

CHAPTER-III

IMPORTANT THEORETICAL DISTRIBUTIONS

We now discuss some generalized distributions which are very useful and relevant to the policy makers. These distributions are based on certain assumptions. Using these distributions, predictions can be made on theoretical grounds.

3.1 Discrete Probability Distributions:

Discrete Uniform Distribution: The probability distribution in which the random variable assumes all its values with equal probabilities, is known as discrete uniform distribution. If the random variable X assumes the value x_1, x_2, \dots, x_k with equal probabilities, then its probability function is given by:

$$P(X = x_i) = \frac{1}{k} \text{ for } i = 1, 2, \dots, k$$

Bernoulli Distribution: A random variable X which takes only two values 1 and 0 (called as success and failure) with respective probabilities p and q such that $P(X=1) = p$, and $P(X = 0) = q$ and $p + q = 1$, is called a Bernoulli Variate and its distribution is called Bernoulli Distribution.

Binomial Distribution: This distribution was given by James Bernoulli (1713) and is based on the assumptions of Bernoulli trials.

Bernoulli Trials: A series of independent trials which can result in one of the two mutually exclusive outcomes called success or failure such that the probability of success (or failure) in each trial is constant, then such repeated independent trials are called Bernoulli trials.

If we perform a series of n Bernoulli trials such that for each trial, p is the probability of success and q is the probability of failure ($p + q = 1$), then probability of x successes in a series of n independent trials is given by

$$p(x) = {}^n C_x p^x q^{n-x} \text{ where } x = 0, 1, 2, \dots, n$$

This is known as binomial distribution and the probabilities $p(0) = q^n$; $p(1) = {}^n C_1 p q^{n-1}$; $p(2) = {}^n C_2 p^2 q^{n-2}$; \dots , $p(n) = p^n$ of 0 success, 1 success, 2 successes \dots n successes are nothing but the first, the second, the third \dots the $(n+1)^{\text{th}}$ terms in the

binomial expansion $(q + p)^n$. For this reason, the distribution is called binomial distribution.

Definition: A random variable X is said to follow Binomial Distribution if it assumes only non-negative values and its probability mass function is given by:

$$P(X = x) = p(x) = {}^n C_x p^x q^{n-x}; \quad x = 0, 1, 2, \dots, n \text{ and } p + q = 1$$

Assumptions:

- i) There are n repeated independent trials.
- ii) Each trial results in two mutually exclusive outcomes called success and failure.
- iii) The probability of success p is constant for each trial.

Properties of Binomial Distribution:

- i) $\sum_{x=0}^n p(x) = \sum_{x=0}^n {}^n C_x p^x q^{n-x} = (q + p)^n = 1$
- ii) The constants n and p are parameters of the Binomial distribution and we denote the distribution as $X \sim B(n, p)$
- iii) Mean = np and variance = npq . Since $q < 1$ being probability, hence for a Binomial distribution mean is always greater than variance.

$$\text{Coefficient of skewness } (\gamma_1) = \frac{q - p}{\sqrt{npq}} \text{ and Coefficient of Kurtosis } (\gamma_2) = \frac{1 - 6pq}{npq}$$

- iv) If $X_1 \sim B(n_1, p)$ and $X_2 \sim B(n_2, p)$ are two independent binomial variates with same probability p , then $Y = X_1 + X_2 \sim B(n_1 + n_2, p)$ i.e. sum of two independent Binomial variates with same probability is again a binomial variate. This property is called **Additive Property of Binomial Distribution**.
- v) For Binomial Distribution, the recurrent relationship is

$$p(x + 1) = \left(\frac{n - x}{x + 1} \times \frac{p}{q} \right) p(x); \quad x = 0, 1, 2, \dots, n-1$$

Frequency Function of Binomial Distribution: Suppose an experiment with n outcomes is repeated N times, then expected frequency function of binomial distribution is given by

$$f(x) = Np(x) = N \cdot {}^n C_x p^x q^{n-x}$$

and is used for fitting the binomial distribution to the observed data.

Example-1: Find the Binomial distribution whose mean is 9 and standard deviation is $3/2$.

Solution: Mean = $np = 9$, variance = $npq = (SD)^2 = 9/4$

$$\text{Thus } \frac{npq}{np} = \frac{9/4}{9} \text{ or } q = 1/4$$

$$p = 1 - q = 1 - 1/4 = 3/4$$

$$\text{Now } np = 9$$

$$\text{Thus } n(3/4) = 9 \Rightarrow n = 12$$

Thus the parameters of binomial distribution are $n = 12$ and $p = 3/4$.

Example-2: If 20% animals in a region are sick, then find the probability that out of 6 animals chosen at random the number of sick animals is i) zero ii) One iii) at the most 1 iv) at least 2.

Solution: Given $n = 6$ and $p = 20/100 = 1/5$

Let X be the number of sick animals out of 6, then $X \sim B(6, 1/5)$

$$\text{i) } P(X = 0) = {}^6C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^6 = \left(\frac{4}{5}\right)^6 = 0.262$$

$$\text{ii) } P(X = 1) = {}^6C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^5 = 6 \times \left(\frac{4}{5}\right)^1 \left(\frac{4}{5}\right)^5 = 0.393$$

$$\begin{aligned} \text{iii) } P(\text{at the most } 1) &= P(X \leq 1) = P(X = 0 \text{ or } 1) \\ &= P(X = 0) + P(X = 1) \\ &= 0.26 + 0.39 = 0.65 \end{aligned}$$

$$\text{iv) } P(X \geq 2) = P(\text{at least } 2) = 1 - P(X = 0 \text{ or } 1) = 1 - 0.65 = 0.35$$

Example-3: The probability that a bulb will fail before 100 hours is 0.2. If 15 bulbs are tested for life length, what is the probability that the number of failures before 100 hours (i) does not exceed 2; (ii) are atleast 1.

Solution: Using binomial distribution we have:

$$n = 15, p = 0.2 \text{ and } q = 1 - 0.2 = 0.8$$

Let X = No. of bulbs failing before 100 hours

Therefore, $X \sim B(15, 0.2)$

$$\begin{aligned}P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\&= {}^{15}C_0(0.2)^0 (0.8)^{15} + {}^{15}C_1(0.2)^1 (0.8)^{14} + {}^{15}C_2(0.2)^2 (0.8)^{13} \\&= (0.8)^{15} + 15(0.2)(0.8)^{14} + 105(0.2)^2(0.8)^{13} = 0.648\end{aligned}$$

$$\begin{aligned}P(X \geq 1) &= 1 - P(X = 0) \\&= 1 - (0.8)^{15} = 1 - 0.035 = 0.965\end{aligned}$$

Example-4: In a farm, 200 rows each of 6 papaya seedlings are sown. Assuming equal chance of a seedling to be male or female, how many rows do you expect to have 4 or 5 female plants?

Solution: Let X be the number of female plants from 6 seedlings in a row.

Here N = 200, n = 6 and p = 1/2

$$\therefore X \sim B(6, 1/2)$$

$$\begin{aligned}P(4 \text{ or } 5 \text{ female plants in a row}) &= P(X = 4 \text{ or } 5) \\&= P(X = 4) + P(X = 5) = {}^6C_4 (1/2)^4 (1/2)^2 + {}^6C_5 (1/2)^5 (1/2)^1 \\&= \frac{15}{64} + \frac{6}{64} = \frac{21}{64}\end{aligned}$$

$$\begin{aligned}\text{Thus, the number of rows with 4 or 5 female plants} &= N \times P(X = 4 \text{ or } 5) \\&= 200 \times \frac{21}{64} \cong 66\end{aligned}$$

Example-5: Assuming equal probabilities of girls and boys in a family of 4 children find the probability distribution of number of boys and also find the expected number of families out of 2000 having (i) at most two boys (ii) at least one boy.

Solution: Let X be the number of boys in a family of 4 children. Possible values of X can be 0, 1, 2, 3 and 4. The required probabilities can be obtained using binomial distribution.

Given n = 4 and p = 1/2 = q

Thus, $X \sim B(4, 1/2)$

Probability of having x boys

$$P(X = x) = p(x) = {}^nC_x p^x q^{n-x} \quad x = 0, 1, 2, \dots, n$$

$$P(X=0) = {}^4C_0 (1/2)^0 (1/2)^{4-0} = 1/16$$

$$P(X=1) = {}^4C_1 (1/2)^1 (1/2)^{4-1} = 1/4$$

$$P(X=2) = {}^4C_2 (1/2)^2 (1/2)^{4-2} = 3/8$$

$$P(X=3) = {}^4C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{4-3} = 1/4$$

$$P(X=4) = {}^4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{4-4} = 1/16$$

The probability distribution of random variable X (number of boys) is given by

x:	0	1	2	3	4
p(x):	1/16	1/4	3/8	1/4	1/16

- i) Probability of getting at most 2 boys = $P(X \leq 2) = P(X = 0, 1 \text{ or } 2)$
 $= P(X = 0) + P(X = 1) + P(X = 2)$
 $= 1/16 + 1/4 + 3/8 = 11/16$

Expected number of families having at most two boys = $N \cdot P(X \leq 2)$
 $= 2000 \times 11/16 = 1375$

- ii) Probability of getting at least one boy = $P(X \geq 1)$
 $= P(X = 1, 2, 3 \text{ or } 4) = 1 - P(X = 0) = 1 - 1/16 = 15/16$

Expected number of families having at least one boy = $N \cdot P(X \geq 1)$
 $= 2000 \times 15/16 = 1875$

Note: Since $p = q = 1/2$, here the binomial distribution is symmetrical and thus the probability distribution of number of boys and number of girls will be same.

Example-6: The screws produced by a certain machine were checked by examining samples of sizes 7. The following table shows the distribution of 128 samples according to the number of defective screws.

No. of Defectives(X):	0	1	2	3	4	5	6	7	Total
No. of Samples(f) :	4	6	19	35	30	26	7	1	128

Fit a binomial distribution and find the expected frequencies. Also find the mean and variance of the distribution.

Solution: Here $N = \sum f_i = 128$, $n = 7$, $\sum_{i=1}^n f_i x_i = 448$

$$\bar{X} = \frac{1}{N} \sum f_i x_i = \frac{448}{128} = \frac{7}{2}$$

If the data follows Binomial distribution, then the mean \bar{X} should be equal to the mean of the Binomial distribution np .

$$\bar{X} = np = \frac{7}{2}; \quad p = \frac{7}{2n} = \frac{7}{14} = \frac{1}{2} \quad \text{and } q = 1 - p = \frac{1}{2}$$

Thus the parameters of the Binomial distributions are $n = 7$ and $p = \frac{1}{2}$

The probability that 0, 1, 2, 3, 4, 5, 6, 7 screws will be defective is given by various terms in the expansion of $(p + q)^n = (\frac{1}{2} + \frac{1}{2})^7$

$$\begin{aligned} (\frac{1}{2} + \frac{1}{2})^7 &= {}^7C_0(\frac{1}{2})^0 (\frac{1}{2})^7 + {}^7C_1(\frac{1}{2})^1 (\frac{1}{2})^6 + {}^7C_2(\frac{1}{2})^2 (\frac{1}{2})^5 + \\ &\quad {}^7C_3(\frac{1}{2})^3 (\frac{1}{2})^4 + {}^7C_4(\frac{1}{2})^4 (\frac{1}{2})^3 + {}^7C_5(\frac{1}{2})^5 (\frac{1}{2})^2 + \\ &\quad {}^7C_6(\frac{1}{2})^6 (\frac{1}{2})^1 + {}^7C_7(\frac{1}{2})^7 (\frac{1}{2})^0 \\ &= (\frac{1}{2})^7 [1 + 7 + 21 + 35 + 35 + 21 + 7 + 1] \end{aligned}$$

The expected frequencies of defective screws are obtained by multiplying with N
 $128 (\frac{1}{2} + \frac{1}{2})^7 = 128 (\frac{1}{2})^7 [1 + 7 + 21 + 35 + 35 + 21 + 7 + 1]$

Thus expected frequencies are

x :	0	1	2	3	4	5	6	7
f :	1	7	21	35	35	21	7	1

Mean of Binomial distribution = $np = 7(\frac{1}{2}) = 3.5$

Variance of Binomial distribution = $npq = 7(\frac{1}{2})(\frac{1}{2}) = 1.75$

Negative Binomial Distribution:

In the last section, binomial distribution is discussed where mean > variance and may be used to find the probability of x success in n trials (fixed). Here the Negative Binomial Distribution is discussed in which mean < variance and may be used to find that how many trials are required to obtain x successes (fixed).

Suppose n trials are required to obtain exactly x successes ($n \geq x$, i.e. $n = x, x + 1, x + 2, \dots$). It will happen only if the n^{th} trial (last trial) results in x^{th} success and the previous (n-1) trials result in (x-1) successes and $(n-1) - (x-1) = (n-x)$ failures. It can happen in ${}^{n-1}C_{x-1}$ ways.

Thus the probability that exactly n trials will be needed to obtain x successes is given by ${}^{n-1}C_{x-1} p^{x-1} q^{n-x}$. $p = {}^{n-1}C_{x-1} p^x q^{n-x}$.

Definition: A discrete random variable N (number of trials needed to obtain x successes) is said to follow negative binomial distribution if its probability function is given by:

$$P(n; x, p) = {}^{n-1}C_{x-1} p^x q^{n-x}; \quad n = x, x+1, \dots$$

where $0 < p < 1$ and $x \geq 1$ is fixed.

Another Form:

Negative binomial distribution may also be defined as the distribution of the number of failures preceding the x^{th} success. If r is the number of failures preceding the x^{th} success, then $x+r$ is the total number of trials required to produce x successes. It will happen only if the last trial i.e. $(x+r)^{\text{th}}$ trial results in success and the first $(x+r-1)$ trials result in $(x-1)$ successes for which the probability is:

$$P(r, x, p) = {}^{x+r-1}C_{x-1} p^x q^r \quad n = x, x+1, \dots \quad r = 0, 1, 2, \dots$$

and may be obtained from $p(n, x, p)$ by putting $n = x + r$.

Properties:

- i) It is a discrete distribution and has two parameters x (number of successes) and p (probability of success)
- ii) Mean = xq/p and Variance = xq/p^2 , thus mean < variance
- iii) It is positively skewed and leptokurtic distribution

Remarks:

- i) The salient feature of negative binomial distribution is that number of successes is fixed while the number of trials (N) and number of failures is a random variable.
- ii) The expected number of trials = x/p whereas expected number of failures = xq/p .
- iii) This distribution is also called as waiting time distribution as it arises in problems where the success takes place as a waiting time phenomenon. For example, it arises in problems concerning inverse sampling where the sampling of items is continued till required numbers (fixed number) of items of a particular type are obtained in the sample. For $x = 1$, this distribution is called **Geometric Distribution**.
- iv) This distribution is also called as Pascal's distribution after the name of French mathematician B. Pascal (1623-62).

Example-7: Find the probability that a man tossing a coin gets the 4th head on the 9th toss.

Solution: No. of trials $n = 9$

Number of successes $x = 4$

$$p = 0.5$$

$$\begin{aligned} \text{Therefore } P(4^{\text{th}} \text{ head on } 9^{\text{th}} \text{ trial}) &= {}^{9-1}C_{4-1} (0.5)^4 (0.5)^5 \\ &= {}^8C_3 (.5)^9 \\ &= 0.1094 \end{aligned}$$

Example-8: A recruitment agency conducts interviews by telephone found from the past experience that only 40% calls are being answered. Assuming it a Bernoulli process, find:

- i) The probability that the agency's 5th answer comes on tenth call
- ii) Expected number of calls necessary to obtain eight answers
- iii) The probability that the agency receives the first answer on third call

Solution:

- i) Consider the answer to a call as a success. Here probability of success (p) = $40/100 = 0.40$

$$x = 5, \quad n = 10$$

$$\begin{aligned} P(5^{\text{th}} \text{ answer coming on } 10^{\text{th}} \text{ call}) &= {}^{10-1}C_{5-1} (0.4)^5 (1-0.4)^{10-5} \\ &= {}^9C_4 (0.4)^5 (0.6)^5 = (126) (0.24)^5 = 0.1003 \end{aligned}$$

- ii) Expected number of calls necessary to obtain 8 answers = $\frac{x}{p} = \frac{8}{0.4} = 20$

- iii) $P(1^{\text{st}} \text{ answer on the } 3^{\text{rd}} \text{ call}) = {}^{3-1}C_{1-1} (0.4)^1 (0.6)^{3-1}$
 $= {}^2C_0 (0.4) (0.6)^2$
 $= 0.144$

Example-9: Suppose in a region, 20% animals are suffering from tuberculosis. If the animals are medically examined until 4 are found to have TB, then (i) what is the probability that 10 animals are examined; (ii) What are the expected number of animals required to be examined to have 5 positive cases of TB.

Solution:

- i) Considering an animal with TB as success

$$\text{Here } p = 20/100 = 0.2$$

$$x = 4 \quad n = 10$$

P (10 animals are examined in order to have 4 positive cases of TB)

$$\begin{aligned} &= {}^{10-1}C_{4-1} (0.2)^4 (0.8)^{10-4} \\ &= {}^9C_3 (0.2)^4 (0.8)^6 = 0.0352 \end{aligned}$$

- ii) Expected number of animals required to be examined to have 5 positive cases =
 $x/p = 5/0.2 = 25$

Poisson Distribution:

It is discrete probability distribution and was developed by a French mathematician S.D. Poisson (1837). It is also a limiting case of binomial distribution. If number of trials (n) is very large and probability of success (p) is very small so that $np = \lambda$ (say) is finite then the Binomial distribution can be approximated by Poisson distribution i.e. $B(n, p) \sim P(\lambda)$ where $\lambda = np$

Note: P.D. is good approximation to B.D. when $n \geq 20$ and $p \leq 0.05$.

Definition: A discrete R.V. X is said to follow Poisson distribution if its probability function is given by

$$P(X = x) = p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \dots, \infty$$

Assumptions:

- i) The occurrence or non-occurrence of an event does not affect the occurrence and non-occurrence of any other event.
- ii) Probability of occurrence of more than one event in a quite small interval/region is very small.
- iii) The probability of success for a short interval or a small region is proportional to the length of the interval or space.

Properties:

- i) The Poisson distribution is identified by only one parameter λ and distribution of Poisson variate is denoted as $X \sim P(\lambda)$.
- ii) Mean = variance = λ
- iii) If $X_1 \sim P(\lambda_1)$ and $X_2 \sim P(\lambda_2)$ are two independent Poisson random variables, then $Y = X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$ and it is often known as called the additive property of Poisson Distribution.

iv) For different values of λ , the values of $e^{-\lambda}$ may be obtained from Fisher and Yates tables, or may be computed by using a calculator. Then various probabilities can be computed using the recurrence relation

$$p(x+1) = \frac{p(x)}{x+1} \text{ and } p(0) = e^{-\lambda}$$

Importance of Poisson Distribution:

Poisson distribution is used in wide variety of problems concerning rare events with respect to time, area, volume or similar unit. Some of the practical situations where Poisson distribution can be used are given below

- i) Number of road accidents per month
- ii) Number of printing mistakes per page
- iii) Number of defective items produced by a standard process per batch etc.

Fitting of Poisson Distribution: Construct the frequency distribution of data (if not given) and find its mean $\bar{X} = \frac{1}{N} \sum_{i=1}^n f_i x_i$. If the data follows Poisson distribution then

$= \bar{X}$ and the expected frequencies are given by

$$f(x) = Np(x) = N \frac{e^{-\lambda} \lambda^x}{x!}; \quad x = 0, 1, 2, \dots$$

and is called the fitting of Poisson distribution to the observed data.

Example-10: Find the distribution of a Poisson random variable X for which $2P(X=1) = P(X=2)$. Also compute the probability that (i) $X = 2$ (ii) $X < 2$, given $e^{-4} = 0.0183$

Solution: Probability mass function of a Poisson distribution is

$$P(X = x) = p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

Also $P(X=1) = \frac{e^{-\lambda} \lambda^1}{1!} = \lambda \cdot e^{-\lambda}$ and

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

Given $2P(X=1) = P(X=2)$

$$\Rightarrow 2(\lambda e^{-\lambda}) = \frac{\lambda^2 e^{-\lambda}}{2} \text{ or } \lambda = 4$$

$$\text{Now } P(X = 2) = \frac{e^{-4} 4^2}{2!} = \frac{16 \times e^{-4}}{2}$$

$$= 8 \times 0.0183 = 0.1465$$

$$P(X < 2) = P[X = 0 \text{ or } 1] = P(X=0) + P(X=1)$$

$$= \left[\frac{e^{-4} 4^0}{0!} + \frac{e^{-4} 4^1}{1!} \right]$$

$$= [e^{-4} + 4e^{-4}]$$

$$= 5 \times 0.0183$$

$$= 0.0915$$

Example-11: If 1.5% animals are suffering from a rare disease, find the probability that out of 200 animals in a region (i) at least two animals are suffering (ii) No animal is suffering from rare disease use $[e^{-3} = 0.0498]$.

Solution: Total number of animals (n) = 200

Let X be the number of suffering animals

Since 1.5% animals are suffering from the disease

$$\therefore p = \frac{1.5}{100} = 0.015$$

Then $X \sim B(200, 0.015)$

Since n is large and p is small, using Poisson approximation to binomial distribution. The mean of the resulting Poisson distribution is

$$\lambda = np = 200 \times 0.015 = 3$$

$$\therefore X \sim P(3)$$

Probability function for Poisson distribution is

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

$$\text{i) } P(\text{at least two animals are suffering}) = P(X \geq 2)$$

$$= 1 - P(X < 2) = 1 - [P(X=0) + P(X=1)]$$

$$= 1 - \left[\frac{e^{-3} 3^0}{0!} + \frac{e^{-3} 3^1}{1!} \right] = 1 - [e^{-3} + 3e^{-3}]$$

$$= 1 - 4e^{-3} = 1 - 4 \times 0.0498 = 0.8008$$

$$\begin{aligned} \text{ii)} \quad & P(\text{no animal is suffering}) = P(X = 0) \\ & = \frac{e^{-3} 3^0}{0!} = \frac{e^{-3} 3^0}{1} = e^{-3} = 0.0498 \end{aligned}$$

Example-12: On an average, one bulb in 400 bulbs is defective. If the bulbs are packed in boxes of 100, what is the probability that any given box of bulbs will contain (i) no defective, (ii) less than two defective bulbs (iii) one or more defective bulbs.

Solution: Probability of a defective bulb $p = 1/400$ is small and number of bulbs in a given packet $n = 100$ are large so using Poisson approximation to binomial distribution, $\lambda = np = 100/400 = 0.25$ is finite.

Thus the required probabilities can be obtained through Poisson distribution having mean 0.25. If X denotes the number of defective bulbs in a batch of 100 bulbs, then $X \sim P(0.25)$.

$$\begin{aligned} \text{i)} \quad & P(X=0) = e^{-\lambda} = e^{-0.25} = 0.779 \\ \text{ii)} \quad & P(X < 2) = e^{-\lambda} + \lambda e^{-\lambda} = 0.779 (1 + 0.25) = 0.973 \\ \text{iii)} \quad & P(X \geq 1) = 1 - P(X=0) = 1 - 0.779 = 0.221 \end{aligned}$$

Example-13: A manufacturer who produces bottles, finds that 0.1% of the bottles are defective. The bottles are packed in boxes containing 500 bottles. A drug company buys 100 boxes from the producer. Using Poisson distribution, find how many boxes will contain (i) exactly one defective (ii) at least two defectives.

Solution: Let X denotes the number of defective bottles in a pack of 500 bottles.

$$\text{Here } p = 0.001 \text{ and } n = 500 \text{ so that } \lambda = np = 500 \times 0.001 = 0.5$$

Therefore, $X \sim P(0.5)$

$$\text{i)} \quad P(X = 1) = \lambda e^{-\lambda} = 0.5e^{-0.5} = 0.5(0.6065) = 0.3032$$

Thus the number of boxes containing exactly one defective
 $= 100 \times 0.3032 = 30.32$ or 30

$$\begin{aligned} \text{ii)} \quad & P(X \geq 2) = 1 - P(X=0) - P(X=1) \\ & = 1 - (e^{-\lambda} + \lambda e^{-\lambda}) = 1 - e^{-\lambda} (1 + \lambda) \\ & = 1 - 0.6065 (1.5) = 1 - 0.9097 = 0.0903 \end{aligned}$$

\therefore Required number of boxes $= 100 \times 0.0903 = 9.03$ i.e. approximately equal to 9

Example-14: The distribution of typing mistakes committed by a typist is given below

No. of mistakes/page (x)	:	0	1	2	3	4	5	
No. of pages (f)	:	142	156	69	27	5	1	= Total 400

Fit a Poisson distribution and find the expected frequencies.

Solution:

fx	:	0	156	138	81	20	5	$\Sigma fx = 400$
----	---	---	-----	-----	----	----	---	-------------------

$$\bar{X} = \lambda = \Sigma fx/N = 400/400 = 1$$

using the recurrence relation

$$p(x+1) = \frac{p(x)}{x+1} \text{ and } p(0) = e^{-\lambda}$$

$$N P(X=0) = 400 e^{-\lambda} = 400 \times 0.3679 = 147.16$$

$$N P(X=1) = 400 P(X=0) \lambda = 147.16$$

$$N P(X=2) = 400 P(X=1) \lambda/2 = 73.58$$

$$N P(X=3) = 400 P(X=2) \lambda/3 = 24.53$$

$$N P(X=4) = 400 P(X=3) \lambda/4 = 6.13$$

$$N P(X=5) = 400 P(X=4) \lambda/5 = 1.23$$

Thus the expected frequencies are:

No. of Mistakes	:	0	1	2	3	4	5	
Expected No. of Pages	:	147	147	74	25	6	1	

3.2 Continuous Probability Distributions:

Normal Distribution:

While dealing with quantities whose magnitude is of continuous nature such as weight, height etc. a continuous probability distribution is needed. Also when an experiment involves discrete phenomena, the appropriate discrete model may be difficult to use if the number of observations are large. In such cases, it is often convenient to use a continuous model to approximate the discrete model. Normal distribution is one of the most useful theoretical distributions for continuous variables. Most of statistical data concerning biological and agricultural research can be assumed to have a normal distribution.

The normal distribution was first described by Abraham De Moivre (1733) as the limiting form of the Binomial model. Normal distribution was rediscovered by Gauss in

1809 and by Laplace in 1812. The normal model has become the most important probability model in statistical analysis.

Definition: A continuous random variable X is said to have a normal distribution with parameters μ (mean) and σ^2 (variance) if its probability density function is given by:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad -\infty < x < \infty$$

Here $\pi = 3.141$ and $e = 2.718$

If variable X is normally distributed and it has mean ' μ ' and variance σ^2 then the variable

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

is called standard normal variate with mean 0 and variance 1 and probability density function of Z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \quad -\infty < z < \infty$$

Normal Probability Curve: The probability curve of normally distributed variable 'X' is called normal probability curve.

Properties of Normal Distribution:

- i) The curve of normal distribution is bell shaped and is symmetrical about $x = \mu$. It can have different shapes depending upon different values of μ and σ but there is one and only one normal distribution for any given pair of μ and σ .
- ii) The maximum height of the normal curve is $\frac{1}{\sqrt{2\pi}}$ and is attained at $x = \mu$.
- iii) Mean, median and mode of the normal distribution coincide.
- iv) Probability that X lies between two values a and b ($a < b$) is given by the area under normal probability curve between $x = a$ and $x = b$.
- v) The two tails of the normal probability curve extend indefinitely and never touches the horizontal axis, which implies a positive probability for all values of random variable ranging from $-\infty$ to $+\infty$.
- vi) Areas covered by the various ranges of variable are given by:

$$P [\mu - \sigma < X < \mu + \sigma] = 0.6826$$

$$P [\mu - 2\sigma < X < \mu + 2\sigma] = 0.9544$$

$$P [\mu - 3\sigma < X < \mu + 3\sigma] = 0.9973$$

vii) Linear combination of normal variates is also distributed normally.

Relationship between Poisson, Binomial and Normal distributions

If in a binomial distribution the number of trials 'n' tends to infinity and neither p nor q is small then the binomial distribution can be approximated by the normal distribution i.e. X is approximately normal with mean np and variance npq and

probability density function of $Z = \frac{X - np}{\sqrt{npq}}$ tends to $\frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$ which is probability

density function of standard normal variate.

Since binomial distribution is discrete while normal distribution is continuous so continuity correction is applied while working probabilities.

$$\text{i.e. } P\{X = x\} = P\left\{x - \frac{1}{2} \leq X \leq x + \frac{1}{2}\right\}$$

$$\text{and } P\{x_1 \leq X \leq x_2\} = P\left\{x_1 - \frac{1}{2} \leq X \leq x_2 + \frac{1}{2}\right\}$$

Similarly if variable X has Poisson distribution with parameter λ and if λ tends to infinity then the p.d.f. of variable

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \text{ tends to } \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$$

Importance of Normal Distribution:

- i) Most of the distributions occurring in practice like binomial, Poisson, hypergeometric distributions etc. can be approximated by normal distribution. Moreover, many of the sampling distributions like student 't' Snedcor's F, chi-square distribution etc. tend to normality for large samples.
- ii) Even if variable is not normally distributed it can be sometimes brought to normal form by simple transformation of variables.
- iii) If a variable X is normally distributed then the property that $P [\mu - 3\sigma < X < \mu + 3\sigma] = 0.9973$, i.e. $P [|Z| < 3] = 0.9973$ form the basis of entire large sample theory.

- iv) The sampling distributions of important statistics have been derived on the fundamental assumption that the population from which sample is drawn is normal.
- v) Normal distribution finds large applications in statistical quality control.

Computation of probabilities

Let $A(z_1)$ be the area between $z = 0$ and $z = z_1$ under the standard normal curve. For different values of z_1 , the areas from 0 to z_1 are given in the tables of standard normal distribution. Further since the curve is symmetrical i.e. $A(-z_1) = A(z_1)$ and 50% (=0.5) area lies on either side of $z = 0$, then with the help of these properties, we can find various probabilities. If X is normally distributed with mean μ and variance σ^2 , then after converting X into standard normal variate $Z = (X-\mu)/\sigma$, we can find various probabilities as shown in the following figure:

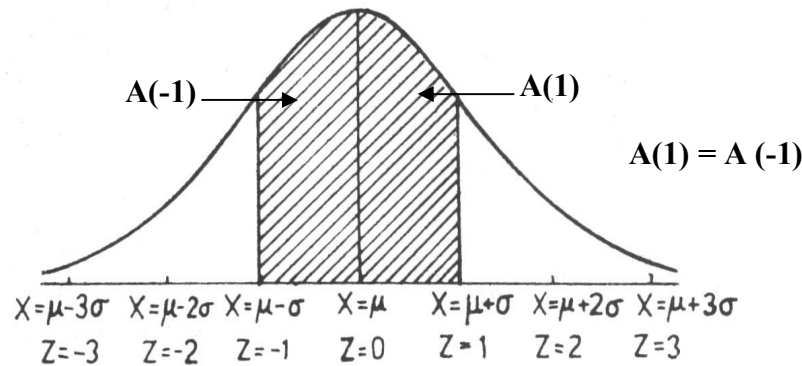
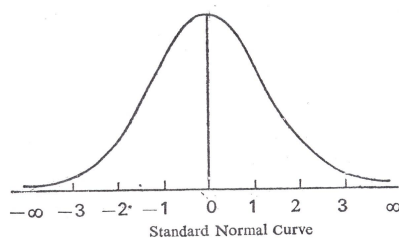


Figure showing area under Normal Curve and Standard Normal Curve

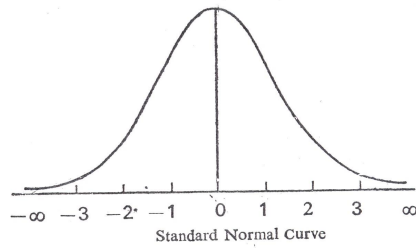
Example-15: Find the probability that standard normal value of Z lies (i) between 0 and 1.8 (ii) from -1.8 to 1.8 (iii) from 1.2 to 2.3 (iv) -2.3 to -1.2 (v) to the right of -1.50 (vi) to the left of -1.82.

Solution: From the shape and tables of area under standard normal curve we find the probabilities

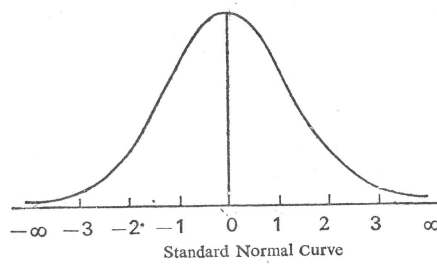
i) $P(0 \leq Z \leq 1.8) = A(1.8) = 0.4656$



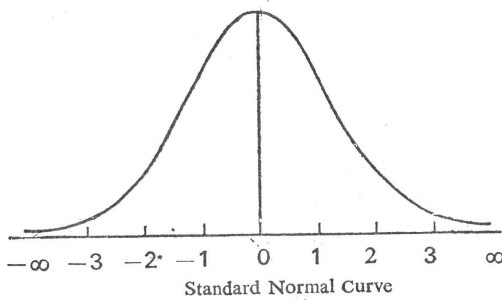
ii) $P(-1.8 \leq Z \leq 1.8) = A(1.8) + A(1.8) = 0.4641 + 0.4641 = 0.9282$



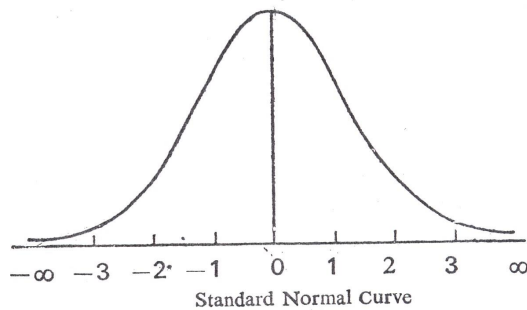
iii) $P(1.2 \leq Z \leq 2.3) = A(2.3) - A(1.2) = 0.4893 - 0.3849 = 0.1044$



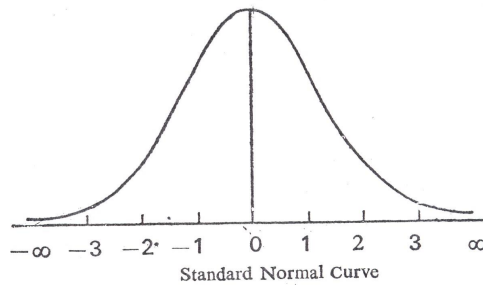
iv) $P(-2.3 \leq Z \leq -1.2) = A(2.3) - A(1.2) = 0.4893 - 0.3849 = 0.1044$



v) $P(Z > -1.50) = 0.5 + A(1.5) = 0.5 + 0.4332 = 0.9332$



vi) $P(Z < -1.82) = 0.5 - A(1.82) = 0.5 - 0.4656 = 0.0344$

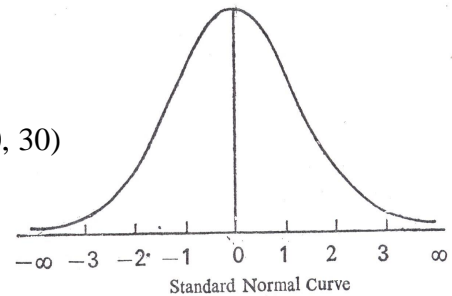


Example-16: Weight of onion bulb is assumed to be normally distributed with mean 200 gm and standard deviation (σ) = 30 gm. What is the probability that a bulb selected at random have weight between 215 gm and 260 gm?

Solution: Given $\mu = 200$ gm; $\sigma = 30$ gm

Let X be the weight of onion bulbs, then $X \sim N(200, 30)$

Also $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$



$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{215 - 200}{30} = \frac{15}{30} = \frac{1}{2} = 0.5$$

and $z_2 = \frac{x_2 - \mu}{\sigma} = \frac{260 - 200}{30} = \frac{60}{30} = 2$

$$P(215 \leq X \leq 260) = P[0.5 \leq Z \leq 2] = P(0 \leq Z \leq 2) - P(0 \leq Z \leq 0.5)$$

$$= 0.4773 - 0.1915 = 0.2858$$

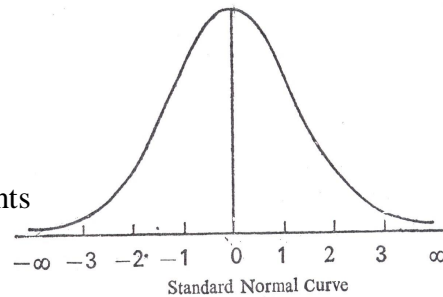
Example-17: Plants height is assumed to be normally distributed with mean 60 cm and s.d.= 5 cm. If 20% shortest plants are to be selected for crossing, find the maximum height of the selected plants.

Solution: Given $\mu = 60$ cm and $\sigma = 5$ cm

20% shortest plants are selected for crossing

Let k be the maximum height of the selected plants

Then $P(X \leq k) = 0.2$



or $P\left(\frac{X - \mu}{\sigma} \leq \frac{k - 60}{5}\right) = 0.2$

We know from table that $P(Z \leq 0.84) = 0.2$

$$\Rightarrow \frac{k - 60}{5} = -0.84$$

$$k = 60 + 5 \times 0.84 = 60 + 4.20 = 64.20$$

Thus maximum height of the selected plant is 64.20 cm.

Example-18: The mean marks for an examination is 72 and the standard deviation is 9. The top 10% of the students are to be awarded A ϕ . Find the minimum marks a student must get in order to receive an A.

Solution: Given $\mu = 72$ and $\sigma = 9$

If X denote the marks obtained by a student, then $X \sim N(72, 81)$. Let x_1 denote the minimum marks a student must get for an A grade.

$$\text{Then, } P(X \geq x_1) = \frac{10}{100} = 0.1$$

$$\text{or } P\left(\frac{x - 72}{9} \geq \frac{x_1 - 72}{9}\right) = 0.1$$

$$\text{From standard normal table we have } \frac{x_1 - 72}{9} = 1.28$$

$$\text{or } x_1 = 9 \times 1.28 + 72 = 83.52$$

Thus the minimum marks for an A grade are $83.52 \cong 84$

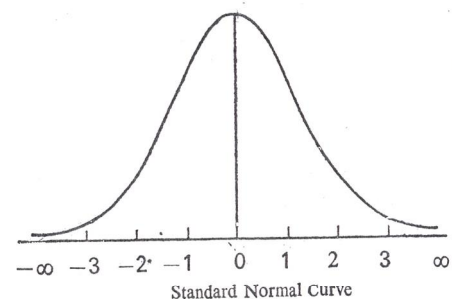
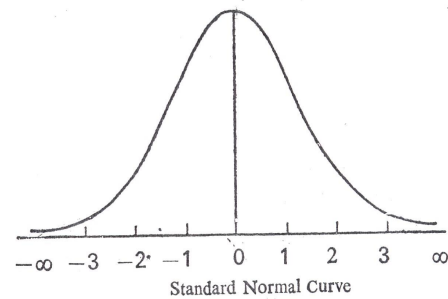
Example-19: The lifetimes of certain kinds of electronic devices have a mean of 300 hours and standard deviation of 25 hours. Assuming normal distribution

- Find the probability that any one of these electronic devices have a lifetime of more than 325 hours.
- What percentage will have lifetimes of 300 hours or less?
- What percentage will have lifetime between 220 and 260 hours?

Solution: Let the random variable X denotes the life time of an electronic device.

$$\text{a) Given } \mu = 300, \sigma = 25 \text{ and } x = 325$$

Converting X into standard normal score



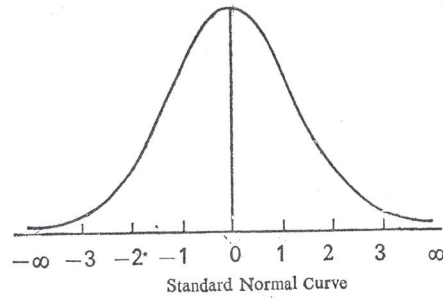
$$z = \frac{x - \mu}{\sigma} = \frac{325 - 300}{25} = 1$$

Thus $P(X > 325) = P(Z > 1) = 0.5 - A(1) = 0.5 - 0.3413 = 0.1587$

b)
$$z = \frac{x - \mu}{\sigma} = \frac{300 - 300}{25} = 0$$

Thus $P[X \leq 300] = P[Z \leq 0] = 0.5$

Therefore, the required percentage is $0.5000 \times 100 = 50\%$



c)
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{220 - 300}{25} = -3.2$$

and
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{260 - 300}{25} = -1.6$$

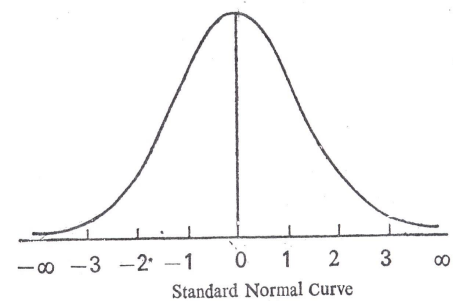
$P(220 \leq X \leq 260) = P(-3.2 \leq Z \leq -1.6)$

$$= A(3.2) - A(1.6)$$

$$= 0.4993 - 0.4552 = 0.0541$$

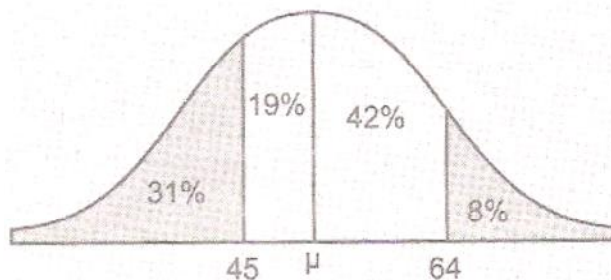
Therefore, the required percentage is

$$0.0541 \times 100 = 5.41\%$$



Example-20: In a normal distribution 31% of the items are under 45 and 8% are over 64. Find the mean and SD of the distribution.

Solution: Let μ and σ be the mean and SD. Since 31% of the items are under 45, therefore, the area to the left of ordinate at $x = 45$ is 0.31. Thus the area to the right of ordinate upto mean is $(0.5 - 0.31) = 0.19$. The value of Z corresponding to this area is -0.5.



Hence $z = (45 - \mu)/\sigma = -0.5$ or $\mu - 0.5\sigma = 45$

As 8% of items are above 64, therefore, the area from $x = 64$ upto mean is $(0.5 - 0.08) = 0.42$ and the value of Z corresponding to this area is 1.4.

$$\text{Hence } z = (64 - \mu) / \sigma = 1.4 \text{ or } \mu + 1.4\sigma = 64$$

From the above two equations we get $1.9\sigma = 19$ i.e. $\sigma = 10$. Putting $\sigma = 10$ in first equation, we get $\mu = 50$

The mean of distribution = 50 and SD = 10

Example-21: A machine produces bolts which are 10% defective. Find the probability that a random sample of 400 bolts produced by the machine the number of defective bolts found (i) will be at most 30 (ii) will be between 30 and 50 (iii) will exceed 55.

Solution: Since the probability of a defective bolt is 0.1 which is considered as large and number of bolts 400 are large, therefore, it is more appropriate to use normal distribution as a limiting case of binomial distribution.

$$\text{Thus } \mu \text{ (mean)} = np = 400 \times 0.1 = 40$$

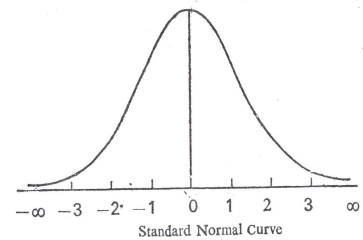
$$= \sqrt{npq} = \sqrt{400 \times 0.1 \times 0.9} = 6$$

Let $X =$ No. of defective bolts

i) $P(X \leq 30) = P(X \leq 30.5)$ (Continuity correction)

$$= P(Z \leq (30.5 - 40) / 6) = P(Z \leq -1.58)$$

$$= 0.5 - A(1.58) = 0.5 - 0.4429 = 0.0571$$



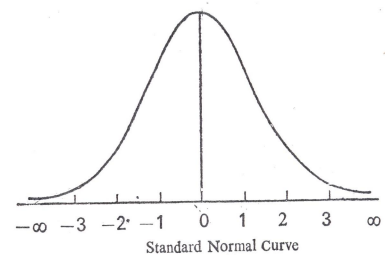
ii) $P(30 < X < 50) = P(29.5 < X < 50.5)$

$$= P\left[\frac{29.5 - 40}{6} \leq Z \leq \frac{50.5 - 40}{6}\right]$$

$$= P(-1.75 < Z < 1.75)$$

$$= A(1.75) + P(1.75)$$

$$= 0.4599 + 0.4599 = 0.9198$$

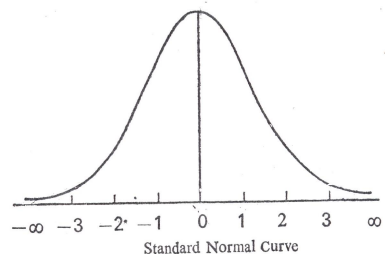


iii) $P(X > 55) = P(X > 54.5) = P\left[Z > \frac{54.5 - 40}{6}\right]$

$$= P(Z > 2.42)$$

$$= 0.5 - A(2.42)$$

$$= 0.5 - 0.4922 = 0.0078$$



Gamma Distribution:

Definition: A continuous random variable X is said to follow gamma distribution with parameters a and $\lambda > 0$, if the p.d.f. is given by:

$$f(x) = \begin{cases} \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1}; & \lambda > 0, \lambda > 0, 0 < x < \infty \\ 0, & \text{otherwise} \end{cases}$$

It is generally denoted by $\Gamma(a, \lambda)$

If $\lambda = 1$, gamma distribution simply denoted by $\Gamma(a)$.

Properties:

- i) It is a continuous distribution with mean $\frac{\lambda}{a}$ and variance $\frac{\lambda}{a^2}$
- ii) If $a < 1$, variance $>$ mean; if $a > 1$, variance $<$ mean and if $a = 1$, variance = mean
- iii) Gamma distribution is positively skewed and leptokurtic in shape

Beta Distribution:

Definition-1: A continuous random variable X is said to have a beta distribution of first kind with parameters μ and ν ($\mu > 0, \nu > 0$) if its p.d.f. is given by:

$$f(x) = \begin{cases} \frac{1}{B(\mu, \nu)} x^{\mu-1} (1-x)^{\nu-1}; & (\mu, \nu) > 0, 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

where, $B(\mu, \nu)$ is the beta function.

Properties:

It is a continuous distribution with mean $\frac{\mu}{\mu + \nu}$ and variance $\frac{\mu\nu}{(\mu + \nu)^2(\mu + \nu + 1)}$

Definition-2: A random variable X is said to have a beta distribution of second kind with parameters μ and ν ($\mu > 0, \nu > 0$) if its p.d.f. is given by:

$$f(x) = \begin{cases} \frac{1}{B(\mu, \nu)} \frac{x^{\mu-1}}{(1+x)^\nu} & (\mu, \nu) > 0, 0 < x < \infty \\ 0, & \text{otherwise} \end{cases}$$

Properties:

It is a continuous distribution with mean $\frac{\mu}{\nu - 1}$ and variance $\frac{(\mu + \nu - 1)}{(\nu - 1)^2(\nu - 2)}$

Note: Both beta type-I and type-II distributions are same except in respect of range. Beta type-I has the range (0, 1) and beta type-II has range (0, ∞).

EXERCISES

1. The probability that an evening college student will become postgraduate is 0.4. Determine the probability that out of 5 students a) none, b) one, c) atleast one will become postgraduate.
2. The incidence of a certain disease is such that on an average 20% of workers suffer from it. If 10 workers are selected at random, find the probability that (i) exactly 2 workers (ii) Not more than 2 workers suffer from the disease.
3. The mean of binomial distribution is 40 and standard deviation 6. Calculate n, p and q.
4. The appearance of 2 or 3 on a die is considered as success. Five dice are thrown 729 times and the following results are obtained

Number of Successes	:	0	1	2	3	4	5
Frequency	:	45	195	237	132	81	39

Assuming dice to be unbiased, fit the binomial distribution.
5. The average number of telephone calls received in an exchange between 2 PM and 3 PM is 2.0 per minute. Find the probability that number of calls received during a particular minute are
(i) 0; (ii) 1; (iii) atleast 2 (Hint: Apply Poisson distribution with $\lambda = 2.0$)
6. If the probability that an individual suffers a bad reaction of a given medicine is 0.002, determine the probability that out of 2000 individuals (i) Exactly 3 individuals (ii) more than 2 individuals will suffer from reaction.
7. The following table gives the number of days in a 50 day period during which automobile accidents occurred in a city.

No. of Accidents	:	0	1	2	3	4
No. of Days	:	21	18	7	3	1

Fit a Poisson distribution to the data.
8. Assuming the mean height of soldiers to be 68.22 inches with a variance of 10.8 inches². How many soldiers in a regiment of 1000 would you expect to be over 6 feet tall?
9. In a certain examination 10% of the students got less than 30 marks and 97% of the students got less than 62 marks. Assuming distribution to be normal, find the mean and standard deviation of marks.
10. How many workers have a salary above Rs 6750 in the distribution whose mean is Rs 5000 and standard deviation is Rs 1000 and the number of workers in the factory is 15000. Assume that salary of workers follows the normal law.