ECON 340
Economic Research Methods

Div Bhagia

Lecture 12: Good Estimators, Sample Mean Distribution, Confidence Intervals

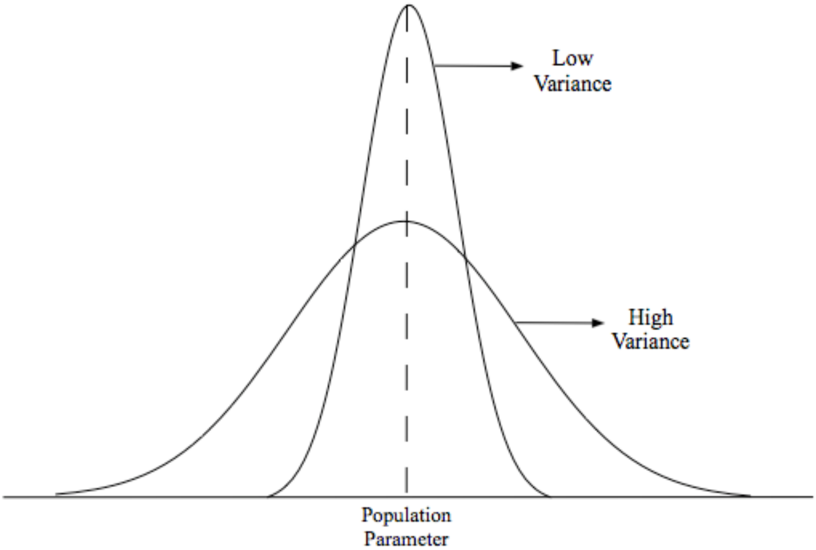
Sampling and Estimation

- We want to learn something about the population
- But often, we can collect data only for a sample of the population
- Good news: if the sample is drawn *randomly* we can use statistical methods to reach *tentative* answers
- Use sample quantities to *estimate* population parameters
- Sample *estimators* are random variables

Estimators

- Denote the population parameter of interest by θ
- And let's denote its sample estimator by $\hat{\theta}$
- Three desirable properties for an estimator:
    - *Unbiasedness*: $E(\hat{\theta}) = \theta$
    - *Efficiency*: lower variance is better
    - *Consistency*: as the sample size becomes infinitely large, $\hat{\theta} \rightarrow \theta$

What is a good estimator?



Expectation and Variance of $\bar{X}$

Let $X_1, X_2, \ldots, X_n$ 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:

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

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*.

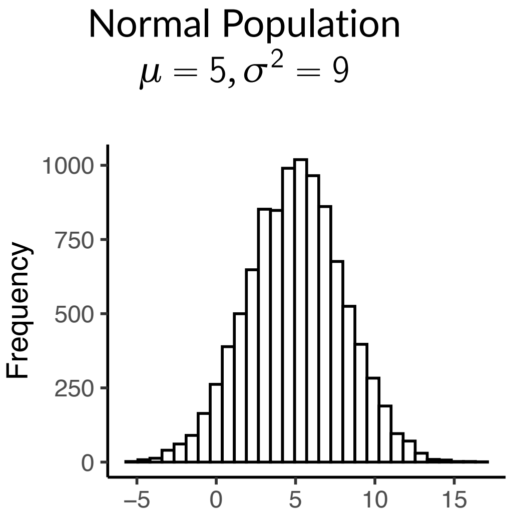
Central Limit Theorem

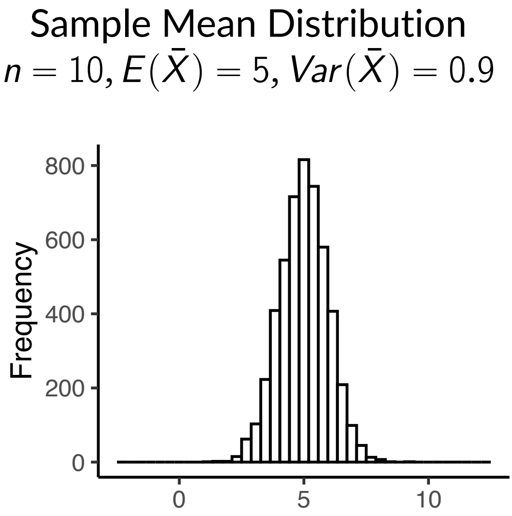
If $X_1, X_2, .., X_n$ are drawn randomly from a population with mean μ and variance σ^2, sample mean $\bar{X}$ is normally distributed with mean μ and variance σ^2/n as long as n is large.

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

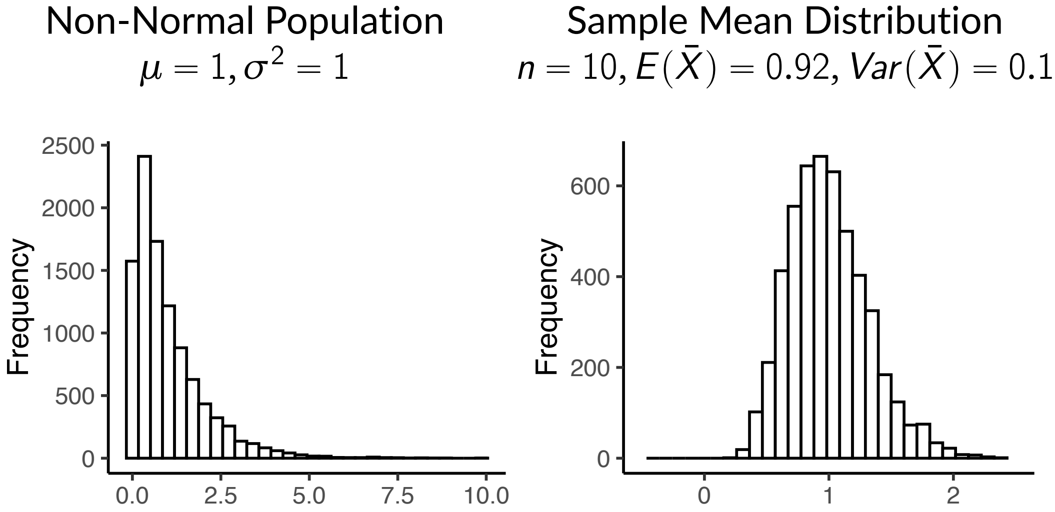
Simulation

Normal Population



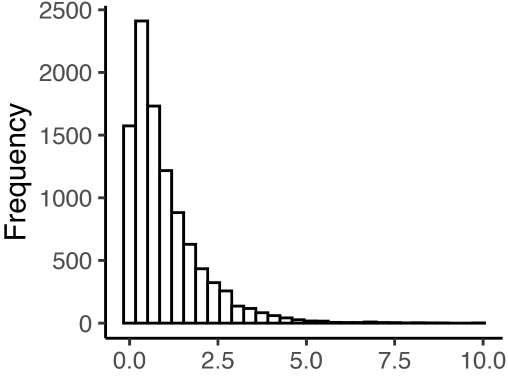


Non-Normal Population

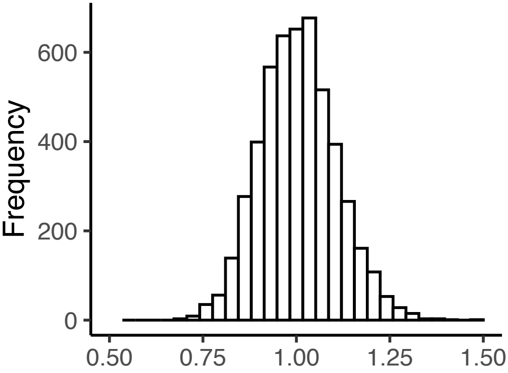


Central Limit Theorem

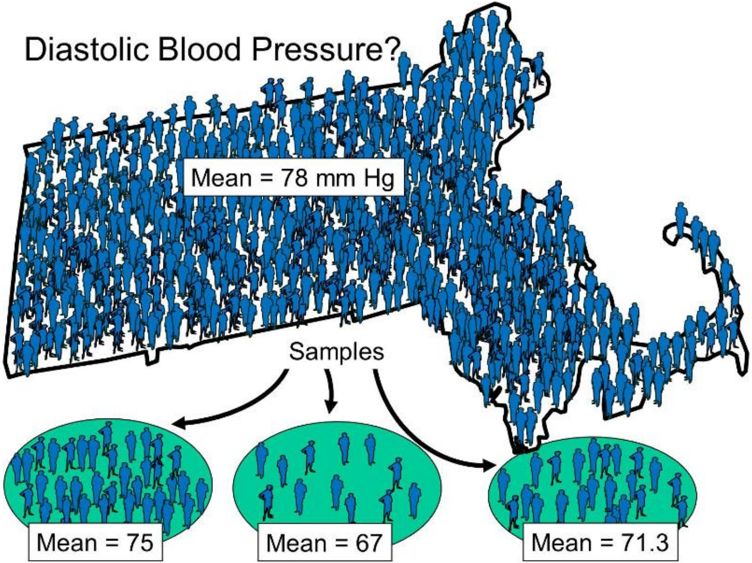
Non-Normal Population
$\mu = 1, \sigma^2 = 1$



Sample Mean Distribution
$n = 100, \bar{X} = 1, Var(\bar{X}) = 0.01$



Example: Blood Pressure in Massachusetts



Confidence Intervals

- Let's say we picked a random sample of 100 people from Massachusetts and took their blood pressure and found $\bar{x} = 75$.
- Given this estimate of 75, what can we say about the true mean?
- Here $n = 100$ so by CLT, $\bar{X} \sim N(\mu, \sigma^2/n)$. For now assume we know $\sigma^2 = 552.25$.
- Then we should be able to say with some certainty that the true mean lies *somewhere* around 75.

Confidence Intervals

- Create an interval around the sample mean that gives us a range of plausible values for the population mean.
- We can have confidence intervals of varying levels of confidence, most common are 90%, 95%, or 99%.
- The level of confidence is the probability that a calculated confidence interval contains the true population parameter.

How to construct a confidence interval?

Say we want to construct a 90% confidence interval for the true mean.

So far we have established that $\bar{X} \sim N(\mu, \sigma^2/n)$.

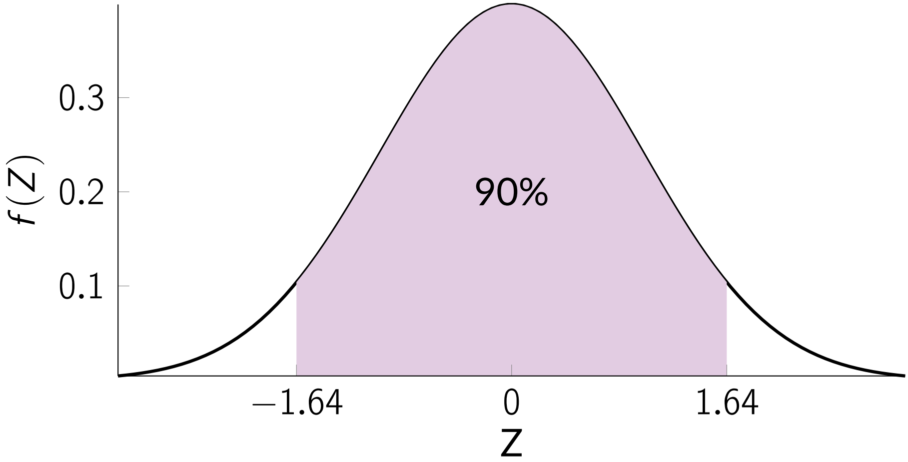
Note that then,

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

From the Standard Normal table, we can find that

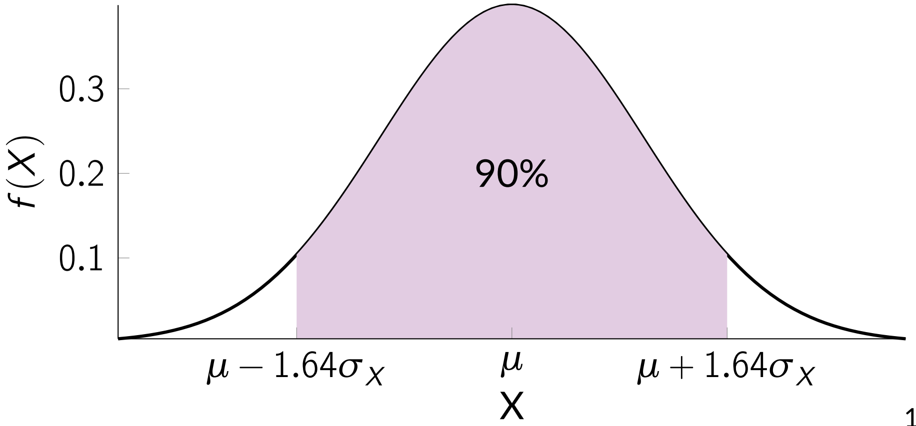
$$P(-1.64 < Z < 1.64) = 0.90$$

Standard Normal Distribution



Normal Distribution

90% of the area under the curve lies within 1.64 standard deviations of the mean.



90% Confidence Intervals

$$Pr(-1.64 < Z < 1.64) = 0.90$$

$$Pr\left(-1.64 < \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} < 1.64\right) = 0.90$$

$$Pr\left(\mu - 1.64\sigma_{\bar{X}} < \bar{X} < \mu + 1.64\sigma_{\bar{X}}\right) = 0.90$$

$$Pr\left(\bar{X} - 1.64\sigma_{\bar{X}} < \mu < \bar{X} + 1.64\sigma_{\bar{X}}\right) = 0.90$$

90% Confidence Intervals

$$Pr\left(\bar{X} - 1.64\sigma_{\bar{X}} < \mu < \bar{X} + 1.64\sigma_{\bar{X}}\right) = 0.90$$

Note that $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, so the 90% confidence interval here is given by:

$$\bar{x} \pm 1.64 \cdot \frac{\sigma}{\sqrt{n}}$$

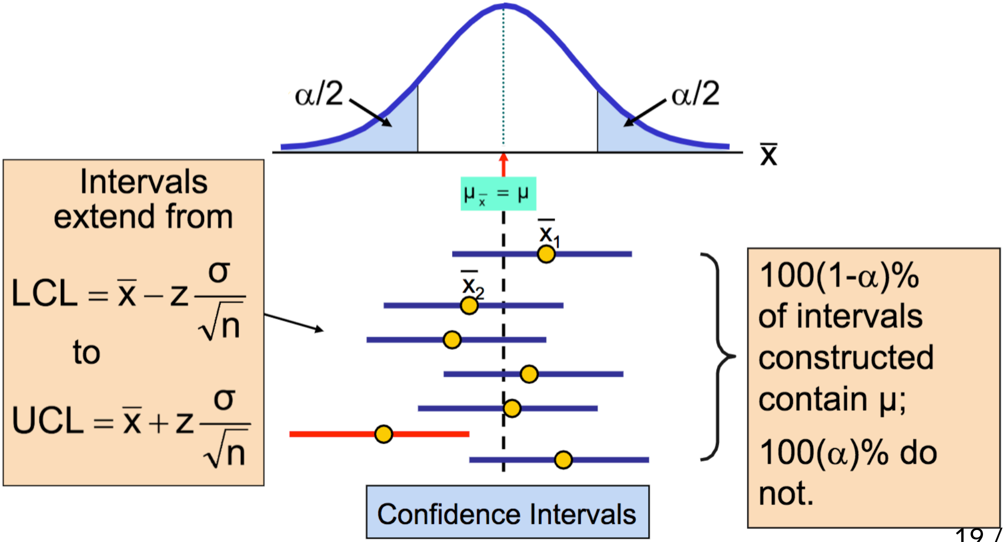
Plugging in $\sigma = \sqrt{552.25}$ and $n = 100$. We get $[71.15, 78.85]$.

Confidence Intervals: Interpretation

There is a 90% chance that the true population average for blood pressure lies in this interval.

What this really means is that if we took 100 random samples from the population and calculated 90% confidence intervals for each sample, we would expect 90 out of 100 intervals to contain the true population mean.

Confidence Intervals: Interpretation



Confidence Intervals: Recipe

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

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

Next up

- Problem Set 3 is due next Tuesday
- Next week: Continue with sampling and estimation
- Week after: Review class and midterm