WEBINAR (ISM _REVISION)

INTRODUCTION TO STATISTICAL METHODS (ISM)

DR. MOHD RIYAZUDDIN

BITS Pilani

Pilani|Dubai|Goa|Hyderabad

PROBABILITY

SAMPLE SPACE AND EVENTS:

1. Random Experiment:

A random experiment is an experiment whose outcome or result is not unique and therefore cannot be predicted with certainty.

Ex: In tossing a coin, one is not sure whether a head or tail will occur.

Trail:

Each performance in a random experiment is called a trial.

Ex: Tossing a coin first time is first trail, second time is second trial.

3. Outcome:

The result of a trial in a random experiment is called an outcome. In coin tossing experiment getting head and tail are outcomes.

4. Sample Space:

The set of all possible outcomes of an experiment is called a sample space.

Ex: In throwing a die, {1,2,3,4,5,6}will form the sample space.

Discrete Sample Space:- A sample space is said to be a discrete sample space if it has finitely many or a countable infinity of elements.

Ex:- a) A sample space consists of finite no of elements.

-2000 students.

b) The sample space consists of countable infinity of elements -- the whole set of natural nos

Continuous Sample Space:-

If the elements of a sample space constitute a continuum – for example, all the points on a line, all the points on a line segment or all the points in plane – the sample space is said to be continuous .

Event:-

Every non empty subset of a sample space of a random experiment is called an event.

Ex: In throwing a die{1,2,3,4,5,6}

-- Sample Space

Event: A = {2,4.6} -- Set of even numbers

B = {1,3.5} -- Set of odd numbers

C= {1,2,3,4,5,6} -- Set of all elements

Mutually exclusive events:

- Events are said to be mutually exclusive if the happening of anyone of them prevents the happening of all the others i.e. if no two or more of them can happen simultaneously in the same trail.
- Ex: In tossing a coin the events Head turning up and tail turning up are mutually exclusive.

The classical probability concept:

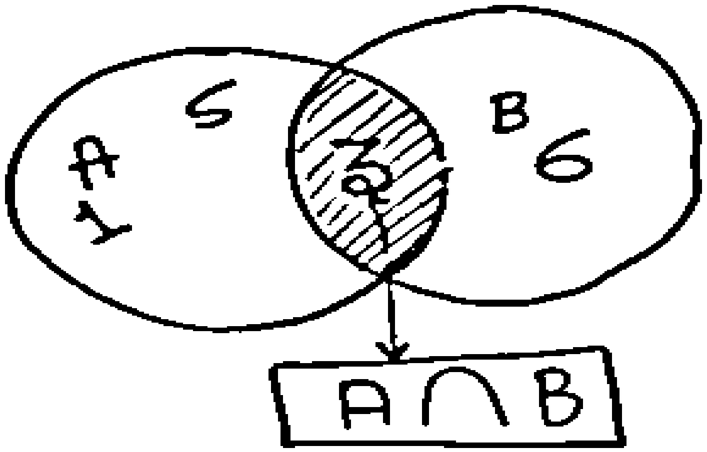
If there are n equally likely possibilities, of which one must occur and S are regard as favorable, or as a “Success”, then the probability of a “Success” is given by S/n.

Ex:-What is the probability of drawing an ace from a well shuffled deck of 52 Playing cards?

Solution:

There are S=4 aces among the n=52 cards, so we get s/n= 4/52 = 1/13.

Intersection



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

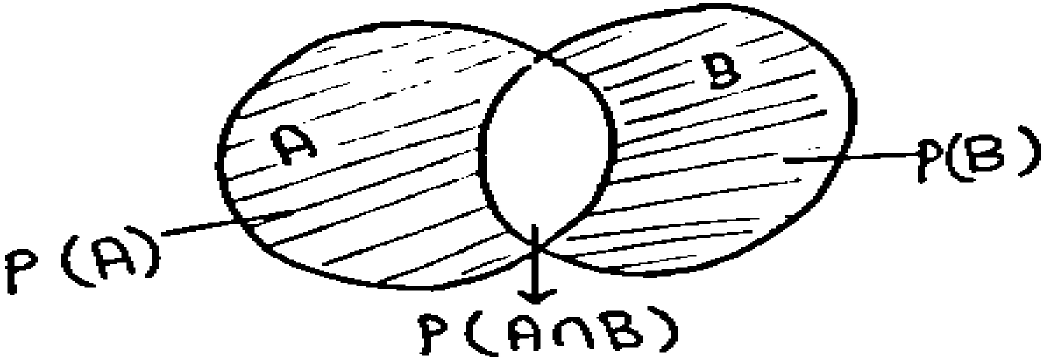
$A = \{1, 3, 5\}$ $\quad$ $A \cap B = \{3\}$

$B = \{3, 6\}$ $\quad$ $n(A \cap B) = 1$

$S = \{1, 3, 5, 6\}$ $\quad$ $n(S) = 4$

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{4}$$

Union $\boxed{A \cup B}$



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

$A = \{1, 2, 3, 4\}$ $B \{4, 5, 6, 7\}$

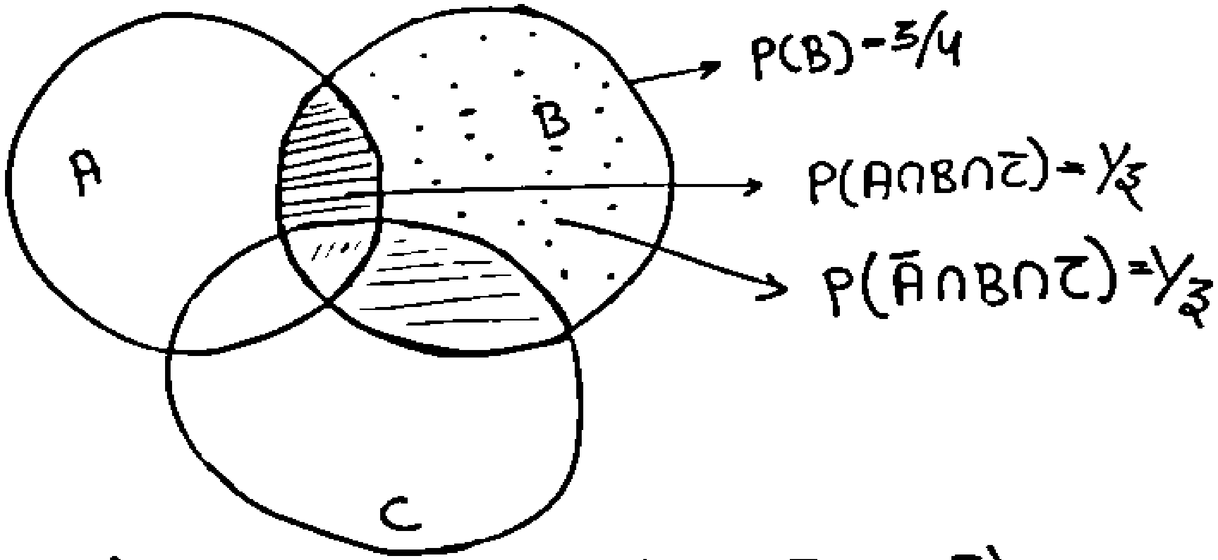
$n(A) = 4$, $n(B) = 4$, $n(S) = 7$, $n(A \cap B) = 1$

$S = \{1, 2, 3, 4, 5, 6, 7\}$, $A \cap B = \{4\}$ P

$$P(A \cup B) = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)} = \frac{4}{7} + \frac{4}{7} - \frac{1}{7} = \frac{4+4-1}{7} = \frac{7}{7} = 1$$

$P(B) = \frac{3}{4}$, $P(A \cap B \cap \bar{C}) = 1/3$, $P(\bar{A} \cap B \cap \bar{C}) = 1/3$

find the value of $P(B \cap C)$.



$$P(B \cap C) = P(B) - P(A \cap B \cap \bar{C}) - P(\bar{A} \cap B \cap \bar{C})$$

$$= \frac{3}{4} - \frac{1}{3} - \frac{1}{3}$$

Permutation:

The number of permutations of r objects taken from a set of "n" distinct object is

$nPr = n! / (n - r)!$

Problem

It repetitions are not allowed, how many 4 digit numbers can be formed from digits 1,2,3,4,5,6,7?

$$= \frac{7!}{(7-4)!} = \frac{7!}{3!} = 7 \times 6 \times 5 \times 4 = 840$$

Combination

The number of ways in which "r" objects can be selected from a set of n distinct objects is

$$nCr = \frac{n!}{r!\,(n-r)!}$$

Ex: In how many ways 3 students can be selected from 15 students.

$$15C_3 = \frac{15!}{3!(15-3)!} = \frac{15!}{3!(12)!} = 455 \text{ ways}$$

Ex: A bag contain 5 red balls, 8 blue balls and 11 white balls. Three balls are drawn together from the box. Find the probability that.

1) One is red, one is blue and one is white.
2) Two whites and one red.
3) Three white.

Solution:

There are $24C_3$ = 2024 equally likely ways of choosing 3 of 24 balls, so n=2024.

1. The number of possible cases

$5C_1 . 8C_1 . 11C_1 = 440.$

Required probability = s/n = 440/2024 = 55/253

2. No. of possible cases = $11C_2 . 5C_1 = 275$.

Required probability = 275/2024 = 25/184

3. No. of possible cases = $11C_3 = 165$

Required probability = 165/2024 = 15/184

This assignment is known as the random variable.

By a random variable we mean a real number X associated with the outcome of a random experiment.

Eg: - Suppose two coins are tossed simultaneously then the sample space is S = $\{HH, HT, TH, TT\}$. We will consider the random variable ,which is the number of heads (0, 1, 2)

Outcome	HH	HT	TH	TT
Value of X	2	1	1	0

Thus to each outcome `S' , there corresponds a real number X(s).

Note:- The real number is denoted by X(s) and it is defined for each $s \in S$.

Random variable is also known as stochastic variable or variable.

The random variables are denoted by upper case letters such as X , Y , Z and the values assumed by them are denoted by lower case letters with subscripts as x_1 , x_2 , x_3 ....

<u>Definition</u> <u>Random variables</u>

A random variable X on a sample space "S" is a function from S to the set of real numbers R, which assign a real number X (s) to each outcome "s" of S.

The function is given as $X : S \rightarrow R$

Example

1) If a coin is tossed , then sample space is

S = { H , T }

Here we consider the random variable

$$X(s) = \begin{cases} 1 & \text{if } s = H \\ 0 & \text{if } s = T \end{cases}$$

2) If a random experiment consists of rolling a die and reading the number of points on the up turned face

Sample space is { 1,2,3,4,5,6 }

Events which are considering is whether number of points is even or odd

We consider the random variable

$$X(s) = \begin{cases} 0 \text{ if } s \text{ is even} & s = 2,4,6 \\ 1 \text{ if } s \text{ is odd} & s = 1,3,5 \end{cases}$$

Types of random variables

There are two types of random variables

1) **Discrete random variable**

2) **Continuous random variable**

Discrete random Variable (Def)

A random variable X is said to be discrete random variable if its set of all possible outcomes (sample space) is countable . (Finite or an un-ending sequence with as many elements as there in whole numbers)

Continuous Random Variables (Def)

A random variable X is said to be continuous random variable if the sample space contains infinite numbers equal to the number of points on a line segment

OR

A random variable X is said to be continuous if it can assume all possible values between certain limits.

Eg:- Weight, height

Definition of Antiderivative: A function F is called an antiderivative of the function f if for every x in the domain of f

F'(x) = f(x) so, dy = f(x) dx

Integration is denoted by an integral sign $\int$.

$$y = \int f(x)dx = F(x) + C$$

Constant of Integration

Integrand

Variable of Integration

F'(x) also = f(x) (first derivative)

Basic Integration Formulas

$$\int 0\,dx = C$$

$$\int k\,dx = kx + C$$

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C$$

$$\int \cos x\,dx = \sin x + C$$

$$\int \sin x\,dx = -\cos x + C$$

$$\int \sec^2 x\,dx = \tan x + C$$

$$\int \sec x \tan x\,dx = \sec x + C$$

$$\int \csc^2 x\,dx = -\cot x + C$$

$$\int \csc x \cot x\,dx = -\csc x + C$$

Integrate

$$\int 3x\,dx \quad = \frac{3x^2}{2} + C$$

$$\int \frac{1}{x^3}dx = \int x^{-3}dx = \frac{x^{-2}}{-2} + C \quad = -\frac{1}{2x^2} + C$$

$$\int \sqrt{x}\,dx = \int x^{1/2}dx = \frac{x^{3/2}}{3/2} + C \quad = \frac{2x^{3/2}}{3} + C$$

$$\int 2\sin x\,dx = 2\int \sin x\,dx \quad = 2(-\cos x) + C = -2\cos x + C$$

$$\int 1dx \quad = x + C$$

$$\int (x+2)dx \quad = \frac{x^2}{2} + 2x + C$$

$$\int \left(3x^4 - 5x^2 + x\right)dx \quad = \frac{3x^5}{5} - \frac{5x^3}{3} + \frac{x^2}{2} + C$$

Find the general solution of the equation F'(x) = $\dfrac{1}{x^2}$ and find the particular solution given the point F(1) = 0.

$$F(x) = \int \frac{1}{x^2}dx = \int x^{-2}dx$$

$$= \frac{x^{-1}}{-1} + C = -\frac{1}{x} + C$$

$$\therefore y = -\frac{1}{x} + C$$

Now plug in (1,0) and solve for C.

0 = -1 + C

C = 1

Final answer.

$$\boxed{y = -\frac{1}{x} + 1}$$

$$\int 12e^{4x}\,dx = 12\frac{e^{4x}}{4} + C$$

$$= 3e^{4x} + C$$

$$z = \int \left(6x^2 + \frac{3}{x} \right) dx$$

$$= \int 6x^2 dx + \int \frac{3}{x} dx$$

$$= \frac{6x^3}{3} + 3\ln x + C$$

$$= 2x^3 + 3\ln x + C$$

$$I = \int_0^1 8xe^{-2x}dx$$

$$= 8\frac{e^{-2x}}{(2)^2}\left[-2x-1\right]_0^1$$

$$= 2e^{-2}\left[-2(1)-1\right] - 2e^{-0}\left[0-1\right]$$

$$= -6e^{-2} + 2 = 1.188$$

When deciding what to choose for u, remember L I P E T.

L - logarithmic function

I - inverse trig function

P - polynomial function

E - exponential function

T - trigonometry function

This is usually the preference order in which you would want to choose u.

Generalized integral.

$$\int uv\,dx = uv_1 - u'v_2 + u''v_3 - u'''v_4 \ldots$$

$$\int uv\,dx = u\int v\,dx - \frac{du}{dx}\iint v\,dx\,dx + \ldots\ldots$$

$$\int_0^\infty x^2 e^{-2x}\,dx$$

$$\left[x^2\left(\frac{e^{-2x}}{-2}\right) - 2x\left(\frac{e^{-2x}}{4}\right) + 2\left(\frac{e^{-2x}}{-8}\right)\right]_0^\infty$$

Problem (14): A continuous random variable has a probability density function,

$$f(x) = \begin{cases} xke^{-\lambda x} & ; \text{ for } x \geq 0 ; \lambda > 0 \\ 0 & , \text{ otherwise} \end{cases}$$

Determine (I) k (II) Mean

(III) Variance (IV) Standard Deviation.

Solution:

(I) $$\int_0^\infty f(x)\,dx = 1$$

$$\Rightarrow \int_0^\infty kxe^{-\lambda x} = 1 \qquad \Rightarrow k\int_0^\infty xe^{-\lambda x}dx = 1$$

$$\Rightarrow k\left[\left(x\left(\frac{e^{-\lambda x}}{-\lambda^2}\right)\right)_0^\infty - 1\left(\frac{e^{-\lambda x}}{\lambda^2}\right)_0^\infty\right] = 1$$

$$\Rightarrow k\left(-\frac{1}{\lambda^2}\left(e^{-\infty} - e^0\right)\right) = 1$$

$$\Rightarrow k = \lambda^2$$

(II) $$\text{Mean} = \int_0^\infty x\,k\,e^{-\lambda x}dx = k\int_0^\infty x^2e^{-\lambda x}\,dx$$

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

$$= k\left(-\frac{2}{\lambda^3}\left(e^{-\infty}-e^{-0}\right)\right)$$

$$= \frac{+2\lambda^2}{\lambda^3} = +\frac{2}{\lambda}$$

III) Variance $= \int_0^\infty x^3 k e^{-\lambda x}\, dx - (\text{mean})^2$

$$= k\left[x^3\left(\frac{e^{-\lambda x}}{-\lambda}\right)_0^\infty - 3x^2\left(\frac{e^{-\lambda x}}{\lambda^2}\right)_0^\infty + 6x\left(\frac{e^{-\lambda x}}{-\lambda^3}\right)\right.$$

$$\left. - 6\left(\frac{e^{-\lambda x}}{\lambda^4}\right)_0^\infty\right] - (\text{mean})^2$$

$$= k\left[-\frac{6}{\lambda^4}\left(e^{-\infty}-e^{0}\right)\right] - \frac{4}{\lambda^2}$$

$$= \frac{6\lambda^2}{\lambda^4} - \frac{4}{\lambda^2} = \frac{6-4}{\lambda^2} = \frac{2}{\lambda^2}$$

IV) Standard deviation $= \sqrt{V(x)}$

$$= \sqrt{2/\lambda^2} = \frac{\sqrt{2}}{\lambda}$$

Learning objectives

- Problems on Basic probability
- Conditional Probability
- Baye's theorem

Problem - 1

In a group there are 3 men and 2 women. Three persons are selected at random from this group. Find the probability that one men two women or two men and one women are selected.

Solution -1

Exhaustive number of cases are $^5c_3 = 10$ ways

Let A be the event of selecting one men and two women

Favourable cases for A are = $^3c_1 \cdot {}^2c_2 = 3$ ways

$$P(A) = \frac{\text{Favourable cases for A}}{\text{Exhaustive number of cases}} = 3/10$$

Solution -1

Let B be the event of selecting two men and one women

Favourable cases for B are = $^3c_2 \, . \, ^2c_1$ = 6 ways

$$P(B) = \frac{\textbf{Favourable cases for B}}{\textbf{Exhaustive number of cases}} = {}^{6}/_{10}$$

Required probability = **P (A U B)** = P(A) + P(B)

[A and B are disjoint events]

$= {}^{3}/_{10} + {}^{6}/_{10}$

$= {}^{9}/_{10}$

Problem - 2

If two dice are thrown, What is the probability that the sum is

(i) greater than 8

(ii) neither 7 nor 11.

Solution -2

When two dice are thrown the sample space contains 36 elements

S = { (1,1) , (1,2), (1,3), (1,4), (1,5), (1,6)

.....

.....

(6,1) , (6,2), (6,3), (6,4), (6,5), (6,6) }

Solution -2

(i) Let A be the event that the sum on the two dice

$P(A > 8) = P(A = 9) + P(A = 10) + P(A = 11) + P(A = 12)$

sum is 9 = { (3,6), (6,3), (4,5), (5,4) }

$$P(A = 9) = \frac{\text{Favourable cases for A}}{\text{Exhaustive number of cases}} = {}^{4}/_{36}$$

sum is 10 = { (4,6), (6,4), (5,5) }

$$P(A = 10) = \frac{\text{Favourable cases for A}}{\text{Exhaustive number of cases}} = {}^{3}/_{36}$$

sum is 11 = { (5,6), (6,5) }

$P(A = 11) = {}^{2}/_{36}$

sum is 12 = { (6,6) }

$P(A) = {}^{1}/_{36}$

Solution -2

P(A > 8) = P(A = 9) + P(A = 10) + P(A = 11) + P(A = 12)

$$= {}^{4}/_{36} + {}^{3}/_{36} + {}^{2}/_{36} + {}^{1}/_{36}$$

$$= {}^{10}/_{36}$$

(ii) **Let B denote the event of getting the sum of 7**

sum is 7 = { (1,6), (6,1), (2,5), (5,2), (3,4), (4,3) }

P(B) = ${}^{6}/_{36}$ = ${}^{1}/_{6}$

Let C denote the event of getting the sum of 11

sum is 11 = { (5,6), (6,5) }

P(C) = ${}^{2}/_{36}$

Solution -2

Required probability = $P(\bar{B} \cap \bar{C}) = P[(B \cup C)^c]$

$= 1 - P(B \cup C)$

$= 1 - [P(B) + P(C)]$ [A and B are disjoint events]

$= 1 - \frac{1}{6} - \frac{1}{18}$

$= \frac{7}{9}$

Problem - 3

The odds that person X speaks the truth are 3:2 and the odds that person Y Speaks the truth are 5:3. In what percentage of cases are they likely to contradict each on an identical point?

Solution -3

Let A = X speaks the truth

$\bar{A}$ = X tell a lie

Let B = Y speaks the truth

$\bar{B}$= Y tell a lie

$$P(A) = \frac{3}{3+2}\ ;\ P(\bar{A}) = \frac{2}{3+2}$$

$$P(B) = \frac{5}{5+3}\ ;\ P(\bar{B}) = \frac{3}{5+3}$$

The event C that X and Y contradict each other on an-identical point.

That can happen in two ways

(i) X speaks the truth and Y tell a lie, i.e, $A \cap \bar{B}$

(ii) X tell a lie and Y speaks the truth, i.e, $\bar{A} \cap \mathrm{B}$

Solution -3

the events $(A \cap \bar{B})$ and $(\bar{A} \cap \mathrm{B})$ mutually exclusive events

P(C) = P$(A \cap \bar{B})$ + P$(\bar{A} \cap \mathrm{B})$

= $P(A)\,.\,P(\bar{B}) + P(\bar{A})\,.\,P(B)$

[A and B are independent events]

$= \frac{3}{5}.\frac{3}{8} + \frac{2}{5}.\frac{5}{8}$

= 0.475

47.5 % of cases are they likely to contradict each on an identical point.

Problem - 4

A certain shop repairs both audio and video components. Let A denote the event that the next component brought in for repair is an audio component, and let B be the event that the next component is a compact disc player (so the event B is contained in A). Suppose that $P(A) = 0.6$ and P$(B) = 0.05$.

What is $P(^{B}/_{A})$?

Solution - 4

Let A denote the event that the next component brought in for repair is an audio component

and

Let B be the event that the next component is a compact disc player

P(A) = 0.6 and P(B) = 0.05

and given that $B \subseteq A$

From sets operations $A \cap B = B$

then P($A \cap B$) = P (B) = 0.05

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.05}{0.6} = 0.0833$$

Problem - 5

Two Vendors X and y are willing to submit their bids for the supply of a large computer system. There is a 50-50 chance for the vender X.

If vender X does not submit the bid, then Y will get the order with a probability $\frac{3}{4}$.

If vender X submits a bid, then Y will get the order with a probability $\frac{1}{4}$.

What is the probability that Y will get the order?

Solution - 5

Let A = Event that vender X submits the bid

$\bar{A}$ = Event that vender X not submits the bid

Let B = Event that vender Y will get the order

Given that $P(A) = {}^{1}/_{2}$ and $P(\bar{A}) = {}^{1}/_{2}$

$P\left({}^{B}/_{\bar{A}}\right) = {}^{3}/_{4}$ and $P\left({}^{B}/_{A}\right) = {}^{1}/_{4}$

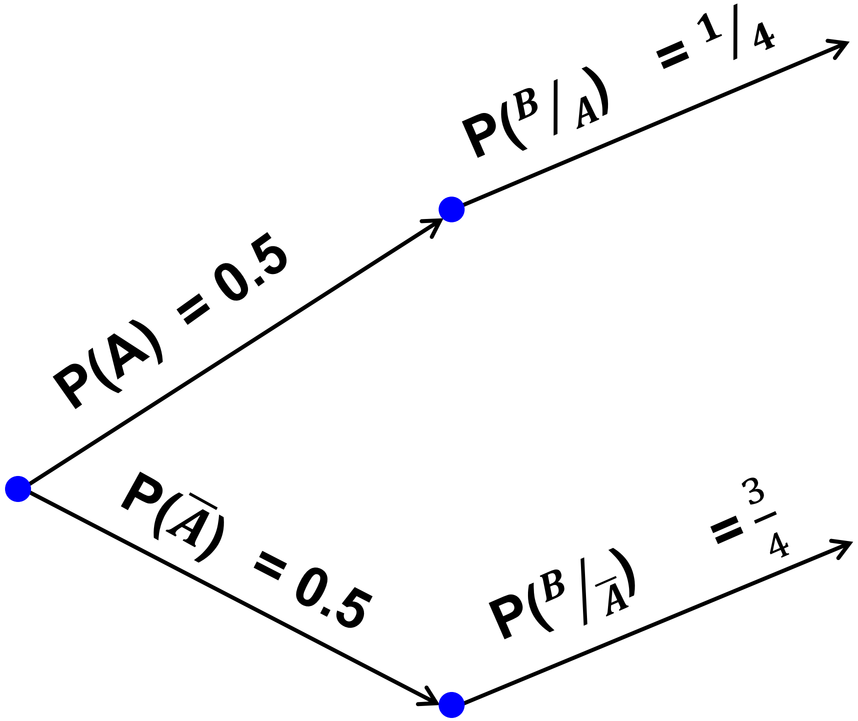
By total probability

$P(B) = P[(A \cap B) \cup (\bar{A} \cap B)] = P(A \cap B) + P(\bar{A} \cap B)$ [Mutually disjoint events]

$= P(A).P\left({}^{B}/_{A}\right) + P(\bar{A}).P\left({}^{B}/_{\bar{A}}\right)$

$= \frac{1}{2}.\frac{1}{4} + \frac{1}{2}.\frac{3}{4} = \frac{1}{2}$

Solution 5



$P(A)\,P(B/A)$ = $P(A \cap B)$ $= \frac{1}{2} \cdot \frac{1}{4}$

$P(\bar{A})\,P(B/\bar{A})$ = $P(\bar{A} \cap B)$ $= \frac{1}{2} \cdot \frac{3}{4}$

$$P(B) = P(A).P(B/A) + P(\bar{A}).P(B/\bar{A}) = P(A \cap B) + P(\bar{A} \cap B) = \frac{1}{2}$$

Problem - 6

An individual has 3 different email accounts. Most of her messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. What is the probability that a randomly selected message is spam?

Solution - 6

A_i = { message is from account i} for i = 1,2,3

B = { message is span}

Given that

$P(A_1) = 0.7, P(A_2) = 0.2, P(A_3) = 0.1$

And $P\left(B/_{A_1}\right) = 0.01, P\left(B/_{A_2}\right) = 0.02, P\left(B/_{A_3}\right) = 0.05$

Solution - 6

By Total probability

$$P(B) = P[(A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)]$$

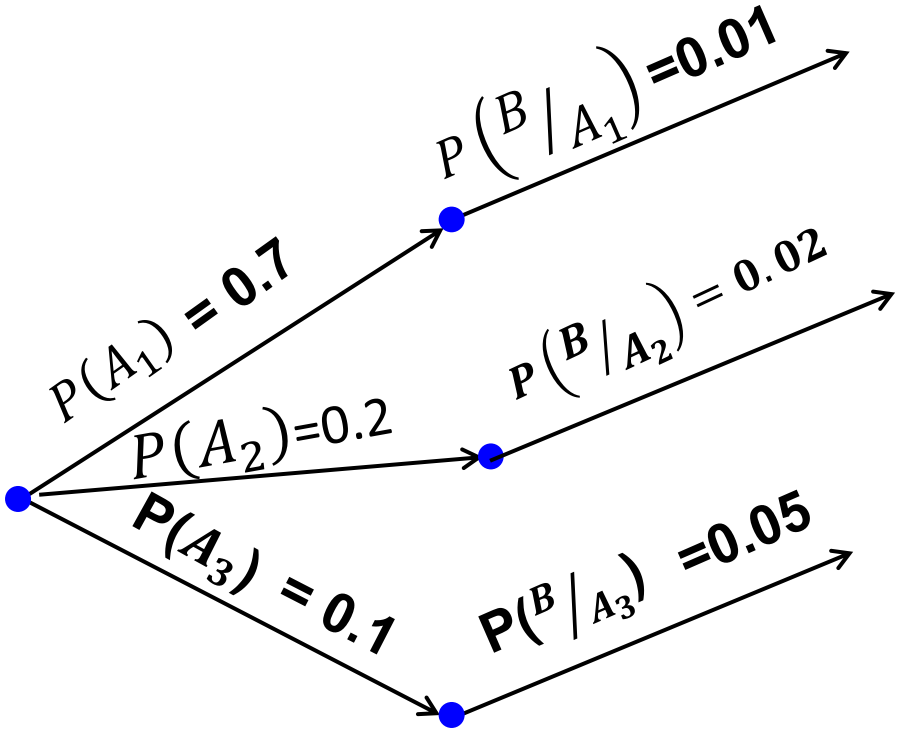
$= P[(A_1 \cap B)] + P[(A_2 \cap B)] + P[(A_3 \cap B)]$ [Mutually disjoint events]

$$= P(A_1)P\left({}^{B}/_{A_1}\right) + P(A_2)P\left({}^{B}/_{A_2}\right) + P(A_3)P\left({}^{B}/_{A_3}\right)$$

= (0.7)(0.01)+(0.2)(0.02)+(0.1)(0.05)

= 0.016

Probability Tree Diagram



$P(A_1)\ P\left(B/A_1\right)$ = (0.7)(0.01)

$P(A_1)\ P\left(B/A_1\right)$ =(0.2)(0.02)

$P(A_1)\ P\left(B/A_1\right)$ = (0.1)(0.05)

P(B) $=P(A_1)P\left(B/A_1\right) + P(A_2)P\left(B/A_2\right) + (A_3)P\left(B/A_3\right)$ = 0.016

Problem - 7

A company that manufactures video camera as produces a basic model and a deluxe model. Over the years 40% of the cameras sold have been the basic model. Of those buying the basic model 30% purchase an extended warranty, whereas 50% of all deluxe purchasers do so.

What is the probability that a randomly selected customer has an extended warranty, how likely is it that he or she has a basic model?

Solution - 7

Let B be the event of a basic model.

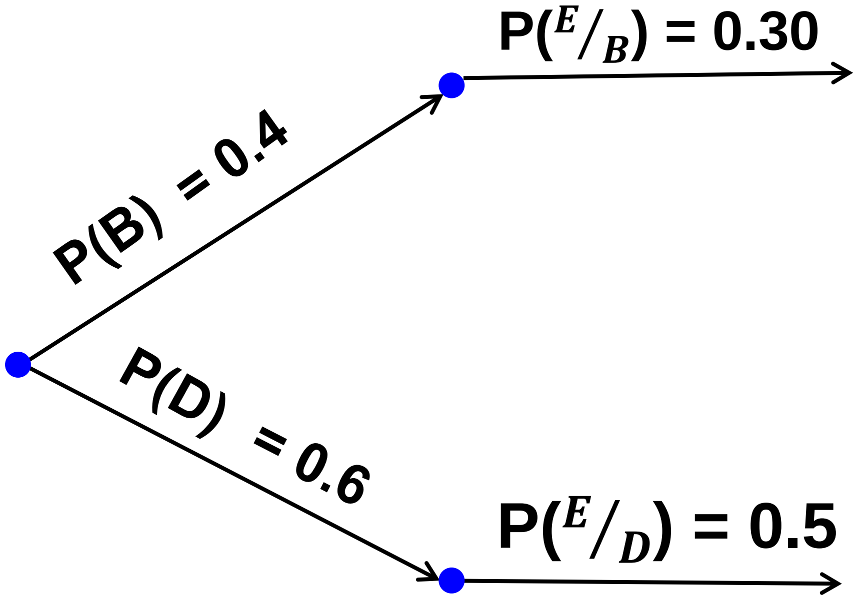
Let D be the event of a deluxe model.

Let E be event of extended warranty.

$P(B) = 0.40$, $P(D) = 0.60$,

$P(E / B) = 0.30$, and $P(E / D) = 0.50$

Probability Tree Diagram



$P(B)\,P(E/B) = P(B \cap E) = (0.4)(0.3)$

$P(D)\,P(E/D) = P(D \cap E) = (0.6)(0.5)$

$P(E) = P(B)\,P(E/B) + P(D)\,P(E/D) = 0.42$

Solution - 7

the probability that a randomly selected customer has an extended warranty, how likely is it that he or she has a basic model is

$$P\left({}^{B}/_{E}\right) = \frac{P(B \cap E)}{P(E)} = \frac{(0.4)(0.3)}{0.42} = 0.2857$$

Problem - 8

In answering a question on a multiple choice test a student either knows the answer or he guesses.

Let p be the probability that he knows the answer and 1-p be the probability that he guesses. Assume that a student who guesses at the answer will be correct probability is 1/5, where 5 is the number of multiple-choice alternatives.

What is the probability that a student knows the answer to a question given that he answer it correctly?

Solution - 8

Let A be the event the student knows the right answer .

$P(A) = p$

Let B be the event the student guesses the right answer .

$P(B) = 1 - p$

And

Let E be event the student gets the right answer .

$P(E/A) = 1$ [P(student gets the right answer given that he knows the right answer)]

$P(E/B) = 1/5$

Probability Tree Diagram

$P(A) = p$

$P(E/A) = 1$

$P(A)\,P(E/A) = P(A \cap E) = (p)(1)$

$P(B) = 1-p$

$P(E/B) = 1/5$

$P(B)\,P(E/B) = P(B \cap E) = (1-p)(1/5)$

$$P(E) = P(A)\,P(E/A) + P(B)\,P(E/B) = (4p+1)/5$$

Solution - 8

The probability that a student knows the answer to a question given that he answer it correctly

$$P(A/E) = \frac{P(A \cap E)}{P(E)} = \frac{p}{(4p+1)/5} = \frac{5p}{(4p+1)}$$

Problem - 9

Verify that $P(X) = \frac{x+3}{25} \;\; for\; x = 1,2,3,4,5$

Serve as Probability mass function?

Solution:

Given that $P(X) = \frac{x+3}{25}\ \ for\ x = 1,2,3,4,5$

$$P(X=1) = \frac{1+3}{25} = \frac{4}{25},$$

$$P(X=2) = \frac{5}{25}$$

$$P(X=3) = \frac{6}{25}$$

$$P(X=4) = \frac{7}{25}$$

$$P(X=5) = \frac{8}{25}$$

$$\sum_{x=1}^{5} P(X=x) = P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

$$= \frac{4}{25} + \frac{5}{25} + \frac{6}{25} + \frac{7}{25} + \frac{8}{25}$$

$$= \frac{30}{25} > 1$$

Given P(X) is not a Probability mass function

[we know that Total probability is 1]

Random variables

Introduction: In random experiments we are interested in the numerical outcomes (i.e numbers associated with the out comes) of the experiment.

Eg: (1) When we throw 2 coins

$S = \{HH, HT, TH, TT\}$

Tails	X —	0	1	1	2
Heads	Y —	2	1	1	0

I. Discrete random variables:-

A random variable which can take only a finite numbers is called discrete random variable.

Ex:- The r.v X = sum of the spots on two dice is discrete, since "S" can assume only the value 2, --12

Probability mass function :— ∴ The no of possibiliti is finite

(Probability Distribution) (P.M.f)

Let 'X' be a discrete random variable taking values $x_1, x_2 -- x_n$ then the probability mass function $[P(X=x)]$ is defined under the following conditions.

(i) $P(X=x) \geq 0$

(ii) $\sum_{\forall x} P(X=x) = 1$.

Mathematical expectation :-

Let 'x' be the discrete random variable taking value $x_1 x_2 \ldots x_n$ then the mathematical ~~expression~~ expectation of X is denoted by $E(x)$

$$E(x) = \sum_{\forall x} (X=x)\, P\,(X=x)$$

Mean :-

$$\text{mean} = \mu = E(x) = \sum_{\forall x} x\,P(x)$$

If $k \geq 0$

$$E(x^k) = \sum_{\forall x} x^k P(x)$$

Variance :- $\sigma^2 = E(x-\mu)^2$

$$\boxed{\sigma^2 = E(x^2) - [E(x)]^2}$$

$$\boxed{V(x) = \sigma^2 = \sum (x-\mu)^2 P(x)}$$

Standard deviation :- $S.D = \sqrt{V(x)}$

① A random variable X has the following probability function

(i) find the value of k

(ii) Mean

(iii) variance

(iv) $P(X \geq 3)$ (v) $P(1 < X \leq 5)$

x	1	2	3	4	5	6
P(x)	k	3k	5k	7k	9k	11k

(i) w.k.t by def of discrete random variable

$$\Rightarrow \sum_{i=1}^{n} P(x_i) = 1$$

$\Rightarrow P(x=1) + P(x=2) + \cdots + P(x=6) = 1,$

$\Rightarrow \therefore K + 3K + 5K + 7K + 9K + 11K = 36K = 1$

$$\boxed{K = \frac{1}{36}}$$

$$E(x)=\sum xP(x)= 1\cdot\frac{1}{36}+2\cdot\frac{3}{36}+3\cdot\frac{5}{36}+4\cdot\frac{7}{36}$$

$$+5\cdot\frac{9}{36}+6\cdot\frac{11}{36}=4.47$$

(iii) <u>Variance</u> $\sigma^2=\sum(x-\mu)^2 p(x)$

$$\Rightarrow \quad \sigma^2=E(x^2)-[E(x)]^2$$

$$\Rightarrow \sigma^2=\sum x^2 p(x)-[E(x)]^2$$

$$\Rightarrow \sigma^2= \frac{1}{36}+4\cdot\frac{(3)}{36}+9\cdot\frac{(5)}{36}+16\frac{(7)}{36}+25\left(\frac{9}{36}\right)$$

$$+36\left(\frac{11}{36}\right)-(4.47)^2$$

$$\sigma^2=\frac{791}{36}-(4.47)^2=1.99$$

$$\therefore \quad \sigma^2=1.99$$

(iv) $P(X \geq 3)$

$$P(X \geq 3) = P(X=3) + P(X=4) + P(X=5) + P(X=6)$$

$$= \frac{5}{36} + \frac{7}{36} + \frac{9}{36} + \frac{11}{36} = \frac{32}{36} = \frac{8}{9}$$

$$P(X \geq 3) = \frac{8}{9}$$

(v) $P(1 < X \leq 5) = P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$$= \frac{3}{36} + \frac{5}{36} + \frac{7}{36} + \frac{9}{36} = \frac{24}{36} = \frac{2}{3}$$

(2) A random variable X has the following probability function

X	-3	-2	-1	0	1	2	3
$P(X)$	K	0.1	K	0.2	$2K$	0.4	$2K$

Find Mean and variance.

(3)* Two dice are thrown & assign to each point if S the sum of the variables on the faces. Find the mean and variance of the random variable.

Sol:-

$$S = \begin{bmatrix} (11) & \cdots & (16) \\ \vdots & & \\ (61) & \cdots & (66) \end{bmatrix}$$

$X(S) = (2, 3, 4\ 5 \cdot 6\ 7, 8\ 9, 10\ 11\ 12)$.

since sum of the variables 2 to 12

$P(2) = P(X=2) = P(11) = \frac{1}{36}$

$P(3) = P(X=3) = P((12)(21)) = \frac{2}{36}$

$P(4) = P(X=4) = P[(13)(31)(22)] = \frac{3}{36}$

$P(5) = P(X=5) = P[(14)(41)(23)(3,2)] = \frac{4}{36}$

$$P(6) = P(X=6) = P[(1\,5)\ (5\,1)\ (2\,4)\ (4\,2)\ (3\,3)] = \frac{5}{36}$$

$$P(7) = P(X=7) = P[(1\,6)\ (6\,1)\ (2\,5)\ (5\,2)\ (3\,4)\ (4\,3)] = \frac{6}{36}$$

$$P(8) = P(X=8) = P[(2\,6)\ (6\,2)\ (3\,5)\ (5\,3)\ (4\,4)] = 5/36$$

$$P(9) = P(X=9) = P[(3\,6)\ (6\,3)\ (4\,5)\ (5\,4)] = \frac{4}{36}$$

$$P(10) = P(X=10) = P[(4,6)\ (6,4)\ (5,5)] = \frac{3}{36}$$

$$P(11) = P(X=11) = P[(5,6)\ (6,5)] = 2/36$$

$$P(12) = P(X=12) = P[(6\,6)] = 1/36$$

II. Continuous Random Variables—

A random variable x is said to be continuous if it can take all possible values in an interval.

Ex:- Age, height, weight etc.

Probability Density ~~Formation~~ Function :—

$f(x)$ is said to PDF if it satisfies :

(i) $f(x) \geq 0 \quad -\infty < x < \infty$.

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

Note:-

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

② variance $= \int_{-\infty}^{\infty} x^2 f(x)\, dx - (\text{mean})^2$

① If a random variable has the probability density $f(x) = \begin{cases} 2e^{-2x} & \text{for } x>0 \\ 0 & \text{for } x\le 0 \end{cases}$

Find the probabilities that it will take on a value

(i) between 1 and 3

(ii) greater than 0.5

3. If p.d.f $f(x) = kx^3$ in $1 \le x \le 3$

0 else where.

Find the value of k and find the probability between $x = \frac{1}{2}$ and $x = \frac{3}{2}$

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

$$\Rightarrow \int_{-\infty}^{1} f(x)\, dx + \int_{1}^{3} f(x)\, dx + \int_{3}^{\infty} f(x)\, dx = 1$$

$$\Rightarrow 0 + \int_{1}^{3} k x^3\, dx + 0 = 1$$

$$\Rightarrow k\left[\frac{x^4}{4}\right]_1^3 = 1 \Rightarrow \frac{k}{4}\left[3^4 - 1\right] = 1$$

$$\Rightarrow \frac{k}{4}[81-1] = 1 \Rightarrow \frac{k}{4}(80) = 1$$

$$\Rightarrow k = \frac{4}{80} \Rightarrow \boxed{k = \frac{1}{20}}$$

$$P\left(\frac{1}{2} \le x \le \frac{3}{2}\right) = \int^{3/2} f(x)\, dx = \int^{3/2} k x^3\, dx$$

(4) The probability density fun

is $y = k(3x^2 - 1)$ in $-1 \leq x \leq 2$

$= 0$ else where.

Find the value of k and find the probability $(-1 \leq x \leq 0)$

(6) The probability density function of a random variable x is

$f(x) = kx(x-1)$ in $1 \le x \le 4$

$= 0$ elsewhere.

$P(1 \le x \le 3) = \frac{28}{3}$. Find the value of k.

Sol:-

$$P(1 \le x \le 3) = \int_1^3 f(x)\,dx \quad \cancel{=2}8$$

$$\frac{28}{3} = \int_1^3 k\, x(x-1)\,dx$$

$$\frac{28}{3} = \int_1^3 k(x^2 - x)\,dx$$

$$\frac{28}{3} = k\left[\frac{x^3}{3} - \frac{x^2}{2}\right]_1^3$$

$$\frac{28}{3} = k\left\{\left[9 - \frac{9}{2}\right] - \left[\frac{1}{3} - \frac{1}{2}\right]\right\}$$

$$\frac{28}{3} = k\left[\frac{9}{2} + \frac{1}{6}\right] \Rightarrow \frac{14}{3}k = \frac{28}{3}$$

$$\Rightarrow \boxed{k = 2}$$

DISTRIBUTION

innovate achieve lead

Binomial distribution is defined under the following assumptions:

1. There are only two possible outcomes for each trial [traials] (i.e) success and failure. [S9k'sel]
2. The probability of a success is the same for each trial. [traial]
3. There are n trials where n is a constant
4. The n trials are independent are known as Bernoulli's trials.

$$P(X=x) = B(x,n,p) = {}^nc_x\, p^x q^{n-x} \quad \forall x$$

Note:-

① $p+q=1$.

② $\sum_{x=0}^{n} b(x, n, p) = 1$

① A coin is tossed 9 times. Find the probability of getting 5 heads.

Sol:- $B(x, n, p) = {}^{n}C_{x}\, p^{x} (1-p)^{n-x}$.

Here n = number of trials $= 9$

$x = 5$

p = probability of getting head $= \frac{1}{2}$

$P(X=5) = {}^{9}C_{5} \left(\frac{1}{2}\right)^{5} \left(\frac{1}{2}\right)^{4}$

$= {}^{9}C_{5} \left(\frac{1}{2}\right)^{9} = \frac{63}{256} = 0.246$

② A die is thrown 8 times. If getting 2 (or) 4 is a success. Find the probability of

(i) 4 success (ii) $P(X \leq 3)$ (iii) $P(X > 3)$

Sol :- $B(x, n, p) = {}^nC_x p^x (1-p)^{n-x}$

n = number of trials = 8.

p = probability of getting 2 (or) 4 $= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

(i) $x = 4$

$$P(x=4) = {}^{8}C_{4} \left(\frac{1}{3}\right)^{4} \left(1-\frac{1}{3}\right)^{8-4}$$

$$= {}^{8}C_{4} \left(\frac{1}{3}\right)^{4} \left(\frac{2}{3}\right)^{4}$$

$$= \frac{1120}{6561}$$

(ii) $P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$$= {}^{8}C_{0} \left(\frac{1}{3}\right)^{0} \left(\frac{2}{3}\right)^{8} + {}^{8}C_{1} \left(\frac{1}{3}\right)^{1} \left(\frac{2}{3}\right)^{7}$$

$$+ {}^{8}C_{2} \left(\frac{1}{3}\right)^{2} \left(\frac{2}{3}\right)^{6}$$

$$+ {}^{8}C_{3} \left(\frac{1}{3}\right)^{3} \left(\frac{2}{3}\right)^{5}$$

(iii) $P(x \geq 2) = 1 - P(x=0) - P(x=1)$

$$= 1 - {}^{8}C_{0}\left(\frac{1}{3}\right)^{0}\left(\frac{2}{3}\right)^{8} - {}^{8}C_{1}\left(\frac{1}{3}\right)^{1}\left(\frac{2}{3}\right)^{7}$$

$$= 1 - \left(\frac{2}{3}\right)^{8} - 8 \cdot \frac{1}{3} \cdot \frac{2^{7}}{3^{7}}$$

$$= 1 - \frac{2^{8}}{3^{8}} - \frac{8(2^{7})}{3^{8}}$$

$$= \frac{5281}{6561}$$

2416

(3) Two dice are thrown 5 times.

If getting a doublet is a [dʌbəlit] success. Find the probability that getting the success atleast once.

Here $n = 5$

P = probability of getting success

= sample space consists of 36 elements among which double (line).

(1 1) (2 2) (3 3) (4 4) (5 5) (6 6)

$$P = \frac{6}{36} = \frac{1}{6} \quad \therefore \quad q = \frac{5}{6}$$

$$P(X \geq 1) = 1 - P(X = 0)$$

$$= 1 - {}^5C_0 \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^5$$

$$= 1 - \left(\frac{5}{6}\right)^5 = \frac{4651}{7776}$$

4. Among the iteems produced in a factory 5% are defective. Find the probability that a sample of 8 contains

(i) exactly 2 defective items

(ii) greater than or equal to 7 defective items

(iii) at least one defective item.

Here $n = 8$

$p =$ probability of defective item $= 5\%$

$$p = \frac{5}{100} = \frac{1}{20}.$$

(i) $x = 2$

$$P(x=2) = {}^{8}C_{2}\left(\frac{1}{20}\right)^{2}\left(\frac{19}{20}\right)^{6}$$

$$= \cancel{28}^{\,7} \cdot \frac{1}{\cancel{400}_{\,100}} \cdot \left(\frac{19}{20}\right)^{6}$$

$$P(x=2) = \frac{7}{100} \cdot \left(\frac{19}{20}\right)^{6}$$

(ii) $P(X \geq 7) = P(X=7) + P(X=8)$

$$= {}^8C_7 \left(\frac{1}{20}\right)^7 \left(\frac{19}{20}\right)^1 + {}^8C_8 \left(\frac{1}{20}\right)^8 \left(\frac{19}{20}\right)^0$$

$$= 8 \frac{1}{20^7} \cdot \frac{19}{20} + 1 \cdot \frac{1}{20^8}.$$

$$= \frac{1}{20^8} \left[19 \times 8 + 1 \right]$$

$$= \frac{1}{20^8} (152 + 1) = \frac{153}{20^8}.$$

(iii) $P(X \geq 1) = 1 - P(X=0)$

$$= 1 - {}^8C_0 \left(\frac{1}{20}\right)^0 \left(\frac{19}{20}\right)^8$$

$$= 1 - 1 \cdot 1 \cdot \left(\frac{19}{20}\right)^8.$$

$$= 1 - \frac{19^8}{20^8}.$$

$\boxed{\text{Variance of Binomial Distribution is } npq}$

$\boxed{\text{The standard deviation is } \sigma = \sqrt{npq}}$

$\boxed{\text{mode} = (n+1)p}$

Note:-

① If $(n+1)p$ is not an integer.
mode is the integral part of $(n+1)p$.
In this case the distribution is called unimodal.

② If $(n+1)p$ is an integer both $(n+1)p$ and $(n+1)p-1$ will represent

① Determine the Binomial distribution for which the mean is 4 and variance 3 and find its mode.

$$\frac{②}{①} \Rightarrow \frac{npq}{np} = \frac{3}{4}$$

$$q = \frac{3}{4} \qquad p = \frac{1}{4}$$

$$np = 4 \Rightarrow n\left(\frac{1}{4}\right) = 4$$

$$\Rightarrow n = 16$$

$$\text{Binomial distribution} = (p+q)^n$$

$$= \left(\frac{1}{4} + \frac{3}{4}\right)^{16}$$

$$\text{mode} = (n+1)p = 17\left(\frac{1}{4}\right) = 4.25$$

$(n+1)p$ is not an integer.

$\therefore$ Integral part of $(n+1)p$

(ie) mode $= 4$.

(4). Six dice are thrown 243 times. How many times do you expect at least two dice to show a 5 (or) 6.

$N = 243.$

Sol: $n = 6$

The probability p of getting 5 or 6 $= \frac{1}{6} + \frac{1}{6}$

$= \frac{1}{3}$

$$P(X \geq 2) = 1 - P(X=0) - P(X=1)$$

$$= 1 - 6_{C_0}\left(\frac{2}{3}\right)^6 - 6_{C_1}\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^5$$

$$= 1 - \left(\frac{2}{3}\right)^6 - 6\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^5$$

$$= 1 - \left(\frac{2}{3}\right)^6 \left[1 + 3\right] = \frac{473}{729}$$

Expected number of dice $= N\,P(x)$

$$= 243 \cdot \frac{473}{729} = \frac{473}{3}$$

$$= 158.$$

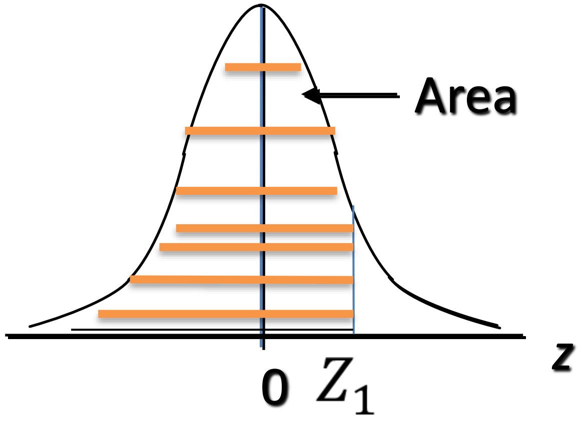
Learning objectives

- Problems on Normal Distributions
- Interval Estimations

Rules for finding Normal Probabilities

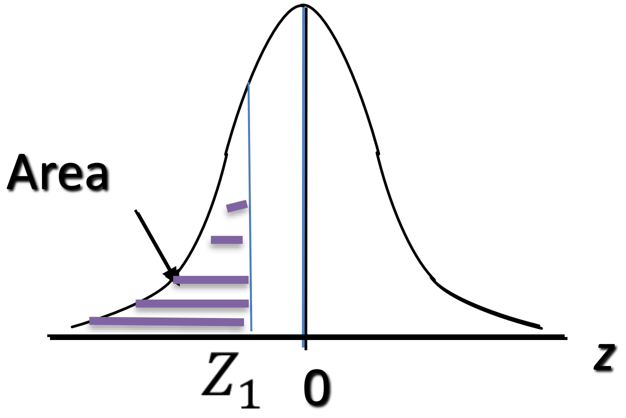
i). If $Z_1\ is + ve$ then

$$P(Z < Z_1) = P(Z < Z_1)$$



ii). If $Z_1\ is - ve$ then

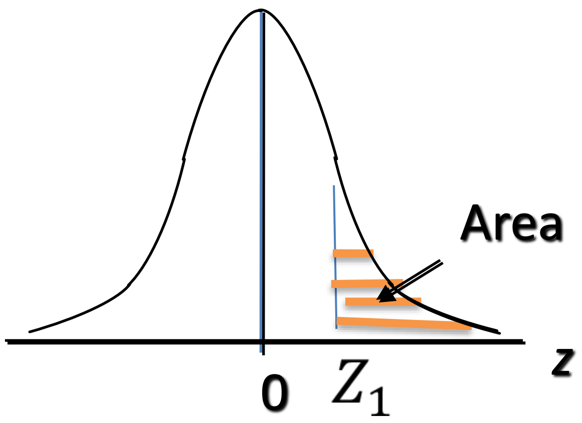
$$P(Z < Z_1) = P(Z < Z_1)$$



Rules for finding Normal Probabilities

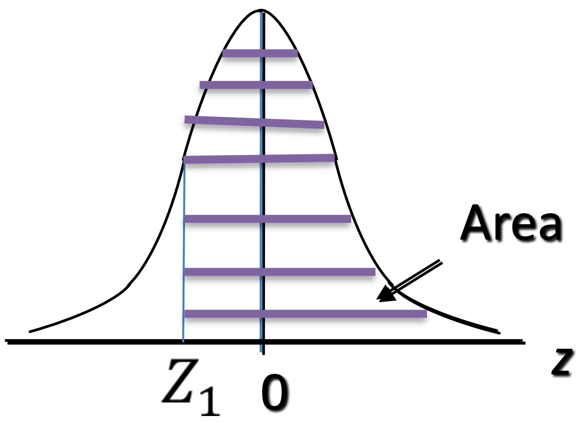
iii). If Z_1 $is + ve$ then

$$P(Z > Z_1) = 1 - P(Z < Z_1)$$



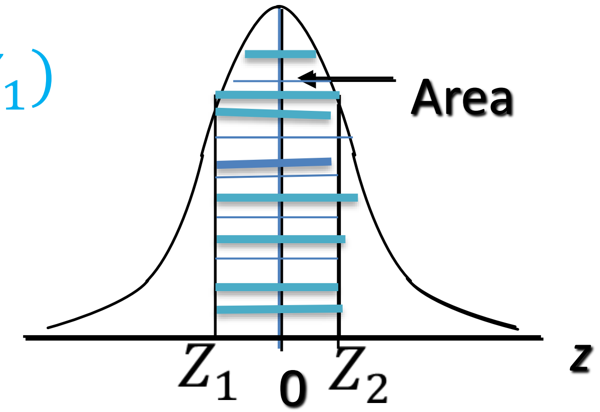
iv). If Z_1 $is - ve$ then

$$P(Z > Z_1) = 1 - P(Z < Z_1)$$



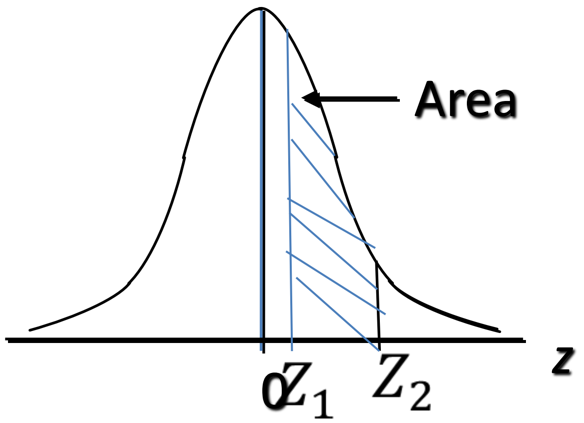
v). If Z_1 $is\ -ve\ and\ Z_2\ is + ve$ then

$$P(Z_1 < Z < Z_2) = P(Z < Z_2) - P(Z < Z_1)$$



vi). If $Z_1\ is + ve\ and\ Z_2\ is + ve$ then

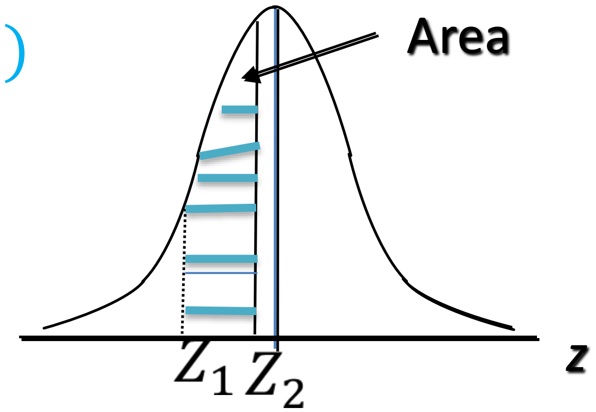
$$P(Z_1 < Z < Z_2) = P(Z < Z_2) - P(Z < Z_1)$$



Rules for finding Normal Probabilities

vii). If Z_1 is $-ve$ and Z_2 is $-ve$ then

$$P(Z_1 < Z < Z_2) = P(Z < Z_2) - P(Z < Z_1)$$



Problem 1

The automatic opening device of a military cargo parachute has been designed to open when the parachute is 200 m above the ground. Suppose opening altitude actually has a normal distribution with mean value 200m and standard deviation 30m. Equipment damage will occur if the parachute opens at an altitude of less than 100m.

What is the probability that there is equipment damage.

Solution 1

innovate achieve lead

Let μ be the mean and σ be the standard deviation

Given that $\mu = 200$m , $\sigma = 30$m

Let X be the altitude above the ground that a parachute opens.

the probability that there is equipment damaged is $= P(X \leq 100) = P(X < x_1)$

$= P(Z \leq z_1)$

The normal variable is

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{100 - 200}{30} = -3.33$$

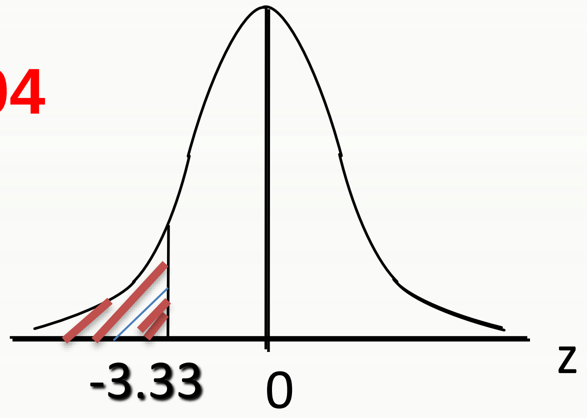
Solution 1

$$\therefore P(X \leq 100) = P(Z \leq -3.33)$$

$$= \quad P(\infty < Z < -3.33)$$

$= \ 0.0004$ (From Normal distribution Table)

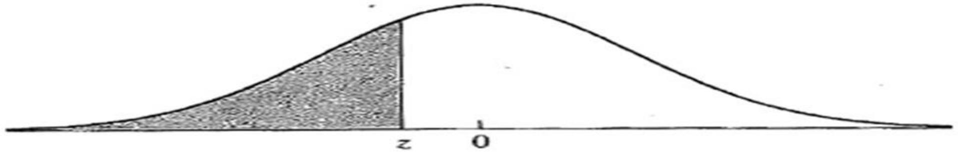
The probability that there is equipment damage is 0.0004



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

TABLE A.2 Cumulative normal distribution (z table)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Problem 2

The mean yield for one – acre plot is 662 kilos with a standard deviation 32 kilos. Assuming normal distribution, how many one – acre plots in a batch of 1000 plots would be expect to have yield Over 700 kilos, Below 650 kilos and

What is the yield of the best 100 plots ?

Solution 2

Let X denotes the yield (in kilos) for one – acre plot.

Let μ be the mean and σ be the standard deviation

Given that $\mu = 662$ and $\sigma = 32$.

N = batch of 1000 plots = 1000

i) The probability that a plot has a yield over 700 kilos is

$$= P(X > 700) = P(X > X_1)$$

Here $X_1 = 700$ then

The normal variable is

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{700 - 662}{32} = 1.1875 = 1.19$$

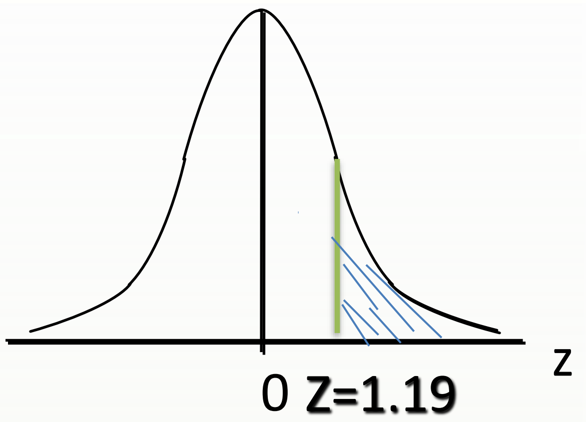
Solution 2

$P(X > 700) = P(Z > 1.19)$

$= 1 - P(\infty < Z < 1.19)$

= 1 – 0.8829 (From Normal distribution Table)

= 0.1171



Expected no.of plots with yield over 700 kilos is = N X $P(X > 700)$

= 1000 X 0.1171= 117 plots

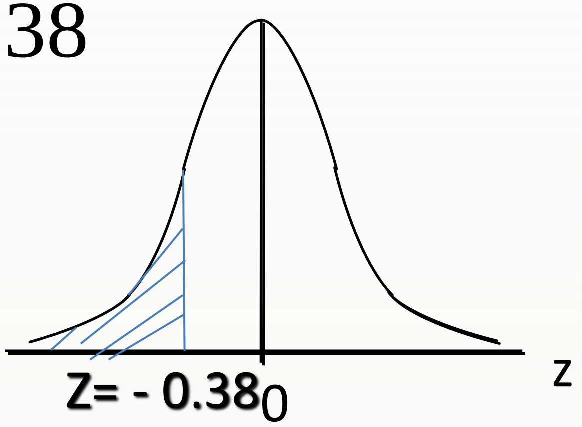
Solution 2

ii) The probability that the plot with yield below 650 kilos is

= P(X < 650) = P(X < X_1)

Here X_1 = 650 then

The normal variable is $Z_1 = \dfrac{X_1 - \mu}{\sigma} = \dfrac{650 - 662}{32} = -0.38$



$P(X < 650) = P(Z < -0.38)$

= 0.352 (From Normal distribution Table)

Expected no.of plots with yield below 650 kilos is = N X $P(X < 650)$

= 1000 X 0.352= 352 plots

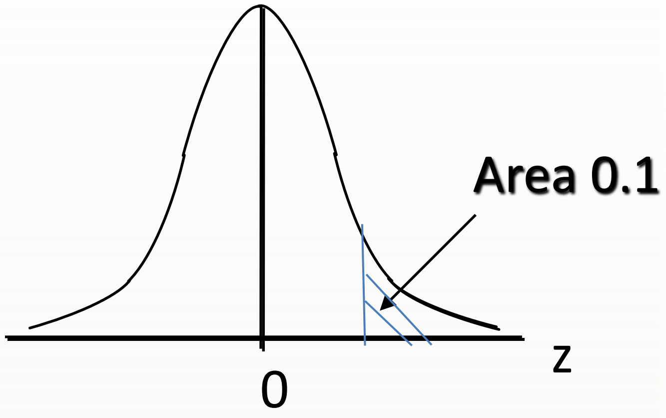
Solution 2

iii) Let X_1 be the yield of the best 100 plots

given that

$$P(X > X_1) = P(Z > Z_1) = \frac{100}{1000} = 0.1$$

Then $P(Z < Z_1) = 1 - 0.1 = 0.9$



for the area 0.9 the Z – Value is 1.29

$$1.29 = \frac{X_1 - \mu}{\sigma} = \frac{X_1 - 662}{32}$$

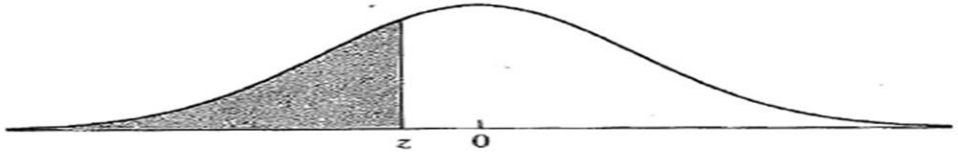
$$X_1 = (1.29\ X\ 32\ ) + 662 = 703.28$$

the best 100 plots have yield over 703.28 kilos.

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

TABLE A.2 Cumulative normal distribution (*z* table)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Problem 3

A soft-drink machine is regulated so that it discharges an average of 200 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,

(a) what fraction of the cups will contain more than 224 milliliters?

(b) what is the probability that a cup contains between

191 and 209 milliliters?

(c) how many cups will probably overflow if 230 milliliters cups are used for the next 1000 drinks?

Solution 3

Let X denotes the amount of drink distribution.

Let μ be the mean and σ be the standard deviation

G0iven that μ = **200** milliliters per cup and σ = **15** milliliters per cup.

a) Let Calculate

what fraction of the cups will contain more than 224 milliliters

$$P(X > 224) = P(X > X_1)$$

Here X_1 = 224 then

The normal variable is

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{224 - 200}{15} = 1.6$$

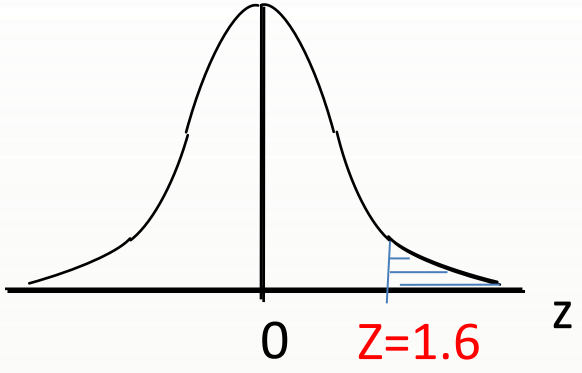
Solution 3

$P(X > 224) = P(Z > 1.6)$

$= 1 - P(\infty < Z < 1.6)$

$= 1 - 0.9452$ (From Normal distribution Table)

$= 0.0548$



b) what is the probability that a cup contains between 191 and 209 millilitres

$P(191 < X < 209) = P(X_1 < X < X_2)$

Here X_1 = 191 and X_2 = 209 then

The normal variable is

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{191 - 200}{15} = -0.6$$ and $$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{209 - 200}{15} = 0.6$$

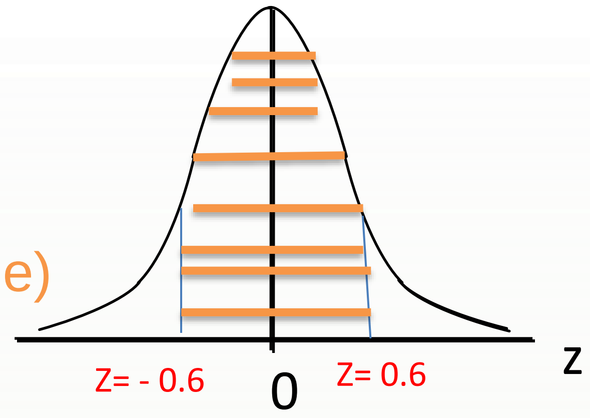
Solution 3

$$P(191 < X < 209) = P(-0.6 < Z < 0.6)$$

$$= P(Z < 0.6) - P(Z < -0.6)$$

= 0.7257 – 0.2743 (From Normal distribution Table)

= 0.4514



The probability that a cup contains between 191 and 209 millilitres is 0.4514

c) probably overflow if 230 milliliters cups are used for the next 1000 drinks

P(X > 230) = P(X > X_1)

Here X_1 = 230 then

The normal variable is

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{230 - 200}{15} = 2$$

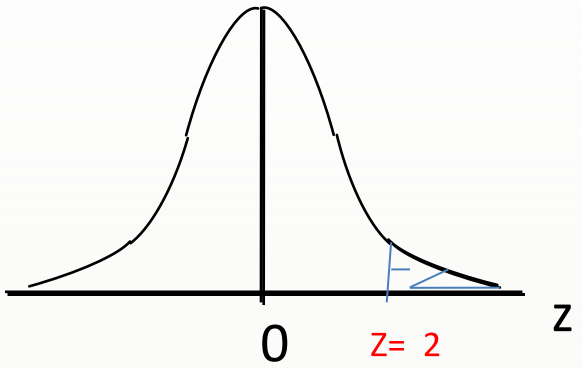
Solution 3

$P(X > 230) = P(Z > 2)$

$= 1 - P(\infty < Z < 2)$

= 1 – 0.9772 (From Normal distribution Table)

= 0.0228



No.of cups probably overflow is = $\mathrm{N}\ \mathrm{X}\ P(X > 230)$

= 1000 X 0.0228

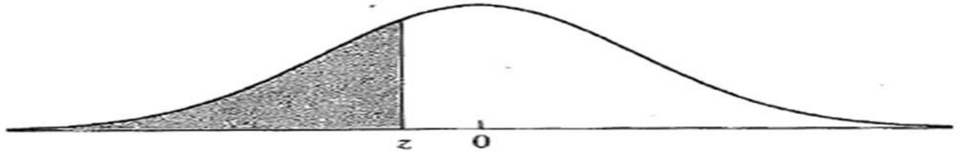
= 22.8 = 23 cups

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

achieve lead

TABLE A.2 Cumulative normal distribution (*z* table)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Problem 4

In an examination it is laid down that a student passes if he secures 40 percent or more. He is placed in the first, second and third division according as he secures 60% or more marks, between 50% and 60% marks and marks between 40% and 50% respectively. He gets a distinction in case he secures 75% or more. It is noticed from the results that 10% of the students failed in the examination, Whereas 5% of them obtained distinction. Calculate the percentage of students placed in the second division.

Solution 4

Let μ be the mean and σ be the standard deviation

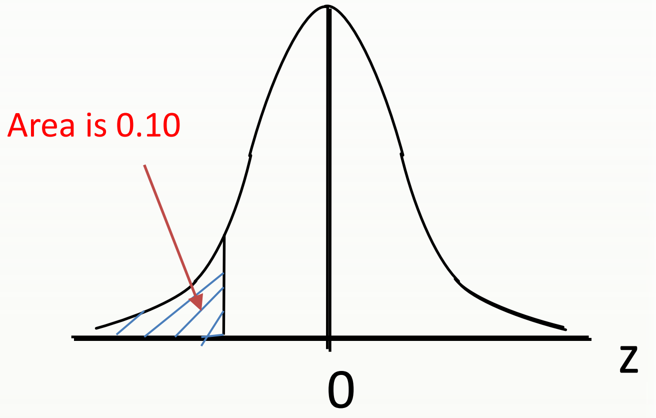
Let X denote the marks in the examination.

Given the results that 10% of the students failed in the examination i.e,

P(X < 40) = 0.10 < 0.5

When X = 40,

$$Z = \frac{X - \mu}{\sigma} = \frac{40 - \mu}{\sigma} = -Z_1$$



Solution 4

From normal tables $Z_1 = 1.28$

Then $$\frac{40 - \mu}{\sigma} = -1.28$$

$$=> \mu - 1.28\sigma = 40 - - - - (1)$$

5% of them obtained distinction i.e,

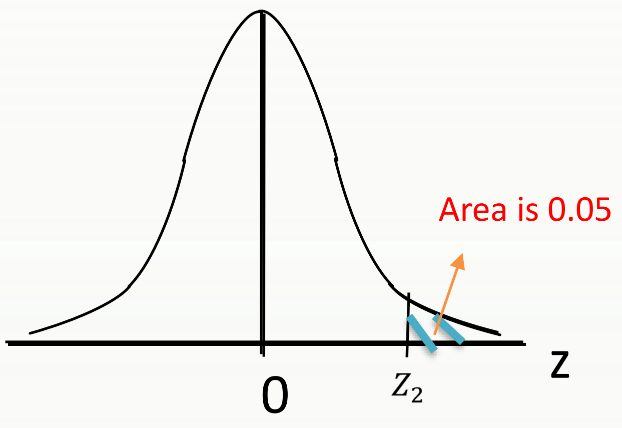
P(X > 75) = 0.05

when X = 75,

$$Z = \frac{X - \mu}{\sigma} = \frac{75 - \mu}{\sigma} = Z_2$$

$$P(\infty < Z < Z_2) = 1 - 0.05$$

$$= 0.95$$



Solution 4

From normal tables $Z_2 = 1.64$

then $\dfrac{75 - \mu}{\sigma} = 1.64$

$$=> \mu + 1.64\sigma = 75 ----(2)$$

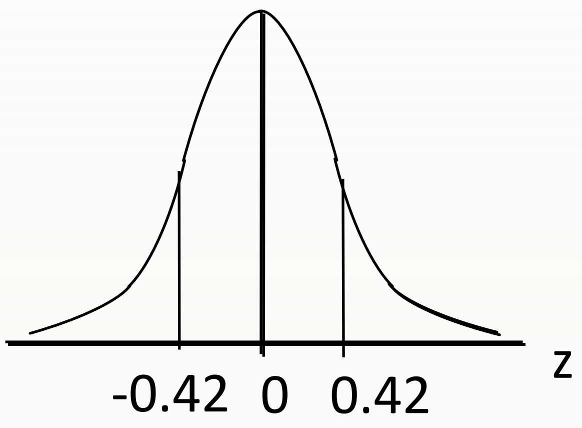
Solving equations (1) and (2) we get μ = 55 , σ = 12

P(students placed in the second division)

= $P(50 < X < 60) = P(X_1 < X < X_2)$

Here X_1 = 50 and X_2 = 60 then

The normal variable is



$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{50 - 55}{12} = -0.42 \quad \text{and} \quad Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{60 - 55}{12} = 0.42$$

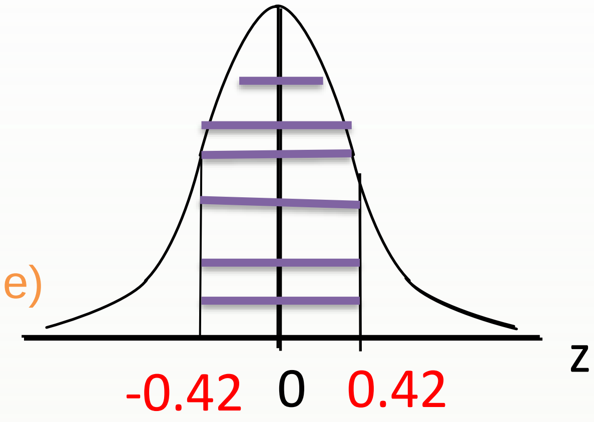
Solution 4

$$P(50 < X < 60) = P(-0.42 < Z < 0.42)$$

$$= P(\infty < Z < 0.42) - P(\infty < Z < -0.42)$$

$= 0.6627 - 0.3372$ (From Normal distribution Table)

$= 0.3255$

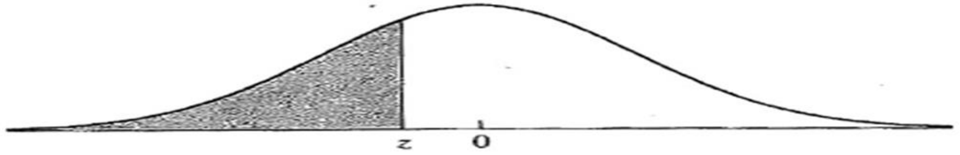


Hence 33% students get second division in the examination.

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

TABLE A.2 Cumulative normal distribution (z table)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Problem 5

A company pays its employees an average wage of $15.90 an hour with a standard deviation of $1.50. If the wages are approximately normally distributed and paid to the nearest cent,

(a) what percentage of the workers receive wages between $13.75 and $16.22 an hour inclusive?

(b) the highest 5% of the employee hourly wages is greater than what amount?

Solution 5

innovate achieve lead

Let X denotes the wages per hour.

Let μ be the mean and σ be the standard deviation

Given that μ = \$15.90 and σ = \$1.50.

a) percentage of the workers receive wages between \$13.75 and \$16.22 an hour inclusive

$$P(13.75 < X < 16.22) = P(X_1 < X < X_2)$$

Here X_1 = 13.75 and X_2 = 16.22 then

The normal variable is

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{13.75 - 15.90}{1.5} = -1.433 \quad \text{and} \quad Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{16.22 - 15.90}{1.5} = 0.213$$

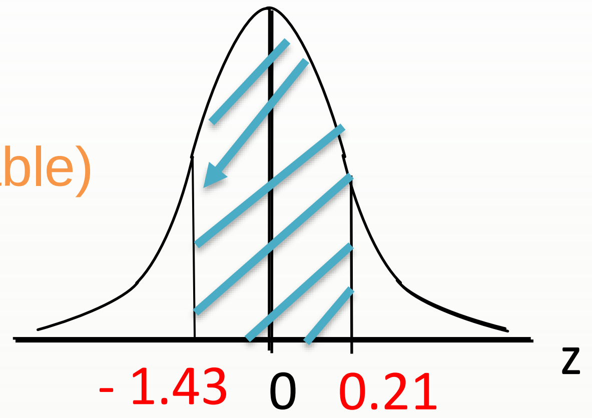
Solution 5

$$P(13.75 < X < 16.22) = P(-1.43 < Z < 0.21)$$

$$= P(Z < 0.21) - P(Z < -1.43)$$

= 0.5832 – 0.0764 (From Normal distribution Table)

= 0.5068



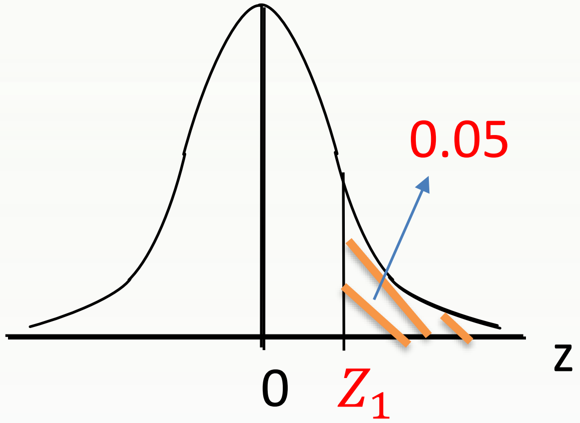
percentage of the workers receive wages between $13.75 and $16.22 an hour inclusive 50.68%

b) the highest 5% of the employee hourly wages is greater than what amount?

$$P(X > X_1) = P(Z > Z_1) = 0.05$$

Then $P(Z < Z_1) = 1 - 0.05 = 0.95$

for the area 0.95 the Z Value is 1.645



Solution 5

$$1.645 = \frac{X_1 - \mu}{\sigma} = \frac{X_1 - 15.90}{1.5}$$

$$X_1 = (1.645\ X\ 1.5\ ) + 15.90 = 18.3675$$

The highest 5% of the employee hourly wages is greater than amount $ 18.37

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

TABLE A.2 Cumulative normal distribution (z table)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Problem 6

The production process for engine control housing units of a particular type has recently been modified. Prior to this modification, historical data had suggested that the distribution of hole diameters for bushings on the housings was normal with a standard deviation of 0.100 mm.

It is believed that the modification has not affected the shape of the distribution or the standard deviation, but that the value of the mean diameter may have changed. A sample of 40 housing units is selected and hole diameter is determined for each one, resulting in a sample mean diameter of 5.426 mm.

Calculate a confidence interval for true average hole diameter using a confidence level of 90%.

Solution 6

innovate achieve lead

Let μ be the mean and σ be the standard deviation of Population

Given that σ = 0.100 and the confidence level is 90% i.e,

$\alpha = 0.1$ then $Z_{\alpha/2} = Z_{0.05} = 1.645$

And sample mean is $\bar{x} = 5.426$, n = 40.

The desired interval is $\bar{x} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$$= 5.426 \pm 1.645\left(\frac{0.100}{\sqrt{40}}\right) = 5.426 \pm 0.026$$

= (5.400, 5.452)

The CI for true average hole diameter is 5.4 < μ < 5.452

Problem 7

Extensive monitoring of a computer time-sharing system has suggested that response time to a particular editing command is normally distributed with standard deviation 25 millisec. A new operating system has been installed, and we wish to estimate the true average response time μ for the new environment. Assuming that response times are still normally distributed with $\sigma = 25$,

what sample size is necessary to ensure that the resulting 95% CI has a width of 10.

Solution 7

Given that $\sigma = 25$

and the confidence level is 95% i.e,

$\alpha = 0.5$ then $Z_{\alpha/2} = Z_{0.25} = 1.96$

The CI is $\left(\bar{x} - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \right)$

and given that length of the interval is 10. i.e,

$$\left(\bar{x} + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \right) - \left(\bar{x} - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \right) = 10$$

$$2Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = 10$$

Solution 7

$$2(1.96)\left(\frac{25}{\sqrt{n}}\right)=10$$

$$\sqrt{n}=9.80$$

$$\Rightarrow n=96.04$$

The Sample size is n = 96

Problem 8

A sample poll of 100 voters chosen at random from all voters in a given district indicated that 55% of them were in favour of a particular candidate. Find 95 % confidence limits for the proportion of all the voters in favour of this candidate.

Solution 8

Given that $P = 0.55$ then $Q = 0.55$

and the confidence level is 95% i.e,

$$\alpha = 0.5 \quad \text{then} \quad Z_{\alpha/2} = Z_{0.25} = 1.96$$

The CI is $\left(P + Z_{\propto/2}\sqrt{\frac{PQ}{n}}, P - Z_{\propto/2}\sqrt{\frac{PQ}{n}}\right)$

$$\left(0.55 + 1.96\sqrt{\frac{(0.55\ X 0.45)}{100}}, 0.55 - 1.96\sqrt{\frac{(0.55\ X 0.45)}{100}}\right)$$

= (0.55 ± 0.0975)

= (0.4525, 0.6525)

95 % confidence limits for the proportion of all the voters in favour of this candidate is (0.4525, 0.6525)

Test of Hypothesis :-

- **Hypothesis** : To arrive at about the decision about the population on the basis of sample information, we make assumptions or guess about the population parameters involved. Such an assumption is called Statistical Hypothesis which may or may not be true.

Ex:- Avg height of soldiers is 165 cm.

→ Test of Hypothesis :- The procedure which enables us to decide on the basis of sample result weather a hypothesis is true or not. is called Test of Hypothesis or test of significance.

There are two types of Hypothesis
1) NULL hypothesis (H_0)
2) Alternative (H_1)

1) Null Hypothesis (H_0) ÷ A definite statement about the population parameter is called Null hypothesis denoted by H_0.

ii H_0 : avg height of Soldiers is 165 cm.

$H_0 : \mu = 165 \text{ cm}$.

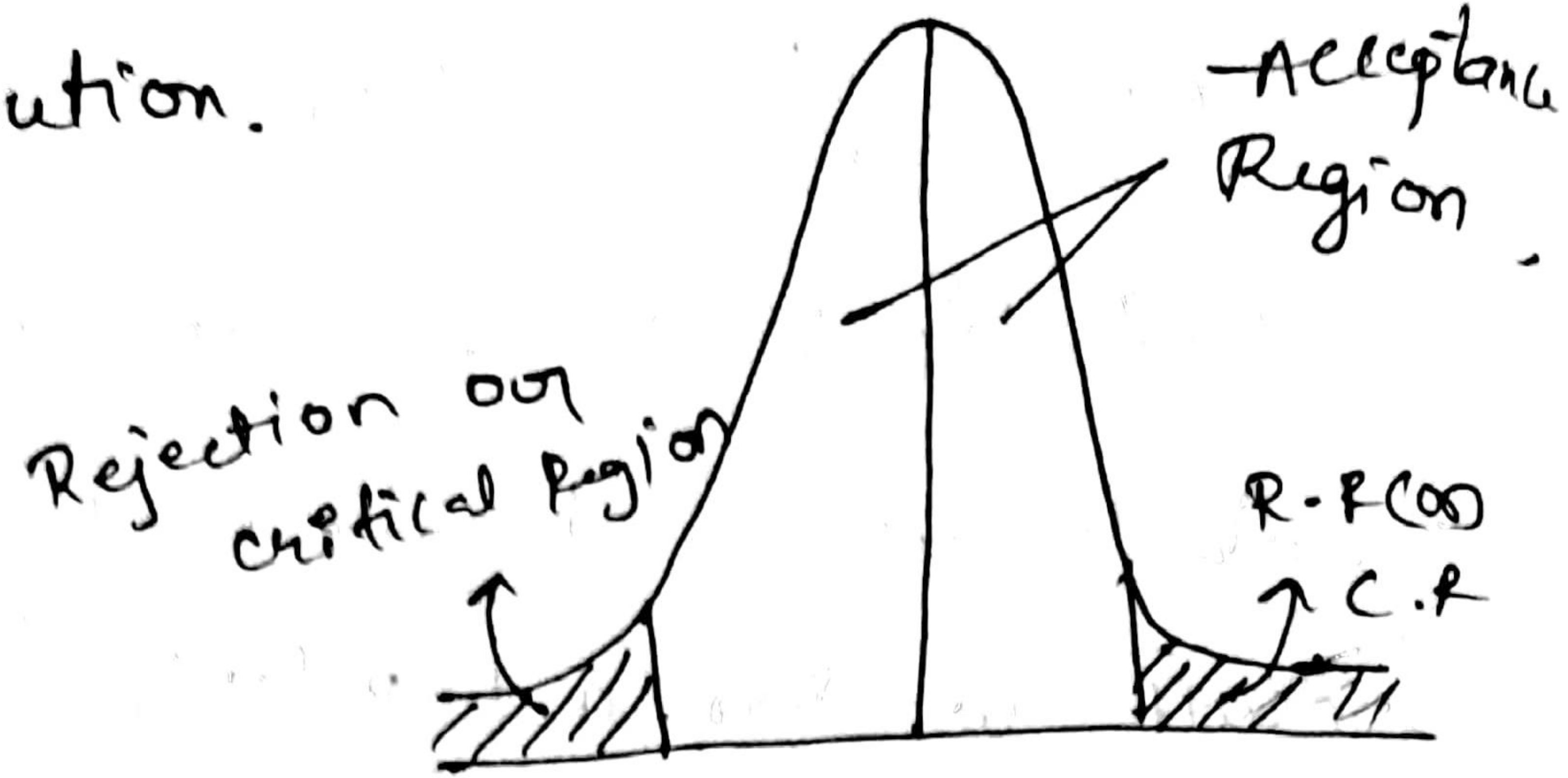
2) Alternative Hypothesis (H_1) ÷ Any hypothesis which contradicts the Null hypothesis is an Alternati hypothesis denoted by H_1.

Ex : $H_1 : \mu \neq 165\text{cm}$ (two tailed test).

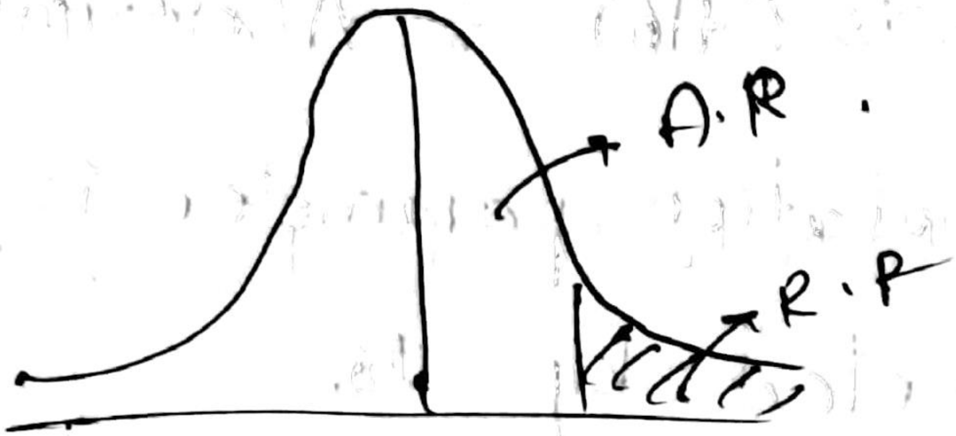
$H_1 : \mu > 165$ (Right one tailed test)

$H_1 : \mu < 165$ (Left one tailed test).

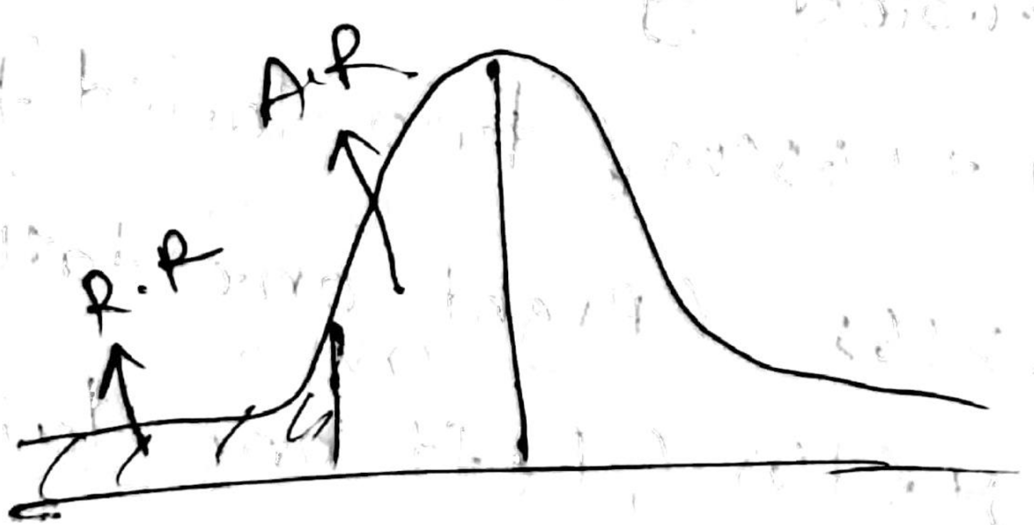
Two tailed test :- If the form has not equal to sign then the critical region (rejection region) is divided equally in the left and Right tails of the distribution.



• **Right One failed test :** If the form of Alternative Hypothesis have greater than sign then the Enter critical Region is taken on the Right tail of the distribution.



• **Left one tailed test** :- If the form of Alternative Hypothesis have less than sign then the critical Region is taken in the left tail of the distribution.



- **Critical or Significant Values**:- The Value of the test statistic which seperates the Rejection Region and acceptance Region is called critical or significant value.

This value depends on

1) the level of significance used.
2) the alternative hypothesis whether it is one tailed or two tailed.

<u>Note</u> ÷ Critical values (Z_α) of Z.

	Level of Significance		
	1% (0.01)	5% (0.05)	10% (0.1)
two-tailed test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Right-tailed test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
left-tailed test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

• Errors in Sampling :

Type I error : Reject Null hypothesis (H_0) when it is true.

Type II error : Accept Null hypothesis (H_0) when it is false.

→ The probability of making type I error is called level of significance and is denoted by α.

* Procedure for testing of hypothesis :

1) Null hypothesis (H_0) : setup taking into consideration the nature of problem and data involved.

2) Alternative Hypothesis (H_1) : setup A.H

So that we could decide wheather we Should use one tail or two tail test.

3) level of significance (α): Select the appropriate level. A Suitable (α) is Selected in advacen it is not given in problem then generally we select 5% level of Significance.

4) Test Statistics: Find test statistics under the Null hypothesis

(z test, t-test, ...)

5) conclusion: If | calculated value | < table Val. then we accept the Null hypothesis.

If | calc. val. | > table val. then we reject the Null hypothesis.

* Test of significance of a single mean (large Sample) :

(or) $n > 30$ $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

Q) According to the norms established for a mechanism aptituted test person who are 18 yrs old having an average ht of 73.2 with a s.d of 8.6. If 40 Randomly selected

person of that age avg. 76.7 test the hypothe
at 1% level of significance.

Sol :- population Mean (μ) = 73.2

S.d (σ) = 8.6.

$n = 40$.

$\bar{x} = 76.7$.

1) Setup null hypothesis. $H_0 : \mu = 73.2$.

2) Setup A.H : $H_1 : \mu < 73.2$.

3) $\alpha = 1\% = 0.01$.

4) $Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{76.7 - 73.2}{8.6 / \sqrt{40}} = 2.57$.

$|Z_{cal}| = 2.57$

$\Rightarrow Z_\alpha = -2.33$.

$|Z_{cal}| > Z\alpha \Rightarrow$ reject Null Hypothesis.

i.e., Accept Alternative hypothesis.

$\mu < 73.2$.

12). An ambulance Service claimed 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 mins and variance of 16 min. test the claim at 0.05 level of significance.

Sol :- $\alpha = 5\%$, $n = 36$, $\bar{x} = 11$, $\mu = 10$

$\sigma = \sqrt{16} = 4$ min.

1) $H_0 : \mu = 10$

2) $H_1 : \mu < 10$

3) $\alpha = 5\%$

4) $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{11 - 10}{4 / \sqrt{36}} = 6/4 = 1.5$

$|Z_{cal}| = 1.5$

$Z_\alpha = -1.64$

$|Z_{cal}| > Z_\alpha \Rightarrow$ Reject Null Hypothesis

3) A sample of 400 items is taken from the population whose s.d is 10. The mean of sample is 40. Test whether the sample das come from population within mean 38. (test of hypothesis). Also calc. 95% Conf. Inter.

Sol: $\bar{x} = 40$, $\mu = 38$, $\sigma = 10$, $n = 400$.

Test of hypothesis

1) $H_0 : \mu = 38$.

2) $H_1 : \mu \neq 38$.

3) $\alpha = 5\%$.

4) $Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{40 - 38}{10 / \sqrt{400}} = \dfrac{1}{5} \cdot 20 = 4$

$|Z_{cal}| = 4$

$|Z_\alpha| = 1.96$.

$|Z_{cal}| > |Z_\alpha| \Rightarrow$ Reject H_0.

$\therefore$ The sample is not from the population whose mean is 38

Confidence Interval: $(\bar{x} \pm E)$

$$E = Z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \quad \alpha/2 = 0.025$

$Z_{\alpha/2} = 1.96$

$$E = Z_{\alpha/2} \cdot \sigma/\sqrt{n} = 1.96 \cdot \left(\frac{10}{\sqrt{400}}\right)$$

$$= 0.98$$

$$C.I = (\bar{x} \pm E) = (40 \pm 0.98)$$

$$= (49.8, 30.2)$$

$$= (40.98, 39.02)$$

14) Test of Significance for difference of Mean of two large Samples :-

* Test for equality of two means :-

→ let $\bar{x}_1$ & $\bar{x}_2$ 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 stand deviations σ_1 and σ_2.

→ To Test weather two population means are equal.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

1) In a Survay of buying habits 400 women Shapers are selected at random in a Super Market A. located in certain Section of city their avg weekly food expenditure is ₹250. with std. dev of Rs 40. For 400 women shopers selected at Random in super M. B in another Section of the city. the avg weekly food expenditure is ₹220. with a standard deviation of Rs. 55. Test at 1% level of Significance weather the avg weekly food expenditure of two population Shopers are equally.

$n_1 = 400$

य : $\bar{x}_1 = 250$ $\sigma_1 = 40$ $n_2 = 400.$

$x_2 = 220$ $\sigma_2 = 58$

Step 1 : $H_0 : \mu_1 = \mu_2$

2. $H_1 : \mu_1 \neq \mu_2.$

3. $\alpha = 0.01.$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{30}{\sqrt{\frac{1600}{400} + \frac{3025}{400}}}$$

$$= \frac{30}{\sqrt{11.5625}} = \frac{30}{3.40036}$$

$$= 8.82$$

$|Z_{cal}| = |8.82|$.

$Z_\alpha = 2.58$

$|Z_{cal}| > Z_\alpha$ (Reject H_0)

Accept H_1

i.e., $\mu_1 \neq \mu_2$

The avg. weekly food expenditure of the two population of shoppers are not equal.

Test of Significance for single proportion (large Sample) :-

Suppose a large random sample of size n has a sample proportion mean P (∴ proportion of success)

To test the hypothesis that the proportion P in the population has specified value P_0.

Step-1 : Null Hypothesis H_0 : P

Step-2 : Alternative Hypothesis H_1 : $P \neq P_0$ or

$P < P_0$ or

$P > P_0$.

Step-3 ÷ level of significance α.

Step-4 ÷ test statistics : $Z = \dfrac{p - P_0}{S.E. \text{ of } p}$

$$= \frac{p - P_0}{\sqrt{\frac{PQ}{n}}}$$

where p = sample proportion.

P = popul. prop.

Step-5 ÷ Conclusion : If $|Z_{cal}| < Z_\alpha$ then accept H_0. If $|Z_{cal}| > Z_\alpha$ then accept H_1.

16) A Manufacturer claimed that atleast 95% of the equipments which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were ~~f~~aulty. Test his claime at 5% level of significance.

$p = 0.95$

Sol ÷ $n = 200$; $\alpha = 5\%$.

population proportion $P = 95\% = 0.95$.

$\Rightarrow Q = 0.05$;

No. of pieces conforming to Specification $= 200 - 18$
$= 182$.

Sample proportion $p =$ proportion of pieces confirming to specificion $= \dfrac{182}{200} = 0.91$.

Step-1 ÷ H_0 ÷ $p = P_0$; $P = 0.95$.

Step-2 ÷ H_1 ÷ $P < 0.95$.

Step-3 ÷ $\alpha = 0.05$.

Step-4 : $z = \dfrac{p - P_0}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.91 - 0.95}{\sqrt{\dfrac{(0.95)(0.05)}{200}}}$

$= -2.59$.

Step-5 : $|Z_{cal}| = |-2.59| = 2.59$.

$Z_\alpha = -1.645$.

$|Z_{cal}| > Z_\alpha$.

$\Rightarrow$ Reject H_0. $\Rightarrow$ Accept H_1

$\Rightarrow P < 0.95$

$\therefore$ The Manufacturer claim is Rejected.

) 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.

Sol : $n = 1000$; $P = 0.540$

population proportion P = proportion of rice eater

$P = 0.54 = \frac{1}{2}$

Sample proportion of rice eaters $= p = \frac{540}{1000} = 0.54$

$q = 0.46$

Step-1 : $H_0 : P = P_0 = Q = 1/2$

Step 2 : $H_1 : P \neq 1/2$

The picture can't be displayed.

The picture can't be displayed.

The picture can't be displayed.

The picture can't be displayed.

The picture can't be displayed.

The picture can't be displayed.

The picture can't be displayed.

Students Single Mean Test :-

Suppose we want to test

1) If a Random Sample x_i of Size 'm' has been has been drawn from a Normal population with a Specified Mean value

2) If the Sample Mean differs Significantly from the hypothetically value of population mean.

Step 1 : $H_0 : \mu = \mu_0$

Step-2 : $H_1 : \mu_1 \neq \mu_2$.

Step-3 : level of Significance (α).

degree of freedom (γ) $= n-1$;

Step-4 : Test statistics (t) $= \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

where s is the Sample Standard deviation.

Step-5 : Conclusion.

if $|t_{cal}| < t_\alpha$; Accept H_0.

if $|t_{cal}| > t_\alpha$; Accept H_1.

Note1: confidence Interval (C.I) $= \left[\bar{x} \pm t_{\alpha} \cdot S/\sqrt{n}\right]$.

Note 2: $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$.

Table 4 Values of t_{α}^{*}

ν	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	ν
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

Q A Sample of 26 bulbs gives a mean life of 990 hours with a stnd. dev. of 20 hrs. the Manufacturer claims that the Mean life of bulb is 1000 hrs. Is the Sample not upto the Standard.

Sol:- $n = 26 < 30$ (Sample - Small Sample).

$\bar{x} = 990$; $\mu = 1000$.

Step-1 : $H_0 : \mu = 1000$

Step-2 : $H_1 : \mu < 1000$.

Step-3 : $\alpha = 5\%$; $\nu = 26 - 1 = 25$.

Step-4 : $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{990-1000}{20/\sqrt{26}} = \dfrac{-1\cancel{0}}{2\cancel{0}}\sqrt{26}$

$= -0.5 \times \sqrt{26}$

$= 2.54$

$|t_{cal}| = |-2.54| = 2.54$.

$t_\alpha = 1.708$.

$t_{cal} > t_\alpha \Rightarrow$ Accept H_1

$\therefore$ company claiming is wrong.

25) A Random Sample of 6 Steel beams has a mean strength of 58392 PSI with a stand. dev. of 648 PSI. use this information and the level of significance $\alpha = 0.05$. To Test weather the true avg strength of the steel from which this sample came is 58000 PSI. also find class interval.

Sol:- $n = 6 < 30$ (Small Sample).

$\bar{x} = 58392$; $S = 648$; $\alpha = 5\%$.

$\mu = 58000$.

Step-1:- $H_0 : \mu_0 = 58000$

$H_1 : \mu_0 < 58000$

$\alpha = 0.05$; $\gamma = 5$;

$t_\alpha = 2.015$.

Step 4: $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{58392 - 58000}{648/\sqrt{6}} = \dfrac{392}{648}\sqrt{6}$

$= 0.604 \times 2.$

$= 1.479 = 1.$

$t_{cal} = 1.48$

$t_{cal} < t_\alpha$.

Students t - test for difference of mean:

Let $\bar{x}_1$ & $\bar{x}_2$ be the means of two independent samples of sizes n_1 and n_2 drawn from two normal populations having mean μ_1 and μ_2.
To test weather the two population means are equal.

Step 1: $H_0 : \mu_1 = \mu_2$

Step 2: $H_1 : \mu_1 \neq \mu_2$

Step 3: Level of significance 'α'

Step 4: Test statistics

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \rightarrow (n_1-1)s_1^2 + (n_2-1)s_2^2$$

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1}, \quad \bar{x}_2 = \frac{\Sigma x_2}{n_2}$$

Step 5: Conclusion.

If $|t_{cal}| < t_\alpha$ Accept Null hypothesis

If $|t_{cal}| > t_\alpha$ Reject H_0 i.e Accept H_1

NOTE.

→ Standard error of $(\bar{x}_1 - \bar{x}_2) = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

→ Confidence Interval $= (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\, s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

→ Degree of freedom (f) $= n + n_2 - 2$

Q) Samples of two types of electric bulbs were tested for length of life and following data were observed.

Type 1	Type 2
$n_1 = 8$	$n_2 = 7$
$\bar{x}_1 = 1234$	$\bar{x}_2 = 1036$
$s_1 = 36$ hrs	$s_2 = 40$ hrs

Is the difference in the mean that type 1 is superior to type 2 regarding length of length

Sol

Step 1: $\mu_1 = \mu_2$

Step 2: $\mu_1 > \mu_2$ (or) $(\mu_1 \neq \mu_2)$

Step 3: $\alpha = 0.05$, $t_\alpha = 1.833$ → $\gamma = 13$

Step 4:

$$t = \frac{1234 - 1036}{S\sqrt{\left(\frac{1}{8} + \frac{1}{7}\right)}}$$

$$t = \frac{198 \times \sqrt{56}}{40.731 \times \sqrt{15}} = 9.39$$

$\Rightarrow |t_{cal}| = 9.39$

Step 5:

$|t_{cal}| > t_\alpha$

Reject H_0, accept H_1

*** Q To examine the hypothesis that the husbands are more intelligent than wives, an investigator took sample of 10 couples and administrated a test which measures the IQ. The results are.

Husband	117	105	97	105	123	109	86	78	103	107
Wife	106	98	87	104	116	95	90	69	108	85

Test hypothesis with a reasonable test at 5% level of significance

Ans

Step 1: $H_0 : \mu_1 = \mu_2$

Step 2: $H_1 : \mu_1 \neq \mu_2$

Step 3: $\alpha = 0.05$, $\gamma = 10 + 10 - 2 = 18$

$t_\alpha = 1.734$

Step 4: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, $\bar{x}_1 = \dfrac{1030}{10} = 103$

$\bar{x}_2 = 95.8$

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

$$S_2 = \sqrt{\frac{10.2^2 + 2.2^2 + 8.8^2 + 8.2^2 + 20.2^2 + 0.8^2 + 5.8^2 + 26.8^2 + 12.2^2 + 10.8}{10}}$$

$$= \sqrt{167.96} = 12.989$$

$$S_1 = 12.67$$

$$S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{10 \times (12.96^2 + 12.67)^2}{20 - 2}}$$

$$= \sqrt{\frac{10}{18}(328.4905)}$$

$$= \cancel{18.925}\ 13.5090$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{7.2}{\sqrt{\cancel{13.18}\ 13.50\left(\frac{2}{10}\right)}}$$

$$= \frac{7.2\sqrt{5}}{\sqrt{13.501}} = 4.38$$

Step 5:

$$t_\alpha = (t_{0.05})_{18} = 1.734$$

$$|t_{cal}| = 4.38$$

$$|t_{cal}| > t_\alpha$$

Reject H_0, accept H_1

Paired sample t-test

Paired observations arise in many practical situations where each homogeneous experimental unit receives both population condition. As a result each experimental unit has a pair of observations.

eg: To test the effectiveness of the drugs. from 11% b.p is measured before and after the intake of certain drug. Here the individual person is the experimental unit and the 2 populations are bp's before and after the drugs is given.

→ If (x_1, y_1) ---- (x_n, y_n) with the data before and after are paired values of data.

→ We apply paired t-test to examine the significance of the difference of the two situations.

→ Let $d_i = x_i - y_i$ (or) $y_i - x_i$ then

Step 1: Null hypothesis $H_0 : \mu_1 = \mu_2$

Step 2: Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$

Step 3: Level of significance α.

Step 4: Test statistics.

Let $d_i = x_i - y_i$ or $= y_i - x_i$

$$t = \frac{\bar{d}}{S/\sqrt{n}}$$

where $\bar{d} = \Sigma d/n$.

$$s^2 = \frac{\Sigma(d_i - \bar{d})^2}{n-1}$$

8 Scores obtained in a shooting competition by 10 soldiers before and after intensive training are given.

before	67	24	57	55	63	54	56	68	33	43
after.	70	38	58	58	56	67	68	75	42	38

Test weather the intensive training is useful at 5% level of significance

Sol, Let us apply paird t-test.

Step 1: $H_0 : \mu_1 = \mu_2$, i.e no significant effect of the training.

Step 2: $H_1 : \mu_1 \neq \mu_2$, i.e intensive training is useful

Step 3: $\alpha = 0.05$

Step 4: $t = \dfrac{\bar{d}}{(s/\sqrt{n})}$ where $d_i = x_i - y_i$ or $y_i - x_i$

where $\bar{d} = \dfrac{\Sigma d}{n}$

$d_1 = 70 - 67 = 3$, $d_2 = 14$, $d_3 = 1$, $d_4 = 3$, $d_5 = -7$

$d_6 = 13$, $d_7 = 12$, $d_8 = 7$, $d_9 = 9$, $d_{10} = -5$

$\bar{d} = \frac{50}{10} = 5$

$s^2 = \frac{\Sigma(d_i - \bar{d})^2}{n-1} = \frac{1}{9}(2^2 + 9^2 + 4^2 + 2^2 + 12^2 + 8^2 + 7^2 + 2^2 + 4^2 + 10^2)$

$s^2 = 53.555$ ~~$\quad$~~ $s = 7.3181$

$t = \frac{\bar{d}}{s/\sqrt{n}}$

$= \frac{5}{7.318} \times \sqrt{10} = 2.160$

Conclusion: $t_\alpha = (t_{0.05})_9 = 1.83$

$|t_{cal}| = 1.999$

$|t_{cal}| > t_\alpha$

reject H_0, accept H_1

Q. The B.P of 5 women before and after intake of certain drug are given below:

before	110	120	125	132	125
after	120	118	125	136	121

Test weather there is significant change in blood pressure at 1% level of significance

Sol

Step 1: $H_0 : \mu_1 = \mu_2$

Step 2: $H_1 = \mu_1 \neq \mu_2$

Step 3: $\alpha = 0.01$, $n = 10$ $(\partial = 9)$

Step 4: $t = \dfrac{\bar{d}}{(s/\sqrt{n})}$

$d_1 = 10, d_2 = -2, d_3 = 0, d_4 = 4, d_5 = -4$

$$\bar{d} = \frac{8}{5} = 1.6$$

$$s^2 = \frac{8.4^2 + 3.6^2 + 1.6^2 + 2.4 + 5.6^2}{5-1}$$

$$= \frac{1}{4}(123.2) = 30.8$$

$$s = 5.549$$

$$t = \frac{1.6}{5.54/\sqrt{5}} = 0.64$$

Step 5:

$$t_\alpha = (t_{0.01})_4 = 3.747$$

$$|t_{cal}| < t_\alpha$$

Accept H_0

F-TEST :

In one sample of 10 observations from a normal population, the sum of the squares of the deviations of the sample values from the sample mean is 102.4 and in another sample of 12 observations from another normal population, the sum of the squares of the deviations of the sample values from the sample mean is 120.5. Examine whether the two normal populations have the 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.

Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_2^2$

Then the Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$

Given $n_1 = 10$, $n_2 = 12$, $\Sigma(x-\bar{x})^2 = 102.4$, $\Sigma(y-\bar{y})^2 = 120.5$

If S_1^2 and S_2^2 be the estimates of σ_1^2 and σ_2^2 then

$$S_1^2 = \frac{\Sigma (x - \bar{x})^2}{n_1 - 1} = \frac{102.4}{9} = 11.37$$

$$S_2^2 = \frac{\Sigma (y - \bar{y})^2}{n_2 - 1} = \frac{120.5}{11} = 10.95$$

The test statistic is $F = \dfrac{S_1^2}{S_2^2}$ $[\because \; S_1^2 > S_2^2]$

$$= \frac{11.37}{10.95} = 1.038$$

i.e, calculated $F = 1.132$

Tabulated value of F for (9, 11) d.f. at 5% level of significance is 2.91 (approx.)

Since calculated $F <$ tabulated F, we accept the null hypothesis H_0 i.e., the two normal populations have the same variance.

Q: 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 ?

Solution : Let the Null Hypothesis be $H_0 : \sigma_1^2 = \sigma_1^2$ where σ_1^2 and σ_2^2 are the variances of the two populations from with the samples are drawn.

The Alternative Hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Calculation of Sample Variances.

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$y - \bar{y}$	$(y - \bar{y})^2$
20	–2.3	5.29	27	–7.4	54.76
16	–6.3	39.69	33	–1.4	1.96
26	3.7	13.69	42	7.6	57.76
27	4.7	22.09	35	0.6	0.36
23	0.7	0.49	32	–2.4	5.76
22	–0.3	0.09	34	–0.4	0.16
			38	3.6	12.96
134		81.34	241		133.72

Given $n_1 = 6, \; n_2 = 7$

$$\therefore \; \bar{x} = \frac{\Sigma x}{n_1} = \frac{134}{6} = 22.3, \qquad \bar{y} = \frac{\Sigma y}{n_2} = \frac{241}{7} = 34.4$$

and $\Sigma(x_i - \bar{x})^2 = 81.34,\ \Sigma(y_i - \bar{y})^2 = 133.72$

If S_1^2 and S_2^2 be the estimates of σ_1^2 and σ_2^2, then

$$S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} = \frac{81.34}{5} = 16.26$$

and $$S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1} = \frac{133.72}{6} = 22.29$$

Let H_0 be true.

Since $S_2^2 > S_1^2$, the statistic is

$$F = \frac{S_2^2}{S_1^2} = \frac{22.29}{16.268} = 1.3699 \simeq 1.37$$

Tabulated value of F for (n_1-1, n_2-1) d.f. *i.e.*, (5, 6) d.f. at 5% level signifcance is 4.39.

Since calculated $F <$ tabulated F, we accept the null hypothesis H_0 at 5% level of significance *i.e.*, there is no significant difference between the variances of the time distribution by the workers.

Table 6 (a) Values of $F^{*}_{0.05}$

v_2 = degrees of freedom for denominator	v_1 = Degrees of freedom for numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.50	19.00	19.20	19.20	19.30	19.30	19.40	19.40	19.40	19.40	19.40	19.40	19.40	19.50	19.50	19.50	19.50	19.50	19.50
3	10.10	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.38	2.38	2.30	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	3.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.93
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Table 6 (b) Tabie Values of $F^{*}_{0.01}$

$_2$ = degrees of eedom for enominator	ν_1 = Degrees of freedom for numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,366
2	98.50	99.00	99.20	99.20	99.30	99.30	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.50	99.50	99.50	99.50	99.50	99.50
3	34.10	30.80	29.50	28.70	28.20	27.90	27.70	27.50	27.30	27.20	27.10	26.90	26.70	26.60	26.50	26.40	26.30	26.20	26.10
4	21.20	18.00	16.70	16.00	15.50	15.20	15.00	14.80	14.70	14.50	14.50	14.20	14.00	13.90	13.80	13.70	13.70	13.60	13.50
5	16.30	13.30	12.10	11.40	11.00	10.70	10.50	10.30	10.20	10.10	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.70	10.90	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.20	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.30	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.83
9	10.60	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.00	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.96	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

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 that there is no significant difference between the observed values and the corresponding expected values.

The 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 set of expected frequencies. Then the test statistic χ^2 is given by

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right] = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)}{E_2} + \ldots\ldots + \frac{(O_n - E_n)^2}{E_n}$$

Assuming that H_0 is true, the test statistic χ^2 follows Chi-squre distribution with $(n-1)$ d.f., where

$$\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i \text{ or } \sum_{i=1}^{n}(O_i - E_i) = 0$$

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 grouping 2 or more classes together before calculating $(O-E)$. Note that the degrees of freedom is determined with the number of classes after regrouping.

Q: 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 : The accidents conditions are different during the 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 $\chi^2 = \Sigma \dfrac{(O_i - E_i)^2}{E_i} = 26.6$ *i.e.*, Calculated $\chi^2 = 26.6$

Here $n = 10$ observations are given

$\therefore$ Degrees of freedom (d.f) = $n - 1 = 10 - 1 = 9$

Tabulated $\chi^2 = 16.9$

Since Calculated $\chi^2 >$ Tabulated χ^2, therefore, the Null Hypothesis is rejected and conclude that the accident conditions were not the same during the 10 week period.

Table 5 Values of χ_α^{2*}

$\nu \downarrow$ $\alpha \rightarrow$	$\alpha = 0.995$	$\alpha = 0.99$	$\alpha = 0.975$	$\alpha = 0.95$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\nu \downarrow$ $\alpha \leftarrow$
1	0.0000393	0.000157	0.000982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.832	15.056	16.750	5
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401	21
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796	22
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.484	36.415	39.364	42.980	45.558	24
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928	25
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645	27
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993	28
29	13.121	14.256	16.047	17.708	42.557	45.772	49.588	52.336	29
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672	30
40	20.706	22.164	24.433	26.509	55.758	59.342	63.691	66.766	40
50	27.991	29.707	32.357	34.764	67.505	71.420	76.154	79.490	50
60	35.535	37.485	40.482	43.118	79.082	83.298	88.379	91.952	60
70	43.275	45.442	48.758	51.739	90.531	95.023	100.425	104.215	70
80	51.172	53.540	57.153	60.391	101.879	106.629	112.329	116.321	80
90	59.196	61.754	65.646	69.126	113.145	118.136	124.116	128.299	90
100	67.328	70.065	74.222	77.929	124.342	129.561	135.807	140.169	100

Q: A pair of dice are 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 ?

Solution :

1. **Null Hypothesis** H_0 **:** The dice are fair
2. **Alternative Hypothesis** H_1 **:** The dice are not fair
3. **Level of significance :** $\alpha = 0.05$

The probabilities of getting a sum 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 are

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Calculations for χ^2

Sum	Observed Frequency (O_i)	Expected Frequency (E_i) $E_i = 360. P(x)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
2	8	10	4	0.4
3	24	20	16	0.8
4	35	30	25	0.833
5	37	40	9	0.225
6	44	50	36	0.72
7	65	60	25	0.417
8	51	50	1	0.02
9	42	40	4	0.1
10	26	30	16	0.53
11	14	20	36	1.8
12	14	10	16	1.6
	360	360		7.445

$$\therefore \quad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 7.445$$

The number of degrees of freedom = $n - 1 = 11$

The tabulated value of χ^2 for 11 d.f. at 5% level of significance is 18.3

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0.

i.e., The dice are fair.

Q: 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

Solution :

x	f	$f.x$
0	419	0
1	352	352
2	154	308
3	56	168
4	19	76
	$N = 1000$	$\Sigma fx = 904$

$$\text{Mean } \lambda = \frac{\sigma f_i \ x_i}{\Sigma f_i} = \frac{904}{1000} = 0.904$$

Theoretical distribution is given by

$$N \times p(x) \ = 1000 \times \frac{e^{-\lambda} \lambda^x}{x!}$$

Hence the theoretical frequencies are given by

$$f(x) = \frac{1000 \, e^{-0.904} (0.904)^x}{x\,!} = \frac{1000 \times 0.4049 \times (0.904)^x}{x\,!} \qquad \text{.... (1)}$$

Putting $x = 0, 1, 2, 3, 4$, we get

x	0	1	2	3	4	Total
f	404.9 (406.2)	366	165.4	49.8	11.3 (12.6)	997.4

In order that the total observed and expected frequencies may agree, we take first and last theoretical frequencies as 406.2 and 12.6 instead of 404.9 and 11.3 as shown in brackets.

$$\therefore \ \chi^2 = \frac{\Sigma(O_i - E_i)^2}{E_i}$$

$$= \frac{(419 - 406.2)^2}{406.2} + \frac{(352 - 366)^2}{366} + \frac{(154 - 165.4)^2}{165.4} + \frac{(56 - 49.8)^2}{49.8} + \frac{(19 - 12.6)^2}{12.6}$$

$$= 0.403 + 0.536 + 0.786 + 0.772 + 3.251$$

$$= 5.748$$

CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

Definition : Literally, an attribute means a quality or characteristic. Examples of attributes are drinking, smoking, blindness, honesty, beauty etc.

An attribute may be marked by its presence (position) or absence in a number of a given population. Let the observations be classified according to two attributes and the frequencies O_i in the different categories be shown in a two-way table, called contigency table. We have to test on the basis of cell frequences whether the two attributes are independent or not. We take the Null - Hypothesis H_0 that there is no association between the attributes *i.e.*, we assume that the two attributes are independent. The expected frequencies (E_i) of any cell $= \dfrac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$

The test statistic $\chi^2 = \sum_i \left[\frac{(O_i - E_i)^2}{E_i}\right]$ approximately follows Chi-square distribution with d.f. = (No. of rows − 1) × (No. of columns −1)

If the calculated value of χ^2 is less than the table value at a specified level (generally 5%) of significance, the hypothesis holds good i.e., the attributes are independent and do not bear any association. On the other hand, if the calculated value of χ^2 is greater than the table value at a specified level of significance, we say that the results of the experiment do not support the hypothesis, in other words, the attributes are associated.

Let us consider two attributes A and B. A is divided into two classes and B is divided into two classes. The various cell frequencies can be expressed in the following table known as 2×2 contigency table.

A	a	b
B	c	d.

a	b	$a + b$
c	d	$c + d$
$a + c$	$b + d$	$N = a + b + c + d$

The expected frequencies are given by

$E(a) = \frac{(a+c)(a+b)}{N}$	$\mathrm{E}(b) = \frac{(b+d)(a+b)}{N}$	$a + b$
$E(c) = \frac{(a+c)(c+d)}{N}$	$\mathrm{E}(d) = \frac{(b+d)(c+d)}{N}$	$c + d$
$a + c$	$b + d$	$N = a+b+c+d$

The value of χ^2 is given by $\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ where $N = a+b+c+d$ with $\text{d.f.} = (2-1)(2-1) = 1$. We use this formula when the expected frequencies are in fractions (or decimals).

Q: On the basis of information given below about the treatment of 200 patients suffering from a disease, state whether the new treatment is comparatively superior to the conventional treatment.

	Favourable	*Not favourable*	*Total*
New	60	30	90
Conventional	40	70	110

Solution :

Null Hypothesis H_0 : No difference between new and conventional treatment (or) New and conventional treatment are independent.

The number of degrees of freedom is $(2 - 1)(2 - 1) = 1$

Expected frequencies are given in the table :

$$\text{Expected frequency} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

$\frac{90 \times 100}{200} = 45$	$\frac{90 \times 100}{200} = 45$	90
$\frac{100 \times 110}{200} = 55$	$\frac{100 \times 110}{200} = 55$	110
100	100	200

Calculation of χ^2 :

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
60	45	225	5
30	45	225	5
40	55	225	4.09
70	55	225	4.09
200	200		18.18

$$\therefore \quad \chi^2 = \Sigma \frac{(O-E)^2}{E} = 18.18$$

Tabulated χ^2 for 1 d.f. at 5% level of significance is 3.841.

Since calculated $\chi^2 >$ tabulated χ^2 we reject the null hypothesis H_0 i.e., new and conventional treatment are not independent. The new treatment is comparatively superior to conventional treatment.

Q: Find if there is any significant correlation between the heights and weights given below.

Height in inches	57	59	62	63	64	65	55	58	57
Weight in lbs	113	117	126	126	130	129	111	116	112

Solution :

Height in inches x	Deviation from Mean (60) $X = x - \overline{x}$	Square of deviations X^2	Weight in lbs y	Deviations from Mean $Y = y - \overline{y}$	Square of deviations Y^2	Product of deviations of X and Y series (XY)
57	–3	9	113	–7	49	21
59	–1	1	117	–3	9	3
62	2	4	126	6	36	12
63	3	9	126	6	36	18
64	4	16	130	10	100	40
65	5	25	129	9	81	45
55	–5	25	111	–9	81	45
58	–2	4	116	–4	16	8
57	–3	9	112	–8	64	24
540	0	102	1080	0	472	216

Coefficient of correlation $r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \times \Sigma Y^2}}$

$$\therefore \quad r = \frac{216}{\sqrt{102 \times 471}} = 0.98$$

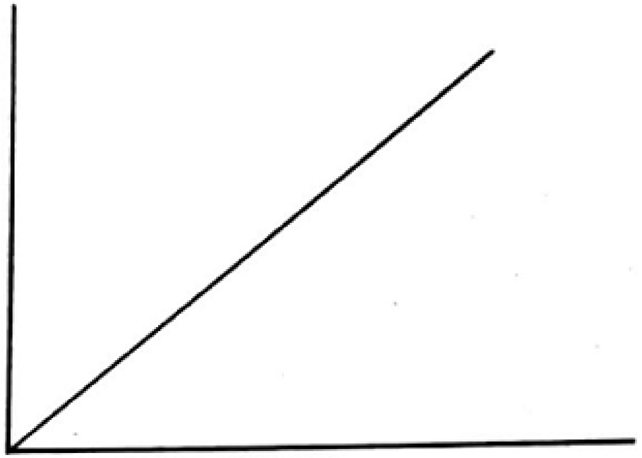
Note : Limits for correlation coefficient are $-1 \le r \le 1$.

Hence correlation coefficient can not exceed one numerically.

If $r = 1$ correlation is perfect and positive.

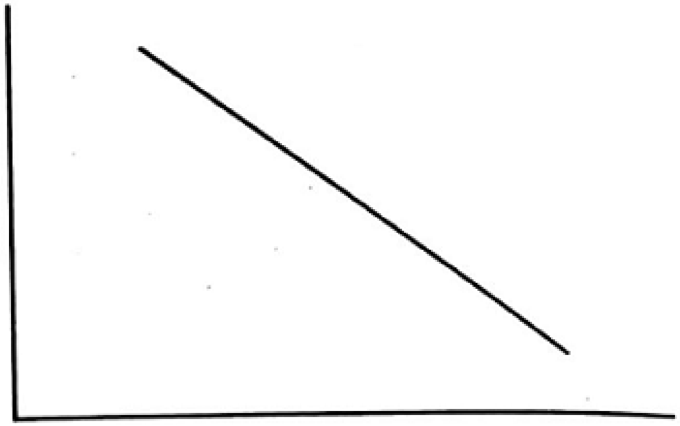
If $r = -1$ correlation is perfect and negative. If $r = 0$, then there is no relationship between the variables.

Perfect positive correlation



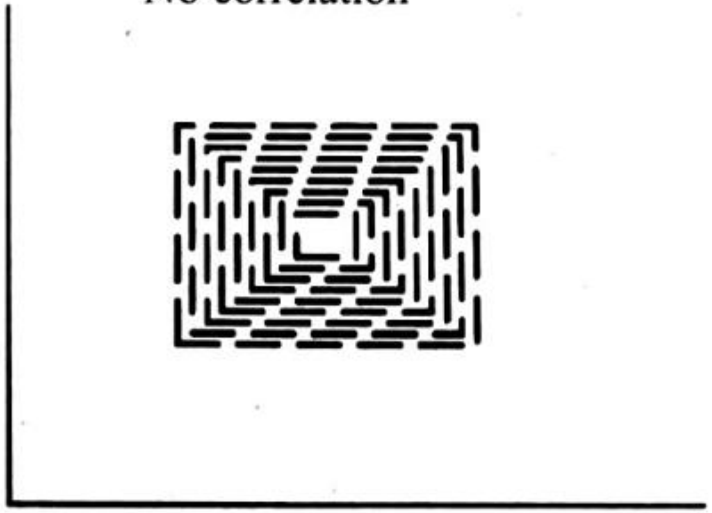
$r = +1$

Perfect Negative correlation

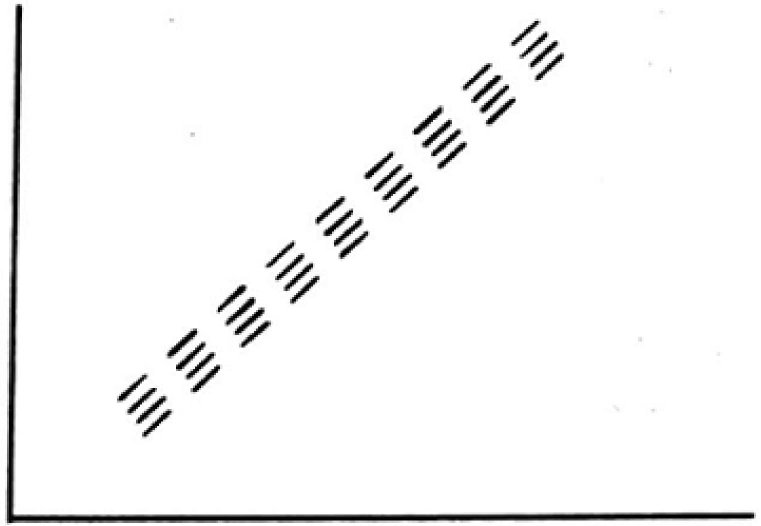


$r = -1$

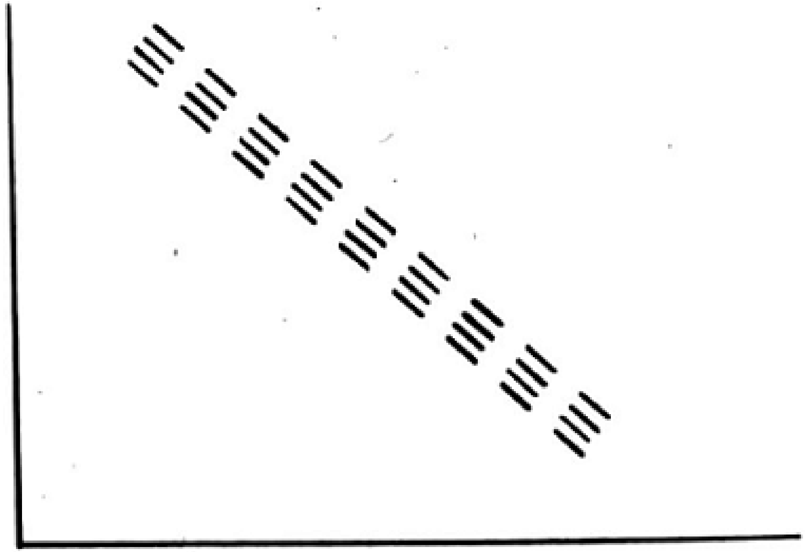
No correlation



High degree of positive correlation



High degree of negitive correlation



RANK CORRELATION COEFFICIENT

A British Psychologist Charles Adward Spearman found out the method of finding the coefficient of correlation by ranks. This method is based on rank and is useful in dealing with qualitative characteristics such as morality, character, intelligence and beauty. It can not be measured quantitatively as in the case of Pearson's coefficient of correlation. It is based on the ranks given to the observations. Rank correlation is applicable only to the individual observations. The formula for Spearman's rank correlation is given by -

$$\rho = 1 - \frac{6\,\Sigma D^2}{N(N^2-1)}$$

where $\rho \rightarrow$ Rank coefficient of correlation

$D^2 \rightarrow$ Sum of the squares of the differences of two ranks.

$N \rightarrow$ Number of paired observations

PROPERTIES OF RANK CORRELATION COEFFICIENT

1. The value of ρ lies between $+1$ and -1
2. If $\rho = 1$, there is complete agreement in the order if the ranks and the direction of the rank is same.
3. If $\rho = -1$ then there is complete disagreement in the order of the ranks and they are in opposite directions.

Procedure to solve problems :

1. When the ranks are given.

Step. 1. Compute the difference of two ranks and denote it by D.

Step. 2. Square D and get ΣD^2

Step. 3. Obtain ρ by substituting the figures in the formula.

2. When the ranks are not given, but actual data are given, then we must give ranks. We can give ranks by taking the highest as 1 or the lowest value as 1, next to the highest (lowest) as 2 and follow the same procedure for both the variables.

Q: Following are the rank 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

Solution :

Rank in statistics (x)	Rank in Mathematics (y)	$D = (x - y)$	D^2
1	2	-1	1
2	4	-2	4
3	1	$+2$	4
4	5	-1	1
5	3	$+2$	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	$+3$	9
10	8	$+2$	4
			$\Sigma D^2 = 40$

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)} = 1 - \frac{6\times 40}{10(10^2-1)} = 1 - \frac{240}{10(100-1)} = 1 - \frac{240}{990} = 1 - 0.24 = 0.76$$

Q: From the following data calculate the rank correlation coefficient after making adjustment for tied ranks.

X	48	33	40	9	16	16	65	24	16	57
Y	13	13	24	6	15	4	20	9	6	19

Solution : First we have to assign ranks to the variables.

X	Rank (x)	Y	Rank (y)	$D = x - y$	D^2
48	8	13	5.5	2.5	6.25
33	6	13	5.5	0.5	0.25
40	7	24	10	−3	9.00
9	1	6	2.5	−1.5	2.25
16	3	15	7	4	16.00
16	3	4	1	2	4.00
65	10	20	9	1	1.00
24	5	9	4	1	1.00
16	3	6	2.5	5	0.25
57	9	19	8	1	1.00
					$\Sigma D^2 = 41$

16 is repeated 3 times in X items hence $m = 3$. Since 13 and 6 are repeated twice in Y items; hence $m = 2$.

$$\therefore\ \rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^2 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right]}{N^3 - N}$$

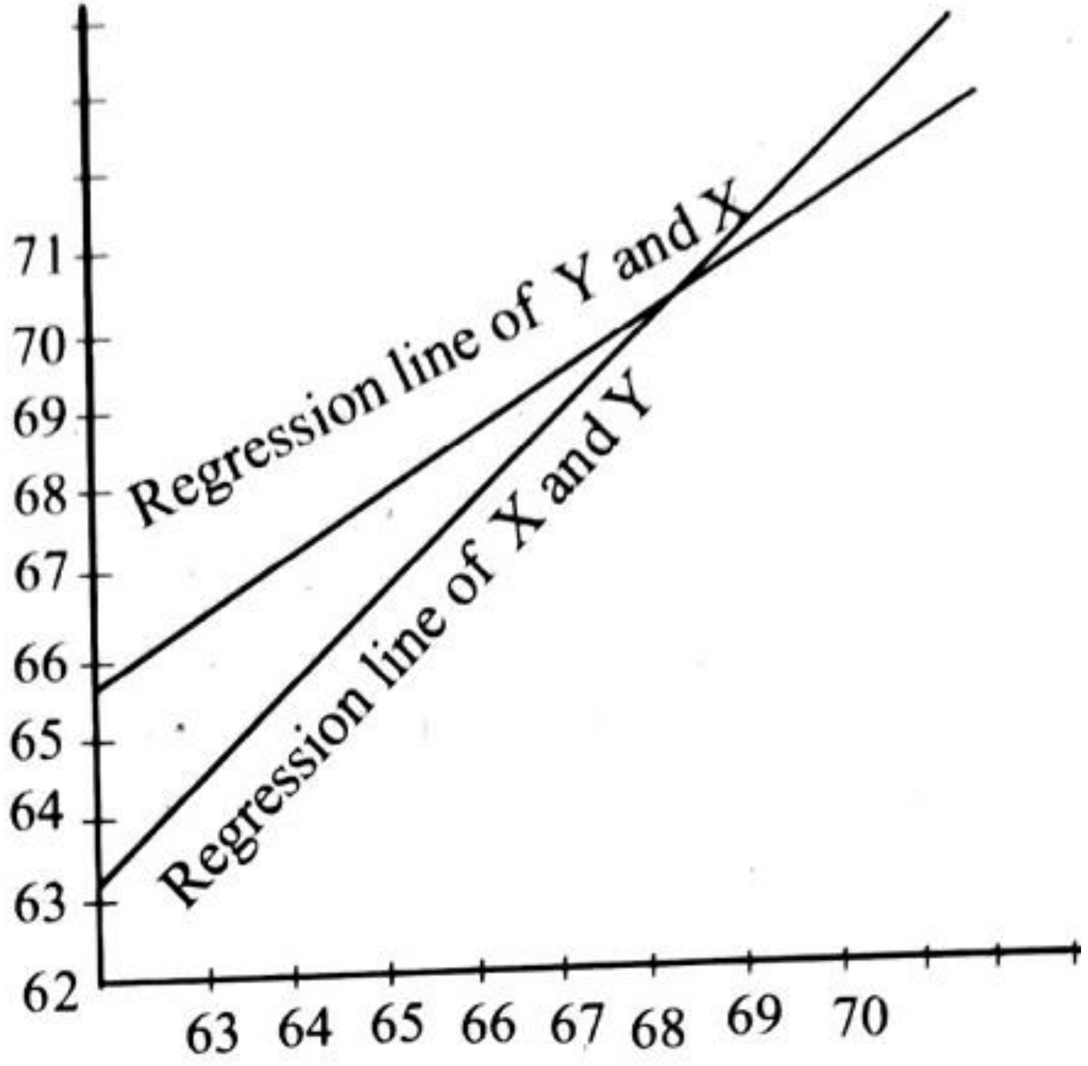
$$= 1 - \frac{6\left[41 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{990}$$

$$= +\,0.733$$

REGRESSION

e.g. Fit a regression line on the scatter diagram for the following data.

X	Y
65	68
67	68
62	66
70	68
67	67
69	68
71	70



Regression Line : A regression line is a straight line fitted to the data by the method of least squares. It indicates the best possible mean value of one variable corresponding to the mean value of the other. There are always two regression lines constructed for the relationship between two variables X and Y. Thus one regression line shows the regression of X upon Y and the other shows the regression of Y on X.

Regression equation of Y on X :

$$\Sigma Y = Na + b\Sigma X$$

$$\Sigma XY = a\,\Sigma X + b\,\Sigma X^2$$

Regression equation of X on Y :

Normal equations are

$$\Sigma X = Na + b\,\Sigma Y$$

$$\Sigma XY = a\,\Sigma Y + b\,\Sigma Y^2$$

Q: Determine the equation of a staight line which best fits the data.

X :	10	12	13	16	17	20	25
Y :	10	22	24	27	29	33	37

Solution : Straight line is $Y = a + bx$ The two normal equations are $\Sigma Y = b\,\Sigma X + N\,a$

$$\Sigma XY = b\,\Sigma X^2 + a\,\Sigma X$$

X	X^2	Y	XY
10	100	10	100
12	144	22	264
13	169	24	312
16	256	27	432
17	289	29	493
20	400	33	660
25	625	37	925
$\Sigma X = 113$	$\Sigma X^2 = 1938$	$\Sigma Y = 182$	$\Sigma XY = 3186$

Substituting the values, we get

$$113b + 7a = 182 \quad \ldots (1)$$

$$1983b + 113a = 3186 \quad \ldots. (2)$$

Solving (1) and (2) , we get $a = 0.82, b = 1.56$

Thus the equation of the straight line is $Y = a + bX$

$$\therefore \; Y = 0.82 + 1.56\, X$$

This is called the regression equation of Y on X.

THANK YOU