1. Introduction

Having studied probability in standard XII, now we shall proceed to study some probability distributions viz. Binomial Distribution, Poisson's Distribution and Normal Distribution.

2. Random Variable

In probability problems it is convenient to think of a variable and consider the outcomes as the values, the variable takes. In a toss of a die we think of a variable x which takes values 1, 2, 3, 4, 5, 6. In a toss of a coin we think of a variable x which takes values 0 and 1 (the number of heads obtained). Such a variable is called **a random variable.** In simple words a variable used to denote the numerical value of the outcome of an experiment is called **a random variable.**

If the random variable x takes only a finite or countably infinite values it is called **a discrete random variable** e.g. if x denotes the result of a toss of a die or if x denotes the result of shots fired till the target is hit, x is a discrete random variable. On the other hand if x takes uncountably infinite values then it is called **a continuous random variable.** For example, if x is the height of a person selected at random or x is the age of the person then x is a continuous random variable because x can take any value between a specified range.

3. Probability distributions

Suppose x denotes the outcome of the toss of a die. Then x takes the values 1, 2, 3, 4, 5, 6 each with probability 1/6. We can put this information in a table. Such a table giving the probabilities of x along with its values is called probability distribution table. The way in which the probabilities are distributed over the values of the variable or the totality of the probabilities together with the corresponding values of x is called the **probability distribution.**

x	$p(x)$
1	1 / 6
2	1 / 6
3	1 / 6
4	1 / 6
5	1 / 6
6	1 / 6
Total	1

In general if x is a discrete random variable and we know the probability p_i that x will take the value x_i for all values of x_i then the probability distribution is called a **discrete probability distribution.** The probabilities p_i may be given in a tabular form or by a formula satisfying the conditions that **(i)** $0 \leq p_i \leq 1$ for all i and **(ii)** $\Sigma\, p_i = 1$. Then p is called **probability density function.** Generally, the capital letter 'P' stands for the term 'probability' and small letter p or p_i or $p\,(x_i)$ for its value.

If x is a continuous random variable and we know the probability that x will be in a given interval (a, b) then the probability distribution is called ***a continuous probability distribution.*** Usually a continuous probability distribution is given in the form of a function $f(x)$ and $f(x)$ is called **probability density function. The area under the curve from a to b is equal to the probability that x will be in interval (a, b). The total area under the curve is equal to 1.**

i.e. $\int_{-\infty}^{\infty} f(x)\, dx = 1$ and $p\,(a < x < b) = \int_a^b f(x)\, dx$ and $f(x)\ {}^{3}\ 0$ for all x.

4. Discrete Probability Distributions

Ex. 1 : The probability density function of a random variable x is zero except at $x = 0, 1, 2$. At these points $P\,(0) = 3c^3$, $P\,(1) = 4c - 10c^2$, $P\,(2) = 5c - 1$. (i) Find c, (ii) Find $P\,(0 < x \leq 2)$

Sol. : (i) Since $\Sigma P_i = 1$, we have

$$3c^3 + 4c - 10c^2 + 5c - 1 = 1 \quad \therefore\ 3c^3 - 10c^2 + 9c - 2 = 0$$

$$\therefore\ (3c - 1)(c - 2)(c - 1) = 0$$

$$\therefore\ c = \frac{1}{3} \quad \text{The other values are inadmissible as } 0 \leq p_i \leq 1$$

(ii) The probability distribution now is

x	0	1	2	Total
$p\,(x)$	1/9	2/9	2/3	1

$$P\,(0 < x \leq 2) = P(x = 1) + P(x = 2) = \frac{2}{9} + \frac{2}{3} = \frac{8}{9}.$$

Exercise

1. A random variable x has the following probability distribution

x	1	2	3	4	5	6	7
$p\,(x)$	k	$2k$	$3k$	k^2	$k^2 + k$	$2k^2$	$4k^2$

6. Binomial Distribution

Definition : If (i) an experiment results only in two ways, success or failure, (ii) the probability of success is p and the probability of failure is q such that $p + q = 1$ and (iii) the experiment is repeated n times then probability of r successes is given by

$$\boxed{\begin{aligned} & p(x = r) = {}^nC_r\, p^r q^{n-r} \\ \text{i.e.}\quad & p(x = r) = \frac{n(n-1)(n-2).....(n-r+1)}{r!} p^r q^{n-r} \end{aligned}}$$

and the resulting probability distribution is called **Binomial.**

The probabilities that x will take the values 0, 1, 2, 3 ..... r ..... n can also be expressed as follows.

Thus, $p(x = 0) = q^n$, $p(x = 1) = n\,p\,q^{n-1}$,

$$p(x = 2) = \frac{n(n-1)}{2!} p^2 q^{n-2} \text{ etc.}$$

A variate having binomial distribution is called **binomial variate.** Further, the characteristics n, p, q **are called the parameters or the constants** of the distribution.

If the values of the parameters n, p, q are known the probability distribution is completely known, because we can then calculate the probabilities of 0, 1, 2, ... n successes by using the above formula.

Uses : Thus, in problems involving **(1)** the tossing of a coin-heads or tails, **(2)** the result of an examination-success or failure, **(3)** the result of a game-win or loss, **(4)** the result of inspection-acceptance or rejection where there are two mutually exclusive and exhaustive outcomes and where trials are finite, the binomial distribution can be used.

Properties of the binomial distribution

We give below some important properties of the binomial distribution without derivation.

(i) If x denotes binomial variate, the mean of the distribution is given by,

$$\boxed{\bar{x} = np}$$

(ii) The variance of the distributions is npq i.e.

$$\boxed{V(x) = npq}$$

Ex. 1 : If 10% bolts produced by a machine are defective, calculate the probability that out of a sample selected at random of 10 bolts, not more than one bolt will be defective. **(S.U. 1986, 90, 91, 92)**

Sol. : We have P (defective bolt) $= p = 0{\cdot}1$, $q = 0{\cdot}9$ and $n = 10$

P (not more than one bolt defective)

$= P$ (Zero defective) + P (one defective)

By Binomial Distribution

$$P(x) = {}^{n}C_{x}\, p^{x} q^{n-x} = {}^{10}C_{x}(0{\cdot}1)^{x}\,(0{\cdot}9)^{10-x}$$

$$\therefore \quad P(x=0) = {}^{10}C_{0}(0{\cdot}1)^{0}\,(0{\cdot}9)^{10} = 0{\cdot}3487$$

$$P(x=1) = {}^{10}C_{1}\,(0{\cdot}1)^{1}\,(0{\cdot}9)^{9} = 0{\cdot}3874$$

$\therefore$ Required probability $= P(x=0) + P(x=1)$

$= 0{\cdot}7361$

Ex. 2 : From a box containing 100 transistors 20 of which are defective, 10 are selected at random. Find the probability that (i) all will be defective (ii) all are non-defective (iii) at least one is defective. **(S.U. 1985, 88, 93, 98)**

Sol.: P(defective transistor) $= p = \dfrac{20}{100} = 0{\cdot}2$, $q = 0{\cdot}8$ and $n = 10$

$$P(x\text{ def.}) = {}^{n}C_{x}\, p^{x} q^{n-x} = {}^{10}C_{x}(0{\cdot}2)^{x}\,(0{\cdot}8)^{10-x}$$

(i) $P(\text{all def.}) = P(x=10) = {}^{10}C_{10}(0{\cdot}2)^{10}\,(0{\cdot}8)^{0} = 0{\cdot}0000001$

(ii) $P(\text{zero def.}) = P(x=0) = {}^{10}C_{0}(0{\cdot}2)^{0}\,(0{\cdot}8)^{10} = 0{\cdot}1074$

(iii) P (at least one defective) $= 1 - P$ (zero defective)

$= 1 - 0{\cdot}1074 = 0{\cdot}8926.$

4. If the probability of defective bulbs is 0·2, find the mean and variance of the distribution of defective bulbs in a lot of 1000 bulbs.

[**Ans. :** $\bar{x} = np = 200$, $var = npq = 160$]

5. If the mean of the binomial distribution is 2 and the variance is 4/3 find the probability of (i) two successes, (ii) less than two successes.

(**Hint :** Find n, p, q.) [**Ans. : (i)** 0·3292, **(ii)** 0·3512]

(B) 1. An examination containing multiple choice questions is designed so that the probability of a correct choice for any question by guessing alone is 0·2. What is the probability that a student will not get more than four questions right out of 20 merely by guessing ? **(S.U. 1986)** [**Ans. :** 0·4114]

2. In a hurdle race a player has to cross 10 hurdles. The probability that he will clear each hurdle is 5 / 6. What is the probability that he will knock down less than two hurdles ? **(S.U. 1997)** [**Ans. :** 0·4845]

3. The probability that a worker will suffer from an occupational disease is 25%. What is the chance than out of 6 workers 4 or more will suffer from the disease ? **(S.U. 1986)** [**Ans. :** 0·0376]

4. 10% of the tools produced in a certain manufacturing process turn out to be defective.

(a) Find the probability that in a sample of 10 tools chosen at random exactly two will be defective. **(S.U. 1987)** [**Ans. :** 0·1937]

(b) Find the probability that out of 20 tools selected at random there are (i) exactly two defectives (ii) at least two defectives.

(S.U. 1989, 2003) [**Ans. : (i)** 0·2852, **(ii)** 0·3918]

5. If the probability that a new-born child is male is 0·6, find the probability that in a family of 5 children there will be exactly 3 boys. **(S.U. 1985, 2001)**

[**Ans. :** 0·3456]

6. If in a lot of 500 solenoids 25 are defective, find the probability that there will be 0, 1, 2 defective solenoids in a random sample of 20 solenoids.

(S.U. 1985) [**Ans. : (i)** 0·3585, **(ii)** 0·3773, **(iii)** 0·1887]

7. In a room there are three lamp-sockets. A bag contains 6 working and 4 non-working bulbs. Three bulbs are selected at random and fitted in the sockets. Find the probability that there will be some light in the room.

(S.U. 1989) [**Ans. :** 29 / 30]

8. If on an average one candidate out of 10 fails in a certain examination, find the chance that out of 5 candidates that have appeared for the examination at least 4 will be successful. **(S.U. 1985)** [**Ans. :** 0·918]

9. In a locality 20% of people smoke. Find the probability that out of 6 persons chosen at random from this locality 4 or more will we found to be smokers. **(S.U. 1985, 2002)** [**Ans. :** 0·0169]

7. Poisson Distribution

Poisson distribution was discovered by the French Mathematician Poisson in 1837. It is the limiting case of binomial distribution when n, the number of trials tends to infinity and p, the probability of success in each trial remains constant and tends to zero such that $np = m$.

Definition : A random x is said to follow Poisson distribution if the probability of x is given by

$$\boxed{P(x) = \frac{e^{-m} m^x}{x!}} \qquad x = 0, 1, 2 \ldots\ldots$$

and m (< 0) is called the parameter of the distribution.

When do we get Poisson distribution : In some cases where an experiment results in two outcomes success or failure, the number of successes only can be observed and not the number of failures. We can observe how many accidents occur on the road but we cannot observe how many accidents do not occur. We can observe how many persons die of cancer but we cannot observe how many do not die of cancer. In such cases Binomial distribution cannot be used. We use Poisson's. For Poisson distribution the following conditions must be satisfied.

(i) $n \to \infty$, (ii) $p \to 0$, (iii) $np = m$.

Uses : We get a Poisson distribution when the above conditions are satisfied. We use it in the investigations involving;

(i) the number of deaths due to a disease such as heart-attack, cancer, etc.

(ii) the number of accidents during a week or a month, etc.

(iii) the number phone-calls received at a particular telephone exchange during a period of time,

(iv) the number of cars passing a particular point on a road during a period of time,

(v) the number of defective articles in a lot,

(vi) the number of printing mistakes on a page of a book.

Constants of Poisson Distribution : We state below without proof the following two properties of Poisson Distribution.

(i) The mean = m
(ii) The variance = m

Ex. 1. Between 2 and 4 p.m. the average number of phone calls per minute coming into a switch board of a company is 2·5. Find the probability during a minute there will be (i) no phone call (ii) exactly 3 calls. **(S.U. 1985)**

Sol. : For a Poisson distribution $P(x) = \dfrac{e^{-m} m^x}{x!}$

We are given that the mean,

m = average number of phone calls = 2·5.

(i) $\therefore \quad P(x=0) = \dfrac{e^{-2\cdot5}(2\cdot5)^0}{0!} = 0\cdot08208$

(ii) $P(x=3) = e^{-2\cdot5}\dfrac{(2\cdot5)^3}{3!} = 0\cdot2138$

Ex. 2 : A firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson variate with mean 1·5. Calculate the probable number of days in a year on which (i) neither car is in demand, (ii) a demand is refused. **(S.U. 1985)**

Sol. : Let x be the number of cars in demand. Since x is Poisson with mean $m = 1\cdot5$.

$$P(x) = e^{-1\cdot5}\frac{(1\cdot5)^x}{x!}$$

$$P(x=0) = e^{-1\cdot5}\frac{(1\cdot5)^0}{0!} = 0\cdot2231$$

$$P(x=1) = e^{-1\cdot5}\frac{(1\cdot5)^1}{1!} = 0\cdot3347$$

$$P(x=2) = e^{-1\cdot5}\frac{(1\cdot5)^2}{2!} = 0\cdot2510$$

$\therefore$ P (demand is refused) = P (demand greater than 2)

$= P(x>2) = 1 - [P(x=0) + P(x=1) + P(x=2)]$

$= 1 - [0\cdot2231 + 0\cdot3347 + 0\cdot2510] = 0\cdot1912$

$\therefore$ No. of days on which there is no demand in a year

$= Np\,(x=0) = 365 \times 0\cdot2231 = 81$ nearly.

No. of days on which demand is more than 2

$= Np\,(x>2) = 365 \times 0\cdot1912 = 70$ nearly.

Ex. 3 : If the probability that an individual suffers a bad reaction from a certain injection is 0·001, determine the probability that out of 2000 individuals (i) exactly 3, (ii) more than 2 will suffer a bad reaction.

(S.U. 1987, 94, 95, 97, 98)

Sol. : We are given $n = 2000$, $p = 0\cdot001$. But $m = np = 2000 \times 0\cdot001 = 2$.

[Ans. : (a) 3,

(B) 1. Accidents occur on a particular stretch of highway at an average rate 3 per week. What is the probability that there will be exactly two accidents in a given week ? [Ans. : 0·224]

2. Find the probability that at most 4 defective bulbs will be found in a box of 400 bulbs, if it is known that 1 per cent of the bulbs are defective.

[Ans. : 0·6282]

3. It is 1 in 1000 that an article is defective. There are in a box 100 articles of this type. Assuming Poisson distribution find the probability that the box contains (i) no detective, (ii) two or more detectives. **(S.U. 1998)**

[Ans. : 0·3679, 0·2642]

4. If the probability that an individual suffers a bad reaction from a particular injection is 0·001, determine the probability that out of 2,000 individuals (i) exactly four, (ii) more than two individuals will suffer a bad reaction.

[Ans. : **(i)** 0·0902, **(ii)** 0·3233]

5. The number of accidents on a particular highway in a month is a Poisson variate with parameter 5. Find the probability that more than 2 accidents have occurred on the road in a given month. [Ans. : 0·8754]

6. A book contains 100 misprints distributed randomly throughout 100 pages. What is the probability that a page observed at random contains at least two misprints ? **(S.U. 1988)** [Ans. : 0·2642]

7. It is known that in a certain plant, there are on an average 4 industrial accidents per month. Find the probability that in a given month there will be less than 4 accidents.

[Ans. : 0·4335]

8. An insurance company found that only 0·01 per cent of the population is involved in a certain type of accident each year. If its 1000 policy-holders were randomly selected from the population, what is the probability that more than two of its clients are involved in such an accident next year ?

(S.U. 1990) [Ans. : 0·0002]

8. Normal Distribution

So far we discussed probability distributions of discrete type. The normal distribution developed by Gauss is a continuous distribution of maximum utility. We first define a continuous probability distribution in general and then the normal distribution in particular.

Definition : If we know a curve such that the area under the curve from $x = \alpha$ to $x = \beta$ is equal to the probability that x will take a value between α and β and that the total area under the curve is unity, then the curve is called the **probability curve**. If the curve is described by a relation between y and x [*i.e.* by $y = \Phi(x)$] then the relation $y = \Phi(x)$ [$\Phi(x) \geq 0$ for all x] is called the **probability density function or simply probability function.**

The most important probability curve is the Normal Curve. The corresponding probability function is called the normal probability function and the probability distribution defined by it is called **Normal Probability Distribution**. The normal distribution can be derived from the binomial distribution if the number of trials n is made to approach infinity.

The normal distribution is given by

$$\boxed{y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}}$$

where y = ordinate or the y-coordinate of a point on the curve,

x = abscissa or the x-coordinate of a point on the curve,

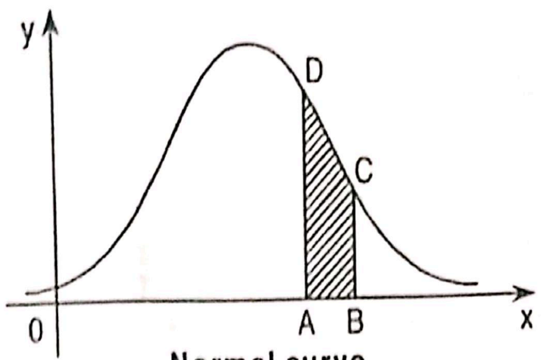
m = the mean of x, σ = the S.D. of x,

π = a constant 3·1416, e = a constant 2·7183.

For example, if x is normally distributed with mean 15 and variance 16, the distribution is given by

$$y = \frac{1}{4\sqrt{(2\pi)}} e^{-\frac{1}{2}\left(\frac{x-15}{4}\right)^2}$$

The normal curve is in the shape of a bell, symmetrical about the highest ordinate. If A and B are two points on the x-axis, the shaded area, $ABCD$, bounded by the curve, the ordinates at A *i.e.* $x = \alpha$ and at B *i.e.* $x = \beta$ and the x-axis is equal to the probability that x lies between α and β. The total area bounded by the curve and the x-axis is equal to 1. The probability distribution defined in this manner by means of a normal curve is called normal probability distribution. The distribution is completely known if the values of m and σ are known.



Normal curve

They are called the **parameters** of the distribution. The normal distribution with mean m and *s.d.* σ is denoted by $N(m, \sigma)$.

The Standardised Normal Variate (S.N.V.)

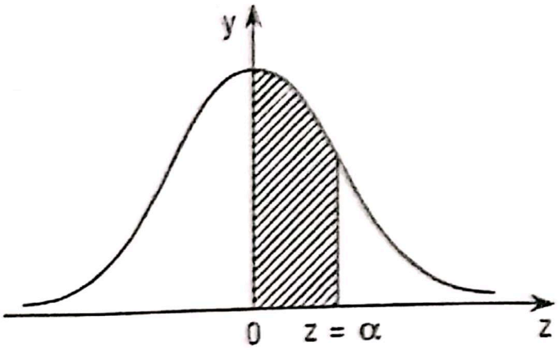
We have seen above that for a normal distribution the probability that the variable x will take a value between α and β is equal to the area under the curve enclosed by the ordinates at $x = \alpha$ and $x = \beta$. As such the problem of finding the probability reduces to that of finding the area between the two ordinates. But for different values of m and σ we get different normal curves. The problem of finding area between two given values will multiply into too many problems if we are to find the area between the given values for different curves with different values of m and σ. All such problems can be reduced to a single one by reducing all normal distributions to a single normal distribution called **Standardised Normal Distribution** by writing $z = \frac{x - m}{\sigma}$. This actually amounts to shifting the origin to m and reducing its scale by σ. The equation of the curve then reduces to.

$$\boxed{y = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}}$$

The mean m, of this distribution is obviously zero and standard deviation σ is one. This is called standard normal distribution. It is denoted by $N(0, 1)$ **or by S.N.V.**

The areas under the curve between the y-axis ($z = 0$) and various ordinates at $z = \alpha$ are given in the table. This area is equal to the probability that z will assume a value between $z = 0$ and $z = \alpha$.

$\therefore\ P(0 < z < \alpha) =$ area between $z = 0$ and $z = \alpha$.



If z_1 and z_2 are two values of z corresponding to the values x_1 and x_2 of x, then

$$P(x_1 < x < x_2) = P(z_1 < z < z_2)$$

= **area under the normal curve from z_1 to z_2.**

Note

1. Total area under the normal curve from $z = -\infty$ to $z = \infty$ is 1.

2. Because of symmetry the area from $z = 0$ to $z = \infty$ is equal to the area from $z = -\infty$ to $z = 0$ is equal to 0·5.

Ex. 1 : Sacks of sugar packed by an automatic loader have an average weight at hundred kilograms with standard deviation of two hundred fifty grams. Assuming a normal distribution, find the chance of getting a sack weighing less than 99·5 kilograms. (Given : For S.N.V. z area from $z = 0$ to $z = 2$ is 0·4772)

Sol. : S.N.V. $z = \dfrac{x - m}{\sigma} = \dfrac{x - 100}{0{\cdot}25}$

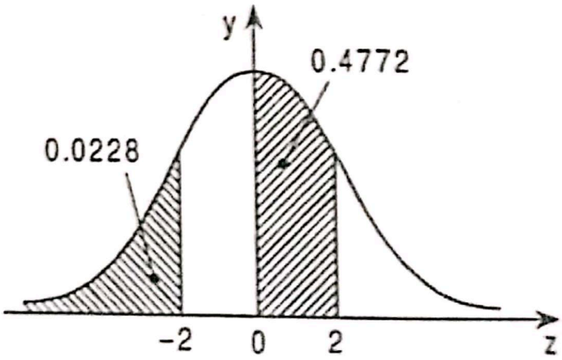
When $x = 99{\cdot}5$,

$$z = \frac{99{\cdot}5 - 100}{0{\cdot}250} = \frac{-0{\cdot}5}{0{\cdot}25} = -2$$

$P(x < 99{\cdot}5) = P(z < -2)$

= area to the left of ($z = -2$)

$= 0{\cdot}5 - 0{\cdot}4772 = 0{\cdot}0228.$



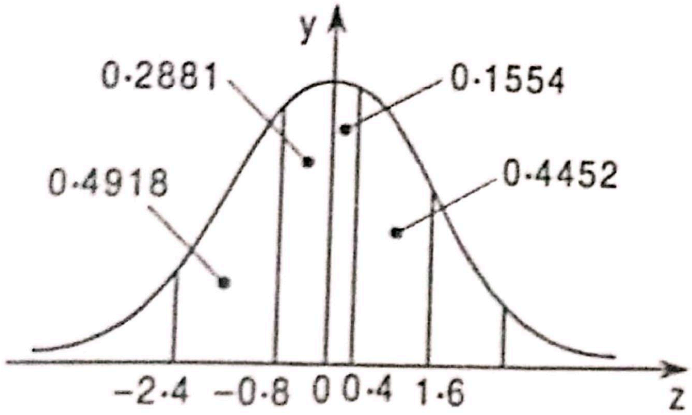
Ex. 2 : In a sample of 1000 students the mean and standard deviation of marks obtained by the students in a certain test are 14 and 2·5. Assuming the distribution to be normal find the number of students getting marks (i) between 12 and 15, (ii) above 18, (iii) below 8.

(Given : For a S.N.V. z area between $z = 0$ and $z = 0{\cdot}4$ is 0·1554, that between $z = 0$ and $z = 0{\cdot}8$ is 0·2881, that between $z = 0$ and $z = 1{\cdot}6$ is 0·4452, that between $z = 0$ and $z = 2{\cdot}4$ is 0·4918.) **(S.U. 1986, 95, 96, 2002)**

Sol. : We have $z = \dfrac{x - m}{\sigma} = \dfrac{x - 14}{2 \cdot 5}$

(i) When $x = 12$

$$z = \frac{12 - 14}{2 \cdot 5} = -\frac{2}{2 \cdot 5} = -0 \cdot 8$$



When $x = 15$,

$$z = \frac{15 - 14}{2 \cdot 5} = \frac{1}{2 \cdot 5} = 0 \cdot 4$$

$P(12 < x < 15) = P(-0 \cdot 8 < z < 0 \cdot 4)$

= area between

$(z = -0 \cdot 8$ and $z = 0 \cdot 4)$

$= 0 \cdot 2881 + 0 \cdot 1554$

$= 0 \cdot 4435.$

$\therefore$ No. of students $= Np = 1000 \times 0 \cdot 4435 = 443.$

(ii) When $x = 18$, $z = \dfrac{18 - 14}{2 \cdot 5} = \dfrac{4}{2 \cdot 5} = 1 \cdot 6$

$\therefore$ $P(x > 18) = P(z > 1 \cdot 6)$

= area to the right of $(z = 1 \cdot 6) = 0 \cdot 5 - 0 \cdot 4452 = 0 \cdot 0548.$

$\therefore$ No. of students $= Np = 1000 \times 0 \cdot 0548 = 55.$

(iii) When $x = 8$, $z = \dfrac{8 - 14}{25} = -\dfrac{6}{2 \cdot 5} = -2 \cdot 4$

$P(x < 8) = P(z < -2 \cdot 4)$

= area to the left of $(z = -2 \cdot 4) = 0 \cdot 5 - 0 \cdot 4918 = 0 \cdot 0082.$

$\therefore$ No. of students $= Np = 1000 \times 0 \cdot 0082 = 8.$