B.Tech II Semester (2020 Batch)
PROBABILITY AND STATISTICS (20BM1104)
(For CSE-3 & CSE-4)
Units – 2: Sampling Distribution

(Population and sample, sampling distribution of the mean (σ known), sampling distribution of the mean (σ unknown), sampling distribution of the variance: Chi-square and F-distributions.)

Sampling Distribution: population and sample, sampling distribution of the mean (σ known), sampling distribution of the mean (σ unknown), sampling distribution of the variance: Chi-square and F-distributions.

Population: A collection of objects (collection of numbers, measurements, observations etc) is called a Population

The number of objects in a population is called its **Size** and is denoted by N

If N is finite then the population is called **Finite population**

If N is infinite then the population is called **Infinite population**

Examples:
(1) Heights of the students in a University
(2) Marks obtained by the students of SSC in Mathematics
(3) Scores of candidates obtained in a competitive exam
(4) The set of outcomes when a coin tossed 1000 times
(5) Collection of all even numbers

Parameters: The statistical measures (Mean, Median, Variance etc.) about a population are called Parameters. The Mean, Variance and Standard deviation of a population are respectively denoted by the symbols μ, σ^2 and σ

Sample: A finite sub collection from a population is called a Sample
The number of objects in a sample is called its **Size** and is denoted by n
If $n \geq 30$ then the sample is called **large sample**
If $n < 30$ then the sample is called **small sample**

Statistics: The statistical measures (Mean, Median, Variance etc.) about a sample are called statistics. The Mean, Variance and Standard deviation of a sample are respectively denoted by the symbols $\bar{x}, s^2$ and s

Examples:
(1) For the population of ‘the Heights of the students in a University’, the heights of the students in class of 40 is a sample
(2) For the population of ‘the Marks obtained by the students of SSC in Mathematics’, the marks of the students of a particular school is a sample
(3) For the population of ‘the Scores of candidates obtained in a competitive exam’, the scores of the candidates from a particular college is a sample
(4) For the population of ‘the set of outcomes when a coin tossed 1000 times’, a collection of 10 outcomes is a sample
(5) For the population of ‘the Collection of all even numbers’, a collection of 20 even numbers is a sample

Random Sample: A sample of size n taken from a population is called a random sample if the probability of any choice of n objects from the population is same.

Sampling: Collecting samples from a given population is called sampling
Large Sampling: Collecting large samples from a given population is called large sampling
Small Sampling: Collecting small samples from a given population is called small sampling
Sampling with replacement: Collecting samples in which the objects may repeat, from a given population is called sampling with replacement
Sampling without replacement: Collecting samples in which the objects not repeat, from a given population is called sampling without replacement

Number of Samples:

(i) The number of samples of size n without replacement, taken from a finite population of size N is ${}^{N}C_{n}$

(ii) The number of samples of size n with replacement, taken from a finite population of size N is N^{n}
(iii) The number of samples of size n with or without replacement, taken from an infinite population is ∞

Examples:
(1) Consider the finite population $\{1,2,3,4\}$ of size $N=4$

The number of samples of size $n=2$ without replacement is given by ${}^{N}C_{n}={}^{4}C_{2}=6$

The samples are $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$, $\{3,4\}$

The means of these 6 samples are respectively 1.5, 2, 2.5, 2.5, 3, 3.5

The frequency distribution of the these sample means is given as follows

Sample mean $\bar{x}$	1.5	2	2.5	3	3.5
Frequency f	1	1	2	1	1

This frequency distribution is called the **sampling distribution of the mean (SDM)**

Note: Here drawing a sample of size $n=2$ without replacement from the above population is a random experiment with the possible outcomes $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$, $\{3,4\}$

That is, the sample space $S=\{\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}\}$

If $\bar{X}$ is the random variable which gives the mean the sample, then the range of $\bar{X}$ is $\{1.5,2,2.5,3,3.5\}$

(2) Consider the finite population $\{1,2,3,4\}$ of size $N=4$

The number of samples of size $n=2$ with replacement is given by $N^{n}=4^{2}=16$

The samples are $\{1,1\}$, $\{1,2\}$, $\{1,3\}$, $\{1,4\}$,
$\{2,1\}$, $\{2,2\}$, $\{2,3\}$, $\{2,4\}$,
$\{3,1\}$, $\{3,2\}$, $\{3,3\}$, $\{3,4\}$,
$\{4,1\}$, $\{4,2\}$, $\{4,3\}$, $\{4,4\}$

The means of these 16 samples are respectively given follows

1, 1.5, 2, 2.5,
1.5, 2, 2.5, 3,
2, 2.5, 3, 3.5,
2.5, 3, 3.5, 4

The frequency distribution of the these sample means is given as follows

Sample mean $\bar{x}$	1	1.5	2	2.5	3	3.5	4
Frequency f	1	2	3	4	3	2	1

This frequency distribution is the **sampling distribution of the mean (SDM)**

Sampling Distribution of the Mean (SDM):
The frequency distribution of the means of all random samples of fixed size, taken from a population is called the Sampling Distribution of the Mean (SDM). It is denoted by $\overline{X}$.

Sampling Distribution of the Mean with σ (SDM with σ):
Theorem: If $\overline{X}$ is the random variable which gives the mean a random sample of size n, taken from a population having mean μ and variance σ^2, then

(i) Mean of SDM $\overline{X}$ is given by $\mu_{\overline{X}} = \mu$ or $E\left[\overline{X}\right] = \mu$

(ii) For infinite population or sampling with replacement,

Variance of SDM $\overline{X}$ is given by $\sigma_{\overline{X}}^2 = \dfrac{\sigma^2}{n}$ or $V\left[\overline{X}\right] = \dfrac{\sigma^2}{n}$

(iii) For finite population of size N and sampling without replacement,

Variance of SDM $\overline{X}$ is given by $\sigma_{\overline{X}}^2 = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$

Note:

(1) The value of $\dfrac{N-n}{N-1}$ is called the **finite population correction factor**

(2) The standard deviation of $\overline{X}$ is called **Standard error of the mean;** that is,

Standard error (SE), $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

(3) The value of $0.6745 \times \dfrac{\sigma}{\sqrt{n}}$ is called the **Probable error of the mean;** that is,

$$\text{Probable error (PE)} = 0.6745 \times \frac{\sigma}{\sqrt{n}}$$

Chebyshev's Theorem:
If $\overline{X}$ is the random variable which gives the mean a random sample of size n, taken from a population having mean μ and variance σ^2, then $P\left(\left|\overline{X} - \mu\right| < k\right) \geq 1 - \dfrac{\sigma^2}{nk^2}$, where k is a positive constant.

Note:

1. $P\left(\left|\overline{X} - \mu\right| < k\right) \geq 1 - \dfrac{\sigma^2}{nk^2}$ or $P\left(\mu - k < \overline{X} < \mu + k\right) \geq 1 - \dfrac{\sigma^2}{nk^2}$
2. $P\left(\left|\overline{X} - \mu\right| \geq k\right) \leq \dfrac{\sigma^2}{nk^2}$

Central Limit Theorem:
If $\overline{X}$ is the random variable which gives the mean a random sample of size n, taken from a population having mean μ and variance σ^2, then $Z = \dfrac{\overline{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$.

Examples:

(1) Find the finite population correction factor for $n = 10$ and $N = 1000$

Finite population correction factor, $\dfrac{N-n}{N-1} = \dfrac{1000-10}{1000-1} = 0.991$

(2) Find the finite population correction factor for $n = 5$ and $N = 100$

Finite population correction factor, $\dfrac{N-n}{N-1} = \dfrac{100-5}{100-1} = 0.9596$

(3) What is the effect on standard error, if a sample is taken from an infinite population and its size is increased from 400 to 900?

Here $n_1 = 400$ and $n_2 = 900$

Initially, the Standard Error $SE_1 = \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n_1}} = \dfrac{\sigma}{\sqrt{400}}$

After the sample size increasing, the Standard Error $SE_2 = \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n_2}} = \dfrac{\sigma}{\sqrt{900}}$

Now consider $\dfrac{SE_2}{SE_1} = \dfrac{\sigma}{\sqrt{900}} \times \dfrac{\sqrt{400}}{\sigma} = \dfrac{2}{3}$

Therefore, the Standard Error is decreased $\dfrac{2}{3}$ times of its original value

(4) What is the effect on standard error, if a sample is taken from an infinite population and its size is decreased from 100 to 25?

Here $n_1 = 100$ and $n_2 = 25$

Initially, the Standard Error $SE_1 = \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n_1}} = \dfrac{\sigma}{\sqrt{100}}$

After the sample size decreasing, the Standard Error $SE_2 = \sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n_2}} = \dfrac{\sigma}{\sqrt{25}}$

Now consider $\dfrac{SE_2}{SE_1} = \dfrac{\sigma}{\sqrt{25}} \times \dfrac{\sqrt{100}}{\sigma} = 2$

Therefore, the Standard Error is increased 2 times of its original value

(5) *A population consists 4 numbers 1, 2, 3, 4 (i) Find the Mean and Variance of the population (ii) Write all possible samples of size 2 without replacement (iii) Write the Sampling distribution of the mean (SDM) (iv) Find Mean and Variance of the Sampling distribution of the mean (SDM) (v) Verify the results in (iv) with suitable formula

Solution: (i) Mean and Variance of the population:

Mean, $\mu = \dfrac{1+2+3+4}{4} = \dfrac{10}{4} = 2.5$

Variance, $\sigma^2 = \dfrac{1}{n}\sum x^2 - \mu^2 = \dfrac{1+4+9+16}{4} - \mu^2 = \dfrac{30}{4} - (2.5)^2 = 7.5 - 6.25 = 1.25$

(ii) Samples of size 2 without replacement:

The number of samples of size $n = 2$ without replacement is given by ${}^{N}C_n = {}^{4}C_2 = 6$

The samples are $\{1,2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2,3\}$, $\{2,4\}$, $\{3,4\}$

(iii) The Sampling distribution of the mean (SDM):

The means of the above 6 samples are respectively 1.5, 2, 2.5, 2.5, 3, 3.5

The frequency distribution of the these sample means is given as follows

Sample mean $\bar{x}$	1.5	2	2.5	3	3.5
Frequency f	1	1	2	1	1

Which is the sampling distribution of the mean (SDM)

(iv) Mean and Variance of the Sampling distribution of the mean (SDM)

$$\text{Mean, } \mu_{\bar{X}} = \frac{1}{6}\sum \bar{x} = \frac{1.5+2+2.5+2.5+3+3.5}{6} = \frac{15}{6} = 2.5$$

$$\text{Variance, } \sigma_{\bar{X}}^2 = \frac{1}{6}\sum \bar{x}^2 - \mu_{\bar{X}}{}^2$$

$$= \frac{2.25+4+6.25+6.25+9+12.25}{6} - \mu_{\bar{X}}{}^2 = \frac{40}{6} - (2.5)^2$$

$$= 6.6667 - 6.25 = 0.4167$$

(v) Verification:

We have $\mu_{\bar{X}} = 2.5 = \mu$

and $\dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right) = \dfrac{1.25}{2}\left(\dfrac{4-2}{4-1}\right) = 0.4167 = \sigma_{\bar{X}}^2$

Therefore, $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$ are verified.

(6) *A population consists 4 numbers 1, 2, 3, 4 (i) Find the Mean and Variance of the population (ii) Write all possible samples of size 2 with replacement (iii) Write the Sampling distribution of the mean (SDM) (iv) Find Mean and Variance of the Sampling distribution of the mean (SDM) (v) Verify the results in (iv) with suitable formula

Solution: (i) Mean and Variance of the population:

$$\text{Mean, } \mu = \frac{1+2+3+4}{4} = \frac{10}{4} = 2.5$$

$$\text{Variance, } \sigma^2 = \frac{1}{n}\sum x^2 - \mu^2 = \frac{1+4+9+16}{4} - \mu^2 \quad = \frac{30}{4} - (2.5)^2 = 7.5 - 6.25 = 1.25$$

(ii) Samples of size 2 with replacement:

The number of samples of size $n = 2$ without replacement is given by $N^n = 4^2 = 16$

The samples are $\{1,1\}$, $\{1, 2\}$, $\{1, 3\}$, $\{1,4\}$,
$\{2,1\}$, $\{2, 2\}$, $\{2, 3\}$, $\{2,4\}$,
$\{3,1\}$, $\{3, 2\}$, $\{3, 3\}$, $\{3,4\}$,
$\{4,1\}$, $\{4, 2\}$, $\{4, 3\}$, $\{4,4\}$

(iii) The Sampling distribution of the mean (SDM):

The means of the above 16 samples are respectively

1, 1.5, 2, 2.5,
1.5, 2, 2.5, 3,
2, 2.5, 3, 3.5,
2.5, 3, 3.5, 4

The frequency distribution of the these sample means is given as follows

Sample mean $\bar{x}$	1	1.5	2	2.5	3	3.5	4
Frequency f	1	2	3	4	3	2	1

Which is the sampling distribution of the mean (SDM)

(iv) Mean and Variance of the Sampling distribution of the mean (SDM)

$$\text{Mean, } \mu_{\bar{X}} = \frac{\sum \bar{x}\, f}{\sum f} = \frac{(1)(1)+(1.5)(2)+(2)(3)+(2.5)(4)+(3)(3)+(3.5)(2)+(4)(1)}{16} = \frac{40}{16} = 2.5$$

$$\text{Variance, } \sigma_{\bar{X}}^2 = \frac{\sum \bar{x}^2 f}{\sum f} - \mu_{\bar{X}}^{\ 2}$$

$$= \frac{(1)^2(1)+(1.5)^2(2)+(2)^2(3)+(2.5)^2(4)+(3)^2(3)+(3.5)^2(2)+(4)^2(1)}{16} - \mu_{\bar{X}}^{\ 2}$$

$$= \frac{110}{16} - (2.5)^2 \qquad = 6.875 - 6.25 \ = 0.625$$

(v) Verification:

We have $\mu_{\bar{X}} = 2.5 = \mu$ and $\dfrac{\sigma^2}{n} = \dfrac{1.25}{2} = 0.625 = \sigma_{\bar{X}}^2$

Therefore, $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \dfrac{\sigma^2}{n}$ are verified.

(7) A random sample of size 100 is taken from an infinite population having the mean 76 and the variance 256. What is the probability that the sample mean will be between 75 and 78

Here $n = 100$, $\mu = 76$, $\sigma^2 = 256$ and $\sigma = 16$, and by central limit theorem, $Z = \dfrac{\bar{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$

$$P(75 < \bar{X} < 78) = P\left(\frac{75-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{\bar{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{78-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right) = P\left(\frac{75-76}{\left(\frac{16}{\sqrt{100}}\right)} < Z < \frac{78-76}{\left(\frac{16}{\sqrt{100}}\right)}\right)$$

$$= P(-0.63 < Z < 1.25) \qquad = F(1.25) - F(-0.63)$$

$$= F(1.25) - [1 - F(0.63)] \qquad = 0.8944 - 1 + 0.7357 \qquad = 0.6301$$

(8) A normal population has mean 0.1 and standard deviation 2.1. Find the probability that the mean of a sample of size 900 will be negative

Here $n = 900$, $\mu = 0.1$ and $\sigma = 2.1$ and by central limit theorem, $Z = \dfrac{\bar{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$

$$P(\bar{X} < 0) = P\left(\frac{\bar{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{0-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right) = P\left(Z < \frac{0-0.1}{\left(\frac{2.1}{\sqrt{900}}\right)}\right) = P(Z < -1.43) = F(-1.43)$$

$$= 1 - F(1.43) \qquad = 1 - 0.9236 \qquad = 0.0764$$

(9) A random sample of size 64 is taken from a normal population with mean 51.4 and standard deviation 6.8. Find the probability that the mean of the sample will (i) exceed 52.9 (ii) fall between 50.5 and 52.3 (iii) less than 50.6

Here $n = 64, \mu = 51.4$ and $\sigma = 6.8$, and by central limit theorem, $Z = \dfrac{\bar{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$

(i) Exceed 52.9, $P(\bar{X} > 52.9) = P\left(\dfrac{\bar{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)} > \dfrac{52.9 - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}\right) = P\left(Z > \dfrac{52.9 - 51.4}{\left(\dfrac{6.8}{\sqrt{64}}\right)}\right)$

$$= P(Z > 1.76) = 1 - P(Z \le 1.76)$$
$$= 1 - F(1.76) = 1 - 0.9608 = 0.0392$$

(ii) Fall between 50.5 and 52.3,

$$P(50.5 < \bar{X} < 52.3) = P\left(\frac{50.5 - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{52.3 - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right) = P\left(\frac{50.5 - 51.4}{\left(\frac{6.8}{\sqrt{64}}\right)} < Z < \frac{52.3 - 51.4}{\left(\frac{6.8}{\sqrt{64}}\right)}\right)$$

$$= P(-1.06 < Z < 1.06) \qquad = F(1.06) - F(-1.06)$$
$$= F(1.06) - [1 - F(1.06)] = 2F(1.06) - 1 = 2(0.8554) - 1 = 0.7108$$

(iii) Less than 50.6 $P(\bar{X} < 50.6) = P\left(\dfrac{\bar{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)} < \dfrac{50.6 - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}\right) = P\left(Z < \dfrac{50.6 - 51.4}{\left(\dfrac{6.8}{\sqrt{64}}\right)}\right)$

$$= P(Z < -0.94) = F(-0.94)$$
$$= 1 - F(0.94) = 1 - 0.8264 = 0.1736$$

(10) If a 1- gallon can of paint covers on an average 513 square feet with a standard deviation of 31.5 square feet, what is the probability that the mean area covered by a sample of 40 of these 1- gallon cans will be anywhere from 510 to 520 square feet

Here $n = 40$, $\mu = 513$ and $\sigma = 31.5$ and by central limit theorem, $Z = \dfrac{\bar{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$

$$P(510 < \bar{X} < 520) = P\left(\frac{510 - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{520 - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right) = P\left(\frac{510 - 513}{\left(\frac{31.5}{\sqrt{40}}\right)} < Z < \frac{520 - 513}{\left(\frac{31.5}{\sqrt{40}}\right)}\right)$$

$$= P(-0.6 < Z < 1.4) = F(1.4) - F(-0.6) = F(1.4) - [1 - F(0.6)]$$
$$= 0.9192 - 1 + 0.7258 = 0.645$$

(11) For a large sample of size n, verify that there is a 50-50 chances that the mean of a random sample from an infinite population with standard deviation σ differ from μ by less than $0.6745 \times \frac{\sigma}{\sqrt{n}}$

We have prove that $P\left(\left|\overline{X}-\mu\right| < 0.6745 \times \frac{\sigma}{\sqrt{n}}\right) = 0.5$

Consider, $P\left(\left|\overline{X}-\mu\right| < 0.6745 \times \frac{\sigma}{\sqrt{n}}\right) = P\left(\left|\frac{\overline{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right| < 0.6745\right) = P(|Z| < 0.6745)$

$$= P(-0.6745 < Z < 0.6745) = F(0.6745) - F(-0.6745)$$
$$= F(0.67) - [1 - F(0.67)] = 0.7486 - [1 - 0.7486] = 0.497$$

(12) If the mean of breaking strength of copper wire is 575 lbs, with a standard deviation of 8.3 lbs. How large a sample must be used in order that there will be one chance in 100 that the mean breaking strength of the sample is less than 572 lbs

Here $\mu = 575$ and $\sigma = 8.3$ By central limit theorem, $Z = \frac{\overline{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$

We have to find the value of n such that $P(\overline{X} < 572) = \frac{1}{100}$

Now $P(\overline{X} < 572) = 0.01$

$$\Rightarrow P\left(\frac{\overline{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{572-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right) = 0.01$$

$$\Rightarrow P\left(Z < \frac{572-575}{\left(\frac{8.3}{\sqrt{n}}\right)}\right) = 0.01$$

$$\Rightarrow P\left(Z < -\frac{3\sqrt{n}}{8.3}\right) = 0.01$$

$$\Rightarrow F\left(-\frac{3\sqrt{n}}{8.3}\right) = 0.01$$

$$\Rightarrow 1 - F\left(\frac{3\sqrt{n}}{8.3}\right) = 0.01$$

$$\Rightarrow F\left(\frac{3\sqrt{n}}{8.3}\right) = 0.99$$

$$\Rightarrow \frac{3\sqrt{n}}{8.3} = 2.33 \quad \Rightarrow n = \left(\frac{2.33 \times 8.3}{3}\right)^2 = 41.55 \approx 42$$

Exercise:

(1) Define (i) Population (ii) Sample (iii) Large sample (iv) Small sample (v) Random sample

(2) Define (i) Sampling (ii) Sampling Distribution of the Mean (iii) finite population correction factor (iv) Standard error of the mean (v) Probable error of the mean

(3) State (i) Chebyshev's Theorem (ii) Central Limit Theorem

(4) Write all possible samples of size 3 without replacement from the population $\{1, 2, 3, 4\}$. Also compute the means of all these samples

(5) Write all possible samples of size 3 without replacement from the population $\{1, 2, 3, 4\}$. Also compute the means of all these samples

(6) Find the finite population correction factor for $n = 5$ and $N = 1000$

(7) Find the finite population correction factor for $n = 25$ and $N = 1000$

(8) What is the effect on standard error, if a sample is taken from an infinite population and its size is decreased from 800 to 200?

(9) What is the effect on standard error, if a sample is taken from an infinite population and its size is increased from 300 to 2700?

(10) A population consists 4 numbers $3, 7, 11, 15$

(i) Find the Mean and Variance of the population

(ii) Write all possible samples of size 2 with replacement

(iii) Write the Sampling distribution of the mean (SDM)

(iv) Find Mean and Variance of the Sampling distribution of the mean (SDM)

(v) Verify the results in (iv) with suitable formula

(11) A population consists 4 numbers $3, 7, 11, 15$

(i) Find the Mean and Variance of the population

(ii) Write all possible samples of size 2 without replacement

(iii) Write the Sampling distribution of the mean (SDM)

(iv) Find Mean and Variance of the Sampling distribution of the mean (SDM)

(v) Verify the results in (iv) with suitable formula

(12) If $\overline{X}$ is the mean of a random sample of size n, taken from the population $1, 2, 3, \cdots N$, find the mean and variance of the Sampling distribution of the mean (SDM)

(13) Construct sampling distribution of means for the population 5, 10, 12, 18 by drawing samples of size 2 without replacement. Find (i) Mean of the population (ii) Standard deviation of the population (iii) Mean of the sampling distribution of means (iv) Standard deviation of the sampling distribution of means

(14) A random sample of size 100 is taken from a normal population with mean 76 and standard deviation 16. Find the probability that the mean of the sample will (i) exceed 77 (ii) fall between 75 and 78

(15) A random sample of size 81 is taken from a normal population with mean 65 and standard deviation 10. Find the probability that the mean of the sample will (i) exceed 67 (ii) fall between 66 and 68

(16) A random sample of size 36 is taken from a normal population with mean 155 and standard deviation 15. Find the probability that the mean of the sample will be (i) less than 157 (ii) in between 153 and 158 (iii) greater than 160

(17) A sample of size 400 is taken from an infinite population with Standard deviation 16. Find the Standard error and Probable error of the mean

(18) If the mean of breaking strength of copper wire is 676 lbs, with a standard deviation of 12 lbs. How large a sample must be used in order that there will be one chance in 100 that the mean breaking strength of the sample is less than 672 lbs

(19) The mean of certain normal population is equal to the standard error of the mean of samples of size 64. Find the probability that the mean of the sample size 36 will be negative

Sampling Distribution of the Mean with unknown σ (SDM with unknown σ):

If we do not know the value of σ , then we cannot use the Central limit theorem $Z = \dfrac{\bar{X}-\mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$.

In this case we use sample standard deviation $'s'$ in place of population standard deviation σ so that we have a random variable different from Z . This new random variable is denoted by t; that is $t = \dfrac{\bar{X}-\mu}{\left(\dfrac{s}{\sqrt{n}}\right)}$.

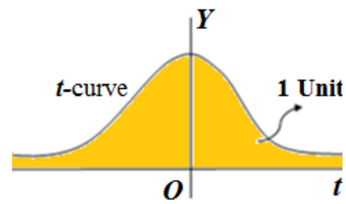
The probability distribution corresponding to this random variable t is called t-distribution with parameter $n-1$. This parameter is known as degrees of freedom, denoted by ν; that is, $\nu = n-1$.

Note: In the t-distribution, for the sample $\{x_1, x_2, x_3, \cdots x_n\}$,

(i) Sample mean $\bar{x}$ is given by $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

(ii) Sample variance is given by $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

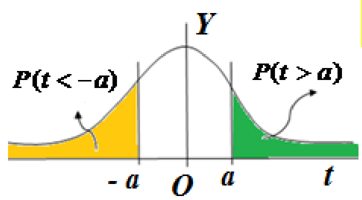
Properties of t - distribution:



(1) The curve given by t-distribution is called t-curve
(2) The t-curve is continuous and above the X-axis (or t- axis)
(3) The curve is symmetric about the Y-axis
(4) The t-curve is similar to Z-curve (standard normal curve)
(5) The area between t - axis and the curve from $-\infty$ to ∞ is 1 unit
(6) The mean of t is 0 and the variance of t is greater than 1; that is, $\mu_t = 0$ and $\sigma_t^2 > 1$
(7) As $n \to \infty$, variance of t tends to 1; that is, $\sigma_t^2 \to 1$. In other words, as $n \to \infty$, $t \to Z$
(8) If n is large ($n \geq 30$), then we can write Z in place of t; that is, $Z = \dfrac{\bar{X}-\mu}{\left(\dfrac{s}{\sqrt{n}}\right)}$

(9) $P(t < -a) = P(t > a)$

(10) $P(t \leq 0) = P(t \geq 0) = \dfrac{1}{2}$



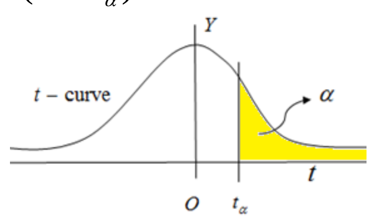
$P(t<-a) = P(t>a)$

t_α - Notation:

If $\alpha \geq 0$ then t_α is a point on t – axis such that $P(t > t_\alpha) = \alpha$ or $P(t < -t_\alpha) = \alpha$

That is, the area between t – axis and the curve from t_α to ∞ is α

(or the area between t – axis and the curve from $-\infty$ to $-t_\alpha$ is α)



$P(t > t_\alpha) = \alpha$

Note:

(1) $t_1 = -\infty$, $t_0 = \infty$, $t_{\frac{1}{2}} = 0$

(2) $t_\alpha + t_{1-\alpha} = 0$ or $-t_\alpha = t_{1-\alpha}$ or $t_\alpha = -t_{1-\alpha}$

Testing a claim using t-distribution: To test a given claim using t-distribution, we follow the rule given below.

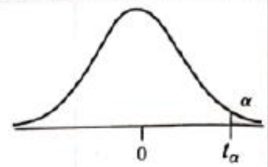
(i) If $|t| < t_{0.005}$ then the claim is accepted

(ii) If $|t| > t_{0.005}$ then the claim is rejected

t_α - Table:

In this table the values of t_α are available for different values of α and ν

Table 4 Values of t_α



ν	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.00833$	$\alpha = 0.00625$	$\alpha = 0.005$	ν
1	3.078	6.314	12.706	31.821	38.204	50.923	63.657	1
2	1.886	2.920	4.303	6.965	7.650	8.860	9.925	2
3	1.638	2.353	3.182	4.541	4.857	5.392	5.841	3
4	1.533	2.132	2.776	3.747	3.961	4.315	4.604	4
5	1.476	2.015	2.571	3.365	3.534	3.810	4.032	5
6	1.440	1.943	2.447	3.143	3.288	3.521	3.707	6
7	1.415	1.895	2.365	2.998	3.128	3.335	3.499	7
8	1.397	1.860	2.306	2.896	3.016	3.206	3.355	8
9	1.383	1.833	2.262	2.821	2.934	3.111	3.250	9
10	1.372	1.812	2.228	2.764	2.870	3.038	3.169	10
11	1.363	1.796	2.201	2.718	2.820	2.891	3.106	11
12	1.356	1.782	2.179	2.681	2.780	2.934	3.055	12
13	1.350	1.771	2.160	2.650	2.746	2.896	3.012	13
14	1.345	1.761	2.145	2.624	2.718	2.864	2.977	14
15	1.341	1.753	2.131	2.602	2.694	2.837	2.947	15

Example:

(i) For $\nu = 7$ (or $n = 8$) and $\alpha = 0.025$, $t_\alpha = t_{0.025} = 2.365$

(ii) For $\nu = 10$ (or $n = 11$) and $\alpha = 0.05$, $t_\alpha = t_{0.05} = 1.812$

(iii) For $\nu = 15$ (or $n = 16$) and $\alpha = 0.005$, $t_\alpha = t_{0.005} = 2.947$

Problems:

(1) Find (i) $t_{0.05}$ when $\nu = 12$ (ii) $-t_{0.01}$ when $\nu = 8$ (iii) $t_{0.995}$ when $\nu = 10$

(i) When $\nu = 12$, $t_{0.05} = 1.782$

(ii) When $\nu = 8$, $-t_{0.01} = -2.896$

(iii) When $\nu = 10$, $t_{0.995} = -t_{1-0.995} = -t_{0.005} = -3.169$ $(\because t_\alpha = -t_{1-\alpha})$

(2) (i) Find $t_{0.975}$ when $\nu = 13$ (ii) Find $-t_{0.99}$ when $\nu = 10$ (iii) Find $t_{0.95}$ when $\nu = 11$

(i) When $\nu = 13$, $t_{0.975} = -t_{1-0.975} = -t_{0.025} = -2.160$ $(\because t_\alpha = -t_{1-\alpha})$

(ii) When $\nu = 10$, $-t_{0.99} = t_{1-0.99} = t_{0.01} = 2.764$ $(\because t_\alpha = -t_{1-\alpha})$

(iii) When $\nu = 11$, $t_{0.95} = -t_{1-0.95} = -t_{0.05} = -1.796$ $(\because t_\alpha = -t_{1-\alpha})$

(3) Find (i) $P(t<2.365)$ when $\nu=7$ (ii) $P(t>1.318)$ when $\nu=24$
(iii) $P(t>-2.567)$ when $\nu=17$ (iv) $P(-1.356<t<2.179)$ when $\nu=12$

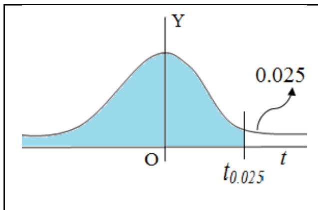
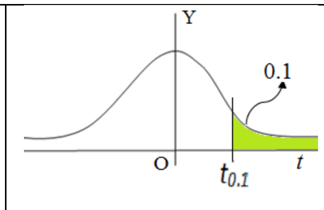
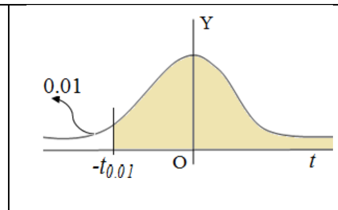
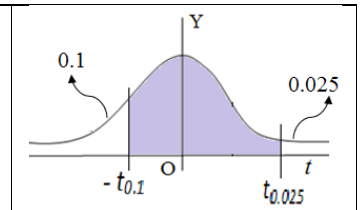
(i) When $\nu=7$, $P(t<2.365)=P(t<t_{0.025})=1-P(t>t_{0.025})=1-0.025=0.975$

(ii) When $\nu=24$, $P(t>1.318)=P(t>t_{0.1})=0.1$

(iii) When $\nu=17$, $P(t>-2.567)=P(t>-t_{0.01})=1-P(t<-t_{0.01})=1-P(t>t_{0.01})=1-0.01=0.99$

(iv) When $\nu=12$, $P(-1.356<t<2.179)=P(-t_{0.1}<t<t_{0.025})=1-(0.1+0.025)=0.875$

$P(-t_{0.1}<t<t_{0.025})=P(t>-t_{0.1})-P(t>t_{0.025})=P(t>t_{0.9})-P(t>t_{0.025})=0.9-0.025=0.875$

<table>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) $P(t<t_{0.025})$</td>
<td>(ii) $P(t>t_{0.1})$</td>
<td>(iii) $P(t>-t_{0.01})$</td>
<td>(iv) $P(-t_{0.1}<t<t_{0.025})$</td>
</tr>
</table>

(4) A random sample of size 25 from a normal population has mean $\bar{x}=47.5$ and the standard deviation $s=8.4$. Does this information tend to support or refute the claim that the mean of the population is $\mu=42.1$

Here $n=25$ (small sample) and $\nu=n-1=24$

Also $\bar{x}=47.5$, $s=8.4$, $\mu=42.1$ and $t_{0.005}=2.797$

We know that given claim is accepted if and only if $|t|<t_{0.005}$

Now $t=\dfrac{\bar{X}-\mu}{\left(\dfrac{s}{\sqrt{n}}\right)}=\dfrac{47.5-42.1}{\left(\dfrac{8.4}{\sqrt{25}}\right)}=3.2143$ and $|t|=3.2143$

Since $|t|>t_{0.005}$, reject the claim

(5) A process for making certain ball bearings is under control if the diameters of the bearings have a mean of 0.5 cm. What can you say about the process if a random sample of 12 of these bearings has a mean diameter of 0.515 cm and standard deviation of 0.017 cm

Here $n=12$ (small sample) and $\nu=n-1=11$

Also $\bar{x}=0.515$, $s=0.017$, $\mu=0.5$ and $t_{0.005}=3.106$

We know that given claim is accepted (the process is under control) if and only if $|t|<t_{0.005}$

Now $t=\dfrac{\bar{X}-\mu}{\left(\dfrac{s}{\sqrt{n}}\right)}=\dfrac{0.515-0.5}{\left(\dfrac{0.017}{\sqrt{12}}\right)}=3.0566$ and $|t|=3.0566$

Since $|t|<t_{0.005}$, accept the claim; that is, the process is under control

Note: In place of $\bar{x}=0.515$, if we take $\bar{x}=0.48$ then $|t|=4.0754$, $|t|>t_{0.005}$ and therefore we need to reject the claim.

(6) The t distribution with 1 degree of freedom is given by $f(t)=\frac{1}{\pi}(1+t^2)^{-1}$ for $-\infty<t<\infty$. Verify the value given for $t_{0.05}$ when $\nu=1$ in the table.

From the tables, when $\nu=1,\ t_{0.05}=6.314$

Now $P(t\geq 6.314)=\int_{6.314}^{\infty} f(t)dt=\frac{1}{\pi}\int_{6.314}^{\infty}\frac{1}{1+t^2}dt=\frac{1}{\pi}\left[\tan^{-1}t\right]_{6.314}^{\infty}$

$=\frac{1}{\pi}\left[\tan^{-1}(\infty)-\tan^{-1}(6.314)\right]=\frac{1}{\pi}\left[\frac{\pi}{2}-1.4137\right]=\frac{7}{22}\left[\frac{11}{7}-1.4137\right]=0.05018\approx 0.05$

Exercise:

(1) (i) Find $t_{0.025}$ when $\nu=14$ (ii) Find $-t_{0.01}$ when $\nu=10$ (iii) Find $t_{0.995}$ when $\nu=7$

(2) (i) Find $t_{0.99}$ when $\nu=6$ (ii) Find $t_{0.975}$ when $\nu=24$ (iii) Find $-t_{0.975}$ when $\nu=19$

(3) Find (i) $P(t>2.821)$ when $\nu=9$ (ii) $P(t<-2.947)$ when $\nu=15$
(iii) $P(t<1.729)$ when $\nu=19$ (iv) $P(-1.714<t<2.069)$ when $\nu=23$

(4) Find k for a sample of size 24 from a normal population such that (i) $P(-2.069<t<k)=0.965$ (ii) $P(k<t<2.807)=0.095$ (iii) $P(-k<t<k)=0.9$

(5) A random sample of size 25 from a normal population has mean $\bar{x}=45.4$ and the standard deviation $s=9.7$. Does this information tend to support or refute the claim that the mean of the population is $\mu=40$

(6) A process for making certain ball bearings is under control if the diameters of the bearings have a mean of 0.5 cm. What can you say about the process if a random sample of 10 of these bearings has a mean diameter of 0.506 cm and standard deviation of 0.004 cm. ($n=10$, $t_{0.005}=3.25$, $|t|=4.7434$ and reject the claim)

(7) The tensile strength (1,000 psi) of a new composite can be modeled as a normal distribution. A random sample of size 25 specimens has mean $\bar{x}=45.3$ and standard deviation $s=7.9$. Does this information tend to support or refute the claim that the mean of the population is $\mu=40.5$?

(8) The process of making concrete in a mixer is under control if the rotations per minute of the mixer have a mean of 22 rpm. What can we say about this process if a sample of 20 of these mixers has a mean rpm of 22.75 rpm and a standard deviation of 3 rpm?

(9) A manufacturer of fuses claims that with 20% overload, the fuses will blow in 12.4 minutes on the average. To test this claim, a sample of 20 of the fuses was subjected to a 20% overload, and the times it took them to blow had a mean of 10.63 minutes and standard deviation of 2.48 minutes. If it can be assumed that the data constitute a random sample from a normal population, do they tend to support or refute the manufacturer's claim

(10) The t distribution with 1 degree of freedom is given by $f(t)=\frac{1}{\pi}(1+t^2)^{-1}$ for $-\infty<t<\infty$. Verify the value given for $t_{0.1}$ when $\nu=1$ in the table.

Sampling Distribution of the Variance (SDV):

The frequency distribution of the variances of all random samples of fixed size, taken from a population is called the Sampling Distribution of the Variance (SDV). It is denoted by S^2.

Examples:

(1) Consider the finite population $\{1,2,3,4\}$ of size $N=4$

The number of samples of size $n=2$ without replacement is given by $^{N}C_n = {}^{4}C_2 = 6$

The samples are $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$, $\{3,4\}$

The means of these 6 samples are respectively 1.5, 2, 2.5, 2.5, 3, 3.5

The variance $= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ or $\frac{1}{n}\sum_{i=1}^{n}x_i^2 - (\bar{x})^2$

The variances of these 6 samples are respectively given as below.

$\frac{1}{2}\left(1^2+2^2\right)-(1.5)^2 = 2.5-2.25 = 0.25$

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

$\frac{1}{2}\left(1^2+4^2\right)-(2.5)^2 = 8.5-6.25 = 2.25$

$\frac{1}{2}\left(2^2+3^2\right)-(2.5)^2 = 6.5-6.25 = 0.25$

$\frac{1}{2}\left(2^2+4^2\right)-(3)^2 = 10-9 = 1$

$\frac{1}{2}\left(3^2+4^2\right)-(3.5)^2 = 12.5-12.25 = 0.25$

The frequency distribution of the these sample variances is given as follows

Sample variance s^2	0.25	1	2.25
Frequency f	3	2	1

Which is the **sampling distribution of the Variance (SDV)**

Theorem: If S^2 is the random variable which gives the variance a random sample of size n, taken from a normal population having variance σ^2, then $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma^2}$ is a random variable having the Chi-square distribution with the parameter $\nu = n-1$.

χ^2 **– Distribution:** A continuous random variable having the probability density function given by $f(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}$ for $x>0$ and $\nu>0$, is called the Chi-square (χ^2) random variable with parameter ν known as degrees of freedom. The probability distribution is called χ^2 – Distribution.

Note:

(i) $\Gamma(n) = \int_0^{\infty} e^{-x}x^{n-1}dx = \int_0^{\infty} e^{-x^2}x^{2n-1}dx$

(ii) $\Gamma(1) = 1$ and $\Gamma(n) = (n-1)!$ for $n = 2, 3, 4, \cdots$

(iii) If $\nu = 2$, then $f(x) = \frac{1}{2\Gamma(1)} e^{-\frac{x}{2}} = \frac{1}{2} e^{-\frac{x}{2}}$ for $x>0$

(iv) If $\nu = 4$, then $f(x) = \dfrac{1}{2^2\Gamma(2)} x e^{-\frac{x}{2}} = \dfrac{1}{4} x e^{-\frac{x}{2}}$ for $x > 0$

(v) If $\nu = 6$, then $f(x) = \dfrac{1}{2^3\Gamma(3)} x^2 e^{-\frac{x}{2}} = \dfrac{1}{16} x^2 e^{-\frac{x}{2}}$ for $x > 0$

Properties of χ^2 – Distribution:

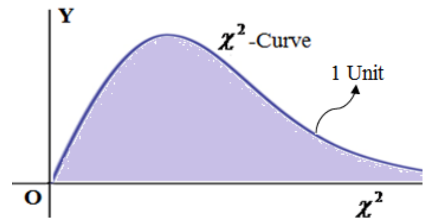
(i) The curve given by the probability density function is called the χ^2 – curve.

(ii) The χ^2 – curve is lies in the 1st quadrant

(iii) It is not symmetric about any axis and has the following shape.

(iv) It depends on the value of ν

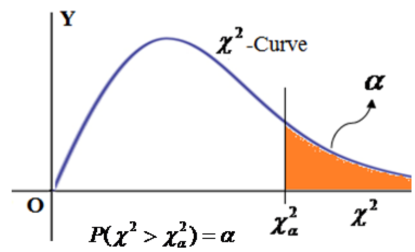
(v) $P(\chi^2 > 0) = 1$; That is, the area between χ^2 – axis and the curve from 0 to ∞ is 1



χ^2_α – Notation:

If $\alpha \geq 0$ then χ^2_α is a point on χ^2 – axis such that $P(\chi^2 > \chi^2_\alpha) = \alpha$

That is, the area between χ^2 – axis and the curve from χ^2_α to ∞ is α



χ^2_α - Table:

In this table the values of χ^2_α are available for different values of α and ν

Table 5 Values of χ^2_α

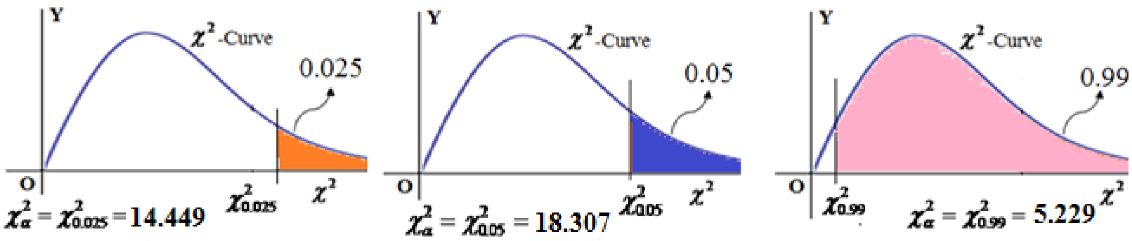
ν	$\alpha = 0.995$	$\alpha = 0.99$	$\alpha = 0.975$	$\alpha = 0.95$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	ν
1	0.0000393	0.000157	0.000982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.833	15.086	16.750	5
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	20

Example:

(i) For $\nu = 6$ (or $n = 7$) and $\alpha = 0.025$, $\qquad \chi^2_\alpha = \chi^2_{0.025} = 14.449$

(ii) For $\nu = 10$ (or $n = 11$) and $\alpha = 0.05$, $\qquad \chi^2_\alpha = \chi^2_{0.05} = 18.307$

(iii) For $\nu = 15$ (or $n = 16$) and $\alpha = 0.99$, $\qquad \chi^2_\alpha = \chi^2_{0.99} = 5.229$



esting a claim using χ^2-distribution: To test a given claim using χ^2-distribution, we follow the rule given below.

(i) If $\chi^2 < \chi^2_{0.005}$ then the claim is accepted

(ii) If $\chi^2 > \chi^2_{0.005}$ then the claim is rejected

Problems:

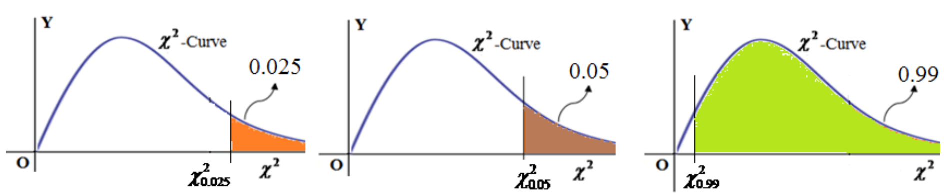
(1) Find (i) $\chi^2_{0.025}$ when $\nu = 12$ (ii) $\chi^2_{0.05}$ when $\nu = 8$ (iii) $\chi^2_{0.99}$ when $\nu = 16$

Solution:

(i) When $\nu = 12$, $\chi^2_{0.025} = 23.337$

(ii) When $\nu = 8$, $\chi^2_{0.05} = 15.507$

(iii) When $\nu = 16$, $\chi^2_{0.99} = 5.812$



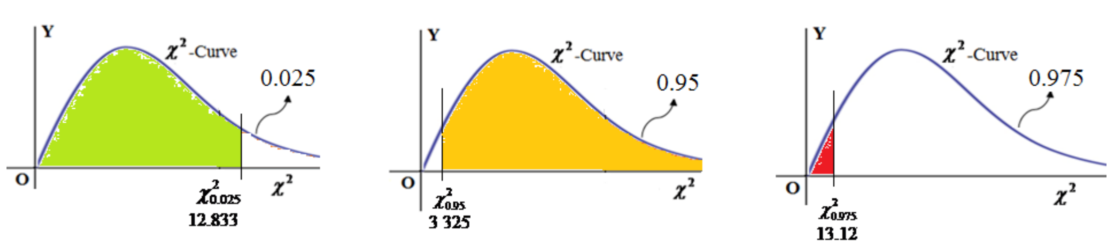
(2) Find (i) $P(\chi^2 < 12.833)$ when $\nu = 5$ (ii) $P(\chi^2 > 3.325)$ when $\nu = 9$ (iii) $P(\chi^2 < 13.12)$ when $\nu = 25$

Solution:

(i) When $\nu = 5$, $P(\chi^2 < 12.833) = P(\chi^2 < \chi^2_{0.025}) = 1 - P(\chi^2 > \chi^2_{0.025}) = 1 - 0.025 = 0.975$

(ii) When $\nu = 9$, $P(\chi^2 > 3.325) = P(\chi^2 > \chi^2_{0.95}) = 0.95$

(iii) When $\nu = 25$, $P(\chi^2 < 13.12) = P(\chi^2 < \chi^2_{0.975}) = 1 - P(\chi^2 > \chi^2_{0.975}) = 1 - 0.975 = 0.025$



(3) Find (i) $P\left(2.088<\chi^2<16.919\right)$ when $\nu=9$ (ii) $P\left(7.564<\chi^2<35.718\right)$ when $\nu=17$

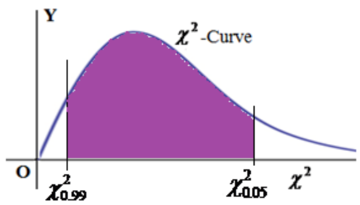
Solution:

(i) When $\nu=9$, $P\left(2.088<\chi^2<16.919\right)=P\left(\chi^2_{0.99}<\chi^2<\chi^2_{0.05}\right)$

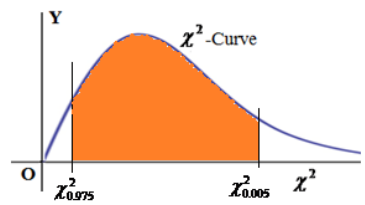
$$=P\left(\chi^2>\chi^2_{0.99}\right)-P\left(\chi^2>\chi^2_{0.05}\right)=0.99-0.05=0.94$$

(ii) When $\nu=17$, $P\left(7.564<\chi^2<35.718\right)=P\left(\chi^2_{0.975}<\chi^2<\chi^2_{0.005}\right)$

$$=P\left(\chi^2>\chi^2_{0.975}\right)-P\left(\chi^2>\chi^2_{0.005}\right)=0.975-0.005=0.97$$



$P\left(\chi^2_{0.99}<\chi^2<\chi^2_{0.05}\right)=0.99 - 0.05 =0.94$



$P\left(\chi^2_{0.975}<\chi^2<\chi^2_{0.005}\right)= 0.975 - 0.005 = 0.97$

(4) The claim that the variance of a normal population $\sigma^2=21.3$ is rejected if the variance of a random sample of size 15 exceeds 39.74. What is the probability that the claim will be rejected even though $\sigma^2=21.3$.

Solution: Here $n=15, \nu=n-1=14$ and $\sigma^2=21.3$

Given that the claim is rejected if $S^2>39.74$

The probability that the claim will be rejected is given by

$$P(S^2>39.74)=P\left(\frac{(n-1)S^2}{\sigma^2}>39.74\frac{(n-1)}{\sigma^2}\right)=P\left(\chi^2>\frac{39.74\times 14}{21.3}\right)$$

$$=P\left(\chi^2>26.120\right)=P\left(\chi^2>\chi^2_{0.025}\right)=0.025$$

(5) The claim that the variance of a normal population $\sigma^2=18.5$ is rejected if the standard deviation of a random sample of size 25 exceeds 5.7559. What is the probability that the claim will be rejected even though $\sigma^2=18.5$.

Solution: Here $n=25, \nu=n-1=24$ and $\sigma^2=18.5$

Given that the claim is rejected if $S>5.7559$

The probability that the claim will be rejected is given by

$$P(S>5.7559)=P(S^2>33.1304)=P\left(\frac{(n-1)S^2}{\sigma^2}>33.1304\frac{(n-1)}{\sigma^2}\right)=P\left(\chi^2>\frac{33.1304\times 24}{18.5}\right)$$

$$=P\left(\chi^2>42.9799\right)=P\left(\chi^2>\chi^2_{0.01}\right)=0.01$$

(6) A random sample of 10 observations is taken from a normal population having the variance $\sigma^2=42.5$. Find the approximate probability of obtaining a sample standard deviation between 3.14 and 8.94.

Solution: Here $n=10$ and $\nu=n-1=9$. Also $\sigma^2=42.5$, we know that $\chi^2=\dfrac{(n-1)S^2}{\sigma^2}$

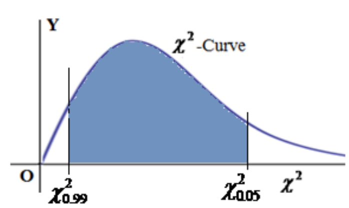
The probability of obtaining a sample standard deviation between 3.14 and 8.94 is given by $P(3.14<S<8.94)=P(9.8596<S^2<79.9236)$

$$=P\left(9.8596\frac{(n-1)}{\sigma^2}<\frac{(n-1)S^2}{\sigma^2}<79.9236\frac{(n-1)}{\sigma^2}\right)$$

$$=P\left(\frac{9.8596\times 9}{42.5}<\chi^2<\frac{79.9236\times 9}{42.5}\right)$$

$$=P\left(2.0879<\chi^2<16.9249\right)$$

$$=P\left(\chi^2_{0.99}<\chi^2<\chi^2_{0.05}\right)=P\left(\chi^2>\chi^2_{0.99}\right)-P\left(\chi^2>\chi^2_{0.05}\right)=0.99-0.05=0.94$$



$P\left(\chi^2_{0.99}<\chi^2<\chi^2_{0.05}\right)=0.99 - 0.05 =0.94$

(7) The Chi-square distribution with 4 degrees of freedom is given by $f(x)=\begin{cases}\frac{1}{4}xe^{-\frac{x}{2}} & x>0\\ 0 & x\le 0\end{cases}$

Find the probability that the variance of a random sample with $\sigma=12$ will exceed 180.

Solution: Here $\nu=4$ and $\sigma=12$

The probability that the variance of a random sample will exceed 180 is given by

$$P(S^2>180)=P\left(\frac{(n-1)S^2}{\sigma^2}>180\frac{(n-1)}{\sigma^2}\right)\quad=P\left(\chi^2>\frac{180\times 4}{144}\right)\quad=P\left(\chi^2>5\right)$$

$$=\int_5^\infty f(x)dx=\int_5^\infty \frac{1}{4}xe^{-\frac{x}{2}}dx=\frac{1}{4}\left[(x)\left(\frac{e^{-\frac{x}{2}}}{\left(-\frac{1}{2}\right)}\right)-(1)\left(\frac{e^{-\frac{x}{2}}}{\left(-\frac{1}{2}\right)^2}\right)\right]_5^\infty$$

$$=\frac{1}{4}\left[(0)-\left(-10e^{-\frac{5}{2}}-4e^{-\frac{5}{2}}\right)\right]=\frac{7}{2}e^{-\frac{5}{2}}$$

Exercise:

(1) Find (i) $\chi^2_{0.05}$ when $\nu=6$ (ii) $\chi^2_{0.99}$ when $\nu=10$ (iii) $\chi^2_{0.025}$ when $\nu=14$ (iv) $\chi^2_{0.01}$ when $\nu=21$

(2) Determine the probabilities: (i) $P(\chi^2<27.688)$ when $\nu=13$ (ii) $P(\chi^2>18.475)$ when $\nu=7$ (iii) $P(\chi^2<14.256)$ when $\nu=29$

(3) Find (i) $P\left(12.461<\chi^2<48.278\right)$ when $\nu=28$ (ii) $P\left(28.685<\chi^2<29.141\right)$ when $\nu=14$

(4) A random sample of 12 observations is taken from a normal population having the variance $\sigma^2=42.5$. Find the approximate probability of obtaining a sample variance between 10.057 and 84.691. (Hint: $P(10.057<s^2<84.691)=P\left(2.6029<\chi^2<21.92\right)=P\left(\chi^2_{0.995}<\chi^2<\chi^2_{0.025}\right)=0.97$)

(5) The claim that the variance of a normal population $\sigma^2=4$ is rejected if the variance of a random sample of size 9 exceeds 7.7535. What is the probability that the claim will be rejected even though $\sigma^2=4$.

(6) A random sample of 15 observations is taken from a normal population having variance $\sigma^2=90.25$. Find the approximate probability of obtaining a sample standard deviation between 7.25 and 10.75.

(7) A manufacturer claims that any of his lists of items cannot have variance more than 1. A sample of 25 items has a variance of 1.2 . Test whether the claim of the manufacturer is correct.

(8) The Chi-square distribution with 2 degrees of freedom is given by $f(x)=\begin{cases}\frac{1}{2}e^{-\frac{x}{2}} & x>0\\ 0 & x\le 0\end{cases}$

Find the probability that the standard deviation of a random sample with $\sigma=10$ will exceed $10\sqrt{2}$.

(Hint: $P(S>10\sqrt{2})=P(S^2>200)=P\left(\chi^2>4\right)=\int_4^\infty f(x)dx$)

$F-$ **Distribution:** A continuous random variable having the probability density function given by $f(x)=\dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\dfrac{x^{\frac{\nu_1}{2}-1}}{\left(1+\frac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}}$ for $x>0,\ \nu_1>0$ and $\nu_2>0$, is called the $F-$random variable with parameters ν_1 and ν_2 known as degrees of freedoms. The probability distribution is called $F-$ Distribution.

Note:

(i) $\Gamma(n)=\int_0^\infty e^{-x}x^{n-1}dx=\int_0^\infty e^{-x^2}x^{2n-1}dx$

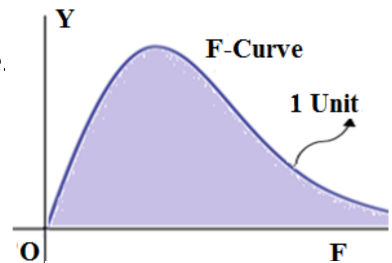
(ii) $\Gamma(1)=1$ and $\Gamma(n)=(n-1)!$ for $n=2,3,4,\cdots$

(iii) If $\nu_1=2$ and $\nu_2=2$, then $f(x)=\dfrac{1}{(1+x)^2}$ for $x>0$

(iv) If $\nu_1=4$ and $\nu_2=4$, then $f(x)=\dfrac{6x}{(1+x)^4}$ for $x>0$

Properties of $F-$Distribution:

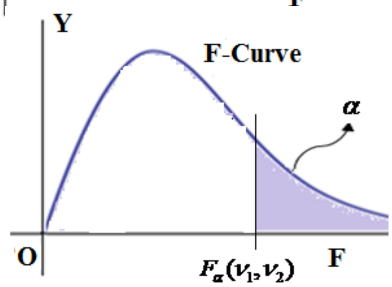
(i) The curve given by the probability density function is called the $F-$curve.
(ii) The $F-$curve is lies in the 1st quadrant
(iii) It is not symmetric about any axis and has the following shape.
(iv) It depends on the values of ν_1 and ν_2
(v) $P(F>0)=1$; That is, the area between F –axis and the curve from 0 to ∞ is 1



$F_\alpha-$**Notation:**

If $\alpha\geq 0$ then F_α is a point on F –axis such that $P(F>F_\alpha)=\alpha$
That is, the area between F –axis and the curve from F_α to ∞ is α.
The value of F_α corresponding to ν_1 and ν_2 is denoted by $F_\alpha(\nu_1,\nu_2)$

Note: $F_\alpha(\nu_1,\nu_2)=\dfrac{1}{F_{1-\alpha}(\nu_2,\nu_1)}$



Theorem: If S_1^2 and S_2^2 are the variances of independent random samples of sizes n_1 and n_2, taken from two normal populations having the same variance, then $F=\dfrac{S_1^2}{S_2^2}$ is a random variable having the $F-$ distribution with the parameters $\nu_1=n_1-1$ and $\nu_2=n_2-1$.

Note: If $F=\dfrac{S_2^2}{S_1^2}$, then the parameters are in the order of $\nu_2=n_2-1$ and $\nu_1=n_1-1$

$F_{0.05}$ – **Tables:** In this table the values of $F_{0.05}(\nu_1, \nu_2)$ are available for different values of ν_1 and ν_2

$F_{0.01}$ – **Tables:** In this table the values of $F_{0.01}(\nu_1, \nu_2)$ are available for different values of ν_1 and ν_2

Table 6(a) Values of $F_{0.05}$

ν_2 = Degrees of Freedom for Denominator	ν_1 = Degrees of Freedom for Numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.50	2.47	2.38	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71

Table 6(b) Values of $F_{0.01}$

ν_2 = Degrees of Freedom for Denominator	ν_1 = Degrees of Freedom for Numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,240	6,261	6,287	6,313	6,339	6,366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.57	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.58	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.91	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.45	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.26	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.31	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.76	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.57	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.41	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.16	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	2.98	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.91	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.84	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.79	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.73	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.69	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.64	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.60	2.54	2.45	2.36	2.27	2.17

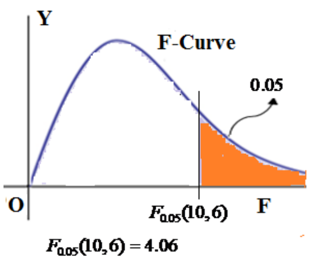
Example:

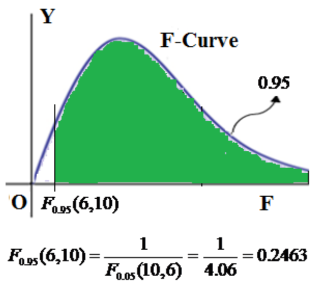
(i) For $\nu_1 = 10, \nu_2 = 6$ and $\alpha = 0.05$, $F_{0.05}(\nu_1, \nu_2) = F_{0.05}(10, 6) = 4.06$

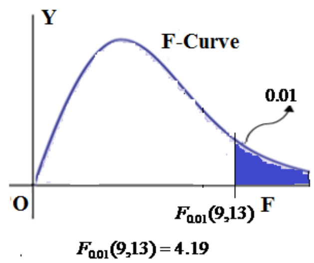
(ii) For $\nu_1 = 6, \nu_2 = 10$ and $\alpha = 0.95$, $F_{0.95}(\nu_1, \nu_2) = F_{0.95}(6,10) = \dfrac{1}{F_{1-0.95}(10,6)} = \dfrac{1}{F_{0.05}(10,6)} = \dfrac{1}{4.06}$

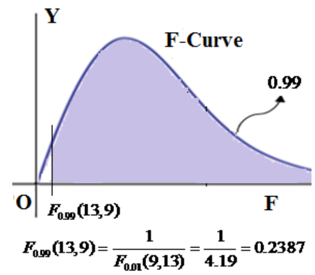
(iii) For $\nu_1 = 9, \nu_2 = 13$ and $\alpha = 0.01$, $F_{0.01}(\nu_1, \nu_2) = F_{0.01}(9,13) = 4.19$

(iv) For $\nu_1 = 13, \nu_2 = 9$ and $\alpha = 0.99$, $F_{0.99}(\nu_1, \nu_2) = F_{0.99}(13,9) = \dfrac{1}{F_{1-0.99}(9,13)} = \dfrac{1}{F_{0.01}(9,13)} = \dfrac{1}{4.19}$









Problems:

(1) For an $F-$distribution find

(i) $F_{0.05}$ with $\nu_1 = 7$ and $\nu_2 = 15$

(ii) $F_{0.01}$ with $\nu_1 = 25$ and $\nu_2 = 19$

(iii) $F_{0.95}$ with $\nu_1 = 19$ and $\nu_2 = 25$

(iv) $F_{0.99}$ with $\nu_1 = 22$ and $\nu_2 = 12$

Solution:

(i) $F_{0.05}(7,15) = 2.71$

(ii) $F_{0.01}(25,19) = 2.91$

(iii) $F_{0.95}(19,25) = \dfrac{1}{F_{1-0.95}(25,19)} = \dfrac{1}{F_{0.05}(25,19)} = \dfrac{1}{2.11} = 0.4739$

(iv) $F_{0.99}(22,12) = \dfrac{1}{F_{1-0.99}(12,22)} = \dfrac{1}{F_{0.01}(12,22)} = \dfrac{1}{3.12} = 0.3205$

(2) If two independent random samples of size $n_1 = 7$ and $n_2 = 13$ are taken from a normal population, what is the probability that the variance of the first sample will be at least three times as large as that of the second sample?

Solution:

Here $n_1 = 7$ and $n_2 = 13$, so $\nu_1 = 6$ and $\nu_2 = 12$

Therefore, the probability that the variance of the first sample will be at least three times as large as that of the second sample is given by

$$P(S_1^2 \ge 3S_2^2) = P\left(\frac{S_1^2}{S_2^2} \ge 3\right) = P(F \ge 3) = P(F \ge F_{0.05}) = 0.05 \qquad (\because F_{0.05}(6,12) = 3)$$

(3) If two independent random samples of size $n_1 = 8$ and $n_2 = 8$ are taken from a normal population, what is the probability that the variance of the first sample will be at least seven times as large as that of the second sample?

Solution:

Here $n_1 = 8$ and $n_2 = 8$, so $\nu_1 = 7$ and $\nu_2 = 7$

Therefore, the probability that the variance of the first sample will be at least seven times as large as that of the second sample is given by

$$P(S_1^2 \geq 7S_2^2) = P\left(\frac{S_1^2}{S_2^2} \geq 7\right) = P(F \geq 7) = P(F \geq F_{0.01}) = 0.01 \qquad (\because F_{0.01}(7,7) = 6.99 \approx 7)$$

(4) If two independent random samples of size $n_1 = 13$ and $n_2 = 26$ are taken from a normal population, what is the probability that the variance of the second sample will be at least 2.5 times as large as that of the first sample?

Solution:

Here $n_1 = 13$ and $n_2 = 26$, so $\nu_1 = 12$ and $\nu_2 = 25$

Therefore, the probability that the variance of the second sample will be at least 2.5 times as large as that of the first sample is given by

$$P(S_2^2 \geq 2.5\, S_1^2) = P\left(\frac{S_2^2}{S_1^2} \geq 2.5\right) = P(F \geq 2.5) = P(F \geq F_{0.05}) = 0.05 \qquad (\because F_{0.05}(25,12) = 2.5)$$

Problems:

(1) For an $F-$distribution find

(i) $F_{0.05}$ with $\nu_1 = 20$ and $\nu_2 = 10$

(ii) $F_{0.01}$ with $\nu_1 = 20$ and $\nu_2 = 5$

(iii) $F_{0.95}$ with $\nu_1 = 15$ and $\nu_2 = 12$

(iv) $F_{0.95}$ with $\nu_1 = 12$ and $\nu_2 = 15$

(v) $F_{0.99}$ with $\nu_1 = 5$ and $\nu_2 = 20$

(vi) $F_{0.99}$ with $\nu_1 = 20$ and $\nu_2 = 5$

(2) If two independent random samples of size $n_1 = 26$ and $n_2 = 8$ are taken from a normal population, what is the probability that the variance of the second sample will be at least 2.4 times as large as that of the first sample?

(3) If two independent random samples of size $n_1 = 9$ 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 four times as large as that of the second sample?

(4) If two independent random samples of size $n_1 = 13$ and $n_2 = 7$ are taken from a normal population, what is the probability that the variance of the first sample will be at least four times as large as that of the second sample?