UNIT-II
PROBABILITY DISTRIBUTIONS

1. Discrete Probability Distributions:
   a) Binomial Distribution
   b) Poisson Distribution
   c) Geometric Distribution
2. Continuous Probability Distributions:
   a) Normal Distribution
   b) Uniform Distribution
   c) Exponential Distribution
   d) Gamma Distribution

TEXT BOOKS:

1. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, Probability & Statistics for Engineers & Scientists, 9th Ed. Pearson Publishers.
2. S C Gupta and V K Kapoor, Fundamentals of Mathematical statistics, Khanna publications

BINOMIAL DISTRIBUTION

- It is a Discrete P.D. and it is also called as Bernouli distribution because it was given by a famous Swiss mathematician "James Bernouli" in 1713.
- It is the distribution of Binomial(i.e., TWO) only Two features(outcomes) which are denoted by **p** – Success and **q** - Failure
- Total probability is i.e., $p + q = 1$

Characteristics of Binomial Distribution

- A finite number of trials.
- Each trial should have exactly two outcomes: success or failure.
- Trials should be independent.
- The probability of success or failure should be the same in each trial.

Note: The trails satisfying the above are known as Bernouli trails

Binomial Distribution(B.D)

- **Definition:** A discrete R.V. X is said have B.D. if it assumes only Non-negative values and with it's p.m.f given by
- $P(X = x) = p(x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$
- for $x = 0, 1, \dots, n$ and $P(X = x) = 0$ otherwise. Here, $q = 1 – p$.
- **Binomial Variate:** Any random variable X which follows B.D. is called a Binomial variate
- It is denoted by $X \sim B.D(n,p)$

FORMULAE:

- The General Binomial Probability Formula:
  P(k out of n) = (*n!*/**k!(n-k)!) x** $p^k(1-p)^{(n-k)}$
- Mean value of B.D: $\mu = np$
- Variance of B.D: $\sigma^2 = np(1-p)$
- Standard Deviation of B.D: $\sigma = \sqrt{np(1-p)}$
- Mode of B.D:
  i) If $(n+1)p = (\mathbf{x_1}+f)$ **Not an integer** then distribution has a unique mode i.e., $\mathbf{x_1}$
  ii) If $(n+1)p =$ (**x**) an **integer** then distribution has a Bi-modal i.e., x-1 and x

Ex: if n=9 and p=0.4 then (n+1)p = (9+1)0.4
= 4 which is an integer.
Therefore it has two modes i.e., 4 and 4-1 = 3

Ex: if n=7 and p=0.6 then (n+1)p = (7+1)0.6
= 4.8 which is not an integer.
Therefore it has Unique mode i.e., 4

Ex:1 what is the probability that in a family of 4 children there will be at least 1 boy?

- Solution :

  n= 4; p=0.5 and q=0.5 (since prob. Is same)

  $P(X \geq 1) = ?$

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

  $= 1 - P(X = 0)$

  $= 1 - [\, 4C_0 * (0.5)^0 * (0.5)^4$

  $= 1 - [0.0625]$

  $= 0.9375$

Ex:2 If the probability that a new born child is a male is 0.6 then find the probability that in a family of 5 children that there are exactly 3 boys?

- Solution :

  $n= 5; p=0.6$ and $q=0.4$

  $P(X =3)= ?$

  $P(X =3)= 5C_3 * (0.6)^3 * (0.4)^2$

  $= 10* (0.216)* (0.16)$

  $= 0.3456$

Example:3 Ram says "70% choose chicken, so 7 of the next 10 customers should choose chicken" ... what are the chances Ram is right?

- **Solution**:
- p = 0.7; n = 10; k = 7
- And
- $P(X=K)= p^k(1-p)^{(n-k)}$
- $=0.7^7(1-0.7)^{(10-7)}$
- $=0.7^7(0.3)^{(3)}$
- =**0.0022235661**
- That is the probability of each outcome.
-
- And the total number of those outcomes is:
- *n!***k!(n-k)!** *=10!***7!(10-7)!**
- *=10×9×8×7×6×5×4×3×2×1***7×6×5×4×3×2×1 × 3×2×1**
- *=10×9×8***3×2×1**
- =**120**
- And Number of outcomes we want Probability of each outcome IS
- 120 × 0.0022235661 = 0.266827932

- So the probability of 7 out of 10 choosing chicken is only about **27%**
-

Poisson Distribution(P.D)

- It is a Discrete P.D. and it was given by a famous English mathematician “Simeon D. Poisson” in 1837.
- It is used to explain the behavior of a discrete R.V. where the probability of occurrence of an event is very small and total no. of possible outcomes is sufficiently large.
- It is the limiting case of Binomial distribution under the conditions that

    1. Number of trials is infinitely large i.e., $n \to \infty$
    2. Probability of success is very small i.e., $p \to 0$
    3. $np = \lambda$ (finite)

- Total probability is i.e., $p + q = 1$

Poisson Distribution(P.D)

- **Definition:** A discrete R.V. X is said have P.D. if it assumes only Non-negative values and with it's p.m.f given by

$$P(X = r) = \frac{e^{-\lambda}\lambda^{r}}{r!}$$

- for r = 0, 1, … ,∞
- **Poisson Variate:** Any random variable X which follows P.D. is called a Poisson variate
- It is denoted by X ~ P.D(λ)

FORMULAE:

- The General Poisson Probability Formula:
- Mean value of P.D: $\mu = np = \lambda$
- Variance of B.D: $\sigma^2 = \lambda$
- Standard Deviation of B.D: $\sigma = \sqrt{\lambda}$
- Mode of P.D:

  i) If $\lambda = k+f$ is **<u>Not an integer</u>** then distribution has a unique mode i.e., k

  ii) If λ= (**x**) an **<u>integer</u>** then distribution has a Bi-modal i.e., x-1 and x

Ex: if n=10 and p=0.5 then (np) = (10)0.5

= 5 which is an integer.

Therefore it has two modes i.e., 5 and 5-1 = 4

Ex: if n=7 and p=0.6 then (np) = (7)0.6

= 4.2 which is not an integer.

Therefore it has Unique mode i.e., 4

EX:1 There are 50 misprints in a book which has 250 pages. Find the probability that page 100 has no misprints.

- SOLUTION:
- The average number of misprints on a page is $\lambda = 50/250 = 0.2$.
- Therefore, if we let X be the random variable denoting the number of misprints on a page.
- X will follow a Poisson distribution with parameter 0.2 .

$$P(X = r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

- $P(X = 0) = \dfrac{(e^{-0.2})(0.2^0)}{0!}$
- $= \underline{0.819}$

Example 2:

A bank is interested in studying the number of people who use the ATM located outside its office late at night. On average, 1.6 customers walk up to the ATM during any 10 minute interval between 9pm and midnight.

a) What is the probability of exactly 3 customers using th ATM during any 10 minute interval?

- Solution: λ = 1.6

  P(X=3) = ?

-

- $$P(X = r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

  $$\frac{e^{-1.6}\ (1.6)^3}{3!}$$

-

- $$= \frac{0.2019(4.096)}{6}$$

- =0.1378

EX: 3 If the probability that an individual suffer a bad reaction from certain injection is 0.001 then find the probability that out of 2000 more than 1 suffers?

- Solution:

- GIVEN : $n= 2000$, $p = 0.001$

$$\lambda = n.p = 2000(0.001) = 2$$

To find $P(X>1) = ?$

$$P(X>1) = P(X=2) + P(X=3) + \dots\dots\dots + P(X=2000)$$

$$= 1 - [\, P(X\leq 1)]$$

$$= 1 - [P(X=0) + P(X=1)\,]$$

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

$$= 1 - [e^{-2} + (e^{-2})(2^{1})]$$

$$= 1 - [\,0.406\,]$$

$$= 0.594$$

EX:4 Average number of accidents on a day on a highway is 1.6. find the probability number of accidents are at most one?

- Solution:
- GIVEN : $\lambda = 1.6$

To find $P(X \leq 1) = ?$

$P(X \leq 1) = P(X=0) + P(X=1)$

$$= \left[\frac{(e^{-1.6})(1.6^{0})}{0!} + \frac{(e^{-1.6})(1.6^{1})}{1!}\right]$$

$$= \left[e^{-1.6} + (e^{-1.6})(1.6^{1})\right]$$

$$= [e^{-1.6}]\,(1 + 1.6)$$

$$= 0.2019(2.6)$$

$$= 0.5249$$

GEOMETRIC DISTRIBUTION

- Geometric distribution is similar to Binomial distribution, but instead of tossing a coin for a fixed no. of times and observing for a "H".
- We toss the coin and count the number of times "TOSSES" until a "H" appears.
- **Definition:** A discrete Random variable X is said to have G.D. if it's p.m.f is given by
- $P(X = r) = p. q^{r-1}$
- For r = 1,2, … where p is success, and $q = 1 – p$.
- r is the number of trails required until to get a first success.
- NOTE: Mean of G.D is 1/p
- Variance of G.D is q/p^2

EX:1 A die is tossed until a '6 'appears. Find the probability that it must be lasts more than "5" times?

- **Solution:**
- Let p be the probability of getting "6"
- $\Rightarrow$ p= 1/6
- Let q be the probability of Not getting "6"
- $\Rightarrow$ q= 5/6
- Let X be the no. times a die is tossed until a "6" appears
- By G.D $P(X = r) = p.\ q^{r-1}$ where r = 1,2,3,..
- $P(X>5) = ?$
- $P(X>5) = 1- P(X\leq5) = 1 - [\ P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)\ ]$

$$= 1 - [p.\ q^0 + p.\ q^1 + p.\ q^2 + p.\ q^3 + p.\ q^4\ ]$$

$$=1 - p\ [\ 1 + q^1 + q^2 + q^3 + q^4\ ]$$

$$= 1 - 1/6[\ 1+ (5/6)^1 + (5/6)^2 + (5/6)^3 + (5/6)^4\ ]$$

$$=1 - 0.59$$

$$= 0.41$$

-

EX:2 Products produced by a machine has a 3% defective rate. What is the probability that the first defective occurs in the fifth item inspected?

- **Solution:**
- Let p be the probability of getting defective
- $\Rightarrow$ p= 3% = 0.03
- Let q be the probability of Not defective
- $\Rightarrow$ q= 97% = 0.97
- Let X be the no. times a defective appears
- By G.D $P(X = r) = p.\ q^{r-1}$ where r = 1,2,3,..

$$P(X = 5) = p.\ q^{5-1}$$

$$= (0.03).\ (0.97)^4$$

$$= 0.0265$$

Continuous Probability Distributions

Uniform Distribution(Rectangular Distribution)

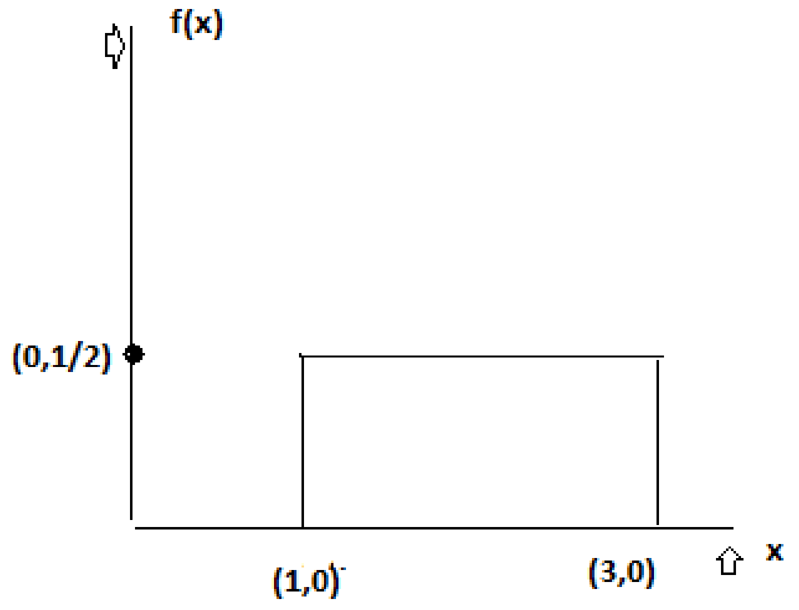
It is one of the simplest one in continuous distributions in all of statistics .This distribution is characterized by density function that is “flat” and the probability is uniform in a closed interval , [A,B]

Definition

The density function of the continuous uniform random variable X on the interval [A,B] is

$$f(x) = \begin{cases} \dfrac{1}{B-A} & ; \quad A \le x \le B \\ 0 & ; \quad elsewhere \end{cases}$$

The density function for a uniform random variable on the interval [1,3] is shown below



EXAMPLES

Let metro trains on a certain line run every half hour between mid night and six in the morning

Selecting a card from a deck of cards has a **uniform distribution**.

This is because an individual has an equal chance of drawing a spade, a heart, a club, or a diamond.

Tossing a coin has a **uniform distribution**.

This is because the likelihood of getting a tail or head is the same.

- **The application of this distribution is based on the assumption that the probability falling in an interval of fixed length within [A,B] is constant.**

Theorem

The mean and variance of the uniform distribution are

$$\mu = \frac{A+B}{2}$$

$$\sigma^2 = \frac{(B-A)^2}{12}$$

Proof

$$E(X) = \int_{-\infty}^{-\infty} x\ f(x)\ dx = \int_{A}^{B} x \frac{1}{B-A}\ dx =$$

$$\frac{1}{B-A}\int_{A}^{B} x\ dx = \frac{1}{B-A}\left(\frac{x^2}{2}\right)_{A}^{B}$$

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

$$= \frac{(B-A)(B+A)}{2(B-A)} = \left(\frac{B+A}{2}\right)$$

$$\mu = \frac{B+A}{2}$$

$$V((X) = E(X^2) - (E(X))^2$$

$$= \int_A^B x^2 f(x)\,dx - \left(\frac{A+B}{2}\right)^2$$

$$= \int_A^B x^2 \frac{1}{B-A}\,dx - \left(\frac{A+B}{2}\right)^2$$

$$= \frac{1}{B-A}\int_A^B x^2\,dx - \left(\frac{A+B}{2}\right)^2$$

$$= \frac{1}{B-A}\left(\frac{x^3}{3}\right)_A^B - \left(\frac{A+B}{2}\right)^2$$

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

$$=\frac{1}{B-A}\left[\frac{(B-A)(B^2+AB+A^2)}{3}\right]-\left(\frac{A+B}{2}\right)^2$$

$$=\left[\frac{(B^2+AB+A^2)}{3}\right]-\left(\frac{A^2+2AB+B^2}{4}\right)$$

$$=\left[\frac{4(B^2+AB+A^2)-3(A^2+2AB+B^2)}{12}\right]$$

$$=\left[\frac{A^2-2AB+B^2}{12}\right]=\left(\frac{(A-B)^2}{12}\right)$$

$$=\left(\frac{(B-A)^2}{12}\right)$$

Problem

Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. Both long and short conference occur quite often.

a)What is the probability density function?

b) What is the probability that any given conference lasts at least 3hours?

Solution

a) Probability density function is

$$f(x) = \begin{cases} \dfrac{1}{4-0} & ; \quad 0 \leq x \leq 4 \\ 0 & ; \quad otherwise \end{cases}$$

b)

$$P[X \geq 3] = \int_3^4 f(x)dx = \int_3^4 \frac{1}{4} dx = \frac{1}{4}(x)_3^4 = \frac{1}{4}$$

Problem

A random variable 'X' has a uniform distribution over (-3,3). find 'k' for which P(X>k) =1/3 . Also evaluate P(X<2) and

$$P\left[\left|X-2\right|<2\right]$$

Solution

1) **Probability density function of "X" is**

$$f(x) = \frac{1}{b-a} = \frac{1}{3-(-3)} = \frac{1}{6}$$

2)

$$P[\,X > k] = 1 - P[X \le k] = 1 - \int_{-3}^{k} f(x)\,dx$$

$$= 1 - \frac{1}{6}\int_{-3}^{k} dx = 1 - \frac{1}{6}(k+3) = \frac{1}{3}$$

$$\Rightarrow k = 1$$

$$3)\,P[\,X < 2\,] = \int_{-3}^{2} f(x)\,dx = \frac{1}{6}\int_{-3}^{2} dx = \frac{5}{6}$$

$$4)\,P\big[|X-2| < 2\big] = P[\,(2-2) < X < (2+2)\,] = P[0 < X < 4] = \frac{1}{6}\int_{0}^{3} dx = \frac{3}{6}$$

PROBLEM

You have been informed that the assessor will visit your home sometime between 10:00 am and 12:00 pm. It is reasonable to assume that his visitation time is uniformly distributed over the specified two-hour interval. Suppose you have to run a quick errand at 10:00 am.

a) If it takes 30 minutes to run the errand, what is the probability that you will be back before the assessor visits

b) If it takes 60 minutes to run the errand, what is the probability that you will be back before the assessor visits?

SOLUTION.

(a)

An accessor is supposed to visit your home sometime between 10 am and 12 pm.

Then, let X represent the minutes after 10 am that the accessor arrives at the location.

Since we need to run a 30-minute errand at 10 am, we are going to calculate the probability that the accessor does not arrive until 10:30 A.M.

X will follow a uniform probability distribution function.

$$f(x) = \frac{1}{B-A} \qquad for\ A \le X \le B$$

$$f(x) = \frac{1}{120-0} \qquad for\ 0 \le X \le 120$$

$$f(x) = 0.0083 \qquad for\ 0 \le X \le 120$$

$$P(X>30)=\int_{30}^{120} f(x)\,dx=\int_{30}^{120} 0.0083\,dx=$$

$$=0.0083\left(x\right)_{30}^{120}$$

$$0.0083\times(120-30)=0.75$$

(b)

Now, according to the question, we need to run a 60-minute errand.

Thus, we need to find the probability that the accessor does not visit until 11 A.M.

P(X>60)= $\int_{60}^{120} f(x)\,dx = \int_{60}^{120} 0.0083\,dx =$

$= 0.0083\left(x\right)_{60}^{120}$

=0.0083×(120−60)=0.50

PROBLEM

Suppose in a quiz there are 30participants. A question is given to all 30 participants and the time allowed to answer it is 25 seconds. Find the probability of participants responds within 6 seconds? Find the number of participants likely to answer it in 6 seconds?

Solution:

Given
Interval of probability distribution =[0,25]
Density of probability = 1/25
Interval of probability distribution of successful event = [0 seconds, 6seconds]

The probability $P(x<6)$ = 6/25
There are 30 participants in the quiz
Hence the number of participants likely to answer it in 6 seconds = 6/25
$6/25 \times 30 \approx 7$

PROBLEM

Let metro trains on a certain line run every half hour between mid night and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least twenty minutes

SOLUTION

Let x denotes the waiting time(in minutes) for the next train,under the assumption that a man arrives at random at the station.X is distributed uniformlyon (0,30) with p.d.f.

$f(x)= 1/30$,$0<x<30$

$= 0$ Otherwise

The probability that he has to wait at least 20 minutes is given by $P(X>20)= 1/30\int 1.dx$ 20 to 30

$=1/30(30-20) =1/3$

EXERCISE PROBLEMS

Q1.Suppose that you arrive at a bus stop randomly, so all arrival times are equally likely. The bus arrives regularly every 70 minutes without delay. What is the expected value of your waiting time?

Q2.According to a recent economist report, an adult driver in America spends between \$500and \$2500 on gasoline per year. Consider this spending pattern to be uniformly distributed.

a. What is the mean amount spent on gasoline?

b. What is the standard deviation of the amount spent?

c. What is the probability of a randomly selected adult spending less than \$1000 on gasoline?

d. What is the probability of a randomly selected adult spending more than \$2200 on gasoline?

NORMAL DISTRIBUTION

Normal distribution is a continuous probability distribution for a real-valued random variable.

Normal distribution, also known as the Gaussian distribution

Example

Heights, Blood pressure, Measurement error, IQ score, Stock market, Shoe size, Student's Average Report follow the normal distribution

Definition

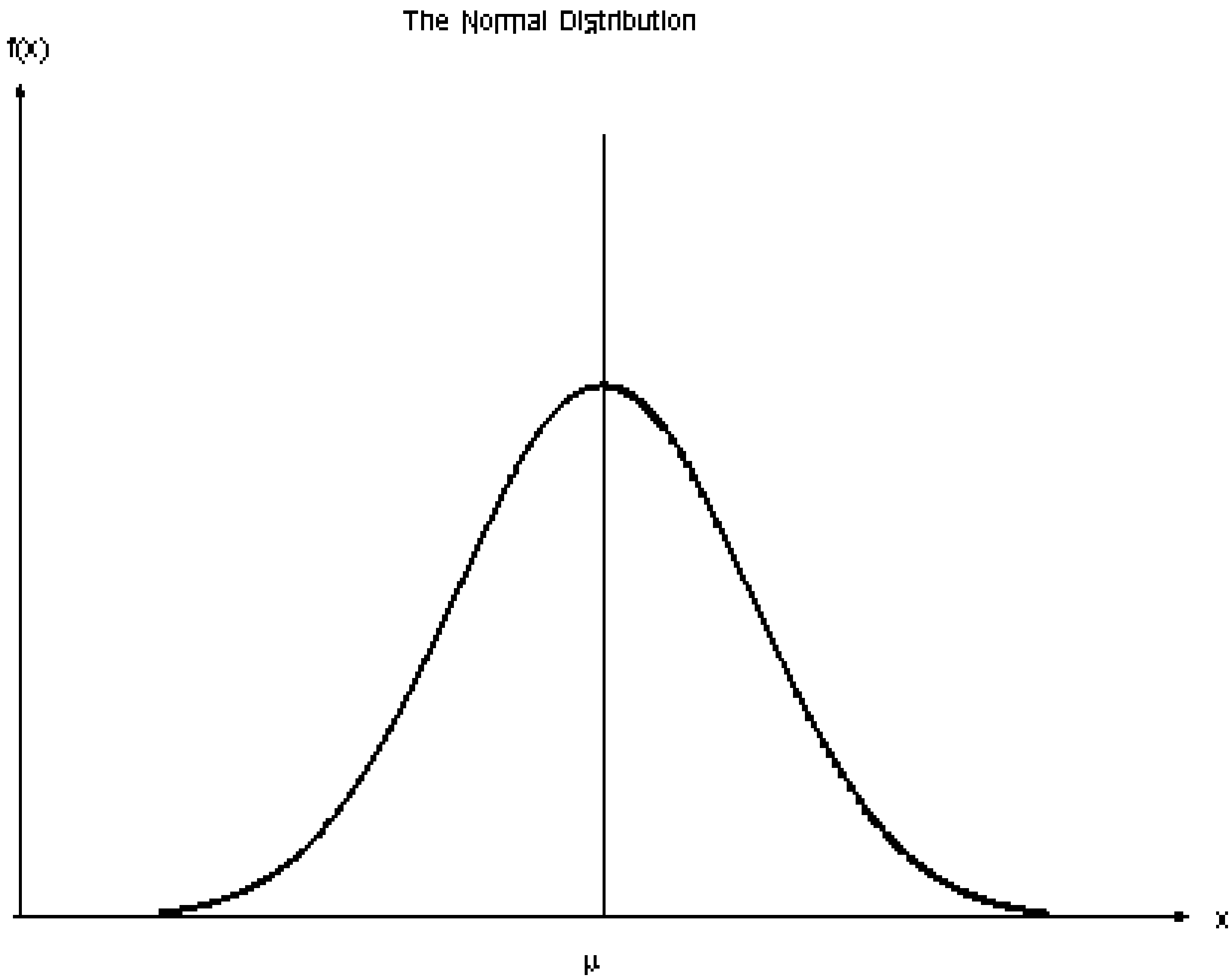
A random variable "X" is said to have a normal distribution , if its density function is given by

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

where $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$

where μ is the mean , σ is the standard deviation ,which are the two parameters of the normal distribution.

The random variable "X" is then a <u>normal random variable.</u>



The Normal Distribution
f(x)
x
μ

Mean of Normal Distribution

Mean =

$$E[X] = \int_{-\infty}^{\infty} x \, f(x)\, dx$$

$$E[X] = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$E[X] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \, e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

put $z = \left(\dfrac{x-\mu}{\sigma} \right)$

so that $dz = \dfrac{dx}{\sigma}$

$$E[X] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \sigma \, dz$$

$$E[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{\frac{-z^2}{2}} \, dz$$

$$E[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z) e^{\frac{-z}{2}^2} dz + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu e^{\frac{-z}{2}^2} dz$$

$$E[X] = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (z) e^{\frac{-z}{2}^2} dz + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-z}{2}^2} dz$$

$$E[X] = 0 + \frac{\mu}{\sqrt{2\pi}} 2\int_{0}^{\infty} e^{\frac{-z}{2}^2} dz$$

$$E[X] = \frac{\mu}{\sqrt{2\pi}} 2\int_{0}^{\infty} e^{\frac{-z}{2}^2} dz$$

$$E[X] = \frac{2\mu}{\sqrt{2\pi}}\left(\sqrt{\frac{\pi}{2}}\right) \quad = \mu$$ (using gamma function)

Mean = μ

$$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx$$

Variance of normal distribution

Variance = $E\left[(X-\mu)^2\right]$

= $\displaystyle\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

= $\displaystyle\int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

put $z = \left(\dfrac{x - \mu}{\sigma} \right)$ so that $dz = \dfrac{dx}{\sigma}$

$$= \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^2 e^{\frac{-z^2}{2}} \sigma \, dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^2 e^{\frac{-z^2}{2}} dz$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (z)^2 e^{\frac{-z^2}{2}} dz$$

since integrand is even function

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^{\infty} (z)^2 e^{\frac{-z^2}{2}} dz$$

put $\frac{z^2}{2} = t$ so that $dz = \frac{dt}{\sqrt{2t}}$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^{\infty} 2t\, e^{-t} \frac{dt}{\sqrt{2t}}$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} \sqrt{t}\, e^{-t} dt$$

using Gamma function

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty e^{-t}\, t^{\frac{3}{2}-1} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right)$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2} \Gamma\left(\frac{1}{2}\right)$$

$\Gamma(n) = (n-1)\Gamma(n-1)$

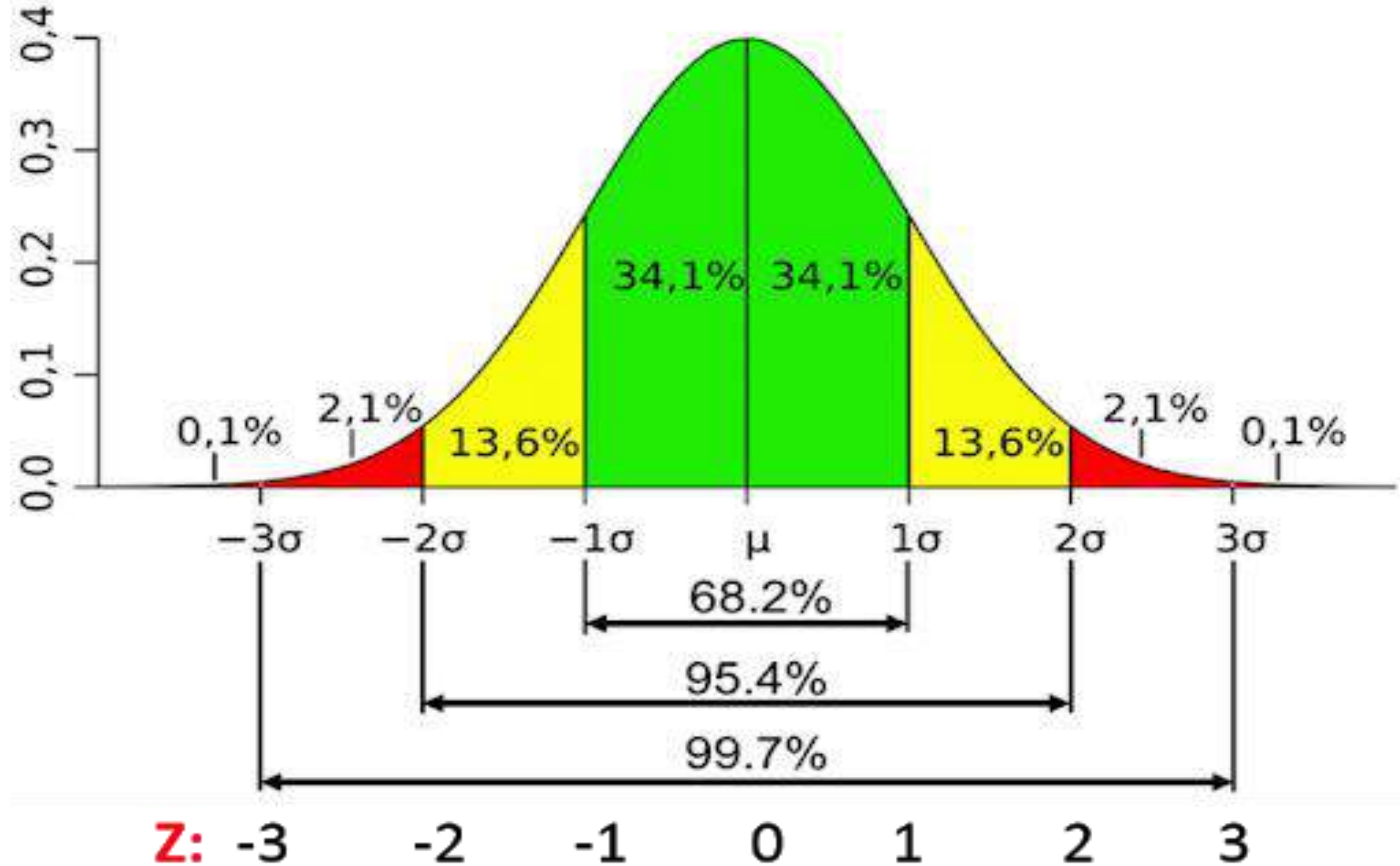
$$= \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2} \sqrt{\pi}$$

$$= \sigma^2$$

The Normal Distribution

Features of Normal Distribution

1. Normal distributions are symmetric around their mean.
2. The mean, median, and mode of a normal distribution are equal.
3. The area under the normal curve is equal to 1.0.
4. Normal distributions are denser in the center and less dense in the tails.
5. Normal distributions are defined by two parameters, the mean (μ) and the standard deviation (σ).
6. 68% of the area of a normal distribution is within one standard deviation of the mean.
7. Approximately 95% of the area of a normal distribution is within two standard deviations of the mean.


0,4
0,3
0,2
0,1
0,0
34,1%
34,1%
13,6%
13,6%
2,1%
2,1%
0,1%
0,1%
−3σ
−2σ
−1σ
μ
1σ
2σ
3σ
68.2%
95.4%
99.7%
Z:
-3
-2
-1
0
1
2
3

Area under Normal Curve

The curve of any continuous probability distribution or density function is constructed so that the area under the curve bounded by the two ordinates $x = x_1$ and $x = x_2$ equals the probability that the random variable X assumes a value between $x = x_1$ and $x = x_2$

$$P(X_1 < X < X_2) = \int_{X_1}^{X_2} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

put $z = \frac{x-\mu}{\sigma}$

$$P(Z_1 < Z < Z_2) = \frac{1}{\sqrt{2\pi}} \int_{Z_1}^{Z_2} e^{\frac{-z^2}{2}} \, dz$$

The distribution of a normal random variable with mean 0 and variance 1 is called a standard normal distribution.

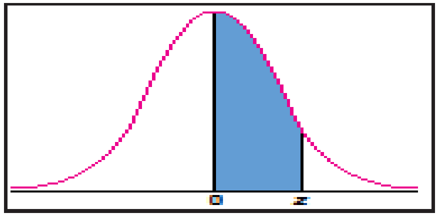


TABLE A Areas of a Standard Normal Distribution (Alternate Version of Appendix I Table 4)

The table entries represent the area under the standard normal curve from 0 to the specified value of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999

For values of z greater than or equal to 3.70, use 0.4999 to approximate the shaded area under the standard normal curve.

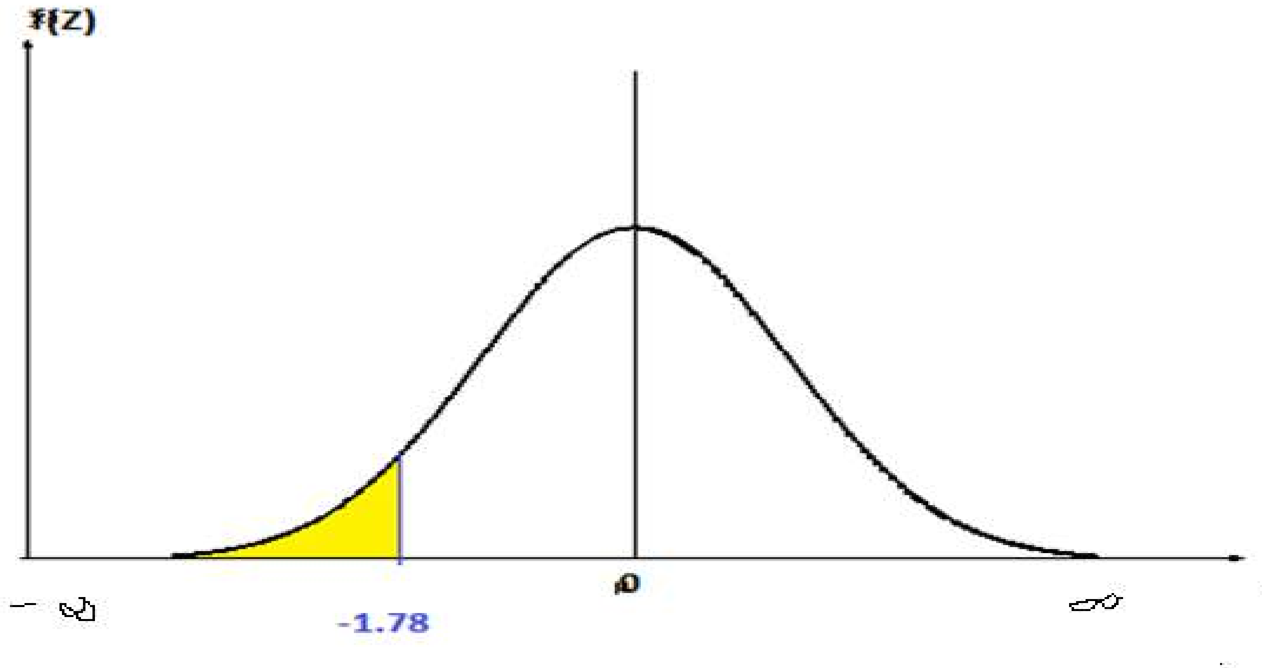
PROBLEM

If X is a normal variate , find the area

1) to the left of $z = -1.78$
2) to the right of $z = -1.45$
3) corresponding to $-0.8 \leq z \leq 1.53$
4) to the left of $z = -2.52$ and to the right of $z = 1.83$

SOLUTION

1)

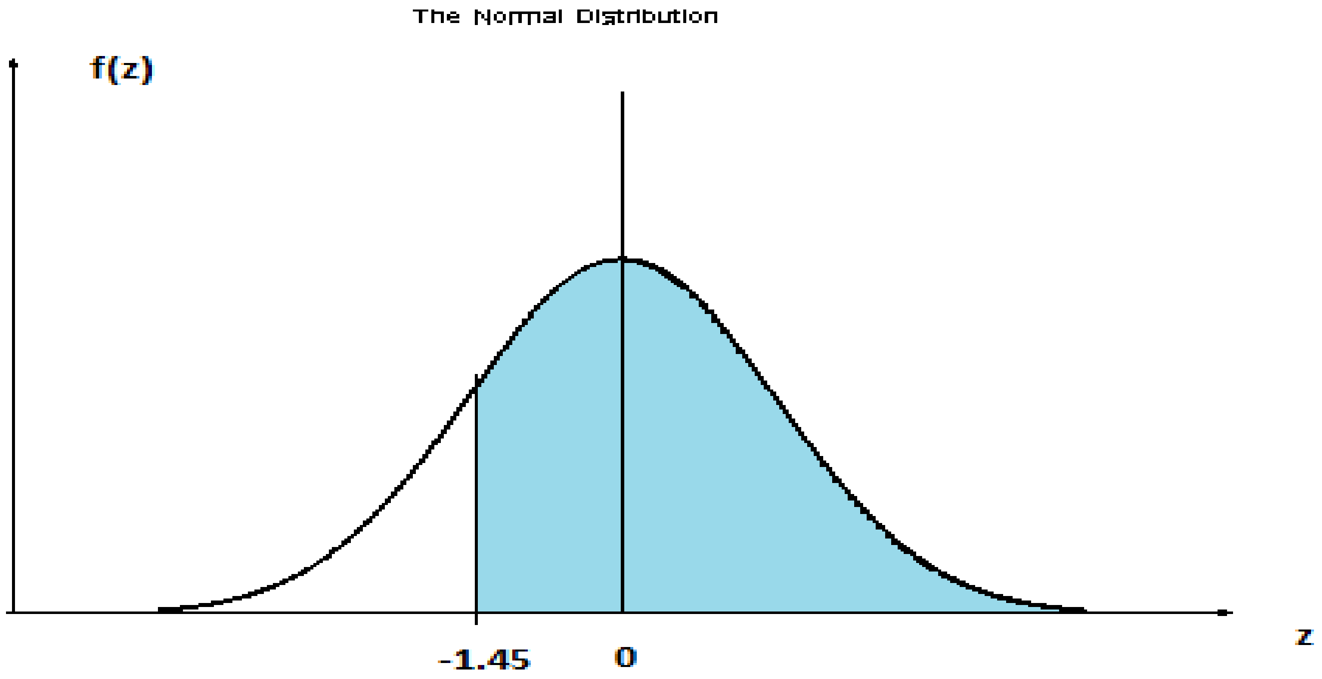


Required area ,

A = 0.5-Area(0 to -1.78)

=0.5-Area(0 to 1.78) (By symmetry)

=0.5-0.4625=0.0375 (SEE TABLE)

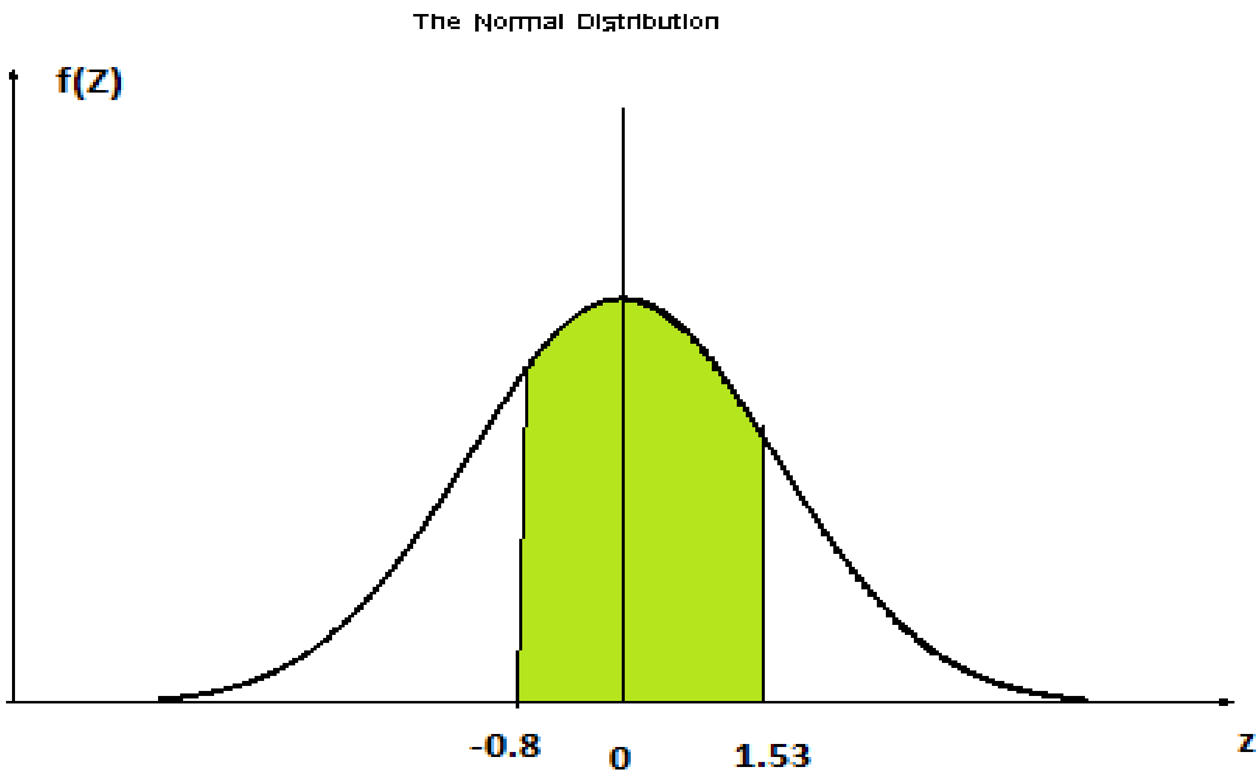


The Normal Distribution
f(z)
-1.45
0
z

Required area, $A=0.5+\text{Area from } 0 \text{ to } -1.45$

$=0.5+\text{Area from } 0 \text{ to } 1.45$ (by Symmetry)

$=0.5+0.4265=0.9265$

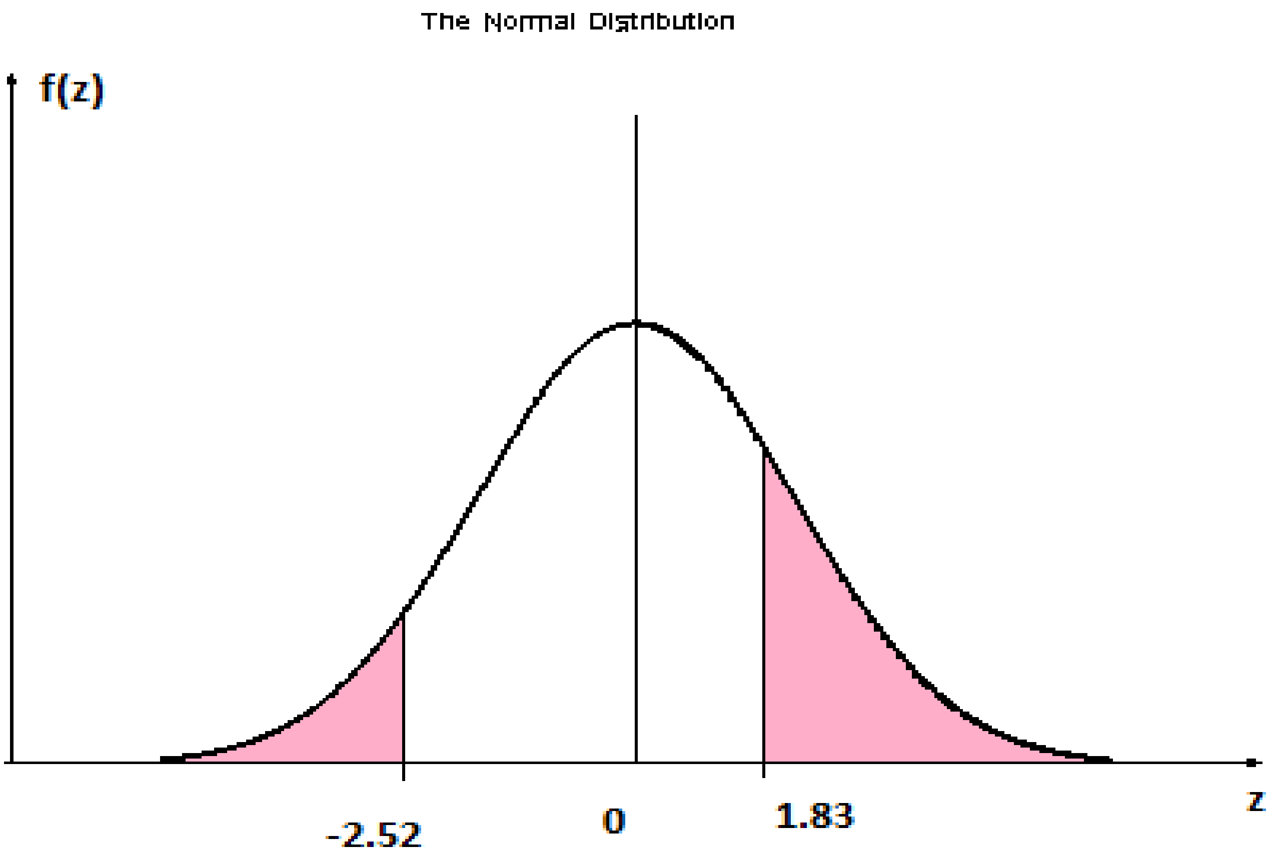


The Normal Distribution
f(Z)
-0.8
0
1.53
z

A=Area from (0 to -0.8)+Area from (0 to 1.53)

= Area from (0 to 0.8)+Area from (0 to 1.53)

=0.4370+0.2881=0.7251


The Normal Distribution
f(z)
-2.52
0
1.83
z

A=(0.5-Area from 0 to 2.52)+(0.5-Area from 0 to 1.83)

=(0.5-0.4941)+(0.5-0.4664)=0.0059+0.0336

=0.0395

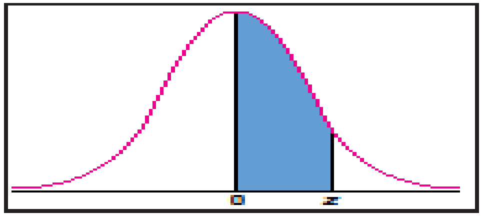


TABLE A **Areas of a Standard Normal Distribution (Alternate Version of**
Appendix I Table 4)

The table entries represent the area under the standard normal curv
to the specified value of z.

z	.00	.01	.02	.03	.04	.05	.06	.07
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4998	.4999	.4999	.4999	.4999	.4999

For values of z greater than or equal to 3.70, use 0.4999 to approximate the shaded area unde
normal curve.

Problem

If X is a normal variate with mean 30 and standard deviation 5. Find the probabilities that

1) $26 \leq X \leq 40$ 2) $X \geq 45$

Solution

1)

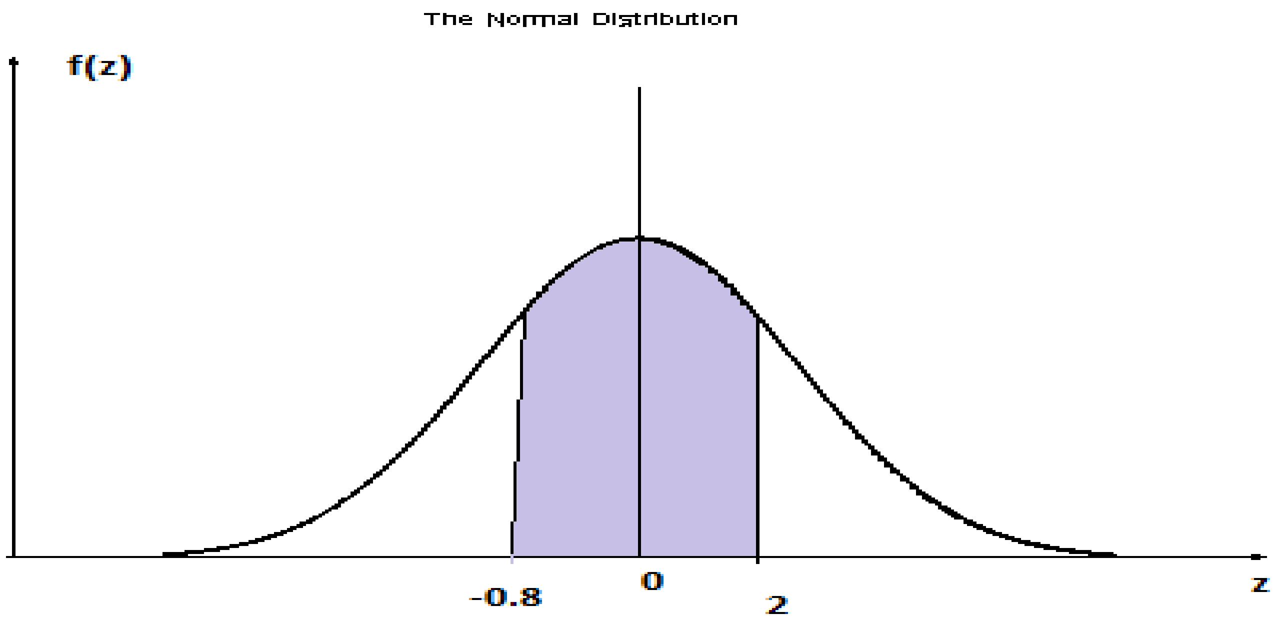
mean= $\mu = 30$, standard deviation $\sigma = 5$

When $x = 26$, $z = \dfrac{x - \mu}{\sigma}$

$$z = \frac{26 - 30}{5} = -0.8$$

When $x = 40$, $z = \dfrac{40 - 30}{5} = 2$

$$\therefore P(26 \le X \le 40) = P(-0.8 \le z \le 2)$$

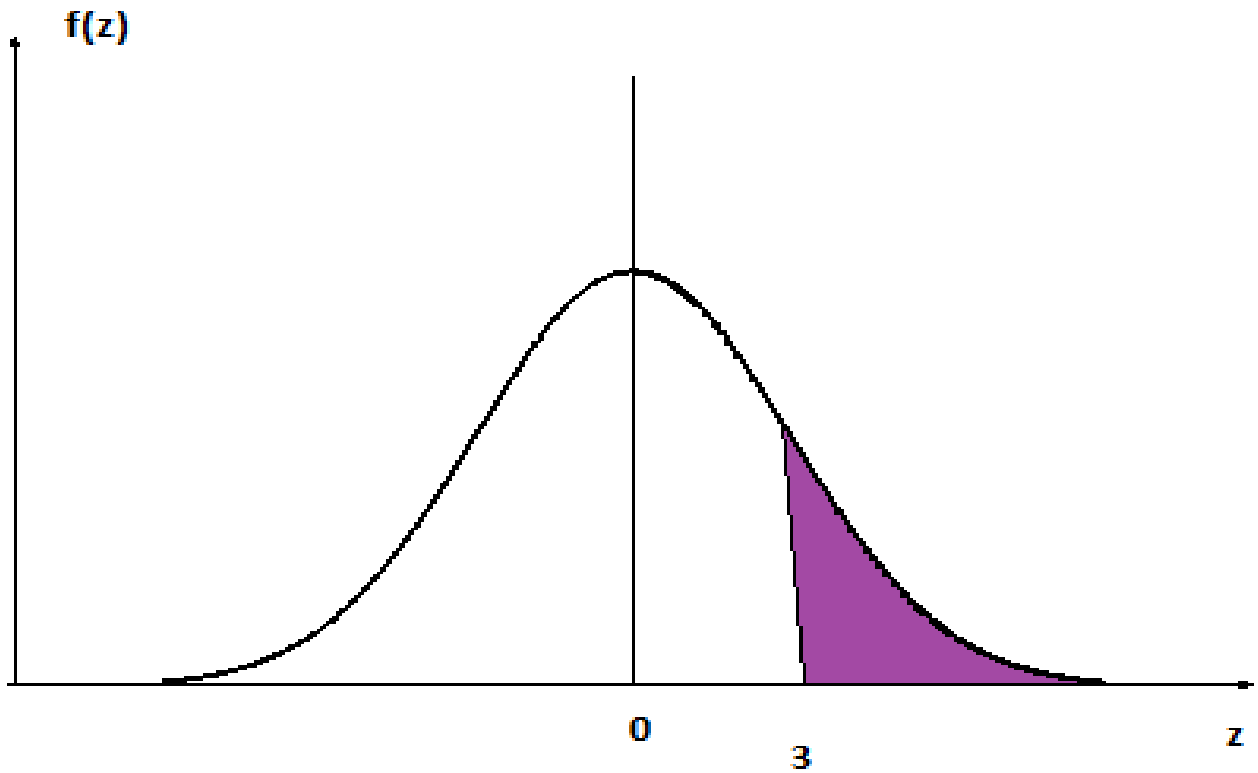

The Normal Distribution
f(z)
-0.8
0
2
z

2)

When $x = 45$, $z = \dfrac{45-30}{5} = 3$

$\therefore P(X \geq 45) = \quad P(z \geq 3)$

=0.5-0.49865=0.00135



f(z)
0
3
z

Problem

Suppose the weight of 800 male students are normally distributed with mean 140 pounds and standarded deviation 10pounds. Find the number of students whose weights are

1) between 138 and 148 pounds
2) more than 152 pounds

Solution

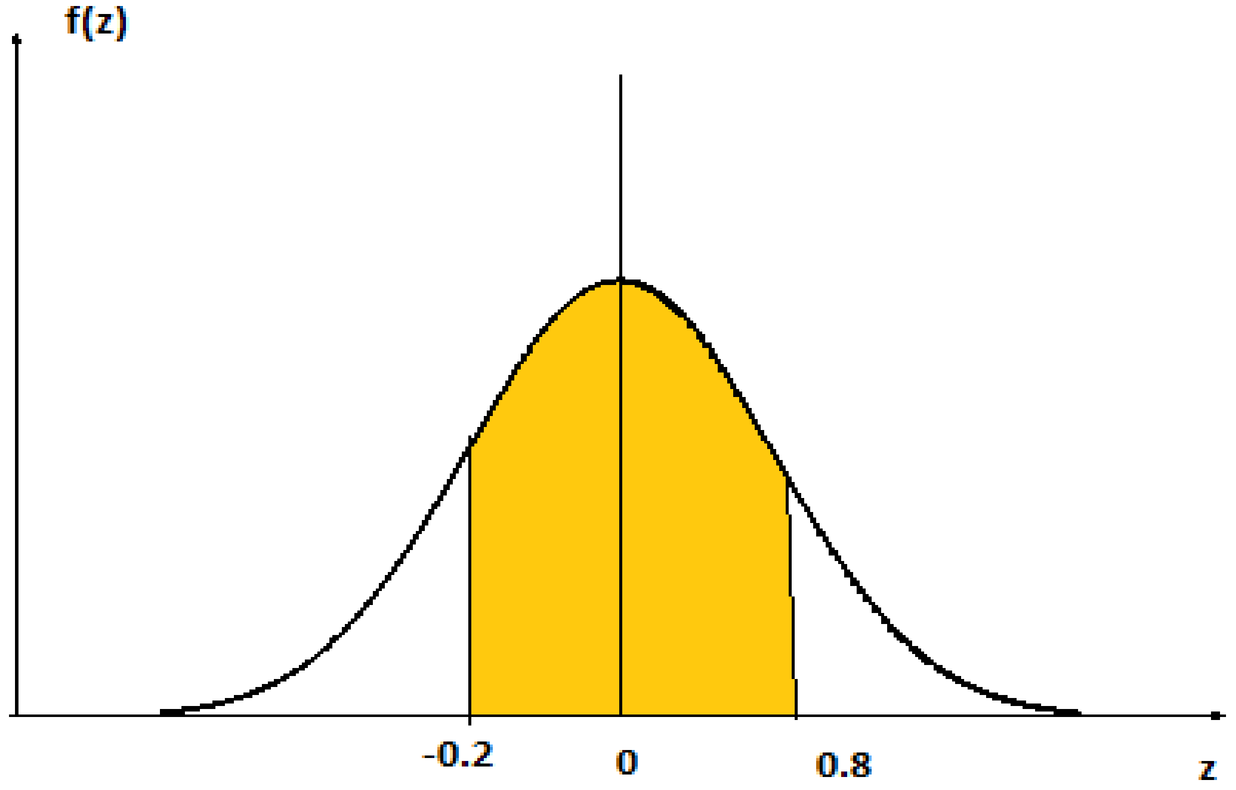
1) mean= $\mu = 140$, standard deviation $\sigma = 10$

When $x = 138$, $z = \dfrac{x - \mu}{\sigma}$

$$z = \frac{138 - 140}{10} = -0.2$$

When $x = 148$, $z = \dfrac{148 - 140}{10} = 0.8$

$$\therefore P(138 \le X \le 148) = \quad P(-0.2 \le z \le 0.8)$$



f(z)
-0.2
0
0.8
z

$A=0.2881+0.0793=0.3674$

hence the number of students whose weights are between 138 pounds and 148 pounds$=0.3674(800)=294$

When $x = 152$, $z = \dfrac{152 - 140}{10} = 1.2$

$\therefore P(X \geq 152) = \quad P(z > 1.2)$

=0.5-0.3849=0.1151

hence the number of students whose weights are more than 152 pounds=800(0.1151)=92

1.2

Problem

The marks obtained in statistics in certain examination found to be normally distributed. If 15% of students ≥ 60 marks, 40%<30 marks , find the mean and standard deviation.

Solution

Let μ be the mean and σ be the standard deviation of the normal distribution.

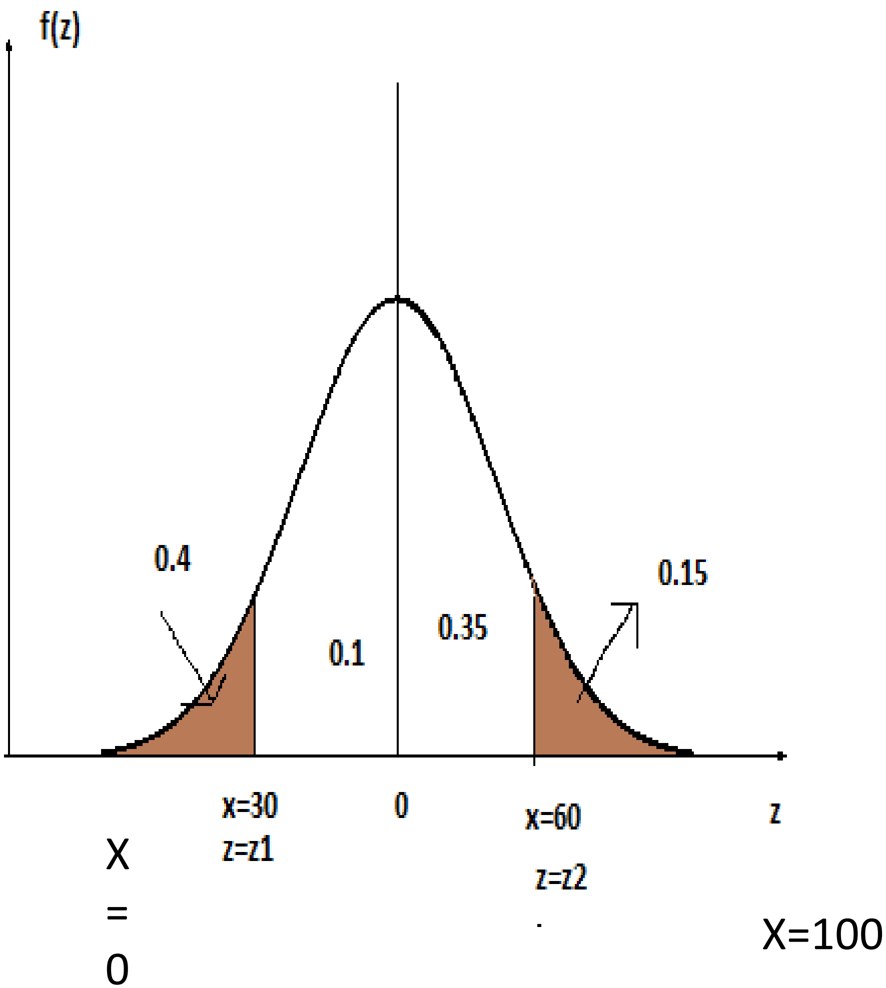
Let the variable X denote the marks in statistics.

Then we are given

$P(X < 30) = 0.4$ and $P(X \geq 60) = 0.15$

When X=30 , $z = \dfrac{x-\mu}{\sigma} = \dfrac{30-\mu}{\sigma} = -z_1 \ (say)$ ------(1)

When X=60 , $z = \dfrac{x-\mu}{\sigma} = \dfrac{60-\mu}{\sigma} = z_2 \ (say)$ -------(2)


f(z)
0.4
0.1
0.35
0.15
x=30
z=z1
0
x=60
z=z2
z
X
=
0
X=100

hence $P(0 < z < z_2)$=0.5-0.15=0.35

and $P(0 < z < z_1)$= $P(-z_1 < z < 0)$ (by symmetry)

=0.5-0.4=0.1

From the table we get $z_1 = 0.25$ and $z_2 = 1.04$

Hence $\frac{30-\mu}{\sigma} = -0.25$ using (1) -----------(3)

and $\frac{60-\mu}{\sigma} = 1.04$ using (2)--------------------(4)

which gives $\mu = 35.81$ and $\sigma = 23.26$

Problem

In a normal distribution, 7% of the items are under 35 and 89% are under 63 . Determine the mean and variance of the distribution.

Solution

Let μ be the mean and σ be the standard deviation of the normal distribution.

Then we are given

$P(X < 35) = 0.07$ and $P(X < 63) = 1 - 0.89 = 0.11$

When X=35 , $z = \dfrac{x-\mu}{\sigma} = \dfrac{35-\mu}{\sigma} = -z_1 \ (say)$ ------(1)

When X=63 , $z = \dfrac{x-\mu}{\sigma} = \dfrac{63-\mu}{\sigma} = z_2 \ (say)$ -------(2)

hence $P\left(0 < z < z_2\right)$=0.39

and $P\left(0 < z < z_1\right)$= $P\left(-z_1 < z < 0\right)$ (by symmetry)

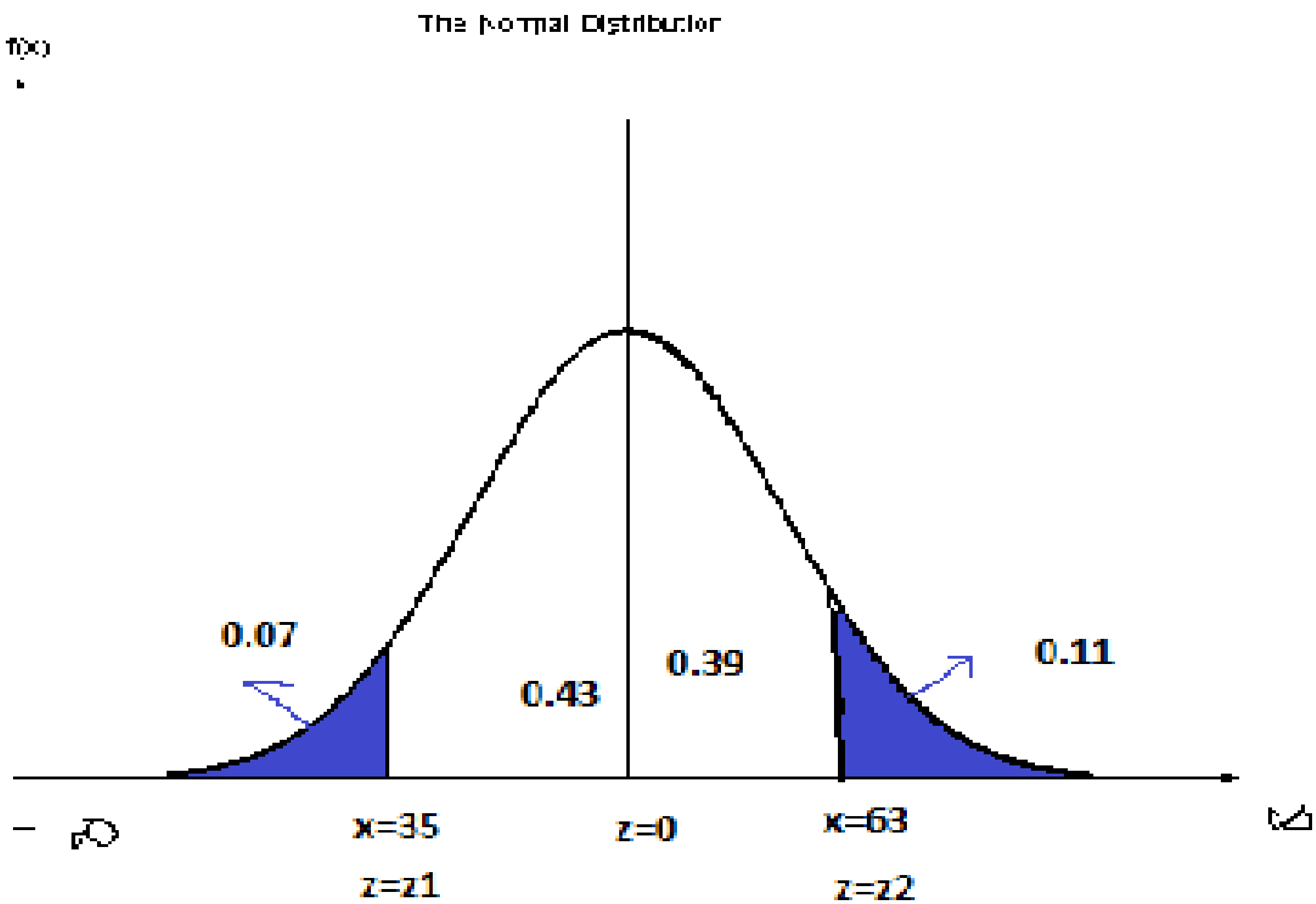
=0.43

From the table we get $z_1 = 1.23$ and $z_2 = 1.48$

Hence $\dfrac{35-\mu}{\sigma} = -1.48$ (using (1)) -----------(3)

and $\dfrac{63-\mu}{\sigma} = 1.23$ using (2)--------------------(4)

which gives $\mu = 50.3$ and $\sigma = 10.33$



The Normal Distribution
f(x)
0.07
0.43
0.39
0.11
x=35
z=z1
z=0
x=63
z=z2

PROBLEM

The mean and standard deviation of the marks obtained by 1000 students in an examination are respectively 34.5 and 16.5. Assuming the normality of the distribution, find the approximate number of students expected to obtain marks between 30 and 60.

Solution

Given mean , $\mu = 34.5$ and standard deviation $\sigma = 16.5$

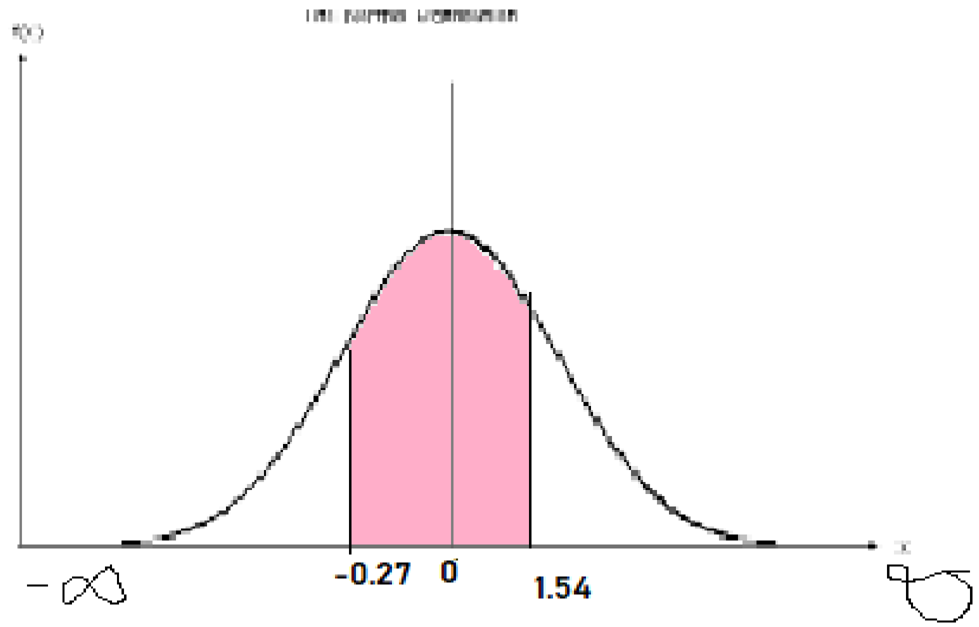
When x=30,

$$z = \frac{x-\mu}{\sigma} = \frac{30-34.5}{16.5} = -0.27 = z_1 \ (say)$$

When x=60,

$$z = \frac{x-\mu}{\sigma} = \frac{60-34.5}{16.5} = 1.54 = z_2 \ (say)$$

$$P(30 \le x \le 60) = P(-0.27 \le z \le 1.54)$$



-0.27
0
1.54

Area from 0 to 0.27+ Area from 0 to 1.54

=0.4382+0.1084

=0.5916

The number of students get marks between 30 and 60= 1000(0.5916)=592

3. In a sample of 1000 cases, the mean of a certain test is 14 and standard deviation is 2.5. Assuming the distribution to be normal, find

(i) How many students score between 12 and 15?

(ii) How many score above 18?

(iii) How many score below 18?

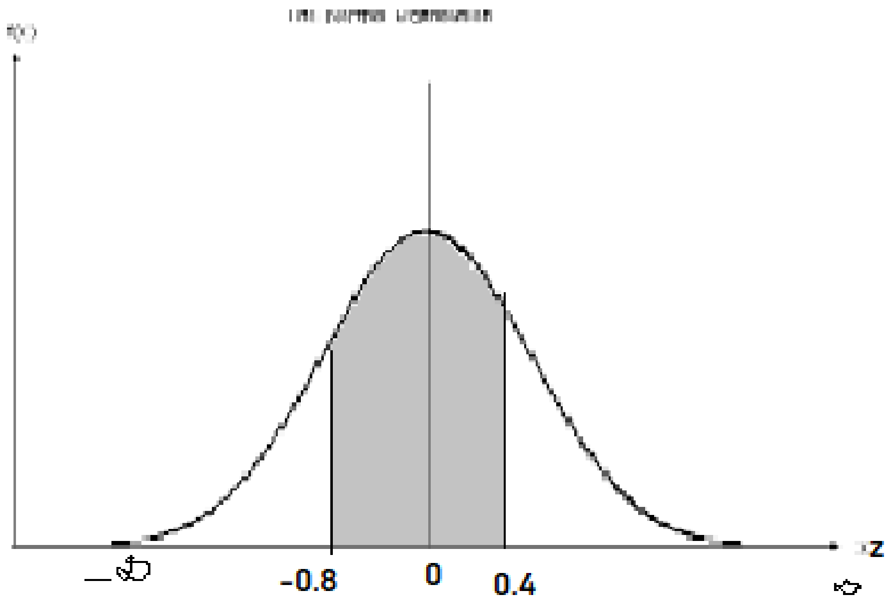
SOLUTION

Given mean, $\mu = 14$ and standard deviation $\sigma = 2.5$

When x=12, $z = \dfrac{x-\mu}{\sigma} = \dfrac{12-14}{2.5} = -0.8 = z_1\ (say)$

When x=15, $z = \dfrac{x-\mu}{\sigma} = \dfrac{15-14}{2.5} = 0.4 = z_2\ (say)$

$P(12 \le x \le 15) = P(-0.8 \le z \le 0.4)$


-0.8
0
0.4
z

Area from 0 to 0.8+ Area from 0 to 0.4

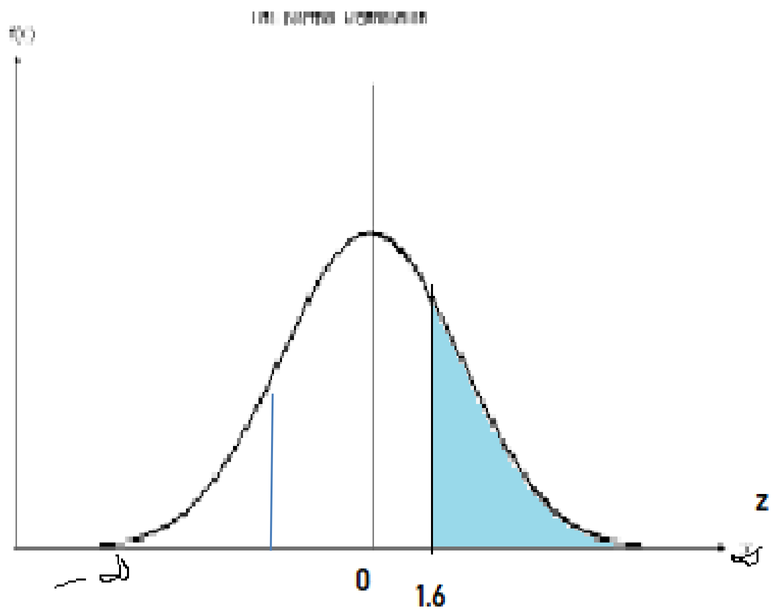
=0.1554+0.2881

=0.4435

The number of students get marks between 12 and 15= 1000(0.4435)=443

When x=18, $z = \dfrac{x-\mu}{\sigma} = \dfrac{18-14}{2.5} = 1.6$

$$P(x > 18) = P(z > 1.6)$$



-2.4

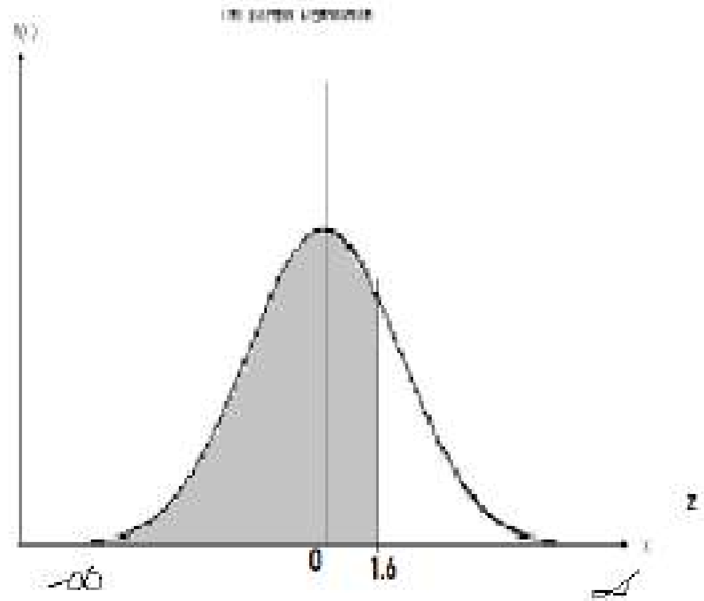
$=0.5-$ (Area from 0 to 1.6) $=$

$=0.5-0.4452=0.0548$

The number of students get marks above 18

$=1000(0.0548)=55$

$$P(x < 18) = P(z < 1.6)$$



$=0.5 + \text{area from } 0 \text{ to } 1.6$

$$=0.5 + 0.4452 = 0.9452$$

The number of students get marksbelow18

$= 1000(0.9452) = 945$

PROBLEM

A sales tax officer has reported that the average sales of the 500 business that has to deal with during a year is Rs. 36,000 with a standard deviation of 10,000. Assuming that the sales in these business are normally distributed. Find

(i) The number of business as the sales of which are over Rs. 40,000.

(ii) The percentage of business of which are likely to range between Rs. 30,000 and Rs. 40,000.

Solution

Given mean , $\mu = 36000$ and standard deviation $\sigma \quad = 10000$

When x=4000, $z = \dfrac{x-\mu}{\sigma} \quad = \dfrac{40000-36000}{10000} \quad = 0.4$

When x=3000,

$$z = \frac{x-\mu}{\sigma} \quad = \frac{30000-36000}{10000} \quad = -0.6$$

$$P(x > 40000) \quad = P(z > 0.4\ )$$

=0.5-(area from 0 to 0.4)

=0.5-0.1554=0.3446

Number of business as the sales of which are Rs40000

500(0.3446)=172

$$P(30000 < x < 40000) \quad = P(\_ -0.6 < z < 0.4)$$

=0.1554+0.2257=0.3811

The required percentage of business = 38.11%

PROBLEM

The current (in mA) measured in a piece of copper wire is known to follow a uniform distribution over the interval [0, 25].
Write down the formula for the probability density function $f(x)$ of the random variable X representing the current.
Calculate the mean and variance of the distribution and find the cumulative distribution function $F(x)$.

SOLUTION

Over the interval [0, 25] the probability density function f(x) is given by the formula

$$f(x) = \begin{cases} \dfrac{1}{25-0} = 0.04 & \text{for } 0 \le x \le 25 \\ 0 & \text{other wise} \end{cases}$$

Mean = $E(X) = \dfrac{25+0}{2} = 12.5\ mA$

Var(X)= $\dfrac{(25-0)^2}{12} = 52.08 \quad mA^2$

The cumulative distribution function is

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

Hence, choosing the three distinct regions x < 0, $0 \leq x \leq 25$ **and x > 25 in turn gives:**

$$F(x) = \quad 0 \quad ; \quad x < 0$$

$$F(x) = \quad \frac{x}{25} \quad ; \quad 0 \leq x \leq 25$$

$$F(x) = \quad 1 \quad ; \quad x > 25$$

PROBLEM

You have been informed that the assessor will visit your home sometime between 10:00 am and 12:00 pm. It is reasonable to assume that his visitation time is uniformly distributed over the specified two-hour interval. Suppose you have to run a quick errand at 10:00 am.

a) If it takes 30 minutes to run the errand, what is the probability that you will be back before the assessor visits

b) If it takes 60 minutes to run the errand, what is the probability that you will be back before the assessor visits?

SOLUTION.

(a)

An accessor is supposed to visit your home sometime between 10 am and 12 pm.

Then, let X represent the minutes after 10 am that the accessor arrives at the location.

Since we need to run a 30-minute errand at 10 am, we are going to calculate the probability that the accessor does not arrive until 10:30 A.M.

X will follow a uniform probability distribution function.

$$f\left(x\right)=\frac{1}{B-A} \qquad for\ A \leq X \leq B$$

$$f\left(x\right)=\frac{1}{120-0} \qquad for\ 0 \leq X \leq 120$$

$$f\left(x\right)=0.0083 \qquad for\ 0 \leq X \leq 120$$

P(X>30)= $\int_{30}^{120} f(x)\,dx = \int_{30}^{120} 0.0083\,dx =$

$$= 0.0083\left(x\right)_{30}^{120}$$

0.0083×(120−30)=0.75

(b)

Now, according to the question, we need to run a 60-minute errand.

Thus, we need to find the probability that the accessor does not visit until 11 A.M.

P(X>60)= $\int_{60}^{120} f(x)\,dx = \int_{60}^{120} 0.0083\,dx =$

$= 0.0083\left(x\right)_{60}^{120}$

=0.0083×(120−60)=0.50

PROBLEM

Suppose in a quiz there are 30participants. A question is given to all 30 participants and the time allowed to answer it is 25 seconds. Find the probability of participants responds
within 6 seconds?

Solution:

Given

Interval of probability distribution =[0,25]

Density of probability = 1/25

Interval of probability distribution of successful event = [0 seconds, 6seconds]

The probability $P(x<6) = 6/25$

There are 30 participants in the quiz

Hence the number of participants likely to answer it in 6 seconds = ~~6/25~~

$6/25 \times 30 \approx 7$

Normal approximation to the Binomial

Theorem

If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$, then the limiting form of the distribution of

$Z = \left(\dfrac{X - np}{\sqrt{npq}} \right)$ as $n \to \infty$, is the standard normal distribution $n(z, 0, 1)$.

Gamma Distribution

The gamma function is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad ; \alpha > 0$$

Result

$$\Gamma(\alpha) \quad = (\alpha - 1)\Gamma(\alpha - 1)$$

Gamma Distribution

Definition

The continuous random variable X has a gamma distribution, with parameters α and β, if its density function is given by

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$.

The mean and variance of the gamma distribution are $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$.

<u>Exponential distribution</u>

The continuous random variable X has an exponential distribution, with parameter β, if its density function is given by $f(x,\beta) = \left(\frac{1}{\beta}\right) e^{-x/\beta}$ $x > 0$, where $\beta > 0$.

The mean and variance of the exponential distribution are $\mu = \beta$ and $\sigma^2 = \beta^2$

Sampling

Population: The totality of observations with which we are concerned, whether this number be finite or infinite, constitutes what we call population. The size of the population is denoted by N

Sample: A portion of the population which is examined with a view to determining the population characteristics is called a sample. The size of the sample is denoted by n

Different methods of sampling

Some important methods of sampling are discussed below.

I. Probability Sampling Methods

1. Random sampling or Probability sampling

It is the process of drawing a sample from a population in such a way that each member of the population has an equal chance of being included in the sample. The sample obtained by the process of random sampling is called a random sample.

If N is the size of a population and n is the sample size, then

(i) The number of sample with replacement = N^n

(i) The number of sample without replacement = ${}^NC_n$

2. Stratified sampling or Stratified Random sampling

This method is useful when the population is heterogeneous. In this type of sampling, the population is first sub divided into several parts or small groups called *strata* according to some relevant characteristics so that each stratum is more or less homogeneous. Each stratum is called a sub-population. Then a small sample called sub-sample is selected from each stratum at random. All the sub samples are combined together to form the stratified sample which represents the population properly. The process of obtaining and examining a stratified sample with a view to estimating the characteristic of the population is known as *Stratified Sampling.*

3. Systematic Sampling or Quasi – Random Sampling

As the name suggests this means forming the sample in some systematic manner by taking items at regular intervals. In this method, all the units of the population are arranged in some order. If the population size is finite, all the units of the population are arranged in some order. Then from the first k items, one unit is selected at random. This unit and every k^{th} unit of the serially listed population combined together constitute a systematic sample. This type of sampling is known as Systematic Sampling.

II. Non – Probability Sampling Methods

1. Purposive Sampling or Judgment Sampling

When the choice of the individual items of a sample entirely depends on the individual judgment of the investigator (or sampler), it is called a purposive or *Judgment Sampling*. For example, if a sample of 20 students is to be selected from a class of 100 to analyze the extra-curricular activities of the students, the investigator would select 20 students who, in his judgment, would represent the class.

2. Sequential Sampling

It consists of a sequence of sample drawn one after another from the population depending on the results of previous sample. If the result of the first sample leads to a decision which is not acceptable, the lot from which the sample was drawn is rejected. But if the result of the first sample is acceptable, no new sample is drawn. But if the first sample leads to no clear decision, a second sample is drawn and as before if required a third sample is drawn to arrive at a final decision to accept or reject the lot. This process is called *Sequential Sampling*

Classification of Samples

Samples are classified in two ways.

1. Large Sample: if the size of the sample $(n) \geq 30$, the sample is said to be large sample

2. Small Sample: if the size of the sample $(n) < 30$, the sample is said to be small sample

Parameter and Statistics

Parameter is statistical measures of the population. Ex: population mean (μ), population variance (σ^2)

Statistic is statistical measures of the Sample. Ex: Sample mean ($\bar{x}$), population variance (s^2)

The Sample Mean :

If X_1, X_2 ……, X_n represents a random sample of size n, then the sample mean is defined by the statistic $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$

The sample Variance:

If $X_1, X_2 \ldots\ldots, X_n$ represents a random sample of size n, then the sample variance is defined by the statistic $s^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$

Central Limit Theorem

If $\bar{x}$ be the mean of a sample of size n drawn from a population with mean μ and S.D. σ then the standardized sample mean $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is a random variable whose distribution function approaches that to the standard normal distribution N(z;0,1) as $n \to \infty$

STANDARD ERROR (S.E.) OF A STATISTIC

1) S.E. $(\bar{x}) = \frac{\sigma}{\sqrt{n}}$
2) S.E. of sample proportion $p = \sqrt{\frac{PQ}{n}}$ where Q=1-P
3) S.E. $(\overline{x_1} - \overline{x_2}) = \sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}$ where $\overline{x_1}$ and $\overline{x_2}$ are the means of two random samples of sizes n_1 and n_2 drawn from two populations with S.D. σ_1 and σ_2 respectively.
4) S.E. $(s_1 - s_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

Finite Population: Consider a finite population of size N with mean μ and S.D. σ. Draw all possible samples of size n without replacement, from this population. Then

(i) The mean of the sampling distribution of means (for N>n) is $\mu_{\overline{X}} = \mu$

(ii) The variance is $\sigma_{\overline{X}}{}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

Note : The factor $\left(\frac{N-n}{N-1}\right)$ is called the finite population **correction factor.**

1. What is the value of correcting factor if n = 5 and N = 200

Sol. Given N = the size of the finite population = 200

n = the size of the sample = 5

∴ Correction factor = $\frac{N-n}{N-1} = \frac{200-5}{200-1} = \frac{195}{199} = 0.98$

2. What is the value of correcting factor if n = 10 and N = 1000

Sol. Given N = the size of the finite population = 1000

n = the size of the sample = 10

∴ Correction factor = $\frac{N-n}{N-1} = \frac{1000-10}{1000-1} = \frac{990}{999} = 0.991$

3. A population consists of five numbers 2, 3, 6, 8 and 11. Consider all possible samples of size two which can be drawn with replacement from this population. Find

(a) The mean of the population
(b) The S.D. of the population
(c) The mean of the sampling distribution of means and
(d) The S.D. of the sampling distribution of means (i.e., the standard error of means)

Sol.

(a) Mean of the population is given by $\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$

(b) Variance of the population σ^2 is given by

$$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i-\overline{x})^2}{n}$$

$$= \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5}$$

$$= \frac{16+9+0+4+25}{5} = 10.8$$

(c) Sampling with replacement (infinite population) :

The total no. of samples with replacement is

$N^n = 5^2 = 25$ samples of size 2

Here N = population size and n = sample size listing all possible samples of size 2 from population 2, 3, 6, 8, 11 with replacement we get 25 samples

$$\begin{Bmatrix} (2,2) & (2,3) & (2,6) & (2,8) & (2,11) \\ (3,2) & (3,3) & (3,6) & (3,8) & (3,11) \\ (6,2) & (6,3) & (6,6) & (6,8) & (6,11) \\ (8,2) & (8,3) & (8,6) & (8,8) & (8,11) \\ (11,2) & (11,3) & (11,6) & (11,8) & (11,11) \end{Bmatrix}$$

Now compute the arithmetic mean for each of these 25 samples. The set of 25 means $\overline{x}$ of these 25 samples, gives rise to the distribution of means of the samples known as sampling distribution of means.

The samples means are

$$\begin{Bmatrix} 2 & 2.5 & 4 & 5 & 6.5 \\ 2.5 & 3 & 4.5 & 5.5 & 7 \\ 4 & 4.5 & 6 & 7 & 8.5 \\ 5 & 5.5 & 7 & 8 & 9.5 \\ 6.5 & 7 & 8.5 & 9.5 & 11 \end{Bmatrix}$$

And the mean of sampling distribution of means is the mean of these 25 means.

$$\mu_{\overline{X}} = \frac{Sum of all sample means}{25} = \frac{150}{25} = 6$$

(d) The variance of the sampling distribution of means is obtained by subtracting the mean 6 from each number and squaring the result, adding all 25 members thus obtained, and dividing by 25

$$\sigma_{\overline{X}}{}^2 = \frac{(2-6)^2 + \ldots\ldots\ldots\ldots + (11-6)^2}{25} = \frac{135}{25} = 5.4$$

$\therefore \sigma_{\overline{X}} = \sqrt{5.40} = 2.32$

Clearly $\sigma_{\overline{X}}{}^2 = \frac{\sigma^2}{n} = \frac{10.8}{2} = 5.4 \Rightarrow \sigma_{\overline{X}} = \sqrt{5.4} = 2.32$

4. Solve the above examples without replacement to find $\mu_{\overline{X}}$ and $\sigma_{\overline{X}}$

Sol.

$\mu = 6$and $\sigma = 3.29$

Sampling without replacement (finite population) :

The total no. of samples without replacement is $^NC_n = {}^5C_2 = 10$ samples of size 2

$$\begin{Bmatrix} (2,3) & (2,6) & (2,8) & (2,11) \\ (3,6) & (3,8) & (3,11) & \\ (6,8) & (6,11) & & \\ (8,11) & & & \end{Bmatrix}$$

The corresponding sample means are

$$\begin{Bmatrix} 2.5 & 4 & 5 & 6.5 \\ 4.5 & 5.5 & 7 & \\ 7 & 8.5 & & \\ 9.5 & & & \end{Bmatrix}$$

The mean of the sampling distribution of means is

$$\mu_{\overline{X}} = \frac{Sum of all sample means}{25} = \frac{(2.5 + 4 + \ldots\ldots 8.5 + 9.5)}{10} = \frac{60}{10} = 6$$

The variance of sampling distribution of means

$$\sigma_{\overline{X}}{}^2 = \frac{(2.5-6)^2 + \ldots\ldots\ldots\ldots + (9.5-6)^2}{10} = \frac{40.5}{10} = 4.05$$

Showing that $\sigma_{\overline{X}}{}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{10.8}{2}\left(\frac{5-2}{5-1}\right) = 4.05$

5(H.W). A population consists of 6 numbers 5, 10, 14, 18, 13, and 24. Consider all possible samples of size two which can be drawn without replacement from this population. Find

(a) The mean of the population
(b) The S.D. of the population
(c) The mean of the sampling distribution of means and
(d) The S.D. of the sampling distribution of means (i.e., the standard error of means)

6. The mean height of students in a college is 155cms and S.D. is 15. What is the probability that the mean height of 36 students is less than 157 cms.

Sol.

μ = mean of the population

= mean height of students of a college = 155 cm

σ = S.D. of population = 15 cms

n = sample size = 36

$\overline{x}$ = Mean of samples = 157 cms

Now $z = \frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{157-155}{15/\sqrt{36}} = 0.8$

$\therefore P(\overline{x} \leq 157) = P(z < 0.8) = 0.5 + A(0.8) = 0.5 + 0.2881 = 0.7881$

Thus the probability that the mean height of 36 students less than 157 = 0.7881

7. A random sample of size 100 is taken from an infinite population having the mean μ = 76 and the variance σ^2 = 256. What is the probability that $\overline{x}$ will be between 75 and 78.

Sol.

μ = mean of the population = 76

σ^2 = variance of population = 256 i.e. σ = 16

n = sample size = 100

$\overline{x}_1$ = 75

Now $z_1 = \frac{\overline{x}_1-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{75-76}{16/\sqrt{100}} = -0.625$

And when $\overline{x}_2$ = 78

Now $z_2 = \frac{\overline{x}_1-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{78-76}{16/\sqrt{100}} = 1.25 \therefore P(75 \leq \overline{x} \leq 78) = P(-0.625 \leq$
$z \leq 1.25) = A(-0.625) + A(1.25) = 0.2334 + 0.3944 = 0.628$

8. A random sample of size 64 is taken from an infinite population having the mean 45 and the S.D. 8. What is the probability that x will be between 46 and 47.5

Sol.

μ = mean of the population = 45

σ = S.D. of the population = 8

n = sample size = 64

$\overline{x}_1$ = 46

Now $z_1 = \frac{\overline{x}_1-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{46-45}{8/\sqrt{64}} = 1$

And when $\overline{x}_2$ = 47.5

Now $z_2 = \frac{\overline{x}_2-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{47.5-45}{8/\sqrt{64}} = 2.5$

$\therefore P(46 \leq \overline{x} \leq 47.5) = P(1 \leq z \leq 2.5) = A(2.5) - A(1) = 0.4938 - 0.3413$
$= 0.1525$

9(H.W). A random sample of size 64 is taken from a normal population with μ = 51.4 and σ = 68. What is the probability that the mean of the sample will (a) exceed 52.9, (b) fall between 50.5 and 52.3, (c) 50.6

ESTIMATION

Estimate: An estimate is a statement made to find an unknown population parameter

Estimator: The procedure or rule to determine an unknown population parameter is called an estimator. For ex. Sample mean $\overline{x}$ is an estimator of population mean μ

Types of Estimation:

Basically there are two kinds of estimates to determine the statistic of the population parameters

(a) Point Estimation : A point estimate of a parameter θ is a single numerical value, which is computed from a given sample.

(b) Interval Estimation : An interval estimation is given by two values between which the parameter may be considered to lie.

Interval Estimation of μ: The interval estimation of μ is given by the interval $(\overline{x} - E_{Max}, \overline{x} + E_{Max})$ where $E_{Max} = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

Interval Estimation of p (proportion): The interval estimation of μ is given by the interval $(p - E_{Max}, p + E_{Max})$ where $E_{Max} = z_{\frac{\alpha}{2}} \sqrt{\frac{PQ}{n}}$

$z_{\frac{\alpha}{2}}$ **values :** 1.96 for 95% confidence
2.58 for 99% confidence
1.64 for 90% confidence

1. A random sample of size 100 has a S.D. of 5. What can you say about the maximum error with 95% confidence.

Sol.

Given σ = 5, n = 100, $z_{\frac{\alpha}{2}}$ for 95% confidence = 1.96

We know that $E_{Max} = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{5}{\sqrt{100}} = 0.98$

2. Assuming that σ = 20.0, how large a random sample be taken to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 3.0

Sol.

Given maximum error E = 3.0, and σ = 2.0, $z_{\frac{\alpha}{2}}$ = 1.96 for 95% confidence

We know that $E_{Max} = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$

$$n = \left(\frac{1.96X20}{3}\right)^2 = 170.74 \simeq 171$$

3. In a study of an automobile insurance a random sample of 80 body repair costs had a mean of Rs. 472.36 and the S.D. of Rs. 62.35. If $\overline{x}$ is used as a point estimate to the true average repair costs, with what confidence we can assert that the maximum error doesn't exceed Rs. 10

Sol.

Given n = 80, $\overline{x}$ = 472.36, σ = 62.35, E_{max} = 10

$$E_{Max} = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow z_{\frac{\alpha}{2}} = \frac{E_{Max} \cdot \sqrt{n}}{\sigma} = \frac{10\sqrt{80}}{62.35} = \frac{89.4427}{62.35} = 1.4345$$

$$z_{\frac{\alpha}{2}} = 1.43$$

The area when $z_{\frac{\alpha}{2}} = 1.43$ from tables is 0.4236

$1 - \alpha = 2 \times 0.4236 = 0.8472$

Confidence = $(1 - \alpha)100\% = 84.72$

Hence we are 84.72% confidence that the maximum error is Rs. 10

4. If we can assert with 95% that the maximum error is 0.05 and P=0.2, find the size of the sample.

Sol.

Given P=0.2, E = 0.05

We have Q = 1- P = 1 – 0.2 = 0.8 and $z_{\frac{\alpha}{2}} = 1.96(for 95\%)$

We know that maximum error, $E_{Max} = z_{\frac{\alpha}{2}}\sqrt{\frac{PQ}{n}} \Rightarrow 0.05 = 1.96\sqrt{\frac{0.2\times0.8}{n}}$

$n = \frac{0.2\times0.8\times(1.96)^2}{(0.05)^2} = 246$

5. What is the size of the smallest sample required to estimate an unknown proportion to within a maximum error of 0.06 with at least 95% confidence

Sol.

Given E = 0.06, Confidence limit = 95%

i.e. $z_{\frac{\alpha}{2}} = 1.96$

here P is not given, So we take $P = \frac{1}{2} \Rightarrow Q = \frac{1}{2}$

Hence $n = \left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2 (PQ) \Rightarrow n = \left(\frac{1.96}{0.06}\right)^2 \left(\frac{1}{2}.\frac{1}{2}\right) = 266.78 \simeq 267$

6. The mean and S.D. of a population are 11,795 and 14054 respectively. What can one assert with 95% confidence about the maximum error if $\overline{x}$ = 11,795 and n = 50. And also construct 95% confidence interval for the true mean

Sol.

Given $\mu = 11795, \sigma = 14054, \overline{x} = 11795, n = 50, z_{\frac{\alpha}{2}} = 1.96$

$E_{Max} = z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \Rightarrow 1.96.\frac{(14054)}{\sqrt{50}} = 3899$

Confidence interval = $\left(\overline{x} - E_{Max}.\frac{\sigma}{\sqrt{n}}, \overline{x} + E_{Max}.\frac{\sigma}{\sqrt{n}},\right)$

$(11795 - 3899, 11795 + 3899)$

$(7896, 15694)$

7.(H.W) A sample of 10 cam shafts intended for use in gasoline engines has an average eccentricity of 1.02 and a S.D. of 0.044 inch. Assuming the data may be treated a random sample from a normal population, determine a 95% confidence interval for the actual mean eccentricity of the cam shaft?

(Ans) = (0.993, 1.047)