Index

Unit – 3 ⇝ Probability Distribution

Method 1 ⇝ Binomial Distribution

Introduction

→ In this chapter we shall study some of the probability distribution that figure most prominently in statistical theory and application. We shall also study their parameters. We shall introduce number of discrete probability distribution that have been successfully applied in a wide variety of decision situations.

→ The purpose of this chapter is to show the types of situations in which these distributions can be applied.

→ Probability function of discrete random variable is known as probability mass function (P.M.F.) and probability function of continuous random variable is known as probability density function (P.D.F.).

→ Some special probability distributions:

(1) Binomial distribution (P.M.F.)

(2) Poisson distribution (P.M.F.)

(3) Normal distribution (P.D.F.)

(4) Exponential distribution (P.D.F.)

Bernoulli Trials

→ Suppose a random experiment has two possible outcomes, which are

(1) Success (S)

(2) Failure (F)

→ If the probability $p(0 < p < 1)$ of getting success at each of the n trials of this experiment is constant, then the trials are known as Bernoulli trials.

Binomial Distribution

→ A random experiment consists of n Bernoulli trials such that the trials are independent.

→ Each trial results in only two possible outcomes, labeled as success and failure.

→ The probability of success in each trial remains constant.

→ The random variable X that equals the number of trials that results in a success is a binomial random variable with parameters $0 < p < 1,\ q = 1 - p$ and $n = 1, 2, 3, \ldots.$

→ The probability mass function of X is

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

→ For Examples:

(1) Number of defective bolts in a box containing n bolts.

(2) Number of post-graduates in a group of n people.

(3) Number of oil wells yielding natural gas in a group of n wells test drilled.

(4) In the next 20 births at a hospital. Let X = the number of female births.

(5) Flip a coin 10 times. Let X = number of heads obtained.

Properties of Binomial Distribution

→ The mean of binomial distribution is defined as $\mu = E(X) = np$.

→ The variance of binomial distribution is defined as $V(X) = npq$.

→ The standard deviation of binomial distribution is defined as $\sigma = \sqrt{npq}$.

Example of Method-1: Binomial Distribution

C	1	For the binomial distribution with $n = 20, p = 0.35$. Find mean, variance and standard deviation. **Answer**: $\textbf{Mean} = 7$, $\textbf{Variance} = 4.55$, $\textbf{Standard Deviation} = 2.1331$
C	2	Obtain the binomial distribution function for which mean is 10 and variance is 5. **Answer**: $P(X = x) = \binom{20}{x} (0.5)^x (0.5)^{20-x} \; ; x = 0, 1, 2, \dots, 20.$
C	3	12% of the tablets produced by a tablet machine are defective. What is the probability that out of a random sample of 20 tablets produced by the machine, 5 are defective? **Answer**: 0.0567

C	4	Find probability of getting a sum of 7 at least once in 3 tosses of a pair of dice. **Answer**: $\frac{91}{216}$
C	5	The probability that India wins a cricket test match against Australia is given to be $\frac{1}{3}$. If India and Australia play 3 test matches, what is the probability (a) India will lose all the three test matches? (b) India will win at least one test match? **Answer**: (a) 0.2963, (b) 0.7037
C	6	The probability that in a university, a student will be a post-graduate is 0.6. Determine probability that out of 8 students none, two and at least two will be post-graduate. **Answer**: 0.0007, 0.0413, 0.9914
C	7	The probability that an infection is cured by a particular antibiotic drug within 5 days is 0.75. Suppose 4 patients are treated by this antibiotic drug. What is the probability that no patient, exactly two patients, and at least two patients, are cured? **Answer**: 0.0039, 0.2109, 0.9492
C	8	By assuming equal probabilities for boys and girls. Out of 800 families with 4 children each, how many would you expect to have (a) 2 boys and 2 girls? (2) at least 1 boy? (3) at most 2 girls? (4) no girl? **Answer**: (a) 300, (b) 750, (c) 550, (d) 50
C	9	Find the probability that in five tosses of a fair die, 3 will appear twice, at most once, at least two times. **Answer**: $\frac{625}{3888}$, $\frac{3125}{3888}$, $\frac{763}{3888}$

Method 2 ⇝ Poisson Distribution

Poisson Distribution

→ A discrete random variable X is said to follow Poisson distribution if it assumes only non-negative values.

→ Its probability mass function is given by

$$\mathbf{P(X = x) = \frac{e^{-\lambda}\,\lambda^{x}}{x\,!}\,;\ x = 0, 1, 2, 3, \ldots}$$

Where, λ = mean of the poisson distribution

→ For Examples:

(1) Number of telephone calls per minute at a switchboard.

(2) Number of cars passing a certain point in one minute.

(3) Number of printing mistakes per page in a large text.

(4) Number of persons born blind per year in a large city.

Properties of Poisson Distribution

→ The random variable X should be discrete.

→ The number of trials n is very large.

→ The probability of success p is very small (very close to zero).

→ The occurrences are rare.

→ The mean and variance of the Poisson distribution with parameter λ are defined as

mean $\mu = E(X) = \lambda = np$ & variance $V(X) = \sigma^2 = \lambda$.

Example of Method-2: Poisson Distribution

C	1	For Poisson variant X, if $P(X = 3) = P(X = 4)$, then find $P(X = 0)$. **Answer: $P(X = 0) = e^{-4}$**
C	2	The number of page requests that arrive at a Web server is a Poisson random variable. Its probability distribution is as follows: Number of x requests/sec. | 0 | 1 | 2 | 3 | 4 | 5 | 6 Probability f(x) | 0.368 | 0.368 | 0.184 | 0.061 | 0.015 | 0.003 | 0.001 Find the mean and variance of this probability distribution. **Answer: 1, 1**
C	3	In sampling a large number of parts manufactured by a machine, the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples, how many would be expected to contain exactly two defective parts? **Answer: 271**
C	4	Potholes on a highway can be serious problems. The past experience suggests that there are, on an average, 2 potholes per mile after a certain amount of usage. It is assumed that Poisson process applies to random variable “no. of potholes”. What is the probability that no more than four potholes will occur in a given section of 5 miles? **Answer: $P(X \leq 4) = 0.0293$**
C	5	In a company, there are 250 workers. The probability of a worker remains absent on any one day is 0.02. Find the probability that on a day, seven workers are absent. **Answer: 0.1044**
C	6	Average number of accidents on any day on a national highway is 1.8. Determine the probability that the number of accidents are at least 1 and at most 1. **Answer: 0.8347, 0.4628**

H	7	The probability that a person catch corona virus is 0.001. Find the probability that out of 3000 persons exactly 3, more than 2 persons will catch the virus. **Answer**: $\mathbf{0.2240}$, $\mathbf{0.5768}$
C	8	Suppose 1% of the items made by machine are defective. In a sample of 100 items find the probability that the sample contains all good, 1 defective and at least 3 defectives. **Answer**: $\mathbf{P(X = 0) = 0.3679}$, $\mathbf{P(X = 1) = 0.3679}$ $\mathbf{P(X \geq 3) = 0.0803}$

Method 3 ⇝ Exponential Distribution

Exponential Distribution

→ A random variable X is said to have an exponential distribution with parameter $\theta > 0$, if its probability density function is given by

$$\mathbf{f(X = x) = \begin{cases} \theta \cdot e^{-\theta x} & ; \quad x \geq 0 \\ 0 & ; \text{ otherwise} \end{cases}}$$

Where, $\theta = \dfrac{1}{\text{mean}}$ and variance $= \dfrac{1}{\theta^2}$.

→ In exponential distribution we can find the probability as given below:

(1) $P(X \leq x) = 1 - e^{-\theta x}$

(2) $P(X \geq x) = e^{-\theta x}$

(3) $P(a \leq X \leq b) = e^{-a\theta} - e^{-b\theta}$

Properties of Exponential Distribution

→ Exponential distribution is used to describe lifespan and waiting times.

→ Exponential distribution can be used to describe (waiting) times between Poisson events.

Example of Method-3: Exponential Distribution

C	1	The lifetime T of an alkaline battery is exponentially distributed with $\theta = 0.05$ per hour. What are mean and standard deviation of batteries lifetime? **Answer: 20, 20**
C	2	The time between breakdowns of a particular machine follows an exponential distribution with a mean of 17 days. Calculate the probability that a machine breakdown in 15-day period. **Answer: 0.5862**

C	3	The lifetime T of an alkaline battery is exponentially distributed with $\theta = 0.05$ per hour. (a) What are the probabilities for battery to last between 10 and 15 hours? (b) What are the probabilities for the battery to last more than 20 hours? **Answer: (a) 0.1342, (b) 0.3679**
C	4	In a large corporate computer network, user log-on to the system can be modeled as a Poisson process with a mean of 25 log-on per hours. (a) What is the probability that there are no log-on in an interval of six min.? (b) What is the probability that time until next log-on is between 2 & 3 min.? **Answer: (a) 0.0821, (b) 0.1481**
C	5	A random variable has an exponential distribution with probability density function given by $f(x) = 3e^{-3x}\,;x > 0$ & $f(x) = 0\,;x \leq 0$. What is the probability that X is not less than 4? **Answer: e^{-12}**
T	6	The income tax of a man is exponentially distributed with $f(x) = \frac{1}{3}e^{-\left(\frac{x}{3}\right)}\,;x > 0.$ What is the probability that his income will exceed Rs. 17000? Assume that the income tax is levied at the rate of 15% on the income above Rs. 15000. **Answer: e^{-100}**
C	7	A random variable has an exponential distribution with probability density $f(x) = \begin{cases} \frac{1}{6} & ;\ -3 < x < 3 \\ 0 & ;\ \text{otherwise} \end{cases}$ Find (a) $P(X < 2)$, (b) $P(\lvert X \rvert \leq 2)$ **Answer: (a) $\frac{5}{6}$, (b) $\frac{2}{3}$**

C	8	A random variable has an exponential distribution with probability density $f(x) = \begin{cases} \frac{1}{5} e^{-\frac{x}{5}} & ; \ x > 0 \\ 0 & ; \ \text{otherwise} \end{cases}$ Find (a) $P(X > 5)$, (b) $P(3 \le X \le 6)$, (c) Mean, (d) Variance **Answer**: **(a) 0.3679, (b) 0.2476, (c) 5, (d) 25**

Method 4 ⇝ Normal Distribution

Normal Distribution

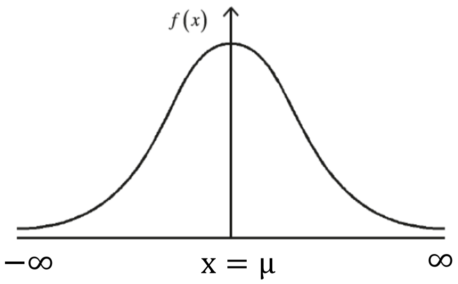
→ A continuous random variable X is said to follows a normal distribution if its probability density function is given by

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

Where, $-\infty < x < \infty$ & $\sigma > 0$

μ = mean of the distribution

σ = standard deviation of the distribution



→ Here, μ (mean) & σ^2 (variance) are known as parameters of the distribution.

→ If X is a normal random variable with mean μ and standard deviation σ, then we can find the random variable defined as below

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

→ If $\boldsymbol{\mu = 0}$ and $\boldsymbol{\sigma = 1}$, then Z is known as the standard (standardized) normal variable and probability destiny function is given by

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

→ The distribution of any normal variate X can always be transformed into the distribution of the standard normal variate Z.

$$P(x_1 \le X \le x_2) = P\left(\frac{x_1-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}\right) = P(z_1 \le Z \le z_2).$$

→ This probability is equal to area under the standard normal curve between the coordinates at $Z = z_1$ and $Z = z_2$.

Methods for Finding the Probability(area) from Z Table

→ Normal curve is a curve of normal density function and it can be seen in figure.

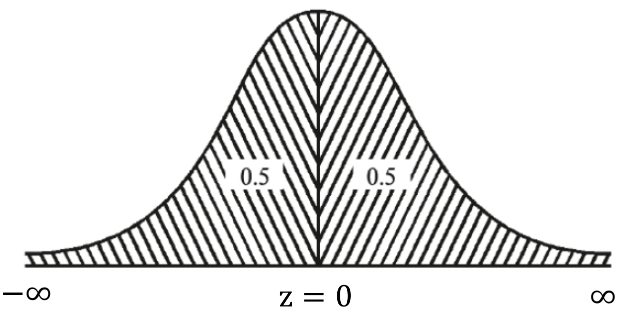
→ The probability(area) of shaded region between the curve and x – axis is equal to 1.

→ The probability(area) to the right side of $X = \mu$ is 0.5 and is denoted by

$P(X \geq \mu) = 0.5$

→ The probability(area) to the left side of $X = \mu$ is also 0.5 and is denoted by

$P(X \leq \mu) = 0.5$



→ Let us discuss some cases which is useful to solve examples.

Important Cases

(1) **Area between $Z = 0$ to $Z = z_1 = P(0 \leq Z \leq z_1)$**

→ Find z_1 by using first row and column of Z table to get the answer of $P(0 \leq Z \leq z_1)$.

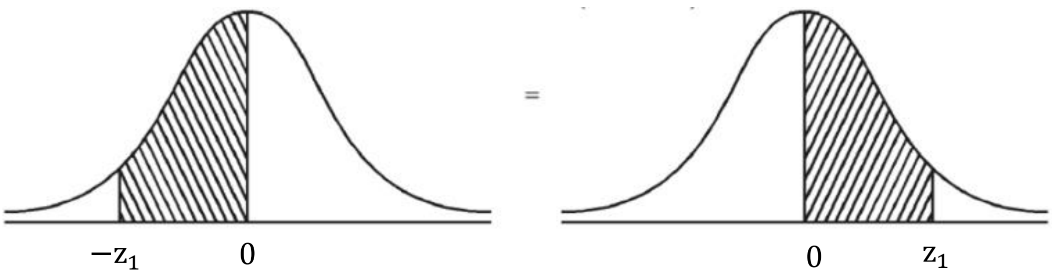
→ For Example:

- Suppose $z_1 = 1.43$. Use below steps to find value at $z = 1.43$.
- Step 1: Split 1.43 as $1.4 + 0.03$.
- Step 2: Find 1.4 in first column and 0.03 in first row.
- Step 3: Intersection of that row and column is answer of $P(0 \leq Z \leq 1.43)$.
- Here, $P(0 \leq Z \leq 1.43) = 0.4236$

→ Note:

- Area between $Z = -z_1$ to $Z = 0 = P(-z_1 \leq Z \leq 0) = P(0 \leq Z \leq z_1)$ as normal distribution is symmetric about x – axis.

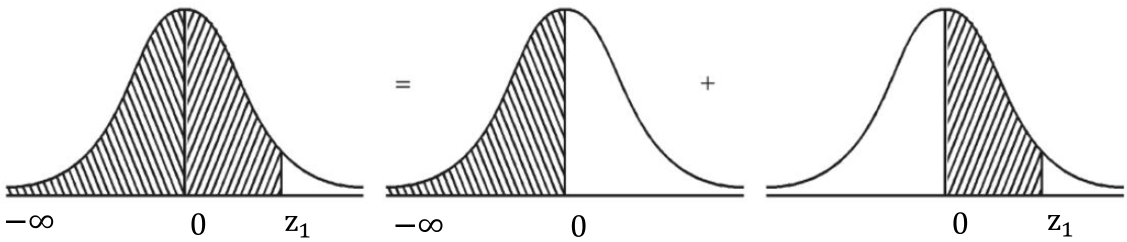
  i.e., $P(0 \leq Z \leq 1.43) = P(-1.43 \leq Z \leq 0) = 0.4236$



(2) **Area between $Z = -\infty$ to $Z = z_1 = P(Z \le z_1)$**

→ $P(Z \le z_1) = P(-\infty \le Z \le 0) + P(0 \le Z \le z_1)$

→ Since, $P(-\infty \le Z \le 0) = 0.5$ and $P(0 \le Z \le z_1)$ can be found by using case – I.



→ Note:

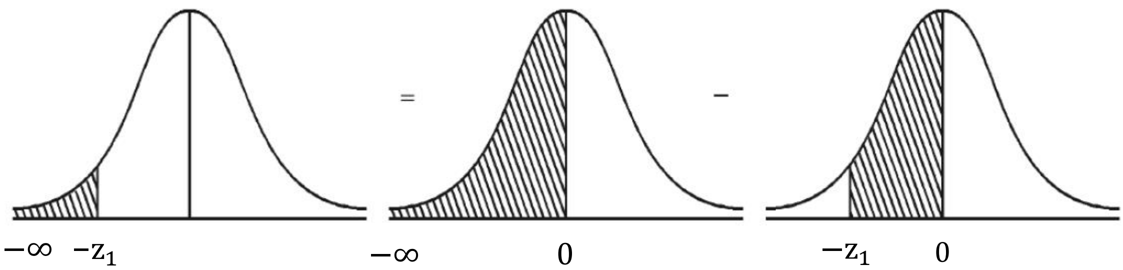
- Area between $Z = -z_1$ to $Z = \infty = P(Z \ge -z_1) = P(Z \le z_1)$ as normal distribution is symmetric about x – axis.

  i.e., $P(Z \ge -1.43) = P(Z \le 1.43) = 0.9236$

(3) **Area between $Z = -\infty$ to $Z = -z_1 = P(Z \le -z_1)$**

→ $P(Z \le -z_1) = P(-\infty \le Z \le 0) - P(-z_1 \le Z \le 0)$

→ Since, $P(-\infty \le Z \le 0) = 0.5$ and $P(-z_1 \le Z \le 0)$ can be found by using case – I.



→ Note:

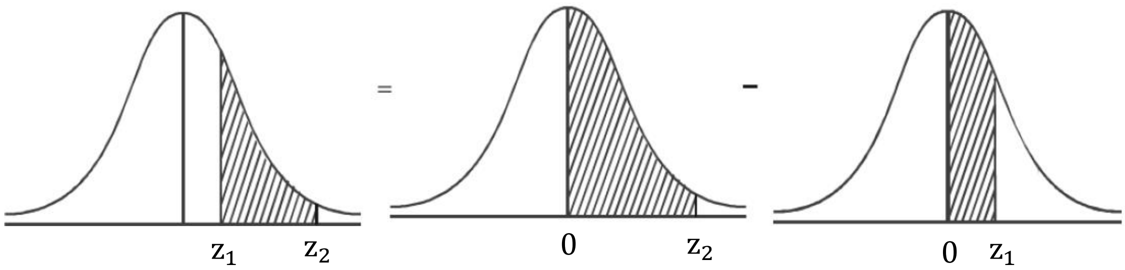
- Area between $Z = z_1$ to $Z = \infty = P(Z \ge z_1) = P(Z \le -z_1)$ as normal distribution is symmetric about x – axis.

  i.e., $P(Z \ge 1.43) = P(Z \le -1.43) = 0.0764$

(4) **Area between $\mathbf{Z = z_1}$ to $\mathbf{Z = z_2 = P(z_1 \le Z \le z_2)}$**

→ $P(z_1 \le Z \le z_2) = P(0 \le Z \le z_2) - P(0 \le Z \le z_1)$

→ Since, $P(0 \le Z \le z_1)$ and $P(0 \le Z \le z_2)$ can be found by using case – I.



→ Note:

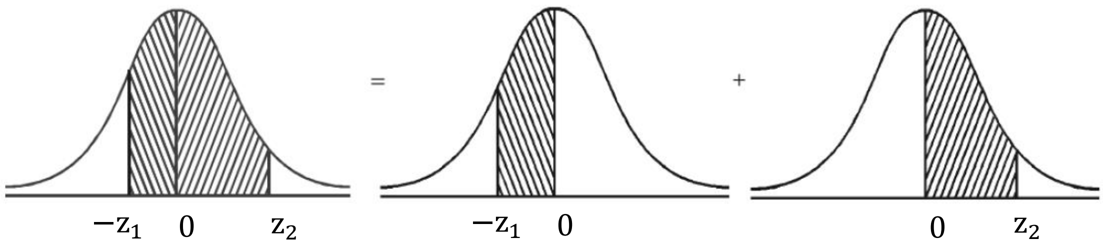
- Area between $Z = -z_2$ to $Z = -z_1 = P(-z_2 \le Z \le -z_1) = P(z_1 \le Z \le z_2)$ as normal distribution is symmetric about x – axis.

  i.e., $P(-2 \le Z \le -1.43) = P(1.43 \le Z \le 2) = 0.0436$

(5) **Area between $\mathbf{Z = -z_1}$ to $\mathbf{Z = z_2 = P(-z_1 \le Z \le z_2)}$**

→ $P(-z_1 \le Z \le z_2) = P(-z_1 \le Z \le 0) + P(0 \le Z \le z_2)$

→ Since, $P(-z_1 \le Z \le 0)$ and $P(0 \le Z \le z_2)$ can be found by using case – I.



→ Note:

- Area between $Z = -z_1$ to $Z = z_2 = P(-z_1 \le Z \le z_2)$ as normal distribution is symmetric about x – axis.

  i.e., $P(-1 \le Z \le 2) = P(-1 \le Z \le 0) + P(0 \le Z \le 2)$

  $= 0.3413 + 0.4772$

  $= 0.8185$

Properties of Normal Distribution

→ The curve is bell shaped and symmetrical about the line $x = \mu$.

→ Mean, median and mode of the distribution coincide. i.e. the normal distribution is symmetrical.

→ As x increases numerically, f(x) decreases rapidly, the maximum probability occurring at the point $x = \mu$.

→ x – axis is an asymptote to the curve.

Importance

→ Most of the distributions occurring in practice, e.g., Binomial, Poisson, etc. can be approximated by normal distribution.

→ Even if a variable is not normally distributed, it can sometimes be brought to normal form by simple transformation of variable.

→ The entire theory of small sample tests is based on the fundamental assumption that the parent populations from which the samples have been drawn follows normal distribution.

Example of Method-4: Normal Distribution

C	1	If X is normally distributed and the mean of X is 12 and the SD is 4. Find out the probability of the following. (a) $X \geq 20$, (b) $X \leq 20$, (c) $0 \leq X \leq 12$ **Answer: (a) $\mathbf{0.0228}$, (b) $\mathbf{0.9772}$, (c) $\mathbf{0.4987}$**
C	2	For a random variable having the normal distribution with $\mu = 18.2$ and $\sigma = 1.25$, find the probabilities that it will take on a value less than 16.5, between 16.5 and 18.8. $[P(z = 1.36) = 0.4131,\quad P(z = 0.48) = 0.1844]$ **Answer: $\mathbf{0.0869}$, $\mathbf{0.5974}$**

C	3	The compressive strength of the sample of cement can be modelled by normal distribution with mean $6000\ \text{kg/cm}^2$ and standard deviation $100\ \text{kg/cm}^2$. (a) What is the probability that a sample strength is less than $6250\ \text{kg/cm}^2$? (b) What is probability if sample strength is between 5800 and $5900\ \text{kg/cm}^2$? (c) What strength is exceeded by 95% of the samples? $[P(z = 2.5) = 0.4938,\quad P(z = 1) = 0.3413,\quad P(z = 2) = 0.4772,$ $P(z = 1.65) = 0.4505]$ **Answer: (a) 0.9938, (b) 0.136, (c) 6165**
C	4	A sample of 100 dry battery cell tested & found that average life is 12 hours & standard deviation 3 hours. Assuming data to be normally distributed what % of battery cells are expected to have life (a) more than 15 hrs.? (b) less than 6 hrs.? (c) between 10 & 14 hrs.? $[P(z = 1) = 0.3413,\quad P(z = 2) = 0.4772,\quad P(z = 0.67) = 0.2486]$ **Answer: (a) 15.87%, (b) 2.27%, (c) 49.72%**
C	5	In a photographic process, the developing time of prints may be looked upon as a random variable having normal distribution with mean of 16.28 seconds and standard deviation of 0.12 seconds. Find the probability that (a) It will take anywhere from 16.00 to 16.50 sec to develop one of the prints (b) At least 16.20 sec to develop one of the prints (c) At most 16.35 sec to develop one of the prints. $[P(z = 1.83) = 0.4664,\quad P(z = 2.33) = 0.4901]$ $[P(z = 0.67) = 0.2486,\quad P(z = 0.58) = 0.2190]$ **Answer: (a) 0.9565, (b)0.7486, (c) 0.7190**

C	6	Assuming that the diameters of 1000 brass plugs taken consecutively from a machine form a normal distribution with mean 0.7515 cm and standard deviation 0.002 cm. Find the number of plugs likely to be rejected if the approved diameter is 0.752 ± 0.004 cm. $[P(z = 1.75) = 0.4599, \quad P(z = 2.25) = 0.4878]$ **Answer: 52**
C	7	In an examination, minimum 40 marks for passing and 75 marks for distinction are required. In this examination 45% students passed and 9% obtained distinction. Find average marks and standard deviation of this distribution of marks. $[P(z = 0.125) = 0.05, \quad P(z = 1.34) = 0.4099]$ **Answer: 36.40, 28.81**

*** * * * * End of the Unit * * * * ***