PDFZilla - Unregistered

PDFZilla - Unregistered

Module-3
Part - 2

Continuous Probability Distribution

For every continuous random variable X, the real number $f(x)$ is said to be continuous probability function or probability density function (p.d.f) if the following conditions are satisfied:

(i) $f(x) \geq 0$ and

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

Note: The probability of x lies in the interval (a, b) is defined as

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

Note: If X is a continuous random variable with p.d.f $f(x)$ then the function $f(x)$ is defined by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

Mean and variance of continuous probability distribution

If X is a continuous random variable with probability density function $f(x)$ where $-\infty < x < \infty$ then the mean, variance and standard deviation of X are given by

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

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

Standard deviation $\sigma = \sqrt{V}$

Example **Find the constant C such that**

$$f(x)=\begin{cases} Cx^2, & 0\le x\le 3 \\ 0, & \text{otherwise} \end{cases}$$

is a probability density function. Also compute $P(1 < X < 2)$.

$f(x)$ is (p.d.f) if

(i) $f(x)\ge 0$ and (ii) $\int_{-\infty}^{\infty} f(x)\,dx=1$

From (ii), $\int_{-\infty}^{\infty} f(x)\,dx=1$

$\Rightarrow \int_{0}^{3} Cx^2\,dx=1$

$$\Rightarrow \qquad \int_0^3 Cx^2\,dx = 1$$

$$\Rightarrow \qquad C\left[\frac{x^3}{3}\right]_0^3 = 1$$

$$\Rightarrow \qquad C\left[\frac{(3)^3 - 0}{3}\right] = 1$$

$$\Rightarrow \qquad C[9] = 1 \qquad \Rightarrow \quad C = \frac{1}{9}$$

We have, $$P(a \le x \le b) = \int_a^b f(x)\,dx$$

$$\Rightarrow \qquad P(1 < x < 2) = \int_1^2 f(x)\,dx$$

$\Rightarrow$ $$P(1 < x < 2) = \int_1^2 f(x)\,dx$$

$$= \int_1^2 \frac{1}{9} x^2\,dx$$

$$= \frac{1}{9}\left[\frac{x^3}{3}\right]_1^2$$

$$= \frac{1}{9}\left[\frac{(2)^3 - (1)^3}{3}\right] = \frac{7}{27}$$

Example **A random variable X has the density function:**

$$f(x)=\begin{cases} kx^2, & -3\le x\le 3 \\ 0, & \text{otherwise} \end{cases}$$

Find k and also find $P(X\le 2)$ and $P(X>1)$.

$f(x)$ is (p.d.f) if

(i) $f(x)\ge 0$ and (ii) $\int_{-\infty}^{\infty} f(x)dx=1$

From (ii), $\int_{-\infty}^{\infty} f(x)dx=1$

$\Rightarrow$ $\int_{-3}^{3} kx^2\,dx=1$

$$\Rightarrow \quad \int_{-3}^{3} kx^2\,dx = 1$$

$$\Rightarrow \quad k\left[\frac{x^3}{3}\right]_{-3}^{3} = 1$$

$$\Rightarrow \quad k\left[\frac{(3)^3 - (-3)^3}{3}\right] = 1$$

$$\Rightarrow \quad k[9+9] = 1$$

$$\Rightarrow \quad k = \frac{1}{18}$$

We have, $$P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

$$\Rightarrow \quad P(X \le 2) = \int_{-\infty}^{2} f(x)\,dx$$

$$= \int_{-3}^{2} \frac{1}{18} x^2\,dx$$

$$= \frac{1}{18}\left[\frac{x^3}{3}\right]_{-3}^{2}$$

$$= \frac{1}{18}\left[\frac{(2)^3 - (-3)^3}{3}\right]$$

$$= \frac{35}{54}$$

We have, $$P(X > x) = \int_x^{\infty} f(x)\,dx$$

$$\Rightarrow \quad P(X > 1) = \int_1^{\infty} f(x)\,dx$$

$$= \int_1^{3} \frac{1}{18} x^2\,dx$$

$$= \frac{1}{18}\left[\frac{x^3}{3}\right]_1^3$$

$$= \frac{1}{18}\left[\frac{(3)^3 - (1)^3}{3}\right] = \frac{13}{27}$$

Example 3.2.13: A random variable X has the density function:

$$f(x)=\begin{cases} k(x+1), & -1<x<1 \\ 0, & \text{otherwise} \end{cases}$$

Find (i) k, (ii) mean and variance of X.

$f(x)$ is (p.d.f) if

(i) $f(x)\geq 0$ and (ii) $\int_{-\infty}^{\infty} f(x)dx=1$

From (ii), $\int_{-\infty}^{\infty} f(x)dx=1$

$\Rightarrow$ $\int_{-1}^{1} k(x+1)dx=1$

$$\Rightarrow \quad \int_{-1}^{1} k(x+1)\,dx = 1$$

$$\Rightarrow \quad k\left[\frac{x^2}{2} + x\right]_{-1}^{1} = 1$$

$$\Rightarrow \quad k\left[\left(\frac{(1)^2}{2} + 1\right) - \left(\frac{(-1)^2}{2} - 1\right)\right] = 1$$

$$\Rightarrow \quad 2k = 1$$

$$\Rightarrow \quad k = \frac{1}{2}$$

We have, Mean $\mu = \int_{-1}^{1} x f(x) dx$

$$= \int_{-1}^{1} x \left[\frac{1}{2}(x+1) \right] dx$$

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

$$= \frac{1}{2} \left[\frac{x^3}{3} + \frac{x^2}{2} \right]_{-1}^{1}$$

$$= \frac{1}{2} \left[\left(\frac{1}{3} + \frac{1}{2} \right) - \left(\frac{-1}{3} + \frac{1}{2} \right) \right] \quad = \frac{1}{3}$$

We have, Variance $V = \int_{-1}^{1} (x-\mu)^2 f(x)\,dx$

$$= \int_{-1}^{1} \left(x - \frac{1}{3}\right)^2 \left[\frac{1}{2}(x+1)\right] dx$$

$$= \frac{1}{2}\int_{-1}^{1} \left(x^2 - \frac{2x}{3} + \frac{1}{9}\right)(x+1)\,dx$$

$$= \frac{1}{2}\int_{-1}^{1} \left(x^3 + \frac{x^2}{3} - \frac{5x}{9} + \frac{1}{9}\right) dx$$

$$= \frac{1}{2}\left[\frac{x^4}{4} + \frac{x^3}{9} - \frac{5x^2}{18} + \frac{1}{9}x\right]_{-1}^{1}$$

$$=\frac{1}{2}\left[\left(\frac{1}{4}+\frac{1}{9}-\frac{5}{18}+\frac{1}{9}\right)-\left(\frac{1}{4}-\frac{1}{9}-\frac{5}{18}-\frac{1}{9}\right)\right] \quad =\frac{2}{9}$$

Example : The function $f(x)$ is defined as

$$f(x)=\begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

Is $f(x)$ a probability density function? If so, determine the probability that the variate having this density will fall in the interval (1, 2). Also find the cumulative probability function F (2).

(i) Clearly, $f(x) \geq 0$ and

(ii) $$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} (0)dx + \int_{0}^{\infty} e^{-x}\, dx$$

$$= 0 + \int_0^{\infty} e^{-x}\, dx$$

$$= \left[\frac{e^{-x}}{-1} \right]_0^{\infty} = -\left[e^{-\infty} - e^{0} \right] = -(0-1) = 1$$

Hence the function $f(x)$ satisfies the requirements for a density function.

$\therefore$ $f(x)$ is a probability density function

We have, $P(a \le x \le b) = \int_a^b f(x)\,dx$

$\Rightarrow$ $P(1 < x < 2) = \int_1^2 f(x)\,dx$

$\Rightarrow$ $$P(1<x<2)=\int_1^2 f(x)dx$$

$$=\int_1^2 e^{-x}\,dx$$

$$=\left[\frac{e^{-x}}{-1}\right]_1^2$$

$$=-\left(e^{-2}-e^{-1}\right)=0.233$$

Also we have, $$F(x)=P(X\le x)=\int_{-\infty}^{x} f(x)dx$$

$\Rightarrow$ $$F(2)=P(X\le 2)=\int_{-\infty}^{2} f(x)dx$$

$$\Rightarrow \quad F(2) = P(X \le 2) = \int_{-\infty}^{2} f(x)\,dx$$

$$= \int_{-\infty}^{0} (0)\,dx + \int_{0}^{2} e^{-x}\,dx$$

$$= 0 + \int_{0}^{2} e^{-x}\,dx$$

$$= \left[\frac{e^{-x}}{-1}\right]_{0}^{2}$$

$$= -\left[e^{-2} - e^{0}\right] = 0.865$$

EXERCISE

1. If a function $f(x)$ defined by $f(x)=\begin{cases}\left(\dfrac{x+1}{8}\right), & 2<x<4\\ 0, & \text{otherwise}\end{cases}$

   (i) Prove that $f(x)$ is a p.d.f

   (ii) Find $P(X < 3.5)$ and $P(X \geq 3.5)$

2. Find the value of k such that $f(x)=\begin{cases}\dfrac{x}{6}+k, & 0\leq x\leq 3\\ 0, & \text{otherwise}\end{cases}$

   is a p.d.f. Also find $P(1 \leq X \leq 2)$.

3. Find the value of k such that $f(x)=\begin{cases}k\,x\,e^{-x}, & 0<x<1\\ 0, & \text{otherwise}\end{cases}$

   is a p.d.f. Also find its mean.

4. Find the value of k such that $f(x)=\begin{cases}k\,e^{-3x}, & x>0\\ 0, & \text{otherwise}\end{cases}$

   is a p.d.f. Also find $P(0.5 \leq X \leq 1)$ and $P(-2 \leq X \leq 1.5)$

Module-3
Part – 6
Normal distribution

Normal Distribution

The normal distribution is a continuous distribution, is the most important of all the distributions.

It can be derived from the binomial distribution in the limiting case when n is very large and p is close to 0.5.

Normal Distribution

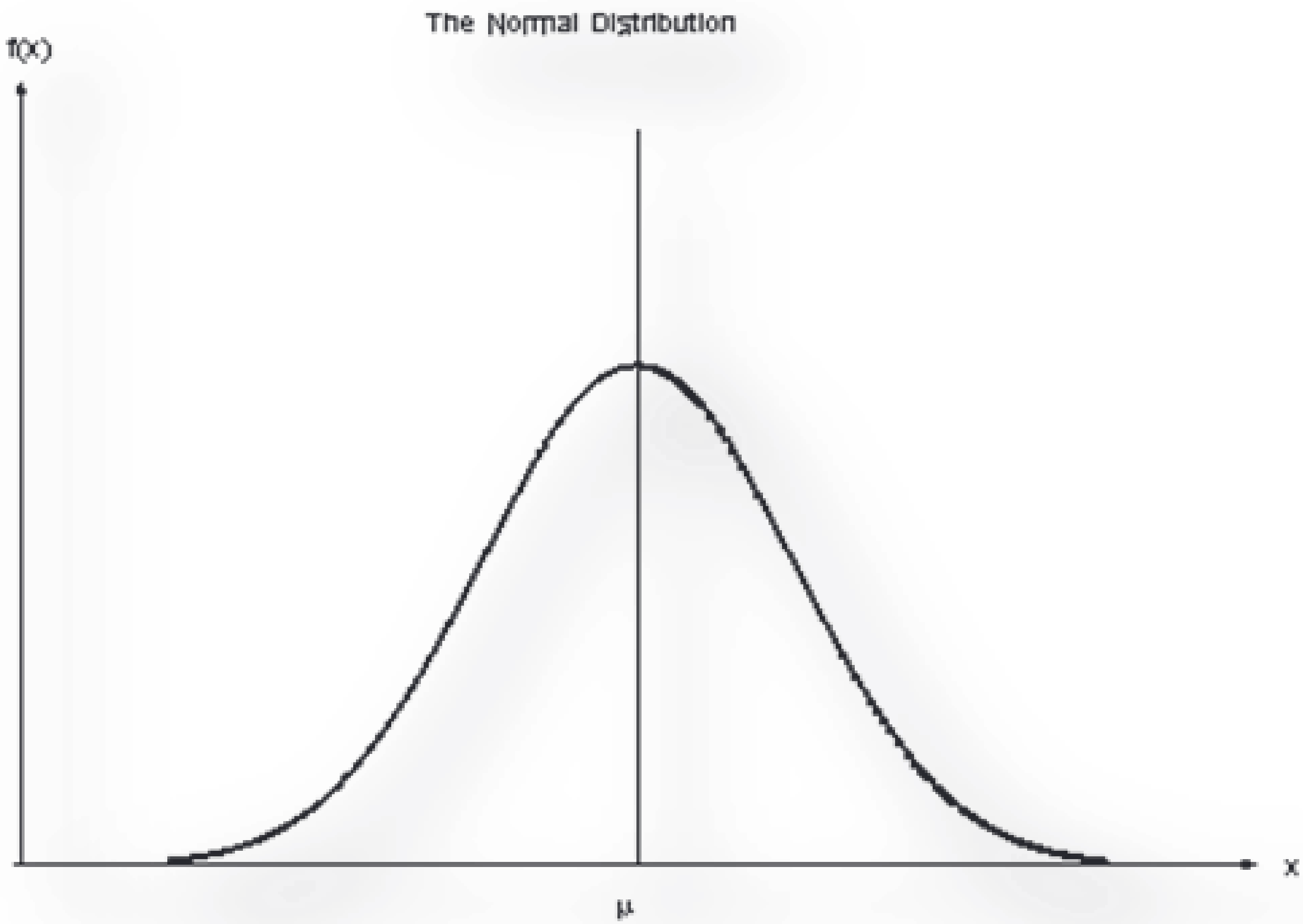
The normal distribution has two parameters (two numerical descriptive measures): the mean (μ) and the standard deviation (σ). If X is a quantity to be measured that has a normal distribution with mean (μ) and standard deviation (σ), we designate this by writing the following formula of the normal probability density function:

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

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

Here *x* is called normal variate and *f* (*x*) is called probability density function of the normal distribution.

Graph of the Normal distribution



- The graph of the normal distribution is called the **normal curve.**
- It is bell-shaped and symmetrical about the line

$$\boldsymbol{x = \mu}$$

- The line $x = \mu$ divides the total area under the curve which is equal to 1 into two equal parts.
- The area to the right as well as to the left of the line $x = \mu$ is 0.5.

Mean and variance of Normal distribution

The mean and variance of the normal distribution are given by

Mean $= \mu$

Variance $V = \sigma^2$

Standard deviation $= \sigma$

Standard form of the Normal distribution

If X is a normal random variable with mean μ and standard deviation σ then the random variable $z = \frac{x - \mu}{\sigma}$ has the normal distribution with mean 0 and standard deviation 1. The random variable z is called the standard normal random variable.

The standard normal distribution is a normal distribution of standardized values called ***z-scores.***
A z-score is measured in units of the standard deviation. The mean for the standard normal distribution is zero, and the standard deviation is one.

If X is a normally distributed random variable and $X \sim N(\mu, \sigma)$, then the z-score for a particular x is:

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

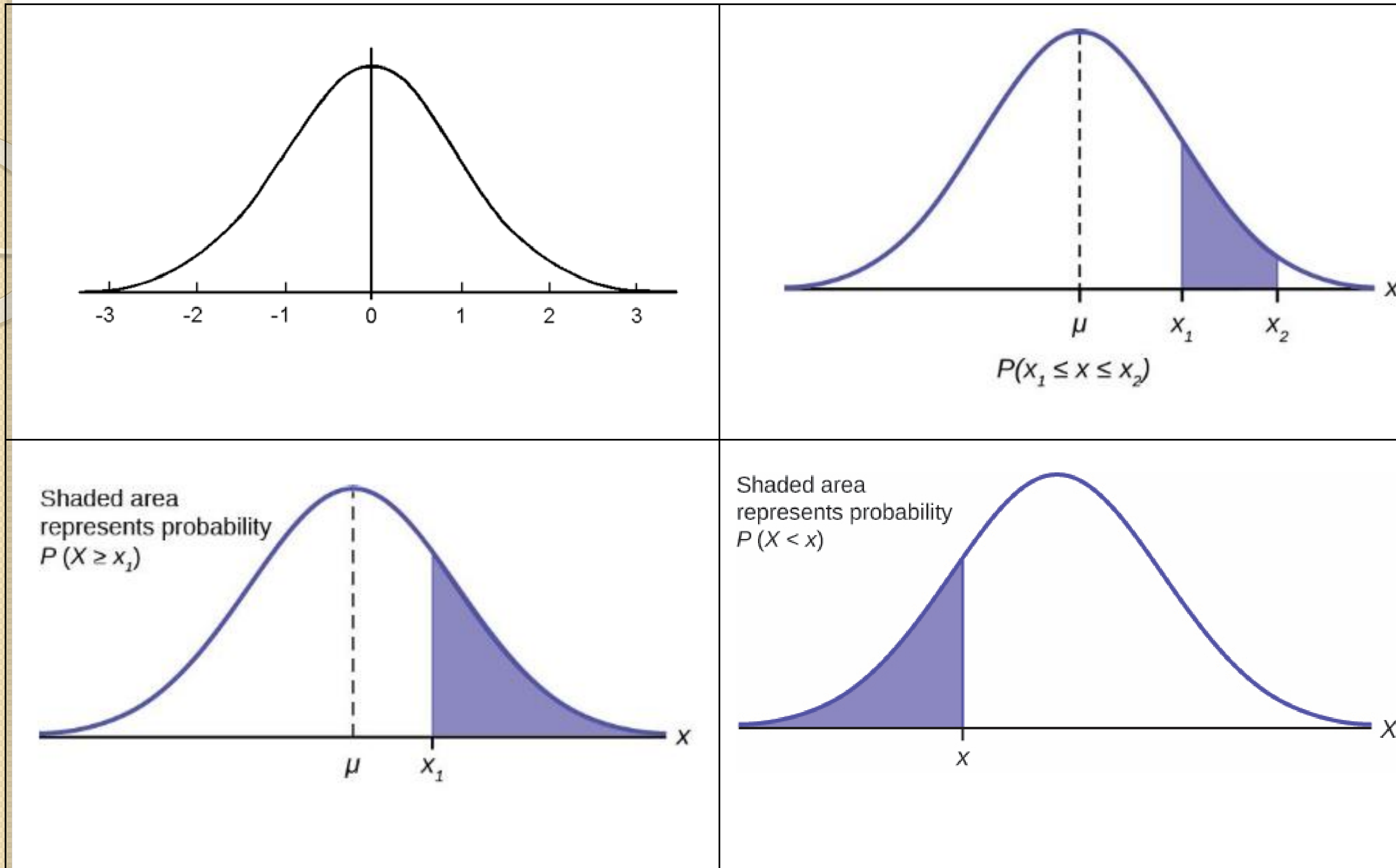
- The *z*-score tells you how many standard deviations the value *x* is above (to the right of) or below (to the left of) the mean, μ.
- Values of *x* that are larger than the mean have positive *z*-scores, and values of *x* that are smaller than the mean have negative *z*-scores.
- If *x* equals the mean, then x has a *z*-score of zero.

The probability density function for the normal distribution in standard form is given by

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

This helps us to compute areas under the normal probability curve by making use of standard tables.

The mathematical tool needed to find the area under a curve is integral calculus. The integral of the normal probability density function between the two points x1 and x2 is the area under the curve between these two points and is the probability between these two points.



Normal probability distribution curves

Note:

(1) If $f(z)$ is the probability density function for the normal distribution, then $P(z_1 < z < z_2) = \int_{z_1}^{z_2} f(z)\,dz = f(z_2) - f(z_1)$ where $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$

The function $f(z)$ defined above is called the distribution function for the normal distribution.

(2) The probabilities $P(z_1 \le z \le z_2)$, $P(z_1 \le z < z_2)$, $P(z_1 < z \le z_2)$ and $P(z_1 < z < z_2)$ are all regarded to be the same.

(3) $P(z \le z_1) = \int_{-\infty}^{z_1} f(z)\,dz$ and $P(z \ge z_1) = \int_{z_1}^{\infty} f(z)\,dz$

(4) $P(-\infty \le z \le 0) = P(0 \le z \le \infty) = 0.5$

(5) $P(z \ge z_1) = P(0 \le z \le \infty) - P(0 \le z \le z_1)$

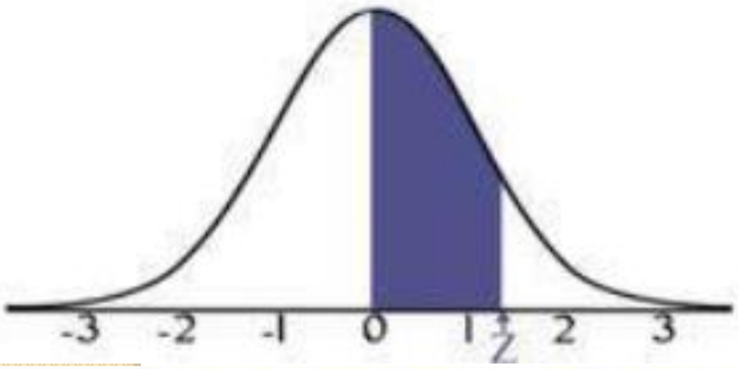
(6) $P(-z_1 \le z \le z_2) = P(-z_1 \le z \le 0) + P(0 \le z \le z_2)$

(7) $P(-z_1 < z < 0) = P(0 < z < z_1)$ (since the normal curve is symmetric)

(8) $P(-z_1 < z < z_1) = 2P(0 < z < z_1)$

(9) $F(-z_1) = 1 - F(z_1)$

(10) The following table called normal probability table gives the area under the standard normal curve from 0 to $z > 0$, for various values of z.



STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621

0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

- It is also important to note that because the normal distribution is symmetrical it does not matter if the z-score is positive or negative when calculating a probability.
- One standard deviation to the left (negative z-score) covers the same area as one standard deviation to the right (positive z-score).
- This fact is why the Standard Normal tables do not provide areas for the left side of the distribution.

Example If X is a normal variate with mean 30 and S.D. 5, find the probabilities that (i) $26 \le X \le 40$, (ii) $X \ge 45$ and (iii) $|X - 30| > 5$.

Given, $\mu = 30$ and S.D. $\sigma = 5$

We have, $$z = \frac{x-\mu}{\sigma}$$

$$\Rightarrow \qquad z = \frac{x-30}{5}$$

(i) When $x = 26 \Rightarrow z = \dfrac{26-30}{5} = -0.8$

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

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

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

$$= P(-0.8 \le z \le 0) + P(0 \le z \le 2)$$

$$= P(0 \le z \le 0.8) + P(0 \le z \le 2)$$

$$= 0.2881 + 0.4772 = 0.7653$$

(ii) When $x = 45 \Rightarrow z = \dfrac{45-30}{5} = 3$

$\therefore \quad P(x \ge 45) = P(z \ge 3)$

$$= P(0 \le z \le \infty) - P(0 \le z \le 3)$$

$$= 0.5 - 0.4986 = 0.0014$$

(iii) $P\left[|X-30|\le 5\right] = P(25 \le X \le 35)$

$$\text{When } x = 25 \Rightarrow z = \frac{25-30}{5} = -1$$

$$\text{When } x = 35 \Rightarrow z = \frac{35-30}{5} = 1$$

$$\therefore \quad P(25 \le X \le 35) = P(-1 \le z \le 1)$$

$$= 2P(0 \le z \le 1)$$

$$= 2(0.3413) = 0.6826$$

$$\therefore \quad P\left[|X-30| > 5\right] = 1 - P\left[|X-30| \le 5\right]$$

$$= 1 - 0.6826 = 0.3174$$

Example In a test on 2000 electric bulbs, it was found that the life of a particular make was normally distributed with an average life of 2040 hours and S.D. of 60 hours. Estimate the number of bulbs likely to burn for (i) more than 2150 hours, (ii) less than 1950 hours, (iii) more than 1920 hours and less than 2160 hours.

Given, $\mu = 2040$ hours and S.D. $\sigma = 60$ hours

We have, $$z = \frac{x-\mu}{\sigma}$$

$$\Rightarrow \qquad z = \frac{x-2040}{60}$$

(i) When $x = 2150 \Rightarrow z = \frac{2150-2040}{60} = 1.83$

$$\therefore \qquad P(x > 2150) = P(z > 1.83)$$

$$\therefore \qquad P(x>2150)=P(z>1.83)$$

$$=P(0<z<\infty)-P(0<z<1.83)$$

$$=0.5-0.4664=0.0336$$

$\therefore$ Number of bulbs expected to burn for more than 2150 hours is

$$=0.0336\times 2000\approx 67$$

(ii) When $x=1950 \Rightarrow z=\dfrac{1950-2040}{60}=-1.5$

$$\therefore \qquad P(x<1950)=P(z<-1.5)$$

$$=P(z>1.5)$$

$$=P(0<z<\infty)-P(0<z<1.5)$$

(ii) When $x = 1950 \Rightarrow z = \dfrac{1950 - 2040}{60} = -1.5$

$$\therefore \quad P(x < 1950) = P(z < -1.5)$$

$$= P(z > 1.5)$$

$$= P(0 < z < \infty) - P(0 < z < 1.5)$$

$$= 0.5 - 0.4332 = 0.0668$$

$\therefore$ Number of bulbs expected to burn for more than 2150 hours is

$$= 0.0668 \times 2000 \approx 134$$

(iii) When $x = 1920 \Rightarrow z = \dfrac{1920 - 2040}{60} = -2$

When $x = 2160 \Rightarrow z = \dfrac{2160 - 2040}{60} = 2$

$$\therefore \quad P(1920 < x < 2160) = P(-2 < z < 2)$$

$$= 2P(0 < z < 2)$$

$$= 2(0.4772) = 0.9544$$

$\therefore$ Number of bulbs expected to burn for more than 1920 hours and less than 2160 hours is

$$= 0.9544 \times 2000 \approx 1909$$

Example A sample of 100 dry battery cells tested to find the length of life produced the following results:

$$\bar{x} = 12 \text{ hours and S.D. } \sigma = 3 \text{ hours}$$

Assuming the data to be normally distributed, what percentage of battery cells are expected to have life (i) more than 15 hours, (ii) less than 6 hours and (iii) between 10 and 14 hours?

Given, $\mu = 12$ hours and S.D. $\sigma = 3$ hours

We have, $$z = \frac{x-\mu}{\sigma}$$

$$\Rightarrow \quad z = \frac{x-12}{3}$$

(i) When $x = 15 \Rightarrow z = \dfrac{15-12}{3} = 1$

$$\therefore \quad P(x > 15) = P(z > 1)$$

$$\therefore \quad P(x > 15) = P(z > 1)$$

$$= P(0 < z < \infty) - P(0 < z < 1)$$

$$= 0.5 - 0.3413 = 0.1587 = 15.87\%$$

(ii) When $x = 6 \Rightarrow z = \dfrac{6-12}{3} = -2$

$$\therefore \quad P(x < 6) = P(z < -2)$$

$$= P(z > 2)$$

$$= P(0 < z < \infty) - P(0 < z < 2)$$

$$= 0.5 - 0.4772 = 0.0228 = 2.28\%$$

(iii) When $x = 10 \Rightarrow z = \dfrac{10-12}{3} = \dfrac{-2}{3} = -0.67$

When $x = 14 \Rightarrow z = \dfrac{14-12}{3} = \dfrac{2}{3} = 0.67$

$$\therefore \quad P(10 < x < 14) = P(-0.67 < z < 0.67)$$

$$= 2P(0 < z < 0.67)$$

$$= 2(0.2486) = 0.4972 = 49.72\%$$

Example **The marks of 1000 students in an examination follow a normal distribution with mean 70 and standard deviation 5. Find the number of students whose marks will be (i) less than 65, (ii) more than 75, (iii) between 65 and 75.**

Given, $\mu = 70$ and S.D. $\sigma = 5$

We have, $$z = \frac{x-\mu}{\sigma}$$

$$\Rightarrow \qquad z = \frac{x-70}{5}$$

(i) When $x = 65 \Rightarrow z = \frac{65-70}{5} = -1$

$$\therefore \qquad P(x < 65) = P(z < -1)$$

$$\therefore \qquad P(x<65)=P(z<-1)$$

$$=P(z>1)$$

$$=P(0<z<\infty)-P(0<z<1)$$

$$=0.5-0.3413=0.1587$$

$\therefore$ Number of students scoring less than 65 marks is

$$=0.1587\times 1000\approx 159$$

(ii) When $x=75 \Rightarrow z=\dfrac{75-70}{5}=1$

$$\therefore \qquad P(x>75)=P(z>1)$$

$$\therefore \quad P(x>75)=P(z>1)$$

$$=P(0<z<\infty)-P(0<z<1)$$

$$=0.5-0.3413=0.1587$$

$\therefore$ Number of students scoring more than 75 marks is

$$=0.1587\times 1000\approx 159$$

(iii) When $x=65 \Rightarrow z=-1$

When $x=75 \Rightarrow z=1$

$$\therefore \quad P(65<x<75)=P(-1<z<1)$$

$$\therefore \quad P(65 < x < 75) = P(-1 < z < 1)$$

$$= 2P(0 < z < 1)$$

$$= 2(0.3413) = 0.6826$$

$\therefore$ Number of students scoring marks between 65 and 75 is

$$= 0.6826 \times 1000 \approx 683$$

Example In an examination taken by 500 candidates, the average and the standard deviation of marks obtained (normally distributed) are 40% and 10% respectively. Find approximately (i) how many will pass, if 50% is fixed as a minimum, (ii) how many have scored marks above 60% and (iii) what should be the minimum percentage if 350 candidates are to pass?

Given, $\mu = 40$ and S.D. $\sigma = 10$

We have, $$z = \frac{x-\mu}{\sigma}$$

$$\Rightarrow \qquad z = \frac{x-40}{10}$$

(i) When $x = 50 \Rightarrow z = \dfrac{50-40}{10} = 1$

$$\therefore \qquad P(x \geq 50) = P(z > 1)$$

$$\therefore \qquad P(x \geq 50) = P(z > 1)$$

$$= P(0 < z < \infty) - P(0 < z < 1)$$

$$= 0.5 - 0.3413 = 0.1587$$

$\therefore$ Number of students passing if 50% is fixed as minimum is

$$= 0.1587 \times 500 \approx 79$$

(ii) When $x = 60 \Rightarrow z = \dfrac{60-40}{10} = 2$

$$\therefore \qquad P(x > 60) = P(z > 2)$$

$$\therefore \qquad P(x>60)=P(z>2)$$

$$=P(0<z<\infty)-P(0<z<2)$$

$$=0.5-0.4772=0.0228$$

$\therefore$ Number of students scoring more than 60% is

$$=0.0228\times 500\approx 11$$

(iii) Let M% of marks is the minimum for passing if 350 candidates are to pass. Then, we should have M < 40, and

$$P(x\geq M)\times 500=350$$

$$\Rightarrow \qquad P(x\geq M)=\frac{350}{500}=0.7$$

We have, $z = \dfrac{x-40}{10}$

$\Rightarrow$ $x = 10z + 40$

$$\therefore P(x \geq M) = P(10z + 40 \geq M) = P\left(z \geq \frac{M-40}{10}\right)$$

$$= P(z \geq -z_1) \text{ where } -z_1 = \frac{M-40}{10}$$

$$= P(-z_1 \leq z \leq 0) + P(0 \leq z \leq \infty)$$

$$= P(0 \leq z \leq z_1) + 0.5$$

Since $P(x \geq M) = 0.7$

$$\Rightarrow P(0 \le z \le z_1) = 0.2$$

The normal probability table shows that $P(0 \le z \le z_1) = 0.2$ for $z_1 = 0.55$

$$\text{So, } -z_1 = \frac{M-40}{10} = -0.55 \Rightarrow M = 34.5 \approx 35$$

$\therefore$ 35% is fixed as minimum marks for passing in order to pass 350 candidates out of 500.

Example In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.

Let μ be the mean and σ be the S.D.

Given, 31% of the items are under 45

$\Rightarrow$ Area to the left of the ordinate $x = 45$ is 0.31

Given, 8% of the items are over 64

$\Rightarrow$ Area to the right of the ordinate $x = 64$ is 0.08

When $x = 45$, let $z = z_1$

$$P(z_1 < z < 0) = 0.5 - 0.31 = 0.19$$

From the normal probability table, the value of z corresponding to this area is 0.5.

$$\text{i.e., } z_1 = -0.5 \qquad [\because z_1 < 0]$$

When $x = 64$, let $z = z_2$

$$P(0 < z < z_2) = 0.5 - 0.08 = 0.42$$

From the normal probability table, the value of z corresponding to this area is 1.4.

$$\text{i.e., } z_2 = 1.4$$

Since, $$z = \frac{x - \mu}{\sigma}$$

$$\Rightarrow \quad -0.5 = \frac{45 - \mu}{\sigma} \text{ and } 1.4 = \frac{64 - \mu}{\sigma}$$

$$\Rightarrow \quad \mu - 0.5\sigma = 45 \text{ and } \mu + 1.4\sigma = 64$$

Solving these equations, we get

$$\mu = 50 \text{ and } \sigma = 10$$

Example In a normal distribution, 7% of the items are under 35 and 89% are under 60. Find the mean and standard deviation of the distribution.

Let μ be the mean and σ be the S.D.

Given, 7% of the items are under 35

$\Rightarrow$ Area to the left of the ordinate $x = 35$ is 0.07

Given, 89% of the items are under 60

$\Rightarrow$ Area to the left of the ordinate $x = 60$ is 0.89

When $x = 35$, let $z = z_1$

$$P(z_1 < z < 0) = 0.5 - 0.07 = 0.43$$

From the normal probability table, the value of z corresponding to this area is 1.48

$$\text{i.e., } z_1 = -1.48 \qquad [\because z_1 < 0]$$

When $x = 60$, let $z = z_2$

$$P(z_2 < z < 0) = 0.5 - 0.89 = -0.39$$

From the normal probability table, the value of z corresponding to this area is 1.23

$$\text{i.e., } z_2 = 1.23 \qquad [\because z_2 > 0]$$

Since, $$z = \frac{x-\mu}{\sigma}$$

$$\Rightarrow \quad -1.48 = \frac{35-\mu}{\sigma} \text{ and } 1.23 = \frac{60-\mu}{\sigma}$$

$$\Rightarrow \quad \mu - 1.48\sigma = 35 \text{ and } \mu + 1.23\sigma = 60$$

Solving these equations, we get

$$\mu = 48.65 \text{ and } \sigma = 9.22$$

Example Fit a normal distribution for the data

x	2	4	6	8	10
f	1	4	6	4	1

$$\sum f_i = 1+4+6+4+1=16$$

$$\sum x_i f_i = 2+16+36+32+10=96$$

$$\sum x_i^2 f_i = 4+64+216+256+100=640$$

$$\therefore \text{ Mean } \mu = \frac{\sum x_i f_i}{\sum f_i} = \frac{96}{16} = 6$$

$$\text{Variance } V = \frac{\sum x_i^2 f_i}{\sum f_i} - \mu^2 = \frac{640}{16} - 36 = 4$$

S.D $\sigma = \sqrt{V} = \sqrt{4} = 2$

The p.d.f. of normal distribution is

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

$$= \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-6}{2}\right)^2}$$

The equation of the normal curve that fits the data is

$$F(x_i) = \left(\sum f_i\right) f(x)$$

$$\Rightarrow \quad F(x_i) = (16)\left[\frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-6}{2}\right)^2}\right] = \frac{8}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-6}{2}\right)^2}$$

$$F(x_i) = (16)\left[\frac{1}{2\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-6}{2}\right)^2}\right] = \frac{8}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-6}{2}\right)^2}$$

$$F(2) = \frac{8}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{2-6}{2}\right)^2} = 0.97$$

$$F(4) = \frac{8}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{4-6}{2}\right)^2} = 3.9$$

$$F(6) = \frac{8}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{6-6}{2}\right)^2} = 6.1$$

$$F(8) = \frac{8}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{8-6}{2}\right)^2} = 3.9$$

$$F(10)=\frac{8}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{10-6}{2}\right)^2}=0.97$$

$\therefore$ The theoretical frequencies are

x	2	4	6	8	10
f	1	4	6	4	1

Module-3
Part – 4
Exponential Distribution

Exponential Distribution

The continuous probability distribution having the probability density function ***f*** (***x***) given by

$$f(x)=\begin{cases}\alpha\, e^{-\alpha x}, & 0\le x<\infty \\ 0, & \text{otherwise}\end{cases}$$

where $\alpha > 0$, is known as the exponential distribution.

Mean and variance of Exponential distribution

The mean and variance of the binomial distribution are given by

$$\text{Mean } \mu = \frac{1}{\alpha}$$

$$\text{Variance } V = \frac{1}{\alpha^2}$$

$$\text{Standard deviation } \sigma = \sqrt{V} = \frac{1}{\alpha}$$

Note: Since the function $f(x)$ is zero for $-\infty < x < 0$, the probabilities of the exponential distribution for various cases are as follows:

(i) $P(0 \le x < a) = \int_0^a f(x)\,dx, \quad a > 0$

(ii) $P(x \ge a) = 1 - P(x < a) = 1 - \int_0^a f(x)\,dx, \quad a > 0$

Example If x is an exponential variate with mean 5, evaluate the following:

(i) $P(0<x<1)$, **(ii)** $P(x>2)$,

(iii) $P(-\infty<x<10)$ **(iv)** $P(x\le 0 \text{ or } x\ge 1)$

The p.d.f of the exponential distribution is $f(x)=\begin{cases}\alpha e^{-\alpha x}, & 0\le x<\infty\\ 0, & \text{otherwise}\end{cases}$

Given, Mean = 5 i.e., Mean $\mu=\dfrac{1}{\alpha}=5$

$$\Rightarrow \qquad \alpha=\frac{1}{5}$$

(i) $P(0<x<1)=\int_0^1 f(x)\,dx=\int_0^1 \alpha e^{-\alpha x}\,dx$

$$=\int_0^1\left(\frac{1}{5}\right)e^{-\frac{1}{5}x}\,dx=\frac{1}{5}\left[\frac{e^{-\frac{1}{5}x}}{-\frac{1}{5}}\right]_0^1=-\left[e^{-0.2}-e^{0}\right]=0.1813$$

(ii) $$P(x>2)=1-P(x\le 2)=1-\int_0^2 f(x)\,dx$$

$$=1-\int_0^2 \alpha\, e^{-\alpha x}\,dx$$

$$=1-\int_0^2 \left(\frac{1}{5}\right)e^{-\frac{1}{5}x}\,dx$$

$$=1-\frac{1}{5}\left[\frac{e^{-\frac{1}{5}x}}{-\frac{1}{5}}\right]_0^2$$

$$=1+\left[e^{-2/5}-e^{0}\right]:=0.6703$$

(iii) $P(-\infty < x < 10) = P(-\infty < x < 0) + P(0 \le x < 10)$

$$= 0 + \int_0^{10} f(x)\,dx$$

$$= \int_0^{10} \alpha\, e^{-\alpha x}\, dx$$

$$= \int_0^{10} \left(\frac{1}{5}\right) e^{-\frac{1}{5}x}\, dx \quad = \frac{1}{5}\left[\frac{e^{-\frac{1}{5}x}}{-\frac{1}{5}}\right]_0^{10}$$

$$= -\left[e^{-2} - e^{0}\right]$$

$$= 0.8647$$

(iv) $$P(x \le 0 \text{ or } x \ge 1) = \int_1^\infty f(x)\,dx$$

$$= \int_1^\infty \alpha\, e^{-\alpha x}\, dx$$

$$= \int_1^\infty \left(\frac{1}{5}\right) e^{-\frac{1}{5}x}\, dx \quad = \frac{1}{5}\left[\frac{e^{-\frac{1}{5}x}}{-\frac{1}{5}}\right]_1^\infty$$

$$= -\left[0 - e^{-0.2}\right]$$

$$= 0.8187$$

Example The length of a telephone conversation has an exponential distribution with a mean of 3 minutes. Find the probability that a call (i) ends in less than 3 minutes, (ii) takes between 3 and 5 minutes.

The p.d.f of the exponential distribution is $f(x)=\begin{cases}\alpha e^{-\alpha x}, & 0\le x<\infty \\ 0, & \text{otherwise}\end{cases}$

Given, Mean = 3 i.e., Mean $\mu=\frac{1}{\alpha}=3$

$$\Rightarrow \qquad \alpha=\frac{1}{3}$$

(i) P (less than 3 minutes) $=P(x<3)$

$$=P(-\infty<x<3)$$

$$=P(-\infty<x<0)+P(0\le x<3)$$

$$= 0 + \int_0^3 f(x)\,dx$$

$$= \int_0^3 \alpha\, e^{-\alpha x}\, dx \quad = \int_0^3 \left(\frac{1}{3}\right) e^{-\frac{1}{3}x}\, dx$$

$$= \frac{1}{3}\left[\frac{e^{-\frac{1}{3}x}}{-\frac{1}{3}}\right]_0^3$$

$$= -\left[e^{-1} - e^{0}\right]$$

$$= 0.6321$$

(ii) P (between 3 and 5 minutes) $= P(3 < x < 5)$

$$= \int_3^5 f(x)\,dx \quad = \int_3^5 \alpha\, e^{-\alpha x}\, dx$$

$$= \int_3^5 \left(\frac{1}{3}\right) e^{-\frac{1}{3}x}\, dx$$

$$= \frac{1}{3}\left[\frac{e^{-\frac{1}{3}x}}{-\frac{1}{3}}\right]_3^5$$

$$= -\left[e^{-\frac{5}{3}} - e^{-1}\right] \quad = 0.179$$

Example **In a certain town, the duration of a shower is exponentially distributed with mean 5 minutes, what is the probability that a shower will last for (i) less than 10 minutes, (ii) 10 minutes or more?**

The p.d.f of the exponential distribution is $f(x) = \begin{cases} \alpha e^{-\alpha x}, & 0 \le x < \infty \\ 0, & \text{otherwise} \end{cases}$

Given, Mean = 5 i.e., Mean $\mu = \frac{1}{\alpha} = 5$

$$\Rightarrow \qquad \alpha = \frac{1}{5}$$

(i) P (less than 10 minutes) $= P(x < 10)$

$$= P(-\infty < x < 10) = P(-\infty < x < 0) + P(0 \le x < 10)$$

$$= 0 + \int_0^{10} f(x)\,dx$$

$$= \int_0^{10} \alpha\, e^{-\alpha x}\, dx \quad = \int_0^{10} \left(\frac{1}{5}\right) e^{-\frac{1}{5}x}\, dx$$

$$= \frac{1}{5}\left[\frac{e^{-\frac{1}{5}x}}{-\frac{1}{5}}\right]_0^{10} = -\left[e^{-2} - e^{0}\right] = 0.8647$$

(ii) P (10 minutes or more) $= P(x \geq 10)$

$$= 1 - P(x < 10)$$

$$= 1 - 0.8647 = 0.1353$$

Example 3.5.4: The sale per day in a shop is exponentially distributed with the average sales amounting to Rs.100 and net profit is 8%. Find the probability that the net profit exceeds Rs.30 on two consecutive days.

The p.d.f of the exponential distribution is $f(x)=\begin{cases}\alpha e^{-\alpha x}, & 0\le x<\infty \\ 0, & \text{otherwise}\end{cases}$

Given, Mean = 100 i.e., Mean $\mu=\frac{1}{\alpha}=100$

$$\Rightarrow \qquad \alpha=\frac{1}{100}$$

Let A be the amount for which profit is 8%

Given, profit = Rs.30

$\therefore$ 8% of A = 30

$$\Rightarrow \qquad A=\frac{30}{8\%}=\frac{30}{0.08}=375$$

Now, P (profit exceeding Rs.30) $= 1 - P$ (profit $\leq$ Rs.30)

$= 1 - P$ (sales $\leq$ Rs.375)

$$= 1 - \int_0^{375} f(x)\,dx$$

$$= 1 - \int_0^{375} \alpha\, e^{-\alpha x}\,dx$$

$$= 1 - \int_0^{375} (0.01) e^{-0.01x}\,dx$$

$$=1-(0.01)\left[\frac{e^{-0.01x}}{-0.01}\right]_0^{375}$$

$$=1+\left[e^{-3.75}-e^{0}\right]$$

$$= 0.0235$$

EXERCISE

1. The length of a telephone conversation has an exponential distribution with a mean of 5 minutes. Find the probability that a call (i) ends in less than 5 minutes, (ii) takes between 5 and 10 minutes. **(VTU 2019)**
2. The average lifetime of a car is 15 years and it is exponentially decreases. If you buy a 10 years old car, what is the probability that it is in service after 10 years of purchase from your side.
3. The mileage (in thousands of kilometres) which car owners get with a certain kind of radial tyre is a random variable having an exponential distribution with mean 40. Find the probabilities that one of these tyres will last (i) at least 20,000 kms, (ii) at most 30,000 kms.
4. The sale per day in a shop is exponentially distributed with mean is Rs.100. If sales tax is levied at the rate of 8%, what is the probability that the sales tax return from that shop will not exceed Rs.60 per day?

Module-5
Part – 1
Joint Probability

Joint Probability Distribution

If X and Y are discrete random variables defined on the samples space S then the joint probability function of X and Y is defined by

$$P(X = x, Y = y) = f(x, y)$$

where $f(x,y)$ satisfy the conditions

(i) $f(x,y) \geq 0$ and

(ii) $\sum_x \sum_y f(x,y) = 1$

Suppose $X = \{x_1, x_2, \ldots, x_m\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$ then

$$P(X = x_i, Y = y_j) = f(x_i, y_j)$$

and is denoted by J_{ij}, i.e., $J_{ij} = P(X = x_i, Y = y_j) = f(x_i, y_j)$

The set of values of this function for i = 1, 2, 3, ... , m and j = 1, 2, 3, ... , n is called the joint probability distribution of X and Y.

These values are presented in the form of a table called joint probability table.

$X \backslash Y$	y_1	y_2	…	y_n	Sum
x_1	J_{11}	J_{12}	…	J_{1n}	$f(x_1)$
x_2	J_{21}	J_{22}	…	J_{2n}	$f(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	J_{m1}	J_{m2}	…	J_{mn}	$f(x_m)$
Sum	$g(y_1)$	$g(y_2)$	…	$g(y_n)$	1

In the table,

$$f(x_1) = J_{11} + J_{12} + \ldots + J_{1n}; \qquad g(y_1) = J_{11} + J_{21} + \ldots + J_{m1}$$

$$f(x_2) = J_{21} + J_{22} + \ldots + J_{2n}; \qquad g(y_2) = J_{12} + J_{22} + \ldots + J_{32}$$

$$\vdots$$

$$f(x_m) = J_{m1} + J_{m2} + \ldots + J_{mn}; \qquad g(y_n) = J_{1n} + J_{2n} + \ldots + J_{mn}$$

The following tables are called marginal probability distributions of X and Y

x_i	x_1	x_2	$\ldots$	x_m
$f(x_i)$	$f(x_1)$	$f(x_2)$	$\ldots$	$f(x_m)$

y_j	y_1	y_2	$\ldots$	y_m
$g(y_j)$	$g(y_1)$	$g(y_2)$	$\ldots$	$g(y_n)$

Note: The total of all the entries in the joint probability table is equal to one.

i.e., $f(x_1)+f(x_2)+...+f(x_m)=1$ and $g(y_1)+g(y_2)+...+g(y_m)=1$

This is equivalent to writing

$$\sum_x \sum_y f(x_i, y_j) = \sum_x \sum_y J_{ij} = 1$$

Independent Random variables

The discrete random variables X and Y are said to be independent random variables if

$$P(X=x, Y=y) = P(X=x).P(Y=y)$$

$$i.e., \ P(X=x_i, Y=y_j) = P(X=x_i).P(Y=y_j)$$

This is equivalent to $f(x_i).g(y_j) = J_{ij}$ in the joint probability table.

Note: If $f(x_i).g(y_j) \neq J_{ij}$ then X and Y are dependent.

Expectation, Variance and Covariance

If X and Y are discrete random variables taking values having probability function f(x) and g (y) then

1. The Expectation of X is $E(X) = \sum_{i=1}^{m} x_i f(x_i)$

2. The Expectation of Y is $E(Y) = \sum_{j=1}^{n} y_j g(y_j)$

3. The Expectation of XY is $E(XY) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i \, y_j \, J_{ij}$

4. The variance of X is $V(X)=E(X^2)-[E(X)]^2$ where $E(X^2)=\sum_{i=1}^{m} x_i^2 f(x_i)$

5. The variance of Y is $V(Y)=E(Y^2)-[E(Y)]^2$ where $E(Y^2)=\sum_{j=1}^{n} y_j^2 g(y_j)$

6. The standard deviation of X is $\sigma_X=\sqrt{V(X)}$

7. The standard deviation of Y is $\sigma_Y=\sqrt{V(Y)}$

8. The covariance of X and Y is $COV(X,Y)=E(XY)-E(X).E(Y)$

9. The correlation of X and Y is $\rho(X,Y)=\dfrac{COV(X,Y)}{\sigma_X \sigma_Y}$

Note: If X and Y are independent random variables then

(i) $E(XY)=E(X).E(Y)$

(ii) $COV(X,Y)=0$ and hence $\rho(X,Y)=0$

Example 5.1.1: The joint distribution of two random variables X and Y is as follows:

X \ Y	−4	2	7
1	1/8	1/4	1/8
5	1/4	1/8	1/8

Find marginal probability distributions of *X* and *Y* and compute the following:

(i) E (*X*) and E (*Y*), (ii) E(*XY*), (iii) σ_X and σ_Y (iv) COV (*X*, *Y*) and (v) $\rho(X,Y)$

Marginal probability distributions of X and Y are got by adding all the respective row entries and the respective column entries.

x_i	1	5
$f(x_i)$	1/2	1/2

y_j	– 4	2	7
$g(y_j)$	3/8	3/8	1/4

(i) $E(X)=\sum_{i=1}^{m} x_i f(x_i)=(1)\left(\frac{1}{2}\right)+(5)\left(\frac{1}{2}\right)=3$

$$E(Y)=\sum_{j=1}^{n} y_j g(y_j)=(-4)\left(\frac{3}{8}\right)+(2)\left(\frac{3}{8}\right)+(7)\left(\frac{1}{4}\right)=1$$

(ii) $E(XY)=\sum_{i=1}^{m}\sum_{j=1}^{n} x_i\, y_j\, J_{ij}$

$$=(1)(-4)\left(\frac{1}{8}\right)+(1)(2)\left(\frac{1}{4}\right)+(1)(7)\left(\frac{1}{8}\right)+(5)(-4)\left(\frac{1}{4}\right)+(5)(2)\left(\frac{1}{8}\right)+(5)(7)\left(\frac{1}{8}\right)$$

$$=\frac{3}{2}$$

(iii) $E\left(X^2\right)=\sum_{i=1}^{m} x_i^2 f\left(x_i\right)=(1)^2\left(\frac{1}{2}\right)+(5)^2\left(\frac{1}{2}\right)=13$

$$E\left(Y^2\right)=\sum_{j=1}^{n} y_j^2 g\left(y_j\right)=(16)\left(\frac{3}{8}\right)+(4)\left(\frac{3}{8}\right)+(49)\left(\frac{1}{4}\right)=\frac{79}{4}$$

$$V(X)=E\left(X^2\right)-[E(X)]^2=13-3^2=4$$

$$V(Y)=E\left(Y^2\right)-[E(Y)]^2=\frac{79}{4}-1^2=\frac{75}{4}$$

$\sigma_X=\sqrt{V(X)}=\sqrt{4}=2$ and $\sigma_Y=\sqrt{V(Y)}=\sqrt{18.75}=4.33$

(iv) $COV(X,Y)=E(XY)-E(X).E(Y)=\frac{3}{2}-(3)(1)=-\frac{3}{2}$

(v) $\rho(X,Y)=\frac{COV(X,Y)}{\sigma_X\sigma_Y}=\frac{-(3/2)}{(2)(4.33)}=-0.1732$

Example 5.1.2: The joint distribution of two random variables *X* and *Y* is as follows:

X \ Y	− 3	2	4
1	0.1	0.2	0.2
3	0.3	0.1	0.1

Find marginal probability distributions of *X* and *Y* and compute the following:

(i) E (*X*) and E (*Y*), (ii) E(*XY*), (iii) σ_X and σ_Y (iv) COV (*X*, *Y*) and (v) $\rho(X,Y)$

Marginal probability distributions of X and Y are got by adding all the respective row entries and the respective column entries.

x_i	1	3
$f(x_i)$	0.5	0.5

y_j	– 3	2	4
$g(y_j)$	0.4	0.3	0.3

(i) $E(X)=\sum_{i=1}^{m} x_i f(x_i)=(1)(0.5)+(3)(0.5)=2$

$E(Y)=\sum_{j=1}^{n} y_j g(y_j)=(-3)(0.4)+(2)(0.3)+(4)(0.3)=0.6$

(ii) $E(XY)=\sum_{i=1}^{m}\sum_{j=1}^{n} x_i\, y_j\, J_{ij}$

$=(1)(-3)(0.1)+(1)(2)(0.2)+(1)(4)(0.2)+(3)(-3)(0.3)+(3)(2)(0.1)+(3)(4)(0.1)$

$=0$

(iii) $E\left(X^2\right)=\sum_{i=1}^{m} x_i^2 f\left(x_i\right)=(1)^2(0.5)+(3)^2(0.5)=5$

$$E\left(Y^2\right)=\sum_{j=1}^{n} y_j^2 g\left(y_j\right)=(-3)^2(0.4)+(2)^2(0.3)+(4)^2(0.3)=9.6$$

$$V(X)=E\left(X^2\right)-\left[E(X)\right]^2=5-2^2=1$$

$$V(Y)=E\left(Y^2\right)-\left[E(Y)\right]^2=9.6-(0.6)^2=9.24$$

$$\sigma_X=\sqrt{V(X)}=\sqrt{1}=1 \text{ and } \sigma_Y=\sqrt{V(Y)}=\sqrt{9.24}=3.04$$

(iv) $COV(X,Y)=E(XY)-E(X).E(Y)=0-1.2=-1.2$

(v) $\rho(X,Y)=\dfrac{COV(X,Y)}{\sigma_X\sigma_Y}=\dfrac{-1.2}{(1)(3.04)}=-0.3947$

Example 5.1.3: A fair coin is tossed thrice. The random variables X and Y are defined as follows: $X = 0$ or 1 according as head or tail occurs on the first toss. Y = Number of heads. Determine (i) the distributions of X and Y, (ii) the joint distribution of X and Y, (iii) the expectations of X, Y and XY, (iv) standard deviations of X and Y, (v) covariance and correlation of X and Y.

The sample space S and the association of random variables X and Y is

S	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
X	0	0	0	1	0	1	1	1
Y	3	2	2	2	1	1	1	0

Here $X = \{0, 1\}$ and $Y = \{0, 1, 2, 3\}$

(i) $P(X=0)=\frac{4}{8}=\frac{1}{2}$; $\quad P(X=1)=\frac{4}{8}=\frac{1}{2}$

$P(Y=0)=\frac{1}{8}$; $\quad P(Y=1)=\frac{3}{8}$; $\quad P(Y=2)=\frac{3}{8}$; $\quad P(Y=3)=\frac{1}{8}$

Thus the probability distribution X and Y are

x_i	0	1
$f(x_i)$	1/2	1/2

y_j	0	1	2	3
$g(y_j)$	1/8	3/8	3/8	1/8

(ii) The joint distribution of X and Y is $J_{ij} = P\left(X = x_i, Y = y_j\right) = f\left(x_i, y_j\right)$

$$J_{11} = P\left(X = 0, Y = 0\right) = 0; \qquad J_{12} = P\left(X = 0, Y = 1\right) = \frac{1}{8};$$

$$J_{13} = P\left(X = 0, Y = 2\right) = \frac{2}{8} = \frac{1}{4}; \qquad J_{14} = P\left(X = 0, Y = 3\right) = \frac{1}{8};$$

$$J_{21} = P\left(X = 1, Y = 0\right) = \frac{1}{8}; \qquad J_{22} = P\left(X = 1, Y = 1\right) = \frac{2}{8} = \frac{1}{4};$$

$$J_{23} = P\left(X = 1, Y = 2\right) = \frac{1}{8}; \qquad J_{24} = P\left(X = 1, Y = 3\right) = 0;$$

∴ The joint probability distribution of X and Y is

$X \backslash Y$	0	1	2	3	Sum
0	0	1/8	1/4	1/8	1/2
1	1/8	1/4	1/8	0	1/2
Sum	1/8	3/8	3/8	1/8	1

(iii) $E(X)=\sum_{i=1}^{m} x_i f(x_i)=(0)(0.5)+(1)(0.5)=\frac{1}{2}$

$$E(Y)=\sum_{j=1}^{n} y_j g(y_j)=(0)\left(\frac{1}{8}\right)+(1)\left(\frac{3}{8}\right)+(2)\left(\frac{3}{8}\right)+(3)\left(\frac{1}{8}\right)=\frac{3}{2}$$

$$E(XY)=\sum_{i=1}^{m}\sum_{j=1}^{n}x_i\,y_j\,J_{ij}$$

$$=(0)(0)(0)+(0)(1)\left(\frac{1}{8}\right)+(0)(2)\left(\frac{1}{4}\right)+(0)(3)\left(\frac{1}{8}\right)+(1)(0)\left(\frac{1}{8}\right)$$

$$+(1)(1)\left(\frac{1}{4}\right)+(1)(2)\left(\frac{1}{8}\right)+(1)(3)\left(\frac{1}{8}\right)$$

$$=\frac{1}{2}$$

(iv) $E\left(X^2\right)=\sum_{i=1}^{m}x_i^2 f\left(x_i\right)=(0)\left(\frac{1}{2}\right)+(1)\left(\frac{1}{2}\right)=\frac{1}{2}$

$$E\left(Y^2\right)=\sum_{j=1}^{n}y_j^2 g\left(y_j\right)=(0)\left(\frac{1}{8}\right)+(1)\left(\frac{3}{8}\right)+(4)\left(\frac{3}{8}\right)+(9)\left(\frac{1}{8}\right)=3$$

$$V(X)=E(X^2)-[E(X)]^2=\frac{1}{2}-\frac{1}{4}=\frac{1}{4}$$

$$V(Y)=E(Y^2)-[E(Y)]^2=3-\frac{9}{4}=\frac{3}{4}$$

$$\sigma_X=\sqrt{V(X)}=\frac{1}{2} \text{ and } \sigma_Y=\sqrt{V(Y)}=\sqrt{3/4}=\frac{\sqrt{3}}{2}$$

(v) $$COV(X,Y)=E(XY)-E(X).E(Y)=\frac{1}{2}-\frac{3}{4}=-\frac{1}{4}$$

$$\rho(X,Y)=\frac{COV(X,Y)}{\sigma_X\sigma_Y}=\frac{-(1/4)}{(1/2)(\sqrt{3}/2)}=-\frac{1}{\sqrt{3}}$$

Example 5.1.4: If the joint probability distribution of X and Y is given by $f(x,y)=c(x^2+y^2)$, for $x=-1, 0, 1, 3$; $y=-1, 2, 3$. Find (i) the value of c, (ii) $P(X=0, Y\le 2)$, (iii) $P(X\le 1, Y>2)$, (iv) $P(X\ge 2-Y)$.

(i) Since $f(x,y)$ is a probability distribution, we have

$$f(x,y)\ge 0 \Rightarrow c\ge 0$$

Also, $\sum_x \sum_y f(x,y)=1$

$$\Rightarrow \quad f(-1,-1)+f(-1,2)+f(-1,3)+f(0,-1)+f(0,2)+f(0,3)+f(1,-1)+f(1,2) +f(1,3)+f(3,-1)+f(3,2)+f(3,3)=1$$

$$\Rightarrow \quad c[2+5+10+1+4+9+2+5+10+10+13+18]=1$$

$$\Rightarrow \quad 89c=1 \Rightarrow c=\frac{1}{89}$$

(ii) $P(X=0, Y \le 2) = f(0,-10) + f(0,2) = \frac{1}{89}[1+4] = \frac{5}{89}$

(iii) $P(X \le 1, Y > 2) = f(-1, 3) + f(0, 3) + f(1, 3) = \frac{1}{89}[10+9+10] = \frac{29}{89}$

(iv) $P(X \ge 2-Y) = f(3,-1) + f(3,2) + f(3,3) + f(1,2) + f(1,3) + f(0,3) + f(0,2) + f(-1,3)$

$$= \frac{1}{89}[10+13+18+5+10+9+4+10] = \frac{79}{89}$$

Example 5.1.5: Suppose X and Y are independent random variables with the following distributions:

x_i	1	2
$f(x_i)$	0.7	0.3

y_j	– 2	5	8
$g(y_j)$	0.3	0.5	0.2

Find the joint distribution of X and Y. Show that X and Y are independent random variables and also find COV (X, Y).

$$E(X)=\sum_{i=1}^{m} x_i f(x_i)=(1)(0.7)+(2)(0.3)=1.3$$

$$E(Y)=\sum_{j=1}^{n} y_j g(y_j)=(-2)(0.3)+(5)(0.5)+(8)(0.2)=3.5$$

$$E(X).E(Y)=(1.3)(3.5)=4.55$$

Suppose X and Y are independent random variables then the joint distribution of X and Y is $J_{ij}=f(x_i).g(y_j)$

$J_{11}=(0.7)(0.3)=0.21;$ $\quad J_{12}=(0.7)(0.5)=0.35;$ $\quad J_{13}=(0.7)(0.2)=0.14$

$J_{21}=(0.3)(0.3)=0.09;$ $\quad J_{22}=(0.3)(0.5)=0.15;$ $\quad J_{23}=(0.3)(0.2)=0.06$

$\therefore$ The joint probability distribution of X and Y is

$X \backslash Y$	-2	5	8	Sum
1	0.21	0.35	0.14	0.7
2	0.09	0.15	0.06	0.3
Sum	0.3	0.5	0.2	1

$$E(XY) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i \, y_j \, J_{ij}$$

$$= (1)(-2)(0.21) + (1)(5)(0.35) + (1)(8)(0.14) + (2)(-2)(0.09)$$

$$+ (2)(5)(0.15) + (2)(8)(0.06) = 4.55$$

$$E(XY)=E(X).E(Y)$$

$\therefore$ X and Y are independent random variables

Hence $COV(X,Y)=0$

Example 5.1.6: The joint probability distribution of two discrete random variables *X* and *Y* is given by $f(x,y)=k(2x+y)$ for $0 \le x \le 2; 0 \le y \le 3$. Find (i) the value of *k*, (ii) the marginal distribution of X and Y, (iii) show that *X* and *Y* are dependent.

Given $X = \{0, 1, 2\}$ and $Y = \{0, 1, 2, 3\}$

$$f(x,y)=k(2x+y)$$

The joint probability distribution of *X* and *Y* is

$X \backslash Y$	0	1	2	3	Sum
0	0	k	$2k$	$3k$	$6k$
1	$2k$	$3k$	$4k$	$5k$	$14k$
2	$4k$	$5k$	$6k$	$7k$	$22k$
Sum	$6k$	$9k$	$12k$	$15k$	$42k$

(i) We must have $42k = 1$

$\Rightarrow \quad k = 1/42$

(ii) Marginal probability distribution X and Y are

x_i	0	1	2
$f(x_i)$	1/7	1/3	11/21

y_j	0	1	2	3
$g(y_j)$	1/7	3/14	2/7	5/14

(iii) Here, $f(x_i).g(y_j) \neq J_{ij}$

Hence X and Y are dependent.

Example 5.1.8: ***X*** **and** ***Y*** **are independent random variables.** ***X*** **take values 2, 5, 7 with probabilities 1/2, 1/4, 1/4 respectively.** ***Y*** **take values 3, 4, 5 with the probabilities 1/3, 1/3, 1/3.**

(i) **Find the joint probability distribution of** ***X*** **and** ***Y*****,**
(ii) **Show that the covariance of** ***X*** **and** ***Y*** **is equal to zero.**
(iii) **Find the probability distribution of** $Z = X + Y$

Given data is as follows

x_i	2	5	7
$f(x_i)$	1/2	1/4	1/4

y_j	3	4	5
$g(y_j)$	1/3	1/3	1/3

(i) The joint distribution of X and Y is $J_{ij} = P\left(X = x_i, Y = y_j\right) = f\left(x_i, y_j\right)$

$$J_{11} = P\left(X = 2, Y = 3\right) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}; \qquad J_{12} = P\left(X = 2, Y = 4\right) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6};$$

$$J_{13} = P\left(X = 2, Y = 5\right) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}; \qquad J_{21} = P\left(X = 5, Y = 3\right) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12};$$

$$J_{22} = P\left(X = 5, Y = 4\right) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}; \qquad J_{23} = P\left(X = 5, Y = 5\right) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12};$$

$$J_{31} = P\left(X = 7, Y = 3\right) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}; \qquad J_{32} = P\left(X = 7, Y = 4\right) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12};$$

$$J_{32} = P\left(X = 7, Y = 5\right) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

∴ The joint probability distribution of X and Y is

X \ Y	3	4	5	Sum
2	1/6	1/6	1/6	1/2
5	1/12	1/12	1/12	1/4
7	1/12	1/12	1/12	1/4
Sum	1/3	1/3	1/3	1

(ii) $E(X) = \sum_{i=1}^{m} x_i f(x_i) = (2)\left(\frac{1}{2}\right) + (5)\left(\frac{1}{4}\right) + (7)\left(\frac{1}{4}\right) = 4$

$$E(Y) = \sum_{j=1}^{n} y_j g(y_j) = (3)\left(\frac{1}{3}\right) + (4)\left(\frac{1}{3}\right) + (5)\left(\frac{1}{3}\right) = 4$$

$$E(XY) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i\, y_j\, J_{ij}$$

$$= (2)(3)\left(\frac{1}{6}\right) + (2)(4)\left(\frac{1}{6}\right) + (2)(5)\left(\frac{1}{6}\right) + (5)(3)\left(\frac{1}{12}\right) + (5)(4)\left(\frac{1}{12}\right)$$

$$+ (5)(5)\left(\frac{1}{12}\right) + (7)(3)\left(\frac{1}{12}\right) + (7)(4)\left(\frac{1}{12}\right) + (7)(4)\left(\frac{1}{12}\right) = 16$$

$$COV(X,Y) = E(XY) - E(X).E(Y) = 16 - (4)(4) = 0$$

(iii) $Z = X + Y$

Let $z_i = x_i + y_i$ and hence $z_i = 5, 6, 7, 8, 9, 10, 11, 12$

The corresponding probabilities are 1/6, 1/6, 1/6, 1/12, 1/12, 1/6, 1/12, 1/12

The probability distribution of $Z = X + Y$ is as follows:

Z	5	6	7	8	9	10	11	12
$P(Z)$	1/6	1/6	1/6	1/12	1/12	1/6	1/12	1/12

Example 5.1.7: If X and Y have the joint probabilities shown in the following table:

$X \backslash Y$	0	1	2	3
0	1/12	1/4	1/8	1/120
1	1/6	1/4	1/20	-
2	1/24	1/40	-	-

Find (i) $P(X = 1, Y = 2)$, (ii) $P(X = 0, 1 \leq Y < 3)$,

(iii) $P(X + Y < 2)$, (iv) $P(X < Y)$

EXERCISE

1. The joint distribution of two random variables X and *Y* is as follows:

$X \backslash Y$	1	3	6
1	1/9	1/6	1/18
3	1/6	1/4	1/12
6	1/18	1/12	1/36

Determine the marginal distribution of *X* and *Y*. Also find whether X and Y are independent or not. **(VTU 2006, 2009)**

2. The joint distribution of two random variables X and Y is as follows:

$X \backslash Y$	-3	2	4
1	0.1	0.2	0.2
3	0.3	0.1	0.1

Find marginal probability distributions of X and Y and compute the following:

(i) E (X) and E (Y), (ii) E(XY), (iii) σ_X and σ_Y (iv) COV (X, Y) and (v) $\rho(X,Y)$

3. The joint distribution of two random variables X and Y is as follows:

X \ Y	1	3	9
2	1/8	1/24	1/12
4	1/4	1/4	0
6	1/8	1/24	1/12

Find marginal probability distributions of X and Y and compute the following:

(i) E (X) and E (Y), (ii) E(XY), (iii) σ_X and σ_Y (iv) COV (X, Y) and (v) $\rho(X,Y)$

4. A fair coin is tossed twice. The random variables X and Y are defined as follows: $X = 1$ or 0 according as head or tail occurs on the first toss, $Y = 1$ if both the tosses are head and $Y = 0$ otherwise. Determine the marginal probability distributions of X and Y and the joint distribution of X and Y and also verify that X and Y are independent or not.

5. Two cards are drawn at random from a box which contains five cards numbered 1, 1, 2, 2 and 3. Let X denotes the sum and Y denote the maximum of the two numbers drawn. Find the joint distribution of X and Y. Also compute COV (X, Y) and $\rho(X,Y)$

6. Let X be random variable with the following distribution and Y defined by X^2

$X(=x_i)$	-2	-1	1	2
$f(x_i)$	$1/4$	$1/4$	$1/4$	$1/4$

Determine; (i) the distribution of X of Y (ii) joint distribution of X and Y (iii) E (XY) **(VTU 2017)**

7. X and Y are independent random variables, X takes the values 1, 2 with probability 0.7, 0.3 and Y take the values –2, 5, 8 with probabilities 0, 3, 0.5, 0.2. Find the joint distribution of X and Y hence find Cov (X, Y).

(VTU 2016)

5.2 Test of significance - t test

Working rule:

- Write the null hypothesis H_0 and the alternative hypothesis H_1.
- Find the calculated value using $|t| = \left|\frac{\bar{x}-\mu}{S.E\ (\bar{x})}\right|$, Where $S.E(\bar{x}) = \sqrt{\frac{s^2}{n-1}}$

$$(or)$$

$$|t| = \left|\frac{\overline{x_1} - \overline{x_2}}{SE(\overline{x_1} - \overline{x_2})}\right|, \text{ where } S.E(\overline{x_1} - \overline{x_2}) = \sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}$$

- Find the critical value using the table at $n-1$ or n_1+n_2-2 degrees of freedom.
- If calculated value < critical value, Accept H_0. H_0 is the conclusion.
- If calculated value > critical value, Reject H_0. H_1 is the conclusion.

t distribution: Critical values of t

		Significance level					
Degrees of freedom	*Two-tailed test:*	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819
22		1.717	2.074	2.508	2.819	3.505	3.792
23		1.714	2.069	2.500	2.807	3.485	3.768
24		1.711	2.064	2.492	2.797	3.467	3.745
25		1.708	2.060	2.485	2.787	3.450	3.725
26		1.706	2.056	2.479	2.779	3.435	3.707
27		1.703	2.052	2.473	2.771	3.421	3.690
28		1.701	2.048	2.467	2.763	3.408	3.674
29		1.699	2.045	2.462	2.756	3.396	3.659
30		1.697	2.042	2.457	2.750	3.385	3.646

1. **A Machinist making engine parts with axle diameter of 0.7 inches. A random sample of 10 parts shows mean diameter 0.742 inches with a SD of 0.04 inches. On the basis of this sample, would you say that the work is inferior at 5% level of significance?**

$$[t_{(.05,\ 9)} = 2.26]$$

Since n = 10, apply t test.

By data, $\bar{x} = 0.742, s = 0.04, \mu = 0.7, \alpha = 0.05$

H0: $\mu = 0.7$, The work is not inferior.

$$S.E(\bar{x}) = \sqrt{\frac{s^2}{n-1}} = \sqrt{\frac{0.04^2}{9}} = 0.0133,$$

$$|t| = \left|\frac{\bar{x}-\mu}{\text{S E}(\bar{x})}\right| = \left|\frac{0.742-0.7}{0.0133}\right| = 3.1579$$

Calculated value of $t = 3.1579$

$\alpha = 0.05, \gamma = n - 1 = 9.$

Therefore, Critical value of $t = 2.26$

Since calculated value > critical value,

Reject H0.

Therefore, the work is inferior.

2. **The nine items of the sample have the following values: 45, 47, 50, 52, 48, 47, 49, 53 and 51. Does the mean of these differ significantly from the assumed mean of 47.5?** $[t_{(.05,\ 8)} = 2.31]$

Since n = 9, apply t test.

By data, $\mu = 47.5, \alpha = 0.05$

$$\bar{x} = \frac{\sum x}{n} = \frac{442}{9} = 49.11$$

$$s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{21762}{9} - 49.11^2 = 6.2079$$

$$S.E(\bar{x}) = \sqrt{\frac{s^2}{n-1}} = \sqrt{\frac{6.2079}{8}} = 0.8809$$

$H_0: \mu = 47.5$, There is no significant difference from the assumed mean 47.5

$$|t| = \left|\frac{\bar{x}-\mu}{S.E(\bar{x})}\right| = \left|\frac{49.11-47.5}{0.8809}\right| = 1.8276$$

Therefore, calculated value of t = 1.8276

$\alpha = 0.05,\ \gamma = n - 1 = 8,$

Therefore, Critical value of t = 2.31

Since calculated value < critical value,

Accept H_0.

∴There is no significant difference from the assumed mean 47.5

3. **A random sample of 10 boys had the following IQ: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Does this data support the hypothesis that the population mean of IQ's is 100 at 5% level of significance?** $[t_{(.05,\ 9)} = 2.26$

Since n = 10, apply t - test.

$$\bar{x} = \frac{\sum x}{n} = \frac{972}{10} = 97.2$$

$$s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{96312}{10} - 97.2^2 = 183.36$$

$$S.E(\bar{x}) = \sqrt{\frac{s^2}{n-1}} = \sqrt{\frac{183.36}{9}} = 4.5136$$

H_0: $\mu = 100$, The population mean of IQ's is 100.

$$|t| = \left|\frac{\bar{x} - \mu}{SE(\bar{x})}\right| = \left|\frac{97.2 - 100}{4.5136}\right| = 0.6203$$

Therefore, calculated value of t = 0.6203

$\alpha = 0.05, \qquad \gamma = n - 1 = 9.$

Therefore, Critical value of t = 2.26

Since calculated value < critical value,

Accept H_0.

Therefore, the population mean IQ is100 at 5% level of significance.

4. **A certain stimulus administered to each of 12 patients resulted in the following increases of blood pressure: 5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4, 6. Can it be concluded that the stimulus will in general be accompanied by an increase in blood pressure?**

$[t_{(.05,\ 11)} = 2.2$

Since n = 12, apply t - test.

$$\bar{d} = \frac{\sum d}{n} = \frac{31}{12} = 2.5833$$

$$s^2 = \frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2 = \frac{185}{12} - \left(\frac{31}{12}\right)^2 = 8.7433$$

$$S.E(\bar{d}) = \sqrt{\frac{s^2}{n-1}} = \sqrt{\frac{8.7433}{11}} = 0.8915$$

$H_0: \mu = 0$

The stimulus will not increase in blood pressure.

$$|t| = \left|\frac{\bar{d}-\mu}{SE(\bar{d})}\right| = \left|\frac{2.5833-0}{0.8915}\right| = 2.8977$$

Therefore, calculated value of t = 2.8977

$\alpha = 0.05, \gamma = n - 1 = 11$

Critical value of t = 2.2

Since calculated value > critical value,

Reject H_0.

∴The stimulus will increase blood pressure.

5. Eleven students were given a test in statistics. They were given a month's further tuition and a second test of equal difficulty was held at the end of it. Do the marks give evidence that the students have benefitted by extra coaching?

Boys:	1	2	3	4	5	6	7	8	9	10	11
Marks I test:	23	20	19	21	18	20	18	17	23	16	19
Marks II test:	24	19	22	18	20	22	20	20	23	20	17

$[t_{(.05,\ 10)} = 2.23]$

Since n = 11, apply t test.

$$\Sigma d = \Sigma(x_2 - x_1) = 11$$

$$\bar{d} = \frac{\Sigma d}{n} = 1$$

$$s^2 = \frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2$$

$$= \frac{61}{11} - 1$$

$$= \frac{50}{11}$$

$$S.E\,(\bar{x}) = \sqrt{\frac{s^2}{n-1}}$$

$$= \sqrt{\frac{50/11}{10}}$$

$$= \sqrt{\frac{50}{110}}$$

$$= 0.6742$$

H_0: $\mu = 0$, The students did not have benefitted by extra coaching.

$$|t| = \left|\frac{\bar{d}-\mu}{SE(\bar{d})}\right| = \left|\frac{1-0}{0.6742}\right| = 1.4832$$

Therefore, calculated value of t = 1.4832

$\alpha = 0.05, \gamma = n - 1 = 10.$

Therefore, critical value of t = 2.23

Since calculated value < critical value,

Accept H_0. ∴ The students did not have benefit by extra coaching.

6. A group of boys and girls were given an intelligent test. The mean score SD's and numbers in each group are as follows:

	Mean	S.D	n
Boys	124	12	18
Girls	121	10	14

Is the mean score of boys significantly different from that of girls? $[t_{(.05,30)} = 2.04]$

Since $n_1 = 18$, $n_2 = 14$, apply t test.

By data, $\overline{x_1} = 124, \overline{x_2} = 121,$

$s_1 = 12, s_2 = 10$

$$SE(\overline{x_1} - \overline{x_2}) = \sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}$$

$$= \sqrt{\frac{144}{17} + \frac{100}{13}}$$

$$= 4.0203$$

H_0: $\mu_1 = \mu_2$, the mean score of boys does not differ significantly from that of girls.

$$|t| = \left|\frac{\overline{x_1} - \overline{x_2}}{S.E(\overline{x_1} - \overline{x_2})}\right| = \left|\frac{124 - 121}{4.0203}\right| = 0.7462$$

Therefore, calculated value of t = 0.77

$\alpha = 0.05, \gamma = n_1 + n_2 - 2 = 30$

Therefore, critical value of t = 2.04

Since calculated value < critical value, Accept H_0.

∴The mean score of boys does not differ significantly from that of girls.

7. The means of two random samples of sizes 9 and 7 are 196.42 and 198.82 respectively. The sum of squares of the deviations from the mean are 26.94 and 18.73 respectively. Can the sample be considered to have been drawn from the same normal population? $[t_{(.05,\ 14)} = 2.14]$

Since $n_1 = 9$, $n_2 = 7$, apply t test.

By data, $\overline{x_1} = 196.42, \overline{x_2} = 198.$

$s_1^2 = \frac{26.94}{9}, s_2^2 = \frac{18.73}{7}$

$$SE(\overline{x_1} - \overline{x_2}) = \sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}$$

$$= 0.9056$$

H_0: $\mu_1 = \mu_2$, sample is drawn from the same normal population.

$$|t| = \left|\frac{\overline{x_1} - \overline{x_2}}{SE(\overline{x_1} - \overline{x_2})}\right| = \left|\frac{196.42 - 198.82}{0.9056}\right| = 2.6502$$

Therefore, calculated value of t = 2.6502

$\alpha = 0.05, \gamma = n_1 + n_2 - 2 = 9 + 7 - 2 = 14$

Critical value of t = 2.14

Since calculated value < critical value,

Accept H_0.

∴ Sample is drawn from the same normal population.

8. **From a random sample of 10 pigs fed on diet A, the increases in weight in a certain period were 10, 6, 16, 17, 13, 12, 8, 14, 15, 9 lbs. For another random sample of 12 pigs fed on diet B, the increases in the same period were 7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17 lbs. Test whether diets A and B differ significantly as regards their effect on increases in weight?** $[t_{(.05,\ 14)} = 2.09]$

Since $n_1 = 10$, $n_2 = 12$, apply t test.

By data, $\overline{x_1} = \frac{\sum x_1}{n_1} = \frac{120}{10} = 12$

$$\overline{x_2} = \frac{\sum x_2}{n_2} = \frac{180}{12} = 15$$

$$s_1^2 = \frac{\sum {x_1}^2}{n_1} - \left(\frac{\sum x_1}{n_1}\right)^2 \qquad s_2^2 = \frac{\sum {x_2}^2}{n_2} - \left(\frac{\sum x_2}{n_2}\right)^2$$

$$= \frac{1560}{10} - 144 \qquad = \frac{3014}{12} - 225$$

$$= 12 \qquad = 26.17$$

$$SE(\overline{x_1} - \overline{x_2}) = \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}$$

$$= \sqrt{\frac{12^2}{9} + \frac{26.17^2}{11}} = 1.93$$

$H_0: \mu_1 = \mu_2$,

diets A and B do not differ significantly.

$$|t| = \left|\frac{\overline{x_1} - \overline{x_2}}{SE(\overline{x_1} - \overline{x_2})}\right| = \left|\frac{12 - 15}{1.93}\right| = 1.6$$

Therefore, calculated value $= 1.6$

$\alpha = 0.05, \gamma = n_1 + n_2 - 2 = 10 + 12 - 2 = 20$

Critical value = 2.09

Since calculated value < critical value,

Accept H_0.

∴Diets A and B do not differ significantly.

Home work

9. A machine is expected to produce nails of length 3 inches. A random sample of 25 nails gave an average length of 31. Inches with standard deviation 0.3. Can it be said that the machine is producing nails as per specifications? Given $t_{0.05}(24) = 2.064$

10. Two horses A and B were tested according to the time (seconds) to run a particular race with the following results:

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whether you can discriminate between the two horses.

Given that $t_{0.05} = 2.20$ for 11 degrees of freedom

Home work

11. Two types of batteries are tested for their length of life and the following results were obtained:

	Battery A	Battery B
Mean	500	500
Variance	100	121
Sample size	10	10

Check whether there is a significant difference between two means. $[t_{0.05}(18) = 0.086]$

12. A sample of 12 measurements of the diameter of a metal ball gave the mean 7.38 mm with standard deviation 1.24 mm. Find 99% confidence limits for actual diameter. $[t_{0.01}(11) = 3.11]$

Note: Confidence limits for the mean are $\bar{x} \pm \frac{s}{\sqrt{n-1}} t_{\alpha}(\gamma)$

5.2 Sampling theory

Introduction:

- Entire group of individuals under study is called **population**. Quantity associated with population like mean(μ), SD(σ) is called **parameter**.

- A small part of the population is called **sample**. Quantity associated with sample like mean($\bar{x}$), SD(s) is called **statistic**.

- The number of units in the sample is called sample size. It is denoted by n. If n < 30, the sample is called **small sample**. If $n \geq 30$ the sample is called **large sample**.

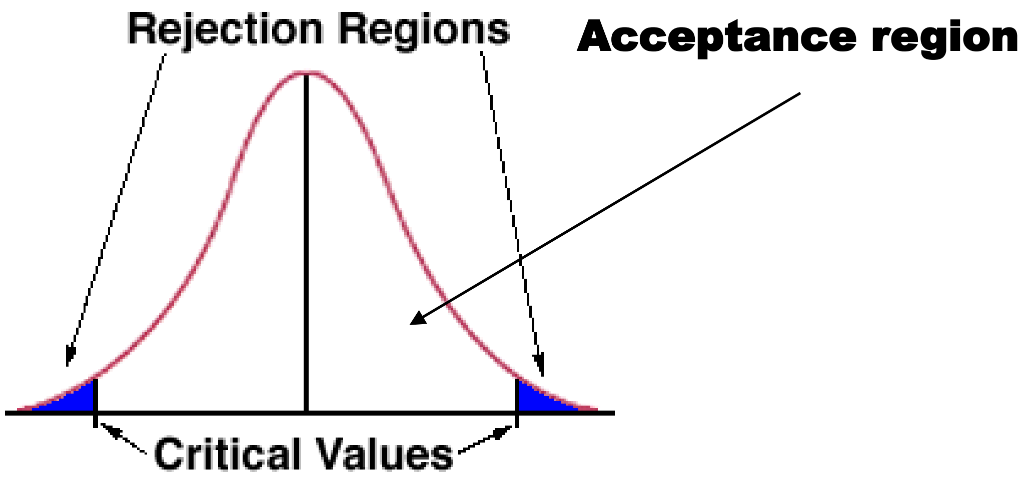
- The distribution of values of the statistic for different samples is called **sampling distribution** of the statistic. The SD of sampling distribution of a statistic is called **standard error** of a statistic.

Test of significance

- Some assumption about the population is called **statistical hypothesis**.
- A statistical hypothesis which we formulate to check whether it can be rejected is called **null hypothesis** (H_0). A hypothesis which differs from the null hypothesis is called **alternative hypothesis** (H_1).
- Procedure which enables to decide whether to accept or reject the null hypothesis is called **test of hypothesis** or **test of significance**.

Acceptance region and critical region:

- The limits of the critical region are called **critical values**.
- Critical value splits the region in to **acceptance region** and **critical region**.



Critical value of z

If $H_1: \bar{x} \neq \mu_0$ then the test is two tailed. Use $z_{\alpha/2}$.

If $H_1: \bar{x} < (>) \mu_0$ then the test is singled tailed. Use z_{α}.

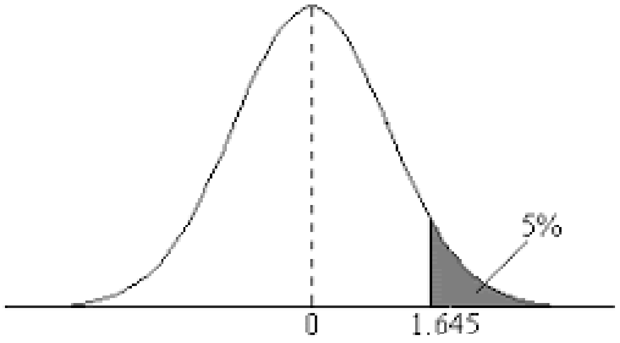
At 1% level of significance

$z_\alpha = z_{0.01}$ $R(x) = 0.01$ $x = 2.33$	$z_{\alpha/2} = z_{0.05}$ $R(x) = 0.005$ $x = 2.58$

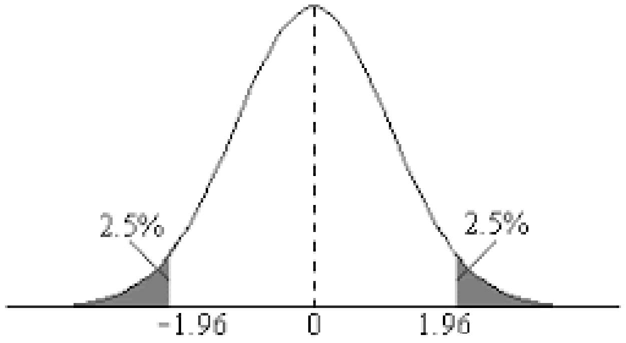
At 5% level of significance

$z_\alpha = z_{0.05}$ $R(x) = 0.05$ $x = 1.65$	$z_{\alpha/2} = z_{0.025}$ $R(x) = 0.025$ $x = 1.96$

	$z_{\alpha/2}$	z_α
$\alpha = 0.05$	1.96	1.65
$\alpha = 0.01$	2.58	2.33



(a) One-tailed test z_α



(b) Two-tailed test $z_{\alpha/2}$

Calculated value of z

$$|z| = \left|\frac{x-\mu}{\sigma}\right| \;(or)\; \left|\frac{\bar{x}-\mu}{S.E\,(\bar{x})}\right| \;(or)\; \left|\frac{\overline{x_1}-\overline{x_2}}{SE(\overline{x_1}-\overline{x_2})}\right|$$

Use the 1st result if x, μ are known.

Use the 2nd result if $\bar{x}, \mu$ are known.

Use the 3rd result if $\bar{x}_1, \bar{x}_2$ are known.

where $S.E(\bar{x}) = \sqrt{\frac{s^2}{n}}$, $S.E(\overline{x_1} - \overline{x_2}) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

	Mean	**S.D**
Sample	$\bar{x}$	s
Population	μ	σ

Type 1 error and Type 2 error

- Rejecting H_0 when it is true is called **Type I error**. P (Type I error) is called **level of significance**. It is denoted by α.

- Accepting H_0 when it is false is called **Type II error**. P (Type II error) is called **power of the test**. It is denoted by β.

	True	False
Accept H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

5.1 Test of significance - z test

Working rule:

- Write the null hypothesis H_0.

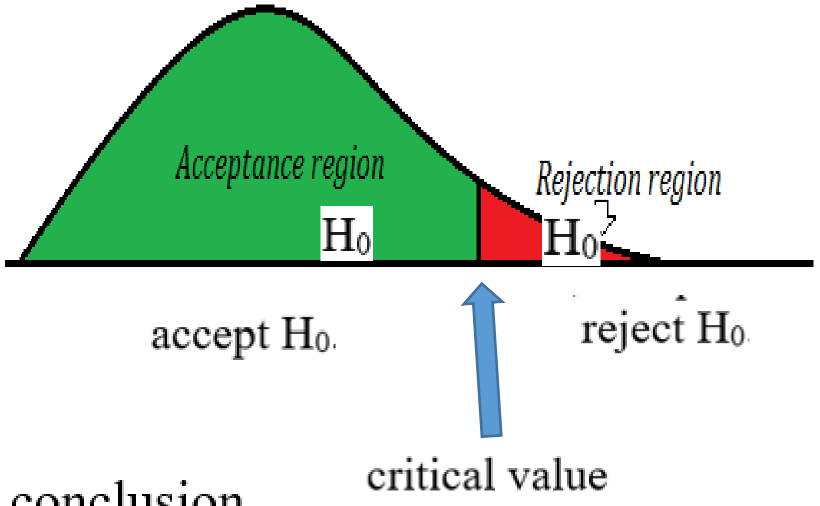
- Find the calculated value using

$$|z| = \left|\frac{x-\mu}{\sigma}\right| \ (or) \ \left|\frac{\bar{x}-\mu}{S.E\ (\bar{x})}\right| \ (or) \ \left|\frac{\overline{x_1}-\overline{x_2}}{SE(\overline{x_1}-\overline{x_2})}\right|$$

where $S.E(\bar{x}) = \sqrt{\frac{s^2}{n}}$, $S.E(\overline{x_1}-\overline{x_2}) = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

- Find the critical value using the table:

	Two tailed	Single tailed
$\alpha = 0.05$	1.96	1.65
$\alpha = 0.01$	2.58	2.33



- If calculated value < critical value, accept H_0. H_0 is the conclusion.

- If calculated value > critical value reject H_0. H_1 is the conclusion.

1. **A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased at 5% level of significance.** $[z_{\frac{\alpha}{2}} = 1.96]$

Since $n = 400$, apply z test.

By data, $x = 216, p = \frac{1}{2}, \ \mu = np = 200$

$H_0: \mu = 200$, The coin is unbiased.

$$|z| = \left|\frac{x-\mu}{\sigma}\right|$$

$$= \left|\frac{x-\mu}{\sqrt{npq}}\right|$$

$$= \left|\frac{216-200}{10}\right|$$

$$= 1.6$$

Therefore, calculated value of $z = 1.6$

At $\alpha = 0.05$, critical value of $z = 1.96$

Since calculated value < critical value,

Accept H_0.

Therefore, the coin is unbiased at 5% level of significance.

2. **A die was thrown 9000 times and a throw of 5 or 6 was obtained 3240 times. On the assumption of random throwing, do the data indicate an unbiased die? $(\alpha = 0.01)$**

$[z_{\frac{\alpha}{2}} = 2.58]$

Since $n = 9000$, apply z test.

By data, $x = 3240,\ p = \frac{1}{3},\ \mu = np = 3000$

$H_0: \mu = 3000$, The die is unbiased.

$$|z| = \left|\frac{x-\mu}{\sigma}\right|$$

$$= \left|\frac{x-\mu}{\sqrt{npq}}\right|$$

$$= \left|\frac{3240-3000}{\sqrt{2000}}\right|$$

$$= 5.4$$

Therefore, calculated value of $z = 5.4$

At $\alpha = 0.01$, critical value of $z = 2.58$

Since calculated value> critical value,

Reject H_0.

Therefore, the die is biased

at 1% level of significance.

3. **In 324 throws of a die, an odd number turned up 181 times. Is it reasonable to think that ay 1% level of significance the die is an unbiased one?**$[z_{\frac{\alpha}{2}} = 2.58]$

Since $n = 324$, apply z test.

By data, $x = 181,\ p = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$,

$\mu = np = 162$

$H_0: \mu = 162$,The die is unbiased.

$$|z| = \left|\frac{x-\mu}{\sigma}\right|$$

$$= \left|\frac{x-\mu}{\sqrt{npq}}\right|$$

$$= \left|\frac{181-162}{\sqrt{162 \times \frac{1}{2}}}\right| = 2.11$$

Therefore, calculated value of $z = 2.11$

At $\alpha = 0.01$, critical value of $z = 2.58$

Since calculated value < critical value,

accept H_0.

Therefore, the die is unbiased

at 1% level of significance.

4. **A sample of 900 members is found to have a mean of 3.4 cm. Can it be reasonably regarded as a truly random sample from a large population with mean 3.25 cm and SD 1.61 cm.**

By data, $\bar{x} = 3.4, n = 900,\ \mu = 3.25, \sigma = 1.61$

This is a large sample. Apply z test.

$$S.E(\bar{x}) = \sqrt{\frac{1.61^2}{900}} = 0.0537$$

$H_0: \mu = 3.25,$

Sample is taken from the population with mean 3.25

$$|z| = \left|\frac{\bar{x} - \mu}{S.E(\bar{x})}\right|$$

$$= \left|\frac{3.4 - 3.25}{0.0537}\right| = 2.8$$

Therefore, calculated value of $z = 2.8$

At $\alpha = 0.05$, critical value of $z = 1.96$

Since calculated value > critical value,

Reject H_0.

Therefore, sample is not taken from

the population with mean 3.25

5. **If a mean breaking strength of copper wire is 575 lbs with a standard deviation 8.3 lbs. How large a sample must be used in order that there be one chance in 100 that the mean breaking strength of the sample is less than 572 lbs. ($Z_\alpha = 2.33$)**

By data, $\bar{x} = 572$, $\mu = 575, \sigma = 8.3$.

This is a large sample. Apply z test.

$$S.E(\bar{x}) = \sqrt{\frac{8.3^2}{n}}$$

$H_0: \mu = 575$,

mean breaking strength of copper wire is 575 lbs.

$$|z| = \left|\frac{\bar{x}-\mu}{S.E(\bar{x})}\right|$$

$$= \left|\frac{572-575}{\sqrt{\frac{8.3^2}{n}}}\right|$$

To find: n such that $\mu > 575$.

Suppose $\mu > 575$

Calculated value > Critical value.

$$\left|\frac{572-575}{\sqrt{\frac{8.3^2}{n}}}\right| > 2.33$$

$$n > 41.56$$

Therefore, $n = 42$.

6. **The means of samples of sizes 1000 and 2000 are 67.5 and 68.0 cm respectively. Can the samples be regarded as drawn from the same population of SD 2.5 cm?[$z_{\frac{\alpha}{2}}(0.05) = 1.96$]**

Since $n_1 = 1000,\ n_2 = 2000$, apply z test.

By data, $\overline{x_1} = 67.5, \overline{x_2} = 68.0$

$\sigma_1 = \sigma_2 = 2.5, \alpha = 0.05$

H0: $\mu_1 = \mu_2$ Both the samples are drawn from the same population.

$$|z| = \left|\frac{\overline{x_1} - \overline{x_2}}{SE(\overline{x_1} - \overline{x_2})}\right|$$

$$= \left|\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right|$$

$$= \left|\frac{67.5 - 68}{\sqrt{2.5/1000 + 2.5/2000}}\right|$$

$$= 5.1$$

Therefore, calculated value of $z = 5.15$

At $\alpha = 0.05$, critical value of $z = 1.96$

Since calculated value > critical value,

Reject H0.

Therefore, Both the samples are not drawn from the same population.

7. **A sample of height of 6400 soldiers has a mean of 67.85 inches and a standard deviation of 2.56 inches while a sample of height of 1600 sailors has a mean of 68.55 inches and a SD of 2.52 inches. Does the data indicate that the sailors are on an average taller than soldiers? Use 0.05 level of significance.** $[z_\alpha = 1.65]$

Since $n_1 = 6400, n_2 = 1600$, apply z test.

By data, $\overline{x_1} = 67.85$, $\overline{x_2} = 68.55$,

$s_1 = 2.56$, $s_2 = 2.52$, $\alpha = 0.05$

H_0: $\mu_1 = \mu_2$, The sailors are not taller than soldiers.

$$|z| = \left|\frac{\overline{x_1} - \overline{x_2}}{SE(\overline{x_1} - \overline{x_2})}\right| = \left|\frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}\right|$$

$$= \left|\frac{67.85 - 68.55}{\sqrt{2.56^2/6400 + 2.52^2/1600}}\right| = \frac{0.7}{0.005} = 140$$

Therefore, calculated value = 140

At $\alpha = 0.05$, critical value = 1.65

Since calculated value > critical value,

Reject H_0.

Therefore, the sailors are taller than soldiers

at 0.05 level of significance.

5.4 Chi square distribution

Test for goodness of fit

Conditions to apply χ^2 test for goodness of fit:

- The observations should be independent.
- The total frequency N should be large.
- If any E_i is less than 5, it should be pooled with the adjacent frequency.
- If any parameter is estimated, corresponding to every such estimation, one degree of freedom should be lessened.

Working rule:

Assume:

Expected frequency distribution is a good fit to the observed frequency distribution.

Calculated value:

Under H_0, $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$ with $n - c$ degrees of freedom.

Where $O_i -$ Observed frequency or tabulated frequency

$E_i -$ Expected frequency or theoretical frequency

$n -$ number of terms, $c -$ number of constraints

Critical value:

Level of significance $\alpha = 0.05$ *or* 0.01 (Always upper tailed)

Degrees of freedom $\gamma = n - c$. Where $c = \begin{cases} 1, & \textit{In general} \\ 2, & \textit{For Poisson distribution} \\ 3, & \textit{For normal distribution} \end{cases}$

Cont.

Conclusion:

If calculated value < critical value, the expected frequency distribution is **a good fit** to the observed frequency distribution.

If calculated value > critical value, the expected frequency distribution **is not a good** fit to the observed frequency distribution.

Critical values of chi square distribution

d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

1. A die is thrown 60 times and the frequency distribution for the number appearing on the face x is given by the following table:

x:	**1**	**2**	**3**	**4**	**5**	**6**
f:	**15**	**6**	**4**	**7**	**11**	**17**

Test the hypothesis that the die is unbiased. $[\chi_{0.05}(5) = 11.07$

$x =$ Number appearing on the face

$= \{1, 2, 3, 4, 5, 6\}$

$P(x) = \frac{1}{6}$, Total frequency $N = 60.$

$$f(x) = N \times P(x) = 60 \times \left(\frac{1}{6}\right) = 10.$$

Put $x = 1, 2, 3, 4, 5, 6$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	15	10	25	2.5
2	6	10	16	1.6
3	4	10	36	3.6
4	7	10	9	0.9
5	11	10	1	0.1
6	17	10	49	4.9
				13.6

Cont.

H_0: The die is unbiased.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 13.6$$

Calculated value = 13.6

Degrees of freedom= $n - 1 = 6 - 1 = 5$

Critical value = $\chi_{0.05}(5) = 11.07$

∴ Calculated value > Critical Value.

Reject H_o.

Therefore, the die is not unbiased.

2. **The following table gives the number of road accidents that occurred in a large city during the various days of a week. Test the hypothesis that the accidents are uniformly distributed over all the days of a week.** $[\chi_{0.05}(6) = 12.59]$

Day:	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
No. of accidents:	14	16	8	12	11	9	14	84

$x =$ Various days of a week

$= \{Sun, Mon, Tue, Wed, Thu, Fri, Sat\}$

$P(x) = \frac{1}{7}$, Total frequency $N = 84$

$f(x) = N \times P(x) = 84 \times \left(\frac{1}{7}\right) = 12$

For $x = Sun, Mon, \ldots$ we get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
Sun	14	12	4	4/12
Mon	16	12	16	16/12
Tue	8	12	16	16/12
Wed	12	12	0	0
Thu	11	12	1	1/12
Fri	9	12	9	9/12
Sat	14	12	4	4/12
				50/12

Cont.

H_o: Accidents are uniformly distributed

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{50}{12}$

Calculated value = 4.165

Degrees of freedom= $n - 1 = 7 - 1 = 6$,

Critical value= $\chi_{0.05}(6) = 12.59$

∴ Calculated value < Critical Value.

Accept H_o.

Therefore, the accidents are uniformly distributed.

3. A set of 5 similar coins is tossed 320 times and the result is

Number of heads:	0	1	2	3	4	5
Frequency:	6	27	72	112	71	32

Test the hypothesis that the data follows a binomial distribution. $[\chi_{0.05}(5) = 11.07]$

By data, $n = 5, p = \frac{1}{2}, q = \frac{1}{2}, N = 320$

$x =$ Number of heads$= \{0, 1, 2, 3, 4, 5\}$

$$f(x) = \mathrm{N} \times P(x)$$

$$= 320 \times nC_x p^x q^{n-x}$$

$$= 320 \times 5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x}$$

$$= 320 \times \frac{1}{32} \times 5C_x$$

$$= 10 \times 5C_x$$

Put $x = 0, 1, 2, 3, 4, 5$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
0	6	10	16	1.6
1	27	50	529	10.58
2	72	100	784	7.84
3	112	100	144	1.44
4	71	50	441	8.82
5	32	10	484	48.4
				78.68

Cont.

H_o: The data follows Binomial distribution.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 78.68$$

Calculated value = 78.68

Degrees of freedom= $n - 1 = 6 - 1 = 5$

Critical value= $\chi_{0.05}(5) = 11.07$

$\therefore$ Calculated value > Critical Value.

Reject H_o.

Therefore, the data does not follow Binomial distribution.

4. Fit a Poisson distribution to the following data and test for its goodness of fit at level of significance 0.05.

x:	0	1	2	3	4
f:	419	352	154	56	19

$[\chi_{0.05}(3) = 7.82]$

$x = \{0, 1, 2, 3, 4\}$

$$m = mean = \frac{\Sigma fx}{\Sigma f} = \frac{904}{1000} = 0.904$$

$$e^{-m} = e^{-0.904} = 0.4049$$

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

$$= \frac{(0.4049)(0.904)^x}{x!}$$

$$f(x) = 1000 \times P(x)$$

$$= 404.9\,\frac{(0.904)^x}{x!}$$

Put $x = \{0, 1, 2, 3, 4\}$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
0	419	405+1	169	0.4033
1	352	366	196	0.5355
2	154	165	121	0.7333
3	56	50	36	0.7200
4	19	11+2	36	2.7692
				5.1613

Numbers added in E_i only to preserve totality.

Cont.

H_o: The data follows Poisson distribution.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 5.1613$$

Calculated value = 5.1613

Degrees of freedom= $n - 2 = 5 - 2 = 3$

($\because$ It follows Poisson distribution)

Critical value = $\chi_{0.05}(3) = 7.82$

$\therefore$ Calculated value < Critical Value.

Accept H_o.

Therefore, the data follows Poisson distribution.

5. Genetic theory states that children having one parent of blood type M and other blood type N will always be one of the three types M, MN, N and that the proportion of these types will on average 1:2:1. A report says that out of 300 children having one M parent and one N parent, 30% were found to be type M , 45% of type MN and remainder of type N. Test the hypothesis by χ^2 test. $[\chi_{0.05}(2) = 5.99]$

x is a set of types of blood

$x = \{M, MN, N\}$

$P(M) = \frac{1}{4}$

$P(MN) = \frac{2}{4}$.

$P(N) = \frac{1}{4}$

$f(x) = 300 \times P(x)$

Put $x = M, MN, N$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
M	30% of 300 $= 90$	$300 \times \frac{1}{4} = 75$	225	3
MN	45% of 300 $= 135$	$300 \times \frac{2}{4} = 150$	225	1.5
M	25% of 300 $= 75$	$300 \times \frac{1}{4} = 75$	0	0
				4.5

Cont.

H_0: The proportion of these types is on average 1:2:1

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 4.5$$

Calculated value = 4.5

Degrees of freedom= $n - 1 = 3 - 1 = 2,$

Critical value= $\chi_{0.05}(2) = 5.99$

$\therefore$ Calculated value < Critical Value.

Accept H_0.

The proportion of these types is on average 1:2:1

6. In experiments on Pea breading, the following frequencies of seeds were obtained:

Round and yellow	Wrinkled and yellow	Round and green	Wrinkled and green	Total
315	101	108	32	556

Theory predicts that the frequencies should be in proportions $9:3:3:1$. Examine the correspondence between theory and experiment. $[\chi_{0.05}(3) = 7.82]$

$$x = Set\ of\ types\ of\ seeds$$

$$= \{RY, WY, RG, WG\}$$

$$P(RY) = \frac{9}{16}, P(WY) = \frac{3}{16},$$

$$P(RG) = \frac{3}{16}, P(WG) = \frac{1}{16}$$

$$f(x) = 556 \times P(x)$$

Put $x = RY, WY, RG, RY$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
RY	315	313	4	0.0128
WY	101	104	9	0.0865
RG	108	104	16	0.1538
WG	32	35	9	0.2571
				0.5102

Cont.

H_o: The frequencies are in proportions $9:3:3:1$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 0.5102$$

Calculated value = 0.5102

Degrees of freedom= $n - 1 = 4 - 1 = 3$

Critical value= $\chi_{0.05}(3) = 7.82$

$\therefore$ Calculated value < Critical Value.

Accept H_o.

The frequencies should be in proportions $9:3:3:1$