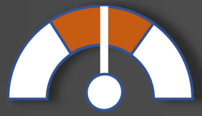
Confidence Intervals

Become Future Fit

You will learn

Learn what are Confidence Intervals & why should you know about it

Level of Difficulty



Medium

Confidence Intervals

It represents the uncertainty in the estimation of true mean(or even the variation).

Confidence Intervals

Sample statistic ± [___ X ___]

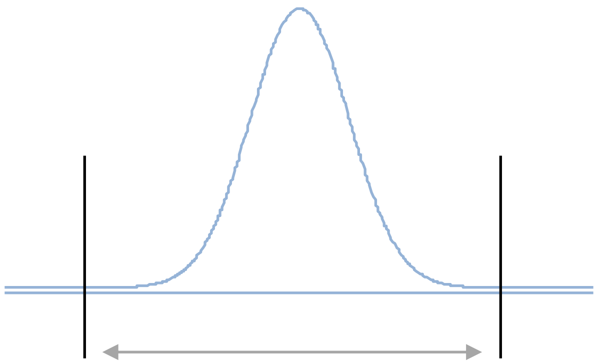
- e.g., $\bar{x}$, s, …
- Confidence factor
- Measure of variability

In some cases the uncertainty is not symmetrical and the ‘+’ term is different from the ‘-’ term; e.g., for σ.

Confidence Intervals

Confidence Intervals are nothing but upper and lower limits of a band within which we find the true mean.

Area under the Normal Curve



99% = 3σ

95.45 % = 2σ

% Area under Curve = $Z\sigma$

Central Limit Theorem

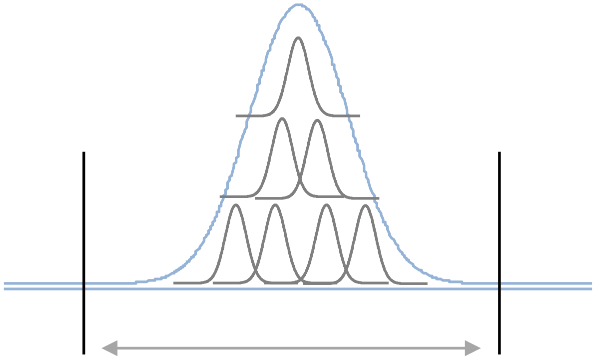
The standard deviation of the SAMPLE MEAN:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Standard Error of Mean (SE)

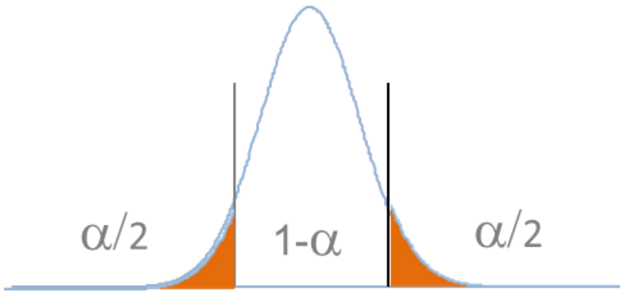
Confidence Intervals

Population Individual Values Vs Sample Means Values



% Area under Curve = $Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

Confidence Intervals



Population Individual Values Vs Sample Means Values

% Area under Cuve = $Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

Confidence Interval for Means

Become Future Fit

Business Management Group

You will learn

Learn how to compute Confidence Intervals for means

Level of Difficulty

Medium

CI for Mean of Continuous Data

$$\mathbf{CI} = \overline{\mathbf{X}} \pm \mathbf{Z}_{\alpha/2}\left(\frac{\boldsymbol{\sigma}}{\sqrt{\mathbf{n}}}\right)$$

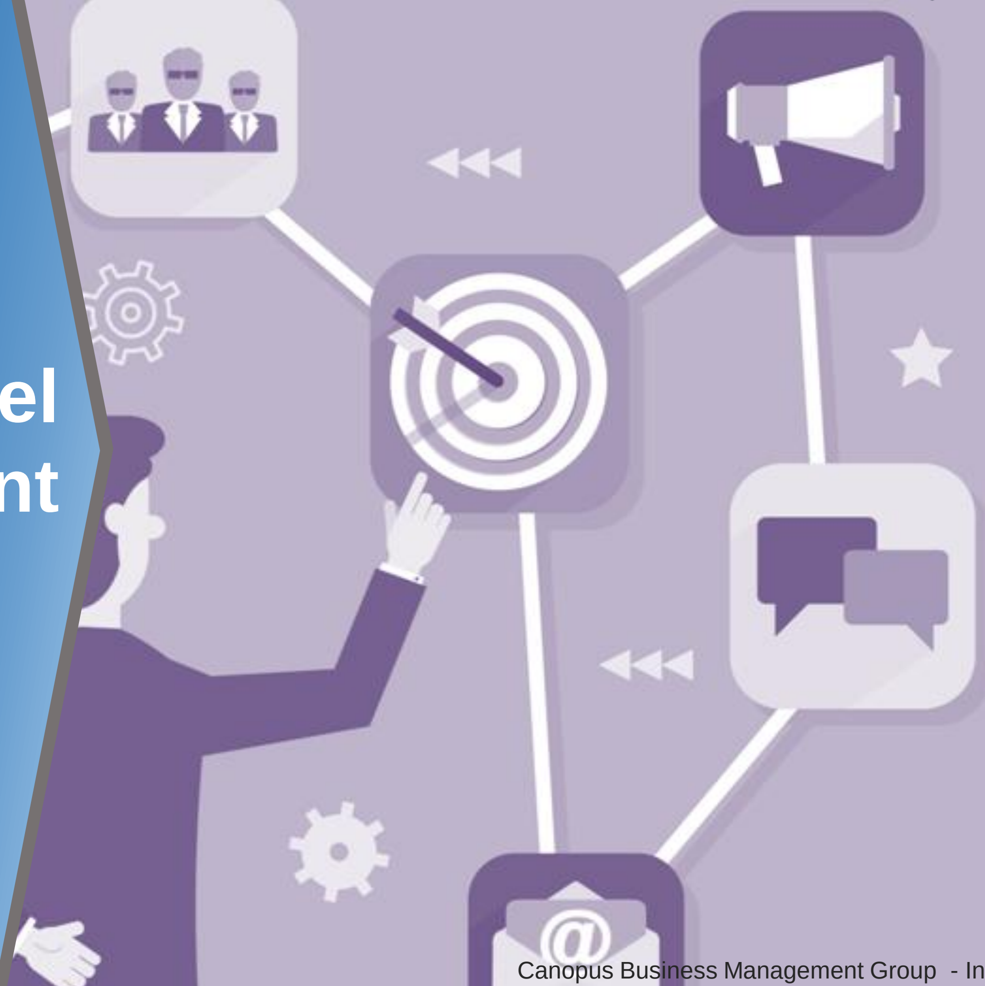
$Z_{\alpha/2}$ = normal distribution value for a given confidence level

$\overline{X}$ = mean of data

σ = population standard deviation

n = sample size

Competency Level Assessment

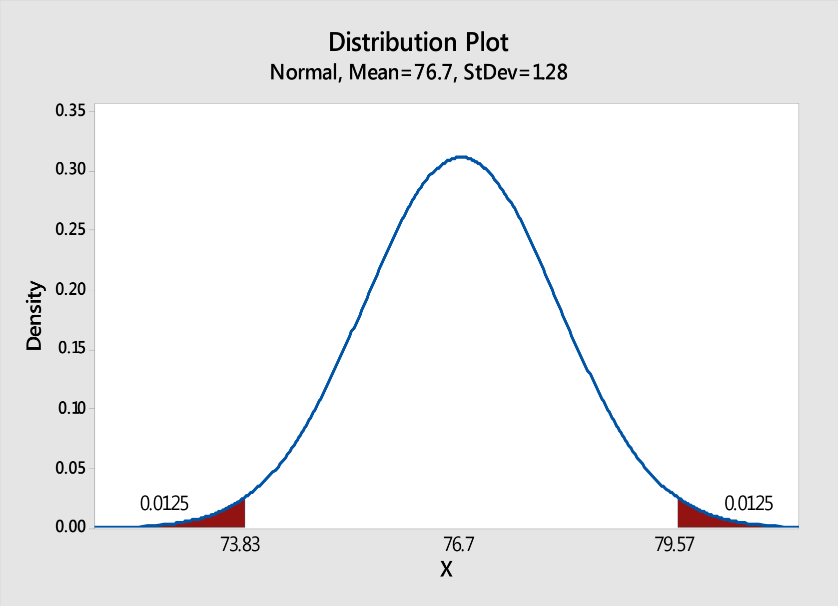


Competency Level Assessment

Company has established norms for the competency of their executives in aptitude test related to a specific subject. The historic population data suggests that average 73.2 with a standard deviation of 8.6.

If 45 randomly selected persons and have a average 76.7, find the confidence interval of the scores at 95% level of confidence

CI for Mean of Continuous Data



CI for Mean of Continuous Data

- This formula only applies when σ is known, which is rare.
- If the sample size is large (exceeds 50), it is a good approximation.

Degrees of Freedom

Become Future Fit

You will learn

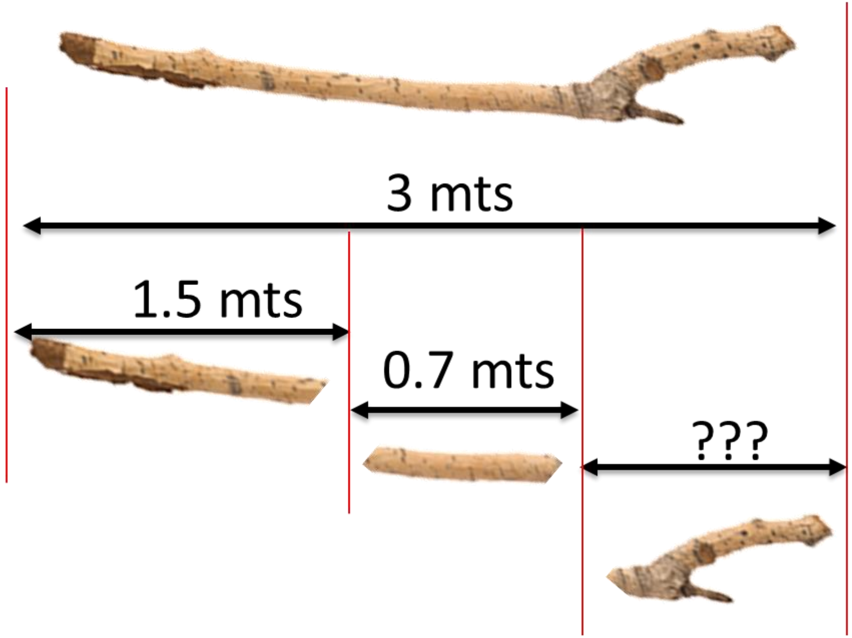
Learn about what is Degrees of Freedom &

Where will it be applied

Level of Difficulty

Medium

Degrees of Freedom



Degrees of Freedom

The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom

Degrees of Freedom

DoF is the number of chances we have to determine the characteristics of a system.

Degrees of Freedom

Example

- If there are 50 samples, then the degree of freedom will be 49
- If there are 5 groups/levels, the degree of freedom will be 4

Sampling Distribution

1. t Distribution
2. F Distribution
3. Chi-Square Distribution

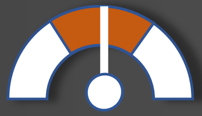
t Distribution

Become Future Fit

You will learn

Learn about t distribution & its relevance to data analysis

Level of Difficulty



Medium

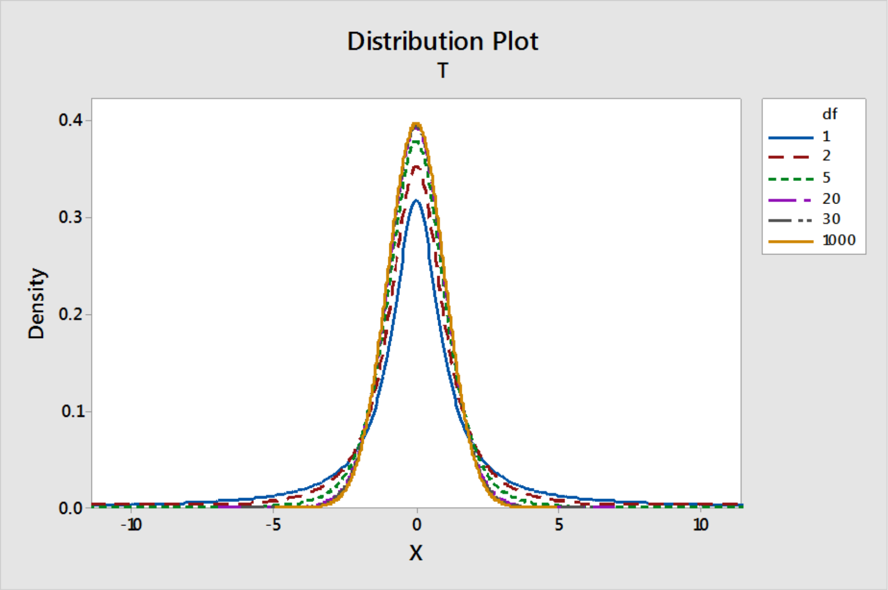
When the Sample Size < 30

&

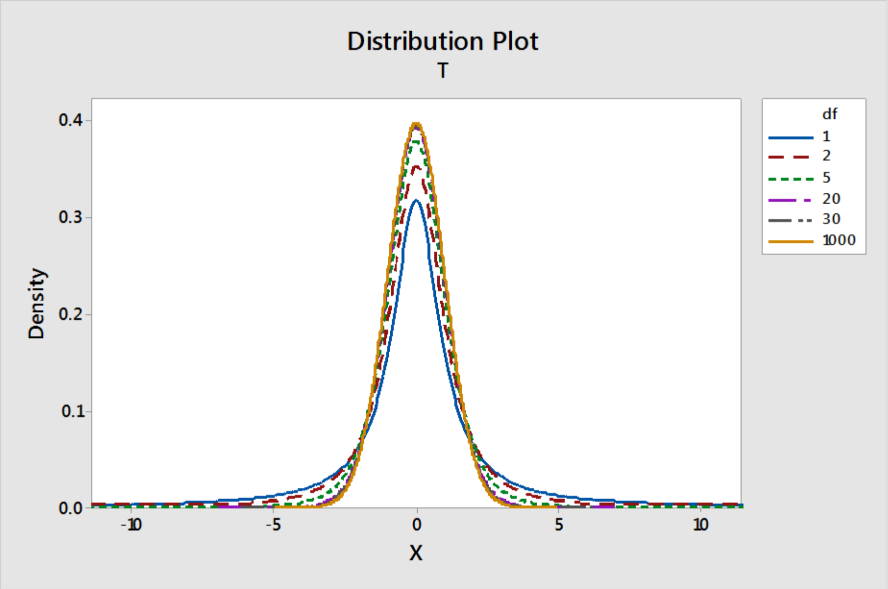
Population SD is not known

Students t-Distribution

t-Distribution

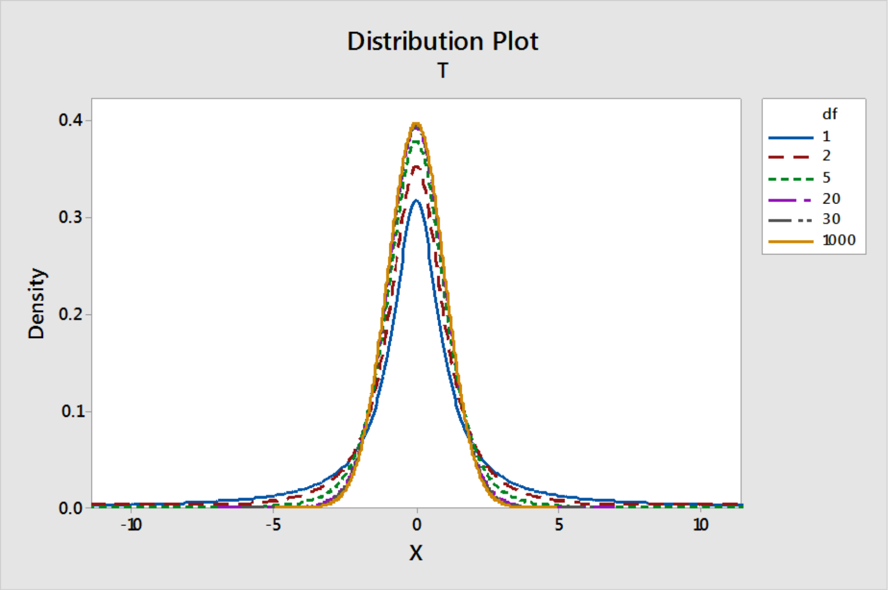


t-Distribution



- Looks like Normal Distribution
- Has fatter tails

t-Distribution



Developed in 1908 by William Gosset

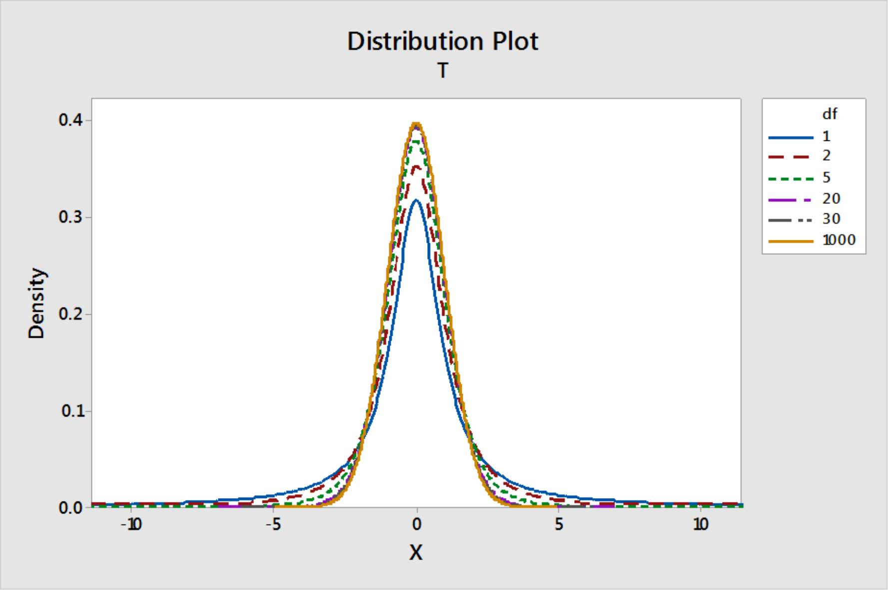
Properties

- Mean of t-distribution = 0
- Variance = $\nu / (\nu - 2)$,
- Where ν is the degrees of freedom
- t-distribution Shape & Scale parameters are dependent on Degrees of Freedom

Properties

- Always Variance > 1
- When ν ~ Large
  - Variation ~ 1 $[\nu / (\nu - 2)]$
  - t Distribution → Standard Normal Distribution

Fatter Tails



t- statistic

Sample Size	t-value ($\alpha/2$ =.025)
2	12.71
3	4.30
5	2.78
10	2.26
20	2.09
30	2.05
100	1.98
1000	1.96

When to use t-Distribution

Use t distribution when population distribution is normal or any bell-shaped distribution & sampling distribution has below characteristics

Sample	Characteristics
< 15	Sampling distribution is symmetric, unimodal, no outliers
16~40	Sampling distribution is moderately skewed, unimodal, without outliers
> 40	Sampling distribution doesn’t have outliers.

Don’t use t distribution when population is not approximately normal

Confidence Intervals for t Distribution

Become Future Fit

You will learn

Learn to compute the Confidence Limits for sample mean using t-distribution

Level of Difficulty

Medium

Confidence Intervals for Sample Data

Inspection of 10 items gave the following measurements :

0.983, 1.002, 0.998, 0.996, 1.002, 0.983, 0.994, 0.991, 1.005 & 0.986.

What is the likely variation in the measurements with 99% confidence ?

CI with t Distribution

Canopus Business Management Group

Formula

$$CI = \overline{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

$\overline{X}$ = Mean

s = Standard deviation

n = Sample size

υ = Degrees of freedom, used in some tables & calculated as n-1 for this test.

$t_{\alpha/2,n-1}$ = Value from t distribution

Confidence Interval for Proportions

Become Future Fit

Business Management Group

You will learn

Learn to compute Confidence Interval for proportion data

Level of Difficulty

Medium

Scenario

An internal auditor randomly selects 200 documents for an audit. He finds 5 defective documents. Calculate a 90% confidence interval for the proportion of good documents.

Assumptions

If n> 5/min(p,q)

Binomial tends to Normal

Mean (μ) = n*p ; SD (σ) = sqrt[n*p*q]

CI for Proportions

Discrete Data

$$\bar{p} \pm Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Where…

$\bar{p}$ = average proportion seen in the sample

n = sample size

α = risk

CI for Proportions

Discrete Data

$$\bar{p} \pm Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Binomial Approximates to Normal Distribution for large Samples.

Use for large sample sizes

Chi-square Distribution

Become Future Fit

You will learn

Learn about Chi-square distribution & its relevance to data analysis

Level of Difficulty

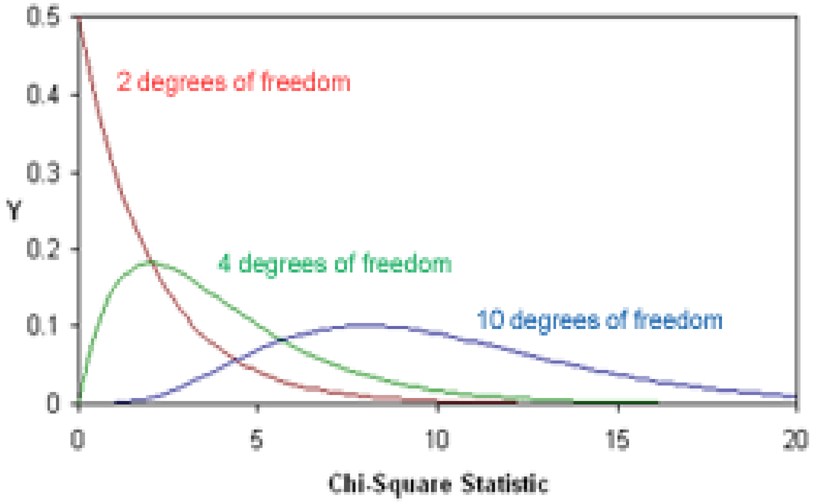
Medium

Chi-Squared Distribution

If a process has cycle time data that is normal, then the SUM OF SQUARES of its standard normal variable ($\mu = 0$ and $\sigma = 1$) forms a Chi-square distribution

Chi-Squared Distribution

Chi-square is a family of distributions



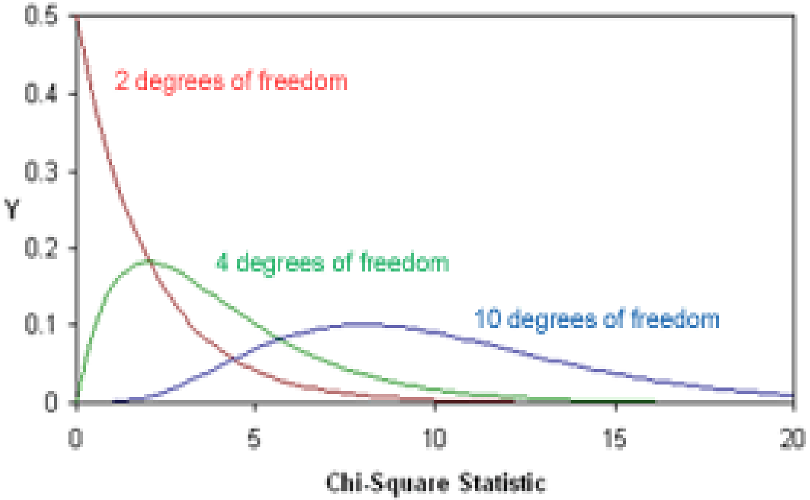
Chi-Squared Application

- For characterizing & analyzing dispersion of any data (Sum of Squares)
- For both Continuous & Discrete data:
  - Variance (Goodness of Fit)
  - Observed Vs Expected (Contingency Tables)

Properties

- Chi-square distribution has one parameter, degrees of freedom(ν)
  - Mean = ν
  - Standard Deviation = Sqrt(2ν)
- Chi-square is always a positive value

Properties



- As Degree of Freedom increases, Chi-square distribution tends to normal curve

CI for Variance

$$\frac{(n-1)s^2}{\chi^2_{(\alpha/2),(n-1)}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2),(n-1)}}$$

Where…

n = sample size

s^2 = variance

α = risk

CI will not be symmetrical on both sides

The standard deviation of process parameter is 0.86 mins as ascertained from 12 samples. Estimate 95% CI for Standard Deviation of this data.

Descriptive Statistics

N	StDev	Variance	95% CI for σ using Chi-Square
12	0.860	0.740	(0.609, 1.460)

F Distribution

Become Future Fit

You will learn

Learn about F Distribution and its relevance to data analysis

Level of Difficulty

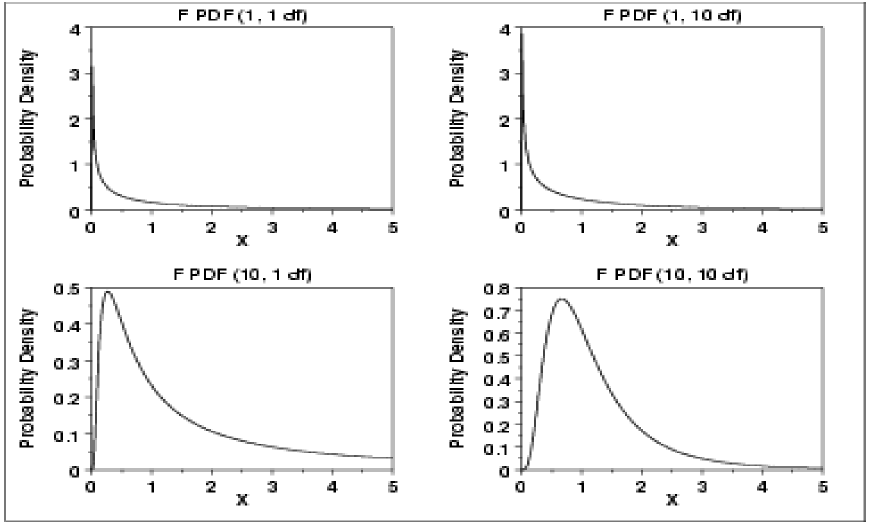
Medium

F-Distribution

The F distribution is the ratio of two chi-square distributions with degrees of freedom (n1-1) and (n2-1) respectively

F-Distribution

F Distribution is a family of distributions



F-Distribution

The ratio of two $\chi 2$ distribution with different degrees of freedom form a F distribution as below:

$$F_{\nu_1,\nu_2} = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2} \quad (\nu_1 \le \nu_2)$$

F-Distribution

Ratio of two sample variances from the same normally distributed population

$$F_{\nu_1,\nu_2} = \frac{s_1^2}{s_2^2} \quad (\nu_1 \le \nu_2)$$

Application

- Used to compare variances of different samples
- Many statistical tests are based on F test