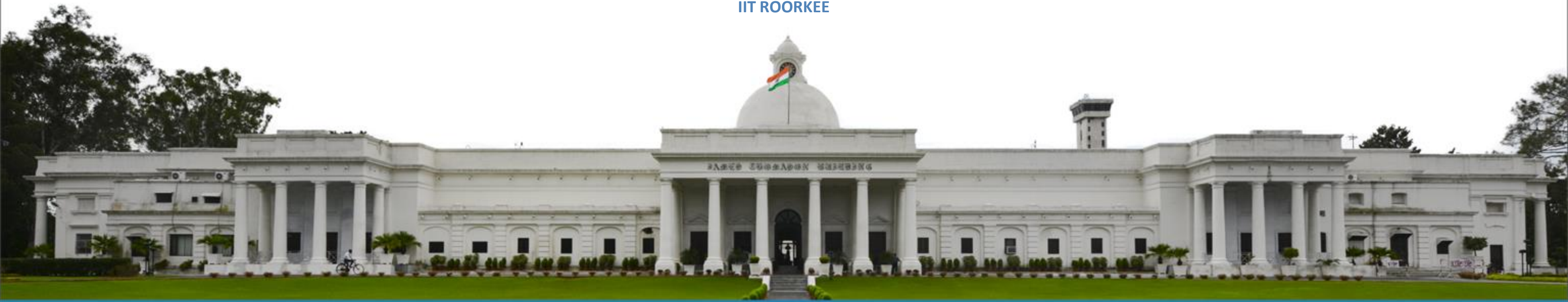
IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

Confidence Interval Estimation: Single Population-II

Dr. A. Ramesh
Department of Management Studies
IIT ROORKEE



Student's t Distribution

- Consider a random sample of n observations
  - with mean $\bar{x}$ and standard deviation s
  - from a normally distributed population with mean μ

- Then the variable

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

follows the Student's t distribution with (n - 1) degrees of freedom

Confidence Interval for μ (σ^2 Unknown)

- If the population standard deviation σ is unknown, we can substitute the sample standard deviation, s
- This introduces extra uncertainty, since s is variable from sample to sample
- So we use the t distribution instead of the normal distribution

Confidence Interval for μ (σ Unknown)

(continued)

- Assumptions
  - Population standard deviation is unknown
  - Population is normally distributed
  - If population is not normal, use large sample
- Use Student's t Distribution
- Confidence Interval Estimate:

$$\boxed{\bar{x} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}}$$

where $t_{n-1,\alpha/2}$ is the critical value of the t distribution with n-1 d.f. and an area of $\alpha/2$ in each tail

Margin of Error

- The confidence interval,

$$\boxed{\bar{x} - t_{n-1,\alpha/2} \frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}}$$

- Can also be written as $\boxed{\bar{x} \pm ME}$

  where ME is called the margin of error:

$$\boxed{ME = t_{n-1,\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

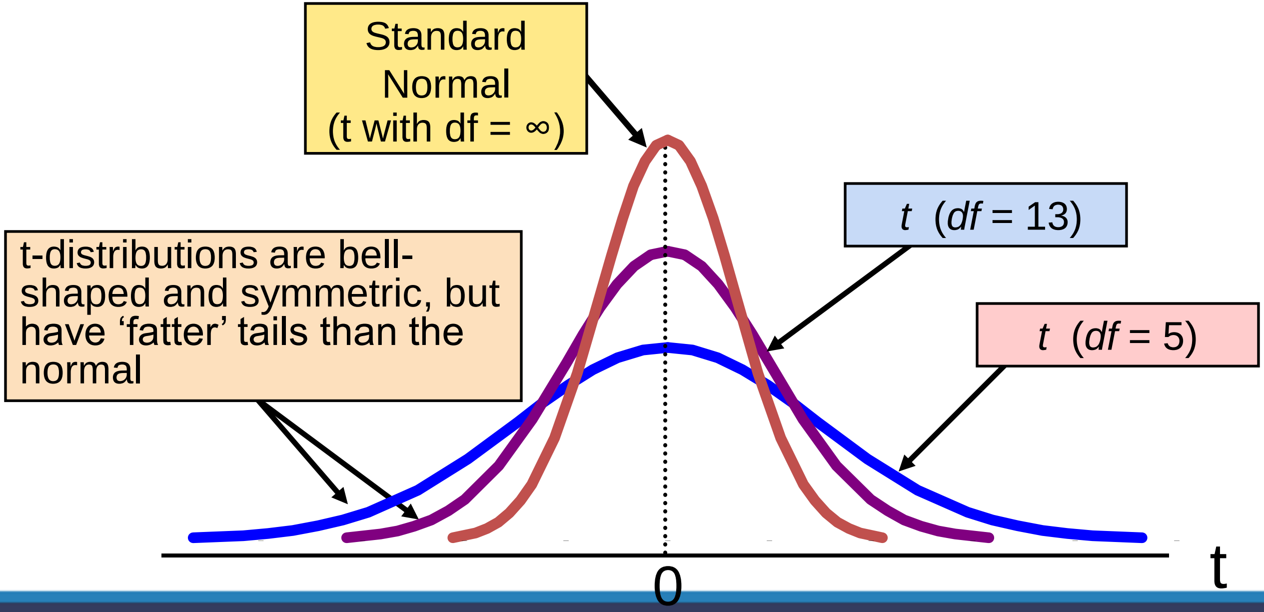
Student's t Distribution

- The t is a family of distributions
- The t value depends on degrees of freedom (d.f.)
  - Number of observations that are free to vary after sample mean has been calculated

$$d.f. = n - 1$$

Student's t Distribution

Note: t $\longrightarrow$ Z as n increases

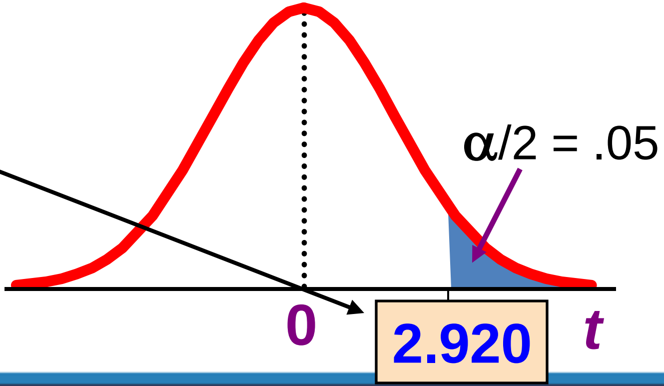


Student's t Table

	Upper Tail Area		
df	.10	**.05**	.025
1	3.078	6.314	12.706
2	1.886	**2.920**	4.303
3	1.638	2.353	3.182

The body of the table contains t values, not probabilities

Let: n = 3
df = *n* - 1 = 2
α = .10
α/2 =.05



t distribution values

With comparison to the Z value

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	Z
.80	1.372	1.325	1.310	1.282
.90	1.812	1.725	1.697	1.645
.95	2.228	2.086	2.042	1.960
.99	3.169	2.845	2.750	2.576

Note: $t \rightarrow Z$ as n increases

Example

A random sample of n = 25 has $\bar{x}$ = 50 and s = 8. Form a 95% confidence interval for μ

– d.f. = n – 1 = 24, so $t_{n-1,\alpha/2} = t_{24,.025} = 2.0639$

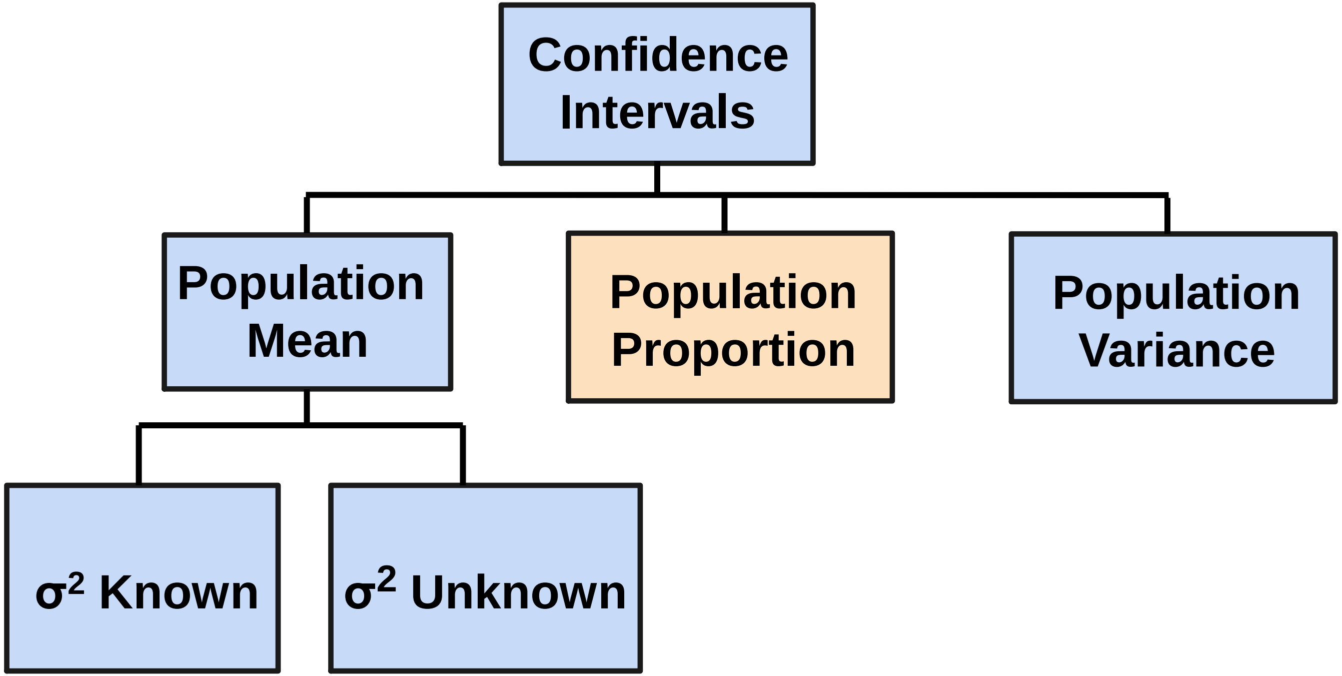
The confidence interval is

$$\bar{x} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

$$50 - (2.0639)\frac{8}{\sqrt{25}} < \mu < 50 + (2.0639)\frac{8}{\sqrt{25}}$$

$$46.698 < \mu < 53.302$$

Confidence Intervals



Confidence Intervals for the Population Proportion

- An interval estimate for the population proportion (P) can be calculated by adding an allowance for uncertainty to the sample proportion ($\hat{p}$)

Confidence Intervals for the Population Proportion, p

(continued)

- Recall that the distribution of the sample proportion is approximately normal if the sample size is large, with standard deviation

$$\sigma_P = \sqrt{\frac{P(1-P)}{n}}$$

- We will estimate this with sample data:

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence Interval Endpoints

- Upper and lower confidence limits for the population proportion are calculated with the formula

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < P < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- where
  - $z_{\alpha/2}$ is the standard normal value for the level of confidence desired
  - $\hat{p}$ is the sample proportion
  - n is the sample size
  - $nP(1-P) > 5$

Example

- A random sample of 100 people shows that 25 are left-handed.
- Form a 95% confidence interval for the true proportion of left-handers

Example

(continued)

- A random sample of 100 people shows that 25 are left-handed. Form a 95% confidence interval for the true proportion of left-handers.

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < P < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\frac{25}{100} - 1.96\sqrt{\frac{.25(.75)}{100}} < P < \frac{25}{100} + 1.96\sqrt{\frac{.25(.75)}{100}}$$

$$0.1651 < P < 0.3349$$

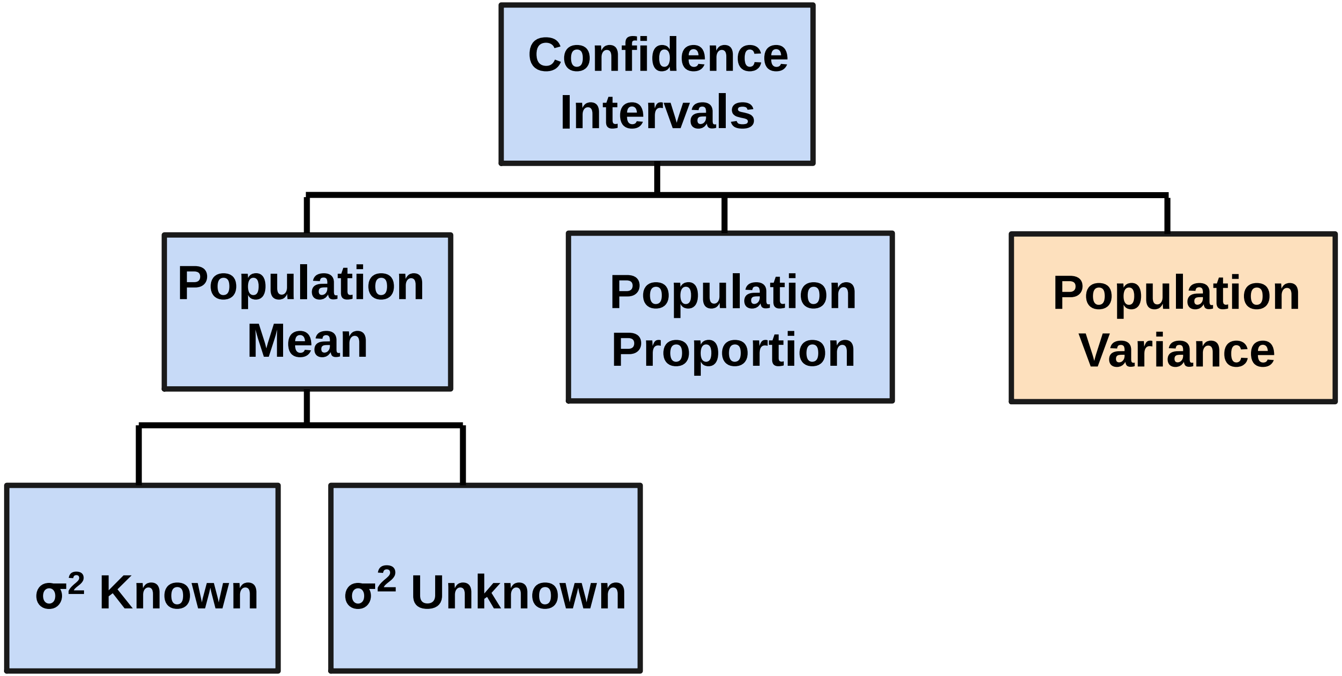
Interpretation

- We are 95% confident that the true percentage of left-handers in the population is between

  16.51% and 33.49%.

- Although the interval from 0.1651 to 0.3349 may or may not contain the true proportion, 95% of intervals formed from samples of size 100 in this manner will contain the true proportion.

Confidence Intervals



Confidence Intervals for the Population Variance

- **Goal:** Form a confidence interval for the population variance, σ^2

- The confidence interval is based on the sample variance, s^2
- Assumed: the population is normally distributed

Confidence Intervals for the Population Variance

(continued)

The random variable

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}$$

follows a chi-square distribution with (n – 1) degrees of freedom

Confidence Intervals for the Population Variance

The (1 - α)% confidence interval for the population variance is

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

Example

You are testing the speed of a batch of computer processors. You collect the following data (in Mhz):

Sample size	17
Sample mean	3004
Sample std dev	74

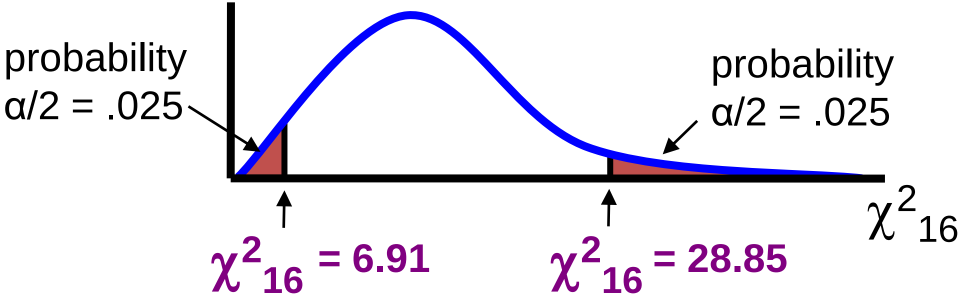
Assume the population is normal. Determine the 95% confidence interval for σ_x^2

Finding the Chi-square Values

- n = 17 so the chi-square distribution has (n – 1) = 16 degrees of freedom
- α = 0.05, so use the the chi-square values with area 0.025 in each tail:

$$\chi^2_{n-1,\alpha/2} = \chi^2_{16,0.025} = 28.85$$

$$\chi^2_{n-1,1-\alpha/2} = \chi^2_{16,0.975} = 6.91$$



Calculating the Confidence Limits

- The 95% confidence interval is

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

$$\frac{(17-1)(74)^2}{28.85} < \sigma^2 < \frac{(17-1)(74)^2}{6.91}$$

$$3037 < \sigma^2 < 12683$$

Converting to standard deviation, we are 95% confident that the population standard deviation of CPU speed is between 55.1 and 112.6 Mhz

Finite Populations

- If the sample size is more than 5% of the population size (and sampling is without replacement) then a finite population correction factor must be used when calculating the standard error

Finite Population Correction Factor

- Suppose sampling is without replacement and the sample size is large relative to the population size
- Assume the population size is large enough to apply the central limit theorem
- Apply the finite population correction factor when estimating the population variance

$$\text{finite population correction factor} = \frac{N-n}{N-1}$$

Estimating the Population Mean

- Let a simple random sample of size n be taken from a population of N members with mean μ
- The sample mean is an unbiased estimator of the population mean μ
- The point estimate is: $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Finite Populations: Mean

- If the sample size is more than 5% of the population size, an unbiased estimator for the variance of the sample mean is

$$\hat{\sigma}_{\bar{x}}^2 = \frac{s^2}{n}\left(\frac{N-n}{N-1}\right)$$

- So the 100(1-α)% confidence interval for the population mean is

$$\bar{x} - t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}}$$

Estimating the Population Proportion

- Let the true population proportion be P
- Let $\hat{p}$ be the sample proportion from n observations from a simple random sample
- The sample proportion, $\hat{p}$, is an unbiased estimator of the population proportion, P

Finite Populations: Proportion

- If the sample size is more than 5% of the population size, an unbiased estimator for the variance of the population proportion is

$$\hat{\sigma}_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n}\left(\frac{N-n}{N-1}\right)$$

- So the 100(1-α)% confidence interval for the population proportion is

$$\hat{p} - z_{\alpha/2}\hat{\sigma}_{\hat{p}} < P < \hat{p} + z_{\alpha/2}\hat{\sigma}_{\hat{p}}$$

Lecture Summary

- Introduced the concept of confidence intervals
- Discussed point estimates
- Developed confidence interval estimates
- Created confidence interval estimates for the mean (σ^2 known)
- Introduced the Student's t distribution
- Determined confidence interval estimates for the mean (σ^2 unknown)

Lecture Summary

(continued)

- Created confidence interval estimates for the proportion
- Created confidence interval estimates for the variance of a normal population
- Applied the finite population correction factor to form confidence intervals when the sample size is not small relative to the population size

Summary

- Introduced sampling distributions
- Described the sampling distribution of sample means
  - For normal populations
  - Using the Central Limit Theorem
- Described the sampling distribution of sample proportions
- Introduced the chi-square distribution
- Examined sampling distributions for sample variances
- Calculated probabilities using sampling distributions

Thank You