UNIT 3

CONTINUOUS PROBABILITY DISTRIBUTIONS, FUNDAMENTAL SAMPLING DISTRIBUTIONS

SIA GROUP

PART-A

SHORT QUESTIONS WITH SOLUTIONS

Q1. Define continuous uniform Distribution.

Answer :

Model Paper-I, Q1(a)

If 'x' is a random variable then it will have a continuous uniform (rectangular) distribution over an interval (a, b) i.e., $(-\infty < a < b < \infty)$ if its Probability Density Function (p.d.f) is constant for all values of x in (a, b). This distribution is given by,

$$\boxed{f(x) = \begin{cases} k \; ; & \text{for } a<x<b \\ 0 \; ; & \text{otherwise} \end{cases}}$$

Here, k is a constant.

Q2. List out the properties of normal distribution.

Answer :

The vairous properties of normal distribution are as follows,

- The curve of normal distribution has a single peak; thus it is unimodal, and is bell-shaped with the highest point over the mean μ.
- It is symmetrical about a vertical line through mean.
- As it is symmetrical about mean, the mean, median and mode of the distribution also have the same value.
- The two tails of the curve extend indefinitely and never touch the horizontal axis.
- The curve height decreases on both the sides of the peak which occurs at the mean.
- The area on either side of the peak is equal to each other.

Q3. Write down the applications of Exponential Distribution.

Model Paper-II, Q1(a)

Answer :

The applications of exponential distribution are as follows,

1. The time required to accomplish the needs of a customer at railway booking counter, banks, petrol pump, ration shops etc.
2. The time during which the electronic components are operated without any breakdown/failure.
3. The time required for any service facility among two successive arrivals.

Q4. What is sampling?

Answer :

Sampling may be defined as the selection of some part of an aggregate or totality on the basis of which a judgement or inference about the aggregate or totality is made. In other words, it is the process of obtaining information about an entire population by examining only a part of it.

In most of the research work and surveys, the usual approach happens to be to 'make generalizations' or to 'draw inferences' based on samples about the parameters of population from which the samples are taken. The researcher quite often selects only a few items from the universe for his study purpose. All this is done on the assumption that the sample data will enable him to estimate the population parameters.

The items so selected constitute what is technically called a sample, their selection process is called sample design and the survey conducted on the basis of sample is described a sample survey. Sample should be truly representative of population characteristics without any bias so that it may result in valid and reliable conclusions.

Q5. Explain the need for sampling.

Answer : Model Paper-III, Q1(e)

Sampling is used in practice for a variety of reasons like,

(i) Sampling can save time and money. A sample study is usually less expensive than a census study and produces results at a relatively faster speed.

(ii) Sampling is the only way when population contains infinitely many members.

(iii) Sampling may enable more accurate measurements for a sample study is generally conducted by trained and experienced investigators.

(iv) Sampling remains the only choice when a test involves the destruction of the items under study.

(v) Sampling usually enables to estimate the sampling error, and thus, assists in obtaining information concerning some characteristics of the population.

Q6. Define sampling distribution. Explain the characteristic of sampling distribution.

Answer :

Sampling Distribution

Samples of a given size may be drawn randomly from a population and the statistical constants like mean, standard deviation may be computed for each such sample. The distribution of each such statistic is called sampling distribution.

Characteristic of a Sampling Distribution

Sampling distribution is a frequency distribution representing the means taken from a great many samples, of the same size. The main characteristic of this is that it approaches normal distribution even went the population distribution is not normal provided the sample size is sufficiently large (greater than 30). The significance of the sampling distribution follows from the fact that the mean of sampling distribution is the same as the mean of the population.

Q7. Describe the sampling distribution of variances.

Answer : Model Paper-I, Q1(f)

The sampling distribution of variance can be obtained by drawing all the possible random samples of size 'n' from the given population and then calculating the variance for each of these samples.

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2}$$

$$= \frac{(x_1-\bar{X})+(x_2-\bar{X})^2+\ldots+(x_n-\bar{X})^2}{\sigma^2}$$

$$\text{Where, } \bar{X} = \frac{x_1+x_2+\ldots+x_n}{n}$$

It is easy to generate the sampling distribution of the related random variable, rather than finding the sampling distribution of S^2 or $\hat{S}^2$.

$$\hat{S}^2 = \frac{n}{n-1}S^2$$

$$\hat{S}^2 = \frac{(x_1-\bar{X})^2+(x^2-\bar{X})^2+\ldots+(x_n-\bar{X})^2}{n-1}$$

$$S^2 = \frac{(x_1-\bar{X})^2+(x_2-\bar{X})^2+\ldots+(x_n-\bar{X})^2}{n}$$

Q8. Prove that the area under the normal curve is unity.

Answer : Model Paper-II, Q1(f)

Proof

The probability density function of the normal distribution is,

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

Therefore, area under the normal curve is equal to,

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

Put $\frac{x-\mu}{\sigma\sqrt{2}} = t$

$$dx = \sigma\sqrt{2}dt$$

$$\therefore \int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2} \sigma\sqrt{2}dt$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt$$

$$= \frac{1}{\sqrt{\pi}} \sqrt{\pi} \quad \left(\because \int_{-\infty}^{\infty} e^{-x^2} dx = 1\right)$$

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

Hence proved.

Q9. Define t-Distribution.

Answer :

For unknown population standard deviation (s_p) and small size (i.e., $n < 30$) sample, 't' distribution (student's 't' distribution) is used for the sampling distribution of mean and workout 't' variable as,

$$t = \frac{(\overline{X} - \mu)}{\frac{S}{\sqrt{n}}}$$

Where, $\sigma_s = \sqrt{\frac{\Sigma(X_i - \overline{X})^2}{n-1}}$

i.e., the sampling standard deviation

t-distribution is also symmetrical and is very close to the distribution of standard normal variate Z, except for small values of 'n'. For large sample size or $n \geq 30$, the distribution of T remains same as that of standard normal. There are different 't' distributions one for each sample size i.e., for different degrees of freedom. The degrees of freedom for a sample of size n is $(n-1)$.

Q10. Two random sample of sizes 15 and 25 are taken from a N(μ, σ^2). Find the probability that the ratio of the sample variances does not exceed 2.28.

Model Paper-III, Q1(f)

Answer :

From the table $F_{0.05} = 2.35 \simeq 2.28$

for $V_1 = n_1 - 1 = 25 - 1 = 24$

$V_2 = n_2 - 1 = 15 - 1 = 14$

$\therefore$ The desired probability is 0.05.

PART-B

ESSAY QUESTIONS WITH SOLUTIONS

3.1 CONTINUOUS PROBABILITY DISTRIBUTIONS

3.1.1 Continuous Uniform Distribution

Q11. Define continuous uniform distribution and derive its mean and variance.

Answer :

Model Paper-II, Q6(a)

Continuous Uniform Distribution

If 'x' is a random variable then it will have a continuous uniform (rectangular) distribution over an interval (a, b) i.e., $(-\infty < a < b < \infty)$ if its Probability Density Function (p.d.f) is constant for all values of x in (a, b). This distribution is given by,

$$\boxed{f(x) = \begin{cases} k \;;\; \text{for } a<x<b \\ 0 \;;\; \text{otherwise} \end{cases}} \quad \text{... (1)}$$

Here, k is a constant.

The total probability of a function is equal to one at all times. Hence,

$$\int_a^b f(x).dx = 1$$

$$\int_a^b k.dx = 1$$

$$k\int_a^b 1.dx = 1$$

$$k[x]_a^b = 1$$

$$k[b-a] = 1$$

$$k = \frac{1}{b-a}$$

Now, substitute the value of k in equation (1), we get,

$$\boxed{f(x) = \begin{cases} \dfrac{1}{b-a} & ; a<x<b \\ 0 & ; \text{otherwise} \end{cases}} \quad \text{... (2)}$$

Here, the variables 'a' and 'b' are referred to as the parameters of the rectangular/uniform distribution.

The Cumulative Distribution Function (CDF) of 'x' is given as follows,

$$F(X) = P[X \leq x]$$

$$= \begin{cases} 0 & ,[X \leq x] \\ \dfrac{x-a}{b-a} & ,a \leq x \leq b \\ 1 & , x \geq b \end{cases}$$

Mean (μ_1')

Now, substitute, $r = 1$ in equation (3) the mean (μ_1') will be,

$$\text{Mean, } \mu_1' = \frac{1}{b-a}\left[\frac{b^{1+1} - a^{1+1}}{1+1}\right]$$

$$= \frac{1}{b-a}\left[\frac{b^2 - a^2}{2}\right]$$

$$= \frac{1}{(b-a)}\left[\frac{(b-a).(b+a)}{2}\right] = \frac{b+a}{2}$$

$$\boxed{\therefore \text{Mean, } \mu_1' = \frac{b+a}{2}} \quad \text{... (4)}$$

Variance (μ_2)

It is known that variance $(\mu_2) = \mu'_2 - (\mu'_1)^2$ ... (5)

Now, substitute $r = 2$ in equation (3), the value of μ'_2 will be,

$$\mu'_2 = \frac{1}{b-a}\left[\frac{b^{2+1} - a^{2+1}}{2+1}\right] = \frac{1}{b-a}\left[\frac{b^3 - a^3}{3}\right]$$

$$= \frac{1}{(b-a)}\left[\frac{(b-a).(b^2+a^2+ab)}{3}\right]$$

$$= \frac{b^2+a^2+ab}{3}$$

$$\therefore \quad \mu'_2 = \frac{b^2+a^2+ab}{3}$$

Now substitute μ'_2 and μ'_1 in equation (5) the variance (μ_2) will be,

$$\mu_2 = \mu'_2 - (\mu'_1)^2$$

$$= \frac{b^2+a^2+ab}{3} - \left[\frac{b+a}{2}\right]^2$$

$$= \frac{b^2+a^2+ab}{3} - \left[\frac{b^2+a^2+2ab}{4}\right]$$

$$= \frac{4b^2+4a^2+4ab-3b^2-3a^2-6ab}{12}$$

$$= \frac{b^2+a^2-2ab}{12} = \frac{(b-a)^2}{12}$$

$$\boxed{\therefore \mu_2 = \frac{(b-a)^2}{12}}$$

3.1.2 Normal Distribution

Q12. (a) Write about normal distribution and its properties.

(b) Write notes on,

(i) Normal distribution curve

(ii) Standard normal curve.

Answer : Model Paper-I, Q6

(a) Normal Distribution

While binomial and Poisson distributions enable us to deal with the occurrence of distinct events, such as the number of defective items in a sample of a given size or the number of accidents occurring in a factory, the normal distribution is a mathematical distribution for dealing with quantities whose magnitude vary continuously.

Normal distribution is widely used in the analysis of the agricultural and the biological data.

Normal distribution was discovered by English mathematician, De-Moivre in the year 1733 for the first time. He obtained this distribution as a limiting case of the binomial distribution.

Later on Laplace and Gauss also derived normal distribution. This distribution has become the most important distribution in statistical analysis.

Any quantity whose variations depends on random causes, is distributed according to the normal law. In most of the biological and agricultural analysis values are shown in accordance with normal distribution curve.

Importance

- In case of large sample size, the normal distribution facilitates a good approximation of discrete distributions.
- Normal distribution is extremely useful in several biological analyses whose values are often distributed with respect to normal distribution. Increase in sample size tends the sample mean towards a normal distribution. In case if large size samples are drawn from not normally distributed population, the successive sample mean will form normal distribution.
- Large sample size decreases the chances of error and produce a smooth curve than small sample size. The following diagram is a normal distribution curve.

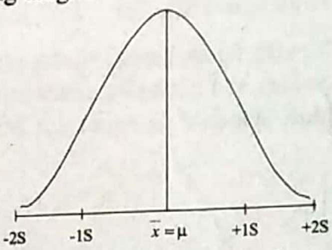


Figure: Normal Probability Curve

Properties

- The curve of normal distribution has a single peak; thus it is unimodal, and is bell-shaped with the highest point over the mean μ.
- It is symmetrical about a vertical line through mean.
- As it is symmetrical about mean, the mean, median and mode of the distribution also have the same value.
- The two tails of the curve extend indefinitely and never touch the horizontal axis.
- The curve height decreases on both the sides of the peak which occurs at the mean.
- The area on either side of the peak is equal to each other.

(b)

(i) Normal Distribution Curve

There exists several shapes of normal distribution curves differentiated only by the parameters such as means and the standard deviation of the sample.

Few shapes of the curve are shown below,

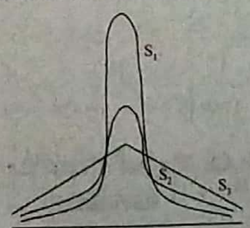


Figure: Normal Probability Curves with Different Standard Deviation and Same Mean Values

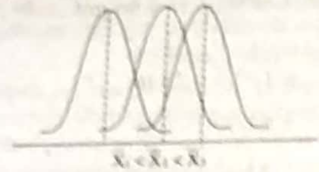


Figure: Normal Probability Curves with Different Mean and Same Standard Values

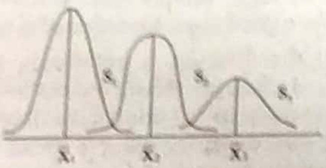


Figure: Normal Probability Curves with Different Mean and Same Standard Deviation Values

Area Under Normal Distribution

The distribution of area under normal curve with respect to mean and standard deviation is:

1. The following figure depicts that approximately 68% of all the values in a normally distributed population that lie within 1 standard deviation (S) from the mean.

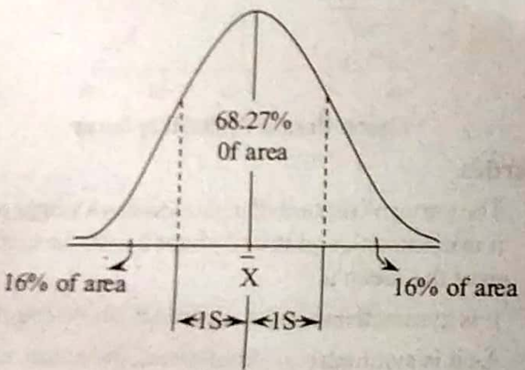


Figure: Area Under Normal Distribution with 1 Standard Deviation

$\overline{X} \pm 1S = 68.27\%$ of area

Area covered on both side of the mean = 34.14%

2. The following figure depicts that approximately 95.45% of all the values in a normally distributed population that lie within 2 standard deviation (S) from the mean.

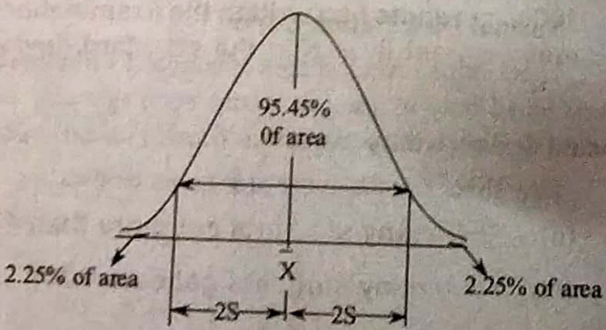


Figure: Area Under Normal Distribution with 2 Standard Deviations

$\overline{X} \pm 2S = 95.45\%$ of area

Area covered on both side of the mean = 47.73%

3. The following figure depicts that approximately 99.73% of all the values in a normally distributed population lie within 3 standard deviations (S) from the mean.

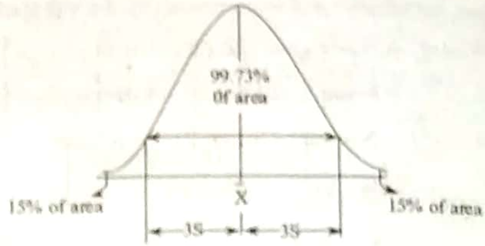


Figure: Area Under Normal Distribution with 3 Standard Deviations

$\overline{X} \pm 3S = 99.73\%$ of area

Area covered on both sides of the mean = 49.87%

(ii) Standard Normal Curve

The standard normal curve is defined as a curve with zero mean and unit standard deviation.

Any curve with mean and standard deviation can be transformed into a standard normal curve by changing the scale and origin from (x-scale) to (z-scale). Also the original parameters the mean ($\overline{X}$) and standard deviation (S) are transformed as zero mean and unit standard deviation.

Therefore, the method of changing the x-scale into z-scale is called z-transformation.

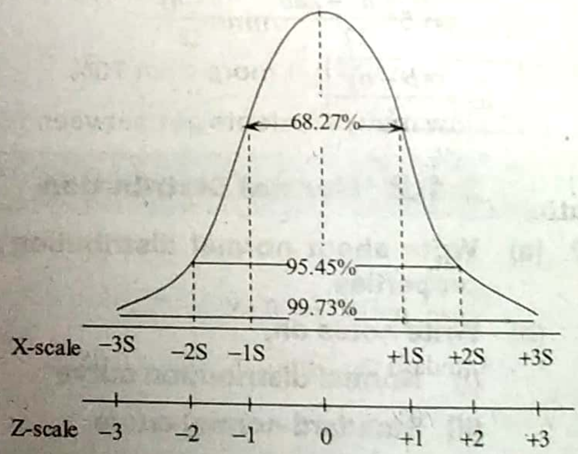


Figure: Normal Probability Curve Showing z Values and Standard Deviation (S)

Formula

Area under any normal curve is found from the table of standard normal probability distribution showing the area between the mean and any value of normally distributed random variable.

For given value of $\overline{X}$ (mean); S (Standard deviation); x (value of random variable, the standardized variate z is derived from the formula)

$$z = \frac{x - \overline{X}}{S}$$

The above formula is based on sample.

The formula for normal distribution based on population is given as,

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

Where, μ = Mean of the population.

σ = Standard deviation of the population.

Calculations for Normal Curve Probabilities

Few points while calculating the normal curve probabilities:

(a) The total area under the normal curve is 1. The area under the curve is divided into two equal parts by $z = 0$. Left hand side area and right hand side area to $z = 0$ is 0.5

(b) The area between the ordinate $z = 0$ and any other ordinate can be noted from the tables.

(c) z represents the number of standard deviations from the mean.

The range between the ordinates $z = 1$ and $z = 2$ shows that the normal variate z, is one standard deviation wide. However, there is a significant differences between the area covered for mean and one z value.

PROBLEMS

Q13. The marks obtained by 100 students is normally distributed with mean 68% and standard deviation 5%. Determine,

(i) How many get more than 70%

(ii) How many students get between 65% and 75%.

Solution :

Given that,

Mean, $\mu = 68\% = 0.68$

Standard deviation, $\sigma = 5\% = 0.05$

(i) When $x = 70\% = 0.70$

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

$$Z = \frac{0.70-0.68}{0.05}$$

$$= \frac{0.02}{0.05}$$

$$= 0.4$$

$$= z_1$$

$$P(x > 0.70) = P(z > z_1)$$

$$= 0.5 - A(z_1)$$

$$= 0.5 - A(0.4)$$

$$= 0.5 - 0.1554$$

$$= 0.3446$$

The number of student who got more than 70% out of 100 students,

$$= 100 \times 0.3446$$

$$= 34.46$$

$$\simeq 34$$

(ii) When $x = 65\% = 0.65$

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

$$= \frac{0.65-0.68}{0.05}$$

$$= \frac{-0.03}{0.05}$$

$$= -0.6$$

$$= z_1$$

When $x = 75\% = 0.75$

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

$$= \frac{0.75-0.68}{0.05}$$

$$= \frac{0.07}{0.05}$$

$$= 1.4$$

$$= z_2$$

$$P(0.65 < x < 0.75) = P(z_1 \le Z \le z_2)$$

$$= A(z_2) + A(z_1) \quad [\text{Since } z_1 < 0 \text{ and } z_2 > 0]$$

$$= A(1.4) + A(0.6)$$

$$= 0.4192 + 0.2258$$

$$= 0.645$$

The number of students who got marks between 65% and 75%. Out of 100 students,

$$= 100 \times 0.645$$

$$= 64.5$$

$$\simeq 65$$

Q14. 10000 students had written the examination the mean of test is 35 and the standard deviation is 5. Find,

(i) How many students marks lie between 25 and 40?

(ii) How many students get more than 40?

(iii) How many students get below 20?

Solution :

Given that,

Mean, $\bar{X} = 35$

Standard deviation, $S = 5$

(i) Marks lying between 25 and 40

$\Rightarrow \; x_1 = 25; \quad x_2 = 40$

We know that,

$$Z = \frac{x - \overline{X}}{S} \qquad \text{... (1)}$$

$$\Rightarrow \; Z_1 = \frac{25 - 35}{5}; \qquad Z_2 = \frac{40 - 35}{5}$$

$$= \frac{-10}{5} \qquad = \frac{5}{5}$$

$$= -2 \qquad = 1$$

$\therefore \; p(25 \le x \le 40) = p(-2 \le Z \le 1)$

= Area EABCDE

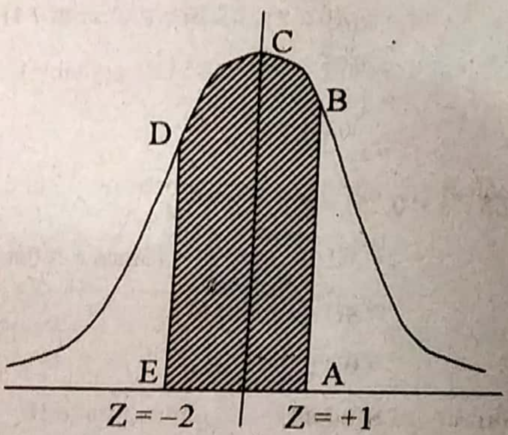
= Area EOCDE + Area OABCO

$= p(-2 \le Z \le 0) + p(0 \le Z \le 1)$

$= p(0 \le Z \le -2) + p(0 \le Z \le 1)$

$= 0.4772 + 0.3415$ [From table]

$= 0.185$



$\therefore$ The number of students whose marks lie between 25 and 40 = $10000 \times 0.8185 = 8185$

(ii) Marks more than 40

$x = 40$

$$Z = \frac{40 - 35}{5} \qquad \text{[From equation (1)]}$$

$$= \frac{5}{5} = 1$$

$\therefore \; p(x \ge 40) = p(Z \ge 1)$

= Area ABC

= 0.5 – Area OABDO

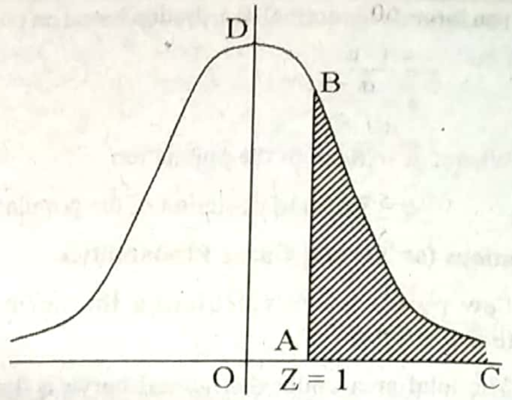
$= 0.5 - p(0 \le Z \le 1)$

$= 0.5 - 0.3413$ [From table]

$= 0.1587$

The number of student who get more than,

$40 = 10000 \times 0.1587 = 587$



(iii) Marks below 20

$x = 20$

$$z = \frac{20 - 35}{5}$$

$$= \frac{-15}{5}$$

$$= -3$$

$\therefore \; P(x < 20) = P(z < -3)$

= Area s

= 0.5 – Area of OABCO

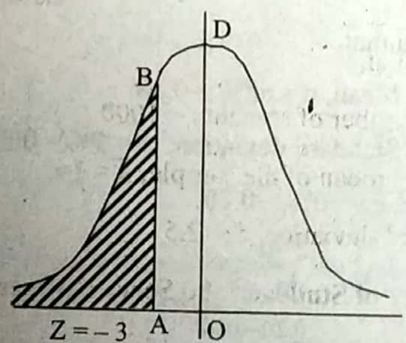
$= 0.5 - P(-3 \le z \le 0)$

$= 0.5 - P(0 \le z \le 3)$

$= 0.5 - 0.4987$ [From table]

$= 0.0013$

$\therefore$ The number of students who got less than 2-marks $= 0.0013 \times 10000 = 13$.



Q15. Students of a class were given an examination. Their marks were found to be normally distributed with mean 55 marks and standard deviation 5. Find the number of students who get the marks more than 60 if 500 students wrote the examination.

Solution :

Given that,

Mean, $\mu = 55$

Standard deviation, $\sigma = 5$

When, $x = 60$

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

$$z = \frac{60-55}{5}$$

$$z = \frac{5}{5}$$

$$z = 1$$

$$P(x > 60) = P(z > 1)$$

$$= 0.5 - A(z)$$

[$\because$ The tabulated value of $A(z) = 0.3413$]

$$= 0.5 - 0.3413$$

$$= 0.1587$$

The number of student who get more than 60 marks out of 500 students.

$$= 500 \times 0.1587$$

$$= 79.35$$

$$\simeq 79$$

Q16. In a sample of 1000 cases, the mean of a certain test is 14 and standard deviation is 2.5. Assuming the distribution to be normal find,

(i) How many students score between 12 and 15?

(ii) How many score above 18?

(iii) How many score below 8?

Solution : **Model Paper-II, Q6(b)**

Given that,

Total number of students = 1000

Average mean of the sample, $\overline{X} = 14$

Standard deviation, $S = 2.5$

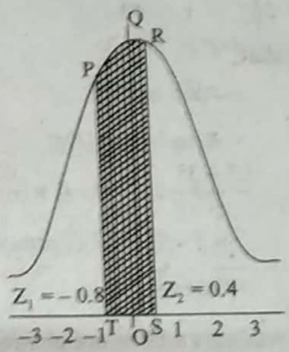
(i) Number of Students who Score Between 12 and 15

Let $x_1 = 12$ and $x_2 = 15$

We know that standardized variate $Z = \frac{x - \overline{X}}{S}$

$$\Rightarrow \quad Z_1 = \frac{x_1 - \overline{X}}{S} = \frac{12-14}{2.5} = \frac{-2}{2.5} = -0.8 \qquad Z_2 = \frac{x_2 - \overline{X}}{S} = \frac{15-14}{2.5} = \frac{1}{2.5} = 0.4$$

Area of the normal curve,



Therefore, probability of students who score between 12 and 15 is given by,

$$p(12 \le x \le 15) = p(12 \le z \le 15)$$

$$= \text{Area of } PQRST$$

$$= \text{Area of } PQOT + \text{Area of } QROS$$

$$= p(-0.8 \le z \le 0) + p(0 \le z \le 0.4)$$

$$= p(0 \le z \le -0.8) + p(0 \le z \le 0.4)$$

$$= 0.2881 + 0.1554 \text{ (From table)}$$

$$= 0.4435$$

$\therefore$ Number of students who score between 12 and 15

$$= 0.4435 \times 1000$$

$$= 0.44 \times 1000$$

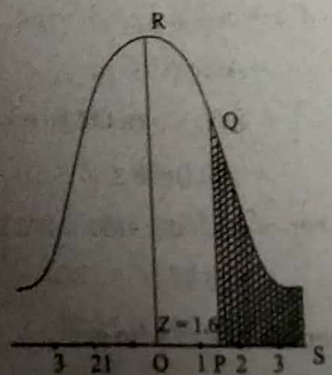
$$= 440$$

(ii) Number of Students who Score Above 18

$$x = 18$$

$$\Rightarrow \quad z = \frac{x - \overline{X}}{S} = \frac{18-14}{2.5} = \frac{4}{2.5} = 1.6$$

Area of the normal curve,



Therefore, probability of students who score above 18 is given by,

$$P(x \geq 18) = P(z \geq 1.6)$$
$$= \text{Area of } PQS$$
$$= \text{Area of } OPSQR - \text{Area of } OPQR$$
$$= 0.5 - P(0 \leq z \leq 1.6)$$
$$= 0.5 - 0.4452 \text{ (From table)}$$
$$= 0.0548$$

$\therefore$ Number of students who score above 18

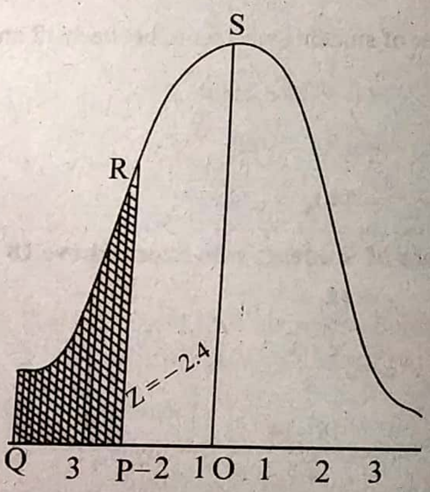
$$= 0.0548 \times 1000$$
$$= 0.055 \times 1000$$
$$= 55$$

(iii) Number of Students who Score Below 8

$x = 8$

$$\Rightarrow \quad z = \frac{x - \overline{X}}{S} = \frac{8-14}{2.5}$$
$$= \frac{-6}{2.5} = -2.4$$

Area of the normal curve,



Therefore, probability of students who score below 8 is given by,

$$P(x \leq 8) = P(z \leq 8)$$
$$= \text{Area of } PQR$$
$$= \text{Area of } OPQRS - \text{Area of } OPRS$$
$$= 0.5 - P(0 \leq z \leq -2.4)$$
$$= 0.5 - 0.4892$$
$$= 0.0108$$

$\therefore$ Number of students who score below 8

$$= 0.0108 \times 1000$$
$$= 0.011 \times 1000$$
$$= 11.$$

3.1.3 Areas Under the Normal Curve

Q17. Write about the area under the normal curve. Also, prove that area under normal curve is unity.

Answer : Model Paper-III, Q6(a)

Area Under Normal Curve

Area under any normal curve is found from the table of standard normal probability distribution showing the area between the mean and any value of the normally distributed random variable.

Since for different values of μ and σ we have different normal curves. Hence it is not possible to draw the normal curves for various values of μ and σ. Thus, the normal curve is transformed into a standardized normal curve.

'x' is transformed into 'z' which is known as 'standard normal variate'.

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

The area under the normal curve between the ordinates $x = $ 'x_1' and $x = $ 'x_2' gives the probability that the normal variate lies between 'x_1' and 'x_2' as shown in figure.

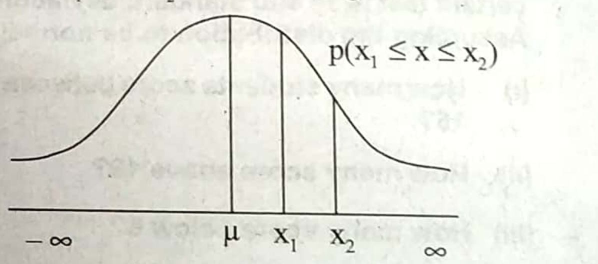


Figure: Probability that Normal Variate Lies between 'x_1' and 'x_2'

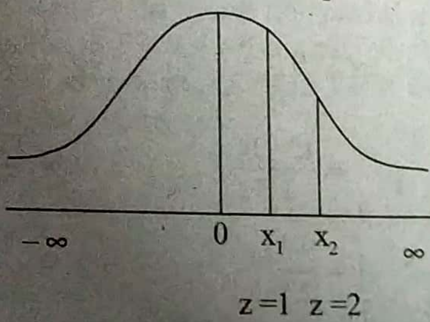
The probability can be evaluated as standardized 'z' as follows,

Substitute, $z = \dfrac{x - \mu}{\sigma}$ $\quad [\mu = 0, \sigma = 1]$

When, $x = x_1$, $z = \dfrac{x_1 - \mu}{\sigma} = z_1$ (suppose)

When, $x = x_2$, $z = \dfrac{x_2 - \mu}{\sigma} = z_2$

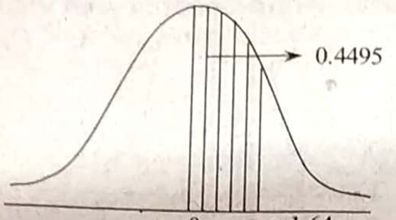
$\therefore \quad p(x_1 \leq x \leq x_2) = p(z_1 \leq z \leq z_2)$



Figure

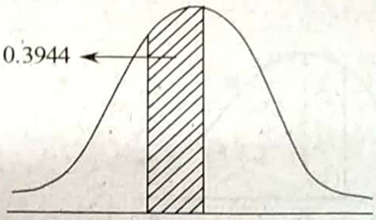
Examples

The area under the normal curve for $z = 1.64$



Figure

The area under the normal curve for $z = -1.25$



Figure

From normal distribution table, area values.

If $z = 1.64$, Area = 0.4495

$z = 1.25$, Area = 0.3944

are taken and shade the region obtained.

Area Under Normal Curve is Unity

The probability density function of the normal distribution is,

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

Therefore, area under the normal curve is equal to,

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

Put $\frac{x-\mu}{\sigma\sqrt{2}} = t$

$$dx = \sigma\sqrt{2}dt$$

$$\therefore \int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2} \sigma\sqrt{2}dt$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt$$

$$= \frac{1}{\sqrt{\pi}} \sqrt{\pi} \quad \left(\because \int_{-\infty}^{\infty} e^{-x^2} dx = 1\right)$$

$$= 1$$

Hence proved.

Q18. Write the procedure of converting the standard normal curve into standard normal variate.

Answer :

Standard normal curve can be converted into standard normal variate by changing the values of mean and standard deviation. This makes the calculations more simpler and also helps to draw normal curves in all cases.

In the original form, the mean 'μ' and standard deviation 'σ' are considered whereas in the standardized form μ is taken as '0' and 'σ' as 1.

Thus,

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

Where,

'x' = Random variable

μ = Mean

σ = Standard deviation

Thus, the transformed 'z' is called as 'standard normal variate'.

The curve is 'bell-shaped', 'symmetrical' and 'asymptotic' (never touches the base).

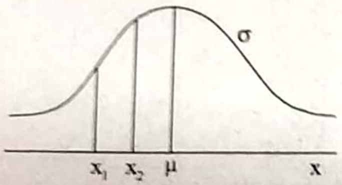


Figure: Original Normal Distribution

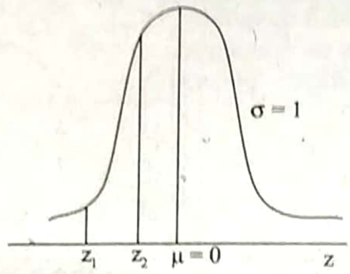


Figure: Standardized or Transformed Normal Distribution

The two parameters are 'μ' and 'σ'.

'μ' is '*location parameter*', it decides the location of the curve.

'σ' is '*shape parameter*', it decides the shape of the curve.

PROBLEMS

Q19. The marks obtained in mathematics by 1000 students is normally distributed with mean 78% and standard deviation 11%. Determine

(i) How many students got marks above 90%

(ii) What was the highest mark obtained by the lowest 10% of the student

(iii) Within what limits did the middle of 90% of the students lie.

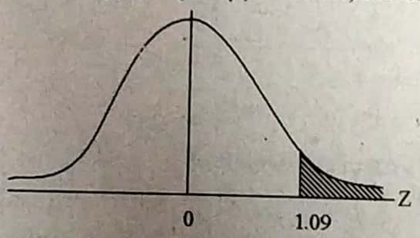
Solution :

$$Z = \frac{X - \bar{x}}{\sigma} = \frac{X - 0.78}{0.11}$$

(i) For $X = 0.9$, $Z = \dfrac{0.9 - 0.78}{0.11} = 1.09$

$P(x > 0.9) = P(Z > 1.09) = 0.5 - 0.3621$

$= 0.1379$ [$\because P(X > \mu) = P(X < \mu) = 0.5$ i.e., the area to the right and left of the ordinate is 0.5]



Number of students with marks above 90% = $1000 \times P(X > 0.9) = 1000 \times 0.1379 \approx 138$.

(ii) The lowest 10% students constitute 0.1 area ($< \frac{1}{2}$) of extreme left tail. So Z_1 must be negative.

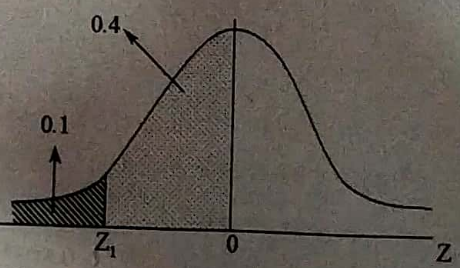
From table $0.4 = 0.5 - 0.1 = 0.5 -$ Area 0.1 from 0 to Z_1

So, $Z_1 = -1.28$

Thus,

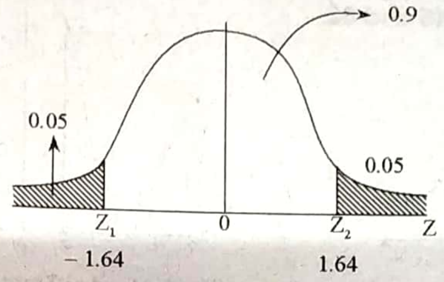
$$-1.28 = \frac{X - 0.78}{0.11}$$

$\therefore$ $X = 0.6392$



Thus, the highest mark obtained by the lowest 10% of students is $63.92 \cong 64\%$

(iii) Middle 90% correspond to 0.9 area, leaving 0.05 area on both sides. Then the corresponding Z's are $\pm$ 1.64



$$1.64 = Z_2 = \frac{X_2 - 0.78}{0.11} \Rightarrow X_2 = 96.04\ \%$$

$$-1.64 = Z_1 = \frac{X_1 - 0.78}{0.11} \Rightarrow X_1 = 59.96\ \%$$

Thus, the middle 90% have marks in between 60 to 90.

Q20. Given a standard normal distribution, find the area under the curve that lies,

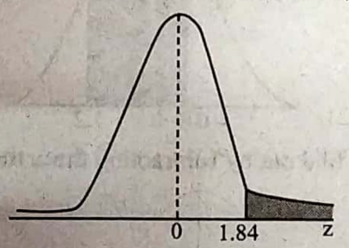
(a) To the right of z = 1.84 and

(b) Between z = – 1.97 and z = 0.86.

Solution :

(a) **To the right of z = 1.84**

Initially, draw the curve that lies to the right of $z = 1.84$ as shown below,



The area under the curve that lies to the left of $z = 1.84$ is equal to,

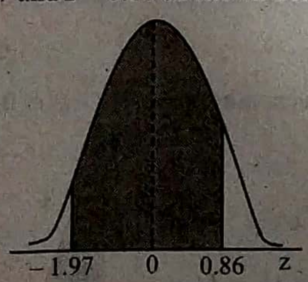
$= 1 -$ Area under the normal curve table to the left of $z = 1.84$

$= 1 - 0.96471$ $\left(\because \text{The corresponding z value for 1.84 is 0.9671 from the table Area under the normal curve}\right)$

$= 0.0329$

(b) **Between z = – 1.97 and z = 0.86**

Draw the curve that lies between $z = -1.97$ and $z = 0.86$ as shown below,



Area under the curve that lies between $z = -1.97$ and $z = 0.86$ (or) Area under the shaded region

$=$ (Area to the left of $z = 0.86$) – (Area to the left of $z = -1.97$)

$= 0.8051 - 0.0244$ $\left(\because \text{The corresponding value of z for 0.86 and } -1.97 \text{ is 0.8051 and 0.0244 respectively}\right)$

$= 0.7807$

Q21. Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 62.

Solution :

Given that,

Mean, $\mu = 50$

Variance, $\sigma = 10$

$X_1 = 45$

$X_2 = 62$

To find : $P(45 < X < 62)$

The respective values of z for $x_1 = 45$ and $x_2 = 62$ are,

$$z_1 = \frac{X_1 - \mu}{\sigma} \qquad z_2 = \frac{X_2 - \mu}{\sigma}$$

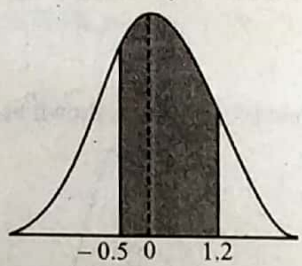
$$= \frac{45 - 50}{10} \qquad = \frac{62 - 50}{10}$$

$$= -\frac{5}{10} \qquad = \frac{12}{10}$$

$$= -0.5 \qquad = 1.2$$

Thus, $P(45 < X < 62) = P(-0.5 < Z < 1.2)$

The area that lies between -0.5 and 1.2 is shown below by the shaded region in the curve,



The area between -0.5 and 1.2 can be find out by subtracting the area to the left of the ordinate $z = -0.5$ from the complete area to the left of $z = 1.2$.

$$P(45 < X < 62) = P(-0.5 < z < 1.2)$$

$$= P(z < 1.2) - P(z < -0.5) \qquad \left(\because \text{The value of z for 1.2 and } -0.5 \text{ is 0.8849 and 0.3685 respectively}\right)$$

$$= 0.8849 - 0.3085$$

$$\therefore \quad P(45 < X < 62) = 0.5764$$

Q22. Given a standard normal distribution, find the value of k such that

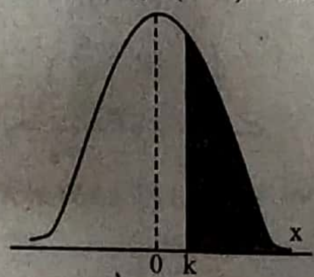
(a) $P(z > k) = 0.3015$ and

(b) $P(k < z < -0.18) = 0.4197$.

Solution :

(a) $P(z > k) = 0.3015$

Initially, draw the curve according the given value (i.e., $P(z > k) = 0.3015$) as shown below,

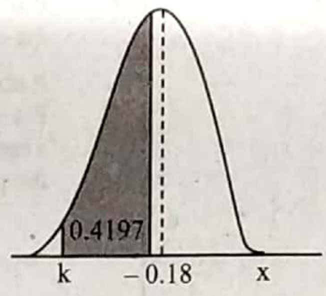


From the above figure, the k value leaves an area of 0.3015 to the right. Likewise, the k value leaves an area of 0.6985 to the left. Since, the area under the normal curve is unity (i.e, 1), $1 - 0.3015 = 0.6985$.

Therefore, the value of $k = 0.52$, which is obtained from the table Areas under the Normal curve for the z-value 0.6985.

(b) $P(k < z < -0.18) = 0.4197$

Initially, draw the curve according to the given condition as shown below,



The corresponding z value for – 0.18 is a 0.4286.

Area to the left of k = (Total area to the left of – 0.18) – (Area between k and – 0.18)

$= 0.4286 - 0.4197$

$= 0.0089$

Therefore, the value of $k = -2.37$ which is obtained from the Table Areas under the Normal curve for the z-value 0.0089.

Q23. Given that X has a normal distribution with $\mu = 300$ and $\sigma = 50$, find the probability that X assumes a value greater than 362.

Solution :

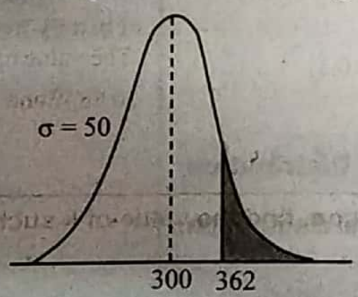
Given that,

Mean $\mu = 300$

Variance $\sigma = 50$

To find $P(X > 362)$,

initially draw the curve of the normal probability distribution according to the given values as shown below,



Then, calculate the area under the normal curve to the right of $x = 362$ by transferring $x = 362$ to its respective 'z' value,

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

$$= \frac{362 - 300}{50}$$

$$= \frac{62}{50}$$

$$z = 1.24$$

Now, the area to the left of z can be obtained by subtracting Area under the Normal curve table from 1.

$P(X > 362) = P(z > 1.24)$

$= 1 - P(z < 1.24)$

$= 1 - 0.8925$ $\left(\because \text{The value of 1.24 in the Area under the normal curve table is 0.8925}\right)$

$= 0.1075$

Q24. Given a normal distribution with μ = 40 and σ = 6, find the value of x that has

(a) 45% of the area to the left and

(b) 14% of the area to the right.

Solution :

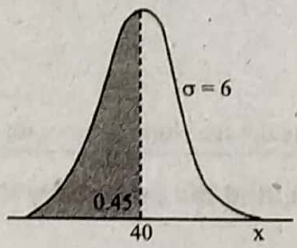
Given that,

Mean, $\mu = 40$

Variance, $\sigma = 6$

(a) 45% of the Area to the Left

Initially, draw the curve and shade 45% (i.e., 0.45) of the area to the left as shown below,



(b) 14% of the Area to the Right

As the shaded region occupies 0.14% the area to the right. The left over area to the left of the shaded region will be $1 - 0.14 = 0.86$.

The corresponding z value for 0.86 is $z = 1.08$ which is taken from the table area under the normal curve.

Therefore, $x = \sigma z + \mu$

$= 6 \times 1.08 + 40$

$= 6.48 + 40$

$= 46.48$

3.1.4 Applications of the Normal Distribution

Q25. State the applications of the normal distribution.

Answer :

The applications of the normal distribution are as follows,

1. Normal distribution is useful for approximating practical distributions like Poisson, binomial, hypergeometric. In addition to this, it is also useful for the large sample distributions like chi-square, student's t, Snedecor's F.

2. It is useful to bring a variable to a normal state by performing simple transformations on that variable. For instance, if a variable 'X' is of skewed distribution then the square root of X i.e., $\sqrt{X}$ may result is a normal distribution form.

3. It is useful for assigning control limits to the large applications in the field of statistical quality control.

4. It is useful in describing large sample statistic distributions that result in normality.

5. It is useful to guide the parent populations from which the small sample tests like t, F, χ^2 tests are derived.

6. It acts as a basis for the complete theory of large sample. For instance, if $X \sim N(\mu, \sigma^2)$ is a normal distribution of a random variable 'X' with mean μ and variance σ^2 then the probability,

$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$

$= P(-3 < Z < 3)$

$= 0.9973$

Thus, $P(|Z| > 3) = 1 - P(|Z| \leq 3)$

$= 1 - 0.9973$

$= 0.0027$

PROBLEMS

Q26. A certain type of storage battery lasts on average of 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

Solution :

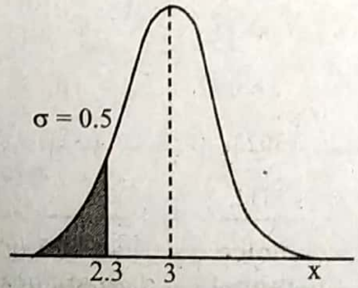
Given that,

Mean, $\mu = 3$

Variance, $\sigma = 0.5$

$x = 2.3$ years

Initially, draw the curve that shows the given distribution of battery lives and the desired area as shown below,



To find $P(X < 2.3)$ evaluate the area under the normal curve to the left of 2.3.

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

$$= \frac{2.3 - 3}{0.5}$$

$$= \frac{-0.7}{0.5}$$

$$= \frac{-7}{5}$$

$$= -1.4$$

Now, $P(X < 2.3) = P(z < -1.4)$

$= 0.0808$ ($\because$ The corresponding z value of -1.4 is 0.0808)

Q27. An electrical firm manufactures light bulbs that have a life, before burn out, that is normally distributed with mean equal to 800 hours a standard deviation of 40 hours. Find the probability that a bulb burns between 778 and 834 hours.

Solution :

Given that,

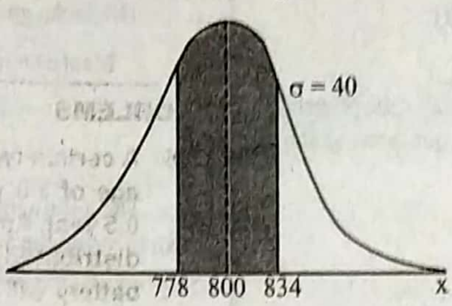
Mean, $\mu = 800$

Variance, $\sigma = 40$

$x_1 = 778$

$x_2 = 834.$

The distribution of the light bulb life is shown in the below figure,



The corresponding z-values for $x_1 = 778$ and $x_2 = 834$ are,

$$z_1 = \frac{x_1 - \mu}{\sigma} \qquad z_1 = \frac{x_2 - \mu}{\sigma}$$

$$= \frac{778 - 800}{40} \qquad = \frac{834 - 800}{40}$$

$$= \frac{-22}{40} \qquad = \frac{34}{40}$$

$$z_1 = -0.55 \qquad z_2 = 0.85$$

Therefore, $P(778 < X < 834) = P(-0.55 < Z < 0.85)$

$= P(z < 0.85) - P(z > -0.55)$

$= 0.8023 - 0.2912$ ($\because$ The corresponding z values for 0.85 and -0.55 are 0.8023 and 0.2912 respectively)

$= 0.5111.$

Q28. A multiple-choice quiz has 200 questions, each with 4 possible answers of which only 1 is correct. What is the probability that sheer guesswork yields from 25 to 30 correct answers for the 80 of the 200 problems about which the student has no knowledge?

Solution : **Model Paper-III, Q6(b)**

The probability of guessing a correct answer for every 80 questions, $P = \frac{1}{4}$

Let 'X' denote the number of correct answers resulting from guesswork,

$$P(25 \le X \le 30) = \sum_{x=25}^{30} b\left(x; 80, \frac{1}{4}\right)$$

Mean of the normal curve, $\mu = np$

$= 80 \times \frac{1}{4}$

$= 20$

Variance of the normal curve,

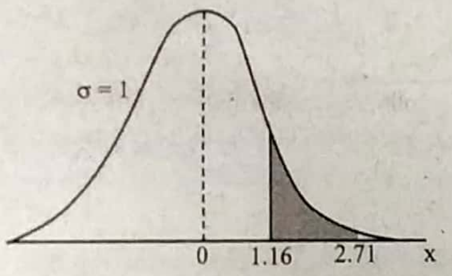
$$\sigma = \sqrt{npq}$$
$$= \sqrt{80 \times \frac{1}{4} \times \frac{3}{4}}$$
$$= \sqrt{20 \times \frac{3}{4}}$$
$$= \sqrt{15}$$
$$= 3.873$$

Now, the corresponding z values for $x_1 = 24.5$ and 30.5 can be computed as,

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{24.5 - 20}{3.873} = \frac{4.5}{3.873}, \quad z_1 = 1.16$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{30.5 - 20}{3.873} = \frac{10.5}{3.873}, \quad z_2 = 2.71$$

The shaded region in the curve shows that the probability of guessing correct answers from 25 to 30 questions as shown below,



Using the Table Area under Normal curve, we get

$$P(25 \le X \le 30) = \sum_{x=25}^{30} b(x; 80, 0.25)$$
$$\approx P(1.16 < z < 2.71)$$
$$= P(z < 2.71) - P(z < 1.16)$$
$$= 0.9966 - 0.8770$$
$$= 0.1196$$

3.1.5 Normal Approximation to the Binomial

Q29. Explain in detail about Normal Approximation to the Binomial.

Answer : Model Paper-I, Q7(a)

If n(Number of cases) is large and if neither probability of success 'p' nor probability of failure 'q' is very close to zero, the binomial distribution can be closely approximated by a normal distribution with standardized random variable.

In case of not large 'n', Binomial Distribution (B.D) can be approximated by Normal Distribution (N.D).

$$\therefore \quad z = \frac{X - np}{\sqrt{npq}}$$

Where,

np = mean of B.D

$\sqrt{npq}$ = Standard deviation of B.D

To explain the normal approximation to the binomial distribution, initially draw the histogram by using the binomial distribution formula $b(x; n, p)$ and then superimpose the normal curve on it that has the mean and variance identical to the binomial variable (X). Thus, the normal curve can be draw by calculating mean and variance which is given by,

$$\text{Mean, } \mu = np$$

$$\text{Variance, } \sigma^2 = npq$$

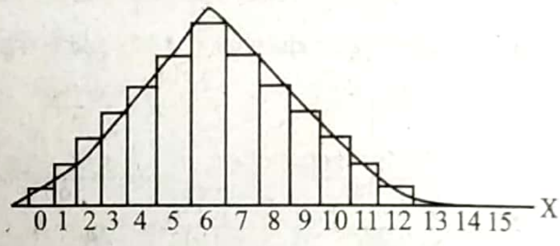
Consider an example in which the value of the binomial distribution $b(x; n, p)$ is taken as $b(x; 15, 0.4)$. The very first step is to draw the histogram based on the given value of binomial distribution and then calculate the mean and variance of the normal curve for superimposing it on the histogram.

The mean and variance are computed as,

$$\text{Man}(\mu) = np = 15 \times 0.4 = 6$$

$$\text{Variance}(\sigma^2) = npq = 15 \times 0.4 \times 0.6 = 3.6$$

The following figure illustrates that the histogram is drawn by using the $b(x; 15, 0.4)$ and the normal curve is drawn by computing the mean and variance,



Binomial random variable (X) assumes that the given value of X is equal to the area of the rectangular bar of the histogram whose base is centered at x. For instance, the exact probability can be find out by taking the value of $X = 4$ which is equal to the rectangular bar with center at $X = 4$. The exact probability of the binomial random variable X for $x = 4$ can be computed by,

$$P(X = 4) = b(4; 15, 0.4)$$
$$= 0.1268$$

($\because$ The corresponding value of $b(4; 15, 0.4) = 0.1268$ taken from the Table Binomial probability sums)

Thus, the resulted value of X (i.e., 0.1268) is approximately equal to the area of the shaded region under the normal curve that lies in between the ordinates $x_1 = 3.5$ and $x_2 = 4.5$

Now, computing the values of z so as to check whether the resulted value of x is identical to z-values or not.

As we known that $z = \dfrac{X - np}{\sqrt{npq}}$

$$z_1 = \frac{3.5 - 6}{1.897} = -1.32 \qquad z_2 = \frac{4.5 - 6}{1.897} = -0.79$$

If binomial random variable and standard normal variable are X and z respectively, then

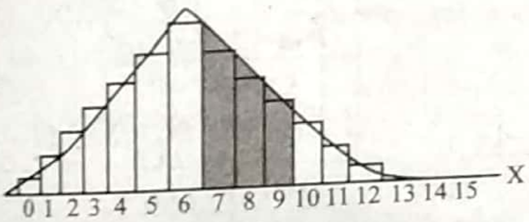
$$P(X=4) = b(4;15,0.4)$$
$$\approx P(-1.32 < z < -0.79)$$
$$= P(z < -0.79) - P(z < -1.32)$$
$$= 0.2148 - 0.0934$$
$$= 0.1214$$

Thus, the above value(i.e., 0.1214) is very close or approximately equal to the value 0.1268 which is obtained by using Table Binomial Probability sums.

Now, consider that the probability of X assumes a value from 7 to 9. Then, the exact probability is given by,

$$P(7 \le X \le 9) = \sum_{x=0}^{9} b(x;15,0.4) - \sum_{x=0}^{6} b(x;15,0.4)$$
$$= 0.9662 - 0.6098$$
$$= 0.3564$$

Now, the area of the shaded region with bases centered at $x = 7, 8$ and 9 is to be find out. This can be done by calculating the area of the shaded region under the curve that lies in between the ordinates $x_1 = 6.5$ and $x_2 = 9.5$ as shown in figure below,



As we know that $z = \dfrac{X - np}{\sqrt{npq}}$

$$z_1 = \frac{6.5-6}{1.897} = 0.26 \qquad z_2 = \frac{9.5-6}{1.897} = 1.85$$

Now,

$$P(7 \le X \le 9) \approx P(0.26 < z < 1.85)$$
$$= P(z < 1.85) - P(z - 0.26)$$
$$= 0.9678 - 0.6026$$
$$= 0.3652$$

Therefore, the resulted value (i.e., 0.3652) of the normal curve approximation is very close to the value of 0.3564.

3.1.6 Gamma and Exponential Distributions

Q30. Discuss about Gamma Distribution with its applications.

Answer :

Gamma Distribution

The name "Gamma Distribution" is derived from the gamma function and is defined as,

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx; \ \alpha > 0$$

Properties

Some of the properties of gamma function are as follows,

1. $\Gamma(n) = (n-1)(n-2)\ldots(1)\Gamma(1)$

   Apply integration by parts to the gamma function wherein $u = x^{\alpha-1}$ and $v = e^{-x}dx$, then

   $$\Gamma(\alpha) = -e^{-x}x^{\alpha-1}\Big|_0^{\infty} + \int_0^{\infty} e^{-x}(\alpha-1)x^{\alpha-2}dx$$
   $$= (\alpha-1)\int_0^{\infty} x^{\alpha-2}e^{-x}dx$$

   If $\alpha > 1$, then it generates the recursion formula as,

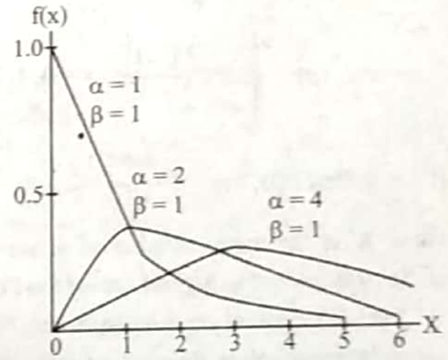
   $$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$$

2. For a positive integer n, $\Gamma(n) = (n-1)!$
3. $\Gamma(1) = 1$
4. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

If 'X' is a continuous random variable, then it will have a gamma distribution with parameters α and β and and its density function is given by,

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta\alpha\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0,\ \alpha < 0, \beta > 0 \\ 0 & \text{, Otherwise} \end{cases}$$

The following graphs illustrates different gamma distributions with certain particular values of the parameters α and β,



Applications

1. It is used for calculating the amount of rainfall which is stored in a reservoir.
2. It is used in the processing of loan and aggregate insurance.
3. It is used in minimizing the errors present in multi level poisson regression model.

Q31. Define Exponential Distribution. Also, explain its application.

Answer :

Exponential Distribution

The exponential distribution can be defined as a special gamma distribution for which the value of $\alpha = 1$. If 'X' is a continuous random variable then it will have exponential distribution with parameter β and its density function is given by,

$$f(x;\beta) = \begin{cases} \dfrac{1}{\beta} e^{-\frac{x}{\beta}}, & x > 0, \beta > 0 \\ 0 & \text{, otherwise} \end{cases}$$

Applications

1. It is used for calculating survival time in biomedical equipments.
2. It is used for calculating computer response time.
3. The time required to accomplish the needs of a customer at railway booking counter, banks, petrol pump, ration shops etc.
4. The time during which the electronic components are operated without any breakdown/failure.
5. The time required for any service facility among two successive arrivals.

Q32. Write short note on Relationship to the Poisson Process.

Answer :

A random variable 'x' is said to have a Poisson distribution if it assumes only the number of arrivals (events) within the time interval 't' and its Probability Mass Function (PMF) is given by,

$$P(x, \lambda t) = \begin{cases} = \dfrac{e^{-\lambda t}(\lambda t)^x}{x!}; & x = 0, 1, 2, 3, \ldots n \\ = 0 \quad ; & \text{Otherwise} \end{cases}$$

If $x = 0$ then $P(0; \lambda t) = \dfrac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$

Here 'λ' is the mean number of arrivals (events) per unit time. So, it is known as arrival rate of the Poisson process. When 'λ' does not depend on time then, the Poisson process is known as homogeneous Poisson process. And, when the 'λ' depends on time (i.e., if it is a function of time) then, it is known as nonhomogeneous Poisson process and it is represented as $\lambda(t)$.

Let 'X' represents the initial Poisson event time. The probability that the time length until the emitical event will exceed x is identical to the probability that no events occurs in x. This is given by $e^{-\lambda x}$ as shown below,

$$P(X > x) = e^{-\lambda x}$$

Therefore, the cumulative distribution function for X is,

$$P(0 \leq X \leq x) = 1 - e^{-\lambda x}$$

Now, differentiating cumulative distribution function in order to obtain density function of the exponential distribution with $\lambda = \dfrac{1}{\beta}$ is given by,

$$f(x) = \lambda e^{-\lambda x}$$

Q33. State and prove the memory less property of exponential distribution.

Answer :

Statement

If 'X' is a random variable representing the lifetime of a product of a system that follows exponential distribution, then conditional probability density function after 'a' time can be given as,

$$P\{X \leq x + a / x > a\} = P\{X \leq x\}$$

Proof

The standard PDF of exponential distribution for X random variable with λ as mean can be given as,

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \ ; & \text{for } \lambda \geq 0 \text{ and } 0 < x < \infty \\ 0 \ ; & \text{otherwise} \end{cases}$$

Conditional probability after completing time 'a' is given as,

$$P(x \leq x + a / X > a\} = \frac{P\{(X \leq x + a) \cap (X > a)\}}{P(X > a)}$$

$$= \frac{[P(a \leq X \leq x + a) \cap P(a < X < \infty)]}{P(a < X < \infty)}$$

$$= \frac{P(a < X \leq x + a)}{1 - P(X < a)}$$

$$= \frac{P(X \leq x + a) - P(X < a)}{1 - F(a)}$$

$$= \frac{F(x + a) - F(a)}{1 - F(a)}$$

P.D.F of exponential distribution is given as,

$$F(x) = 1 - e^{-\lambda}x$$

$$F(a) = 1 - e^{-\lambda}a$$

$$F(x + a) = 1 - e^{-\lambda}(x + a)$$

$$P\{X \leq x + a\ X > a\} = \frac{(1 - e^{-\lambda(x+a)}) - (1 - e^{-\lambda a})}{1 - (1 - e^{-\lambda a})}$$

$$= \frac{1 - e^{-\lambda x} . e^{-\lambda a} - 1 + e^{-\lambda a}}{e^{-\lambda a}}$$

$$= \frac{e^{-\lambda a}(1 - e^{-\lambda x})}{e^{-\lambda a}}$$

$$= 1 - e^{-\lambda x}$$

$$= P(X \leq x) \quad \ldots (1)$$

Hence, it can be concluded that the conditional probability life time product of the system does not rely on the past information of lifetime. This property of exponential distribution is known as memoryless property.

Consider a variable,

$y = x + a$, then $x = y - a$ then equation (1) can be written as,

$$P\{X \le y/X > a\} = 1 - e^{-\lambda(y-a)}$$

$$= F(x / X > a)$$

The condition, PDF, $f(x / x > a)$

$$= \frac{d}{dx} F(x / x > a)$$

$$= \frac{d}{dx}(1 - e^{-\lambda(y-a)})$$

$$= 0 - e^{-\lambda(y-a)} . (-\lambda)$$

$$= \lambda e^{-\lambda(y-a)}$$

PROBLEMS

Q34. Suppose that a system contains a certain type of component whose time, in years to failure is given by T. The random variable T is modeled nicely by the exponential distribution with mean time to failure $\beta = 5$. If 5 out of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years?

Solution :

Let P be the probability of the working component after 8 years which is given by,

$$P(T > 8) = \frac{1}{5}\int_8^\infty e^{-\frac{t}{5}} dt$$

$$= \frac{1}{5}\left[\frac{e^{-\frac{t}{5}}}{-\frac{1}{5}}\right]_8^\infty$$

$$= \frac{1}{5} \times 5\left[e^{-\frac{t}{5}}\right]_8^\infty$$

$$= \left[e^{-\frac{t}{5}}\right]_8^\infty$$

$$= \left[e^{-\frac{\infty}{5}} - e^{-\frac{8}{5}}\right]$$

$$= \left[0 - e^{-\frac{8}{5}}\right]$$

$$= e^{-\frac{8}{5}} \approx 0.2$$

The number of components that are working after 8 years be 'X'. By using binomial distribution we get,

$$P(X \ge 2) = \sum_{x=2}^{5} b(x; 5, 0.2)$$

$$= 1 - \sum_{x=0}^{5} b(x; 5, 0.2)$$

$= 1 - 0.7373$ ($\because$ The corresponding value of the $b(x; 5, 0.2)$ in the Binomial Distribution Table is 0.7373)

$$P(X \ge 2) = 0.2627$$

Q35. Suppose that telephone calls arriving at a particular switchboard follow a Poisson process with an average of 5 calls coming per minute. What is the probability that up to a minute will elapse by the time 2 calls have come into the switchboard?

Solution : **Model Paper-I, Q7(b)**

Given that,

$$\alpha = 2, \beta = \frac{1}{5}$$

Let 'X' denote the time in minutes before two calls come into the switchboard.

The required probability can be calculated as,

$$P(X \le 1) = \int_0^1 \frac{1}{\beta^2} x e^{-\frac{x}{\beta}} dx$$

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

$$= 25\int_0^1 x e^{-5x} dx$$

Apply u - substitution where $u = -5x$

$$= 25\int_0^1 \frac{e^u u}{25} du$$

$$= 25 \times \frac{1}{25}\int_0^1 e^u u du$$

$$= \int_0^1 e^u u du$$

Apply Integration by parts: $u = 4$, $v' = e^4$

$$= e^4 u - \int e^4 du$$

$$= e^4 u - e^4 \qquad \left(\because \int e^4 du = e^u\right)$$

Substitute back $u = -5x$

$$= \int_0^1 e^{-5x}(-5x) - e^{-5x}$$

$$= \int_0^1 -5e^{-5x}x - e^{-5x}$$

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

$$= (-5e^{-5(1)}(1) - e^{-5(1)}) - (-5e^{-5(0)}(0) - e^{-5(0)})$$

$$= (-5e^{-5} - e^{-5}) - (0 - e^0)$$

$$= e^{-5}(-5 - 1) + e^0$$

$$= -6(e^{-5}) + 1 \qquad (\because e^0 = 1)$$

$$= \frac{-6}{e^5} + 1$$

$$= -0.040 + 1$$

$$= 0.959.$$

Q36. In a biomedical study with rats, a dose-response investigation is used to determine the effect of the dose of a toxicant on their survival time. The toxicant is one that is frequently discharged into the atmosphere from jet fuel. For a certain dose of the toxicant, the study determines that the survival time in weeks, has a gamma distribution with $\alpha = 5$ and $\beta = 10$. What is the probability that a rat survives no longer than 60 weeks?

Solution :

Given that,

$\alpha = 5$

$\beta = 10$

Let 'X' is a random variable which represents the survival time of rats.

The required probability can be computed as,

$$P(X \leq 60) = \frac{1}{\beta^5} \int_0^{60} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{(\Gamma(5))} dx \quad \text{... (1)}$$

The above equation can be computed by using incomplete gamma function.

$$f(x; \alpha) = \int_0^x \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy$$

Let $y = \frac{x}{\beta}$ then $x = \beta y$. The equation (1) changes to,

$$P(X \leq 60) = \int_0^6 \frac{y^4 e^{-y}}{\Gamma(5)} dy$$

The above equation denotes $F(x ; \alpha)$ as $F(6 ; 5)$ function. Thus, the probability that the rat survives not more than 60 days is given by,

$P(X \leq 60) = F(6 ; 5) = 0.715$ ($\because$ The corresponding value of $F(6 ; 5)$ in the Incomplete Gamma table is 0.715)

3.2 FUNDAMENTAL SAMPLING DISTRIBUTIONS

3.2.1 Random Sampling

Q37. Discuss in brief about the following,

(a) Population

(b) Sample

(c) Sample unit

(d) Parameters.

Answer :

(a) Population

Galton defines 'Population' as "totality of individual observations about which inferences are to be made, existing anywhere in the world, or within a definitely specified area limited by space and time".

When the term 'Population' is used in statistics, it corresponds to the observation of all the individuals that are used in giving conclusion at a specific time. The term 'Population' does not refer only to the human or animal population, instead it also refers to the biostatistical population. There are two different types of populations. They are,

(i) Finite population

(ii) Infinite population.

(i) Finite Population

A population whose values can be theoretically observed as it has fixed number is called a finite population. The size of the finite population is limited.

Example: Number of students in a school, number of rice plants in a field etc.

(ii) Infinite Population

A population whose values cannot be theoretically observed as it does not have a fixed number is called an infinite population. The size of the infinite population is unlimited.

Example: Number of White Blood Cells (WBC) in the human body.

The size of the population is always very large, even if it is finite or infinite. So, the collection of data from each of the individuals is very difficult. Because of this, samples are selected from the population and based on their results conclusions are made regarding that population.

(b) Sample

In biostatistical studies, the data collected depends on the observation of several individuals of a population. Here, sampling can also be performed i.e., few individuals can be selected from a population to represent the complete population. These selected individuals are together called as 'sample'.

Random samples is a sample selected from a population in such a way that each individual in the population has a fair chance of getting selected.

There are two ways into which samples are classified,

(i) Large Sampling

The sample comprising of objects which are more than 30 (i.e., $n \geq 30$), it is known as large sampling.

(ii) Small Sampling

The sample comprising of objects less than 30 (i.e., $n < 30$), it is known as small sampling.

(c) Sample Unit

It can be defined as the distinguishable and observable part of the population which is used for gathering information.

Example

1. In a university (population) survey, each student can be a sample unit.
2. In a crop (population) estimation survey, a specific part of the land can be a sample unit.

Q38. What are the different methods of sampling?

Answer :

The different methods of sampling are as follows,

(i) Purposive sampling

(ii) Random sampling

(iii) Simple sampling

(iv) Stratified sampling

(i) **Purposive Sampling**

If the sample of elements are selected with some purpose then it is said to be purposive sampling.

Example

If there is a complaint against the defectiveness of components produced, then sample of elements which are defective are considered, whereas others are not considered. This is purposive sampling.

(ii) **Random Sampling**

If every element in (space) sample space have equal chance of being included in test, then it comes under random sampling.

(iii) **Simple Sampling**

It is a special type of random sampling in which selection chance of element in a sample is not dependent on the previous selection made is known as simple sampling.

Example

The selection made with a coin or a die comes under, simple sampling.

(iv) **Stratified Sampling**

Heterogeneous population is divided into subpopulation of states which are homogeneous within itself than in whole population.

Example

The population of people watching movies can be subdivided into stratus of people watching Hindi movies and people watching English movies. After dividing into strates, random selection of individuals are made from (these) each stratum.

The aggregate of sampled individuals of each stratum is known to be stratified sample. This technique of selecting is known as stratified sampling technique.

Q39. Explain about measures of location.

Answer :

Measures of Location

Measures of location can be used to estimate about the cluster of values or observations in the central part of the distribution. These are also known as 'Averages'. Averages are the values that lie between the smallest and the largest observations. It is the Mean of the given data. Some of the important measures of location are Arithmetic mean, median and mode.

1. **Arithmetic Mean**

The centre of data set for symmetric data is computed by a conventional, simple and most effectual approach known as 'Arithmetic mean'. This approach is based on numeric quantities which are used to analyze the co-relation between different data values.

Mathematically, the mean of a sample for 'n' observations is computed using the formula,

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Where, X_i represents the set of n data values.

Calculating mean is similar to that of average calculation by using built-in function AVERAGE(data range) present in Excel. It operates based on the property that sum of the deviations for every single observation obtained from the mean is equal to 0.

i.e., $\sum_i (X_i - \overline{X}) = 0$

The mean of a population is given by,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i$$

2. **Median**

Another approach for estimating centre of data set for asymmetric data is 'median' which is an example of holistic measure. There are two cases for calculating median for k-data values which are distinctive and sequentially ordered.

(i) **When n is Odd**

The median is the center value of entire sorted data set. If the value of the 'n' is odd then the median can be calculated as $\tilde{x} = x_{(n+1)/2}$.

(ii) **When n is Even**

The median is the mean of two center data value. If the value of the 'n' is even then the median can be calculated as $\frac{1}{2}(x_{n/2} + x_{n/2+1})$

3. **Mode**

Mode measure is another way of computing central tendency. If a data value occurs more number of times than any other values in data set, then that data value is referred as mode. If a data set contains values which are distinctive then there is no mode in that data set. However, there are situations, where more than one data value may be occurring frequently in the same data set because of which there is a possibility of getting more than one mode.

Q40. Discuss about Measures of variability.

Answer :

The following are the different measures of variability,

1. Variance
2. Standard deviation
3. Range.

1. Variance

Variance is defined as the measure of variability which uses complete data. It depends on the deviation about the mean i.e., difference between the value of every observation (X_i) and the mean. It is the most frequently used measure of dispersion. If the variance is maximum then the data obtained from the mean is more. The variance for a population is given by,

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}$$

Where, x_i is the value of an i^{th} element or an observation k is the total number of observations in the population. The variance of a sample is computed using the formula,

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

Where, X_i is the total number of observations in an sample and $\overline{X}$ is mean for a given sample.

2. Standard Deviation

Standard deviation is defined as a measure of variability that is the positive square root of variance. Standard deviation for a sample is given by,

$$S = \sqrt{S^2}$$

The standard deviation for a population can be computed by the formula,

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}}$$

The standard deviation for a samples is given by,

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

3. Range

Range for n-data values $a_1, a_2, ..., a_n$ can be computed by finding out the difference between the maximum value and minimum value i.e., subtracting maximum data value from minimum data value.

Range = Max (data range) – MIN (data range)

PROBLEM

Q41. A comparison of coffee prices at 4 randomly selected grocery stores in San Diego showed increases from the previous month of 12, 15, 17 and 20 cents for a 1-pound bag. Find the variance of this random sample of price increases.

Solution :

Mathematically, the mean of a sample for n observations is computed using the formula,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Now, substitute $n = 4$ and $X_i = 12, 15, 17$ and 20 in the above equation, then,

$$\overline{X} = \frac{1}{4}(12+15+17+20)$$

$$= \frac{64}{4}$$

$$\boxed{\overline{X} = 16}$$

The variance of a sample in computed by using the formula,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$= \frac{1}{4-1}\sum_{i=1}^{4}(X_i - 16)^2 \qquad (\because \text{ The value of } \overline{X} = 16)$$

$$= \frac{(12-16)^2 + (15-16)^2 + (17-16)^2 + (20-16)^2}{3}$$

$$= \frac{(-4)^2 + (-1)^2 + (1)^2 + (4)^2}{3}$$

$$\boxed{S^2 = \frac{34}{3}}$$

3.2.2 Some Important Statistics

Q42. Define the following terms,

(a) Statistic

(b) Sampling frame.

Answer :

(a) Statistic

Statistic acts as a helpful practice used for the purpose of collecting, classifying, presenting and interpreting the data. It is otherwise known as 'Analytical statistics'.

Some of the definitions of statistics are presented as follows,

(i) According to A.L.Bowley,

"Statistics is the science of counting".

(or)

"Statistics may rightly be called as the science of averages".

(ii) According to Turtle, "Statistics is a body of principles and techniques of collecting, classifying, presenting, comparing and interpreting the quantitative data".

Thus, the above definitions specify that statistics in singular sense, is a science which comprises different statistical methods used for collection, organization, classification, presentation and interpretation of data.

In plural sense, the term "Statistics" implies systematical collection of numerical facts like statistics of population, production, price-level, national income and so on.

(b) Sampling Frame

Sampling frame is a collection of all the members/units of population.

It should not contain any error and duplication of units. Each and every individual sampling unit is selected from a sampling frame.

It possess the following characteristics,

1. **Complete**

   A frame should consist of all legal units of the population.

2. **Accuracy**

   A frame should not consist of any non-existing units of population.

3. **Adequate**

   The structure of a frame must be sufficient enough to cover the total population.

4. **Updated**

   The unit and its content present in a frame must be up to date.

3.2.3 Sampling Distributions, Sampling Distribution of Means and the Central Limit Theorem

Q43. Define sampling distribution. Explain the characteristic of sampling distribution. Discuss about sampling distribution of mean with known σ.

Answer : Model Paper-II, Q7(a)

Sampling Distribution

Samples of a given size may be drawn randomly from a population and the statistical constants like mean, standard deviation may be computed for each such sample. The distribution of each such statistic is called sampling distribution.

Characteristic of a Sampling Distribution

Sampling distribution is a frequency distribution representing the means taken from a great many samples, of the same size. The main characteristic of this is that it approaches normal distribution even went the population distribution is not normal provided the sample size is sufficiently large (greater than 30). The significance of the sampling distribution follows from the fact that the mean of sampling distribution is the same as the mean of the population.

The mean of a sample of size n is denoted by $\overline{x}$ and the probability distribution of $\overline{x}$ is called the sampling distribution of the mean. It depends upon the population size, samples size and the method of selecting the samples.

The mean, variance and Standard Deviation (S.D) of the samples drawn from an infinite population are given by,

$$\text{Mean, } \mu_{\overline{x}} = \frac{\mu+\mu+...+\mu}{n} = \frac{\mu}{n}$$

$$\text{Variance, } \sigma^2_{\overline{x}} = \frac{\sigma^2+\sigma^2+...+\sigma^2}{n} = \frac{\sigma^2}{n}$$

$$\text{S.D} = \sqrt{\text{Variance}} = \sqrt{\sigma^2_{\overline{x}}}$$

$$\sigma_{\overline{x}} = \sqrt{\frac{\sigma^2}{n}}$$

$$\therefore \quad \text{S.D} = \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Therefore, the sampling distribution of $\overline{x}$ is approximately normal with the mean and variance, if the sample size is large.

Similarly, the mean, variance and S.D of the samples drawn from finite population are given by,

$$\text{Mean, } \mu_{\overline{x}} = \mu$$

$$\text{Variance, } \sigma^2_{\overline{x}} = \frac{\sigma^2}{n} = \left[\frac{N-n}{N-1}\right] \quad \text{(Correction factor)}$$

$$\text{S.D} = \sigma_{\overline{x}} = \sqrt{\frac{N-n}{N-1}}$$

Where,

N = Size of the finite population

n = Size of the sample.

Q44. State and explain central limit theorem.

Answer :

If $\bar{x}$ is the random sample mean of size n taken from a population having the mean μ and finite variance σ^2, then the limiting form of the distribution of $Z = \dfrac{\overline{x}-\mu}{\sigma/\sqrt{n}}$ as $n \to \infty$, is the standard normal distribution $N(0, 1)$.

The central limit theorem is one of the important concepts in statistics. It states that the distribution of sample means tends to be always normal distribution. This is true and also is free from shape of the population distribution, from where the sample is taken. The mean of sampling distribution is equal to the population mean and is independent of its sample size whether the population is normal (or) not.

As the sample size increases, the sampling distribution of the mean will approach normality. The relationship between shape of sampling distribution of mean and shape of population distribution is called as Central Limit Theorem.

If the random sample comes from a normal population, the sampling distribution of the mean is normal regardless of the size of the sample.

Q45. Derive the expression for,

(i) Sampling distribution of difference between two means

(ii) Sampling distribution of sums of two means.

Answer :

(i) Sampling Distribution of Difference Between Two Means

Consider two populations in which $\mu_{\overline{X_1}}$ and $\sigma_{\overline{X_1}}$ denote the mean and standard deviation of a sampling distribution $\overline{X_1}$ obtained by calculating $\overline{X_1}$ for all possible samples of size n_1 drawn from population 1. Similarly, for each sample of size n_2 taken from the population 2, S_2 is computed whose mean and standard deviation are $\mu_{\overline{X_2}}$ and $\sigma_{\overline{X_2}}$ respectively.

From the two populations, all the possible combination of the samples are taken and then the distribution of the difference is calculated (i.e.,) $\overline{X_1} - \overline{X_2}$ which is known as the sampling distribution of differences of the statistics. These selected sample do not depend on each other.

The mean of the sampling distribution of difference is,

$$\boxed{\mu_{\overline{X_1} - \overline{X_2}} = \mu_{\overline{X_1}} - \mu_{\overline{X_2}}}$$

The standard deviation of the sampling distribution of difference is,

$$\boxed{\sigma_{\overline{X_1} - \overline{X_2}} = \sqrt{\sigma^2_{\overline{X_1}} + \sigma^2_{\overline{X_2}}}}$$

If S_1 and S_2 are the sample means from the two populations, denoted by $\overline{X_1}$, $\overline{X_2}$ respectively, then the sampling distribution of the differences of means is given for infinite populations with mean and standard deviation μ_1, σ_1 and $\mu_2 \sigma_2$ by

$$\mu_{\overline{X_1} - \overline{X_2}} = \mu_{\overline{X_1}} - \mu_{\overline{X_2}} \; \sigma = \mu_1 - \mu_2$$

$$\sigma_{\overline{X_1} - \overline{X_2}} = \sqrt{\sigma^2_{\overline{X_1}} + \sigma^2_{\overline{X_2}}}$$

$$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

(ii) Sampling Distribution of Sum Between Two Means

Similarly to the difference between two means, the mean of the sampling distribution of sums is

$$\boxed{\mu_{\overline{X_1} + \overline{X_2}} = \mu_{\overline{X_1}} + \mu_{\overline{X_2}}}$$

The standard deviation of the sampling distribution of sums is

$$\boxed{\sigma_{\overline{X_1} + \overline{X_2}} = \sqrt{\sigma^2_{\overline{X_1}} + \sigma^2_{\overline{X_2}}}}$$

If S_1 and S_2 are the sample means from two populations, denoted by $\overline{X_1}$ and $\overline{X_2}$ respectively, then the sampling distribution of the sums of the means is given for infinite population with means and standard deviation μ_1, σ_1 and μ_2, σ_2 by,

$$\mu_{\overline{X_1} + \overline{X_2}} = \mu_{\overline{X_1}} + \mu_{\overline{X_2}} = \mu_1 + \mu_2$$

$$\sigma_{\overline{X_1} + \overline{X_2}} = \sqrt{\sigma^2_{\overline{X_1}} + \sigma^2_{\overline{X_2}}}$$

$$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

PROBLEMS

Q46. The following are the times between six calls for an ambulance in a city and the patient's arrival at the hospital 27, 15, 20, 32, 18 and 26 minutes. Use these figures to judge the reasonableness of the ambulance services claim that it take on an average 20 minutes between the call for an ambulance and patients arrival at hospital.

Solution :

Given that,

$$n = 6 \text{ and } \mu = 20$$

Let $x_1, x_2, x_3, x_4, x_5, x_6$ be the time between six calls for an ambulance and the patient arrival at hospital.

Here,

$$x_1 = 27$$
$$x_2 = 15$$
$$x_3 = 20$$
$$x_4 = 32$$
$$x_5 = 18$$
$$x_6 = 26$$

Test statistic is expressed as,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Where,

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6}$$

$$= \frac{27 + 15 + 20 + 32 + 18 + 26}{6}$$

$$= \frac{138}{6}$$

$$= 23$$

Variance (S^2) is given by,

$$S^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

$$= \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+(x_4-\bar{x})^2+(x_5-\bar{x})^2+(x_6-\bar{x})^2}{n-1}$$

$$= \frac{(27-23)^2+(15-23)^2+(20-23)^2+(32-23)^2+(18-23)^2+(26-23)^2}{6-1}$$

$$= \frac{(4)^2+(-8)^2+(-3)^2+(9)^2+(-5^2)+(3)^2}{5}$$

$$= \frac{16+64+9+81+25+9}{5}$$

$$= \frac{204}{5}$$

$$= 40.8$$

$\therefore$ Variance, $S^2 = 40.8$

$\therefore$ Standard deviation, $s = \sqrt{S^2}$

$$= \sqrt{40.8}$$

$$= 6.4$$

Substituting $\bar{x} = 23$, $\mu = 20$, $s = 6.4$ and $n = 6$ in equation (1), we get,

$$t = \frac{23-20}{6.4/\sqrt{6}}$$

$$= \frac{3}{6.4/2.44} = \frac{3\times 2.44}{6.4}$$

$$= \frac{7.32}{6.4}$$

$$= 1.14$$

From the table of t distribution, the probability that 't' will exceed is 1.48 when degrees of freedom is 6 is 0.1.

$1.14 < 1.48$

$\therefore \quad \mu \neq 20$

Thus, the given information refuse the claim that the mean of the population is $\mu = 20$.

Q47. A random sample of size 25 from a normal population has mean 47.5 and standard deviation 8.4. Does this information tend to support or refuse the claim that the mean of the population is 42.1?

Solution :

Given that,

Size of sample, $n = 25$

Mean of sample, $\bar{X} = 47.5$

Standard deviation of sample, $S = 8.4$

Mean of population, $\mu = 42.1$

t-distribution is expressed as,

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$$

$$= \frac{47.5 - 42.1}{\frac{8.4}{\sqrt{25}}}$$

$$= \frac{5.4}{\frac{8.4}{5}}$$

$$= \frac{5.4 \times 5}{8.4}$$

$$= \frac{27}{8.4}$$

$\therefore\ t = 3.21$

When, $t = 3.21$, there are $(n - 1)$ i.e., $(25 - 1)$

$= 24$ degrees of freedom.

For $\mu = 24$, a nearer value to 3.21 is taken from the table of t-distribution. This value is 2.797 at $\alpha = 0.995$ i.e., $\alpha = 1 - 0.995 = 0.005$

$\therefore\quad \mu = 24$ and $t = 2.797$, $\alpha = 0.005$

Similarly, when $v = 24$ and $t = 3.21$,

$\alpha =$ Negligible (i.e., approximately equal to zero).

Thus, the given information refuse the claim that the mean of the population is $\mu = 42.1$.

Q48. Travelling between two campuses of a university in a city via shuttle bus takes on average, 28 minute with a standard deviation of 5 minutes. In a given week, a bus transported passengers 40 times. What is the probability that the average transport time was more than 30 minutes? Assume the mean time is measured to the nearest minute.

Solution :

Given that,

Mean, $\mu = 28$

Variance, $\sigma = 5$

$n = 40$

To find $P(\overline{X} > 30)$ with $n = 40$,

Consider the value of $\bar{x} = 30.5$. Since it is given that the $\bar{x}$ is greater than 30.

By using central limit theorem $z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$

We get, $P(\overline{X} > 30) = P\left(\dfrac{\overline{X} - 28}{\frac{5}{\sqrt{40}}} \geq \dfrac{30.5 - 28}{\frac{5}{\sqrt{40}}}\right)$

Let $z = \dfrac{\overline{X} - 28}{\frac{5}{\sqrt{40}}}$

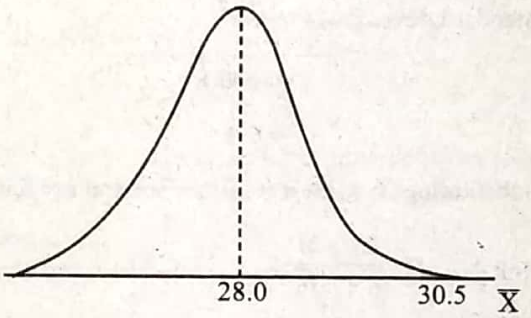
$$= P\left(z \geq \frac{2.50}{\frac{5}{\sqrt{40}}}\right)$$

$$= P(z \geq \sqrt{10})$$

$$= P(z \geq 3.16)$$

$$= 0.0008$$

$\therefore$ It has very less probability that the average time of a bus trip will be more than 30 minutes which is shown in the following figure,



Q49. Two independent experiments are run in which two different types of paints are compared. Eighteen specimens are painted using type A and the drying time in hours, is recorded for each. The same is done with type B. The population standard deviations are both known to be 1.0.

Assuming that the mean drying time is equal for the two types of paint, find $P(\overline{X}_A - \overline{X}_B > 1.0)$, where $\overline{X}_A$ and $\overline{X}_B$ are average drying time for samples of size $n_A = n_B = 18$.

Solution :

Given that,

$\overline{X_A} - \overline{X_B} = 1.0$

$n_A = n_B = 18$

The sampling distribution of mean is given by,

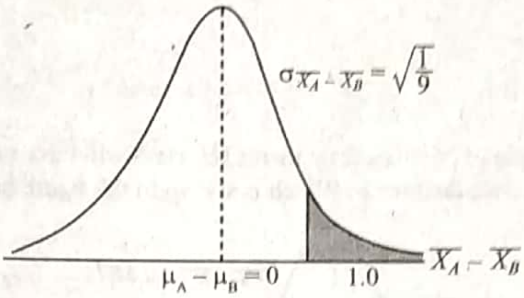
$\mu_{\overline{X_A} - \overline{X_B}}$

$= \mu_A - \mu_B \Rightarrow 0$

The sampling distribution of variance is given by,

$$\sigma^2_{\overline{X}_A - \overline{X}_B} = \frac{\sigma_A^2}{nA} + \frac{\sigma_B^2}{n_B}$$
$$= \frac{1}{18} + \frac{1}{18}$$
$$= \frac{1+1}{18}$$
$$= \frac{2}{18}$$
$$= \frac{1}{9}$$

The shaded region in the following figure shows the required probability.



By using the theorem, sampling distribution of the difference between two means, we have,

$$Z = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\left(\frac{\sigma_A^2}{n_A}\right) + \left(\frac{\sigma_B^2}{n_B}\right)}}$$

$$Z = \left(\frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sigma_{X_A - X_B}}\right) \qquad \left(\because \sigma_{X_A - X_B} = \sqrt{\left(\frac{\sigma_A^2}{n_A}\right) + \left(\frac{\sigma_B^2}{n_B}\right)}\right) \quad \text{... (1)}$$

Substitute the values of $\overline{X}_A - \overline{X}_B = 1.0$ and $\sigma_{\overline{X}_A - \overline{X}_B} = \sqrt{\frac{1}{9}}$

We get,

$$Z = \frac{1 - (\mu_A - \mu_B)}{\sqrt{\frac{1}{9}}}$$
$$= \frac{1 - 0}{\sqrt{\frac{1}{9}}}$$
$$= 3.0$$

Now, $P(z > 30) = 1 - P(z < 30)$

$= 1 - 0.9987$ $\qquad$ ($\because$ The corresponding value of $P(z < 30) = 0.9987$)

$= 0.0013$

Q50. The television picture tubes of manufacturer A have a mean lifetime of 6.5 years and a standard deviation of 0.9 year, while those of manufacturer B have a mean lifetime of 6.0 years and a standard deviation of 0.8 year. What is the probability that a random sample of 36 tubes from manufacturer A will have a mean lifetime that is atleast 1 year more than the mean lifetime of a sample of 49 tubes from manufacturer B?

Solution :

Given that,

Manufacture A	Manufacture B
$\mu_1 = 6.5$	$\mu_2 = 6.0$
$\sigma_1 = 0.9$	$\sigma_2 = 0.8$
$n_1 = 36$	$n_2 = 49$

The mean of the sampling distribution for the difference between two means is given by,

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$$

$$= 6.5 - 6.0$$

$$= 0.5$$

The standard deviation of the sampling distribution for difference between two means is given by,

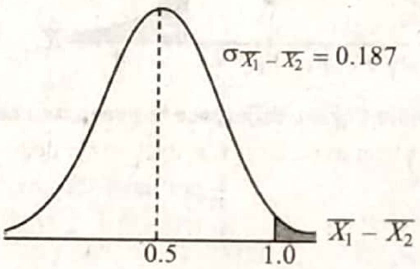
$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$= \sqrt{\frac{0.8}{36} + \frac{0.64}{49}}$$

$$= \sqrt{0.022 + 0.013}$$

$$= 0.187$$

The probability that a random sample of 36 tubes from manufacturer A will have mean lifetime that is at least 1 year more than the mean lifetime for 49 tubes from manufacturer B. Which is shown in the figure below,



The corresponding value of $\bar{x}_1 - \bar{x}_2$ is equal to 1.0

Now, by using the formula

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1}\right) + \left(\frac{\sigma_2^2}{n_2}\right)}}$$

$$Z = \frac{1.0 - 0.5}{0.187}$$

$$= 2.674$$

Therefore, $P(\overline{X}_1 - \overline{X}_2 \geq 1.0) = P(z > 2.65)$

$$= 1 - P(z < 2.65)$$

$$= 1 - 0.9960$$

$$= 0.0040$$

3.2.4 Sampling Distribution of S²

Q51. Explain in detail about sampling distribution of S².

Answer :

Sampling Distribution of S²

The sampling distribution of variance can be obtained by drawing all the possible random samples of size 'n' from the given normal population with mean μ and variance $\sigma 2$ and then calculating the variance for each of the sample is given by,

$$\sum_{i=1}^{n}(X_i - \mu)^2$$

Now, adding and subtracting sample mean (i.e., $\overline{X}$) in the above equation then the equation becomes,

$$\sum_{i=1}^{n}(X_i-\mu)^2=\sum_{i=1}^{n}(X_i-\overline{X}+\overline{X}-\mu)^2$$

$$=\sum_{i=1}^{n}[(X_i-\overline{X})+(\overline{X}-\mu)]^2$$

$$=\sum_{i=1}^{n}(X_i-\overline{X})^2+\sum_{i=1}^{n}(\overline{X}-\mu)^2+2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\overline{X}) \qquad [\because (a+b)^2=a^2+b^2+2ab]$$

$$=\sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-\mu)^2+0 \qquad \left[\because 2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\overline{X})=0\right]$$

$$\sum_{i=1}^{n}(X_i-\mu)^2=\sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-\mu)^2$$

Dividing L.H.S and R.H.S by σ^2 and substituting $(n-1)s^2$ for $\sum_{i=1}^{n}(X_i-\overline{X})^2$ then the equation changes to,

$$\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}=\frac{(n-1)s^2}{\sigma^2}+\frac{n(\overline{X}-\mu)^2}{\sigma^2}$$

$$=\frac{(n-1)s^2}{\sigma^2}+\frac{(\overline{X}-\mu)^2}{\frac{\sigma^2}{n}}$$

As we know that, $\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}$ is a chi-squared random variable with n degrees of freedom having the variance σ^2, then the statistic.

$$\chi^2=\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2}=\frac{(n-1)s^2}{\sigma^2}$$

Contains a chi-squared distribution with $\gamma=n-1$ degrees of freedom.

Q52. Discuss about Degrees of freedom as a Measure of sample information.

Answer :

Degree of freedom is defined as the number of observations that is one less than the total number of observations in a sample. The value is always in the denominator and is much more preferable than the sample size. It is given as,

$df=n-1$

Where,

df = Degree of freedom

n = Number of daservations in a sample

According to theorem, $\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}$ that contains a χ^2 – distribution with n degrees of freedom. This equation is equal to $\frac{(n-1)s^2}{\sigma^2}$ and has a χ^2-distribtion with $n-1$ degree of freedom. In case, if the value of the μ is unknown in $\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}$ then one can use the distribution of $\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2}$.

PROBLEM

Q53. A manufacturer of car batteries guarantees that the batteries will last, on average, 3 years with a standard deviation of 1 year. If five of these batteries have lifetimes of 1.9, 2.4, 3.0, 3.5 and 4.2 years, should the manufacturer still be convinced that the batteries have a standard deviation of 1 year. Assume that the battery lifetime follows a normal distribution.

Model Paper-II, Q7(b)

Solution :

Given that,

$n=5,\ X_i^2=48.26,\ X_i=15,\ \sigma_2=1$

By using theorem of statistic, we have, $S_1^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]$ ... (1)

Now, substitute the respective values in equation (1), then,

$$S^2 = \frac{1}{5(5-1)}\left[(5)(48.26) - (15)^2\right]$$

$$= \frac{1}{5 \times 4}[241.30 - 225]$$

$$= \frac{16.30}{20}$$

$$\boxed{S^2 = 0.82}$$

From sampling distribution of S^2 we have,

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \quad \text{... (2)}$$

Substitute the respective values in equation (2) then,

$$\chi^2 = \frac{(5-1)(0.82)}{1}$$

$$= 4 \times 0.82$$

$$\chi^2 = 3.28$$

Therefore, 3.28 is the value of the chi-squared distribution with $\gamma = n - 1$ degrees of freedom i.e., 4 degrees of freedom. Since 95% of the random variable (χ^2) values with 4 degrees of freedom lies between 0.484 and 11.143, the calculated value with σ^2 is 1 appropriately.

Thus, the manufacturer cannot respect that the standard deviation is more than 1 year.

3.2.5 t-Distribution

Q54. Define t-distribution. State its properties and applications.

Answer : Model Paper-III, Q7(a)

t-Distribution

For unknown population standard deviation (s_p) and small size (i.e., $n < 30$) sample, 't' distribution (student's 't' distribution) is used for the sampling distribution of mean and workout 't' variable as,

$$t = \frac{(\overline{X} - \mu)}{\frac{S}{\sqrt{n}}}$$

Where, $\sigma_s = \sqrt{\frac{\Sigma(X_i - \overline{X})^2}{n-1}}$

i.e., the sampling standard deviation

t-distribution is also symmetrical and is very close to the distribution of standard normal variate Z, except for small values of 'n'. For large sample size or $n \geq 30$, the distribution of T remains same as that of standard normal. There are different 't' distributions one for each sample size i.e., for different degrees of freedom. The degrees of freedom for a sample of size n is $(n - 1)$.

The sampling distribution of T can be developed by selecting the random sample from a normal population and it is written as,

$$T = \frac{\frac{(\overline{X} - \mu)}{(\sigma - \sqrt{n})}}{\sqrt{\frac{S^2}{\sigma^2}}}$$

Where, $z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ contains the standard normal distribution and $v = \frac{(n-1)S^2}{\sigma^2}$ contains chi-squared distribution with $\gamma = n - 1$ degrees of freedom.

Properties

(i) The probability curve of t is symmetric, like in standard normal distribution (z).

(ii) The t-distribution ranges from $-\infty$ to ∞ just as does a normal distribution.

(iii) The t-distribution is bell shaped and symmetrical around mean zero, like normal distribution.

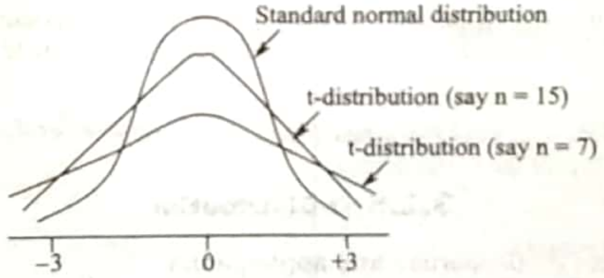
(iv) The shapes of the t-distribution changes as the sample size changes (the number of degrees of freedom changes) whereas it is same for all sample sizes in z-distribution.

(v) The variance of t-distribution is always greater than one and is defined only when $n \geq 3$.

(vi) The t-distribution is more of platykurtic (less packed at centre and higher in tails) than the normal distribution.

(vii) The t-distribution has a greater dispersion than the normal distribution. As n becomes larger, the t-distribution approaches the standard normal distribution.

(viii) There is a family of t-distribution one for each sample size whereas, there is only one standard normal distribution.



Figure

Applications

The following are some important applications of t-distribution.

(i) Test of hypothesis about the population mean.

(ii) Test of hypothesis about the difference between two means.

(iii) Test of hypothesis about the difference between two means with dependent samples.

(iv) Test of hypothesis about coefficient of correlation.

PROBLEMS

Q55. Find $P(-t_{0.025} < T < t_{0.05})$.

Solution :

In a curve, $t_{0.05}$ leaves an area of 0.05 to the right and $-t_{0.025}$ leaves an area of 0.025 to the left. Then, the total area is given by,

$= 1 - 0.05 - 0.025$

$= 1 - 0.025$

$= 0.925$

Therefore, the total area that lies in between $-t_{0.025}$ and $t_{0.05}$ is,

$P(-t_{0.025} < T < t_{0.05}) = 0.925.$

Q56. Find k such that $P(k < T < -1.761) = 0.045$ for a random sample of size 15 selected from a normal distribution and $\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$.

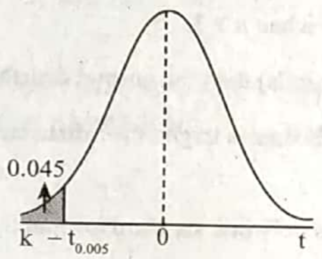
Solution :

Given that, $P(k < T < -1.761) = 0.045$

The value -1.761 corresponds to $-t_{0.05}$ when $v = 14$ obtained from critical values of the

t-Distribution Table

Consider $k = -t\alpha$ and then, draw the curve based on the given values as shown below,



From the above figure, we can compute that,

$0.045 = 0.05 - \alpha$ or $\alpha = 0.005$

The corresponding value of k for $-t_{0.005}$ is -2.977 when $v = 14$ is obtained from the critical values of the t-Distribution Table.

$\therefore \quad P(-2.977 < T < -1.761) = 0.045.$

3.2.6 F-Distribution

Q57. Explain about F-distribution with two sample variances. Also, explain its properties.

Answer :

Consider S_1^2 and S_2^2 as the sample variances of different random sample of sizes n_1 and n_2 respectively. These samples are drawn from two different normal population $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, where (μ_1, σ_1^2) and (μ_2, σ_2^2) denotes the mean and variances of S_1^2 and S_2^2 respectively. Inorder to determine whether the samples (S_1^2, S_2^2) are drawn from two different population, having equal variances. It is necessary to compute the ratio of variances related to two independent random sample. This ratio is computed as,

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

If it is assumed that normal population have equal variance, then,

$$F = \frac{S_1^2}{S_2^2} \qquad [\because \sigma_1^2 = \sigma_2^2] \qquad \ldots (1)$$

The above equation is a random variable having F distribution with parameters $V_1 = n_1 - 1$ and $V_2 = n_2 - 1$.

$$F = \frac{S_1^2}{S_2^2} \Rightarrow \frac{\left[\frac{n_1}{n_1-1}(S_1^2)\right]}{\left[\frac{n_2}{n_2-1}(S_2^2)\right]} \Rightarrow \frac{\left[\frac{n_1(S_1^2)}{\sigma_1^2}\cdot\frac{1}{n_1-1}\right]}{\left[\frac{n_2(S_2^2)}{\sigma_2^2}\cdot\frac{1}{n_2-1}\right]}$$

Since, $\frac{n_1(S_1^2)}{\sigma_1^2}$ and $\frac{n_2(S_2^2)}{\sigma_2^2}$ are independent of chi-square variant with $(n_1 - 1)$ and $(n_2 - 1)$ degree of freedom, F follows F-distribution.

In equation (1), greater of two variance S_1^2 and S_2^2 is to be taken in the numerator and n_1 corresponds to greater variance.

By comparing the calculated value of F which is obtained by using equation (1) for the two given samples with the tabulated value of F for (n_1, n_2) d.o.f at a certain level significance (5% or 1%), M_0 is either rejected or accepted.

The significant value $F_\alpha(n_1, n_2)$ at a level of significance α_1 and (n_1, n_2) d.o.f is determined by, $P[F > F_\alpha(n_1, n_2)] = \alpha$, as shown below,

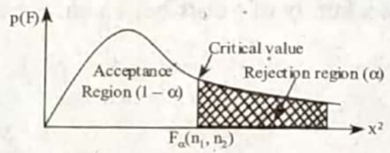


Figure: Critical Value of F-distribution

$$F_\alpha(n_1, n_2) = \frac{1}{F_{1-\alpha}(n_2, n_1)}$$

$$\Rightarrow \quad F_\alpha(n_1, n_2) \times F_{1-\alpha}(n_2, n_1) = 1$$

Properties

1. The F-distribution curve lies in only first quadrant (Q_1) and is unimodal.
2. The F-distribution is independent (free) of population parameter and depends only on the degree of freedom (i.e., V_1 and V_2) according to its order.
3. The F-distribution mode is less than unity (i.e., mode < 1).
4. In the F-distribution, figure.

$$\boxed{F_1 - \alpha(V_1, V_2) = \frac{1}{F_\gamma(V_2, V_1)}}$$

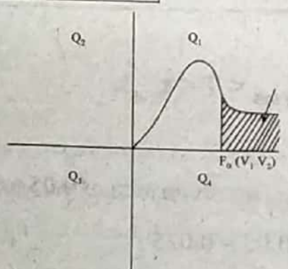


Figure: F-Distribution Curve

Where,

$F_\alpha(V_1, V_2)$ = Value of F.

So, α is at the right of $F_\alpha(V_1, V_2)$ under the F-distribution curve.

Q58. What is the F-Distribution used for?

Answer :

Uses of F-Distribution

1. It is used for drawing inferences that are related to the population variance in a two sample condition.
2. It is used in different types of problems that include sample variance.
3. It is used for testing the equality of many population means.
4. It is used for comparing the sample variances.
5. It is used for performing analysis of variance.
6. It is used for testing the significance of regression equation.
7. It is used for determining whether the ratio incrementally changes from unity at any level chosen randomly.

PROBLEMS

Q59. A random sample of size 25 from a normal population has mean 47.5 and standard deviation 8.4. Does this information tend to support or refuse the claim that the mean of the population is 42.1?

Solution : Model Paper-III, Q7(b)

Given that,

Size of sample, $n = 25$

Mean of sample, $\overline{X} = 47.5$

Standard deviation of sample, $S = 8.4$

Mean of population, $\mu = 42.1$

We know that t-distribution,

$$t = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$$

$$= \frac{47.5-42.1}{\frac{8.4}{\sqrt{25}}}$$

$$= \frac{5.4}{\frac{8.4}{5}}$$

$$= \frac{5.4\times 5}{8.4}$$

$$= \frac{27}{8.4}$$

$$t = 3.21$$

When, $t = 3.21$, there are $(n-1)$ i.e., $(25-1)$

$= 24$ degrees of freedom.

For $v = 24$, a nearer value to 3.21 is taken from the table of t-distribution. This value is 2.797 at $\alpha = 0.995$ i.e., $\alpha = 1-0.995 = 0.005$

$\therefore$ When, $v = 24$ and $t = 2.797$, $\alpha = 0.005$

Similarly, when $v = 24$ and $t = 3.21$, $\alpha =$ Negligible (i.e., approximately equal to zero).

Thus, the given information refuse the claim that the mean of the population is $\mu = 42.1$.

Q60. If two independent random samples of sizes $n_1 = 9$ and $n_2 = 16$ are taken from a normal population. What is the probability that the variance of the first sample will be at least 4 times as large as the variance of the second sample?

Solution :

Given that,

$$n_1 = 9, n_2 = 16$$

∴ Degrees of freedom

$$V_1 = n_1 - 1 = 9 - 1 = 8$$

$$V_2 = n_2 - 1 = 16 - 1 = 15$$

$$S_1^2 = 4S_2^2$$

$$\therefore \quad F = \frac{S_1^2}{S_2^2} = \frac{4S_2^2}{S_2^2} = 4.00$$

The value of $F = 4.00$ follows F-distribution with $V_1 = 8$ and $V_2 = 15$ degrees of freedom. Hence, from the table we get,

$F_{0.01}$ (8, 15) = 4.00

∴ The required probability is 0.01.