Probability Distribution

Probability Distribution

A probability distribution shows the possible outcomes of an experiment and the probability of each of these outcomes.

Or

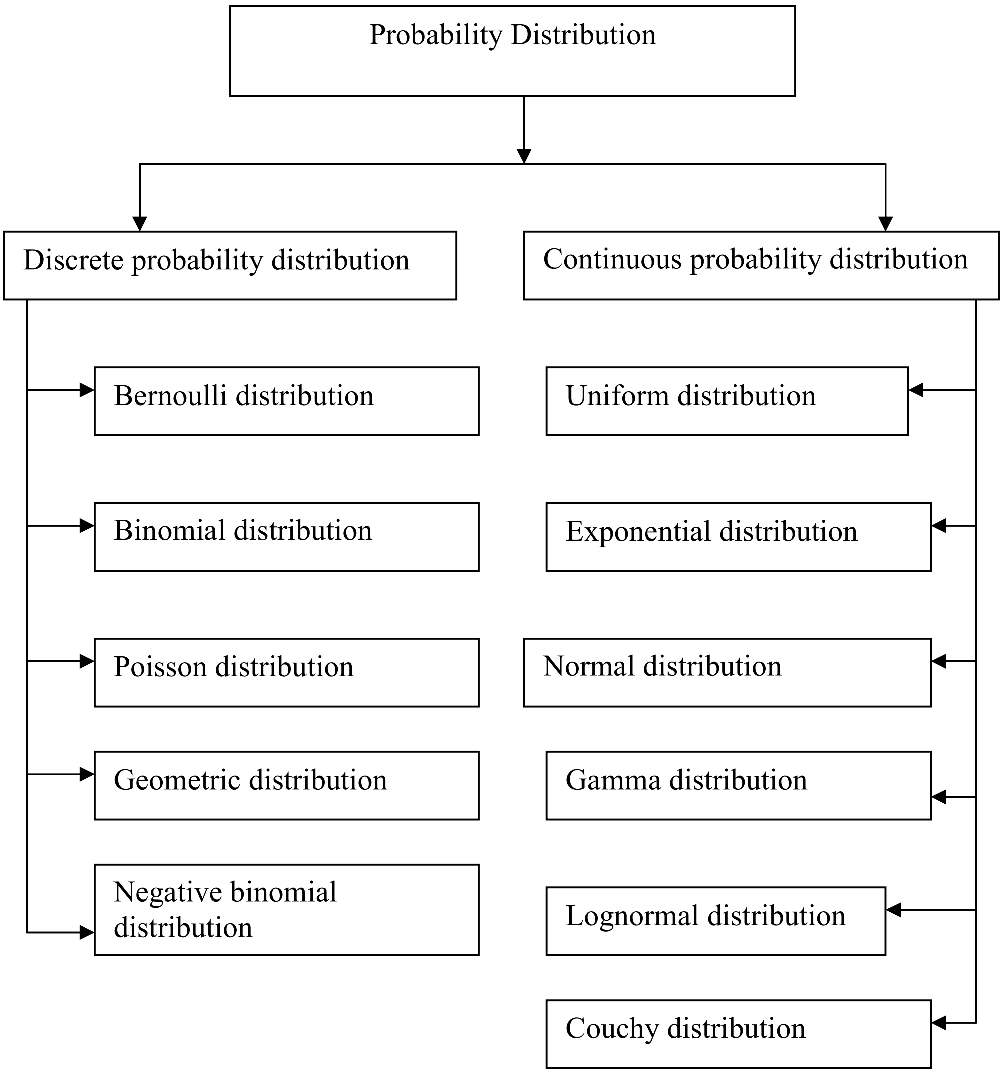
A listing of all the outcomes of an experiment and the probability associated with each outcome.

Example

To begin our study of probability distribution, let’s go back to the idea of a fair coin, suppose we toss a fair coin twice the possible outcomes are:

	First toss	Second toss	Number of heads on two tosses	Probability of the four possible outcomes
Possible outcomes from two tosses of a fair coin	T	T	0	$0.5*0.5=0.25$
	T	H	1	$0.5*0.5=0.25$
	H	T	1	$0.5*0.5=0.25$
	H	H	2	$0.5*0.5=0.25$
Total				1.0

Types of probability distribution



Discrete probability distribution

A discrete probability can take on only a limited number of values which can be listed.

Example: The probability that you were born in a given month is also discrete because there are 12 possible values.

Continuous probability distribution

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

In a continuous probability distribution the variable under consideration is allowed to take on any value within a given range. So we can not list all the possible values.

Example: Suppose we were examining the level of effluent in a variety of streams and we measured the level of effluent by parts of effluent per million parts of water. We would expect quite a continuous range of parts per million (ppm), all the way from very low levels is clear mountains streams of extremely high levels in polluted streams. We would call the distribution of this variable (ppm) a continuous distribution.

Bernoulli distribution

Bernoulli trial

A random experiment whose outcomes have been classified into two categories namely "success" and "failure" represented by letters S and F respectively is called a Bernoulli trail.

Bernoulli distribution

A discrete random variable X is said to have a Bernoulli distribution if its probability function is given by

$$f(x,p) = \begin{cases} p^x q^{1-x} & \text{for } x = 0,1 \\ 0, & \text{otherwise} \end{cases}$$

where p is the parameter of the distribution satisfying $0 \le p \le 1$ and $p + q = 1$.

Example: A coin is tossed in which the outcome "head" is a success and the probability of head is p. Then $q = 1 - p$ is the probability of failure or tail. If the number of heads or success is a random variable X, the X can take values 0 or 1 according to the outcome is tail (failure) or head (success). Then the probability function of X is:

$$f(x,p) = \begin{cases} p^x q^{1-x} & \text{for } x = 0,1 \\ 0, & \text{otherwise} \end{cases}$$

Binomial distribution

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Introduction

Binomial distribution was first derived by Swiss mathematician James Bernoulli (1654-1705) and was first published posthumously in 1913, eight years after his death.

Definition

A discrete random variable X is said to have a binomial distribution if its probability function is defined by

$$f(x;n,p)=\begin{cases}\binom{n}{x}p^x q^{n-x} & \text{for } x=0,1,2,\ldots,n\\ 0; \text{ otherwise}\end{cases}$$

where the two parameters n and p satisfy $0\le p\le 1$ and $p+q=1$, also n is positive integer.

Conditions or Assumptions of Binomial distribution:

a) There are two outcomes in each and every trial: one is termed as success and another is failure.

b) The number of trials n is fixed or finite.

c) The trials are independent of each other.

d) The probability of success p is constant for each and every trial.

Important properties of Binomial distribution:

Mean and variance of the distribution: If n and p are two parameters of the binomial distribution, then mean and variance of the distribution are:

$$Mean,\quad \mu = E(X) = np, \quad and\ Variance, \quad \sigma^2 = E(X^2) - \{E(X)\}^2 = npq$$

Some important properties of binomial distribution:

i) It is a discrete probability with parameter *n* and *p*.

ii) The mean of distribution is np and the variance is npq. The mean of the distribution is greater then variance since q < 1.

iii) The distribution is positively skewed $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(1-2p)^2}{npq}$ if *p* < ½ and negatively skewed if *p* > ½

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

iv) The distribution is symmetric $\beta_1 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{(1\ \ 6pq)}{npq}$ if $p = q = ½$. The distribution is mesokurtic if $pq = 1/6$, is platykurtic if $pq > 1/6$ and uptokurtic if $pq < 1/6$.

v) The distribution tends to Poisson distribution if the number of trials n tends to infinity.

vi) The distribution tends to normal distribution if n tends to infinity and neither p nor q are so small.

vii) $P(X = n) = p^n$ and $P(X = 0) = q^n$

Some practical applications of Binomial distribution:

i) Number of defective items in a randomly selected simple of 12 products.

ii) Number of days of increasing price index in share market in a randomly selected 15 days.

iii) Number of correct answers in a multiple choice test, if the students answer all the questions randomly.

iv) Number of worker suffers from occupational discuses in a randomly selected sample of 10 workers.

v) Number of female babies in a society.

vi) Number of successful hits in a target out of fixed number of hits.

vii) Number of customers purchase a particular commodity.

Example: In a community, the probability that a newly born child will be boy $\frac{2}{5}$. Among the 4 newly born children in that community, what is the probability that

(a) All the four boys

(b) No boys

(c) Exactly one boy.

Solution: Let us consider the event that a newly born child is a boy as success in Bernoulli trial with probability of success $\frac{2}{5}$. Let the number of boys be a random variable X. Then X can take values 0, 1, 2, 3, and 4.

According to binomial law, the probability function of X is

$$f\left(x, 4, \frac{2}{5}\right) = \binom{4}{x}\left(\frac{2}{5}\right)^x\left(\frac{3}{5}\right)^{4-x} \quad \text{for } x = 0,1,2,3,4.$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

a) $p(\text{all boys}) = p(x=4) = \binom{4}{4}\left(\frac{2}{5}\right)^4\left(\frac{3}{5}\right)^{4-4} = 0.0256$.

b) $p(\text{no boys}) = p(x=0) = \binom{4}{0}\left(\frac{2}{5}\right)^0\left(\frac{3}{5}\right)^{4-0} = 0.1296$.

c) $p(\text{exactly one boy}) = p(x=1) = \binom{4}{1}\left(\frac{2}{5}\right)^1\left(\frac{3}{5}\right)^{4-1} = 0.3456$.

Example: A fair coin is tossed 5 times. Find the probability of

a) exactly two heads

b) no head

Solution: Let the number of heads be a random variable X which can take values 0, 1, 2, 3, 4 and 5. Then X is binomial variate with $p = \frac{1}{2}$ and $n = 5$.

The probability function of X is

$$f\left(x, 5, \frac{1}{2}\right) = \binom{5}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x} \quad \text{for } x = 0,1,2,3,4,5$$

a) $p(\text{exactly two heads}) = p(x=2) = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{5-2} = 0.3125$.

b) $p(\text{no heads}) = p(x=0) = \binom{5}{2}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{5-0} = 0.03125$.

Example: Determine the binomial distribution for which mean is 4 and variance is 3.

Solution: Let X be a binomial variate with parameters n and p. Here, we have, $np = 4$ and $npq = 3$. Thus $\frac{npq}{np} = \frac{3}{4} \Rightarrow q = \frac{3}{4}$ and $p = 1 - q = 1 - \frac{3}{4} = \frac{1}{4}$. Then $n = \frac{4}{p} = \frac{4}{\frac{1}{4}} = 16$.

Hence, the binomial distribution is

$$f(x; n, p) = \begin{cases} \binom{16}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{16-x} & \text{for } x = 0,1,2,\ldots,16. \\ 0; \text{ otherwise} \end{cases}$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Example: We can consider the probability of passing a multiple choice exam paper comprising 20 questions each with 5 answers by random guessing.

The Bernoulli trial is a guess at the question, for which the probability of a pass is 0.2 (one fifth). It is important to distinguish between the fact there happen to be five answers but only two outcomes (you pick the right answer or you do not). If there were five possible outcomes, then this would not be a Bernoulli trial.

There are 20 trials (questions) so the probability of getting none right is,

$$p(0)=\binom{20}{0}0.2^0\left(1-0.2\right)^{20}=0.8^{20}=0.011529$$

The probability of getting one right is,

$$p(1)=\binom{20}{1}0.2^1\left(1-0.2\right)^{19}=0.2\times0.8^{19}=0.057646$$

As you can see, getting 4 out of 20 is the most likely score.

The probability of getting a score of less than 10, in other words the probability of failing, is the sum of the probabilities $p(0)$ through to $p(9)$ which is,

$$0.011529+0.057646+0.136909+\ldots+0.022161+0.007387=0.997405$$

So the probability of getting 10 or more answers right and therefore passing the examination by random guesswork is,

$$1-0.997405=0.002595\cong 0.26\%$$

Poisson distribution

Introduction: Poisson distribution was developed by France mathematician and physicist Simeon Denis Poisson (1781-1840), who published it in 1837.

Definition: A discrete random variable X is said to have a Poisson distribution if its probability function is given by

$$f\left(x;\lambda\right)=\begin{cases}\dfrac{\mathrm{e}^{-\lambda}\lambda^x}{x!} & \text{for } x=0,1,2,\ldots,\infty.\\ 0; & \text{otherwise}\end{cases}$$

where, $e=2.71828$ and λ is the parameter of the distribution which is the mean number of success and $\lambda=np$.

Theorem: If X is a poisson variate with parameter λ, then mean $=\lambda$ and variance $=\lambda$. Hence, mean and variance of poisson distribution are equal.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Examples

- The number of cars passing a certain street in time t.
- Number of suicide reported in a particular day.
- Number of faulty blades in a packet of 100.
- Number of printing mistakes at each page of a book.
- Number of air accidents in some unit of time.
- Number of deaths from a disease such as heart attack or cancer or due to snake bite.
- Number of telephone calls received at a particular telephone exchange in some unit of time.
- The number of defective materials in a packing manufactured by a good concern.
- The number of letters lost in a mail on a given day in a certain big city.
- The number of fishes caught in a day in a certain city.
- The number of robbers caught on a given day in a certain city.

Example: Suppose that the number of emergency patients in a given day at a certain hospital is a Poisson variable X with parameter $\lambda = 20$. What is the probability that in a given day there will be

a) 15 emergency patients.

b) At least 3 emergency patients.

c) More than 20 but less than 25 patients.

Solution: We know that,

$$f(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$$

Here, $\lambda = 20$, $\therefore f(x;20) = \begin{cases} \dfrac{e^{-20}(20)^x}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$

a) $p(\text{15 emergency patients}) = p(x = 15) = \dfrac{e^{-20}(20)^{15}}{15!} = 0.0516$.

b) $p(\text{at least 3 patients}) = p(x \geq 3) = 1 - p(x < 3)$

$$= 1 - p(x = 0) - p(x = 1) - p(x = 2)$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

$$= 1 - \frac{e^{-20}(20)^0}{0!} - \frac{e^{-20}(20)^1}{1!} - \frac{e^{-20}(20)^2}{2!} = 1.$$

c) $p(20 < x < 25) = p(x = 21) + p(x = 22) + p(x = 23) + p(x = 24)$

$$= \frac{e^{-20}(20)^{21}}{21!} + \frac{e^{-20}(20)^{22}}{22!} + \frac{e^{-20}(20)^{23}}{23!} + \frac{e^{-20}(20)^{24}}{24!} = 0.2841.$$

Example: If the probability that a car accident happens is a very busy road in on hour is 0.001. If 2000 cars passed in one hour by the road, what is the probability that

a) exactly 3
b) more than 2 car accidents happened on that hour of the road.

Solution: We know that,

$$f(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$$

Here, $p = 0.001$, $n = 2000$. $\therefore \lambda = np = 2000 * 0.001 = 2$.

$$\therefore f(x;2) = \begin{cases} \dfrac{e^{-2}(2)^x}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$$

a) $p(\text{exactly 3 accidents}) = p(x = 3) \; = \dfrac{e^{-2}(2)^3}{3!} = 0.18$.
b) $p(\text{more than 2 accidents}) = p(x > 2) = 1 - p(x \le 2)$

$$= 1 - p(x = 0) - p(x = 1) - p(x = 2)$$

$$= 1 - \frac{e^{-2}(2)^0}{0!} - \frac{e^{-2}(2)^1}{1!} - \frac{e^{-2}(2)^2}{2!} = 0.323.$$

Example: A factory produces blades in a packet of 10. The probability of a blade to be defective is 0.2%. Find the number of packets having two defective blades in a consignment of 10,000 packets.

Solution: We know that,

$$f(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$$

Here, $p = 0.2\% = 0.002$, $n = 10$. $\therefore \lambda = np = 10 * 0.002 = 0.02$.

$$\therefore p(\text{2 defective blades}) = p(x = 2) \; = \frac{e^{-0.02}(0.02)^2}{2!} = 0.000196.$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

Therefore, the total number of packets having two defective blades in a consignment of 10,000 packet is $10000 \times 0.000196 = 1.96 \approx 2$.

Example: What probability model is appropriate to describe a situation where 100 misprints are distributed randomly throughout the 100 pages of a book? For this model, what is the probability that a page observed at random will contain at least three misprints?

Solution: We know that,

$$f(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$$

we have,

$p = \dfrac{1}{100} = 0.01$ (because there is only one mistake on the average in a page), $n = 100$.

$\therefore \lambda = np = 100 \times 0.01 = 1$.

$$\begin{aligned} \therefore p(\text{at least 3 misprints}) &= p(x \geq 3) = 1 - p(x < 3) \\ &= 1 - p(x=0) - p(x=1) - p(x=2) \\ &= 1 - \frac{e^{-1}(1)^0}{0!} - \frac{e^{-1}(1)^1}{1!} - \frac{e^{-1}(1)^2}{2!} = 0.0803 \end{aligned}$$

Normal Distribution

Introduction: Normal distribution is the most important probability distribution in Statistics. The normal distribution was first developed in 1733 by English mathematician De Moivre.

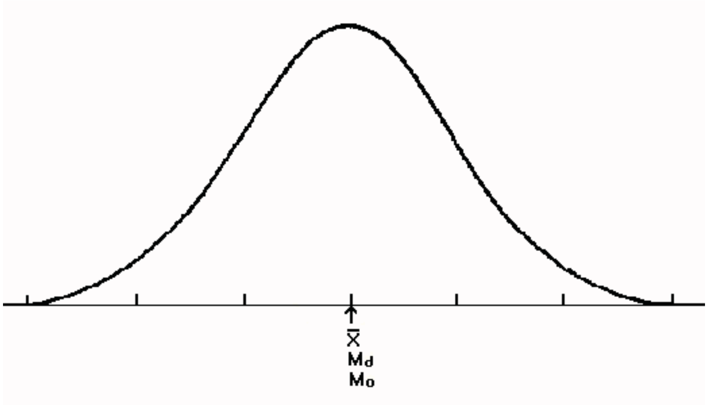
Definition: A continuous random variable X is said to have a normal distribution if its density function is given by

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

where, the parameters μ and σ^2 satisfy $-\infty < \mu < \infty$ and $\sigma^2 > 0$.

The variable X whose density function given in (1) is called normal variate with parameters μ and σ^2 and is denoted by $N(\mu,\sigma^2)$. The parameters μ and σ^2 are actually the mean and variance of the normal variate X. The graph of the normal curve is

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU



Standard Normal Variate: If X is a normal variate with parameters μ and σ^2, then $Z=\dfrac{X-\mu}{\sigma}$ is a standard normal variate with mean zero and variance unity. The density function of Z is

$$f(z,0,1)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}; \qquad -\infty<z<\infty$$

Important Properties of Normal Distribution:

a) The distribution is symmetric about μ.

b) The mean, median and mode of the distribution is equal.

c) The mean of the distribution is μ and variance is σ^2.

d) The curve has a single peak, i.e. it is unimodal.

e) $\mu\pm\sigma$, $\mu\pm2\sigma$, $\mu\pm3\sigma$ covers 68.27%, 95.45% and 99.73% area respectively.

f) All odd central moments of the distribution are zero.

g) Most of the distributions occurring in practice can be approximated by normal distribution. Moreover, many of the sampling distributions e., g., Student's t, Snedecor's F, Chi-square distributions etc tend to normal for large samples.

h) Normal distribution finds large applications in Statistical Quality Control in industry for setting control limits.

i) Skewness is zero that is $\beta_1=0$ and kurtosis is 3 that is $\beta_2=3$.

Important Uses of Normal Distribution:

a) Normal distribution is the basis of all sampling distribution. Without the assumption of normality, sampling distribution has no existence.

b) Assumption of normality is the basis of all parametric test of significance.

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

c) Normal distribution finds its application in industrial statistics such as quality control.

d) According to central limit theorem, if mean and variance of a distribution exists, then the distribution converted to normal distribution.

Note: Let X be a continuous random variable with a cumulative distribution function $F(x)$ and let a and b be two possible values of X, with $a < b$. The probability that X lies between a and b is

$$p(a < x < b) = F(b) - F(a)$$

Example: A company produces light bulbs whose life times follows a normal distribution with mean 1200 hours and standard deviation 250 hours. If a light bulb is chosen randomly from the company's output, what is the probability that it's life time will be between 900 and 1300 hours?

Solution: Let X represent life time in hours. Then

$$\begin{aligned} p(900 < x < 1300) &= p\left(\frac{900-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{1300-\mu}{\sigma}\right) \\ &= p\left(\frac{900-1200}{250} < z < \frac{1300-1200}{250}\right) \\ &= p(-1.2 < z < 0.4) \\ &= p(-\infty < z < 0.4) - p(-\infty < z < -1.2) \\ &= 0.65542 - 0.11507 \quad \text{(By using Normal table)} \\ &= 0.54035 \end{aligned}$$

Hence, the probability is approximately 0.54 that a light bulb will last between 900 and 1300 hours.

Example: A very large group of students obtains test scores that are normally distributed with mean 60 and standard deviation 15. What proportion of students obtained scores

a) Less than 85.
b) More than 90.
c) Between 85 and 95.

Solution: Let X denote the test score. Then

a) $p(x < 80) = p\left(\frac{X-\mu}{\sigma} < \frac{85-\mu}{\sigma}\right) = p\left(z < \frac{85-60}{15}\right)$

$$= p(z < 1.67) = p(-\infty < z < 1.67)$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

$$= 0.9525 \text{. (By using Normal table).}$$

That is 95.25% of the students obtained scores less than 80.

b) $$p(x > 90) = p\left(\frac{X-\mu}{\sigma} > \frac{90-\mu}{\sigma}\right) = p\left(z > \frac{90-60}{15}\right)$$
$$= p(z > 2) = 1 - p(z < 2) = 1 - p(-\infty < z < 2)$$
$$= 1 - 0.9772 = 0.0228 \text{. (By using Normal table).}$$

That is 2.28% of the students obtained scores more than 90.

c) $$p(85 < x < 95) = p\left(\frac{85-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{95-\mu}{\sigma}\right)$$
$$= p\left(\frac{85-60}{15} < z < \frac{95-60}{15}\right)$$
$$= p(1.67 < z < 2.33)$$
$$= p(-\infty < z < 2.33) - p(-\infty < z < 1.67)$$
$$= 0.9901 - 0.9525$$
$$= 0.03756$$
(By using Normal table)

That is 3.76% of the students obtained scores in the range 85 to 95.

Example: The average daily sales of 500 branch office were Tk. 150 thousands and the standard deviation Tk. 15 thousands. Assuming the distribution to be normal indicate how many branches have sales between

a) Tk. 120 thousands and Tk. 145 thousands.
b) Tk. 140 thousands and Tk. 165 thousands.

Solution: Let X be the average daily sales of 500 branch office.

a) $$p(120 < x < 145) = p\left(\frac{120-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{145-\mu}{\sigma}\right)$$
$$= p\left(\frac{120-150}{15} < z < \frac{145-150}{15}\right)$$
$$= p(-2 < z < -0.33)$$
$$= p(-\infty < z < -0.33) - p(-\infty < z < -2)$$
$$= 0.3707 - 0.02275 = 0.34795 \text{ (By using Normal table)}$$

Hence, the expected number of branches having sales between Tk. 120 thousands and Tk. 145 thousands are

$$0.3479 \times 500 = 173.95 \approx 174 .$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU

b) $$p(140 < x < 165) = p\left(\frac{140-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{165-\mu}{\sigma}\right)$$
$$= p\left(\frac{140-150}{15} < z < \frac{165-150}{15}\right)$$
$$= p(-0.67 < z < 1)$$
$$= p(-\infty < z < 1) - p(-\infty < z < -0.67)$$
$$= 0.84434 - 0.25143 = 0.58991 \text{ (By using Normal table)}$$

Hence, the expected number of branches having sales between Tk. 140 thousands and Tk. 165 thousands are

$$0.58991 \times 500 = 294.955 \approx 295.$$

Dr. Mohd. Muzibur Rahman
Professor, Department of Statistics, JU