \text{---}①$$

W.K.T Sample space of race $S = A \cup B \cup C$

$$P(S) = P(A \cup B \cup C)$$

Here A, B, C are Mutually exclusive events

$\therefore P(A) + P(B) + P(C) = 1$.

$P(A) + P(A) + 2P(A) = 1$

$4P(A) = 1 \Rightarrow P(A) = \frac{1}{4}$

$\therefore P(B) = \frac{1}{4}$ & $P(C) = \frac{1}{2}$.

$\therefore$ The probability that B or C wins $= P(B \cup C)$

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

$= \frac{1}{4} + \frac{1}{2} - 0$

$= \frac{3}{4}$

* From a city 3 news papers A, B, C are being published. A is read by 20%, B is read by 16%, C is read by 14%, both A and B are read by 8%, both A and C are read by 5%, both B and C are read by 4% and all three A, B, C are read by 2%, what is the percentage of the population that read at least one paper.

Sol: Given $P(A) = \frac{20}{100}$, $P(B) = \frac{16}{100}$, $P(C) = \frac{14}{100}$ and $P(A \cap B) = \frac{8}{100}$,

$P(A \cap C) = \frac{5}{100}$, $P(B \cap C) = \frac{4}{100}$, $P(A \cap B \cap C) = \frac{2}{100}$

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

$$= \frac{20}{100} + \frac{16}{100} + \frac{14}{100} - \frac{8}{100} - \frac{5}{100} - \frac{4}{100} + \frac{2}{100}$$

$$= \frac{35}{100}$$

$\therefore$ percentage of the population that read atleast one paper $= \frac{35}{100} \times 100 = 35\%$.

* Let a die be rolled and $A = \{1,2,3\}$, $B = \{2,3\}$ and $C = \{2,3,4,5\}$ be the events associated with the random experiment

Sol: Now, $P(1) = \frac{1}{6}$, $P(2) = \frac{1}{6}$, $P(3) = \frac{1}{6}$, $P(4) = \frac{1}{6}$, $P(5) = \frac{1}{6}$, $P(6) = \frac{1}{6}$.

$$P\left(\frac{B}{A}\right) = \frac{P(B \cap A)}{P(A)} = \frac{P(\{2,3\})}{P(\{1,2,3\})} = \frac{\frac{2}{6}}{\frac{1}{6}+\frac{1}{6}+\frac{1}{6}} = \frac{2}{3}.$$

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2,3\})}{P(\{2,3\})} = \frac{\frac{2}{6}}{\frac{1}{6}+\frac{1}{6}} = 1.$$

lly $P(C/A) = 2/3$, $P(A/C) = 1/2$, $P(C/B) = 1$, $P(B/C) = 1/2$.

* If $P(A^c) = \frac{3}{8}$, $P(B^c) = \frac{1}{2}$ and $P(A \cap B) = \frac{1}{4}$, then find (i) $P\left(\frac{A}{B}\right)$ & $P\left(\frac{B}{A}\right)$ (ii) $P\left(\frac{A^c}{B^c}\right)$ & $P\left(\frac{B^c}{A^c}\right)$.

Sol: $P(A) = 1 - P(A^c) = 1 - \frac{3}{8} = \frac{5}{8}$

$P(B) = 1 - P(B^c) = 1 - \frac{1}{2} = \frac{1}{2}$

$$P\left(\frac{A^c}{B^c}\right) = \frac{1 - P(A \cup B)}{1 - P(B)} \text{ (or) } \frac{P(A^c \cap B^c)}{P(B^c)} = \frac{1 - [P(A) + P(B) - P(A \cap B)]}{1 - P(B)}$$

$$= \frac{1 - \left[\frac{5}{8} + \frac{1}{2} - \frac{1}{4}\right]}{1 - \frac{1}{2}} = \frac{1}{4}$$

(By Multiplication theorem)

$$P\left(\frac{B^c}{A^c}\right) = \frac{P(B^c \cap A^c)}{P(A^c)} = \frac{P[(B \cup A)^c]}{P(A^c)} = \frac{1 - P(A \cup B)}{1 - P(A)} = \frac{1 - \left[\frac{1}{2} + \frac{5}{8} - \frac{1}{4}\right]}{\frac{3}{8}} = \frac{1}{3}$$

pairwise Independent events: (with Replacement)

A, B, C are events of a Sample Space 'S'. They are said to be pairwise Independent if

$P(A \cap B) = P(A) \cdot P(B)$, $P(B \cap C) = P(B) \cdot P(C)$

$P(A \cap C) = P(A) \cdot P(C)$ when $P(A) \neq 0$, $P(B) \neq 0$, $P(C) \neq 0$.

pairwise dependent events (without Replacement):

If A & B are said to be dependent events, Then

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

* In a certain town 40% have brown hair, 25% have brown eyes & 15% both brown hair & brown eyes. A person is selected at random from the town.

i) If he has brown hair, what is the probability that he has brown eyes also?

ii) If he has brown eyes, determine the probability that he doesn't have brown hair.

Sol: i) If he has brown hair, then the probability that he has brown eyes also.

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{\left(\frac{15}{100}\right)}{\left(\frac{40}{100}\right)} = 0.375$$

ii) If he has brown eyes, then the probability that he doesnot have brown hair is.

$$P(A^c/B) = \frac{P(A^c \cap B)}{P(A)} = \frac{P(B) - P(A \cap B)}{P(A)} = \frac{\frac{25}{100} - \frac{15}{100}}{\frac{25}{100}} = 0.4$$

* Among the workers in a factory only 50% received bonus. Among those receiving the bonus 30% are skilled. What is the probability of a randomly selected worker whose skilled & receiving bonus.

Sol: Let 'A' be the event that workers received bonus.

$P(A) = 50\% = 0.5$

Let 'B' be the event that the worker is skilled.

$P(B/A) = 30\% = 0.3$

The probability that a person is receiving bonus & skilled.

$$P(A \cap B) = P(A) \cdot P(B/A)$$
$$= (0.5)(0.3)$$
$$= 0.15.$$

BAYES THEOREM:

Statement: $E_1, E_2, \ldots E_n$ are 'n' mutually exclusive and exhaustive events such that $P(E_i) > 0$ $(i = 1, 2, \ldots, n)$ in a sample space S and A is any other event in 'S' intersecting with every E_i.

If E_i is any of the events of $E_1, E_2, \ldots E_n$ where $P(E_1), P(E_2), \ldots P(E_n)$ and $P(A/E_1), P(A/E_2), P(A/E_3), \ldots, P(A/E_n)$ are known then

$$P(E_i/A) = \frac{P(E_i) \cdot P(A/E_i)}{\sum_{i=1}^{n} P(E_i) \cdot P(A/E_i)}$$

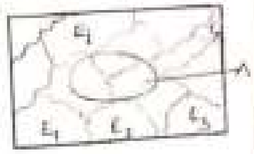
(or)

Suppose $E_1, E_2, \ldots E_n$ are mutually exclusive events of a sample space S such that $P(E_i) > 0$, $i = 1, 2, \ldots n$ and A is any arbitrary event of 'S' such that $P(A) > 0$, and $A \subseteq \bigcup_{i=1}^{n} E_i$. Then conditional probability of E_i given $P(A)$ for $i = 1, 2, \ldots n$ is equal to

$$P(E_i/A) = \frac{P(A/E_i) \cdot P(E_i)}{\sum_{i=1}^{n} P(A/E_i) \cdot P(E_i)}$$

(i.e., A can occur in any combination with any one of the events $E_1, E_2, \ldots E_n$).

proof:-



Let $E_1, E_2, \ldots E_n$ be 'n' mutually exclusive events.

Let 'A' be an independent event.

Given that, $A \subseteq \left(\bigcup_{i=1}^{n} E_i\right)$

i.e., $A \subseteq (E_1 \cup E_2 \cup E_3 \cdots \cup E_n)$ $[\because B \subseteq A \Rightarrow B \cap A = B]$.

So, A can be written as $A = A \cap (E_1 \cup E_2 \cup E_3 \cdots \cup E_n)$

$A = (A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \cdots \cup (A \cap E_n)$ (By distributive law).

$$P(A) = P\{(A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_n)\}$$
$$= P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_n).$$

$$P(A) = \sum_{i=1}^{n} P(A \cap E_i)$$

$$P(A) = \sum_{i=1}^{n} P(E_i) \cdot P(A/E_i) \quad \text{—— (1)}$$

we have from the definition of conditional probability.

$$P(A \cap E_i) = P(E_i) \cdot P(A/E_i) \quad \text{—— (2)}$$

$$P(A \cap E_i) = P(A) \cdot P(E_i/A) \quad \text{—— (3)}$$

$$P(E_i/A) = \frac{P(A \cap E_i)}{P(A)} \quad \text{—— (4)}$$

Substituting (1) & (3) in (4), we get

$$P(E_i/A) = \frac{P(E_i) \cdot P(A/E_i)}{\sum_{i=1}^{n} P(E_i) \cdot P(A/E_i)}$$

Hence proved.

Problems:

1. In a bolt factory machines 'A, B, C' manufacture 20%, 30%, 50% of the total of their output and 6%, 3% and 2% are defective. A bolt is drawn at random and found to be defective. what is the probability that it is manufactured by machines A, B and C?

Sol: Suppose the probability that the bolt was manufactured by machine A is $P(A) = \frac{20}{100} = \frac{1}{5}$

probability that the bolt was manufactured by machine B is

$P(B) = \frac{30}{100} = \frac{3}{10}$

probability that the bolt was manufactured by machine C is

$P(C) = \frac{50}{100} = \frac{1}{2}$

Suppose the probability that the bolt drawn is defective is $P(D)$.

The probability that a defective bolt is drawn from the bolts manufactured by $A = P(D/A) = \frac{6}{100} = 0.06$.

The probability that the defective bolt is from machine $B = P(D/B) = \frac{3}{100} = 0.03$

The probability that the defective bolt is from machine $C = P(D/C) = \frac{2}{100} = 0.02$

probability that the bolt which is defective manufactured from A is $P(A/D)$ by Baye's theorem

$$P(A/D) = \frac{P(A) \cdot P(D/A)}{P(A) \cdot P(D/A) + P(B) \cdot P(D/B) + P(C) \cdot P(D/C)}.$$

$$= \frac{\frac{1}{5}(0.06)}{\frac{1}{5} \cdot \frac{6}{100} + \frac{3}{10} \cdot \frac{3}{100} + \frac{1}{2} \cdot 0.02} = \frac{12}{31}.$$

lly $P(B/D) = \frac{9}{31}$

& $P(C/D) = \frac{10}{31}$.

2. A businessman goes to hotels X, Y, Z, 20%, 50%, 30% of the time respectively. It is known that 5%, 4%, 8% of the rooms in X, Y, Z hotels have faulty plumbings. What is the probability that business-man's room having faulty plumbing is assigned to hotel 'Z'?

Sol: Let the probabilities of business man going to hotels X, Y, Z be respectively $P(X), P(Y), P(Z)$. Then

$$P(X) = \frac{20}{100} = \frac{2}{10} \; ; \; P(Y) = \frac{50}{100} = \frac{5}{10} \; ; \; P(Z) = \frac{30}{100} = \frac{3}{10}.$$

Let E be the event that the hotel room has faulty plumbing. Then the probabilities that hotels X, Y, Z have faulty plumbing are

$$P(E/X) = \frac{5}{100} = \frac{1}{20} \; ; \; P(E/Y) = \frac{4}{100} = \frac{1}{25} \; ; \; P(E/Z) = \frac{8}{100} = \frac{2}{25}$$

The probability that the business man's room having faulty plumbing is assigned to hotel Z $= P\left[\frac{Z}{E}\right] = \dfrac{P(Z) \cdot P(E/Z)}{P(Z)P(E/Z) + P(Y) \cdot P(E/Y) + P(X) \cdot P(E/X)}$

$$= \frac{\frac{3}{10} \times \frac{2}{25}}{\frac{3}{10} \times \frac{2}{25} + \frac{5}{10} \cdot \frac{1}{25} + \frac{2}{10} \cdot \frac{1}{20}}$$

$$= \frac{4}{9}.$$

3. Suppose 5 men out of 100 and 25 women out of 10,000 are colour blind. A color blind person is chosen at random. What is the probability of the person being a male (Assume male and female to be in equal numbers)?

Sol: Given that 5 men out of 100 and 25 women out of 10000 are cold blind.

A colour blind person is chosen at random.

The probability that the chosen person is male $= P(M) = \frac{1}{2}$

The probability that the chosen person is female $= P(W) = \frac{1}{2}$.

Let B represent a blind person. Then

$$P(B/M) = \frac{5}{100} = 0.05$$

$$P(B/W) = \frac{25}{10000} = \frac{1}{400} = 0.0025.$$

The probability that the chosen person is male is given by (13)

$$P(M/B) = \frac{P(B/M) \cdot P(M)}{P(M) \cdot P(B/M) + P(W) \cdot P(B/W)}.$$

$$= \frac{0.05 \times 0.5}{(0.05 \times 0.5) + (0.5 \times 0.0025)} = 0.95.$$

(4) A bag A contains 2 white and 3 red balls and a bag B contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and it is found to be red. Find the probability that the red ball drawn is from bag B.

Sol: Let A and B denote the events of selecting bag A and bag B respectively.

Then $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{2}$

Let R denote the event of drawing ~~a dra~~ red ball.

Having selected bag A, the probability to draw a red ball from A

$$= P(R/A) = \frac{3}{5}.$$

IIly $P(R/B) = \frac{5}{9}$

One of the ball is selected at random and from it a ball is drawn at random.

It is found to be red. Then the probability that the selected bag is $B = P(B/R) = \dfrac{P(B) \cdot P(R/B)}{P(A) \cdot P(R/A) + P(B) \cdot P(R/B)}$

$$= \frac{\frac{1}{2} \cdot \frac{5}{9}}{\frac{1}{2} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{5}{9}}$$

$$\therefore P(B/R) = \frac{25}{52}$$

(5) In a factory, machine A produces 40% of the output and machine B produces 60%. On the average, 9 items in 100 produced by A are defective and 1 item in 250 produced by B is defective. An item drawn at random from a day's output is defective. What is the probability that it was produced by A & B?

Sol: output produced by A = 40%

$\therefore p(A) = 0.4$

output produced by B = 60%

$\therefore p(B) = 0.6$

$P(D/A)$ = probability that items produced by A are defective $= \frac{9}{100}$

ly $p(D/B)$ = " " " " B " " $= \frac{1}{25}$

$\therefore p(A/D)$ = probability that the ball is produced by A given that it is defective

$$= \frac{p(A) \times p(D/A)}{p(A)p(D/A) + p(B) \cdot p(D/B)} = \frac{0.4 \times 0.009}{(0.4 \times 0.009) + (0.6 \times 0.004)} = 0.6$$

lly $P[B/D] = 0.4$

Random variables:

A real variable X whose value is determined by the outcome of a random experiment is called a random variable.

(or)

Random variable is any function that assigns with a numerical value to each possible outcome.

Example:

The sample space corresponding to tossing of two coins.

When two coins are tossed, its outcomes or sample points can be (H,H), (H,T), (T,H), (T,T) i.e., S = {HH, HT, TH, TT}. After the performance of the experiment, we count the number of tails and denote it by X.

The first outcome HH has 0 tail. So X = 0.

Similarly X = 1, denote the outcomes HT or TH and X = 2, represents the outcome TT.

probability of getting zero heads is 1/4; [P(X=0)]

$P[X=1] = 2/4$

$P[X=2] = 1/4$.

Now the probability distribution of the random variable 'X' is given by

X	P(X=x)
0	1/4
1	2/4
2	1/4
Total	1

Types of Random variables:

Random variable is of two types

(i) Discrete Random variable

(ii) Continuous Random variable.

(1) Discrete Random variable:

A random variable X which can take only a finite number of discrete values in an interval of domain is called a discrete random variable.

Ex: 1. No. of children in a family.

2. The no. of customers who visit a bank during any given hour.

3. The no. of heads obtained in three tosses of a coin.

4. The no. of students in a class.

(2) Continuous Random variable:

A random variable X which can take values continuously i.e., which takes all possible values in a given interval is called a continuous random variable.

(or)

A continuous random variable is one which takes an infinite no. of possible values.

* Continuous random variables are usually measurements.

Ex: 1. The amount of sugar in an orange.

2. The height, age & weight of a person.

3. pressure, volume.

probability Distribution function:

Let x be a random variable. Then the probability distribution function associated with x is defined as the probability that the outcome of an experiment will be one of the outcomes for which $x(s) \leq x, x \in R$.

Discrete probability Distribution: (probability Mass function) (2)

Let 'X' be a discrete random variable with possible outcomes (values) with their corresponding probability $p(x_1), p(x_2), \ldots p(x_n)$ then $p(X = x_i) = p(x_i)$ for $i = 1, 2, 3, \ldots$ is called the probability mass function of the random variable 'X' if it satisfies the following properties.

i) $p(x_i) \geq 0 \ \forall \ i; \ 0 \leq p_i \leq 1$ (non -ve)

ii) $\sum p(x_i) = 1, \ i = 1, 2, 3, \ldots$ (total probability).

Expectation, Mean, variance and Standard deviation of a probability distribution:

Expectation: Suppose a random variable 'X' assumes the values $x_1, x_2, \ldots, x_n$ w.r.t. the probabilities $p(x_1), p(x_2), \ldots, p(x_n)$. Then the expectation (or) expected value of x denoted by $E(x)$ is defined as the sum of products of different values of x and the corresponding probabilities.

$$\text{i.e., } E(X) = \sum_{i=1}^{n} p_i x_i$$

In general, the expected value of any function $g(x)$ of a random variable x is defined as

$$E[g(x)] = \sum_{i=1}^{n} p_i \, g(x_i).$$

Mean: The mean value μ of the discrete distribution function is given by

$$\mu = \frac{\sum p_i x_i}{\sum p_i} = \sum p_i x_i = E(x) \qquad \left[\because \sum_{i=1}^{n} p_i = 1\right]$$

Variance: The variance σ^2 of the discrete distribution function is given by $\sigma^2 = var[X] = E[X - E(x)]^2$

(or) $$\sigma^2 = E[X - \mu]^2 = E(x^2) - [E(x)]^2$$

Standard Deviation:

It is the positive square root of the variance

$$\therefore S.D = \sigma = \sqrt{\sum_{i=1}^{n} p_i x_i^2 - \mu^2} = \sqrt{E(x^2) - \mu^2}$$

$$= \sqrt{E[X - E(X)]^2}$$

Results on Expectation:

* $E[X+K] = E[X] + K$; K is a Constant.
* $E[KX] = K\,E(X)$; " "
* $E[aX+b] = a\cdot E[X] + b$; a & b are Constants.
* $E[aX+bY] = a\,E[X] + b\cdot E[Y]$
* $E[X-\bar{X}] = 0$
* $E[X+Y] = E[X] + E[Y]$
* $E[XY] = E[X]\cdot E[Y]$; x & y are Independent variables.
* Variance of Constant is zero. i.e, $V(K) = 0$.
* $V[X+K] = V(X)$.
* $V[aX+b] = a^2\cdot V(X)$.

Addition Theorem on Expectation:

Statement: If x and y are two discrete random variables then $E(X+Y) = E(X) + E(Y)$ provided $E(X)$ & $E(Y)$ exist.

proof: Let 'x' be a random variable taking the values $x_1, x_2, \ldots, x_n$ with the corresponding probabilities $p(x_1), p(x_2), \ldots, p(x_n)$.

Let 'y' be a random variable taking the values $y_1, y_2, \ldots, y_m$ with the corresponding probabilities $p(y_1), p(y_2), \ldots, p(y_m)$

Then by definition, $E(X) = \sum_{i=1}^{n} x_i p_i$

$$E(Y) = \sum_{j=1}^{m} y_j p_j$$

Let $P_{ij} = P(X = x_i \cap Y = y_j)$

$$\therefore E(X+Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i + y_j)\, P_{ij}(x_i y_j).$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i P_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} y_j P_{ij}$$

$$= \sum_{i=1}^{n} x_i \left[\sum_{j=1}^{m} P_{ij}\right] + \sum_{j=1}^{m} y_j \left[\sum_{i=1}^{n} P_{ij}\right]$$

$$= \sum_{i=1}^{n} x_i\, p(x_i) + \sum_{j=1}^{m} y_j\, p(y_j)$$

$\therefore E(X+Y) = E(X) + E(Y)$

Product theorem on Expectation:

Statement: If x and y are two independent random variables then $E(XY) = E(X)\cdot E(Y)$.

proof: Let x be a random variable taking the values $x_1, x_2, \ldots, x_n$ in the corresponding probabilities $p(x_1), p(x_2) \ldots p(x_n)$.

Let y be a random variable taking the values $y_1, y_2, \ldots, y_m$ in the corresponding probabilities $p(y_1), p(y_2) \ldots p(y_m)$.

By definition, $E(X) = \sum_{i=1}^{n} x_i\, p(x_i)$; $E(Y) = \sum_{j=1}^{m} y_j p_j$

$$E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i y_j)\, p(x_i y_j).$$

$$= \left[\sum_{i=1}^{n} x_i\, p(x_i)\right]\left[\sum_{j=1}^{m} y_j\, p(y_j)\right]$$

$$E(XY) = E(X)\cdot E(Y).$$

* If 'x' is a discrete random variable and 'k' is a Constant, then $V(X+K) = V(X)$.

proof: Let $Y = X + K$ —①

Then $E(Y) = E(X+K) = E(X) + K$ —②

①-② gives, $Y - E(Y) = X - E(X)$

By squaring on both sides, we get then taking expectation of both sides

$$E[Y - E(Y)]^2 = E[X - E(X)]^2 \Rightarrow V(Y) = V(X) \Rightarrow V(X+K) = V(X)$$

Scanned with CamScanner

* If 'x' is a discrete random variable, then $V(ax+b) = a^2 V(X)$, where $V(X)$ is a variance of X and a,b are constants.

proof: Let $y = ax+b$ —— ①

Then $E(Y) = a \cdot E(X) + b$ —— ②

① − ② gives, $Y - E(Y) = a[X - E(X)]$

Squaring and taking expectation of both sides, we get

$E[Y-E(Y)]^2 = a^2 E\{X-E(X)\}^2$

i.e, $V(Y) = a^2 V(X)$.

Problems:

1. Let x denote the number of heads in a single toss of 4 fair coins. Determine i) $P(X<2)$ (ii) $P(1<X\leq 3)$.

Sol: The probability distribution is

x	0	1	2	3	4
P(x)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

HHHH
HHHT
HHTH
HTHH
THHH
TTTT
TTTH
TTHH
THTT
TTHT
HTTT
THTH
HTHT
HTTH
THHT
HHTT

1) $P(X<2) = P(X=0) + P(X=1)$

$= \frac{1}{16} + \frac{4}{16} = \frac{5}{16}$

(ii) $P(1<X\leq 3) = P(X=2) + P(X=3)$

$= \frac{6}{16} + \frac{4}{16}$

$= \frac{10}{16} = \frac{5}{8}$.

2. Two dice are thrown. Let X assign to each point (a,b) in S the maximum of its numbers i.e, $X(a,b) = \max(a,b)$. Find the probability distribution. X is a random variable with $X(S) = \{1,2,3,4,5,6\}$. Also find the mean and variance of the distribution.

(OR)

A random variable X has the following distribution?

x	1	2	3	4	5	6
P(x)	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Find a) the mean b) variance c) $P(1<X<6)$.

Sol: The total no. of cases are 6×6 (or) $6^2 = 36$.

The maximum number could be 1,2,3,4,5,6.

i.e, $X(S) = X(a,b) = \max(a,b)$.

In this case, Sample space S = { (1,1),(1,2),(1,3),(1,4),(1,5),(1,6)
(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)
(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)
(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)
(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)
(6,1),(6,2),(6,3),(6,4),(6,5),(6,6) }

For maximum 1, favourable case is one case (1,1).

So, $P(1) = P(X=1) = P(1,1) = \frac{1}{36}$.

For maximum 2, $P(X=2) = \frac{3}{36} = P(2)$

Similarly, $P(3) = P(X=3) = \frac{5}{36}$

$P(4) = P(X=4) = \frac{7}{36}$

$P(5) = P(X=5) = \frac{9}{36}$

and $P(6) = P(X=6) = \frac{11}{36}$.

∴ The required probability distribution is

$X = x_i$	1	2	3	4	5	6
$P(X=x_i) = P(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Mean $\mu = \sum_{i=1}^{6} p_i x_i = 1\cdot\frac{1}{36} + 2\cdot\frac{3}{36} + 3\cdot\frac{5}{36} + 4\cdot\frac{7}{36} + 5\cdot\frac{9}{36} + 6\cdot\frac{11}{36} = \frac{161}{36}$

$= 4.47$

Variance $\sigma^2 = \sum_{i=1}^{6} p_i x_i^2 - \mu^2 = \frac{1}{36}(1)^2 + \frac{3}{36}(2)^2 + \frac{5}{36}(3)^2 + \frac{7}{36}(4)^2 + \frac{9}{36}(5)^2 + \frac{11}{36}(6)^2 - (4.47)^2$

$= 1.9912$

5. Let X denote the minimum of the two numbers that appear when a pair of fair dice is thrown once. Determine the

i) Discrete probability distribution ii) Expectation iii) Variance

Sol: When two dice are thrown, total number of outcomes is $6^2 = 36$

In this case, Sample space S = { (1,1), (1,2), (1,3), (1,4), (1,5), (1,6)
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) }

If the random variable X assigns the minimum of its number in S, then the sample space S = {1, 2, 3, 4, 5, 6}.

$P(1) = P(X=1) = \frac{11}{36}$ $\quad$ $P(4) = P(X=4) = \frac{5}{36}$

$P(2) = P(X=2) = \frac{9}{36}$ $\quad$ $P(5) = P(X=5) = \frac{3}{36}$

$P(3) = P(X=3) = \frac{7}{36}$ $\quad$ $P(6) = P(X=6) = \frac{1}{36}$

i) The probability distribution is

X	1	2	3	4	5	6
P(X)	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

ii) Expectation = Mean = $\sum p_i x_i$

i.e. $E(X) = 1\cdot\frac{11}{36} + 2\cdot\frac{9}{36} + 3\cdot\frac{7}{36} + 4\cdot\frac{5}{36} + 5\cdot\frac{3}{36} + 6\cdot\frac{1}{36}$

(or) $\mu = \frac{91}{36} = 2.5278$

(iii) Variance $= \sum p_i x_i^2 - \mu^2$

$= \frac{11}{36}\cdot 1 + \frac{9}{36}\cdot 4 + \frac{7}{36}\cdot 9 + \frac{5}{36}\cdot 16 + \frac{3}{36}\cdot 25 - \left[\frac{91}{36}\right]^2$

i.e., $\sigma^2 = 1.9713$

4. A random variable X has the following probability function:

x	0	1	2	3	4	5	6	7
p(x)	0	K	2K	2K	3K	K^2	$2K^2$	$7K^2+K$

i) Determine K (ii) Evaluate $P(X<6)$, $P(X\geq 6)$, $P(0<X<5)$ and $P(0\leq X\leq 4)$

(iii) If $P(X\leq K) > \frac{1}{2}$, find the minimum value of K and

(iv) Determine the distribution function of X

(v) Mean (vi) Variance.

Sol: i) Since $\sum_{x=0}^{7} P(x) = 1$, we have

$K + 2K + 2K + 3K + K^2 + 2K^2 + 7K^2 + K = 1$

$\Rightarrow 10K^2 + 9K - 1 = 0$

i.e., $(10K-1)(K+1) = 0$

$\therefore K = \frac{1}{10} = 0.1$ (Since $P(x) \geq 0$, So $K \neq -1$)

ii) $P(X<6) = P(X=0) + P(X=1) + \cdots + P(X=5)$

$= (0.1) + 2(0.1) + 2(0.1) + 3(0.1) + (0.1)^2 + 2(0.1)^2 + 7(0.1)^2 + (0.1)$

$= 0.81$

$P(X\geq 6) = 1 - P(X<6) = 1 - 0.81 = 0.19$

$P(0<X<5) = P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= (0.1) + 2(0.1) + 2(0.1) + 3(0.1)$

$= 0.8$

$P(0\leq X\leq 4) = P(X=0) + P(X=1) + P(X=2) + \cdots + P(X=4)$

$= 0 + (0.1) + 2(0.1) + 2(0.1) + 3(0.1)$

$= 8(0.1) = 0.8$

(iii) The required minimum value of k is obtained as below.

$p(x \le 1) = p(x=0) + p(x=1)$

$= 0 + k = \frac{1}{10} = 0.1$

$p(x \le 2) = [p(x=0) + p(x=1) + p(x=2)]$ ~~tip~~

$= \frac{1}{10} + \frac{2}{10} = \frac{3}{10} = 0.3$

$p(x \le 3) = [p(x=0) + p(x=1) + p(x=2)] + p(x=3)$

$= 0.3 + 0.2$

$= 0.5$

$p(x \le 4) = p(x \le 3) + p(x=4) = 0.5 + \frac{3}{10} = 0.5 + 0.3 = 0.8 > 0.5 = \frac{1}{2}$

$\therefore$ The minimum value of k for which $p(x \le k) > \frac{1}{2}$ is $k=4$.

(v) The distribution function of x is given by the following table:

x	$p(x)$	$F(x) = p(x \le x)$ (cumulative d.f of a R.V)
0	0	0
1	0.1	$k = \frac{1}{10} = 0.1$
2	0.2	$3k = \frac{3}{10} = 0.3$
3	0.2	$5k = \frac{5}{10} = 0.5$
4	0.3	$8k = \frac{8}{10} = 0.8$
5	0.01	$8k + k^2 = \frac{81}{100} = 0.81$
6	0.02	$8k + 3k^2 = \frac{83}{100} = 0.83$
7	0.17	$9k + 10k^2 = 1$

Mean $\mu = \sum_{i=0}^{7} p_i x_i = 1(0.1) + 2(0.2) + 3(0.2) + 4(0.3) + 5(0.01) + 6(0.02) + 7(0.17)$.

$= 3.66$

Variance $\sigma^2 = \sum_{i=0}^{7} p_i x_i^2 - \mu^2$

$= (0.1) + (0.2)(4) + (0.2)(9) + (0.3)(16) + (0.01)(25) + (0.02)(36) + (0.17)(49)$

$= 5.4044$.

5. The probability distribution function of a variable X is

X :	0	1	2	3	4	5	6
$p(x)$:	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

(i) Find k (ii) Find $p(x<4)$ (iii) $p(x \ge 5)$, $p(3<x \le 6)$

(iv) What will be the minimum value of k so that $p(x \le 2) > 0.3$

[Ans: $k = \frac{1}{49}$, $p(x<4) = \frac{16}{49}$, $p(x \ge 5) = \frac{24}{49}$, $p(3<x \le 6) = \frac{33}{49}$

$p(x \le 2) > 0.3 \Rightarrow$ The minimum value of k is $\frac{1}{30}$].

6. A Sample of 3 items is selected at random from a box containing 10 items of which 4 are defective. Find the expected number of defective items?

Sol: The probability of defective is $p = \frac{4}{10} = \frac{2}{5}$

Number of items chosen, $n = 3$

The expected number of defective is $E(x) = \mu = np = 3(\frac{2}{5}) = 1.2 \approx 1$.

7. A Sample of 4 items is selected at random from a box containing 12 items of which 5 are defective. Find the expected number E of defective items.

Sol: Let x denote the number of defective items among 4 items drawn from 12 items.

Obviously x can take the values 0, 1, 2, 3, or 4

No. of good items = 7.

No. of defective items = 5 & $n = 4$

The probability of defective is $p = \frac{5}{12}$.

The expected no. of defective is $E(x) = \mu = np$

$= 4(\frac{5}{12})$

$= 1.6667$

8. From a lot of 10 items containing 3 defectives, a sample of 4 items is drawn at random. Let the random variable X denote the number of defective items in the sample. Find the probability distribution of X when the sample is drawn without replacement.

Sol: X can take the values 0, 1, 2 or 3

Given total no. of items = 10

No. of good items = 7

No. of defective items = 3.

$P(X=0) = P(\text{no defective}) = \frac{^{7}C_{4}}{^{10}C_{4}} = \frac{1}{6}$

$P(X=1) = P(1 \text{ defective } \& \text{ } 3 \text{ good items}) = \frac{^{3}C_{1} \times {}^{7}C_{3}}{^{10}C_{4}} = \frac{1}{2}$.

$P(X=2) = P(2 \text{ defective } \& \text{ } 2 \text{ good items}) = \frac{^{3}C_{2} \times {}^{7}C_{2}}{^{10}C_{4}} = \frac{3}{10}$.

$P(X=3) = P(3 \text{ defective } \& \text{ } 1 \text{ good item})$

$= \frac{^{3}C_{3} \times {}^{7}C_{1}}{^{10}C_{4}} = \frac{1}{30}$.

∴ The probability distribution of random variable X is as follows:

x	0	1	2	3
$P(x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

Continuous Probability distribution:

Probability Density function: Let 'X' be a continuous random variable taking all possible values in the interval $[a, b]$. i.e., $a \le x \le b$ then $f(x) = P(X = x)$ is defined as probability density function of X if it satisfies the following properties.

(i) $f(x) \ge 0 \ \forall x \in R$ (ii) $\int_a^b f(x)\,dx = 1$ (or) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

(iii) The probability $P(E)$ is given by $P(E) = \int_E f(x)\,dx$.

Cumulative Distribution Function of a Continuous Random Variable:

The cumulative distribution function or simply the distribution function of a continuous random variable X is denoted by $F(x)$ and is defined as

$$F(x) = P(X \le x) = \int_{-\infty}^{\infty} f(x)\,dx.$$

Thus $F(x)$ gives the probability that the value of the variable X will be $\le x$.

Properties of $F(x)$:

i) $0 \le F(x) \le 1, \ -\infty < x < \infty$

ii) $F'(x) = f(x) \ge 0$, so that $F(x)$ is a non-decreasing function.

iii) $F(-\infty) = 0$ (iv) $F(\infty) = 1$

v) $P(a \le X \le b) = \int_a^b f(x)\,dx = F(b) - F(a)$.

Measures of Central Tendency for Continuous Probability Distribution:

On replacing f_i by $f(x)\,dx$, x_i by x and the summation over 'i' by integration over the specified range of the variable X in the formulae of probability distribution.

Let $f(x)$ be the probability density function of a random variable X. Then

i) **Mean:**

Mean of a distribution is given by $\mu = E(x) = \int_{-\infty}^{\infty} x f(x) dx$.

If x is defined from a to b then $\mu = E(x) = \int_a^b x f(x) dx$.

ii) **Median:**

Median is the point which divides the entire ~~dis~~ distribution into two equal parts.

If x is defined from a to b and M is the median then

$\int_a^M f(x) dx = \int_M^b f(x) dx = \frac{1}{2}$. Solving for M, we get median.

iii) **Mode:**

Mode is the value of x for which f(x) is maximum.

Mode is thus given by $f'(x)=0$ and $f''(x)<0$ for $a<x<b$.

iv) **Variance:**

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \text{ (or) } \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2.$$

Suppose that the variate x is defined from a to b.

Then $\sigma^2 = \int_a^b (x-\mu)^2 f(x) dx$ (or) $\sigma^2 = \int_a^b x^2 f(x) dx - \mu^2$.

v) **Mean deviation:**

$$\int_{-\infty}^{\infty} |x-\mu| f(x) dx.$$

Problems:

1) A random variable x has the following function:

x :	0	1	3	4	5	6	7
p(x) :	0	k	2k	2k	3k	k^2	$7k^2+k$

i) find the value of k ii) evaluate $p(x<6), p(x\geq 6)$

iii) evaluate $p(0<x<5)$.

Sol: If 'x' is a random variable, then $\Sigma p(x_i) = 1$

$\Sigma p(x_i) = 0+k+2k+2k+3k+k^2+7k^2+k = 1$

$\Rightarrow 8k^2+9k-1=0$

$\Rightarrow k = \frac{-9 \pm \sqrt{113}}{16} = 0.102$ (nearly) [$\because p \geq 0$ & if k is negative, $p<0$].

ii) $p(x<6) = p(x=0)+p(x=1)+p(x=3)+p(x=4)+p(x=5) = 0.816$.

$p(x\geq 6) = 1-p(x<6) = 1-0.816 = 0.184$.

iii) $p(0<x<5) = k+2k+2k = 5k = 5(0.102) = 0.51$ (Repeated)

2. If a random variable has the probability density f(x) as

$f(x) = \begin{cases} 2e^{-2x}, & \text{for } x>0 \\ 0, & \text{for } x\leq 0 \end{cases}$ Find the probabilities that it will take on a value (i) between 1 and 3 (ii) greater than 0.5

Sol: i) The probability that ~~a variable~~ the variate takes a value b/w 1 and 3 is given by

$P(1\leq x\leq 3) = \int_1^3 f(x) dx = \int_1^3 2e^{-2x} dx = 2\left[\frac{e^{-2x}}{-2}\right]_1^3 = -[e^{-6}-e^{-2}]$

$= e^{-2}-e^{-6}$.

ii) The probability that a variable takes a value greater than 0.5 is

$P(x\geq 0.5) = \int_{0.5}^{\infty} f(x) dx = \int_{0.5}^{\infty} 2e^{-2x} dx = 2\left[\frac{e^{-2x}}{-2}\right]_{0.5}^{\infty}$

$= -[e^{-\infty}-e^{-1}] = e^{-1}$.

(2) probability density function of a random variable x is

$$f(x)=\begin{cases}\frac{1}{2}\sin x, & \text{for } 0\le x\le\pi\\ 0, & \text{elsewhere}\end{cases}$$

Find the mean, mode and median of the distribution and also find the probability between 0 and $\pi/2$.

Sol: i) Mean of the distribution $= \int_{-\infty}^{\infty} x f(x)\,dx$

$$= \int_{-\infty}^{0} x(0)\,dx + \int_{0}^{\pi} x\cdot\frac{1}{2}\sin x\,dx + \int_{\pi}^{\infty} x\cdot(0)\,dx$$

$$= \frac{1}{2}\int_0^{\pi} x\sin x\,dx$$

$$= \frac{1}{2}\left[-x\cos x+\sin x\right]_0^{\pi}$$

$$= \frac{\pi}{2}$$

ii) Mode is the value of x for which $f(x)$ is maximum.

Now $f'(x)=\frac{1}{2}\cos x$

for $f(x)$ to be maximum, $f'(x)=0$.

i.e, $\cos x=0 \quad \therefore x=\frac{\pi}{2}$.

$f''(x)=-\frac{1}{2}\sin x$. At $x=\frac{\pi}{2}$, $f''(x)=-\frac{1}{2}<0$.

Hence $f(x)$ is maximum at $x=\frac{\pi}{2}$

$\therefore$ Mode of the distribution is given by $x=\frac{\pi}{2}$.

iii) If median of the distribution is M, then

$$\int_0^M f(x)\,dx = \int_M^{\pi} f(x)\,dx = \frac{1}{2}$$

i.e. $\int_0^M \frac{1}{2}\sin x = \int_M^{\pi}\frac{1}{2}\sin x\,dx = \frac{1}{2}$

Solving $\int_0^M \frac{1}{2}\sin x\,dx = \frac{1}{2}$, we get

$\Rightarrow -\frac{1}{2}(\cos x)_0^M = \frac{1}{2} \Rightarrow -\frac{1}{2}(\cos M-1)=\frac{1}{2} \Rightarrow 1-\cos M=1$

$\Rightarrow \cos M=0$

$\Rightarrow M=\pi/2$.

$\therefore$ Median of the distribution $=\frac{\pi}{2}$

Thus Mean $=$ Mode $=$ Median $=\frac{\pi}{2}$

iv) $P(0<x<\frac{\pi}{2}) = \int_0^{\pi/2} f(x)\,dx = \int_0^{\pi/2}\frac{1}{2}\sin x\,dx = \frac{-1}{2}(\cos x)_0^{\pi/2} = \frac{1}{2}$

(3) A Continuous random variable has the probability density function

$$f(x)=\begin{cases}kxe^{-\lambda x}, & \text{for } x\ge 0,\ \lambda>0.\\ 0, & \text{otherwise}\end{cases}$$

Determine i) k ii) Mean iii) variance.

Sol: i) Since the total probability is unity, we have

i.e, $\int_{-\infty}^{\infty} f(x)\,dx=1$

$$\Rightarrow \int_{-\infty}^{0} 0\cdot dx + \int_0^{\infty} kxe^{-\lambda x}\,dx = 1$$

$$\Rightarrow k\int_0^{\infty} xe^{-\lambda x}\,dx=1 \Rightarrow k\left[x\frac{e^{-\lambda x}}{-\lambda} - \frac{e^{-\lambda x}}{\lambda^2}\right]_0^{\infty}=1$$

$$\Rightarrow k\left[(0-0)-(0-\tfrac{1}{\lambda^2})\right]=1 \quad (\text{or})\ k=\lambda^2.$$

Now $f(x)$ becomes

$$f(x)=\begin{cases}\lambda^2 x e^{-\lambda x}; & \text{for } x\ge 0,\ \lambda>0\\ 0; & \text{otherwise.}\end{cases}$$

ii) Mean, $\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{0} x\cdot 0\,dx + \int_0^{\infty} x\lambda^2 x e^{-\lambda x}\,dx = \lambda^2\int_0^{\infty} x^2 e^{-\lambda x}\,dx$

$$= \lambda^2\left[x^2\frac{e^{-\lambda x}}{-\lambda} - 2x\cdot\frac{e^{-\lambda x}}{\lambda^2} + 2\cdot\frac{e^{-\lambda x}}{-\lambda^3}\right]_0^{\infty}$$

$$= \lambda^2\left[(0-0+0)-(0-0-\tfrac{2}{\lambda^3})\right]$$

$$= \frac{2}{\lambda}$$

(ii) variance of the distribution, $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$

i.e., $\sigma^2 = \int_0^{\infty} x^2 \cdot \lambda^2 x e^{-\lambda x} dx - \frac{4}{\lambda^2}$

$= \lambda^2 \int_0^{\infty} x^3 e^{-\lambda x} dx - \frac{4}{\lambda^2}$

$= \lambda^2 \left[x^3 \cdot \frac{e^{-\lambda x}}{-\lambda} - 3x^2 \cdot \frac{e^{-\lambda x}}{\lambda^2} + 6x \cdot \frac{e^{-\lambda x}}{-\lambda^3} - 6 \frac{e^{-\lambda x}}{\lambda^4} \right]_0^{\infty}$

$= \lambda^2 \left[\frac{6}{\lambda^4} \right] - \frac{4}{\lambda^2}$

$= \frac{2}{\lambda^2}$.

(4) For the continuous probability function $f(x) = kx^2 e^{-x}$ when $x \geq 0$, find (i) k (ii) Mean (iii) variance

[Ans: $k = \frac{1}{2}$, Mean $\mu = 3$, Variance $\sigma^2 = 3$].

(5) Is the function defined by $f(x) = \begin{cases} 0, & x < 2 \\ \frac{1}{18}(2x+3), & 2 \leq x \leq 4 \\ 0, & x > 4 \end{cases}$ a probability density function? Find the probability that a variate having f(x) as density function will fall in the interval $2 \leq x \leq 3$.

Sol: i) For all points x in $-\infty \leq x \leq \infty$, $f(x) \geq 0$ and

$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{2} 0 \cdot dx + \int_2^4 \frac{1}{18}(2x+3)dx + \int_4^{\infty} 0 \cdot dx$

$= \frac{1}{18}\left[2\left(\frac{x^2}{2}\right)_2^4 + 3(x)_2^4 \right]$

$= \frac{1}{18}\left[(16-4) + 3(4-2) \right] = \frac{1}{18}[12+6] = 1$.

Hence f(x) is a probability density function.

ii) The probability that the density will fall in the interval $2 \leq x \leq 3$.

(ii) The probability that the density will fall in the interval $2 \leq x \leq 3$ is (60)

$P(2 \leq x \leq 3) = \int_2^3 f(x)dx = \frac{1}{18}\int_2^3 (2x+3)dx = \frac{1}{18}\left[2\left(\frac{x^2}{2}\right) + 3x \right]_2^3$

$= \frac{1}{18}[5+3] = \frac{4}{9}$

(6) A continuous random variable x has the distribution function $F(x) = \begin{cases} 0, & \text{if } x \leq 1 \\ k(x-1)^4, & \text{if } 1 < x \leq 3 \\ 1; & \text{if } x > 3 \end{cases}$

Determine (i) f(x) (ii) k (iii) Mean.

Sol: (i) We know that $f(x) = \frac{d}{dx}[F(x)]$.

$\therefore f(x) = \begin{cases} 0; & \text{if } x \leq 1 \\ 4k(x-1)^3, & \text{if } 1 < x \leq 3 \\ 0; & \text{if } x > 3. \end{cases}$

(ii) Since total probability is unity, we have

$\int_1^3 f(x)dx = 1$ i.e., $4k\int_1^3 (x-1)^3 dx = 4k\left[\frac{(x-1)^4}{4}\right]_1^3 = 1$

$\Rightarrow k(16-0) = 1$ (or) $k = \frac{1}{16}$.

(iii) Mean of x, $\mu = \int_{-\infty}^{\infty} x f(x)dx = \int_1^3 x \cdot 4k(x-1)^3 dx$

$= 4k\left[x \cdot \frac{(x-1)^4}{4} - \frac{(x-1)^5}{20} \right]_1^3$

$= 4k\left[\frac{1}{4} \cdot 3(16) - \frac{1}{4}(0) - \frac{1}{20}(32) + 0 \right]$

$= 4k\left[12 - \frac{8}{5} \right]$

$= 4\left(\frac{1}{16}\right)\left[\frac{52}{5}\right]$

$= \frac{13}{5}$.

(7) For the continuous random variable x whose probability density function given by $f(x) = \begin{cases} cx(2-x); & \text{if } 0 \le x \le 2 \\ 0; & \text{otherwise} \end{cases}$ where 'c' is a constant. Find c, Mean and variance of x.

[Ans: $c = \frac{3}{4}$, $E(x) = \mu = 1$, $V(x) = \sigma^2 = \frac{1}{5}$]

(8) The probability density function $f(x)$ of a continuous random variable is given by $f(x) = k e^{-|x|}$, $-\infty < x < \infty$. Find k, Mean and variance of the distribution and also find the probability that the variate lies b/n 0 & 4.

Sol: Given $f(x) = k e^{-|x|}$, $-\infty < x < \infty$

we have $\int_{-\infty}^{\infty} f(x)\,dx = 1$

[$\because$ If $x<0$, $e^{-|x|} = e^{x}$; $x \ge 0$, $e^{-|x|} = e^{-x}$]

i.e., $\int_{-\infty}^{\infty} k e^{-|x|}\,dx = 1 \Rightarrow k\left[\int_{-\infty}^{0} e^{-(-x)}\,dx + \int_{0}^{\infty} e^{-x}\,dx\right]$

$\Rightarrow k\left[\int_{-\infty}^{0} e^{x}\,dx + \int_{0}^{\infty} e^{-x}\,dx\right] = 1 \Rightarrow k\left[(e^{x})_{-\infty}^{0} + \left(\frac{e^{-x}}{-1}\right)_{0}^{\infty}\right] = 1$

$\Rightarrow k(1+1) = 1$ (or) $k = \frac{1}{2}$

Hence $f(x) = k e^{-|x|} = \frac{1}{2} e^{-|x|}$.

Mean of the distribution, $\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \frac{1}{2}\int_{-\infty}^{\infty} x e^{-|x|}\,dx = 0$

[$\because$ Integrand is an odd fun]

Variance $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{\infty} (x-0)^2 \cdot \frac{1}{2} e^{-|x|}\,dx$.

$= \frac{1}{2}\int_{-\infty}^{\infty} x^2 e^{-|x|}\,dx = 2 \cdot \frac{1}{2}\int_{0}^{\infty} x^2 e^{-|x|}\,dx$ [$\because$ Integrand is an even function]

$\Rightarrow \int_{0}^{\infty} x^2 e^{-x}\,dx = \left[x^2 \cdot \frac{e^{-x}}{-1} - 2x \cdot \frac{e^{-x}}{1} + 2 \cdot \frac{e^{-x}}{-1}\right]_{0}^{\infty} = 2$

[$\because$ In $0 \le x \le 4$, $|x| = x$]

$P(0 \le x \le 4) = \frac{1}{2}\int_{0}^{4} e^{-|x|}\,dx = \frac{1}{2}\int_{0}^{4} e^{-x}\,dx = \frac{1}{2}(1 - e^{-4}) = 0.4908$ (nearly)

UNIT - III

Probability Distributions

1. probability Distribution
2. Binomial Distribution
3. poisson Distribution
4. Normal distribution
5. properties

Binomial Distribution:

· Binomial distribution is a discrete distribution. It was introduced by James Bernouli in the year 1970.

Definition:

A random variable 'x' has Binomial distribution if it assumes only non -ve values and its probability distribution $p(X=x)$.

$$\therefore \; p(X=x) = {}^{n}C_{x} \, p^{x} \cdot q^{n-x} \; ; x = 0,1,2,\ldots,n \; ; q = 1-p$$

Here n, p are called parameters.

Assumptions (or) Conditions (or) postulates

We get the Binomial Distribution under the following experiment conditions.

1. Each trial has only two possible outcomes called as "SUCCESS" (or) "FAILURE".
2. No. of trials 'n' is finite.
3. The trials are independent to each other.
4. The probability of success 'p' is constant in each trial.

EX:
1. A coin toss has only two possible outcomes: heads or tails and taking a test could have two possible outcomes: pass or fail.
2. If a new drug is introduced to cure a disease, it either cures the disease (it's successful) or it doesn't cure the disease (it's failure
3. Lottery ticket.
4. Result of an exam (pass or fail).

Alternate notation:

$$P(X=x)=p(x)=\begin{cases} {}^nC_x\, p^x\, q^{n-x} & ; x=0,1,2,\ldots,n \; ; q=1-p \\ 0 & ; \text{otherwise.} \end{cases}$$

Constants of Binomial distribution:

1. Mean of the Binomial distribution:

The Binomial probability distribution is given by

$$p(x) = {}^nC_x\, p^x\, q^{n-x} \; ; x=0,1,2,\ldots,n \;\&\; q=1-p.$$

Mean of X, $\mu = E(X) = \sum_{x=0}^{n} x \cdot {}^nC_x \cdot p^x \cdot q^{n-x}$ $\quad [\because E(X)=\sum_{x=0}^{n} x\, p(x)]$

$$= 0\cdot q^n + 1\cdot {}^nC_1\cdot p^1\cdot q^{n-1} + 2\cdot {}^nC_2\cdot p^2\cdot q^{n-2} + \cdots + n\cdot {}^nC_n\cdot p^n\cdot q^{n-n}$$

$$= npq^{n-1} + n(n-1)\cdot p^2 q^{n-2} + \cdots + n\cdot p^n$$

$$= np\left[q^{n-1} + (n-1)pq^{n-2} + \cdots + p^{n-1}\right]$$

$$= np(q+p)^{n-1} \quad [\because \text{using binomial theorem}]$$

$$= np(1) \quad [\because p+q=1]$$

$$\boxed{\therefore \mu = np}$$

2. Variance of Binomial distribution:

Variance, $V(X) = E(X^2) - [E(X)]^2$

$$= \sum x^2 p(x) - \mu^2$$

$$V(X) = \sum_{x=0}^{n} \left[{}^nC_x\, p^x\, q^{n-x}\right] x^2 - \mu^2$$

$$= \sum_{x=0}^{n} \left[{}^nC_x\cdot p^x\cdot q^{n-x}\right]\left[x(x-1)+x\right] - \mu^2$$

$$= \sum_{x=0}^{n} {}^nC_x\, p^x\cdot q^{n-x}\, x(x-1) + \sum_{x=0}^{n} {}^nC_x\cdot p^x\cdot q^{n-x}\cdot x - \mu^2$$

$$= 1\cdot 2\cdot {}^nC_2\cdot p^2 q^{n-2} + 2\cdot 3\cdot {}^nC_3\cdot p^3 q^{n-3} + 3\cdot 4\cdot {}^nC_4\cdot p^4 q^{n-4} + \cdots + {}^nC_n\, n(n-1)p^n + \mu - \mu^2$$

$$\Rightarrow \left[2\cdot \frac{n(n-1)}{2}\, p^2 q^{n-2} + 6\cdot \frac{n(n-1)(n-2)}{6}\, p^3 q^{n-3} + \cdots + n(n-1)p^n\right] + \mu - \mu^2$$

$$\Rightarrow n(n-1)p^2\left[q^{n-2} + (n-2)p\, q^{n-3} + \cdots + p^{n-2}\right] + \mu - \mu^2$$

$$= n(n-1)p^2(q+p)^{n-2} + \mu - \mu^2 \quad \text{(by Binomial Theorem)}$$

$$= n(n-1)p^2(1)^{n-2} + np - (np)^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np(1-p)$$

$$= npq$$

$\therefore$ variance of the Binomial distribution $= \sqrt{npq}$.

Mode of the Binomial distribution:

Mode of the binomial distribution is the value of x at which $p(x)$ has maximum value.

$$\text{Mode} = \begin{cases} \text{Integral part of } (n+1)p, & \text{if } (n+1)p \text{ is not an integer.} \\ (n+1)p \text{ and } (n+1)p-1, & \text{if } (n+1)p \text{ is an integer.} \end{cases}$$

Recurrence relation for Binomial distribution:

$$p(x+1) = \frac{(n-x)p}{(x+1)q}\cdot p(x).$$

① A fair coin is tossed six times. Find the probability of getting four heads.

Sol: p = probability of getting a head = $1/2$

q = probability of not getting a head = $1/2$

and $n = 6$, $r = 4$

We know that $p(r) = {}^{n}C_{r} \cdot p^{r} \cdot q^{n-r}$

$\therefore p(4) = {}^{6}C_{4}\left(\frac{1}{2}\right)^{4}\left(\frac{1}{2}\right)^{(6-4)} = \frac{6!}{4!\,2!}\left(\frac{1}{2}\right)^{6} = \frac{15}{64} = 0.2344.$

② Ten coins are thrown simultaneously. Find the probability of getting at least (i) seven heads (ii) ≤ 2 heads

Sol: p = probability of getting a head = $1/2$

q = probability of not getting a head = $1/2$

The probability of getting 'r' heads in a throw of 10 coin is

$P(X=x) = p(x) = {}^{10}C_{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{10-x}$; $x = 0, 1, 2, \ldots, 10$.

i) probability of getting at least seven heads is given by $p(x \geq 7)$

$= p(x=7) + p(x=8) + p(x=9) + p(x=10)$.

$= {}^{10}C_{7}\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{3} + {}^{10}C_{8}\left(\frac{1}{2}\right)^{8}\left(\frac{1}{2}\right)^{2} + {}^{10}C_{9}\left(\frac{1}{2}\right)^{9}\left(\frac{1}{2}\right) + {}^{10}C_{10}\left(\frac{1}{2}\right)^{10}$

$= \frac{1}{2^{10}}\left[{}^{10}C_{7} + {}^{10}C_{8} + {}^{10}C_{9} + {}^{10}C_{10}\right]$

$= \frac{1}{2^{10}}\left[120 + 45 + 10 + 1\right]$

$= \frac{176}{1024} = 0.1719.$

③ A die is thrown 6 times. If getting an even number is a success, find the probabilities of (i) at least one success (ii) ≤ 3 success (iii) 4 successes.

Sol: In a single of throw of a die, an even number can occur in 3 ways out of 6 ways. Thus

p = probability of occurrence of an even number in one throw

$= \frac{3}{6} = \frac{1}{2}$

n = no of trials = 6

(i) $P(x \geq 1) = 1-$

ii) $p(x \leq 3) = p(x=0) + p(x=1) + p(x=2) + p(x=3)$

$= {}^{6}C_{0}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{6} + {}^{6}C_{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{5} + {}^{6}C_{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{4} + {}^{6}C_{3}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right)^{3}$

$= \left(\frac{1}{2}\right)^{6}\left[1 + 6 + \frac{6(5)}{2} + \frac{6(5)(4)}{6}\right]$

$= \left(\frac{1}{2}\right)^{6}\left[1 + 6 + 15 + 20\right]$ $\quad \because {}^{n}C_{r} = \frac{n!}{(n-r)!\,r!}$

$= \frac{21}{32} = 0.6562$

(iii) $P(x=4) = {}^{6}C_{4}\left(\frac{1}{2}\right)^{4}\left(\frac{1}{2}\right)^{2} = \frac{15}{64} = 0.2344$.

i) $P(X \geq 1) = 1 - P(X=0) = 1 - {}^{6}C_{0}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{6}$

$= 1 - \left(\frac{1}{2}\right)^{6}$

$= \frac{63}{64} = 0.9844$.

④ If the probability of a defective bolt is 0.2, find (i) mean (ii) standard deviation for the distribution of bolts in a total of 400.

Sol: Given $n = 400$, $p = 0.2$. $\quad \therefore q = 1-p = 1-0.2 = 0.8$

(i) Mean $= np = 400(0.2) = 80$ — ①

(ii) S.D $= \sqrt{npq} = \sqrt{80(0.8)} = \sqrt{64} = 8$ [by ①].

⑤ If the probability of a defective bolt is $1/8$, find: (i) the mean (ii) the variance for the distribution of defective bolts of 640.

Sol: Given $n = 640$, p = the probability of a defective bolt = $1/8$

$q = 7/8$

$\therefore \mu = np = 640/8 = 80$ & variance $= npq = 80\left(\frac{7}{8}\right) = 70$.

⑥ The mean and variance of a binomial distribution are 4 and $\frac{4}{3}$ respectively. Find $p(x \geq 1)$

Sol: Mean of the binomial distribution $= 4$, i.e, $np = 4$ —①

& variance of the binomial distribution $= \frac{4}{3}$, i.e, $npq = \frac{4}{3}$ —②

② ÷ ① gives, $\frac{npq}{np} = \frac{(4/3)}{4} = 1/3$

$q = 1/3$ & $p = 1 - q = 2/3$.

From ①, $n = \frac{4}{p} = 4\left(\frac{3}{2}\right) = 6$.

$\therefore P(x \geq 1) = 1 - p(x < 1) = 1 - p(x=0) = 1 - {}^6C_0\, p^0 q^{6-0} = 1 - \left(\frac{1}{3}\right)^6$
$= 0.9986$

⑦ In eight throws of a die 5 or 6 is considered a success. Find the mean number of success and standard deviation.

Sol: p = The probability of success $= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

q = The probability of failure $= 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

n = no of throws $= 8$.

$\therefore$ mean $= np = 8\left(\frac{1}{3}\right) = \frac{8}{3}$

Variance $= npq = (np)q = \frac{8}{3}\left(\frac{2}{3}\right) = \frac{16}{9}$

Hence Standard deviation $= \sqrt{\text{variance}} = \sqrt{\frac{16}{9}} = \frac{4}{3}$.

⑧ Out of 800 families with 5 children each, how many would you expect to have (a) 3 boys (b) 5 girls (c) either 2 or 3 boys (d) atleast one boy? Assume equal probabilities for boys and girls

Sol: p = The probability of each boy $= \frac{1}{2}$

No. of children, $n = 5$

The P.D is $p(x) = {}^nC_x\, p^x q^{n-x} = {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x} = \frac{1}{2^5} \cdot {}^5C_x$

a) $P(3 \text{ boys}) = p(x=3) = p(3) = \frac{1}{2^5}\, {}^5C_3 = \frac{5}{16}$ per family.

Thus for 800 families the probability of no. of families having 3 boys $= \frac{5}{16}(800) = 250$ family

b) $P(5 \text{ girls}) = P(\text{no boys}) = P(x=0) = p(0) = \frac{1}{2^5}\, {}^5C_0 = \frac{1}{32}$ per family

Thus for 800 families the probability of no of families having 5 girls $= \frac{1}{32}(800) = 25$ family.

c) $P(\text{either 2 or 3 boys}) = p(x=2) + p(x=3) = p(2) + p(3)$

$= \frac{1}{2^5} \cdot {}^5C_2 + \frac{1}{2^5} \cdot {}^5C_3$

$= \frac{5}{8}$ per family.

$\therefore$ Expected no of families with 2 or 3 boys $= \frac{5}{8}(800)$
$= 500$ family

d) $P(\text{at least one boy}) = P(x \geq 1)$

$= 1 - P(x=0)$

$= 1 - \frac{1}{32} = \frac{31}{32}$.

$\therefore$ expected no of families with atleast one boy $= \frac{31}{32}(800)$
$= 775$.

⑨ Out of 800 families with 4 children each, how many families would be expected to have (a) 2 boys and 2 girls (b) atleast one boy (c) no girl (d) atmost two girls? Assume equal probabilities for boy and girls?

⑩ 20% of items produced from a factory are defective. Find the probability that in a sample of 5 chosen at random.
(i) none is defective (ii) one is defective (iii) $p(1<x<4)$.

Sol: probability of defective items $= p = 20\% = 0.2$.

probability of non defective items $= q = 1 - p = 1 - 0.2 = 0.8$.

Total no. of items, $n = 5$.

Total no. of items, $n=5$

(i) probability that none is defective = probability of 0 defective item

$= p(0)$

$= {}^5C_0 (0.2)^0 (0.8)^5$

$= (0.8)^5 = 0.3277$

(ii) probability of 1 defective item $= p(1) = {}^5C_1 (0.2)^1 (0.8)^4$

$= 5(0.2)(0.4096)$

$= 0.4096$.

(iii) $p(1<x<4) = p(2) + p(3)$

$= {}^5C_2 (0.2)^2 (0.8)^3 + {}^5C_3 (0.2)^3 (0.8)^2$

$= (0.2)^2 (0.8)^2 [10(0.8) + 10(0.2)]$

$= 0.0256 [8+2]$

$= 0.256$.

(11) Fit a Binomial distribution to the following frequency distribution

x	0	1	2	3	4	5	6
f	13	25	52	58	32	16	4

Sol: Here n = no. of trials $=6$ and N = total frequency $= \Sigma f_i = 200$

$\therefore$ Mean $= \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{25+104+174+128+80+24}{200} = \dfrac{535}{200} = 2.675$

Now mean of the binomial distribution $= np$

i.e., $np = 6p = 2.675$

$\therefore p = \dfrac{2.675}{6} = 0.446$, and $q = 1-p = 1-0.446 = 0.554$

Hence the binomial distribution to be fitted is

$N(q+p)^n = 200(0.554+0.446)^6 = 200[(0.554)^6 + {}^6C_1 (0.554)^5 (0.446)$

$+ {}^6C_2 (0.554)^4 (0.446)^2 + \cdots + {}^6C_6 (0.446)^6]$

$= 5.782 + 27.92 + 56.18 + 60.32 + 36.42 + 11.728 + 1.5732$

The expected frequencies can be rounded off to the nearest integer to get expected frequencies as whole numbers.

$\therefore$ The successive terms in the expansion give the expected or theoretical frequencies which are

x	0	1	2	3	4	5	6
f	13	25	52	58	32	16	4
Expected or Theoretical frequency	6	28	56	60	36	12	2

(12)

poisson Distribution

poisson distribution was discovered by the french mathematician "Simeon Denis poisson in 1837."

The poisson distribution Can be derived as a limiting Case of the Binomial Distribution under the following conditions.

(i) p, the probability of the occurrence of the event is very Small.

(ii) n is very very large, where n is number of trials. i.e, $n \to \infty$

(iii) np is a finite quantity, Say $np = \lambda$, then λ is Called the parameter of the poisson distribution.

Definition:

A random variable x is said to follow a poisson distribution if it assumes only non-negative values and its probability density function is given by

$$p(x,\lambda) = p(X=x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; x = 0,1,2,3, \dots \\ 0 & ; \text{otherwise.} \end{cases}$$

Here $\lambda > 0$ is Called the parameter of the distribution.

Examples of poisson Distribution:

i) The number of persons born blind per year in a large city.

ii) The no. of Cars passing a Certain point in one minute.

iii) The no. of printing mistakes per page in a large text.

Constants of poisson Distribution:

1. The Mean of the poisson Distribution.

$$\text{Mean} = E(X) = \sum_{x=0}^{\infty} x \cdot p(x) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^{x}}{x!} = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^{x}}{x(x-1)!} = e^{-\lambda}\sum_{x=1}^{\infty} \frac{\lambda^{x}}{(x-1)!}$$

[putting $x-1=y$]

$$= e^{-\lambda}\sum_{y=0}^{\infty} \frac{\lambda^{y+1}}{y!}$$

$$= e^{-\lambda}\lambda\sum_{y=0}^{\infty} \frac{\lambda^{y}}{y!}$$

$= \lambda e^{-\lambda} \cdot e^{\lambda}$ $\qquad \left[\because \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{\lambda}\right]$

$= \lambda \quad (=np)$ $\qquad \left[\because e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots\right]$

Thus the parameter 'λ' is the "Arithmetic Mean" of the Poisson distribution.

2. **Variance of poisson Distribution:**

$V(x) = E(x^2) - [E(x)]^2$

$= \sum_{x=0}^{\infty} x^2 p(x) - \lambda^2 \qquad [\because \lambda = \text{mean of P.D}]$

$= \sum_{x=0}^{\infty} x^2 \cdot \frac{e^{-\lambda} \lambda^x}{x!} - \lambda^2$

$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{x^2 \lambda^x}{x(x-1)!} - \lambda^2$

$= e^{-\lambda} \sum_{x=1}^{\infty} \frac{x \cdot \lambda^x}{(x-1)!} - \lambda^2$

$= e^{-\lambda} \sum_{x=1}^{\infty} \frac{[x-1+1] \lambda^x}{(x-1)!} - \lambda^2$

$= e^{-\lambda} \left\{ \sum_{x=1}^{\infty} [(x-1)+1] \frac{\lambda^x}{(x-1)!} \right\} - \lambda^2$

$= e^{-\lambda} \left\{ \sum_{x=1}^{\infty} (x-1) \frac{\lambda^x}{(x-1)!} + \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \right\} - \lambda^2$

$= e^{-\lambda} \left\{ \sum_{x=2}^{\infty} \frac{(x-1)\lambda^x}{(x-1)(x-2)!} + \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \right\} - \lambda^2$

$= e^{-\lambda} \left\{ \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!} + \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \right\} - \lambda^2$

$= e^{-\lambda} \left[\sum_{y=0}^{\infty} \frac{\lambda^{y+2}}{y!} + \sum_{z=0}^{\infty} \frac{\lambda^{z+1}}{z!} \right] - \lambda^2$ (putting $y = x-2$; $z = x-1$)

$= e^{-\lambda} \lambda^2 \left[\sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \right] + e^{-\lambda} \cdot \lambda \left[\sum_{z=0}^{\infty} \frac{\lambda^z}{z!} \right] - \lambda^2$

$= e^{-\lambda} \lambda^2 e^{\lambda} + e^{-\lambda} \lambda \, e^{\lambda} - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Thus variance $= \lambda$

Hence the variance of the distribution = mean of the distribution $= \lambda$

Further, Standard deviation of the Poisson distribution $\sigma = \sqrt{\lambda}$.

3. **Mode of the poisson Distribution:**

Mode is the value of x for which the probability $p(x)$ is maximum.

Mode of the poisson distribution lies between $(\lambda-1)$ and λ.

i.e., $\lambda - 1 \le \text{Mode} \le \lambda$

I) If 'λ' is an Integer then $\lambda-1$ is also an Integer. So we have two maximum values and the distribution is bimodal and the two modes are $(\lambda-1)$ and λ.

II) If 'λ' is not an Integer, the mode of poisson distribution is Integral part of λ.

Recurrence relation for the poisson distribution:

$$p(x+1) = \left[\frac{\lambda}{x+1}\right] p(x)$$

(or) $\quad p(x) = \frac{\lambda}{x} p(x-1)$.

problems:

① 2% of the items of a factory are defective. The items are packed in boxes. what is the probability that there will be

(i) 2 defective items (ii) at least three defective items in a box of 100 items?

Sol: Given $n = 100$

and p = The probability of defective items $= 2\% = 0.02$

λ = Mean no of defective items in a box of $100 = np = (0.02)(100) = 2$.

Since 'p' is small, we may use poisson distribution. probability of 'x' defective items in a box of 100 is

$$P(X=x) = p(x) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-2} 2^x}{x!}$$

i) $p(2 \text{ defective items}) = p(2) = \frac{e^{-2} 2^2}{2!} = \frac{2}{e^2} = 0.2706$

ii) $p(\text{at least } 3) = p(x \geq 3) = 1 - [p(x=0) + p(x=1) + p(x=2)]$

$$= 1 - e^{-2}\left[\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!}\right]$$

$$= 1 - 5e^{-2} = 0.3233$$

(2) Average number of accidents on any day on a national highway is 1.8. Determine the probability that the no. of accidents are (i) atleast one (ii) at most one

Sol: Mean, $\lambda = 1.8$

we have, $p(x=x) = p(x) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-1.8} (1.8)^x}{x!}$

i) $p(\text{atleast one}) = p(x \geq 1) = 1 - p(x=0) = 1 - e^{-1.8} = 1 - 0.1653 = 0.8347$

ii) $p(\text{at most one}) = p(x \leq 1) = p(x=0) + p(x=1) = e^{-1.8} + e^{-1.8}(1.8) = 0.4628$

(3) A sample of 3 items is selected at random from a box containing 10 items of which 4 are defective. Find the expected no of defective items?

Sol: The probability of defective is $p = \frac{4}{10} = \frac{2}{5}$

No of items chosen, $n = 3$

The expected no of defective is $E(x) = \mu = np = 3\left(\frac{2}{5}\right) = \frac{6}{5} = 1.2 \approx 1$

(4) The wireless sets are manufactured with 25 soldered joints each on the average 1 joint in 500 is defective. How many sets can be expected to be free from defective joints in a consignment of 10,000 sets.

Sol: on average, one joint in 500 joints is defective

probability that a joint is defective $= \frac{1}{500}$

No of soldered joints $= 25 = n$

$\therefore$ Mean $= np = 25 \times \frac{1}{500} = 0.05 = \lambda$

$p(x \text{ defective joints}) = \frac{e^{-\lambda} \lambda^x}{x!} \Rightarrow p(x=0) = \frac{e^{-\lambda} \lambda^0}{0!} = e^{-\lambda} = e^{-0.05}$

Hence expected no of sets free from defective joints among 10000 sets $= 10000 \times e^{-0.05} = 9512.29$

(5) If a bank received on the average 6 bad cheques per day, find the probability that it will receive 4 bad cheques on any given day. (28)

Sol: We have $p(x=x) = \frac{e^{-\lambda} \lambda^x}{x!}$

Here $\lambda = 6$.

$\therefore p(x=4) = \frac{e^{-6} 6^4}{4!} = \frac{54}{e^6} = 0.1339$

(6) Using recurrence formula find the probabilities when $x = 0, 1, 2, 3, 4$ and 5; If the mean of poisson distribution is 3.

Sol: Given mean of the poisson distribution is 3

$\Rightarrow \lambda = 3$.

Now the poisson distribution is

$$p(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

$\therefore p(x=0) = \frac{e^{-3} 3^0}{0!} = e^{-3} = 0.0498$

By recurrence formula,

$p(x+1) = \frac{\lambda}{x+1} p(x) \Rightarrow p(x+1) = \frac{3}{x+1} p(x);\ [\because \lambda = 3]$ ---(1)

$\Rightarrow$ put $x=0$ in (1), then

$p(1) = 3 \cdot p(0) = 0.1494$

put $x=1$ in (1), then $p(2) = \frac{3}{2}(0.1494) = 0.2241$

Similarly, $p(3) = 0.2241$, $p(4) = 0.1681$, $p(5) = 0.1008$.

(7) Fit a poisson distribution to the following data and calculate the expected frequencies

x	0	1	2	3	4
f(x)	109	65	22	3	1

Sol: Here N = total frequency $= \Sigma f_i = 109 + 65 + 22 + 3 + 1 = 200$

Mean $= \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{0 + 65 + 44 + 9 + 4}{200} = \frac{122}{200} = 0.61$

$\therefore$ Mean of P.D, $\lambda = 0.61$

Hence the theoretical frequencies are given by $N \cdot P(x)$, where $x = 0, 1, 2, 3, 4$

i.e., $200 \cdot \frac{e^{-0.61}(0.61)^x}{x!}$, where $x = 0, 1, 2, 3, 4$

i.e., $200 \cdot e^{-0.61}$, $200 e^{-0.61}(0.61)$, $200 e^{-0.61}\frac{(0.61)^2}{2!}$, $200 e^{-0.61}\frac{(0.61)^3}{3!}$, $200 e^{-0.61}\frac{(0.61)^4}{4!}$

i.e., 108.67, 66.29, 20.22, 4.11, 0.63

Since frequencies are always integers, therefore by converting them to nearest integers, we get

observed frequency	109	65	22	3	1
expected frequency	109	66	20	4	1

(6) If a random variable has a P.D s.t $p(1) = p(2)$, find i) mean ii) $p(4)$ (iii) $p(x \geq 1)$ iv) $p(1 < x < 4)$.

Sol: Since $p(1) = p(2)$

$$\frac{e^{-\lambda}\lambda^1}{1!} = \frac{e^{-\lambda}\lambda^2}{2!}$$

$\Rightarrow \lambda^2 - 2\lambda = 0$

$\therefore \lambda = 0$ (or) 2

Since $\lambda \neq 0$, $\therefore \lambda = 2$

i) $P(x=4) = \frac{e^{-2}2^4}{4!} = 0.0902$

ii) $P(x \geq 1) = 1 - P(x=0) = 1 - \frac{e^{-2}2^0}{0!} = 0.8647$

iii) $P(1<x<4) = P(x=2) + P(x=3)$

$$= \frac{e^{-2}2^2}{2!} + \frac{e^{-2}2^3}{3!}$$

$= 0.4511$

(7) If a P.D is such that $\frac{3}{2}P(x=1) = P(x=3)$, find i) $P(x \geq 1)$ (ii) $P(x \leq 3)$ (iii) $P(2 \leq x \leq 5)$.

NORMAL DISTRIBUTION

The normal distribution was first discovered by English Mathematician De-Moivre (1667-1754) in 1733 and further refined by French Mathematician Laplace (1749-1827) in 1774 and independently by Karl Friedrich Gauss (1777-1855). Normal distribution is also known as Gaussian distribution. Normal distribution is a continuous distribution.

The normal distribution is, therefore derived from the B.D by increasing the large no. of trials indefinitely.

Definition: A random variable 'X' is said to have a normal distribution, if its density function or probability distribution is given by

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad -\infty < x < \infty, \; -\infty < \mu < \infty, \; \sigma > 0$$

where μ is the mean, σ is the standard deviation of x.

* N.D is a limiting case of B.D under the following conditions:

i) n, the no. of trials is indefinitely large i.e, $n \to \infty$.

ii) neither p nor q is very small.

Constants of Normal Distribution:

1. Mean of Normal Distribution

Consider the Normal Distribution with b, σ as the parameters.

Then $f(x;b,\sigma) = \frac{e^{-\frac{(x-b)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$

The mean $\mu = E(x)$ is given by

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x \cdot e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2}dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma z + b)e^{-z^2/2}dz \qquad \left[\text{putting } z = \frac{x-b}{\sigma} \Rightarrow dz = \frac{dx}{\sigma} \Rightarrow \sigma z + b = x\right]$$

$$= \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z e^{-z^2/2}dz + \frac{b}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-z^2/2}dz$$

$$= 0 + \frac{b}{\sqrt{2\pi}} \cdot 2\int_0^\infty e^{-\frac{z^2}{2}}\,dz \qquad \left[\because z e^{-\frac{z^2}{2}} \text{ is odd function and } e^{-\frac{z^2}{2}} \text{ is even function}\right]$$

$$= \frac{2b}{\sqrt{2\pi}} \cdot \frac{\sqrt{\pi}}{\sqrt{2}} \qquad \left[\because \int_0^\infty e^{-\frac{z^2}{2}}\,dz = \sqrt{\frac{\pi}{2}}\right]$$

$$= b$$

$\therefore$ **Mean, $\mu = b$**

2. Variance of Normal Distribution:

By definition, variance $= E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-b)^2 \, e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^2 \, e^{-\frac{z^2}{2}} dz \qquad \left[\text{putting } z = \frac{x-b}{\sigma} \text{ so that } dz = \frac{dx}{\sigma}\right]$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \, e^{-\frac{z^2}{2}} dz$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{-\frac{z^2}{2}} dz \qquad [\because \text{Integrand is even function}]$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^{\infty} 2t \, e^{-t} \frac{dt}{\sqrt{2t}} \qquad \left[\text{putting } \frac{z^2}{2} = t \text{ so that } z\,dz = dt,\ dz = \frac{dt}{\sqrt{2t}}\right]$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} \sqrt{t}\, e^{-t} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} e^{-t}\, t^{\frac{3}{2}-1} dt \qquad \left[\because \Gamma(n) = \int_0^\infty e^{-x} x^{n-1} dx,\ \Gamma(n) = (n-1)\Gamma(n-1)\right]$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right)$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2}\sqrt{\pi} = \sigma^2$$

Hence, variance $= \sigma^2$.

Thus the standard deviation of the N.D is σ.

3. Mode of Normal Distribution:

Mode is the value of x for which $f(x)$ is maximum. i.e, mode is the solution of $f'(x) = 0$ & $f''(x) < 0$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

Diff. w.r.t x, we get

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot \left[-\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}\right] = -\frac{x-\mu}{\sigma^2} f(x).$$

Now, $f'(x) = 0 \Rightarrow x - \mu = 0$. i.e, $x = \mu$.

$$f''(x) = -\frac{1}{\sigma^2}\left[(x-\mu)f'(x) + f(x)\right]$$

$$= -\frac{1}{\sigma^2}\left[(x-\mu)\frac{-(x-\mu)}{\sigma^2} f(x) + f(x)\right]$$

$$= \frac{-f(x)}{\sigma^2}\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]$$

At the point $x = \mu$, we have

$$f''(x) = -\left[\frac{f(x)}{\sigma^2}\right]_{x=\mu} = -\frac{1}{\sigma^2\sqrt{2\pi}\,\sigma} < 0$$

Hence $x = \mu$ is the mode of the N.D.

4. Median of Normal distribution:

If M is the median of the N.D, we have

$$\int_{-\infty}^{M} f(x)\,dx = \frac{1}{2}$$

i.e, $$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{2}$$

Scanned with CamScanner

i.e., $\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx + \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx = \frac{1}{2}$ —— ①

Consider $\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$

put $\frac{x-\mu}{\sigma} = z$. Then $dx = \sigma\, dz$

$$\therefore \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{z^2}{2}}\sigma\, dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{z^2}{2}}dz$$

$$= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{z^2}{2}}dz \quad \text{(by symmetry)}$$

$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\pi}{2}} = \frac{1}{2} \text{ --- ②}$$

From ① & ②, we have

$$\frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx = \frac{1}{2}$$

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx = 0 \Rightarrow \int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx = 0$$

$\Rightarrow \mu = M$ [$\because$ If $\int_a^b f(x)\,dx = 0$ then $a = b$, where $f(x) > 0$]

Note:

For the normal distribution,

Mean = Median = Mode

Hence the distribution is Symmetrical.

5. **Mean deviation from the mean for Normal Distribution:**

By definition, mean deviation (about mean)

$$= \int_{-\infty}^{\infty} |x-\mu|\, f(x)\,dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} |x-\mu| \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

[putting $\frac{x-\mu}{\sigma} = z$]

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} |\sigma z| \cdot e^{-\frac{z^2}{2}}\sigma\, dz$$

$$= \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} |z| e^{-z^2/2}dz = \frac{\sigma}{\sqrt{2\pi}}\cdot 2\int_{0}^{\infty} z\, e^{-z^2/2}dz$$

$$= \sqrt{\frac{2}{\pi}}\,\sigma\int_{0}^{\infty} z\, e^{-z^2/2}dz$$

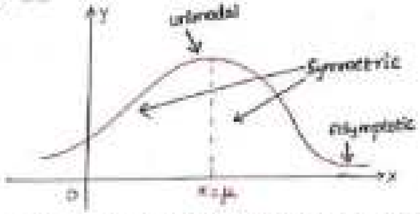
$$= \sqrt{\frac{2}{\pi}}\,\sigma\int_{0}^{\infty} \sqrt{2t}\cdot e^{-t}\,dt \quad \left[\text{putting } \frac{z^2}{2} = t,\ dz = dt\right]$$

$$= \sqrt{\frac{2}{\pi}}\,\sigma\left(-e^{-t}\right)_0^{\infty} = \sqrt{\frac{2}{\pi}}\,\sigma(0+1) = \frac{4}{5}\sigma \text{ (approximately)}$$

Hence the mean deviation from the mean for N.D is equal to $\frac{4}{5}$ times of standard deviation approximately.

Chief Characteristics of the Normal Distribution:

1. The graph of the normal distribution $y = f(x)$ in the xy-plane is known as the normal curve.



2. The curve is a bell shaped curved and symmetrical w.r.t mean i.e., about the line $x=\mu$ and the two tails on the right and the left sides of the mean (μ) extends to infinity. The top of the bell is directly above the mean μ.

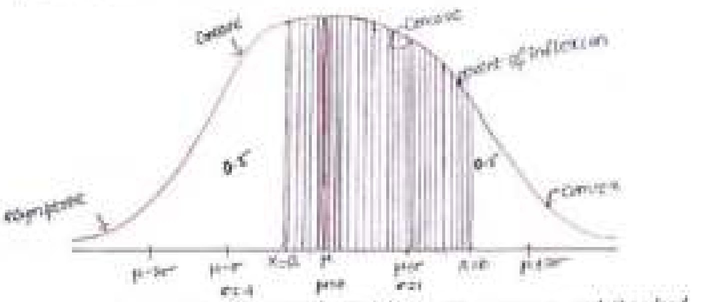
3. Area under the normal curve represents the total population.

4. Mean, median and mode of the distribution coincide at $x=\mu$ as the distribution is symmetrical. So normal curve is unimodal (has only one maximum point)

5. x-axis is an asymptote to the curve (the normal curve approaches, but never touches the x-axis).

6. Linear combination of independent normal variables (variates) is also a normal random variate.

7. The points of inflexion of the curve are at $x=\mu\pm\sigma$ and the curve changes from concave to convex at $x=\mu+\sigma$ to $x=\mu-\sigma$.



8. The probability that the normal variate X with mean μ and standard deviation σ lies b/w x_1 & x_2 is given by

$$P(x_1 \le X \le x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad \text{— ①}$$

By putting $z=\frac{x-\mu}{\sigma}$, the R.H.S of eq ① becomes independent of the two parameters μ & σ. Here z is known as the standard variable.

9. Area of normal curve between

$\mu-\sigma$ and $\mu+\sigma$ is 68.27%. i.e., $P(\mu-\sigma < X < \mu+\sigma) = 0.6826$

$\mu-2\sigma$ and $\mu+2\sigma$ is 95.45%.

$\mu-3\sigma$ and $\mu+3\sigma$ is 99.73%.

10. Mean $\mu=0$ & S.D $\sigma=1$

Importance and applications of the Normal Distribution:

Normal distribution plays a very important role in statistical theory because of the following reasons:

1. Data obtained from psychological, physical and biological measurements approximately follow normal distribution.
   For example, heights, blood pressure, weights of individuals, IQ scores, & measurement errors follow the normal distribution.
2. Since the normal distribution is a limiting case of the Binomial distribution for exceptionally large numbers, it is applicable to many applied problems in kinetic theory of gases and fluctuations in the magnitude of an electric current.
3. Even if a variable is normally distributed, it can sometimes be brought to normal form by simple transformation of the variable.
4. Normal curve is used to find confidence limits of the population parameters.
5. Normal distribution finds large applications in statistical quality control in industry for finding control limits.

Standard Normal Distribution:

The Normal Distribution with mean $\mu = 0$ and $S.D(\sigma) = 1$, is known as standard Normal Distribution.

The random variable that follows this distribution is denoted by z. If a variable x follows normal distribution with mean μ and SD σ, the variable z defined as

$$z = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma z + \mu$$

has standard normal distribution with mean 0 and S.D as 1. This is also referred as z-score.

From the formula of z, we draw conclusions:

i) when $x < \mu$, the value of z is -ve.

ii) when $x > \mu$, the value of z is +ve.

iii) when $x = \mu$, the value of $z = 0$.

To find N.D values in calculator:

Mode – 2times
↓
Reg (2)
exp (3)
Shift (3)

$P \rightarrow -\infty$ to up to the region (or) ∞ to up to the region

$Q \rightarrow 0$ to up-to the value.

$R \rightarrow z$ to ∞ (or) $-z$ to $-\infty$.

Area under the Normal curve (Normal probability Integral) Conversion into standard Normal form:

The probability that random value of x will lie between $x = \mu$ and $x = x_1$ is given by

$$P(\mu < x < x_1) = \int_{\mu}^{x_1} f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{x_1} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= P(0 \le z \le z_1)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{\frac{x_1-\mu}{\sigma}} e^{-\frac{1}{2}z^2} \cdot \sigma\, dz$$

[put $\frac{x-\mu}{\sigma} = z$

$x - \mu = \sigma z$

$dx = \sigma\, dz$]

Limits are 0 to z_1 where $z_1 = \frac{x_1-\mu}{\sigma}$

i.e., $P(\mu < x < x_1) = P(0 < z < z_1)$ $\left[\mu < x < x_1;\ \mu-\mu < x-\mu < x_1-\mu;\ 0 < x-\mu < x_1-\mu;\ 0 < \sigma z < \sigma z_1;\ 0 < z < z_1\right]$

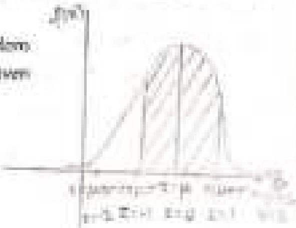
$$= \frac{1}{\sqrt{2\pi}} \int_0^{z_1} e^{-\frac{z^2}{2}} dz = A(z_1) \text{ (say)}.$$

i.e., $A(z_1) = \frac{1}{\sqrt{2\pi}} \int_0^{z_1} e^{-\frac{z^2}{2}} dz.$

The function $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ is the probability function of standard normal variate.

The definite integral $\int_0^{z_1} e^{-\frac{z^2}{2}} dz$ is known as the normal probability integral and gives the area under standard normal curve between $z=0$ and $z=z_1$.

In particular, the probability that a random variable x lies in the interval $(\mu-\sigma, \mu+\sigma)$ is given by $P(\mu-\sigma < x < \mu+\sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x)\,dx$



$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-\sigma}^{\mu+\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$\Rightarrow P(-1 < z < 1) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-z^2/2} dz \qquad \left(\because z = \frac{x-\mu}{\sigma}\right)$$

$$= \frac{1}{\sqrt{2\pi}} 2 \int_0^1 e^{-z^2/2} dz$$

$$= 2\int_0^1 \phi(z)\,dz = 2(0.3413) = 0.6826.$$

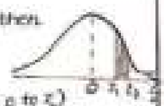
How to find probability density of normal curve:

The probability that the normal variate x with mean μ and S.D σ lies between two specific values x_1 and x_2 with $x_1 < x_2$ can be obtained using area under the standard normal curve as follows:

Step 1: perform the change of scale $z = \frac{x-\mu}{\sigma}$ and find z_1 and z_2 corresponding to the values of x_1 and x_2 respectively.

Step 2(a): To find $P(x_1 < X < x_2) = P(z_1 \le z \le z_2)$

Case 1: If both z_1 and z_2 are positive (or both negative), then



$P(x_1 \le X \le x_2) = |A(z_2) - A(z_1)|$

= (Area under the normal curve from 0 to z_2)

− (Area under the normal curve from 0 to z_1)

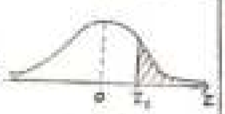
Case 2: If both $z_1 < 0$ and $z_2 > 0$ are positive, then

$P(x_1 \le X \le x_2) = A(z_2) + A(z_1)$.

Step 2(b): To find $P(z > z_1)$

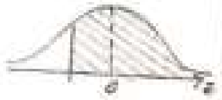
Case 1: If $z_1 > 0$ then,

$P(z > z_1) = 0.5 - A(z_1)$ $\left[\because P(z \le 0) = P(z \ge 0) = \frac{1}{2}\right]$



Case 2: If $z_1 < 0$ then,

$P(z > z_1) = 0.5 + A(z_1)$.

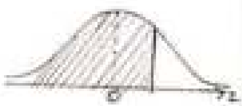


Step 2(c): To find $P(z < z_1) = 1 - P(z > z_1)$

Case 1: If $z_1 > 0$, then $P(z < z_1) = 1 - P(z > z_1)$

$= 1 - [0.5 - A(z_1)]$

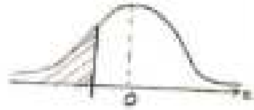
$= 0.5 + A(z_1)$



Case 2: If $z_1 < 0$, then $P(z < z_1) = 1 - P(z > z_1)$

$= 1 - [0.5 + A(z_1)]$

$= 0.5 - A(z_1)$



Problems:

① If x is a normal variate with mean 30 and standard deviation 5. Find the probabilities that (i) $26 \le x \le 40$ (ii) $x \ge 45$.

Sol: Given that, x is a normal variate with mean $\mu = 30$ and S.D $\sigma = 5$.

when $x = 26$, $z = \frac{x-\mu}{\sigma} = \frac{26-30}{5} = -0.8 = z_1$ (say)

when $x = 40$, $z = \frac{x-\mu}{\sigma} = \frac{40-30}{5} = 2 = z_2$ (say)

$\because z_1 < 0$ & $z_2 > 0$

$\therefore p(26 \le x \le 40) = p(-0.8 \le z \le z_2)$

$= A(z_2) + A(z_1)$

$= A(2) + A(-0.8)$

$= 0.4772 + 0.2881$

$= 0.7653$

when $x = 45$, $z = \frac{x-\mu}{\sigma} = \frac{45-30}{5} = 3 = z_1$ (say)

$\therefore p(x \ge 45) = p(z_1 \ge 3) = 0.5 - A(z_1)$

$= 0.5 - A(3)$

$= 0.5 - 0.49865$

$= 0.0013$

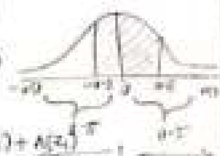
② Suppose the weights of 800 male students are normally distributed with mean $\mu = 140$ pounds and standard deviation 10 pounds. Find the no of students whose weights are

(i) Between 138 and 148 pounds (ii) more than 152 pounds.

Sol: Given that $\mu = 140$ pounds & $\sigma = 10$ pounds.

i) when $x = 138$, $z = \frac{x-\mu}{\sigma} = \frac{138-140}{10} = -0.2 = z_1$ (say)

$x = 148$, $z = \frac{x-\mu}{\sigma} = \frac{148-140}{10} = 0.8 = z_2$ (say)



$\because z_1 < 0$ and $z_2 > 0$.

$\therefore p(138 \le x \le 148) = p(-0.2 \le z \le 0.8) = A(z_2) + A(z_1)$

$= A(0.8) + A(-0.2)$

= Area (-0.2 to 0) + Area (0 to 0.8) = $0.2881 + 0.0793 = 0.3674$

Hence the no. of students whose weights are between 138 pounds and 148 pounds = $0.3674 \times 800 \approx 294$

ii) when $x = 152$, $z = \frac{x-\mu}{\sigma} = \frac{152-140}{10} = 1.2 = z_1$ (say)

$\therefore p(x > 152) = p(z > z_1)$ = Area ($z > 1.2$) = Area (from 0 to ∞) − Area (0 to 1.2)

$= 0.5 - A(z_1) = 0.5 - A(1.2)$

$= 0.5 - 0.3849$

$= 0.1151$

$\therefore$ no. of students whose weights are more than 152 pounds

$= 800 \times 0.1151 = 92$.

③ If the masses of 300 students are normally distributed with mean 68 kgs and standard deviation 3 kgs, how many students have masses

i) Greater than 72 kg (ii) less than or equal to 64 kg

(iii) Between 65 and 71 kg inclusive.

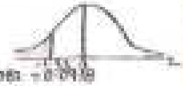
Sol: Here $\mu = 68$ & $\sigma = 3$. Let the variable x denote the masses of students.

i) when $x = 72$, $z = \frac{x-\mu}{\sigma} = \frac{72-68}{3} = 1.33$

$p(x > 72) = p(z > 1.33) = 0.5 - A(1.33) = 0.5 - 0.4082 = 0.0918$

no. of students with more than 72 kgs $= 300 \times 0.0918 = 28$

ii) when $x = 64$, $z = \frac{x-\mu}{\sigma} = \frac{64-68}{3} = -1.33$



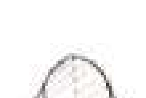
$\therefore p(x \le 64) = p(z \le -1.33) = 0.5 - A(1.33) = 0.5 - 0.4082 = 0.0918$

no. of students have masses less than or equal to 64 kg

$= 300 \times 0.0918 = 28$ (approximately)

iii) when $x = 65$, $z = \frac{x-\mu}{\sigma} = \frac{65-68}{3} = -1 = z_1$ (say)

when $x = 71$, $z = \frac{x-\mu}{\sigma} = \frac{71-68}{3} = 1 = z_2$ (say)



$\therefore p(65 \le x \le 71) = p(-1 \le z \le 1) = A(z_2) + A(z_1) = A(1) + A(-1)$

$= A(1) + A(1)$

$= 2A(1) = 2(0.3413)$

$= 0.6826$.

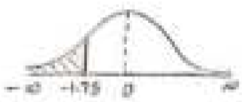
∴ Required no. of Students = 300 × 0.6836 = 205 (approximately).

(4) If x is a normal variate, find the area A

i) to the left of $z = -1.78$ ii) to the right of $z = -1.45$

iii) corresponding to $-0.8 \le z \le 1.53$ iv) to the left of $z = -2.52$ and to the right of $z = 1.83$.

Sol: i) to the left of $z = -1.78$.

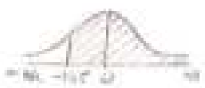


$P(z \le -1.78)$ = Area (from 0 to ∞) − Area (from 0 to 1.78)

= 0.5 − (0.4625) (By Symmetry)

= 0.0375

ii) to the right of $z = -1.45$



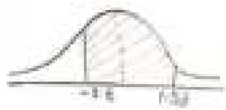
Required Area, A = Area (from 0 to ∞) + Area (from 0 to −1.45)

= Area (from 0 to ∞) + Area (from 0 to 1.45) (By symmetry)

= 0.5 + 0.4265

= 0.9965

iii) Corresponding to $-0.8 \le z \le 1.53$



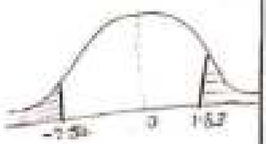
Required Area, A = Area (from 0 to −0.8) + Area (from 0 to 1.53)

= Area (from 0 to 0.8) + Area (from 0 to 1.53)

= 0.2881 + 0.4370

= 0.7251

iv) to the left of $z = -2.52$ and right of $z = 1.83$



Required Area = [Area (from 0 to ∞) − Area (from 0 to −2.52)] + [Area (from 0 to ∞) − Area (from 0 to 1.93)]

= [0.5 − 0.4941] + [0.5 − 0.4664]

= 0.0059 + 0.0336

= 0.0395

(5) A sales tax officer has reported that the average sales of the 500 business that he has to deal with during a year is Rs. 36,000 with a S.D of 10,000. Assuming that the sales in these business are normally distributed, find.

i) the number of business as the sales of which are Rs. 40,000.

ii) the percentage of business the sales of which are likely to range between Rs. 30,000 and Rs. 40,000.

Sol: Let μ be the mean and σ the Standard deviation of the Sales. Then we are given that $\mu = 36{,}000$ & $\sigma = 10{,}000$.

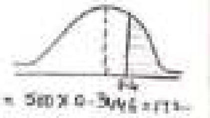
Let the variable x denote the Sales in the business.

when $x = 40000$, $z = \frac{x-\mu}{\sigma} = \frac{40000 - 36000}{10000} = 0.4$

when $x = 30000$, $z = \frac{x-\mu}{\sigma} = \frac{30000 - 36000}{10000} = -0.6$

i) $P(x > 40000) = P(z > 0.4)$ = Area (from 0 to ∞) − Area (from 0 to 0.4)

= 0.5 − 0.1554 = 0.3446



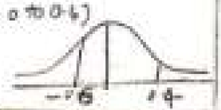
No. of business as the sales of which are Rs. 40,000 = 500 × 0.3446 = 172.

ii) $P(30000 < x < 40000) = P(-0.6 < z < 0.4)$

= A(from 0 to 0.6) + A(from 0 to 0.4)

= 0.1554 + 0.2257

= 0.3811



∴ The required percentage of business = 38.11%

⑥ If X is normally distributed with mean 2 and variance 0.1, then find $P(|X-2| \geq 0.01)$?

Sol: Here $\mu = 2$ & $\sigma = 0.1$

when $X = 1.99$, $z = \frac{1.99-2}{0.1} = -0.1$

$X = 2.01$, $z = \frac{2.01-2}{0.1} = 0.1$

∴ $P(|X-2| < 0.01) = P(-0.1 < z < 0.1)$

= Area from (0 to 0.1) + Area from (0 to 0.1)

$= 2(0.0398) = 0.0796$

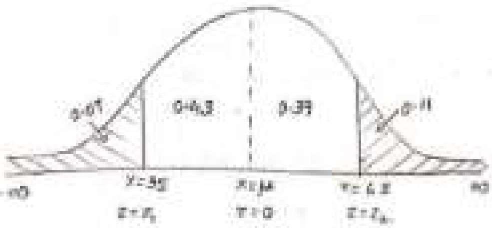
$P(|X-2| \geq 0.01) = 1 - P(|X-2| < 0.01)$

$= 1 - 0.0796$

$= 0.9204$.

⑦ In a normal distribution, 7% of the items are under 35 and 89% are under 63. Determine the mean and variance of the distribution.
(or) Find the mean and S.D of a normal distribution in which 7% of items are under 35 and 89% are under 63.

Sol:



when $x = 35$, $z = \frac{x-\mu}{\sigma} = \frac{35-\mu}{\sigma} = -z_1$ —①

when $x = 63$, $z = \frac{63-\mu}{\sigma} = z_2$ (say) —②

From the figure, we have

$P(0 < z < z_1) = 0.43 \Rightarrow z_1 = 1.48$

$P(0 < z < z_2) = 0.39 \Rightarrow z_2 = 1.23$ (from tables)

From ①, we have $\frac{35-\mu}{\sigma} = -1.48$ —③

From ②, we have $\frac{63-\mu}{\sigma} = 1.23$ —④

④ − ③ gives

$\frac{28}{\sigma} = 2.71 \Rightarrow \sigma = \frac{28}{2.71} = 10.332$

from ③, $35 - \mu = -1.48\sigma = -1.48(10.332) = -15.3$

∴ $\mu = 35 + 15.3 = 50.3$ and variance $= \sigma^2 = 106.75$

⑧ In a normal distribution 31% of the items are under 45 and 8% are over 64. Find the mean and variance of the distribution.

[$\mu = 50$ & $\sigma = 10$]

⑨ The marks obtained in mathematics by 1000 students is normally distributed with mean 78% and standard deviation 11%. Determine
i) How many students got marks above 90%
ii) what was the highest mark obtained by the lowest 10% of the students
iii) within what limits did the middle of 90% of the students lie.

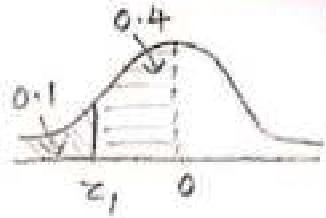
Sol: Given mean, $\mu = 78\% = 0.78$ and S.D, $\sigma = 11\% = 0.11$

i) when $x = 0.9$, $z = \frac{x-\mu}{\sigma} = \frac{0.9-0.78}{0.11} = 1.09 = z_1$ (say)

Hence the no. of students with marks above 90% is 0.1379×1000

∵ Area = 0.3621 & $0.5 - 0.3621 = 0.1379$ $= 137.9 \simeq 138$

ii) The 0.1 area to the left of z corresponds to the lowest 10% of the students. (for 40%)



From figure,

$$0.4 = 0.5 - 0.1 = 0.5 - \text{Area from 0 to } z_1$$

$$\therefore z_1 = -1.28 \text{ (from tables)}$$

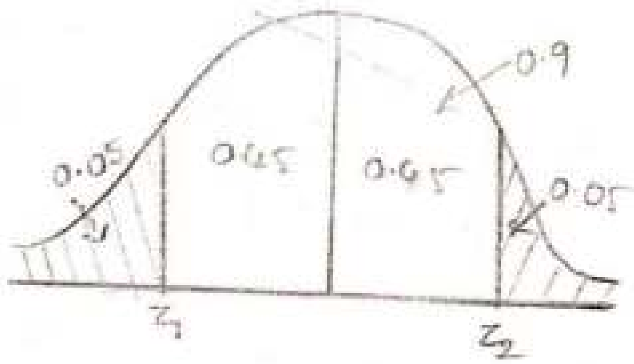
Thus $-1.28 = \frac{x-\mu}{\sigma} = \frac{x-0.78}{0.11} \Rightarrow x = 0.78 - 1.28(0.11) = 0.6392$

Hence the highest mark obtained by the lowest 10% of students

$$= 0.6392 \times 1000 \simeq 64\%.$$

iii) Middle 90% correspond to 0.9 area, leaving 0.05 area on both sides. Then the corresponding z's are ± 1.64

[Since if the area from 0 to z is 0.45, then z = 1.64]



$$\therefore -1.64 = z_1 = \frac{x_1-\mu}{\sigma} = \frac{x_1-0.78}{0.11}$$

$$\Rightarrow x_1 = 0.78 - 1.64(0.11)$$

$$= 0.5996 \text{ or } 59.96\%.$$

and $1.64 = z_2 = \frac{x_2-\mu}{\sigma} = \frac{x_2-0.78}{0.11} \Rightarrow x_2 = 0.9604$ (or) 96.04%.

Thus the middle 90% have marks b/n 60 to 96.

Estimation

In statistics, estimation refers to the process by which one makes inferences about a population, based on information obtained from a sample.

Estimate: An estimate is a statement made to find an unknown population parameter.

EX: Sample mean, $\bar{x}$ is an estimator of population mean, μ because Sample mean is a method of determining the population mean.

i.e., If $\bar{x} = 30$, 30 is called as the estimate of the population mean and $\bar{x}$ is called the estimator.

A parameter can have one or two or many estimators.

Types of Estimation:

Basically there are two kinds of estimates to determine the statistic of the population parameters namely,

(a) point Estimation (b) Interval Estimation.

If an estimate of the population parameter is given by a single value, then the estimate is called a "point Estimation" of the parameter. But if an estimate of a population parameter is given by two different values between which the parameter may be considered to lie, then the estimate is called an "interval estimation" of the parameter.

EX: If the height of a student is measured as 162 cm, then the measurement gives a point estimation.

But if the height is given as (163 ± 3.5) cms, then the height lies between 159.5 cms and 166.5 cms and the measurement gives an interval estimation.

The Sample mean $\bar{x}$ is a point estimate of population mean μ.

Sample variance s^2 is a point estimate of population variance σ^2.

Definition: A point estimate of a parameter 'θ' is a single numerical value, which is computed from a given sample and serves as an approximation of the unknown exact value of the parameter.

Definition: A point estimator is a static for estimating the population parameter θ and will be denoted by $\hat{\theta}$ (read as theta hat)

Properties of Estimation: An estimator is not expected to estimate the population parameter without error. An estimator should be close to the true value of unknown parameter.

(or)

A good estimator is one which is close to the true value of the parameters as possible. Prof R.A. Fisher suggested that a good estimator should satisfy the following properties.

1. Unbiasedness 2. Consistency 3. Efficiency 4. Sufficiency.

Unbiased estimator: A statistic or point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if $E(\hat{\theta}) = \theta$.

In other words, if E (statistic) = parameter, then statistic is said to be an unbiased estimator of the parameter.

Consistency: If an estimator, say θ, approaches the parameter 'θ' closer and closer as the sample size 'n' increases, 'θ' is consistent estimator of 'θ'. [An estimator $\hat{\theta}_n$ of a parameter θ is consistent if it cgs to θ, as $n \to \infty$]

Most efficient estimator: If we consider all possible unbiased estimators of some parameter θ, the one with the smallest variance is called the most efficient estimator of 'θ'.

(or)

A statistic $\hat{\theta}_1$ is said to be a more efficient unbiased estimator of the parameter θ than the statistic $\hat{\theta}_2$ if

a) $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ.

b) $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.

Sufficient Estimator:

An estimator is said to be sufficient for a parameter, if it Contains all the Information in the Sample regarding the parameter.

Interval Estimation:

point estimates rarely Coincide with quantities they are intended to estimate. So instead of point estimation where the quantity to be estimated is replaced by a single value a better way of estimation is interval estimation, which determines an interval in which the parameter lies.

Interval estimates are intervals for which one can be $(1-\alpha)$ 100% Confident that the parameter lies under investigation lies in this interval. Such an interval is known as confidence interval for the parameter with (having) $1-\alpha$ & $(1+\alpha)$ 100% degree of confidence. The two end points of the confidence interval are known as Confidence limits or fiducial limits or critical values or confidence coefficients. Confidence level denoted by α is the percentage of confidence.

when $\alpha = 0.05$, we have a 95% confidence interval, and when $\alpha = 0.01$, we have 99% confidence interval.

1. S.T Sample mean $\bar{x}$ is an unbiased estimator of population mean 'μ'.

[Since $E(\bar{x}) = \mu$].

proof :- Let $x_1, x_2, \ldots, x_n$ be a random sample drawn from a given population with mean μ and variance σ^2. Then.

$$E(\bar{x}) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n} E(x_1 + x_2 + \cdots + x_n)$$

$$= \frac{1}{n}\left[E(x_1) + E(x_2) + \cdots + E(x_n)\right]$$

$$= \frac{1}{n}\left[\mu + \mu + \cdots + \mu\right] \qquad [\because E(x_i) = \mu]$$

$$= \frac{1}{n} \cdot n\mu = \mu$$

Hence the Sample mean $\bar{x}$ is an unbiased estimator of the population mean μ.

$\therefore \bar{x}$ is an unbiased estimator of 'μ'.

(2) prove that for a random Sample of Size n, $x_1, x_2, \ldots, x_n$ taken from an infinite population $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is not unbiased estimator of the parameter σ^2 but $\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is unbiased.

proof :- Let 'μ' be the population mean. Then

$$E(x_i) = \mu \text{ and } var(x_i) = E[(x_i-\mu)^2] = \sigma^2 \text{ for } i=1,2,\ldots,n$$

If 's' be the Sample standard deviation, then $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ — (1)

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

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\bar{x}\cdot\frac{\Sigma x_i}{n} + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2$$

$$= \frac{\Sigma x_i^2}{n} - 2\bar{x}\cdot\bar{x} + \frac{n\bar{x}^2}{n} \qquad \left[\because \frac{\Sigma x_i}{n} = \bar{x} \text{ and } \sum_{i=1}^{n}1 = n\right]$$

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

$$= \Sigma\frac{x_i^2}{n} - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 - (\bar{x}-\mu)^2$$

$$\left[\because (s^2) = \frac{\Sigma(x_i-\mu)^2}{n} - (\bar{x}-\mu)^2 = \frac{\Sigma x_i^2}{n} - 2\mu\,\Sigma\frac{x_i}{n} + n\cdot\frac{\mu^2}{n} - (\bar{x}^2+\mu^2-2\bar{x}\mu) = \Sigma\frac{x_i^2}{n} - \bar{x}^2\right]$$

$$\therefore E(s^2) = E\left[\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} - (\bar{x}-\mu)^2\right] = E\left[\frac{\Sigma(x_i-\mu)^2}{n}\right] - E(\bar{x}-\mu)^2$$

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

$$= \frac{\Sigma\, var(x_i)}{n} - Var(\bar{x})$$

$$= \frac{\Sigma\sigma^2}{n} - \frac{\sigma^2}{n} \qquad \left(\because \bar{x} = \frac{\Sigma x_i}{n}\right)$$

$$= \frac{n\sigma^2}{n} - \frac{\sigma^2}{n}$$

$$= \sigma^2\left[1 - \frac{1}{n}\right]$$

$$\therefore E(s^2) = \left[\frac{n-1}{n}\right] \cdot \sigma^2$$

Thus $E(s^2) \neq \sigma^2$

Hence, s^2 is a biased estimator of σ^2

Let us consider $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \cdot n s^2$ [by ①]

Thus $E(S^2) = E\left[\frac{n}{n-1} s^2\right] = \frac{n}{n-1} E(s^2) = \frac{n}{n-1} \times \frac{n-1}{n} \cdot \sigma^2 = \sigma^2$

Hence $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ is an unbiased estimator of σ^2.

Bayesian Estimation:

Suppose prior distribution has a mean μ_0 and a standard deviation σ. In Bayesian estimation prior feelings about the possible values of μ combined with direct sample evidence. This leads to a posterior distribution of μ, which under fairly general conditions, can be approximated by a normal distribution with

$$\text{Mean of posterior distribution} = \mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

where 'n' is sample size, s = S.D of sample, (use $s = \sigma$)

$\bar{x}$ = Sample mean

σ^2 = variance of population

[Note: when σ^2 is unknown, is replaced by s^2 (sample variance) provided $n \geq 30$ (large sample)]

$$\text{Standard deviation of posterior distribution } \sigma_1 = \sqrt{\frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}}.$$

Bayesian interval for μ:

$(1-\alpha)100\%$ Bayesian interval for μ is given by $\mu_1 - z_{\alpha/2}\cdot\sigma_1 < \mu < \mu_1 + z_{\alpha/2}\cdot\sigma_1$

population and Sample:

population is the Set or collection or totality of objects. Thus mainly population consists of Sets of numbers, measurements or observations which are of interest.

Size: Size of the population N is the number of objects or observations is the population.

population is said to be finite (or) infinite depending on the size N being finite or infinite.

Since it is is impracticable or uneconomical or time consuming to study the entire population, a finite Subset of the population is known as Sample is studied. Size of the Sample is denoted by n. Sampling is the process of drawing Samples from a given population."

Example:

(i) population of India, population of A.P. Estate (Sample).

(ii) Engineering colleges recognised by AICTE, Engineering colleges affiliated to JNTU (Sample).

Large Sampling:

If $n \geq 30$, the Sampling is said to be large Sampling.

Small Sampling:

If $n < 30$, the Sampling is said to be Small Sampling or exact Sampling.

Confidence interval for μ: (For large Samples)

A $(1-\alpha)100\%$ Confidence interval for μ is given by

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

So the maximum error of estimate E with $(1-\alpha)$ probability is given by $E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Confidence interval when 'S' is unknown (Small Samples):

When α, E, σ are known, the Sample size 'n' is given by

$$n = \left[\frac{Z_{\alpha/2} \cdot \sigma}{E}\right]^2$$

when σ is known (or $n<30$ Small Sample)

In this case, σ is replaced by s, the standard deviation of Sample to determine E.

Thus the maximum error estimate $E = t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$ with $(1-\alpha)$ probability

Here t-distribution is with $(n-1)$ degrees of freedom.

Confidence interval for μ (σ unknown):

A $(1-\alpha)\,100\%$ confidence interval μ (σ unknown) is

$$\bar{x} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

Here n = Sample size, S = Standard deviation of Sample.

Problems:

① What is the size of the smallest sample required to estimate an unknown proportion to within a maximum error of 0.06 with atleast 95% confidence.

Sol: We are given, the maximum error, $E = 0.06$

Confidence limit $= 95\%$

i.e., $(1-\alpha)100\% = 95\%$

$\Rightarrow (1-\alpha)100 = 95$

$\Rightarrow 1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

$\therefore Z_{\alpha/2} = 1.96$

Here p is not given. So we take $p = \frac{1}{2}$. Thus $Q = \frac{1}{2}$.

Hence $n = \left[\frac{Z_{\alpha/2}}{E}\right]^2 (pQ)$

When p is unknown, Sample size $n = \frac{1}{4}\left[\frac{Z_{\alpha/2}}{E}\right]^2$

$\therefore n = \frac{1}{4}\left[\frac{1.96}{0.06}\right]^2 = 266.78 \cong 267.$

② If we can assert with 95% that the maximum error is 0.05 and $p = 0.2$, find the size of the sample.

Sol: Given $p = 0.2$, $E = 0.05$

We have $Q = 1-p = 0.8$ and $Z_{\alpha/2} = 1.96$ (for 95%)

We know that maximum error, $E = Z_{\alpha/2} \cdot \sqrt{\frac{pQ}{n}}$

$$\Rightarrow 0.05 = 1.96\sqrt{\frac{(0.2)(0.8)}{n}}$$

$$\Rightarrow \text{Sample size}, n = \frac{0.2 \times 0.8 \times (1.96)^2}{(0.05)^2}$$

$$\boxed{n = 246}$$

③ Assuming that $\sigma = 20.0$, how large a random sample be taken to assert with probability 0.95 that the mean will not differ from the true mean by more than 3.0 points?

Sol: $n \cong 171$

(4) It is desired to estimate the mean number of houses of continuous use until a certain computer will first require repairs. If it can be assume that $\sigma = 48$ hours, how large a sample be needed so that one will be able to assert with 90% confidence that the sample mean is off by at most 10 hours.

Sol It is given that,

Maximum error, $E = 10$ hours

$\sigma = 48$ hours and $Z_{\alpha/2} = 1.645$ (for 90%)

$$\therefore n = \left[\frac{Z_{\alpha/2} \cdot \sigma}{E}\right]^2 = \left[\frac{1.645 \times 48}{10}\right]^2 = 62.3 = 62.$$

Hence sample size $= 62$.

(5) It is desired to estimate the mean time of continuous use until an answering machine will first require service. If it can be assumed that $\sigma = 60$ days, how large a sample is needed so that one will be able to assert with 90% confidence that the sample mean is off by at most 10 days.

Sol $n = 97$.

(6) The mean and standard deviation of a population are 11,795 and 14,054 respectively. What can one assert that 95% confidence about the maximum error if $\bar{x} = 11,795$ and $n = 50$.

Sol Mean of population $\mu = 11795$

Standard deviation of the population $\sigma = 14054$

n = Sample size $= 50$

Maximum error $= Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$Z_{\alpha/2}$ for 95% confidence $= 1.96$

Maximum error $E = 1.96 \cdot \frac{14054}{\sqrt{50}}$

$= 3896.$

(7) What is the maximum error one can expect to make with probability 0.9, when using the mean of a random sample of size $n = 64$ to estimate the mean of a population with $\sigma^2 = 2.56$?

Sol: Maximum error $= E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$Z_{\alpha/2} = 1.645$ (90% confidence)

σ = standard deviation $= \sqrt{2.56} = 1.6$

$n = 64$

$\therefore E = 1.645 \times \frac{1.6}{8} = 0.329$

$\therefore$ Maximum error $= 0.329$.

(8) An industrial engineer intends to use the mean of a random sample of size $n = 150$ to estimate the average mechanical aptitude of assembly line works in a large industry. If on the basis of experience, the engineer can assume that $\sigma = 6.2$ for such data, what can he assert with probability 0.99 about the maximum size of his error?

Sol: Maximum Error $E = 1.3$.

(9) The mean and the standard deviation of a population are 11,795 and 14054 respectively. If $n = 50$, find 95% confidence interval for the mean.

Sol: Here Mean of population, $\mu = 11795$

Standard deviation of population, $\sigma = 14054$

$\bar{x} = 11795$

n = Sample size $= 50$

maximum error $= Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{14054}{\sqrt{50}} = 3899$

$\therefore$ Confidence interval $= \left(\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$

$= (11795 - 3899, 11795 + 3899)$

$= (7896, 15694)$

(10) A random sample of size 81 was taken whose variance is 20.25 and mean is 32, construct 98% confidence interval.

Sol: Given $\bar{x} = 32$, $n = 81$, $\sigma^2 = 20.25 \Rightarrow \sigma = 4.5$

and $Z_{\alpha/2} = 2.33$ (for 98%)

W.K.T 98% confidence interval is $\left[\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$

Now, $Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 2.33 \cdot \frac{4.5}{\sqrt{81}} = 1.165$

$\therefore$ Confidence interval $= (32 - 1.165, 32 + 1.165)$

$= (30.835, 33.165)$

(11) A random sample of 400 items is found to have mean 82 and S.D of 18. Find the maximum error of estimation at 95% confidence interval. Find the confidence limits for the mean if $\bar{x} = 82$.

(12) If S.D is of the sample $S = 4$, $n = 25$ Construct 95% confidence interval if $\bar{x} = 35$.

Sol: Sample size $n = 25$

$t_{\alpha/2} = 2.06$

S = Standard deviation $= 4$

$t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} = 2.06 \cdot \frac{4}{5} = 1.648$

Confident interval $\bar{x} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$

$= (35 - 1.648, 35 + 1.648)$

$= (33.352, 36.648)$

(13) A sample of 11 rats from a central population had an average blood viscosity of 3.92 with a S.D of 0.61. Estimate the 95% confidence limits for the mean blood viscosity of the population.

Sol: $n = 11$ (Small sample)

$S = S.D = 0.61$, $\bar{x} = 3.92$

$t_{\alpha/2}$ for 95% and $v = n - 1 = 11 - 1 = 10$ is 2.23

Confidence limits are $\left(\bar{x} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}\right) = \left(3.92 - 2.23 \cdot \frac{0.61}{\sqrt{11}}, 3.92 + 2.23 \frac{0.61}{\sqrt{11}}\right)$

$= (3.51$ and 4.3[illegible]$)$

(14) A Sample of Size 10 was taken from a population S.D. of Sample is 0.03. Find the maximum error with 99% Confidence.

Sol: S = Standard deviation = 0.03

Sample Size $n = 10$

$t_{\alpha/2}$ for $v = n-1 = 10-1 = 9$, $99\% = 3.25$

$\therefore E = t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} = 3.25 \times \frac{0.03}{3} = 0.0325.$

(15) The mean mark in mathematics in common entrace that will vary from year to year. If this variation of the mean mark is expressed Subjectively by a normal distribution with mean $\mu_0 = 72$ and variance $\sigma_0^2 = 5.76$.

(i) what probability Can we assign to the actual mean mark being somewhere between 71.8 and 73.4 for the next years test?

(ii) Constructed a 95% Bayesian interval for μ if the test is conducted for a random sample of 100 Students from the next incoming class yields a mean mark of 70 with S.D. of 8.

(iii) what posterior probability should we assign to the event of part (i)

Sol: a) Here μ_0 = mean = 72 and $\sigma_0 = S.D = \sqrt{5.76} = 2.4$, $n = 100$.

we have $z = \frac{x - 72}{2.4}$

when $x = 71.8$, $z_1 = \frac{71.8 - 72}{2.4} = -0.0833$

when $x = 73.4$, $z_2 = \frac{73.4 - 72}{2.4} = 0.5833.$

Let x be the mean mark obtained in the common entrace test. Then

Prior probability $= P(71.8 < x < 73.4) = P(-0.083 < z < 0.5833)$

$= 0.0319 + 0.2190 = 0.2509$

b) Here $n = 100$, $\bar{x} = 70$, $\sigma_0 = 2.4$, $\mu_0 = 72$, $\sigma = 8$

$\therefore$ posterior mean, $\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} = \frac{100(70)(2.4)^2 + 72(8)^2}{100(2.4)^2 + (8)^2} = 70.2$

posterior standard deviation, $\sigma_1 = \sqrt{\frac{\sigma^2 \cdot \sigma_0^2}{n\sigma_0^2 + \sigma^2}} = \sqrt{\frac{(8)^2(2.4)^2}{100(2.4)^2 + (8)^2}} = 0.76$

$\therefore$ 95% Bayesian interval for 'μ' is given by $\mu_1 \pm z_{\alpha/2} \cdot \sigma_1 = 70.2 \pm 1.96(0.759)$

$= 70.2 \pm 1.4874$

i.e., $68.71 < \mu < 71.69$.

(C) Here $\mu_1 = 70.2$, $\sigma_1 = 0.76$

when $x = 71.8$, $z_1 = \dfrac{71.8 - 70.2}{0.76} = 2.105$

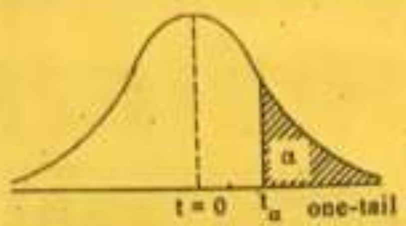
$x = 73.4$, $z_2 = \dfrac{73.4 - 70.2}{0.76} = 4.2105 \simeq 4.21$

$\therefore$ posterior probability $= P(71.8 < x < 73.4) = P(2.105 < z < 4.21)$

$= 0.5 - 0.4821$

$= 0.0179$

t_α - Critical Values of the t-Distribution with ν Degrees of Freedom Table – 4



	α								
ν	0.40	0.30	0.20	0.15	0.10	0.05	0.025	0.01	0.005
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Note : The above table gives the values of t for one-tail test (either left-tail or right-tail test). If we have to find the value of t for a two-tail test at a level, we take the value of $\alpha/2$ for α. For example, the value of t at 5% level with 9 d.f. is $t_{0.025} = 2.262$ and the value of t at 1% level with 11 d.f. is $t_{0.005} = 3.106$

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

PROCEDURE FOR TESTING OF HYPOTHESIS :

[JNTU (H) Nov. 2009 (Set No.2,4), (K) Nov. 2011, Dec. 2013 (Set No. 3)]

Various steps involved in testing of Hypothesis are given below : Infact, the same steps are followed for conducting all tests of significance.

Step 1 : ***Null Hypothesis :*** Define or set up a Null Hypothesis H_0 taking into consideration the nature of the problem and data involved.

Step 2 : ***Alternative Hypothesis :*** Set up the Alternative Hypothesis H_1 so that we could decide whether we should use one-tailed or two-tailed test.

Step 3 : ***Level of Significance :*** Select the appropriate level of significance(α) depending on the reliability of the estimates and permissible risk. That is a suitable α is selected in advance if it is not given in the problem. (Usually we choose 5% level of significance)

Step 4 : ***Test Statistic*:** Compute the test statistic $Z = \dfrac{t - E(t)}{\text{S.E of } t}$ under the null hypothesis.

Here t is a sample statistic and S. E. is the standard error of t.

Step 5 : ***Conclusion :*** We compare the computed value of the test statistic Z with the critical value Z_α at given level of significance (α).

If $|Z| < Z_\alpha$, (that is, if the absolute value of the calculated value of Z is less than the critical value Z_α) we conclude that it is not significant. We accept the null hypothesis.

If $|Z| > Z_\alpha$ then the difference is significant and hence the null hypothesis is rejected at the level of significance α.

Clearly,

For two-tailed test :

If $|Z| < 1.96$ accept H_0 at 5% level of significance.

If $|Z| > 1.96$ reject H_0 at 5% level of significance.

If $|Z| < 2.58$, accept H_0 at 1% level of significance.

If $|Z| > 2.58$ reject H_0 at 1% level of significance.

For single-tailed (right or left) test :

If $|Z| < 1.645$, accept H_0 at 5% level of significance.

If $|Z| > 1.645$, reject H_0 at 5% level of significance.

If $|Z| < 2.33$ accept H_0 at 1% level of significance.

If $|Z| > 2.33$ reject H_0 at 1% level of significance.

Test of Hypothesis

Test of Hypothesis: We need to decide whether to accept or reject a Stam statement about the parameter. This statement is called hypothesis, testing and the decision-making procedure about the hypothesis is called hypothesis testing.

The main object of the Sampling theory is the Study of the Tests of Hypothesis or Tests of Significance.

For example, on the basis of Sample data,

i) a drug chemist is to decide whether a new drug is really effective in curing a disease.

ii) a Statistician has to decide whether a given Coin is biased, etc.

Such decisions are called Statistical decisions (or Simply decisions)

Statistical Hypothesis:

A statistical hypothesis is an assumption about a population parameter. Such an assumption (or guesses or statement) is called a statistical Hypothesis which may or may not be true.

Ex: 1. The majority of men in the City are Software engineers.

2. The teaching methods in both the Schools are effective.

procedure for Testing A hypothesis:

Steps:

Step 1: Statement (or assumption) of hypothesis

There are two types of hypothesis:

(i) Null Hypothesis (ii) Alternative Hypothesis.

Null Hypothesis: For applying the test of Significance, we first Set up a hypothesis a definite Statement about the population parameter. Such a hypothesis of no-difference is Called Null-Hypothesis. It is in the form $H_0 : \mu = \mu_0$.

μ_0 is the value which is assumed or claimed for the population characteristic.

Alternative Hypothesis:

Any hypothesis which contradicts the Null Hypothesis is Called an Alternative Hypothesis, usually denoted by "H_1".

The two hypothesis H_0 and H_1 are such that if one is true, the other is false and vice versa.

Alternative Hypothesis would be

(i) $H_1 : \mu \neq \mu_0$ (i.e, either $\mu > \mu_0$ or $\mu < \mu_0$) (or)

(ii) $H_1 : \mu > \mu_0$ (Right-tailed)

(iii) $H_1 : \mu < \mu_0$ (left-tailed)

Step : 2 Specification of the level of Significance

The level of Significance denoted by 'α' is the confidence with we ~~ma~~ rejects or accepts the null hypothesis H_0.

The level of Significance is always fixed in advance before collecting the Sample information.

The levels of significance is ~~always~~ usually employed in testing of hypothesis are 5% and 1%.

Step : 3 Identification of the Test Statistic

There Are Several tests of Significance, viz., z, t, F etc. First we have to Select the right test depending on the nature of the information given in the problem. Then we construct the test criterion and select the appropriate probability distribution

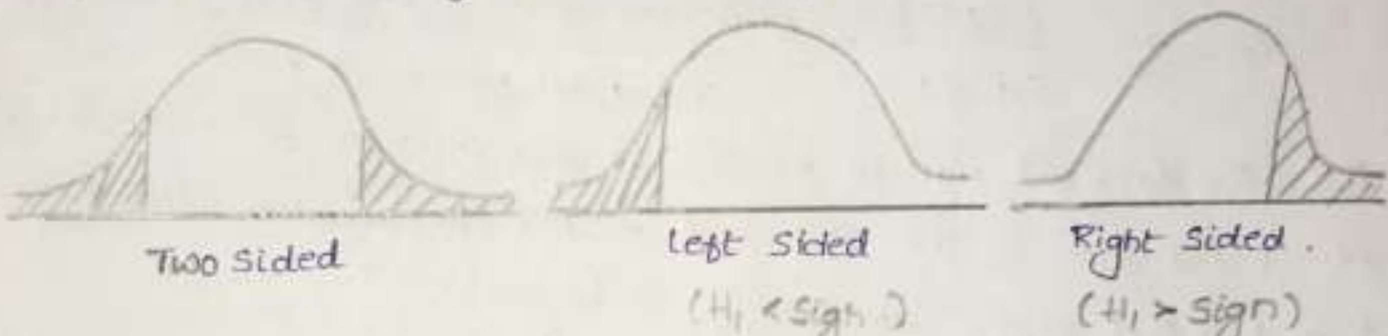
Step: 4 Critical region:

The critical region is formed based on following factors.

a) Distribution of the statistic i.e, whether the statistic follows the normal, t, χ^2 (or) F distribution.

b) form of Alternative hypothesis:

If the form has ≠ sign, the critical region is divided equally in the left and right tails, sides of the distribution.



Step: 5 Making Decision

If the computed value < critical value, we accept H_0, otherwise we reject H_0.

Errors of Sampling:

The test is based on sample observations, the decision of acceptance or rejection of the null hypothesis is always subjected to some error. i.e, some amount of risk.

Types of errors in test of hypothesis:

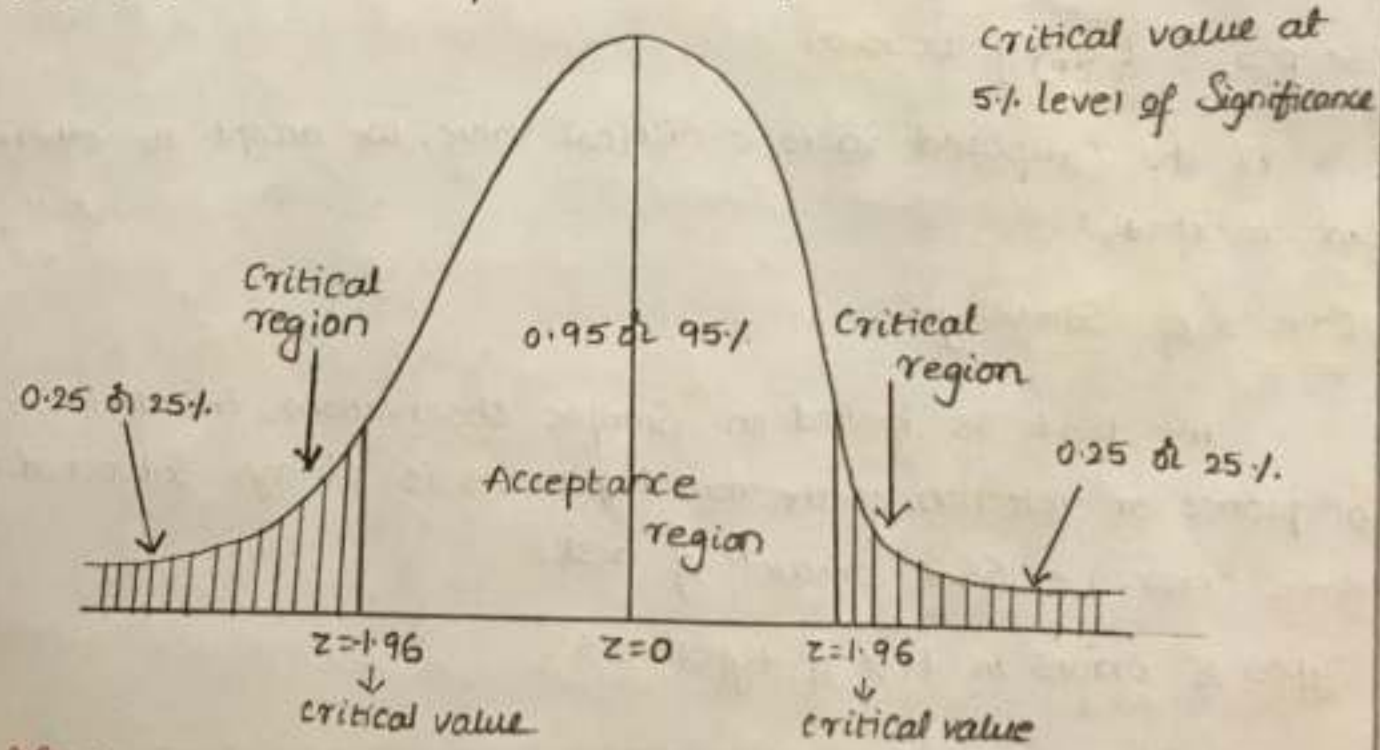
	Accept H_0	reject H_0
H_0 is true	Correct decision	Type I error
H_0 is false	Type II error	Correct decision

The statistical testing of hypothesis aims at limiting the Type I error to a preassigned value (Say: 1% or 5%) and to minimize the type II error. The only way to reduce both types of errors is to increase the Sample size, if possible.

Critical Region: (or region of rejection or the region of Significance)

under a given hypothesis let the Sampling distribution of a Statistic 't' is approximatey a normal distribution with mean $E(t)$ and $S.D. = \sigma_t$ = Standard Error of 't'.

Then $z = \dfrac{t - E(t)}{S.E \text{ of } t} = \dfrac{\text{observed value} - \text{Expected value}}{\text{Standard Error of } t}$ is called the Standardized normal variate or test Statistic or z-score and its distribution is the Standard normal distribution with mean 0 and S.D. 1 when Sample Size is large.



Critical values (or) Significant values:

The value of the test Statistic, which Seperates the critical region (or rejection region) and the acceptance region is called the critical value or Significant value. The value is dependent on

i) the level of Significance used, and

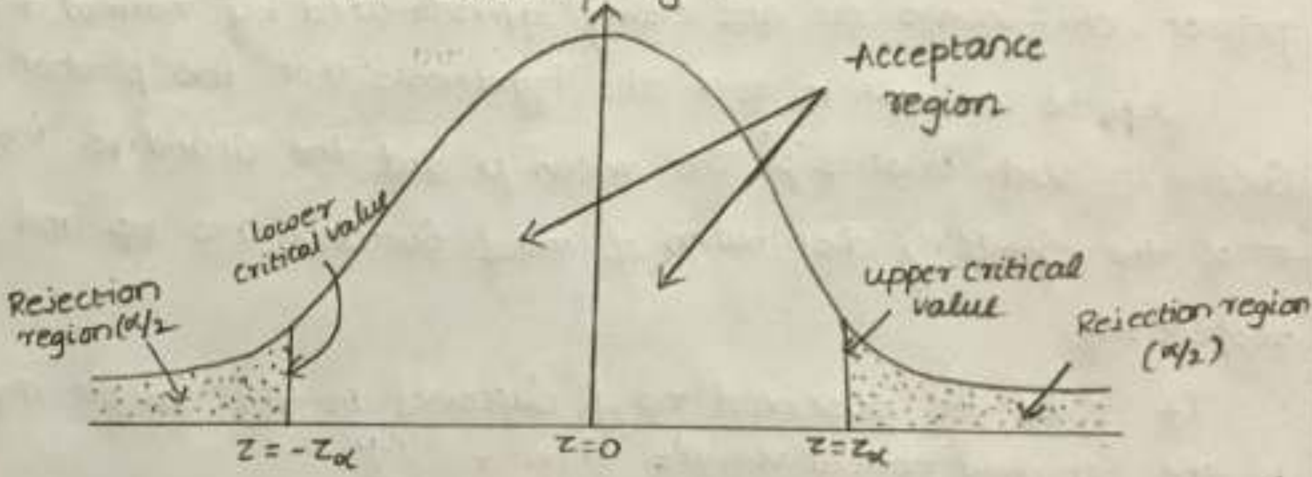
ii) the alternative hypothesis, whether it is one-tailed or two-tailed.

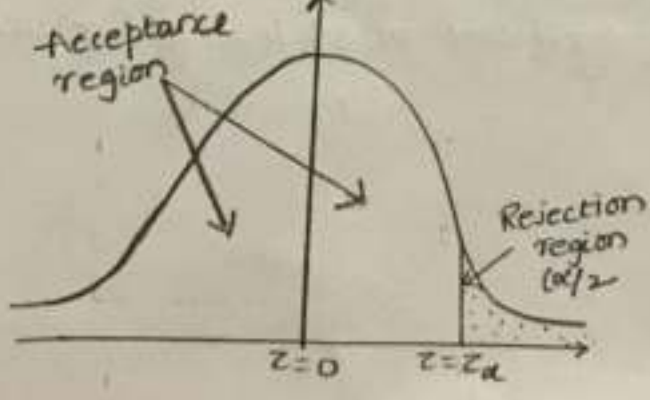
One Tailed Test (O.T.T) and Two Tailed Test (T.T.T):

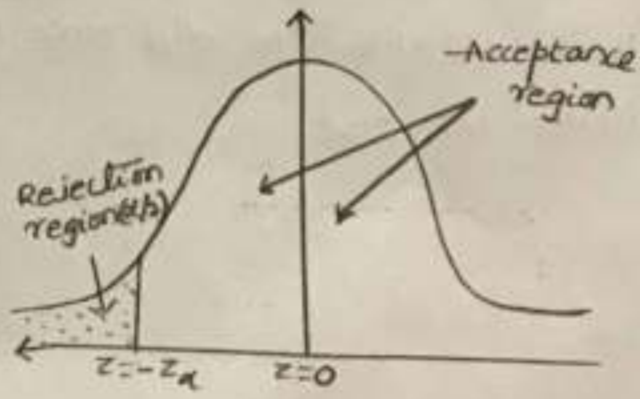
If we have to test whether the population mean μ has a specified value μ_0, then the null hypothesis is $H_0: \mu = \mu_0$ and the Alternative Hypothesis may be (i) $H_1: \mu \neq \mu_0$ (i.e., $\mu > \mu_0$ or $\mu < \mu_0$) (ii) $H_1: \mu > \mu_0$ (or) (iii) $H_1: \mu < \mu_0$.

The Alternative Hypothesis in (i) is known as a two-tailed (i.e., both right and left tail) alternatives and the alternative hypothesis in (ii) and (iii) are known as right-tailed and left-tailed alternatives respectively.

For level of Significance 'α'







Critical values (z_α) of z :

Critical values of z			
Level of Significance α	1%	5%	10%
Critical values for two-tailed test	$\lvert z_\alpha \rvert = 2.58$	$\lvert z_\alpha \rvert = 1.96$	$\lvert z_\alpha \rvert = 1.645$
Critical values for Right-tailed test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Critical values for left-tailed test	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$.

Test of Significance for large Samples:

If the Sample Size $n \geq 30$, then we consider such samples as large samples. If n is large, the distributions, such as Binomial, poisson, chi-Square etc. are closely approximated by normal distributions

Suppose we wish to test the hypothesis that the probability of Success in such trial is 'p'. The mean μ and the Standard deviation σ of the Sampling distribution of no.of Successes are np and $\sqrt{npq}$ respectively.

If x be the observed no.of Successes in the Sample and z is the Standard normal variate then $z = \frac{x-\mu}{\sigma}$.

Thus we have the following test of Significance:

i) If $|z| < 1.96$, the difference between the observed and expected no.of Successes is not Significant.

ii) If $|z| > 1.96$, the difference is Significant at 5% level of Significance.

iii) If $|z| > 2.58$, " " " " " 1% " " ".

problems:

① A coin was tossed 960 times and returned heads 183 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance.

Sol: Here $n = 960$, p = probability of getting head $= 1/2$.

$$\therefore q = 1-p = 1-1/2 = 1/2 \; ; \; \mu = np = 960(1/2) = 480.$$

$$\sigma = \sqrt{npq} = \sqrt{(np)q} = \sqrt{480 \times \tfrac{1}{2}} = \sqrt{240} = 15.49.$$

x = no. of successes $= 183$

1. Null Hypothesis H_0: The coin is unbiased.
2. Alternative Hypothesis H_1: The coin is biased.
3. Level of Significance 'α': $\alpha = 0.05$
4. The test statistic is $z = \frac{x-\mu}{\sigma} = \frac{183-480}{(15.49)} = -19.17$

$$\therefore |z| = 19.17$$

As $|z| > 1.96$, the null hypothesis H_0 has to be rejected at 5% level of significance and we conclude that the coin is biased.

② A die is tossed 256 times and it turns up with an even digit 150 times. Is the die biased?

Sol: Here $n = 256$, p = The probability of getting an even digit (2 or 4 or 6)

$$= \frac{3}{6} = \frac{1}{2}$$

$$q = 1-p, \; p = \tfrac{1}{2} \text{ so } q = 1-\tfrac{1}{2} = \tfrac{1}{2} \; ; \; \mu = np = 256\left(\tfrac{1}{2}\right) = 128$$

$$\sigma = \sqrt{npq} = \sqrt{np(q)} = \sqrt{128 \times \tfrac{1}{2}} = \sqrt{64} = 8$$

x = no. of successes $= 150$

1. Null Hypothesis H_0: The die is unbiased
2. Alternative Hypothesis H_1: The die is biased
3. Level of significance: $\alpha = 0.05$
4. The test statistic is $z = \frac{x-\mu}{\sigma}$ i.e., $z = \frac{150-128}{8} = 2.75$

As $|z| > 1.96$, the H_0 has to be rejected at 5% L.O.S and the die is biased.

Under large Sample tests, we will see four important tests to test the Significance:

1. Testing of Significance for Single proportion.
2. Testing of Significance for difference of proportion.
3. Testing of Significance for Single mean.
4. Testing of Significance for differnce of means.

Test of Significance of a Single mean - Large Samples:

Let a random Sample of Size $n (n \geq 30)$ has the Sample mean $\bar{x}$, and 'μ' be the population mean. Also the population mean μ has a Specified value μ_0.

Working Rule:

1. The Null Hypothesis is H_0: $\bar{x} = \mu$ i.e., " there is no significance difference between the Sample mean and population mean" or the Sample has been drawn from the parent population.
2. The Alternative Hypothesis is (i) $H_1: \bar{x} \neq \mu$ $(\mu \neq \mu_0)$ or (ii) $H_1: \bar{x} > \mu$ $(\mu > \mu_0)$ (iii) $H_1: \bar{x} < \mu$ $(\mu < \mu_0)$

Since n is large, the Sampling distribution of $\bar{x}$ is approximately normal.

3. Level of Significance: Set the level of Significance 'α'.
4. The test statistic: we have the following two Cases.

Case I: When the standard deviation σ of population is known

In this Case, Standard Error mean, $S.E.(\bar{x}) = \frac{\sigma}{\sqrt{n}}$, when n = Sample size, σ = S.d of population

$\therefore$ The test statistic is $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$, where μ is the Sample mean.

Case II: when the standard deviation σ of population is not known

In this Case, we take S, S.d of Sample to compute the S.E of mean (2)

$$\therefore S.E.(\bar{x}) = \frac{S}{\sqrt{n}}$$

Hence the test Statistic is $z = \frac{\bar{x}-\mu}{S.E.(\bar{x})} = \frac{\bar{x}-\mu}{S/\sqrt{n}}$

5. Find the critical value z_α of z at the Level of Significance α from the normal table.

6. Decision :

a) If $|z| < z_\alpha$, we accept the Null Hypothesis H_0.

b) If $|z| > z_\alpha$, we reject the Null Hypothesis H_0.

Problems

(1) A Sample of 64 Students have a mean weight of 70 kgs. Can this be regarded as as a Sample from a population with mean weight 56 kgs and Standard deviation 25 kgs.

Sol Given $\bar{x}$ = mean of the Sample = 70 kgs

μ = mean of the population = 56 kgs

σ = S.D of population = 25 kgs

and n = Sample size = 64.

1. Null Hypothesis H_0 : $\mu = 56$

2. Alternative Hypothesis H_1 : $\mu \neq 56$

3. Level of Significance: $\alpha = 0.05$ (assumption)

4. The test Statistic is $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{70-56}{25/\sqrt{64}} = 4.48$

5. The null Hypothesis H_0 is rejected, Since $|z| > 1.96$

(2) A Sample of 900 members has a mean of 3.4 cms and S.D 2.61 cms. Is this Sample has been taken from a large population of mean 3.25 cms and S.D 2.61 cms. If the population is normal and its mean is unknown find the 95% fiducial limits of true mean.

Sol Given $n = 900$, $\mu = 3.25$

$\bar{x} = 3.4$ cm, $\sigma = 2.61$ and $s = 2.61$

1. Null Hypothesis H_0 : $\mu = 3.25$
2. Alternative Hypothesis H_1 : $\mu \neq 3.25$
3. The test statistic is $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{3.4-3.25}{2.61/\sqrt{900}} = 1.724$

i.e., $z = 1.724 < 1.96$

$\therefore$ we accept the null hypothesis H_0.

i.e, The Sample has been drawn from the population with mean $\mu = 3.25$

95% Confidence limits are given by

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 3.4 \pm 1.96 \times \frac{2.61}{\sqrt{900}} = 3.4 \pm 0.1705$$

i.e., 3.57 and 3.2295.

3. An ambulance Service claims that it takes on the average less than 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the claim at 0.05 level of Significance.

Sol Given $n = 36$, $\bar{x} = 11$, $\mu = 10$ and $\sigma = \sqrt{16} = 4$.

1. Null Hypothesis H_0 : $\mu = 10$.
2. Alternative Hypothesis H_1 : $\mu < 10$.
3. Level of Significance : $\alpha = 0.05$
4. The test statistic is, $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{11-10}{4/\sqrt{36}} = 1.5$

Tabulated value of z at 5% L.O.S is 1.645

Hence Calculated z < tabulated z

$\therefore$ we accept the null Hypothesis H_0.

4. In a random Sample of 60 workers, the average time taken by them to get to work is 33.8 minutes, with a standard deviation of 6.1 min. Can we reject the null hypothesis $\mu = 32.6$ minutes in favour of alternative null hypothesis $\mu > 32.6$ at $\alpha = 0.025$ level of Significance.

Sol: Given $n = 60$, $\bar{x} = 33.8$, $\mu = 32.6$ and $\sigma = 6.1$

1. Null Hypothesis H_0: $\mu = 32.6$
2. Alternative Hypothesis H_1: $\mu > 32.6$ (Right-tailed test)
3. Level of Significance $\alpha = 0.025$
4. The test statistic is $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{33.8-32.6}{6.1/\sqrt{60}} = \frac{1.2}{0.7875} = 1.5238$

Tabulated value of z at 0.025 level of Significance (L.O.S) is 2.58

Hence Calculated $z <$ tabulated z.

$\therefore$ The null hypothesis H_0 is accepted.

5. It is Claimed that a random Sample of 49 tyres has a mean life of 15200 Km. This Sample was drawn from a population whose mean is 15150 kms and a standard deviation of 1200 km. Test the significance at 0.05 level.

Sol: Given $n = 49$, $\bar{x} = 15200$, $\mu = 15150$ and $\sigma = 1200$.

1. Null Hypothesis H_0: $\mu = 15150$
2. Alternative Hypothesis H_1: $\mu \neq 15150$.
3. Level of significance: $\alpha = 0.05$.
4. Critical region: Accept the null hypothesis if $-1.96 < z < 1.96$.
5. The test statistic is, $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{15200 - 15150}{1200/\sqrt{49}} = 0.2917$.

Since $|z| < 1.96$ therefore, we accept the null hypothesis.

6. The mean life time of a Sample of 100 light tubes produced by a company is found to be 1560 hrs with a population S.D of 90 hrs. Test the hypothesis for $\alpha = 0.05$ that the mean life time of the tubes produced by the Company is 1580 hrs.

7. An oceanographer wants to check whether the depth of the ocean in a certain region is 57.4 fathoms, as had previously been recorded. What can he conclude at the 0.05 level of Significance, if readings taken at 40 random locations in the given region yielded a mean of 59.1 fathoms with a standard deviation of 5.2 fathoms.

TEST FOR EQUALITY OF TWO MEANS-LARGE SAMPLES (TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS OF TWO LARGE SAMPLES)

Let $\overline{x_1}$ and $\overline{x_2}$ be the sample means of two independent large random samples sizes n_1 and n_2 drawn from two populations having means μ_1 and μ_2 and standard deviations σ_1 and σ_2. To test whether the two population means are equal.

Step:1 Null Hypthesis $H_0 : \mu_1 = \mu_2$

Step:2 Alternative Hypthesis $H_1 : \mu_1 \neq \mu_2$

Step:3 Set the level of significance α.

Step:4 The test statistic $z = \dfrac{\left(\overline{x_1} - \overline{x_2}\right) - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$; where $\delta = \mu_1 - \mu_2$

If $\delta = o,$ the two populations have the same means

If $\delta \neq o,$ the two populations are different.

Step:5 Rejection rule for $H_0 : \mu_1 = \mu_2$.

i. If $|z| > 1.96$, reject H_0 at 5% level of significance.

ii. If $|z| > 2.58$, reject H_0 at 1% level of significance.

iii. If $|z| > 1.645$, reject H_0 at 10% level of significance.

Note : If the two samples are drawn from population with unknown Standard deviations σ_1^2 and σ_2^2, then σ_1^2 and σ_2^2 can be replaced by sample by variances s_1^2 and s_2^2 provided both the samples n_1 and n_2 are large.

In this case, the test statistic is $z = \dfrac{\left(\overline{x_1} - \overline{x_2}\right) - \delta}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$.

TEST FOR EQUALITY OF TWO MEANS-LARGE SAMPLES (TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS OF TWO LARGE SAMPLES)

Solved Problems

1. The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches.

Solution: Let μ_1 and μ_2 be the means of the two populations.

Given $n_1 = 1000$ and $n_2 = 2000$ and $\overline{x_1} = 67.5$ and $\overline{x_2} = 68.0$ inches.

Population Standard deviation, $\sigma = 2.5$ inches.

Step:1 Null Hypothesis H_0**:** The samples have been drawn from the same population of S.D 2.5 inches. *i.e.,* $\mu_1 = \mu_2$ and $\sigma = 2.5$ inches.

Step:2 Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

Step:3 Set the level of significance: α =0.05.

Step:4 The test statistic $z = \frac{\left(\overline{x_1} - \overline{x_2}\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{67.5 - 68}{\sqrt{(2.5)^2\left(\frac{1}{1000} + \frac{1}{2000}\right)}}$

$$z = \frac{-0.5}{0.0968} = -5.16$$

$$\therefore |z| = 5.16$$

Tabulated value of Z at 5% level of significance is 1.96.

Hence calculated Z > tabulated Z.

Step:5 Hence the null hypothesis H_0 is Rejected at 5% level of significance and we conclude that the samples are not drawn from the sample population of S.D. 2.5 inches.

2. A researcher wants to know the intelligence of students in a school. He selected two groups of students. In the first group there 150 students having mean IQ (intelligence quotient) of 75 with a S.D. of 15 in the second group there are 250 students having mean IQ of 70 with S.D. of 20.

Solution: Let μ_1 and μ_2 be the means of the two populations.

Given $n_1 = 150$ and $n_2 = 250$ and $\overline{x_1} = 75$ and $\overline{x_2} = 70$.

Population Standard deviation, $\sigma_1 = 15$ & $\sigma_2 = 20$

Step: 1 Null Hypothesis H_0**:** The groups have been came from the same population.

$$i.e., \ \mu_1 = \mu_2 \text{ and } \sigma_1 = 15 \ \& \ \sigma_2 = 20.$$

Step: 2 Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

Step: 3 Set the level of significance: α =0.05.

Step: 4 The test statistic $z = \frac{(\overline{x_1} - \overline{x_2})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{75-70}{\sqrt{\left(\frac{225}{150} + \frac{400}{250}\right)}} = \frac{5\sqrt{5}}{\sqrt{17}} = 2.7116$

$$\therefore |z| = 2.7116$$

Tabulated value of Z at 5% level of significance is 1.96.

Hence calculated Z > tabulated Z.

Step: 5 Hence the null hypothesis H_0 is Rejected at 5% level of significance and we conclude that the groups have not been from the same population.

3. Samples of students were drawn from two universities and from their weights in kilograms, mean and S.D. are calculated and shown below. Make a large sample test to the significance of the difference between the means.

	Mean	S.D.	Size of the sample
University A	55	10	400
University B	57	15	100

Solution: Let $\overline{x_1}$ and $\overline{x_2}$ be the means of the two samples.

Given $n_1 = 400$ and $n_2 = 100$ and $\overline{x_1} = 55$ and $\overline{x_2} = 57$.

Standard deviation, $s_1 = 10$ & $s_2 = 15$

Step:1 Null Hypothesis H_0**:** $\overline{x_1} = \overline{x_2}$. *i.e* there is no difference.

Step:2 Alternative Hypothesis $H_1 : \overline{x_1} \neq \overline{x_2}$

Step:3 Set the level of significance: α =0.05.

Step:4 The test statistic $z = \dfrac{(\overline{x_1} - \overline{x_2})}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{55-57}{\sqrt{\left(\dfrac{100}{400} + \dfrac{225}{100}\right)}} = \dfrac{-2}{\sqrt{2.5}} = -1.26$

$\therefore |z| = 1.26$

Tabulated value of Z at 5% level of significance is 1.96.

Hence calculated Z < tabulated Z.

Step:5 Hence the null hypothesis H_0 is accepted at 5% level of significance and we conclude that there is no significant difference between the means.

4. The average marks scored by 32 boys is 72 with S.D. of 8. While that for 36 girls is70 with a S.D. of 6. Does this indicate that the boys perform better than girls at level of significance 0.05?

Solution: Let μ_1 and μ_2 be the means of the two populations.

Given $n_1 = 32$ and $n_2 = 36$ and $\overline{x_1} = 72$ and $\overline{x_2} = 70$.

Population Standard deviation, $\sigma_1 = 8$ & $\sigma_2 = 6$

Step:1 Null Hypothesis H_0**:** $\mu_1 = \mu_2$.

Step:2 Alternative Hypothesis $H_1 : \mu_1 > \mu_2$.

Step:3 Set the level of significance: α =0.05.

Step:4 The test statistic $z = \dfrac{(\overline{x_1} - \overline{x_2})}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{72-70}{\sqrt{\left(\dfrac{64}{32} + \dfrac{36}{36}\right)}} = \dfrac{2}{\sqrt{3}} = 1.1547$

$\therefore |z| = 1.1547$

Tabulated value of Z at 5% level of significance is 1.96.

Hence calculated Z < tabulated Z.

Step:5 Hence the null hypothesis H_0 is accepted at 5% level of significance and we conclude that the performance of boys and girls is the same.

5. A sample of the height of 6400 Englishmen has a mean of 67.85 inches and a S.D. of 2.56 inches while a simple sample of heights of 1600 Austrians has a mean of 68.55 inches and

S.D. of 2.52 inches. Do the data indicate the Austrians are on the average taller than the Englishmen? (Use α as 0.01).

Solution: Let $\overline{x_1}$ and $\overline{x_2}$ be the means of the two samples.

Given $n_1 = 6400$ and $n_2 = 1600$ and $\overline{x_1} = 67.85$ and $\overline{x_2} = 68.55$.

Standard deviation, $s_1 = 2.56$ & $s_2 = 2.52$

Step:1 Null Hypothesis H_0: $\overline{x_1} = \overline{x_2}$. *i.e* there is no difference.

Step:2 Alternative Hypothesis $H_1 : \overline{x_1} < \overline{x_2}$

Step:3 Set the level of significance: α =0.01.

Step:4 The test statistic $z = \dfrac{\left(\overline{x_1} - \overline{x_2}\right)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{67.85 - 68.55}{\sqrt{\left(\dfrac{6.5536}{6400} + \dfrac{6.35}{1600}\right)}} = \dfrac{-0.7}{\sqrt{0.005}} = -9.9$

$$\therefore |z| = 9.9$$

Tabulated value of Z at 5% level of significance is 1.96.

Hence calculated Z >tabulated Z.

Step:5 Hence the null hypothesis H_0 is rejected at 5% level of significance and we conclude that the Austrians are taller than Englishmen.

6. In a certain factory there are two independent processes for manufacturing the same item. The average weights in a sample of 700 items produced from one process is found to be 250 gms with a S.D. of 30 gms while the corresponding gures in a sample of 300 items from the other process are 300 and 40. Is there significant difference between the mean at 1% level of significance.

Solution: Let μ_1 and μ_2 be the means of the two populations.

Given $n_1 = 700$ and $n_2 = 300$ and $\overline{x_1} = 250$ and $\overline{x_2} = 300$.

Population Standard deviation, $\sigma_1 = 30$ & $\sigma_2 = 40$

Step:1 Null Hypothesis H_0: $\mu_1 = \mu_2$ and $\sigma_1 = 30$ & $\sigma_2 = 40$.

Step:2 Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

Step:3 Set the level of significance: α =0.01.

Step:4 The test statistic $z = \frac{\left(\overline{x_1} - \overline{x_2}\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{250-300}{\sqrt{\left(\frac{900}{700} + \frac{1600}{300}\right)}} = -19.43$

$$\therefore |z| = 19.43$$

Tabulated value of Z at 1% level of significance is 2.58.

Hence calculated $Z >$ tabulated Z.

Step:5 Hence the null hypothesis H_0 is Rejected at 5% level of significance and we conclude that there is a significant difference between the means.

7. The mean yield of wheat from a district A was 210 pounds with S.D. 10 pounds per acre from a sample of 100 plots. In another district the mean yield was 220 pounds with S.D. 12 pounds from a sample of 150 plots. Assuming that the S.D. of yield in the entire state was 11 pounds, test whether there is any significant difference between the mean yield of crops in the two districts.

8. The research investigator is interested in studying whether there is a significant difference in the salaries of MBA grades in two metropolitan cities. A random sample of size 100 from Mumbai yields on average income of Rs. 20,150. Another random sample of 60 from Chennai results in an average income of Rs. 20,250. If the variances of both the populations are given as $\sigma_1^2 = Rs.\ 40,000$ and $\sigma_2^2 = Rs.\ 32,400$ respectively.

How to find critical value for Two-Tailed and One-Tailed Test

NOTE: 1. Find a critical value for a 95% confidence level (Two-Tailed Test).

Procedure: Confidence limit = 95%

$$i.e, (1-\alpha)100\% = 95\%$$

$$\Rightarrow (1-\alpha)100 = 95$$

$$\Rightarrow (1-\alpha) = \frac{95}{100} = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \boxed{\frac{\alpha}{2} = 0.025}$$

$$1 - 0.025 = 0.975, z-value\ \ at\ \ 0.975\ \ is\, 1.96$$

$$\boxed{\therefore z_{\alpha/2} = 1.96}$$

2. Find a critical value for a 95% confidence level (One-Tailed Test).

Procedure: Confidence limit = 95%

$$i.e, (1-\alpha)100\% = 95\%$$

$$\Rightarrow (1-\alpha)100 = 95$$

$$\Rightarrow (1-\alpha) = \frac{95}{100} = 0.95 \Rightarrow \boxed{\alpha = 0.05}$$

$$1 - 0.05 = 0.95, z-value\ \ at\ \ 0.95\ \ is\, 1.645$$

$$\boxed{\therefore z_{\alpha} = 1.645}$$

Sample Proportion:

$$p = \frac{Count\ of\ successes\ in\ a\ sample}{sample\ size\ 'n'} = \frac{x}{n};$$

$$q = 1 - p = sample\ of\ proportion\ of\ failures\ in\ a\ sample\ of\ size\ n.$$

Large Proportion:

$$P = \frac{Count\ of\ successes\ in\ a\ population}{Population\ size\ 'N'} = \frac{X}{N};$$

$$Q = 1 - P = Population\ of\ proportion\ of\ failures\ in\ a\ population\ of\ size\ N$$

TEST FOR EQUALITY OF SIGNIFICANCE FOR SINGLE PROPORTION-LARGE SAMPLES

Suppose a large random sample of size n has a sample proportion p of members possessing a certain attribute (*i.e,* proportion of successes). To test the hypothesis that the proportion P in the population has a specified value P_0.

Step:1 Let us set the Null Hypthesis be $H_0 : P = P_0$ (P_0 is a particular value of P).

Step:2 The Alternative Hypothesis is $H_1 : P \neq P_0$ $i.e, H_1 : P > P_0, H_1 : P < P_0$

Step:3 Set the level of significance α.

Step:4 The test statistic $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$; where p is the sample proportion is approximately normally distributed.

Step:5 The critical Rejection for z depending on the nature of H_1 and level of significance α is given in the following table.

Rejection Rule for $H_0 : P = P_0$			
Level of significance	1%	5%	10%
Critical region for $P \neq P_0$ (Two-tailed test)	$\lvert z\rvert > 2.58$	$\lvert z\rvert > 1.96$	$\lvert z\rvert > 1.645$
Critical region for $P > P_0$ (Right-tailed test)	$z > 2.33$	$z > 1.645$	$z > 1.28$
Critical region for $P < P_0$ (Left-tailed test)	$z < -2.33$	$z < -1.645$	$z < -1.28$

Note: 1. without any reference to the level of significance, we may reject the Null Hypothesis H_0 when $|z| > 3$.

2. (i) Limits for population proportion P are given by $p \pm 3\sqrt{\dfrac{pq}{n}}$ where $q = 1 - p$.

(ii) Confidence interval for proportion P are given by $p - z_{\alpha/2}\sqrt{\dfrac{pq}{n}} < z < p + z_{\alpha/2}\sqrt{\dfrac{pq}{n}}$

where $Q = 1 - P$.

Solved Problems

1. A manufacturer claimed that at least 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance.

Solution: Given sample size, n=200.

Number of pieces confirming to specifications=x=200-18=182

$\therefore$ p =Proportion of pieces confirming to specifications= $\frac{x}{n} = \frac{182}{200} = 0.91$

Let P = Population proportion= $\frac{95}{100} = 0.95$; Q=1-P=1-0.95=0.05

Step:1 Null Hypothesis H_0**:** The proportion of pieces confirming to specifications.

$i.e., P = 95\%$.

Step:2 Alternative Hypothesis $H_1 : P < 0.95$. (Left – tail test)

Step:3 Set the level of significance: α =0.05.

Step:4 The test statistic $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91-0.95}{\sqrt{\left(\frac{0.95\times0.05}{200}\right)}} = \frac{-0.04}{0.0154} = -2.59$

$$\therefore |z| = 2.59$$

Since alternative hypothesis is left tailed, the tabulated value of z at 5% level of significance is 1.645.

Hence calculated value of Z > tabulated Z.

Step:5 Hence the null hypothesis H_0 is Rejected at 5% level of significance and we conclude that the manufacture's claim is rejected.

2. In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance?

Solution: Given sample size, n=1000.

Let p =Sample proportion of rice eaters = $\frac{540}{1000} = 0.54$

Let P = Population proportion of rice eaters = $\frac{1}{2} = 0.5$.

$$Q = 1 - P = 1 - 0.5 = 0.5.$$

Step:1 Null Hypothesis H_0**:** Both rice and wheat are equally popular in the state.

$i.e., P = 0.5$.

Step:2 Alternative Hypothesis $H_1 : P \neq 0.5$. (Two– tailed test)

Step:3 Set the level of significance: α =0.01.

Step:4 The test statistic $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54-0.5}{\sqrt{\left(\frac{0.5\times0.5}{1000}\right)}} = 2.532$

$\therefore |z| = 2.532$

Since alternative hypothesis is left tailed, the tabulated value of z at 1% level of significance is 2.58.

Hence calculated Z < tabulated Z.

Step:5 Hence the null hypothesis H_0 is accepted at 1% level of significance and conclude that both rice and wheat are equally popular in the state.

3. A random sample of 500 pineapples was taken from a large consignment and 65 were found bad. Find the percentage of bad pineapples in the consignment.

Solution: Given sample size, n=500.

Let p = Proportion of bad pineapples in the sample = $\frac{65}{500} = 0.13$

$q = 1 - p = 0.87$.

We know that the limits for population proportion P are given by

$$p \pm 3\sqrt{\frac{pq}{n}} = 0.13 \pm 3\sqrt{\frac{0.13\times0.87}{500}} = 0.13 \pm 0.045 = (0.085, 0.175)$$

$\therefore$ The percentage of bad pineapples in the consignment lies between 8.5 and 17.5.

4. A manufacturer claims that only 4% of his products are defective. Test the hypothesis at random sample of 500 were taken among which 100 were defective. Test the hypothesis at 0.05 level.

Solution: Given sample size, n=500, x=100

Let $p = \frac{x}{n} = \frac{100}{500} = 0.2$

Let P =4%=0.04 & $Q = 1 - P = 0.96$

Step:1 Null Hypothesis H_0**:** $P = 0.04$.

Step:2 Alternative Hypothesis $H_1 : P \neq 0.04$. (Two– tailed test)

Step:3 Set the level of significance: α =0.05.

Step:4 The test statistic $z = \dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.2-0.04}{\sqrt{\left(\dfrac{0.04\times 0.96}{500}\right)}} = -18.26$

$$\therefore |z| = 18.26$$

Since alternative hypothesis is right tailed, the tabulated value of z at 5% level of significance is 1.96.

Hence calculated Z > tabulated Z.

Step:5 Hence the null hypothesis H_0 is rejected at 5% level of significance.

5. In a random sample of 100 packages shipped by air freight 13 had some damage. Construct 95% confidence interval for the true proportion of damage package.

Solution: Given sample size, n=100, x=13

Let p =sample proportion of damage packages= $\dfrac{x}{n} = \dfrac{13}{100} = 0.13$

$$q = 1 - p = 1 - 0.13 = 0.87$$

$$p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.13\times 0.87}{100}} = 0.034 \ \ (\because P\, is\ not\ known, we\ take\ p\ \ for\ \ P)$$

$\therefore$ 95% confidence interval for the population proportion of P of damage package

$$p \pm 1.96\sqrt{\frac{pq}{n}} = 0.13 \pm 1.96(0.034) = 0.13 \pm 0.067 = (0.063, 0.197)$$

Hence the 95% confidene limits for the true proportion of damage packages is (0.063, 0.197)

6. In a hospital 480 females and 520 male babies were born in a week. Do these figures confirm the hypothesis that males and females are born in equal number?

Solution: Given sample size, n=Total number of births=480+520=1000, x=480

Let p =proportion of females born= $\dfrac{x}{n} = \dfrac{480}{1000} = 0.48$

Let P =0.5 & $Q = 1 - 0.5 = 0.5$

Step:1 Null Hypothesis H_0**:** The probability of equal proportion *i.e,* $P = \dfrac{1}{2} = 0.5$.

Step:2 Alternative Hypothesis $H_1 : P \neq 0.5$. (Two– tailed test)

Step:3 Set the level of significance: α =0.05.

Step:4 The test statistic $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.48-0.5}{\sqrt{\left(\frac{0.5\times0.5}{1000}\right)}} = -1.265$

$$\therefore |z| = 1.265$$

Since alternative hypothesis is two-tailed, the tabulated value of z at 5% level of significance is 1.96.
Hence calculated Z < tabulated Z.

Step:5 Hence the null hypothesis H_0 is accepted at 5% level of significance.

7. In a random sample of 125 cool drinkers, 68 said they prefer thumsup to pepsi. Test the null hypothesis P=0.5 against the alternative hypothesis P>0.5.

Solution: Given sample size, n=125, x=68 and $p = \frac{x}{n} = \frac{68}{125} = 0.544$

Let P =0.5 & $Q = 1-0.5 = 0.5$

Step:1 Null Hypothesis H_0: $P = 0.5$

Step:2 Alternative Hypothesis $H_1 : P > 0.5$. (Right– tailed test)

Step:3 Set the level of significance: α =0.05.

Step:4 The test statistic $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.544-0.5}{\sqrt{\left(\frac{0.5\times0.5}{125}\right)}} = 0.9839$

$$\therefore |z| = 0.9839$$

Hence calculated Z < tabulated Z.

Since alternative hypothesis is right tailed, the tabulated value of z at 5% level of significance is 1.645.
Step:5 Hence the null hypothesis H_0 is accepted at 5% level of significance.

8. In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?
9. A die was thrown 9000 times and of these 3220 yielded a 3 or 4. Is this consistent with the hypothesis that the die was unbiased?
10. In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. Construct a 99% confidence interval for the corresponding true percentage.

TEST FOR EQUALITY OF TWO PROPORTION (OR SIGNIFICANCE OF DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS-LARGE SAMPLES)

Let p_1 and p_2 be the sample proportions in two large random samples of sizes n_1 and n_2 drawn from two populations having proportions P_1 and P_2.

To test whether the two samples have been drawn from the sample population.

Step:1 Let us set the Null Hypothesis be $H_0 : P_1 = P_2$.

Step:2 The Alternative Hypothesis is $H_1 : P_1 \neq P_2$

Step:3 Set the level of significance α.

Step:4 There are two ways of computing a test statistic z.

(a) **When the population proportions P_1 and P_2 are known.**

In this case $Q_1 = 1 - P_1$ and $Q_2 = 1 - P_2$ and p_1, p_2 are sample proportions.

The test statistic $z = \dfrac{p_1 - p_2}{\sqrt{\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}}}$

(b) **When the population proportions P_1 and P_2 are not known but sample proportions p_1 and p_2 are known .**

In this case we have two methods to estimate P_1 and P_2.

***(i)* Method of Substitution:**

In this method, sample proportion p_1 and p_2 are substituted for P_1 and P_2.

The test statistic $z = \dfrac{p_1 - p_2}{\sqrt{\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}}}$

***(i)* Method of Pooling:**

In this method, the estimated value for the two population proportions is obtained by pooling the two sample proportions p_1 and p_2 into a single proportion p by the formula given below.

Sample proportion of two samples or estimated value of p is given by

$$p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}; q = 1 - p$$

The test statistic $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Step:5 Rejection rule for $H_0: P_1 = P_2$.

i. If $|z| > 1.96$, reject H_0 at 5% level of significance.

ii. If $|z| > 2.58$, reject H_0 at 1% level of significance.

iii. If $|z| > 1.645$, reject H_0 at 10% level of significance.

Solved Problems

1. A manufacturer of electronic equipment subjects samples of two completing brands of transistors to an accelerated performance test. If 45 of 180 transistors of the first kind and 34 of 120 transistors of the second kind fail the test, what can he conclude at the level of significance *a=0.05* about the difference between the corresponding sample proportions?

Solution: We have $n_1 = 180, n_2 = 120, x_1 = 45$ and $x_2 = 34$

$$p_1 = \frac{x_1}{n_1} = \frac{45}{180} = 0.25,\ p_2 = \frac{x_2}{n_2} = \frac{34}{120} = 0.283$$

$$\therefore p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{45 + 34}{180 + 120} = \frac{79}{300} = 0.263\ ;$$

$$q = 1 - p = 1 - 0.263 = 0.737$$

Step:1 Let us set the Null Hypothesis be $H_0: p_1 = p_2$.

Step:2 The Alternative Hypothesis is $H_1: p_1 \neq p_2$

Step:3 Set the level of significance $\alpha = 0.05$.

Step:4 Method of Pooling:

The test statistic $$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.25 - 0.283}{\sqrt{(0.263)(0.737)\left(\frac{1}{180} + \frac{1}{120}\right)}} = -0.647$$

$$\therefore |z| = 0.647$$

Since alternative hypothesis is two tailed, the tabulated value of z at 5% level of significance is 1.96.

Hence calculated $Z <$ tabulated Z.

Step:5 Hence the null hypothesis H_0 is accepted at 5% level of significance.

2. In two large population, there are 30% and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?

Solution: We have $n_1 = 1200, n_2 = 900, x_1 = 30$ and $x_2 = 25$

$P_1 = \text{Proportion of fair haired people in the first population} = 30\% = \frac{30}{100} = 0.3,$

$P_2 = \text{Proportion of fair haired people in the second population} = 25\% = \frac{25}{100} = 0.25$

$Q_1 = 1 - P_1 = 1 - 0.3 = 0.7;$

$Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$

Step:1 Let us set the Null Hypothesis be $H_0 : P_1 = P_2$.

Step:2 The Alternative Hypothesis is $H_1 : P_1 \neq P_2$

Step:3 Set the level of significance $\alpha = 0.05$.

Step:4 The test statistic $z = \frac{P_1 - P_2}{\sqrt{\left(\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}\right)}} = \frac{0.3 - 0.25}{\sqrt{\left(\frac{0.3 \times 0.7}{1200} + \frac{0.25 \times 0.75}{900}\right)}} = 2.56$

$$\therefore |z| = 2.56$$

Since alternative hypothesis is two tailed, the tabulated value of z at 5% level of significance is 1.96.

Hence calculated $Z >$ tabulated Z.

Step:5 Hence the null hypothesis H_0 is rejected at 5% level of significance.

3. In an investigation on the machine performance the following results are obtained.

	No. of units inspected	No. of defectives
Machine 1	375	17
Machine 2	450	22

Test whether there is any significant performance of two machines at $\alpha = 0.05$.

Solution: We have $n_1 = 375, n_2 = 450, x_1 = 17$ and $x_2 = 22$

$p_1 = \frac{x_1}{n_1} = \frac{17}{375} = 0.045, \; p_2 = \frac{x_2}{n_2} = \frac{22}{450} = 0.049$

$\therefore p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{17 + 22}{375 + 450} = \frac{39}{825} = 0.047$;

$q = 1 - p = 1 - 0.047 = 0.953$

Step:1 Let us set the Null Hypothesis be $H_0 : p_1 = p_2$.

Step:2 The Alternative Hypothesis is $H_1 : p_1 \neq p_2$

Step:3 Set the level of significance $\alpha = 0.05$.

Step:4 Method of Pooling:

The test statistic $z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.045-0.049}{\sqrt{(0.047)(0.953)\left(\frac{1}{375}+\frac{1}{450}\right)}} = 0.015$

Since alternative hypothesis is right tailed, the tabulated value of z at 5% level of significance is 1.96.

Hence calculated Z < tabulated Z.

Step:5 Hence the null hypothesis H_0 is accepted at 5% level of significance.

4. During a country wide investigation the incidence of tuberculosis was found to be 1%. In a college of 400 students 3 reported to be affected, where as in another college of 1200 students 10 were affected. Does this indicate any significant difference?

Solution: We have $n_1 = 400, n_2 = 1200, x_1 = 3$ and $x_2 = 10$

$$p_1 = \frac{x_1}{n_1} = \frac{3}{400} = 0.0075,\ p_2 = \frac{x_2}{n_2} = \frac{10}{1200} = 0.0083$$

Given

$$p = 1\% = 0.01\ ;\ q = 1 - p = 1 - 0.01 = 0.99$$

Step:1 Let us set the Null Hypothesis be $H_0 : P_1 = P_2$.

Step:2 The Alternative Hypothesis is $H_1 : P_1 \neq P_2$

Step:3 Set the level of significance $\alpha = 0.05$.

Step:4 Method of Pooling:

The test statistic $z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.0075-0.0083}{\sqrt{(0.01)(0.99)\left(\frac{1}{400}+\frac{1}{1200}\right)}} = -0.14$

$$\therefore |z| = 0.14 < 1.96$$

Since alternative hypothesis is two tailed, the tabulated value of z at 5% level of significance is 1.96.

Step:5 Hence the null hypothesis H_0 is accepted at 5% level of significance.

5. A sample poll of 300 voters from district A and 200 voters from district B showed that 56% and 48% respectively, were in favour of a given candidate. At a 0.05 level of significance, test the hypothesis that there is a difference in the districts.

Solution: We have $n_1 = 300, n_2 = 200, P_1 = 56\% = 0.56$ and $P_2 = 48\% = 0.48$

$$Q_1 = 1 - P_1 = 1 - 0.56 = 0.44;\ Q_2 = 1 - P_2 = 1 - 0.48 = 0.52$$

Step:1 Let us set the Null Hypothesis be $H_0 : P_1 = P_2$.

Step:2 The Alternative Hypothesis is $H_1 : P_1 \neq P_2$

Step:3 Set the level of significance $\alpha = 0.05$.

Step:4 The test statistic $z = \dfrac{P_1 - P_2}{\sqrt{\left(\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}\right)}} = \dfrac{0.56 - 0.48}{\sqrt{\left(\dfrac{0.56 \times 0.44}{300} + \dfrac{0.48 \times 0.52}{200}\right)}} = 1.78$

$$\therefore |z| = 1.78$$

Since alternative hypothesis is two tailed, the tabulated value of z at 5% level of significance is 1.96.

Hence calculated Z < tabulated Z.

Step:5 Hence the null hypothesis H_0 is accepted at 5% level of significance.

6. A random sample of 300 shoppers at a supermarket includes 204 who regularly use cents off coupons. In another sample of 500 shoppers at a supermarket includes 75 who regularly use cents off coupons. Construct confidence interval for the probability that any one shopper at the supermarket, selected at random, will regularly use cents off coupons.

Solution: Here $n_1 = 300$, $n_2 = 500$, $x_1 = 204$ and $x_2 = 75$

$$p_1 = \text{Proportion of shoppers who use cents off coupons in the first sample} = \frac{x_1}{n_1} = \frac{204}{300} = 0.68,$$

$$p_2 = \text{Proportion of shoppers who use cents off coupons in the first sample} = \frac{x_2}{n_2} = \frac{75}{500} = 0.15$$

$$q_1 = 1 - 0.68 = 0.32; q_2 = 1 - 0.15 = 0.85$$

The 98% confidence interval for the probability that any one shopper in sample selected at random is

$$\left[p_1 \pm z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1}}\right] = 0.68 \pm (2.33)\sqrt{\frac{0.68 \times 0.32}{300}} = 0.68 \pm 0.063 = (0.62, 0.74)$$

7. A study shows that 16 of 200 tractors produced on one assembly line required extensive adjustments before they could be shipped, while the same was true for 14 of 400 tractors produced on another assembly line. At the 0.01 level of significance, does this support the claim the second production line superior work?

8. On the basis of their total scores, 200 candidates of a civil service examination are divided into two groups, the upper 30% and the remaining 70%. Consider the first question of the examination. Among the first group, 40had the correct answer, where as among the second group, 80 had the correct answer. On the basis of these results, can one conclude that the first question is not good at discriminating ability of the type being examined here?

F Distribution Table at α = 0.01 or 1

m / n	1	2	3	4	5	6	7
1	4069.7377	4999.5	5403.2423	5624.5833	5764.2278	5921.8678	6401.986
2	98.6873	99	99.1663	99.2494	99.3118	102.4074	125.3596
3	34.1509	30.8165	29.4567	28.7099	28.2383	28.2052	29.8751
4	21.2076	18	16.6944	15.977	15.5222	15.2788	15.4921
5	16.2587	13.2739	12.06	11.3919	10.9671	10.6993	10.6456
6	13.7404	10.9248	9.7795	9.1483	8.746	8.4791	8.3496
7	12.2382	9.5466	8.4513	7.8466	7.4605	7.1987	7.0421
8	11.2476	8.6491	7.591	7.0061	6.6318	6.3753	6.2078
9	10.5481	8.0215	6.9919	6.4221	6.057	5.8049	5.6329
10	10.0289	7.5594	6.5523	5.9943	5.6363	5.3881	5.2142
11	9.6287	7.2057	6.2167	5.6683	5.316	5.071	4.8966
12	9.3112	6.9266	5.9525	5.412	5.0644	4.822	4.6476
13	9.0531	6.701	5.7394	5.2053	4.8616	4.6216	4.4475
14	8.8393	6.5149	5.5639	5.0354	4.695	4.4568	4.2833
15	8.6594	6.3589	5.417	4.8932	4.5556	4.3192	4.1461
16	8.5058	6.2262	5.2922	4.7726	4.4374	4.2025	4.0299
17	8.3732	6.1121	5.185	4.669	4.3359	4.1023	3.9303
18	8.2575	6.0129	5.0919	4.579	4.2479	4.0153	3.8439
19	8.1557	5.9259	5.0103	4.5003	4.1708	3.9392	3.7682
20	8.0654	5.8489	4.9382	4.4307	4.1027	3.8721	3.7015
21	7.9848	5.7804	4.874	4.3688	4.0421	3.8123	3.6422
22	7.9124	5.719	4.8166	4.3134	3.988	3.7589	3.5891
23	7.8469	5.6637	4.7649	4.2636	3.9392	3.7108	3.5414
24	7.7875	5.6136	4.7181	4.2184	3.8951	3.6673	3.4982
25	7.7333	5.568	4.6755	4.1774	3.855	3.6277	3.459

getcalc.com

t-Distribution (or). Student's t-Distribution :-

Definition :-

If $\{x_1, x_2, \ldots x_n\}$ be any random Sample size 'n' drawn from a normal population with mean μ and variance σ^2, then the test statistic 't' is defined by $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ where $\bar{x}$ = Sample mean and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})$ is an unbiased estimate of σ^2. The test statistic $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ is a random variable having the t-distribution with $v = n-1$ Degrees of freedom and with probability density function $f(t) = y_0 \left(1 + \frac{t^2}{v}\right)^{\frac{v+1}{2}}$ where $v = n-1$ and y_0 is a Constant got by $\int_{-\infty}^{\infty} f(t)\, dt = 1$.

This is Known as "Student's t-distribution" (or) t-distribution.

Degree of Freedom (d.f) :-

The number of independent variates which make up the statistic is Known as the degrees of freedom (d.f). and it is denoted by 'v'.

Student's t-Test (or) t-Test for Single Mean :-

Suppose we want to test

(a). if a random Sample x_i of size 'n' has been drawn from a normal population with a specified mean 'μ',

(b). if the Sample mean differs Significantly from the hypothetical value 'μ' of the population mean

In this case the statistic is given by

$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{(n-1)}$ where $\bar{x}, \mu, s, n$ have usual meanings.

let a random Sample of Size 'n' $(n<30)$ has a Sample mean $\bar{x}$. To test the hypothesis that the population mean 'μ' has a Specified value μ_0 when population S.D σ is not Known.

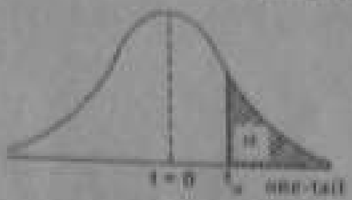
(1). Null Hypothesis be $H_0 : \mu = \mu_0$

(2). Alternative Hypothesis is $H_1 : \mu \neq \mu_0$

(3). The test statistic given by $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$

where s is the sample S.D. follows t-distribution with $\nu = (n-1)$ d.f.

t_α - Critical Values of the t-Distribution with v Degrees of Freedom Table – 4



v	α 0.40	0.30	0.20	0.15	0.10	0.05	0.025	0.01	0.005
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Note : The above table gives the values of t for one-tail test (either left-tail or right-tail test). If we have to find the value of t for a two-tail test at a level, we take the value of $\alpha/2$ for α. For example, the value of t at 5% level with 9 d.f. is $t_{0.025} = 2.262$ and the value of t at 1% level with 11 d.f. is $t_{0.005} = 3.106$

problem:-1:- A Sample of 26 bulbs gives a mean life of 990 hours with a S.D. of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours. Is the Sample not upto the standard.

Solⁿ:-

Here Sample Size, $n = 26 < 30$

∴ The Sample is Small Sample.

also given, Sample mean, $\bar{x} = 990$

Population mean, $\mu = 1000$ and

Standard Deviation, $s = 20$

Degrees of freedom $= n-1 = 26-1$
$= 25$

Here we know $\bar{x}$, μ, S.D. and n

∴ we use Students t-test

(1). Null Hypothesis H_0: The Sample is upto the standard. (i.e $\mu = \mu_0 = 1000$)

(2). Alternative Hypothesis H_1: $\mu < 1000$ (or) ($\mu \neq 1000 \neq \mu_0$)
(left-tail test)

(3). Level of Significance : ~~$\alpha = 95\%$ Assuming~~
$\alpha = 0.05$ (Assume)
(5%)

(4). The test statistic is $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$

$$= \frac{990 - 1000}{20/\sqrt{26-1}} = -2.5$$

$\therefore$ $|t| = |-2.5| = 2.5$

$\therefore$ calculated value of $t = 2.5$

Tabulated value of 't' at 5% L.O.S with 25 D.o.f for left tailed test is 1.708

$\therefore$ $|t| >$ tabulated value 't'

So we reject the null Hypothesis H_0 and conclude that the sample is not upto the standard.

Problem:-2:- The average breaking strength of the steel rods is specified to be 18.5 thousand pounds. To test this sample of 14 rods were tested. The mean and standard deviations obtained were 17.85 and 1.955 respectively. Is the result of experiment significant?

Sol:-

Solⁿ:- Here Sample Size, $n = 14$

Sample mean, $\bar{x} = 17.85$

S.D, $s = 1.955$

Population mean, $\mu = 18.5$

Degrees of freedom $= n-1 = 14-1 = 13$

(1). Null Hypothesis $H_0 : \mu = 18.5$ (or)

The result of the experiment is Significant

(2) Alternative Hypothesis $H_1 : \mu \neq 18.5$

(3) L.O.S $: \alpha = 0.05$ (5% L.O.S) (Assume)

(4). The test Statistic is, $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$

$= \frac{17.85 - 18.5}{1.955/\sqrt{13}}$

$t = \frac{0.65}{0.542} \sim -1.199$

$\therefore |t| = |-1.99| = 1.99$

Calculated value $t = 1.99$

Tabulated value t at 5% L.O.S for 13 D.f for two-tailed test = ~~1.771~~ 2.16

Since calculated t < tabulated t,

~~we accept~~ we ~~reject~~ accept the null Hypothesis H_0 at 5% L.O.S and ~~Calculo~~ Conclude that the result of the Experiment is ~~not~~ Significant.

Q. A random Sample of size 16 values from a normal population showed a mean of 53 and a Sum of Squares of deviations from the mean equals to 150. Can this Sample be regarded as taken from the population having 56 as mean?. Obtain 95% Confidence limits of the mean of population.

Sol:- Here Sample Size, $n = 16$

Sample mean, $\bar{x} = 53$

Now $\sum (x_i - \bar{x})^2 = 150$

$$\therefore S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{150}{15} = 10$$

$$\Rightarrow S = \sqrt{10}$$

Degree of freedom, $\nu = n-1 = 16-1 = 15$

Population mean, $\mu = 56$

(a). (1). Null Hypothesis H_0: The sample is taken from the population having 56 as mean

(or)

$\mu = 56$

(2) Alternative Hypothesis H_1: $\mu \neq 56$

(3) L.O.S $\therefore \alpha = 0.05$ (Assume) (5% L.O.S)

(4). The test statistic is $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

(Note that S.D is not given directly)

$$t = \frac{53-56}{\sqrt{10}/\sqrt{16}} = -3.79$$

$\therefore |t| = |-3.79| = 3.79$

The ~~-~~tabulated value of 't' at 5% L.O.S for 15 d.f for two tailed test is ~~1.753~~ 2.13

Since calculated t > tabulated t,

the null hypothesis H_0 is reject and

Conclude that the sample is ~~taken from~~

Can not regarded as taken from the population (b). The 95% C.I of the mean of population are given by

$$C.I \Rightarrow \left(\bar{x} \pm t_{0.05} \frac{s}{\sqrt{n}}\right)$$

$$\left(53 \pm 2.13 \times \frac{\sqrt{10}}{\sqrt{16}}\right)$$

$$(53 \pm 1.6827) \Rightarrow (53-1.6827, 53+1.6827)$$

~~(51.31, 54.68)~~

$\Rightarrow$ C.I is $(51.31, 54.68)$

(7). A random Sample of six steel beams has a mean Compressive strength of 58,392 Pounds per Square inch with a standard deviation of 648 P.S.i. use the information and the level of Significant $\alpha = 0.05$ to test the true average Compressive strength of the Steel from which the Sample came is 58000 P.S.i Assume normality.

Sol:

Here Sample mean, $\bar{x} = 58392$

Standard deviation, $\sigma = 648$

population mean, $\mu = 58000$

Sample size, $n = 6$, D.f = n-1
= 6-1
D.f = 5

Paired-Sample t-Test (or) Two Sample t-Test :-

If $(x_1, y_1), (x_2, y_2), ---, (x_n, y_n)$ be the pairs of sales data before and after the sales promotion in a business concern, we apply paired t-test to examine the significance of the difference of the two situations.

let $d_i = x_i - y_i$ (or) $y_i - x_i$ for $i = 1, 2 --- n$

(1). Null Hypothesis $H_0 : \mu_1 = \mu_2$ (i.e $\mu = 0$)

There is no significant diff b/w the mean in two situations

(2). Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

(3). Level of Significance : α

(4). The Test statistic is, $t = \dfrac{\bar{d} - \mu}{s/\sqrt{n}}$ ($\because \mu = 0$)

where $\bar{d} = \dfrac{\sum_{i=1}^{n} d_i}{n}$ and

$$s^2 = \frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}$$

(or)

$$s^2 = \frac{\left\{ \sum d^2 - \frac{(\sum d)^2}{n} \right\}}{n-1}$$

are the mean and variance of the differences $d_1, d_2, --- d_n$ respectively and μ is the mean of the population of differences.

The above statistic follows Student's t- test with (n-1) degrees of freedom.

Problem:-1. The Blood pressure of 5 women before and After intake of a certain drug are given below:

Before	110	120	125	132	125
After	120	118	125	136	121

Test whether there is Significant change in Blood pressure at 1% Level of Significance.

(1). The null Hypothesis be $H_0 : \mu_1 = \mu_2$ or $\mu = 0$

(There is no Significant difference in blood pressure before and After intake of a drug.)

(2). The Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

(3). Level of Significance : $\alpha = 0.01$ (1% L.O.S)

(4). The Test Statistic is, $t = \frac{\bar{d} - \mu}{s/\sqrt{n}}$ ($\because \mu = 0$)

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

where $\bar{d} = \frac{\sum d}{n}$, $d = y - x$

and $s^2 = \frac{\sum (d - \bar{d})^2}{n-1}$

Calculations for $\bar{d}$ and s

womens	B.P before intake of drug (x)	B.P After intake of drug (y)	$d = y - x$	$d_i - \bar{d}$ $d_i - 1.6$
1	110	120	10	8.4
2	120	118	−2	−3.6
3	125	125	0	−1.6
4	132	136	4	2.4
5	125	121	−4	−5.6
			$\sum d = 8$	$\sum d^2 = 140$ $\sum d_i - \bar{d} = 0$

$\therefore \bar{d} = \frac{\sum d}{n} = \frac{8}{5} = 1.6$

$s^2 = \frac{\sum (d - \bar{d})^2}{n-1} = \frac{64+16+4+36}{5-1}$ $70.56 + 12.96 + 2.56 + 5.76 + 31.36$

$= \frac{123.2}{4} = 30.8$

$s^2 = 30.8 \Rightarrow s = \sqrt{30.8} = 5.5498$

$t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{1.6}{\sqrt{30.8}/\sqrt{5}} = \frac{1.6\sqrt{5}}{\sqrt{30.8}} = 0.6447$

~~t = 0.82~~ $t = 0.6447$

Degrees of freedom is $n-1 = 5-1 = 4$

Thus $|t| =$ ~~0.6447 3.106~~ $0.6447 < 3.106$

Since the calculated value of 't' < the tabulated value of 't' with 4 d.f at 1% L.O.S

So, we accept the H_0 at 1% L.O.S

Conclude that there is no significant difference in blood pressure before and after intake of a drug.

Problem :- 2. Memory capacity of 10 students were tested before and after training. State whether the training was effective or not from the following score.

Before training	12	14	11	8	7	10	3	0	5	6
After training	15	16	10	7	5	12	10	2	3	8

Sol :-

(1). Null Hypothesis $H_0 : \mu_1 = \mu_2$

(2). Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

(3). Level of Significant : $\alpha = 0.05$ (5% L.O.S) (Assume)

(4). The test statistic is, $t = \frac{\bar{d}}{s/\sqrt{n}}$.

where $\bar{d} = \frac{\sum d}{n}$

$S^2 = \frac{\sum (d_i - \bar{d})}{n-1}$.

Memory Capacity before training (x)	Memory Capacity After training (y)	d=y-x	$d_i - \bar{d}$ $= d_i - 1.2$	$(d_i - \bar{d})$
12	15	3	1.8	3.24
14	16	2	0.8	0.64
11	10	-1	-2.2	4.84
8	7	-1	-2.2	4.84
7	5	-2	-3.2	10.24
10	12	2	0.8	0.64
3	10	7	5.8	33.64
0	2	2	0.8	0.64
5	3	-2	-3.2	10.24
6	8	2.	0.8	0.64
		$\sum d$. = 12		$\sum(d_i - \bar{d})^2$ = 69.6

$\bar{d} = \frac{\sum d}{n} = \frac{12}{10} = 1.2$

$S^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} = \frac{69.6}{10-1} = \frac{69.6}{9}$

$S^2 = 7.7333 \Rightarrow S = \sqrt{7.7333}$

$S = 2.7809$

Hence

$$t = \frac{\bar{d}}{S/\sqrt{n}} = \frac{1.2}{2.7809/\sqrt{10}} = \frac{1.2 \times \sqrt{10}}{2.7809}$$

$$t = \frac{3.7947}{2.7809}$$

$$\boxed{t = 1.3646}$$

$d.f = n-1 = 10-1 = 9$

Thus $|t| = 1.3646 < 2.262$ at 5% L.o.S with 9 degrees of freedom.

Since the calculated value of $t <$ tabulated value of 't' with 9 d.f at 5% L.o.S

So, we accept H_0 at 5% L.o.S

Problem-3. The average losses of workers, before and After certain program are given below. Use 0.05 Level of Significance of test whether the program is effective (paired Sample t-test). 40 and 35, 70 and 65, 45 and 42, 120 and 116, ~~35 and 55~~, 55 and 50, 35 and 33, 77 and 73.

Ans

Sol:

$\therefore t = 4.16 > 2.262$ //.

F-Test (OR) Snedecor's F-Test of Significane

The test is named in the honor of the great statistician R.A. Fisher.

Objective of F-Test:

- To find out whether the two independent estimates of population variance differ significantly.

(OR)

- To find out whether the two samples may be regarded as drawn from the normal populations having same variance.

Test for equality of two population variances:

Let two independent random samples of sizes n_1 and n_2 be drawn from two normal populations.

To test the hypothesis that the two population variances σ_1^2 and σ_2^2 are equal.

Step:1 Let the null hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$.

Step:2 Then the Alternative hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Step:3 The estimates of σ_1^2 and σ_2^2 are given by

$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} (or) \frac{\sum(x_i - \bar{x})^2}{n_1 - 1}$ and $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} (or) \frac{\sum(y_i - \bar{y})^2}{n_2 - 1}$, where s_1^2 and s_2^2 are the variances of the two samples.

Assuming that H_0 is true, the test statistic is

$\boxed{F = \frac{S_1^2}{S_2^2}}$ when $S_1^2 > S_2^2$ (or) $\boxed{F = \frac{S_2^2}{S_1^2}}$ when $S_2^2 > S_1^2$ follows F-distribution with $(n_1 - 1, n_2 - 1)$ degrees of freedom.

Step:4 Set the level of significance α.

Step:5 If the calculated value of F> the tabulated value of F at α, we reject the null hypothesis H_0 and conclude that the variances σ_1^2 and σ_2^2 are not equal . Otherwise, we accept the null hypothesis H_0 and conclude that the variances σ_1^2 and σ_2^2 are equal.

Note:

- In numerical problem, we take the greater of the two variances S_1^2 and S_2^2 in the numerator and the other in the denominator.

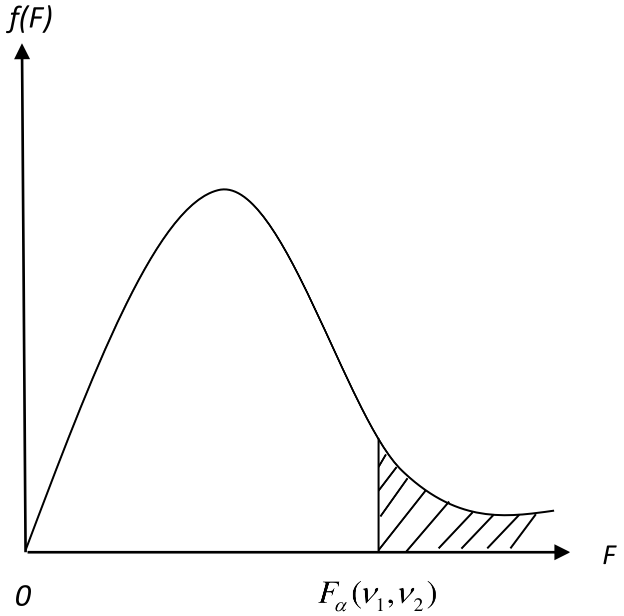
$$i.e,\ F = \frac{Greater\ variance}{Smaller\ variance}$$

- When F is close to 1, the two sample variances S_1 and S_2 are nearly same.

- If sample variance S^2 is given, we can obtain population variance σ^2 by using the relation $n\sigma^2 = (n-1)S^2$ and vice-versa.

Properties of F-distribution:

1. F-distribution curve is Skewed towards right with range 0 to ∞ and having the roughly median value 1.
2. Value of F will always be more than 0.
3. Shape of F-distribution curve is dependent on – d.f. of numerator & d.f. of denominator.
4. F-distribution curve is never symmetrical, but if d.f. will be increased then it will be more similar to the symmetrical shape.
5. F cannot be negative, and it is a continuous distribution.



F-distribution

Solved Problems

1. In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level whether the populations have same variance.

Solution: Let σ_1^2 and σ_2^2 be the variances of the two normal populations from which the samples are drawn.

Here $n_1 = 8, n_2 = 10$

Step:1 Let the null hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$.

Step:2 Then the Alternative hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Step:3 The estimates of σ_1^2 and σ_2^2 are given by

$$S_1^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n_1 - 1} = \frac{84.4}{8-1} = 12.057 \text{ and } S_2^2 = \frac{\sum\left(y_i - \bar{y}\right)^2}{n_2 - 1} = \frac{102.6}{10-1} = 11.4.$$

Assuming that H_0 is true. Since $S_1^2 > S_2^2$, the test statistic is $F = \frac{S_1^2}{S_2^2} = \frac{12.057}{11.4} = 1.057$

with ($(v_1, v_2) = n_1 - 1, n_2 - 1$)=(8-1,10-1)=(7,9) degrees of freedom.

Step:4 Set the level of significance $\alpha = 0.05$.

Step:5 The tabulated value of F at 5% level for (7,9) degrees of freedom is 3.29.

The calculated value of F< the tabulated value of F at $\alpha = 0.05$, we accept the null hypothesis H_0 and conclude that the variances σ_1^2 and σ_2^2 are equal.

2. Two random samples reveal the following results:

Sample	Size	Sample Mean	Sum of Squares of deviations from the mean
1	10	15	90
2	12	14	108

Test whether the samples came from the same normal population.

Solution: Let σ_1^2 and σ_2^2 be the variances of the two normal populations from which the samples are drawn.

Here we have to use two tests *(i)* To test equality of variances by F-test *(ii)* To test equality of means by t-test.

(i) ***F*-test (equality of variances)**

Given $n_1 = 10, n_2 = 12, \bar{x} = 15, \bar{y} = 14$

$$S_1^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n_1 - 1} = \frac{90}{10-1} = 10 \text{ and } S_2^2 = \frac{\sum\left(y_i - \bar{y}\right)^2}{n_2 - 1} = \frac{108}{12-1} = 9.82$$

$$F = \frac{S_1^2}{S_2^2} = \frac{10}{9.82} = 1.018$$

i.e, Calculated F= 1.018. Assuming that H_0 is true. Since $S_1^2 > S_2^2$, the test statistic is with ($n_1 - 1, n_2 - 1$)=(9,11) degrees of freedom. Set the level of significance $\alpha = 0.05$.

The tabulated value of F at 5% level for (9,11) degrees of freedom is 2.89.

The calculated value of F< the tabulated value of F at $\alpha = 0.05$, we accept the null hypothesis H_0 and conclude that the samples came from the same normal populations with same variances.

(ii) **t-test (to test equality of means)**:

Null hypothesis: $H_0 : \mu_1 = \mu_2$

Given $\bar{x} = 15, \bar{y} = 14, n_1 = 10, n_2 = 12$

$$\text{Now } S^2 = \frac{1}{n_1 + n_2 - 1}\left[\sum\left(x - \bar{x}\right)^2 + \sum\left(y - \bar{y}\right)^2\right] = \frac{1}{10+12-1}[90+108] = 9.9$$

$\therefore S = 3.15$

$$\text{The test statistic is } t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{15-14}{3.15} = 0.74.$$

Tabulated value of t for 20 d.f. (n_1+n_2-2) at 5% level of significance is 2.086.

Since calculated value of t < tabulated value of t , we accept the null hypothesis.

Hence from *(i)* and *(ii),* the given samples have been drawn from the same normal populations. Hence we accept the null hypothesis that $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$

3. The nicotine contents in milligrams in two samples of tobacco were found to be as follows:

Sample A	24	27	26	31	25	---
Sample B	27	30	28	31	22	36

Can it be said that the two samples have come from the same normal population.

4. The measurements of the output of two units have given the following results. Assuming that both samples have been obtained from the normal population at 10% significant level, test whether the two populations have the same variances.

Unit-A	14.1	10.1	14.7	13.7	14.0
Unit-B	14.0	14.5	13.7	12.7	14.1

Solution: Let the Null hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$

Then the Alternate hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

Given $n_1 = 5 \& n_2 = 5$

Now

$$\bar{x} = \frac{\sum x}{n} = \frac{1}{5}(14.1 + 10.1 + 14.7 + 13.7 + 14.0) = 13.32$$

$$\bar{y} = \frac{\sum y}{n} = \frac{1}{5}(14.0 + 14.5 + 13.7 + 12.7 + 14.1) = 13.8$$

x	$x-\bar{x}$ =x-13.32	$(x-\bar{x})^2$	y	$y-\bar{y}$ =y-13.8	$(y-\bar{y})^2$
14.1	0.78	0.6084	14.0	0.2	0.04
10.1	-3.22	10.3684	14.5	0.7	0.49
14.7	1.38	1.9044	13.7	-0.1	0.01
13.7	0.38	0.1444	12.7	-1.1	1.21
14.0	0.68	0.4624	14.1	0.5	0.09
$\sum x$=66.6		$\sum(x-\bar{x})^2$=13.488	$\sum y = 69$		$\sum(y-\bar{y})^2$=1.84

$$S_1^2 = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1} = \frac{13.4888}{5-1} = 3.372 \text{ and } S_2^2 = \frac{\sum (y_i - \bar{y})^2}{n_2 - 1} = \frac{1.84}{5-1} = 0.46$$

$$F = \frac{S_1^2}{S_2^2} = \frac{3.372}{0.46} = 7.33$$

Since $S_1^2 > S_2^2$, the test statistic is with ($n_1 - 1, n_2 - 1$)=(4,4) degrees of freedom.

Tabulated value of F for (4,4) d.f. at 10% =0.01 level of significance is 15.97.

Since calculated F < tabulated F, we accept the Null hypothesis H_0.

i.e, There is no significant difference between the variances.

5. Two independent samples of 8 & 7 items respectively had the following values of the variables.

Sample I	9	11	13	11	16	10	12	14
Sample II	11	13	11	14	10	8	10	---

Do the estimates of the population variance differ significantly?

Solution: Let the Null hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$

Then the Alternate hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

Given $n_1 = 8 \& n_2 = 7$

Now

$$\bar{x} = \frac{\sum x}{n_1} = \frac{1}{8}(9+11+13+11+16+10+12+14) = \frac{96}{8} = 12$$

$$\bar{y} = \frac{\sum y}{n_2} = \frac{1}{7}(11+13+11+14+10+8+10) = 11$$

x	$x-\bar{x}$ =x-12	$(x-\bar{x})^2$	y	$y-\bar{y}$ =y-13.8	$(y-\bar{y})^2$
9	-3	9	11	0	0
11	-1	1	13	2	4
13	1	1	11	0	0
11	-1	1	14	3	9
16	4	16	10	-1	1
10	-2	4	8	-3	9
12	0	0	10	-1	1
14	2	4	---	---	---
$\sum x$=96		$\sum(x-\bar{x})^2$=36	$\sum y = 77$		$\sum(y-\bar{y})^2$=24

$$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1} = \frac{36}{8-1} = 5.14 \text{ and } S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1} = \frac{24}{7-1} = 4$$

$$F = \frac{S_1^2}{S_2^2} = \frac{5.14}{4} = 1.285$$

Since $S_1^2 > S_2^2$, the test statistic is with ($n_1 - 1, n_2 - 1$)=(7,6) degrees of freedom.

Tabulated value of F for (7, 6) d.f. at 5% =0.05 level of significance is 4.21.

Since calculated F < tabulated F, we accept the Null hypothesis H_0.

i.e, There is no significant difference between the variances.

6. It is known that the mean diameters of rivets produced by two firms A and B are practically the same, but the standard deviation may differ. For 22 rivets produced by firm A, the S.D. is 2.9 mm, while for 16 rivets manufactured by firm B, the S.D. is 3.8 mm, compute the statistic you would use to test whether the products of firm A have the same variability as those of firm B and test its significance.

Solution: Given $n_1 = 22, n_2 = 16, s_1 = 2.9mm, s_2 = 3.8mm$

Since the S.D's of the samples $s_1 \& s_2$ are given.

Step:1 Let the null hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$.

Step:2 Then the Alternative hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Step:3 The population variances S_1^2 and S_2^2 are obtained by using the relations $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{22(2.9)^2}{21} = 8.805$ and $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{16(3.8)^2}{15} = 15.393$, where s_1^2 and s_2^2 are the variances of the two samples.

Assuming that H_0 is true and $S_2^2 > S_1^2$ then the test statistic is

$$F = \frac{S_2^2}{S_1^2} = \frac{15.393}{15} = 1.74$$

Tabulated value of F- with (21, 15) d.f. at 0.05 level of significance is 2.31.

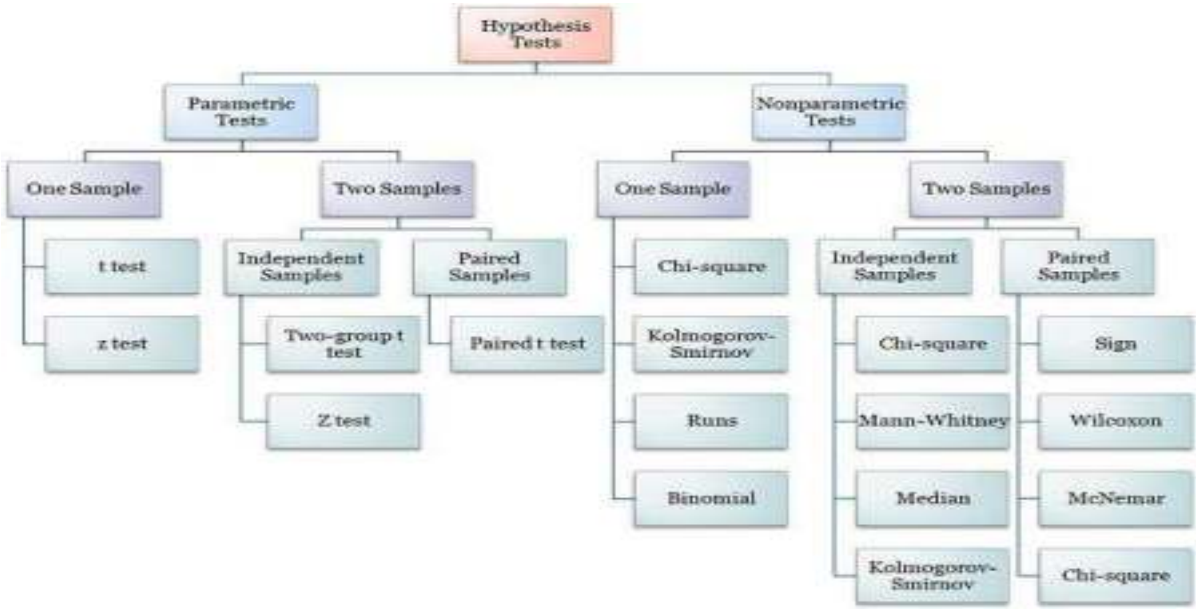
Step:4 The calculated value of F< the tabulated value of F at $\alpha = 0.05$, we accept the null hypothesis H_0 *i.e,* the products of both the firms A and B have the same variability. So we may conclude that the products of firm A are not superior to those of firm B .

7. Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumpkins, show the sample standard deviations of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test hypothesis that the true variances are equal.

8. The time taken by workers in performing a job by Method I and Method II is given below.

Method I	20	16	26	27	23	22	---
Method II	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly?



To make the generalization about the population from the sample, statistical tests are used. A statistical test is a formal technique that relies on the probability distribution, for reaching the conclusion concerning the reasonableness of the hypothesis. These hypothetical testing related to differences are classified as **parametric and nonparametric** tests.

The **parametric test** is one which has information about the population.

Ex: z-test, t-test and F-test.

On the other hand, the **nonparametric test** is one whether no exact information about the population . Chi-square test is commonly used non-parametric test.

χ^2 - test:

- First used by Karl Pearson in 1900.
- Denoted by square of the Greek letter χ .
- The quantity χ^2 describes the magnitude of the discrepancy between theory and observations.

Properties:

1. χ^2 **-** distribution curve is not symmetrical, lies entirely in the first quadrant, and hence not a normal curve, since χ^2 varies from 0 to ∞.
2. As the number of degrees of freedom increases, the chi-square distribution becomes more symmetric.
3. It depends only on the degree of freedom ν .
4. The values are non-negative. *i.e,* the values of are greater than or equal to 0.
5. Mean= ν and variance=2ν .

$f(\chi^2)$

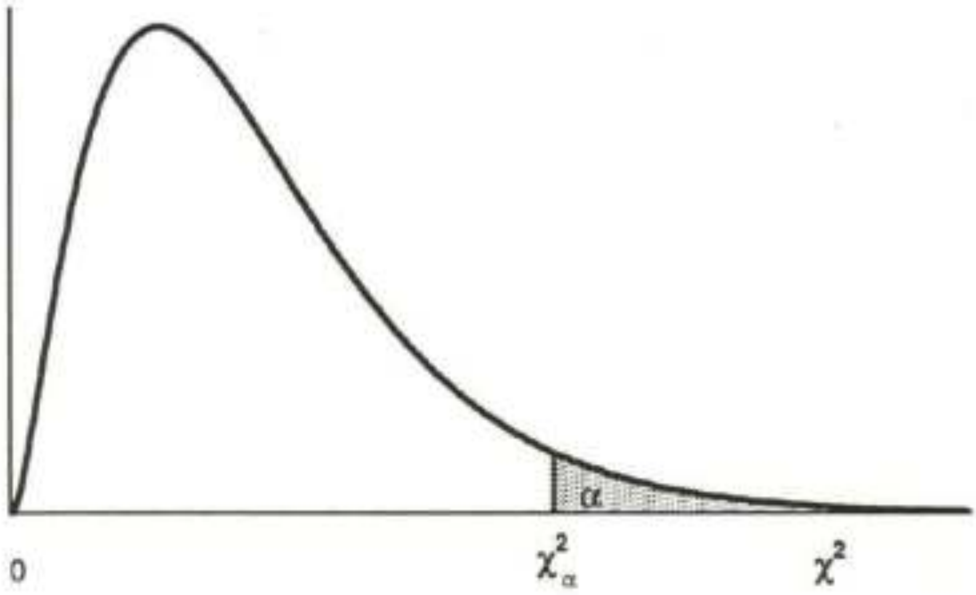


Figure J.1: The χ^2 distribution

Applications of χ^2 distribution:

1. To test the goodness of fit.
2. To test the independence of attributes.
3. To test the homogeneity of independent estimation of the population variances.
4. To test the homogeneity of independent estimation of the population Correlation coefficient.

Conditions of validity :

Following are the conditions which should be satisfied before χ^2 test can be applied.

1. The sample observations should be independent.
2. N, the total frequency is large, *i.e*, >50.
3. The constraints on the cell frequencies, if any, are linear.
4. No theoretical (or expected) frequency should be less than 10. If small theoretical frequencies occur, the difficulty is overcome by regrouping 2 or more classes together before calculating (*O-E*). Note that the degrees of freedom is determined with the number of classes after regrouping.

Definition: If a set of events $A_1, A_2, \ldots, A_n$ are observed to occur with frequencies $O_1, O_2, \ldots, O_n$ respectively and according to probability rules $A_1, A_2, \ldots, A_n$ are expected to occur with frequencies $E_1, E_2, \ldots, E_n$ respectively with $O_1, O_2, \ldots, O_n$ are called observed frequencies and $E_1, E_2, \ldots, E_n$ are called expected frequencies.

If O_i (i=1,2,…..,n) is a set of observed (experimental) frequencies and E_i (i=1,2,…..,n) is the corresponding set of expected (theoretical) frequencies, then χ^2 is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$ with *(n-1)* degrees of freedom.

Note:

If the data is given in a series of ‘n’ numbers then degrees of freedom = *n*-1.

In case of Binomial distribution, d.f $=n-1$.
In case of Poisson distribution, d.f $=n-2$.
In case of Normal distribution, d.f $=n-3$.

χ^2 -**Test as a test of Goodness of fit:**

We use this test to decide whether the discrepancy between theory and experiment is significant or not. i.e, to test whether the difference between the theoretical and observed values can be attributed to chance or not.

Let the Null hypothesis H_0 be there is no significant difference between the observed values and the corresponding expected values.

Then the Alternative hypothesis H_1 is that the above difference is significant.

Let $O_1, O_2, \ldots, O_n$ be a set of observed frequencies and $E_1, E_2, \ldots, E_n$ the corresponding expected frequencies. Then the test statistic χ^2 is given by

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Assuming that H_0 is true, the test statistic χ^2 follows Chi-square distribution with (n-1) d.f. where

$$\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i \ (or) \sum_{i=1}^{n} (O_i - E_i) = 0$$

Solved Problems

1. The number of automobile accidents per week in a certain community are as follows: 12,8,20,2,14,10,15,6,9,4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.

Solution: Expected frequency of accidents each week= $\frac{100}{10} = 10$.

Null hypothesis H_0: The accident conditions were the same during the 10 week period.

Alternative hypothesis H_0: The accident conditions are different during 10 week period.

Observed Frequency (O_i)	Expected Frequency (E_i)	(O_i-E_i)	$\frac{(O_i - E_i)^2}{E_i}$
12	10	2	0.4
8	10	-2	0.4
20	10	10	10.0
2	10	-8	6.4
14	10	4	1.6
10	10	0	0.0
15	10	5	2.5
6	10	-4	1.6
9	10	-1	0.1
4	10	-6	3.6
100	100		26.6

Now, $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 26.6.$

i.e, Calculate value of $\chi^2 = 26.6$

Here n=10 observations are given.

$\therefore$ Degrees of freedom (d.f)=n-1=10-1=9.

Tabulated value at 0.05 with 9 d.f. is $\chi^2 = 16.9$

Since Calculated χ^2 > Tabulated χ^2, therefore the Null hypothesis is rejected and concluded that the accident conditions were not the same during the 10 week period.

2. The following figures show the distribution of digits in numbers chosen at random from a telephone directory.

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur equally in the directory.

Solution: **Null hypothesis H_0**: The digits occur equally frequently in the directory.

Alternative hypothesis H_1: The digits occur differently frequently in the directory.

Digits	Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
0	1026	1000	676	0.676
1	1107	1000	11449	11.449
2	997	1000	9	0.009
3	996	1000	1156	1.156
4	1075	1000	5625	5.625
5	933	1000	4489	4.489
6	1107	1000	11449	11.449
7	972	1000	784	0.784
8	964	1000	1296	1.296
9	853	1000	21609	21.609
Total	10000	10000		58.542

Now, $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 58.542$

i.e, Calculate value of $\chi^2 = 58.542$

Here n=10 observations are given.

$\therefore$ Degrees of freedom (d.f)=n-1=10-1=9.

Tabulated value at 0.05 with 9 d.f. is $\chi^2 = 16.9$

Since Calculated χ^2 > Tabulated χ^2, therefore the Null hypothesis is rejected and concluded that the digits do not occur equally frequently in the directory.

3. A die is thrown 264 times with the following results. Show that the die is biased. [Given $\chi^2_{0.05} = 11.07$ *for* $5d.f$]

No. appeared on the die	1	2	3	4	5	6
Frequency	40	32	28	58	54	52

4. A survey of 240 families with 4 children each revealed the following distribution.

Male Births	4	3	2	1	0
Observed frequency	10	55	105	58	12

Solution: **Null hypothesis H_0**: The male and female births are equally probable.

$i.e, \ p = q = \frac{1}{2}$

The expected frequency *x* of male births is given by

$f(x) = N \times C_k^n p^x q^{n-k}$, where N=240, *n*=4, *x*=0,1,2,3,4

$$\therefore f(0) = 240 \times C_0^4 \times \left(\frac{1}{2}\right)^0 \times \left(\frac{1}{2}\right)^{4-0} = 240 \times 1 \times 1 \times \frac{1}{16} = 15$$

$$f(1) = 240 \times C_1^4 \times \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^{4-1} = 240 \times 4 \times \frac{1}{2} \times \frac{1}{8} = 60$$

$$f(2) = 240 \times C_2^4 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^{4-2} = 240 \times 6 \times \frac{1}{4} \times \frac{1}{4} = 90$$

$$f(3) = 240 \times C_3^4 \times \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^{4-3} = 240 \times 4 \times \frac{1}{8} \times \frac{1}{2} = 60$$

$$f(4) = 240 \times C_4^4 \times \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^{4-4} = 240 \times 1 \times \frac{1}{16} \times 1 = 15$$

The expected or theoretical (Binomial) frequencies of male births are:

x	4	3	2	1	0
f(x)	15	60	90	60	15

Let us now apply χ^2 test to examine the goodness of fit of the given data to the above Binomail distribution.

No.of families				
Observed Frequency (O_i)	Expected Frequency (E_i)	(O_i-E_i)	$(O_i-E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
10	15	-5	25	1.67
55	60	-5	25	0.42
105	90	15	225	2.5
58	60	-2	4	0.07
12	15	-3	9	0.6
Total: 240	240			5.26

Now, $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 5.26$

i.e, Calculate value of $\chi^2 = 5.26$

Here n=5 observations are given.

$\therefore$ Degrees of freedom (d.f)=n-1=5-1=4. Tabulated value at 0.05 with 4 d.f. is $\chi^2 = 9.488$

Since Calculated χ^2 < Tabulated χ^2, therefore the Null hypothesis is accepted and conclude that the male and female births are equally probable.

5. 4 coins were tossed 160times and the following results were obtained.

No.of Heads	0	1	2	3	4
Observed frequencies	17	52	54	31	6

Under the assumption that coins are balanced, find the expected frequencies of 0,1,2,3 or 4 heads and test the goodness of fit $(\alpha = 0.05)$.

6. Fit a Poisson distribution to the following data and for its goodness of fits at level of significance 0.05?

x	0	1	2	3	4
f(x)	419	352	154	56	19

Solution:

x	*f*	*f.x*
0	419	0
1	352	352
2	154	308
3	56	168
4	19	76
	N= $\sum f$ =1000	$\sum fx = 904$

Mean $\lambda = \frac{\sum f_i x_i}{\sum f} = \frac{904}{1000} = 0.904$

Theoretical distribution is given by $N \times p(x) = 1000 \times \frac{e^{-\lambda} \lambda^x}{x!}$

Hence the theoretical frequencies are given by

$$f(x) = 1000 \times \frac{e^{-0.904}(0.904)^x}{x!} = \frac{1000 \times 0.4049 \times (0.904)^x}{x!} \text{-----} (1)$$

Putting *x=0,1,2,3,4*, we get

x	0	1	2	3	4
f(x)	406.2	366	165.4	49.8	12.6

Now,

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \frac{(419-406.2)^2}{406.2} + \frac{(352-366)^2}{366.6} + \frac{(154-165.4)^2}{165.4} + \frac{(56-49.8)^2}{49.8} + \frac{(19-12.6)^2}{12.6}$$

$$= 5.748$$

$\therefore$ Degrees of freedom (d.f)=n-2=5-2=3. Tabulated value at 0.05 with 3 d.f. is $\chi^2 = 7.82$

Since Calculated χ^2 < Tabulated χ^2, therefore the Null hypothesis is accepted.

7. A sample analysis of examination results of 500 students was made. It was found that 220 students had failed, 170 had secured a third class, 90 were placed in second class and 20 got a first class. Do these figures commensurate with the general examination result which is in the ratio of 4:3:2:1 for the various categories respectively.

Solution: Null Hypothesis: H_0: The observed results commensurate with the general examination results.

Expected frequencies are in the ratio of 4:3:2:1

Total frequency=500

If we divide the total frequency 500 in the ratio 4:3:2:1, we get the expected frequencies as

$$500\times\frac{4}{10}=200;\ 500\times\frac{3}{10}=150;500\times\frac{2}{10}=100 \text{ and } 500\times\frac{1}{10}=50$$

Class	Observed Frequency (O_i)	Expected Frequency (E_i)	(O_i-E_i)	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
Failed	220	200	20	400	2.06
Third	170	150	20	400	2.667
Second	90	100	-10	100	1.000
First	20	50	-30	900	18.00
	500	500			23.667

$$\chi^2=\sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i}=23.667$$

i.e, Calculate value of $\chi^2=23.667$

Here n=4 observations are given.

$\therefore$ Degrees of freedom (d.f)=n-1=4-1=3. Tabulated value at 0.05 with 3 d.f. is $\chi^2=7.82$

Since Calculated χ^2 >Tabulated χ^2, therefore the Null hypothesis is rejected and conclude that the observed results are not commensurate with the general examination.

8. A pair of dice is thrown 360 times and the frequency of each sum is indicated below.

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of the Chi-Square test at 0.05 level of significance?

One Tailed Student's t-Distribution Table

α / df	0.01	0.03	0.05	0.1	0.2	0.25	0.5
1	127.32	63.66	21.20	12.71	6.31	5.03	2.41
2	14.09	9.92	5.64	4.30	2.92	2.56	1.60
3	7.45	5.84	3.90	3.18	2.35	2.11	1.42
4	5.60	4.60	3.30	2.78	2.13	1.94	1.34
5	4.77	4.03	3.00	2.57	2.02	1.84	1.30
6	4.32	3.71	2.83	2.45	1.94	1.78	1.27
7	4.03	3.50	2.71	2.36	1.89	1.74	1.25
8	3.83	3.36	2.63	2.31	1.86	1.71	1.24
9	3.69	3.25	2.57	2.26	1.83	1.69	1.23
10	3.58	3.17	2.53	2.23	1.81	1.67	1.22
11	3.50	3.11	2.49	2.20	1.80	1.66	1.21
12	3.43	3.05	2.46	2.18	1.78	1.65	1.21
13	3.37	3.01	2.44	2.16	1.77	1.64	1.20
14	3.33	2.98	2.41	2.14	1.76	1.63	1.20
15	3.29	2.95	2.40	2.13	1.75	1.62	1.20
16	3.25	2.92	2.38	2.12	1.75	1.62	1.19
17	3.22	2.90	2.37	2.11	1.74	1.61	1.19
18	3.20	2.88	2.36	2.10	1.73	1.61	1.19
19	3.17	2.86	2.35	2.09	1.73	1.60	1.19
20	3.15	2.85	2.34	2.09	1.72	1.60	1.18
21	3.14	2.83	2.33	2.08	1.72	1.60	1.18
22	3.12	2.82	2.32	2.07	1.72	1.59	1.18
23	3.10	2.81	2.31	2.07	1.71	1.59	1.18
24	3.09	2.80	2.31	2.06	1.71	1.59	1.18
25	3.08	2.79	2.30	2.06	1.71	1.59	1.18
26	3.07	2.78	2.30	2.06	1.71	1.59	1.18
27	3.06	2.77	2.29	2.05	1.70	1.58	1.18
28	3.05	2.76	2.29	2.05	1.70	1.58	1.17
29	3.04	2.76	2.28	2.05	1.70	1.58	1.17

One Tailed Student's t-Distribution Table

30	3.03	2.75	2.28	2.04	1.70	1.58	1.17
31	3.02	2.74	2.27	2.04	1.70	1.58	1.17
32	3.01	2.74	2.27	2.04	1.69	1.58	1.17
33	3.01	2.73	2.27	2.03	1.69	1.57	1.17
34	3.00	2.73	2.27	2.03	1.69	1.57	1.17
35	3.00	2.72	2.26	2.03	1.69	1.57	1.17
36	2.99	2.72	2.26	2.03	1.69	1.57	1.17
37	2.99	2.72	2.26	2.03	1.69	1.57	1.17
38	2.98	2.71	2.25	2.02	1.69	1.57	1.17
39	2.98	2.71	2.25	2.02	1.68	1.57	1.17
40	2.97	2.70	2.25	2.02	1.68	1.57	1.17
41	2.97	2.70	2.25	2.02	1.68	1.57	1.17
42	2.96	2.70	2.25	2.02	1.68	1.57	1.17
43	2.96	2.70	2.24	2.02	1.68	1.56	1.17
44	2.96	2.69	2.24	2.02	1.68	1.56	1.17
45	2.95	2.69	2.24	2.01	1.68	1.56	1.17
46	2.95	2.69	2.24	2.01	1.68	1.56	1.17
47	2.95	2.68	2.24	2.01	1.68	1.56	1.16
48	2.94	2.68	2.24	2.01	1.68	1.56	1.16
49	2.94	2.68	2.24	2.01	1.68	1.56	1.16
50	2.94	2.68	2.23	2.01	1.68	1.56	1.16
51	2.93	2.68	2.23	2.01	1.68	1.56	1.16
52	2.93	2.67	2.23	2.01	1.67	1.56	1.16
53	2.93	2.67	2.23	2.01	1.67	1.56	1.16
54	2.93	2.67	2.23	2.00	1.67	1.56	1.16
55	2.92	2.67	2.23	2.00	1.67	1.56	1.16
56	2.92	2.67	2.23	2.00	1.67	1.56	1.16
57	2.92	2.66	2.23	2.00	1.67	1.56	1.16
58	2.92	2.66	2.22	2.00	1.67	1.56	1.16
59	2.92	2.66	2.22	2.00	1.67	1.56	1.16
60	2.91	2.66	2.22	2.00	1.67	1.56	1.16

One Tailed Student's t-Distribution Table

61	2.91	2.66	2.22	2.00	1.67	1.56	1.16
62	2.91	2.66	2.22	2.00	1.67	1.56	1.16
63	2.91	2.66	2.22	2.00	1.67	1.55	1.16
64	2.91	2.65	2.22	2.00	1.67	1.55	1.16
65	2.91	2.65	2.22	2.00	1.67	1.55	1.16
66	2.90	2.65	2.22	2.00	1.67	1.55	1.16
67	2.90	2.65	2.22	2.00	1.67	1.55	1.16
68	2.90	2.65	2.22	2.00	1.67	1.55	1.16
69	2.90	2.65	2.22	1.99	1.67	1.55	1.16
70	2.90	2.65	2.22	1.99	1.67	1.55	1.16
71	2.90	2.65	2.21	1.99	1.67	1.55	1.16
72	2.90	2.65	2.21	1.99	1.67	1.55	1.16
73	2.89	2.64	2.21	1.99	1.67	1.55	1.16
74	2.89	2.64	2.21	1.99	1.67	1.55	1.16
75	2.89	2.64	2.21	1.99	1.67	1.55	1.16
76	2.89	2.64	2.21	1.99	1.67	1.55	1.16
77	2.89	2.64	2.21	1.99	1.66	1.55	1.16
78	2.89	2.64	2.21	1.99	1.66	1.55	1.16
79	2.89	2.64	2.21	1.99	1.66	1.55	1.16
80	2.89	2.64	2.21	1.99	1.66	1.55	1.16
81	2.89	2.64	2.21	1.99	1.66	1.55	1.16
82	2.88	2.64	2.21	1.99	1.66	1.55	1.16
83	2.88	2.64	2.21	1.99	1.66	1.55	1.16
84	2.88	2.64	2.21	1.99	1.66	1.55	1.16
85	2.88	2.63	2.21	1.99	1.66	1.55	1.16
86	2.88	2.63	2.21	1.99	1.66	1.55	1.16
87	2.88	2.63	2.21	1.99	1.66	1.55	1.16
88	2.88	2.63	2.21	1.99	1.66	1.55	1.16
89	2.88	2.63	2.21	1.99	1.66	1.55	1.16
90	2.88	2.63	2.21	1.99	1.66	1.55	1.16
91	2.88	2.63	2.20	1.99	1.66	1.55	1.16
92	2.88	2.63	2.20	1.99	1.66	1.55	1.16

One Tailed Student's t-Distribution Table

93	2.88	2.63	2.20	1.99	1.66	1.55	1.16
94	2.87	2.63	2.20	1.99	1.66	1.55	1.16
95	2.87	2.63	2.20	1.99	1.66	1.55	1.16
96	2.87	2.63	2.20	1.98	1.66	1.55	1.16
97	2.87	2.63	2.20	1.98	1.66	1.55	1.16
98	2.87	2.63	2.20	1.98	1.66	1.55	1.16
99	2.87	2.63	2.20	1.98	1.66	1.55	1.16
100	2.87	2.63	2.20	1.98	1.66	1.55	1.16

Two Tailed Student's t-Distribution Table

α / df	0.01	0.03	0.05	0.1	0.2	0.25	0.5
1	63.66	31.82	12.71	6.31	3.08	2.41	1.00
2	9.92	6.96	4.30	2.92	1.89	1.60	0.82
3	5.84	4.54	3.18	2.35	1.64	1.42	0.76
4	4.60	3.75	2.78	2.13	1.53	1.34	0.74
5	4.03	3.36	2.57	2.02	1.48	1.30	0.73
6	3.71	3.14	2.45	1.94	1.44	1.27	0.72
7	3.50	3.00	2.36	1.89	1.41	1.25	0.71
8	3.36	2.90	2.31	1.86	1.40	1.24	0.71
9	3.25	2.82	2.26	1.83	1.38	1.23	0.70
10	3.17	2.76	2.23	1.81	1.37	1.22	0.70
11	3.11	2.72	2.20	1.80	1.36	1.21	0.70
12	3.05	2.68	2.18	1.78	1.36	1.21	0.70
13	3.01	2.65	2.16	1.77	1.35	1.20	0.69
14	2.98	2.62	2.14	1.76	1.35	1.20	0.69
15	2.95	2.60	2.13	1.75	1.34	1.20	0.69
16	2.92	2.58	2.12	1.75	1.34	1.19	0.69
17	2.90	2.57	2.11	1.74	1.33	1.19	0.69
18	2.88	2.55	2.10	1.73	1.33	1.19	0.69
19	2.86	2.54	2.09	1.73	1.33	1.19	0.69
20	2.85	2.53	2.09	1.72	1.33	1.18	0.69
21	2.83	2.52	2.08	1.72	1.32	1.18	0.69
22	2.82	2.51	2.07	1.72	1.32	1.18	0.69
23	2.81	2.50	2.07	1.71	1.32	1.18	0.69
24	2.80	2.49	2.06	1.71	1.32	1.18	0.68
25	2.79	2.49	2.06	1.71	1.32	1.18	0.68
26	2.78	2.48	2.06	1.71	1.31	1.18	0.68
27	2.77	2.47	2.05	1.70	1.31	1.18	0.68
28	2.76	2.47	2.05	1.70	1.31	1.17	0.68
29	2.76	2.46	2.05	1.70	1.31	1.17	0.68

Two Tailed Student's t-Distribution Table

30	2.75	2.46	2.04	1.70	1.31	1.17	0.68
31	2.74	2.45	2.04	1.70	1.31	1.17	0.68
32	2.74	2.45	2.04	1.69	1.31	1.17	0.68
33	2.73	2.44	2.03	1.69	1.31	1.17	0.68
34	2.73	2.44	2.03	1.69	1.31	1.17	0.68
35	2.72	2.44	2.03	1.69	1.31	1.17	0.68
36	2.72	2.43	2.03	1.69	1.31	1.17	0.68
37	2.72	2.43	2.03	1.69	1.30	1.17	0.68
38	2.71	2.43	2.02	1.69	1.30	1.17	0.68
39	2.71	2.43	2.02	1.68	1.30	1.17	0.68
40	2.70	2.42	2.02	1.68	1.30	1.17	0.68
41	2.70	2.42	2.02	1.68	1.30	1.17	0.68
42	2.70	2.42	2.02	1.68	1.30	1.17	0.68
43	2.70	2.42	2.02	1.68	1.30	1.17	0.68
44	2.69	2.41	2.02	1.68	1.30	1.17	0.68
45	2.69	2.41	2.01	1.68	1.30	1.17	0.68
46	2.69	2.41	2.01	1.68	1.30	1.17	0.68
47	2.68	2.41	2.01	1.68	1.30	1.16	0.68
48	2.68	2.41	2.01	1.68	1.30	1.16	0.68
49	2.68	2.40	2.01	1.68	1.30	1.16	0.68
50	2.68	2.40	2.01	1.68	1.30	1.16	0.68
51	2.68	2.40	2.01	1.68	1.30	1.16	0.68
52	2.67	2.40	2.01	1.67	1.30	1.16	0.68
53	2.67	2.40	2.01	1.67	1.30	1.16	0.68
54	2.67	2.40	2.00	1.67	1.30	1.16	0.68
55	2.67	2.40	2.00	1.67	1.30	1.16	0.68
56	2.67	2.39	2.00	1.67	1.30	1.16	0.68
57	2.66	2.39	2.00	1.67	1.30	1.16	0.68
58	2.66	2.39	2.00	1.67	1.30	1.16	0.68
59	2.66	2.39	2.00	1.67	1.30	1.16	0.68
60	2.66	2.39	2.00	1.67	1.30	1.16	0.68

Two Tailed Student's t-Distribution Table

61	2.66	2.39	2.00	1.67	1.30	1.16	0.68
62	2.66	2.39	2.00	1.67	1.30	1.16	0.68
63	2.66	2.39	2.00	1.67	1.30	1.16	0.68
64	2.65	2.39	2.00	1.67	1.29	1.16	0.68
65	2.65	2.39	2.00	1.67	1.29	1.16	0.68
66	2.65	2.38	2.00	1.67	1.29	1.16	0.68
67	2.65	2.38	2.00	1.67	1.29	1.16	0.68
68	2.65	2.38	2.00	1.67	1.29	1.16	0.68
69	2.65	2.38	1.99	1.67	1.29	1.16	0.68
70	2.65	2.38	1.99	1.67	1.29	1.16	0.68
71	2.65	2.38	1.99	1.67	1.29	1.16	0.68
72	2.65	2.38	1.99	1.67	1.29	1.16	0.68
73	2.64	2.38	1.99	1.67	1.29	1.16	0.68
74	2.64	2.38	1.99	1.67	1.29	1.16	0.68
75	2.64	2.38	1.99	1.67	1.29	1.16	0.68
76	2.64	2.38	1.99	1.67	1.29	1.16	0.68
77	2.64	2.38	1.99	1.66	1.29	1.16	0.68
78	2.64	2.38	1.99	1.66	1.29	1.16	0.68
79	2.64	2.37	1.99	1.66	1.29	1.16	0.68
80	2.64	2.37	1.99	1.66	1.29	1.16	0.68
81	2.64	2.37	1.99	1.66	1.29	1.16	0.68
82	2.64	2.37	1.99	1.66	1.29	1.16	0.68
83	2.64	2.37	1.99	1.66	1.29	1.16	0.68
84	2.64	2.37	1.99	1.66	1.29	1.16	0.68
85	2.63	2.37	1.99	1.66	1.29	1.16	0.68
86	2.63	2.37	1.99	1.66	1.29	1.16	0.68
87	2.63	2.37	1.99	1.66	1.29	1.16	0.68
88	2.63	2.37	1.99	1.66	1.29	1.16	0.68
89	2.63	2.37	1.99	1.66	1.29	1.16	0.68
90	2.63	2.37	1.99	1.66	1.29	1.16	0.68
91	2.63	2.37	1.99	1.66	1.29	1.16	0.68
92	2.63	2.37	1.99	1.66	1.29	1.16	0.68

Two Tailed Student's t-Distribution Table

93	2.63	2.37	1.99	1.66	1.29	1.16	0.68
94	2.63	2.37	1.99	1.66	1.29	1.16	0.68
95	2.63	2.37	1.99	1.66	1.29	1.16	0.68
96	2.63	2.37	1.98	1.66	1.29	1.16	0.68
97	2.63	2.37	1.98	1.66	1.29	1.16	0.68
98	2.63	2.37	1.98	1.66	1.29	1.16	0.68
99	2.63	2.36	1.98	1.66	1.29	1.16	0.68
100	2.63	2.36	1.98	1.66	1.29	1.16	0.68

getcalc.com

Chi-square (χ^2) Distribution Table

α / df	0.1	0.05	0.025	0.01	0.005	0.001
1	2.706	3.841	5.024	6.635	7.879	10.828
2	4.605	5.991	7.378	9.21	10.597	13.816
3	6.251	7.815	9.348	11.345	12.838	16.266
4	7.779	9.488	11.143	13.277	14.86	18.467
5	9.236	11.07	12.833	15.086	16.75	20.515
6	10.645	12.592	14.449	16.812	18.548	22.458
7	12.017	14.067	16.013	18.475	20.278	24.322
8	13.362	15.507	17.535	20.09	21.955	26.124
9	14.684	16.919	19.023	21.666	23.589	27.877
10	15.987	18.307	20.483	23.209	25.188	29.588
11	17.275	19.675	21.92	24.725	26.757	31.264
12	18.549	21.026	23.337	26.217	28.3	32.909
13	19.812	22.362	24.736	27.688	29.819	34.528
14	21.064	23.685	26.119	29.141	31.319	36.123
15	22.307	24.996	27.488	30.578	32.801	37.697
16	23.542	26.296	28.845	32	34.267	39.252
17	24.769	27.587	30.191	33.409	35.718	40.79
18	25.989	28.869	31.526	34.805	37.156	42.312
19	27.204	30.144	32.852	36.191	38.582	43.82
20	28.412	31.41	34.17	37.566	39.997	45.315
21	29.615	32.671	35.479	38.932	41.401	46.797
22	30.813	33.924	36.781	40.289	42.796	48.268
23	32.007	35.172	38.076	41.638	44.181	49.728
24	33.196	36.415	39.364	42.98	45.559	51.179
25	34.382	37.652	40.646	44.314	46.928	52.62
26	35.563	38.885	41.923	45.642	48.29	54.052
27	36.741	40.113	43.195	46.963	49.645	55.476
28	37.916	41.337	44.461	48.278	50.993	56.892
29	39.087	42.557	45.722	49.588	52.336	58.301

Chi-square (χ^2) Distribution Table

30	40.256	43.773	46.979	50.892	53.672	59.703
31	41.422	44.985	48.232	52.191	55.003	61.098
32	42.585	46.194	49.48	53.486	56.328	62.487
33	43.745	47.4	50.725	54.776	57.648	63.87
34	44.903	48.602	51.966	56.061	58.964	65.247
35	46.059	49.802	53.203	57.342	60.275	66.619
36	47.212	50.998	54.437	58.619	61.581	67.985
37	48.363	52.192	55.668	59.893	62.883	69.346
38	49.513	53.384	56.896	61.162	64.181	70.703
39	50.66	54.572	58.12	62.428	65.476	72.055
40	51.805	55.758	59.342	63.691	66.766	73.402
41	52.949	56.942	60.561	64.95	68.053	74.745
42	54.09	58.124	61.777	66.206	69.336	76.084
43	55.23	59.304	62.99	67.459	70.616	77.419
44	56.369	60.481	64.201	68.71	71.893	78.75
45	57.505	61.656	65.41	69.957	73.166	80.077
46	58.641	62.83	66.617	71.201	74.437	81.4
47	59.774	64.001	67.821	72.443	75.704	82.72
48	60.907	65.171	69.023	73.683	76.969	84.037
49	62.038	66.339	70.222	74.919	78.231	85.351
50	63.167	67.505	71.42	76.154	79.49	86.661
51	64.295	68.669	72.616	77.386	80.747	87.968
52	65.422	69.832	73.81	78.616	82.001	89.272
53	66.548	70.993	75.002	79.843	83.253	90.573
54	67.673	72.153	76.192	81.069	84.502	91.872
55	68.796	73.311	77.38	82.292	85.749	93.168
56	69.919	74.468	78.567	83.513	86.994	94.461
57	71.04	75.624	79.752	84.733	88.236	95.751
58	72.16	76.778	80.936	85.95	89.477	97.039
59	73.279	77.931	82.117	87.166	90.715	98.324
60	74.397	79.082	83.298	88.379	91.952	99.607

Chi-square (χ^2) Distribution Table

61	75.514	80.232	84.476	89.591	93.186	100.888
62	76.63	81.381	85.654	90.802	94.419	102.166
63	77.745	82.529	86.83	92.01	95.649	103.442
64	78.86	83.675	88.004	93.217	96.878	104.716
65	79.973	84.821	89.177	94.422	98.105	105.988
66	81.085	85.965	90.349	95.626	99.33	107.258
67	82.197	87.108	91.519	96.828	100.554	108.526
68	83.308	88.25	92.689	98.028	101.776	109.791
69	84.418	89.391	93.856	99.228	102.996	111.055
70	85.527	90.531	95.023	100.425	104.215	112.317
71	86.635	91.67	96.189	101.621	105.432	113.577
72	87.743	92.808	97.353	102.816	106.648	114.835
73	88.85	93.945	98.516	104.01	107.862	116.092
74	89.956	95.081	99.678	105.202	109.074	117.346
75	91.061	96.217	100.839	106.393	110.286	118.599
76	92.166	97.351	101.999	107.583	111.495	119.85
77	93.27	98.484	103.158	108.771	112.704	121.1
78	94.374	99.617	104.316	109.958	113.911	122.348
79	95.476	100.749	105.473	111.144	115.117	123.594
80	96.578	101.879	106.629	112.329	116.321	124.839
81	97.68	103.01	107.783	113.512	117.524	126.083
82	98.78	104.139	108.937	114.695	118.726	127.324
83	99.88	105.267	110.09	115.876	119.927	128.565
84	100.98	106.395	111.242	117.057	121.126	129.804
85	102.079	107.522	112.393	118.236	122.325	131.041
86	103.177	108.648	113.544	119.414	123.522	132.277
87	104.275	109.773	114.693	120.591	124.718	133.512
88	105.372	110.898	115.841	121.767	125.913	134.745
89	106.469	112.022	116.989	122.942	127.106	135.978
90	107.565	113.145	118.136	124.116	128.299	137.208
91	108.661	114.268	119.282	125.289	129.491	138.438
92	109.756	115.39	120.427	126.462	130.681	139.666

Chi-square (χ^2) Distribution Table

93	110.85	116.511	121.571	127.633	131.871	140.893
94	111.944	117.632	122.715	128.803	133.059	142.119
95	113.038	118.752	123.858	129.973	134.247	143.344
96	114.131	119.871	125	131.141	135.433	144.567
97	115.223	120.99	126.141	132.309	136.619	145.789
98	116.315	122.108	127.282	133.476	137.803	147.01
99	117.407	123.225	128.422	134.642	138.987	148.23
100	118.498	124.342	129.561	135.807	140.169	149.449
101	119.589	125.458	130.7	136.971	141.351	150.667
102	120.679	126.574	131.838	138.134	142.532	151.884
103	121.769	127.689	132.975	139.297	143.712	153.099
104	122.858	128.804	134.111	140.459	144.891	154.314
105	123.947	129.918	135.247	141.62	146.07	155.528
106	125.035	131.031	136.382	142.78	147.247	156.74
107	126.123	132.144	137.517	143.94	148.424	157.952
108	127.211	133.257	138.651	145.099	149.599	159.162
109	128.298	134.369	139.784	146.257	150.774	160.372
110	129.385	135.48	140.917	147.414	151.948	161.581
111	130.472	136.591	142.049	148.571	153.122	162.788
112	131.558	137.701	143.18	149.727	154.294	163.995
113	132.643	138.811	144.311	150.882	155.466	165.201
114	133.729	139.921	145.441	152.037	156.637	166.406
115	134.813	141.03	146.571	153.191	157.808	167.61
116	135.898	142.138	147.7	154.344	158.977	168.813
117	136.982	143.246	148.829	155.496	160.146	170.016
118	138.066	144.354	149.957	156.648	161.314	171.217
119	139.149	145.461	151.084	157.8	162.481	172.418
120	140.233	146.567	152.211	158.95	163.648	173.617
121	141.315	147.674	153.338	160.1	164.814	174.816
122	142.398	148.779	154.464	161.25	165.98	176.014
123	143.48	149.885	155.589	162.398	167.144	177.212
124	144.562	150.989	156.714	163.546	168.308	178.408

Chi-square (χ^2) Distribution Table

125	145.643	152.094	157.839	164.694	169.471	179.604
126	146.724	153.198	158.962	165.841	170.634	180.799
127	147.805	154.302	160.086	166.987	171.796	181.993
128	148.885	155.405	161.209	168.133	172.957	183.186
129	149.965	156.508	162.331	169.278	174.118	184.379
130	151.045	157.61	163.453	170.423	175.278	185.571
131	152.125	158.712	164.575	171.567	176.438	186.762
132	153.204	159.814	165.696	172.711	177.597	187.953
133	154.283	160.915	166.816	173.854	178.755	189.142
134	155.361	162.016	167.936	174.996	179.913	190.331
135	156.44	163.116	169.056	176.138	181.07	191.52
136	157.518	164.216	170.175	177.28	182.226	192.707
137	158.595	165.316	171.294	178.421	183.382	193.894
138	159.673	166.415	172.412	179.561	184.538	195.08
139	160.75	167.514	173.53	180.701	185.693	196.266
140	161.827	168.613	174.648	181.84	186.847	197.451
141	162.904	169.711	175.765	182.979	188.001	198.635
142	163.98	170.809	176.882	184.118	189.154	199.819
143	165.056	171.907	177.998	185.256	190.306	201.002
144	166.132	173.004	179.114	186.393	191.458	202.184
145	167.207	174.101	180.229	187.53	192.61	203.366
146	168.283	175.198	181.344	188.666	193.761	204.547
147	169.358	176.294	182.459	189.802	194.912	205.727
148	170.432	177.39	183.573	190.938	196.062	206.907
149	171.507	178.485	184.687	192.073	197.211	208.086
150	172.581	179.581	185.8	193.208	198.36	209.265
151	173.655	180.676	186.914	194.342	199.509	210.443
152	174.729	181.77	188.026	195.476	200.657	211.62
153	175.803	182.865	189.139	196.609	201.804	212.797
154	176.876	183.959	190.251	197.742	202.951	213.973
155	177.949	185.052	191.362	198.874	204.098	215.149

Chi-square (χ^2) Distribution Table

156	179.022	186.146	192.474	200.006	205.244	216.324
157	180.094	187.239	193.584	201.138	206.39	217.499
158	181.167	188.332	194.695	202.269	207.535	218.673
159	182.239	189.424	195.805	203.4	208.68	219.846
160	183.311	190.516	196.915	204.53	209.824	221.019
161	184.382	191.608	198.025	205.66	210.968	222.191
162	185.454	192.7	199.134	206.79	212.111	223.363
163	186.525	193.791	200.243	207.919	213.254	224.535
164	187.596	194.883	201.351	209.047	214.396	225.705
165	188.667	195.973	202.459	210.176	215.539	226.876
166	189.737	197.064	203.567	211.304	216.68	228.045
167	190.808	198.154	204.675	212.431	217.821	229.215
168	191.878	199.244	205.782	213.558	218.962	230.383
169	192.948	200.334	206.889	214.685	220.102	231.552
170	194.017	201.423	207.995	215.812	221.242	232.719
171	195.087	202.513	209.102	216.938	222.382	233.887
172	196.156	203.602	210.208	218.063	223.521	235.053
173	197.225	204.69	211.313	219.189	224.66	236.22
174	198.294	205.779	212.419	220.314	225.798	237.385
175	199.363	206.867	213.524	221.438	226.936	238.551
176	200.432	207.955	214.628	222.563	228.074	239.716
177	201.5	209.042	215.733	223.687	229.211	240.88
178	202.568	210.13	216.837	224.81	230.347	242.044
179	203.636	211.217	217.941	225.933	231.484	243.207
180	204.704	212.304	219.044	227.056	232.62	244.37
181	205.771	213.391	220.148	228.179	233.755	245.533
182	206.839	214.477	221.251	229.301	234.891	246.695
183	207.906	215.563	222.353	230.423	236.026	247.857
184	208.973	216.649	223.456	231.544	237.16	249.018
185	210.04	217.735	224.558	232.665	238.294	250.179
186	211.106	218.82	225.66	233.786	239.428	251.339
187	212.173	219.906	226.761	234.907	240.561	252.499

Chi-square (χ^2) Distribution Table

188	213.239	220.991	227.863	236.027	241.694	253.659
189	214.305	222.076	228.964	237.147	242.827	254.818
190	215.371	223.16	230.064	238.266	243.959	255.976
191	216.437	224.245	231.165	239.386	245.091	257.135
192	217.502	225.329	232.265	240.505	246.223	258.292
193	218.568	226.413	233.365	241.623	247.354	259.45
194	219.633	227.496	234.465	242.742	248.485	260.607
195	220.698	228.58	235.564	243.86	249.616	261.763
196	221.763	229.663	236.664	244.977	250.746	262.92
197	222.828	230.746	237.763	246.095	251.876	264.075
198	223.892	231.829	238.861	247.212	253.006	265.231
199	224.957	232.912	239.96	248.329	254.135	266.386
200	226.021	233.994	241.058	249.445	255.264	267.541

Standard Normal Distribution Table for Z = 0.00 to 3.59

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981

Standard Normal Distribution Table for Z = 0.00 to 3.59

2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

Standard Normal Distribution Table for Z = -3.59 to 0.00

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183

Standard Normal Distribution Table for Z = -3.59 to 0.00

-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002

ANNAMACHARYA INSTITUTE OF TECHNOLOGY & SCIENCES:: TIRUPATI

AUTONOMOUS

AK 20 Regulations

Year : I B.Tech **Semester** : II **Branch of Study :** CSE,AI&DS,CIC,AI &ML

Subject Code:20ABS9911	**Subject Name:** Probability and Statistics	L T P **3 0 0**	Credits 3

Course Outcomes:

1. Interpret the association of characteristics and through correlation and regression tools.
2. Make use of the concepts of probability and their applications.
3. Apply discrete and continuous probability distributions.
4. Design the components of a classical hypothesis test for large sample.
5. Design the components of a classical hypothesis test for small samples.

UNIT -1

Measures of Central Tendency, Measures of Dispersion, Correlation,Regression

10 marks

1.Compute the arithmetic mean of the following data:

Roll No.s	1	2	3	4	5	6	7	8	9	10
Marks(x)	40	50	55	78	58	60	73	35	43	48

2: From the following data find the mean profits

Profit per shop(Rs)	100-200	200-300	300-400	400-500	500-600	600-700	700-800
Number of shops	10	18	20	26	30	28	18

3:Compute the arithmetic mean of the following by direct and short -cut methods both:

Class	20-30	30-40	40-50	50-60	60-70
Frequency	8	26	30	20	16

4:Calculate the arithmetic mean of the following distribution

Variate	6	7	8	9	10	11	12
Frequency	20	43	57	61	72	45	39

5: Calculate the median from the following data

Marks	10-25	25-40	40-55	55-70	70-85	85-100
Frequency	6	20	44	26	3	1

6: Calculate the median from the following data

Wages in Rs.	0-10	10-20	20-30	30-40	40-50
No. of workers	22	38	46	35	20

7.Compute the mode of the following distribution:

Class	0-7	7-14	14-21	21-28	28-35	35-42	42-49
Frequency	19	25	36	72	51	43	28

8: Compute the geometric mean of the following distribution:

Wheat (kg)	7.5-10.5	10.5-13.5	13.5-16.5	16.5-19.5	19.5-22.5	22.5-25.5	25.5-28.5
No. of farms	5	9	19	23	7	4	1

9_:Claculate coefficient of correlation from the following data

X	12	9	8	10	11	13	7
Y	14	8	6	9	11	12	3

10._ Find the coefficient of correlation between the two variables

X	50	50	55	60	65	65	65	60	60	60
Y	11	13	14	16	16	15	15	14	13	13

11.Find if there is any significant correlation between the heights and weights given below:

Height inches(x)	57	59	62	63	64	65	55	58	57
Weight in lbs(y)	113	117	126	126	130	129	111	116	112

12. Claculate coefficient of correlation between age of cars and annual maintenance cost and comment:

Age of cars in years(x)	2	4	6	7	8	10	12
Annual maintenance cost in Rs.(y)	1600	1500	1800	1900	1700	2100	2000

13. Problem 1 :Claculate Karl Pearson's correlation coefficient for the following paired data:

X	28	41	40	38	35	33	40	32	36	33
Y	23	34	33	34	30	26	28	31	36	38

14.The following table gives the distribution of the total population and those who are totally and partially blind among them. Find out if there is any relation between age and blindness

Age	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No.ofpersons '(000)	100	60	40	36	24	11	6	3

Blind	55	40	40	40	36	22	18	15

.15. Following are the ranks obtained by 10 students in two subjects, Statistics and Mathematics. To what extent the knowledge of the students in two subjects is related?

Statistics	1	2	3	4	5	6	7	8	9	10
Mathematics	2	4	1	5	3	9	7	10	6	8

16. The ranks of 16 students in Mathematics and Statistics are as follows:
(1,1) (2, 10) (3, 3) (4, 4) (5, 5) (6, 7) (7, 2) (8, 6) (9, 8) (10, 11) (11, 15) (12, 9) (13, 14) (14, 12) (15, 16) (16, 13) . Calculate the rank correlation coefficient for proficiencies of this group in Mathematics and Statistics

17. Ten competitors in a musical test were ranked by the three judges A, B and C in the following order.

Ranks by A	1	6	5	10	3	2	4	9	7	8
Ranks by B	3	5	8	4	7	10	2	1	6	9
Ranks by C	6	4	9	8	1	2	3	10	5	7

Using rank correlation method, discuss which pair of judges has the nearest approach to common likings in music.

18. A sample of 12 fathers and their elder sons gave the following data about their elder sons. Calculate the rank correlation coefficient

X	65	63	67	64	68	62	70	66	68	67	69	71
Y	68	66	68	65	69	66	68	65	71	67	68	70

19.Find the standard deviation of the following data (use the step deviation method):

Wages (Rs.)	125-175	175-225	225-275	275-325	325-375	375-425	425-475	475-525	525-575
no. of workers	2	22	19	14	3	4	6	1	1

20.Goals scored by two teams A and B in foot ball season are as follows.

Number of goals scored in match	Number of matches	
	Team A	Team B
0	24	25
1	9	9
2	8	6
3	5	5
4	4	5

By calculating the standard deviations in each case find which team be consider more consistent

21.The scores of two cricketers A and B in 10 innings are given here. Find who is a better and who is more a consistent player.

Section of A x_i	40	25	19	80	38	8	67	121	66	76
Section of A y_i	28	70	31	0	14	111	66	31	25	4

22. Calculate Karl Pearson's coefficient of skewness for the following data: 25, 15, 23, 40, 27, 25, 23, 25, 20.

23. Calculate Karl Pearson's coefficient of skewness for the following data:

Variable	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	2	5	7	13	21	16	8	3

24. From the following distribution, calculate (*i*) First 4 moment about the mean (*ii*) Skewness based on moments (*iii*) Kurtosis

Income (Rs)	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

2Marks

Problem 1: According to the census of 1991, following are the population figure, in thousands, of 10 cities : 1400, 1250, 1670, 1800, 700, 650, 570, 488, 2100, 1700. Find the median

2: Find the mode of the following salaries:
(i) 850, 750, 600, 825, 850, 725, 600, 850, 640, 530
(ii) 40, 45, 48, 57, 78

3.Given n=10, $\sigma_x = 5.4$, $\sigma_y = 6.2$ and sum of the product of deviations from the mean of X and Y is 66 find the correlation co-efficient

4. Find the variance and standard deviation of the following data: 5, 12, 3, 18, 6, 8, 2, 10.

5.Find the variance and standard deviation of the following data: 45, 60, 62, 60, 50, 65, 58, 68,44, 48.

6.Calculate the standard deviation for the following distribution:

x_i	4	8	11	17	20	24	32
f_i	3	5	9	5	4	3	1

7. Calculate the mean and standard deviation for the following frequency distribution:

x_i	6	10	14	18	24	28	30
f_i	2	4	7	12	8	4	3

Compute the range for the following observation 15, 20, 25, 25, 30, 35

8. The following table gives the daily sales (Rs.) of two firms A and B for five days.

Firm A	5050	5025	4950	4835	5140
Firm B	4900	3100	2200	1800	13000

Calculate the mean deviation of the variates 40, 62, 54, 68, 76 from A.M

9. Find the mean deviation from the mean for the following data: 38, 70, 48, 40, 42, 55, 63, 46, 54, 44.

10.Find the mean deviation about the mean for the following data

x_i	2	5	7	8	10	35
f_i	6	8	10	6	8	2

11.Find the mean deviation about the median for the following data

x_i	6	9	3	12	15	13	21	22
f_i	4	5	3	2	5	4	4	3

12.The following table gives the sales of 100 companies. Find the mean deviation from the mean.

Sales in thousands	40-50	50-60	60-70	70-80	80-90	90-100
,Number of companies	5	15	25	30	20	5

13.Find the mean deviation about the mean for the following data

Classes	0-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800
Freq.	4	8	9	10	7	5	4	3

:

14.Find the mean deviation from median for the following data

Age of workers	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No. of workers	120	125	175	160	150	140	100	30

15.Write the Types of (i) correlation (ii) Regression.

16. Explain Positive and Negative Correlation

17. Explain Simple and Multiple Correlation

18. Explain Partial and Total Correlation

19. Explain Linear and Non-linear Correlation

20.Write the formula for Regression lines (i) Y on X (ii) X on Y

21.Define Co-Variance.

22.Define Skewness

23.Define kurtosis

UNIT-II & III

PROBABILITY

10 Marks

1. A class consists of 6 girls and 10 boys. If a committee of 3 is chosen at random from the class, find the probability that (a) 3 boys are selected,(b) exactly two girls are selected.

2. A and B throw alternately with a pair of ordinary dice. A wins if he throws 6 before B throws 7 and B wins if he throws 7 before A throws 6.If A begins , show that his chance of winning is 30/61.

3. A,B and C in order toss a coin. The first one to toss head wins the game. What are the probabilities of winning, assuming that the game may continue indefinitely.

4. Two cards are selected at random from 10 cards numbered 1 to 10.Find the probability that the sum is even if (a)The two cards are drawn together. (b)The two cards are drawn one after other with replacement.

5. A box contains n tickets marked 1 through n. Two tickets are drawn in succession without replacement. Determine the probability that the number on the tickets are consecutive integers.

6. Determine the probability for each of the following events: A non-defective bolt will be found if out of 600 bolts already examined 12 were defective.

7. What is the probability of picking an ace and a king from a 52 cards deck?

8. Out of 15 items 4 are not in good condition 4 are selected at random. Find the probability that (a) All are not good (b) Two are not good

9. Five persons in a group 20 are engineers. If three persons are selected at random, determine the probability that all engineers and the probability that at least one being an engineer.

10. Three students A,B,C are in running race. A and B have the same probability of winning and each is twice as likely to win as C. Find the probability that B or C wins.

11. A class has 10 boys and 5 girls. Three students are selected at random one after another. Find the probability that (a) first two are boys and third is girl (b) first and third are of same sex and the second is of opposite sex.

12. Two aeroplanes bomb a target in succession. The probability of each correctly scoring a hit is 0.3 and 0.2 respectively. The second will bomb only if the first misses the target. Find the probability that (a)target is hit (b) both fails to score hits.

13. Box A contains 5 red and 3 white marbles and box B contains 2 red and 6 white marbles. If a marble is drawn from each box, what is the probability that they are both of same colour.

14. Two marbles are drawn in succession from a box containing 10 red, 30 white,20 blue and 15 orange marbles, with replacement being made after each draw. Find the probability that (a) both are white (b) first is red and second is white.

15. A can hit a target 3 times in 5 shots, B hits target in 5 shots ,C hits target 3 times in 4 shots. Find the probability of the target being hit when all of them try.

16. Two dice are thrown. Let A be the event that the sum of the points on the faces is 9.Let B be the event that at least one number is 6.Find

(a) P(A∩B) (b) P(AUB) (c) P($A^c U \boldsymbol{B}^c$)

17. If the probability that a communication system will have high fidelity is 0.81 and the probability that it will have high fidelity and selectivity is 0.81.What is the probability that a system with high fidelity will also have high selectivity?

18. Suppose 5 men out of 100 and 25 women out of 10,000 are colour blind. A colour blind person is choosen at random. What is the probability of the person being a male(Assume male and female to be in equal numbers)?

19. In a bolt factory machines A,B,C manufacture 20%,30% and 50% of the total of their output and 6%,3% and 2% are defective. A bolt is drawn at random and found to be defective. Find the probabilities that it is manufactured from (a)Machine A. (b)Machine B. (c)Machine C.

20. Of the three men,the chances that a politician,a business man or an academician will be appointed as a vice-chancellor (V.C) of a University are 0.5,0.3,0.2 respectively.Probability that research is promoted by these persons if they are appointed as V.C are 0.3,0.7,0.8 respectively.

(a) Determine the probability that research is promoted

(b) If research is promoted, what is the probability that V.C is an academician?

DISCRETE PROBABILITY DISTRIBUTION

22. Two dice are thrown. Let X assign to each point (a,b) in S the maximum of its numbers i.e., X(a,b)=max.(a,b). Find the probability distribution. X is a random variable with X(s) = {1,2,3,4,5,6}. Also find the mean and variance of the distribution.
(OR)

A random variable X has the following distribution?

X	1	2	3	4	5	6
P(x)	1/36	3/36	5/36	7/36	9/36	11/36

Find (a) the mean (b) variance (c) $P(1<X<6)$

23. A random variable X has the following probability function:

x	0	1	2	3	4	5	6	7
P(x)	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

(a) Determine K (b) Evaluate $P(X<6)$,$P(X\geq6)$,$P(0<X<5)$ and $P(0\leq X\leq4)$ (c) if $P(X\leq K)$ >1/2, find the minimum value of K and, (d) Determine the distribution of X (e) Mean (f) Variance.

24. The probability density function of a variate X is

X	0	1	2	3	4	5	6
P(X)	K	3k	5k	7k	9k	11k	13k

(a) Find K (b) Find $P(X<4)$, $P(X\geq5)$, $P(3<X\leq6)$ (c) what will be the minimum value of k so that $P(X\leq2)>0.3$?

25. A random variables X has the following probability function

x_i	-3	-2	-1	0	1	2	3
$P(x_i)$	K	0.1	K	0.2	2K	0.4	2K

Find (a) K (b) Mean (c) Variance

26. From a lot of 10 items containing 3 defectives, a sample of 4 items is drawn at random.Let the random variable X denote the number of defective items in the sample. Find the probability distribution of X when the sample is drawn without replacement.

27. Let X denote the minimum of the two numbers that appear when a pair of fair dice is thrown once. Determine the

(a) Discrete probability distribution (b) Expectation (c) Variance

28. A sample of 4 items is selected at random from a box containing 12 items of which 5 are defective. Find the expected number E of defective items.

29. A fair coin is tossed until a head or five tails occurs. Find the expected number E of tosses of the coin.

30. Find the mean of the probability distribution of the number of heads obtained in three flips of a balanced coin.

CONTINUOUS PROBABILITY DISTRIBUTION

1. If the probability density of a random variable is given by

$$f(x) = \begin{cases} k(1 - x^2), & for\ 0 < x < 1 \\ 0, & otherwise \end{cases}$$

2. The probability density f(x) of a continuous random variable is given by

$f(x) = ce^{-|x|}, -\infty < x < \infty$. Show that c=1/2 and find that the mean and variance of the distribution. Also find the probability that the variate lies between 0 and 4.

3. Probability density function of a random variable X is $f(x) = \begin{cases} \frac{1}{2} sinx\ , & for\ 0 \le x \le \pi \\ 0, elsewhere \end{cases}$.

   Find the mean,mode and median of the distribution and also find the probability between 0 and $\frac{\pi}{2}$?

4. A continuous random variable X has the distribution function

$f(x) = \begin{cases} 0, if\ x \le 1 \\ k(x-1)^4, if\ 1 < x \le 3 \\ 1, if\ x > 3 \end{cases}$. Determine (i) f(x) (ii) k (iii) Mean

5. If X is a continuous random variable and Y = aX + b, prove that E(Y)=a E(X) + b and V(Y) = a^2. V(X),where V stands for variance and a, b are constants?
6. If X is a continuous random variable and k is a constant, the prove that (i) Var(X+k) = Var(X) (ii) Var(kX) = k^2 Var(X)?
7. For the continuous random variable X whose probability density function is given by

$f(x) = \begin{cases} cx(2-x), if\ 0 \le x \le 2 \\ 0, otherwise \end{cases}$ where c is constant. Find c, Mean and Variance of X?

8. The daily consumption of electric power (in millions of kW-hours) is a random variable having the probability density function $f(x) = \begin{cases} \frac{1}{9} xe^{-x/3}, x > 0 \\ 0, x \le 0 \end{cases}$. If the total production is 12 million kW-hours, determines the probability that there is power cut on any given day.
9. The density function of a random variable X is $f(x) = \begin{cases} e^{-x}, x \ge 0 \\ 0, otherwise \end{cases}$. Find E(X), E($X^2$), Var(X).
10. The cumulative distribution function for a continuous random variable X is

$F(x) = \begin{cases} 1 - e^{-2x}, x \ge 0 \\ 0, x < 0 \end{cases}$. Find (i) the density function f(x) (ii) Mean (iii) Variance of the density function.

BINOMIAL DISTRIBUTION

1. 10 coins are thrown simultaneously. Find the probability of getting at least (i) 7 heads (ii) 6 heads
2. 2 dice are thrown 5 times. Find the probability of getting 7 as sum (i) at least once (ii) two times (iii) $p(1<X<5)$
3. In 256 sets of 12 tosses of a coin, how many cases one can expect 8 heads and 4 tails
4. Out of 800 families with 4 children each, how many families would be expected to have (a) 2 boys and 2 girls (b) at least one boy (c) no girl (d) at most 2 girls? Assume equal probabilities for boys and girls
5. In a binomial distribution consisting of 5 independent trails, probabilities of 1 and 2 success are 0.4096 and 0.2048 respectively. Find the parameter p of the distribution.
6. Fit a binomial distribution to the following data

x	0	1	2	3	4	5
f	2	14	20	34	22	8

7. The mean of binomial distribution is 3 and variance is 9/4. Find (i) the value of n (ii) $p(x \geq 7)$ (iii) $p(1 \leq X < 6)$
8. The probability that the life of a bulb is 100 days is 0.05. find the probability that out of 6 bulbs (i) at least one (ii) greater than 4 (iii) none, will be having a life of 100 days

POISSON DISTRIBUTION

1. A car-hire firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as a poisson distribution with mean 1.5. Calculate the proportion of days (a) on which there is no demand (b) on which demand is refused.
2. A hospital switch board receives an average of 4 emergency calls in a 10 minute interval. What is the probability that (a) there are at most 2 emergency calls in a 10 minute interval (b) there are exactly 3 emergency calls in a 10 minute interval.
3. If a random variable has a poisson distribution such that P(1)=P(2), find (a) mean of the distribution (b) P(4) (c) $P(x \geq 1)$ (d) $P(1<x<4)$.
4. The average number of phone calls / minute coming into a switch board between 2 p.m and 4 p.m is 2.5. Determine the probability that during one particular minute there will be (a) 4 or fewer (b) more than 6 calls.

5. Average number of accidents on any day on a national highway is 1.8. Determine the probability that the number of accidents are (a) at least one (b) at most one
6. If X is a poisson variate such that P(x=0)=P(x=1),find P(x=0) and using recurrence formula find the probabilities at x=1,2,3,4 and 5.
7. If the variance of a poisson variate is 3,then find the probability that (a) x=0 (b) $0<x\leq 3$ (c) $1\leq x<4$.
8. If X is a poisson variate such that 3P(x=4)=1/2 P(x=2) + P(x=0), find (a) the mean of x (b) $P(x\leq 2)$
9. Wireless sets are manufactured with 25 soldered joints each. On the average 1 joint in 500 is defective. How many sets can be expected to be free from defective joints in a consignment of 10,000 sets.

NORMAL DISTRIBUTION

10. If X is a normal variate with mean 30 and standard deviation 5. Find the probabilities that (a) $26\leq X\leq 40$ (b) $X\geq 45$
11. In a Normal distribution, 7% of the items are under 35 and 89% are under 63. Determine the mean and variance of the distribution.

(OR)

Find the mean and standard deviation of a normal distribution in which 7% of items are under 35 and 89% are under 63.

12. The mean and standard deviation of the marks obtained by 1000 students in an examination are respectively 34.5 and 16.5. Assuming the normality of the distribution, find the approximate number of students expected to obtain marks between 30 and 60.
13. The marks obtained in mathematics by 1000 students is normally distributed with mean 78% and standard deviation 11%.Determine
    (a) How many students got marks above 90%
    (b) What was the highest mark obtained by the lowest 10% of the students
    (c) Within what limits did the middle of 90% of the students lie.
14. Suppose the weights of 800 male students are normally distributed with mean μ=140 pounds and standard deviation 10 pounds. Find the number of students whose weights are (a) between 138 and 148 pounds (b) more than 152 pounds.

15. The marks obtained in statistics in a certain examination found to be normally distributed. If 15% of the students $\geq$ 60 marks,40% < 30 marks, find the mean and standard deviation.

16. A sales tax officer has reported that the average sales of the 500 business that he has to deal with during a year is Rs. 36,000 with a standard deviation of 10,000. Assuming that the sales in these business are normally distributed,find
    (a) the number of business as the sales of which are Rs.40,000.
    (b) the percentage of business the sales of which are likely to range between Rs.30,000 and Rs.40,000.

17. If the masses of 300 students are normally distributed with mean 68 kgs and standard deviation 3kgs, how many students have masses
    (a) greater than 72 kg
    (b) less than or equal to 64 kg
    (c) between 65 and 71 kg inclusive.

18. Given that the mean height of students in a class is 158cms with standard deviation of 20cms. Find how many students heights lie between 150cms and 170cms, if there are 100 students in the class.

2marks

1. A box contains n tickets marked 1 through n. Two tickets are drawn in succession without replacement. Determine the probability that the number on the tickets are consecutive integers.

2. Determine the probability for each of the following events: A non-defective bolt will be found if out of 600 bolts already examined 12 were defective.

3. What is the probability of picking an ace and a king from a 52 cards deck?

4. Out of 15 items 4 are not in good condition 4 are selected at random. Find the probability that (a) All are not good (b) Two are not good.

5. Let X denote the number of heads in a single toss of 4 fair coins. Determine (a) $P(X<2)$ (b) $P(1<X\leq 3)$.

6. The mean and variance of a binomial distribution are 4 and 4/3 respectively. Find $p(x \geq 1)$.

7. Using recurrence formula find the probabilities when x=0,1,2,3,4 and 5;if the mean of poisson distribution is 3.

8. Find the probability of getting an even number 3 or 4 or 5 times in throwing 10 dice. Using binomial distribution.

9. Determine the probability that getting an even number on face 3 to 5 times in throwing 10 dice together.

UNIT-IV

TEST OF HYPOTHESIS

1.A sample of 64 students have a mean weight of 70 kgs. Can this be regarded as a sample from a population with mean weight 56 kgs and standard deviation 25kgs.

2.A sample of 900 members has a mean of 3.4 cms and S.D 2.61 cms. Is this sample has been taken from a large population of mean 3.25 cm and S.D 2.61 cms. If the population is normal and its mean is unknown find the 95% fiducial limits of true mean.

3.A sample of 400 items is taken from a population whose standard deviation is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38. Also calculate 95% confidence interval for the population.

4.An ambulance service claims that it takes on the average less than 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the claim at 0.05 level significance.

5.It is claimed that a random sample of 49 tyres has a mean life of 15200 km. This sample was drawn from a population whose mean is 15150 kms and a standard deviation of 1200km. Test the significance at 0.05 level.

6.The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches.

7.A researcher wants to the intelligence of students in a school. He selected two groups of students. In the first group there 150 students having mean IQ of 75 with a S.D. Of 15 in the second group there are 250 students having men IQ of 70 with S.D. of 20.

8.The mean life of a sample of 10 electric bulbs (or motors) was found to be 1456 hours with S.D. Of 423 hours. A second sample of 17 bulbs chosen from a different batch showed a mean life 1280 hours with S.D. of 398 hours. Is there a significant difference between the means of two batches.

9.The mean height of 50 male students who participated in sports is 68.2 inches with a S.D of 2.5. The mean height of 50 male students who have not participated in sport is 67.2 inches with a S.D of 2.8. Test the hypothesis that the height of students who participated in sports is more than the students who have not participated in sports.

10.A manufacturer claimed that at least 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faculty. Test his claim at 5% level of significance.

11.In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance.

12.In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?

13.Experience had shown that 20% of a manufactured product is of the top quality. In one day's production of 400 articles only 50 are of top quality. Test the hypothesis at 0.05 level.

14.In a sample of 500 from a village in Rajasthan, 280 are found to be wheat eaters and the rest rice eaters. Can we assume that the both articles are equally popular.

15.20 people were attacked by a disease and only 18 survived. Will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favour of the hypothesis that is more at 5% level.

16.A random sample of 500 apples was taken from a large consignment of 60 were found to be bad, obtain the 98% confidence limits for the percentage number of bad apples in the consignment.

17.Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and women in favour of the proposal are same, at 5% level.

18.On the basis of their total scores, 200 candidates of a civil service examination are divided into two groups, the upper 30% and the remaining 70%. Consider the first question of the examination. Among the first group, 40 had the correct answer, whereas among the second group, 80 had the correct answer. On the basis of these results, can one conclude that the first question is not good at discriminating ability of the type being examined here?

19.In two large populations, there are 30%, and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations.

20.In a random sample of 1000 persons from town A, 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as the proportion of wheat consumers is concerned?

21.Before an increase on excise duty on tea 500 people out of a sample of 900 found to have the habit of having tea. After an increase on excise duty 250 are have the habit of having tea among 1100. Is there any decrease in the consumption of tea. Test at 5% level.

2Marks

1Find the value of the finite population correction factor for n=10 and N=100.

2.A random sample of size 81 was taken whose variance is 20.25 and mean is 32, construct 98% confidence interval .

3.In a random sample of 100 packages shipped by air freight 13 had some damage. Construct 95% confidence interval for the true proportion of damage package.

4.Define (i)Sample (ii) Population

5.Define (i) Large Sample (ii) Small Sample.

6.Define Sample Variance.

7.Define Central Limit Theorem.

8.Define Standard Error.

9.Write the confidence limits for (i) single mean (ii) Difference of Means (iii)Single Proportion (iv) Difference of Proportion.

10. Define (i)Null Hypothesis (ii) Alternative Hypothesis

11.Define (I) Type I error (ii) Type II error.

12.Define (i)critical region (ii) Acceptance Region.

13.Define (i) Two Tailed Test (ii)One-Tailed Test

UNIT-V

TEST OF SIGNIFICANCE (small samples)

1>Find (a) $P(t < 2.365)$ when $v = 7$

(b) $P(t > 1.318)$ when $v = 24$

(c) $P(-1.356 < t < 2.179)$ when $v = 12$

(d) $P(t > -2.567)$ when $v = 17$

2> A random of sample of size 25 from a normal population has the mean$\overline{x}$ =4.75 and the standard deviation S = 8.4 .Does this information tend to support or refute the claim that mean of the population is μ = 42.5?

3> Ten bearings made by a certain process have a mean diameter of 0.5060cm with a standard deviation of 0.0040cm. Assuming that the data may be taken as a random sample from a normal distribution, construct a 95% confidence interval for the actual average diameter of the bearings?

4> A sample of size 10 was taken from a population S.D of sample is 0.03.Find maximum error with 99% confidence.

5>A sample of 11 rats from a central population had an average blood viscosity of 3.92 with a S.D of0.61. Estimate the 95% confidence limits for the mean blood viscosity of the population.

6>A mechanist is making engine parts with axle diameters of 0.700inch.A random sample of 10 parts shows a mean diameter of 0.742inch with a S.D of 0.040inch. Compute the statistics you would use to test whether the work is meeting the specification at 0.05 level of significance.

7>A sample of 26 bulbs gives a mean life of 990 hours with a S.D of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours. Is the sample not upto the standard .

8> The average breaking strength of the steel rods is specified to be 18.5 thousand pounds. To test this sample of 14 rods were tested .The mean and standard deviations obtained were 17.85 and 1.955 respectively. Is the result of experiment significant?

9>> A random of sample of size 16 values from a normal population showed a mean of 53 and a sum of squares of deviations from the mean equals to 150. Can this sample be regarded as taken from population having 50 as mean? Obtain 95% confidence limits of the mean of the population.

10>A random of sample of six steel beams has a mean compressive strength of 58,392 p.s.i (pounds per square inch) with a standard deviation of 648 p.s.i. Use this information and the level of significance α = 0.05 to test whether the true average compressive strength of the steel from which this sample came is 58,000 p.s.i. Assume normality.

11> A random of sample of 10 boys had the following I.Q’s : 70,120,110,101,88,83,95,98,107 and 100.
(a)Do these data support the assumption of a population mean I.Q of 100
(b)Find the reasonable range in which most of the mean I.Q values of samples of 10 boys lie.

11>Producer of ‘gutkha’, claims that the nicotine content in his gutkha on the average is 1.83mg.Can this claim accepted if a random sample of 8 gutkha of this type have the nicotine contents of 2.0,1.7,2.1,1.9,2.2,2.1,2.0,1.6mg? Use a 0.05 level of significance.

12>Two horses A and B were tested according to the time (in sec) to run a particular track with the following results.

Horse A	28	30	32	33	33	29	34

Horse B	29	30	30	24	27	29	

Test whether two horses have the same running capacity.

13>To examine the hypothesis that the husbands are more intelligent than the wives, an investigator took a sample of 10 couples and administered them a test which measures the I.Q. The results are as follows:

Husbands	117	105	97	105	123	109	86	78	103	107
Wives	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with a reasonable test at the level of significance of 0.05.

14>To compare two kinds of bumper guards,6 of each kind were mounted on a car and then the car was run into a concrete wall. The following are the costs of repairs.

Guard 1	107	148	123	165	102	119
Guard 2	134	115	112	151	133	129

Use the 0.01 level of significance to test whether the difference between two sample mean is significant.

15>Scores obtained in a shooting competition by 110 soliders before and after intensive training are given below:

Before	67	24	57	55	63	54	56	68	33	43
After	70	38	58	58	56	67	68	75	42	38

Test whether the intensive training is useful at 0.05 level of significance.

16>The blood pressure of 5 women before and after intake of a certain drug are given below:

Before	110	120	125	132	125
After	120	118	125	136	121

Test whether tere is significant change in blood pressure at 1% level of significance.

17>Memory capacity of 10 students were tested before and after training. State whether the training was effective or not from the following scores.

Before training	12	14	11	8	7	10	3	0	5	6
After training	15	16	10	7	5	12	10	2	3	8

18>In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level whether the populations have the same variance.

19>The nicotine contents in milligrams in two samples of tobacco were found to be as follows:

Sample A	24	27	26	21	25	—

Sample B	27	30	28	31	22	36

Can it be said that the two samples have come from the same normal population?

20>The measurements of the output of the two units have given the following results. Assuming that both samples have been obtained from the normal populations at 10% significant level, test whether the two populations have the same variance.

Unit-A	14.1	10.1	14.7	13.7	14.0
Unit-B	14.0	14.5	13.7	12.7	14.1

21>In one sample of 10 observations, the sum of the squares of the deviations of the sample values from sample mean was 120 and in the other sample of 12 observations, it was 314. Test whether the difference is significant at 5% level?

22> Two independent samples of 8 and 7 items respectively had the following values of the variables

Sample I	9	11	13	11	16	10	12	14
SampleII	11	13	11	14	10	8	10	—

Do the estimates of the population variance differ significantly.

23>The number of automobile accidents per week in a certain community are as follows: 12,8,20,2,14,10,15,6,9,4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.

24>A die is thrown 264 times with the following results. Show that die is biased. [Given $\chi^2_{0.05}$ =11.07 for 5 d.f]

No .appeared on the die	1	2	3	4	5	6
Frequency	40	32	28	58	54	52

25> A pair of dice are thrown 360 times and the frequency of each sum is indicated as below:

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of the chi-square test at 0.05 level of significance?

26>4 coins were tossed 160 times and the following results were obtained:

No. of heads	0	1	2	3	4
Observed frequencies:	17	52	54	31	6

Under the assumption that coins are balanced, find the expected frequencies of 0,1,2,3 or 4 heads, and test the goodness of fit (α= 0.05).

27> Fit a poisson distribution to the following data and for its goodness of fit at level of significance 0.05?

x	0	1	2	3	4
f	419	352	154	56	19

28>The following table gives the classification of 100 workers according to gender and nature of work. Test whether the nature of work is independent of the gender of the worker.

	Stable	**Unstable**	**Total**
Males	40	20	60
Females	10	30	40
Total	50	50	100

29>Given the following contingency table for hair colour and eye colour. Find the value of χ^2 . Is there good association between the two?

		Hair colour			
		Fair	**Brown**	**Black**	**Total**
Eye colour	**Blue**	15	5	20	40
	Grey	20	10	20	50
	Brown	25	15	20	60
	Total	60	30	60	150

30> From the following data find whether there is any significant liking in the habit of taking soft drinks among the categories of employees.

	Employees		
Soft Drinks	**Clerks**	**Teachers**	**Officers**
Pepsi	10	25	65
Thumsup	15	30	65
Fanta	50	60	30

31> In an investigation on the machine performance, the following results are obtained:

	No . of units inspected	No . of defectives
Machine 1	375	17
Machine 2	450	22

Test whether there is any significant performance of two machines at $\alpha=0.05$.

32>A firm manufacturing rivets wants to limit variations in their length as much as possible. The lengths (in cms) of 10 rivets manufactured by a new process are

2.15	1.99	2.05	2.12	2.17
2.01	1.98	2.03	2.25	1.93

Examine whether the new process can be considered superior to the old if the old population has standard deviation 0.145 cm?

2Marks

1.Define Degrees of Freedom.

2.Define t- Distribution and uses.

3.Write properties of t-Distribution and Applications.

4.Define Chi-Square Distribution.

5. Write properties of Chi-Distribution and Applications.

6. Write properties of F-Distribution .

7. Write Test of significance of small samples.

8.write the conditions of validity of test of Goodness of Fit.

9.Write confidence or Fiducial limits for μ.