ECON 340
Economic Research Methods

Div Bhagia

Lecture 13: Confidence Intervals

Expectation and Variance of $\bar{X}$

Let $X_1, X_2, ..., X_n$ denote independent random draws (random sample) from a population with mean μ and variance σ^2.

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

Then $\bar{X}$ is also a random variable with:

$$E(\bar{X}) = \mu \qquad \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

So $\bar{X}$ is an unbiased and consistent estimator for μ.

Sample Mean Distribution

The distribution of the sample mean is normal if *either* of the following is true:

- The underlying population is normal
- The sample size is large, say $n \geqslant 100$

The first one follows from the sample mean being a linear combination of normally distributed variables.

The latter is implied by the *Central Limit Theorem*.

Central Limit Theorem

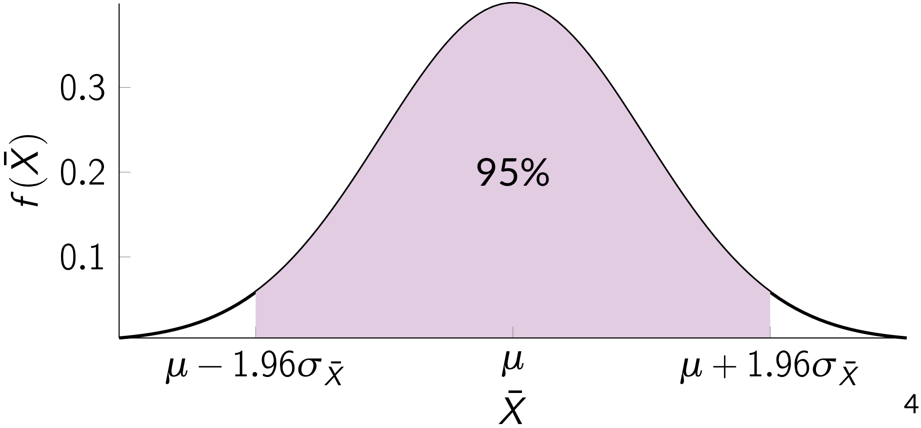
If $X_1, X_2, .., X_n$ are drawn randomly from a population with mean μ and variance σ^2, sample mean $\bar{X}$ is normally distributed with mean μ and variance σ^2/n as long as n is large.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Simulation

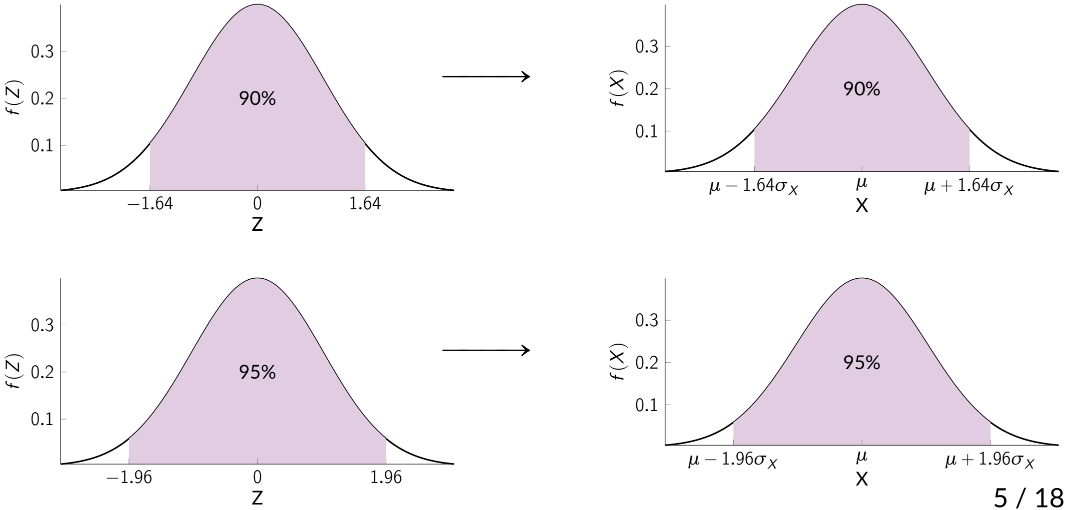
Normal Distribution

95% of the area under the curve lies within 1.96 standard deviations of the true mean.



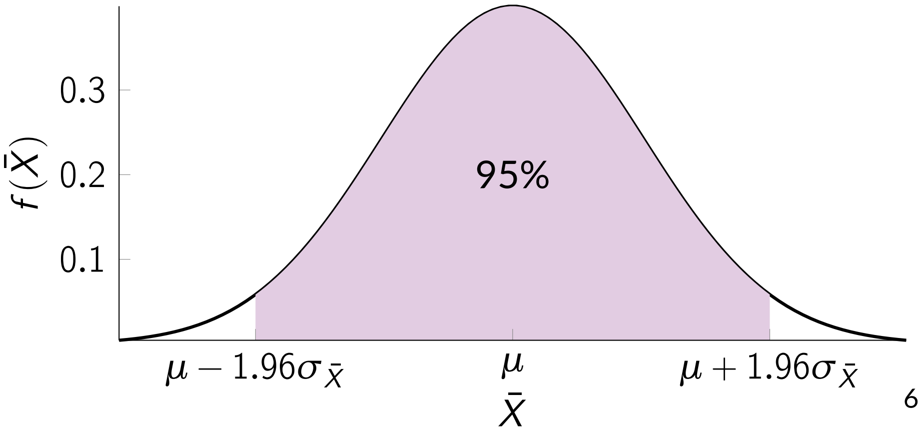
Aside

Standard Normal → Normal Distribution



Normal Distribution

95% of the time that we take a sample and calculate the sample mean, it will be within 1.96 standard deviations of μ.



Confidence Intervals

- If 95% of the time, the sample mean $\bar{X}$ will be within 1.96 standard deviations of the true mean μ.
- Then 95% of the time, the true mean μ will be within 1.96 standard deviations of the sample mean $\bar{X}$.
- Use this logic to create a 95% confidence interval for the true population mean μ.
- Say in your sample you found sample mean $\bar{x}$, then 95% confidence interval for μ:

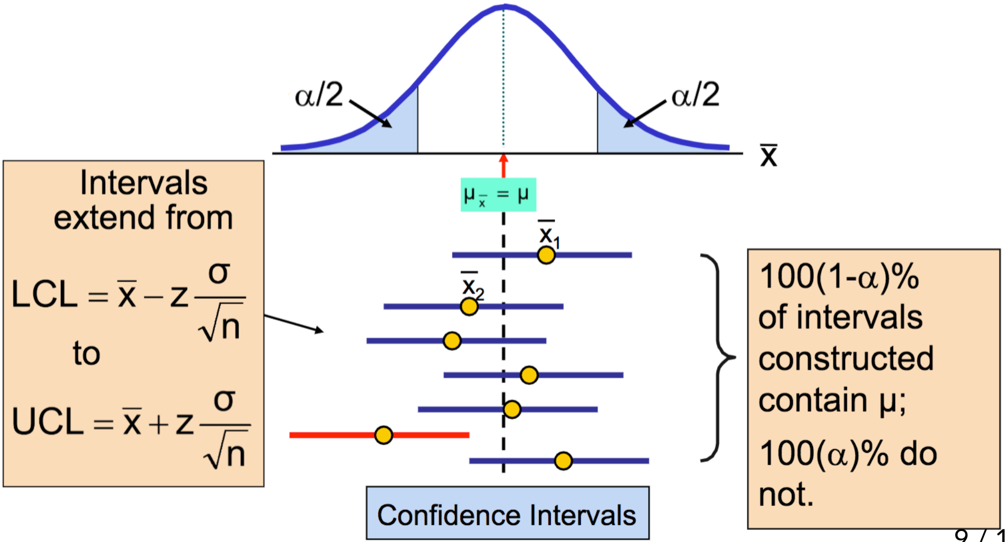
$$\bar{x} \pm 1.96\sigma_{\bar{X}}$$

Confidence Intervals: Interpretation

There is a 95% chance that the true population average lies in the interval $\bar{x} \pm 1.96\sigma_{\bar{X}}$.

What this really means is that if we took 100 random samples from the population and calculated 95% confidence intervals for each sample, we would expect 95 out of 100 intervals to contain the true population mean.

Confidence Intervals: Interpretation



Recipe: Confidence Intervals

Let $z_{\alpha/2}$ be the z-value that leaves area $\alpha/2$ in the upper tail of the normal distribution.

Then $1 - \alpha$ confidence interval is given by

$$\bar{x} \pm \underbrace{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{Margin of Error}}$$

Margin of Error

The margin of error will be reduced if

- Population standard deviation is reduced ($\downarrow \sigma$)
- The sample size is increased ($\uparrow n$)
- The confidence level is decreased ($\downarrow (1 - \alpha)$)

Population variance is not known!

- So far, we have assumed that we know the true population variance σ^2
- This is obviously not realistic!
- Most times we have to use the sample variance S^2 instead of σ^2.
- How do we create a confidence interval in this case?

Population variance is not known

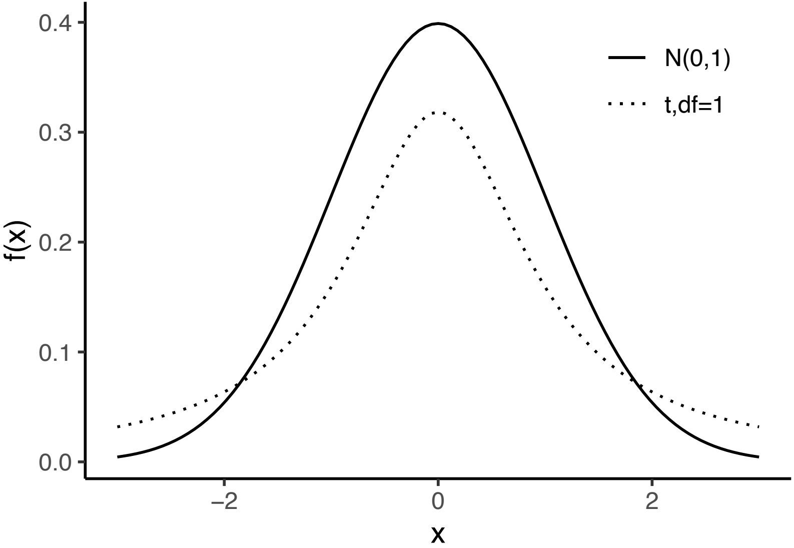
Instead of the Z statistic, we can use the T statistic

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

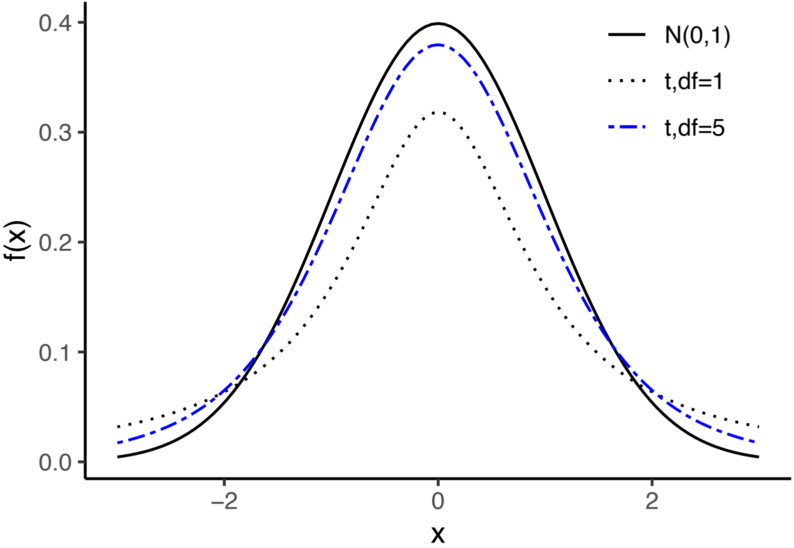
It can be shown that this statistic follows a t distribution with $n-1$ degrees of freedom.

The t-distribution is similar to the normal distribution but has thicker tails to account for the greater uncertainty in smaller samples. However, in large samples t-distribution can be approximated by the standard-normal.

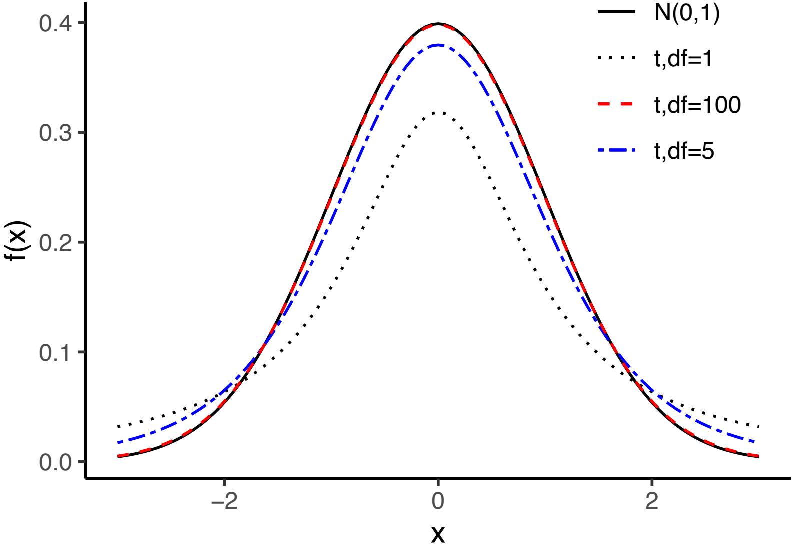
Student's T Distribution



Student's T Distribution



Student's T Distribution



Confidence Intervals: Unknown Variance

- Construct the confidence interval as before but now use T statistic instead of Z

- So need to use critical value for t

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

- But since we said that in large samples, t is approximated by z, we can continue using the standard normal table for the critical values if $n \geqslant 100$.

Next up

- Problem Set 3 is due on Thursday
- Next class: Hypothesis testing and p-values
- Next week:
  - Review class on Tuesday
  - Midterm exam on Thursday