Week 3: Learning from Random Samples

Brandon Stewart

Princeton

September 14-18, 2020

1These slides are heavily influenced by Matt Blackwell, Adam Glynn, Justin Grimmer, Jens Hainmueller, Erin Hartman and Matt Salganik. Some illustrations by Shay O’Brien.
Where We’ve Been and Where We’re Going...

- Last Week

- This Week
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- Next Week
  - hypothesis testing
  - what is regression?

- Long Run
  - probability
  → inference
  → regression
  → causal inference

Stewart (Princeton)
Where We’ve Been and Where We’re Going...

- Last Week
  - random variables

- This Week
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- Next Week
  - hypothesis testing
  - what is regression?

- Long Run
  - probability → inference → regression → causal inference
Where We’ve Been and Where We’re Going...

- Last Week
  - random variables
  - joint distributions

This Week
- estimators and sampling distributions
- estimator properties (bias, variance, consistency)

Next Week
- hypothesis testing
- what is regression?

Long Run
- probability
- → inference
- → regression
- → causal inference
Where We’ve Been and Where We’re Going…

- Last Week
  - random variables
  - joint distributions
- This Week
Where We’ve Been and Where We’re Going...

- Last Week
  - random variables
  - joint distributions
- This Week
  - estimators and sampling distributions

Stewart (Princeton)
Where We’ve Been and Where We’re Going...

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)

- **Next Week**
  - hypothesis testing
  - what is regression?

- **Long Run**
  - probability → inference → regression → causal inference

Stewart (Princeton)
Where We’ve Been and Where We’re Going...

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
  - hypothesis testing
  - what is regression?

- **Long Run**
  - probability → inference → regression → causal inference
Where We’ve Been and Where We’re Going...

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
Where We’ve Been and Where We’re Going...

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
  - hypothesis testing
Where We’ve Been and Where We’re Going...

- Last Week
  - random variables
  - joint distributions
- This Week
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals
- Next Week
  - hypothesis testing
  - what is regression?
Where We’ve Been and Where We’re Going…

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
  - hypothesis testing
  - what is regression?

- **Long Run**
  - probability
  - inference
  - regression
  - causal inference
Where We’ve Been and Where We’re Going...

- Last Week
  - random variables
  - joint distributions
- This Week
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals
- Next Week
  - hypothesis testing
  - what is regression?
- Long Run
  - probability → inference → regression → causal inference
1. Estimation
   - Populations and Samples
   - Estimators
   - Analytical

2. Weak Law of Large Numbers and the Central Limit Theorem
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3. Properties of Estimators
   - Four Desirable Properties
   - Example

4. Interval Estimation
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5. Plug-In Principle
1. **Estimation**
   - Populations and Samples
   - Estimators
   - Analytical

2. **Weak Law of Large Numbers and the Central Limit Theorem**
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3. **Properties of Estimators**
   - Four Desirable Properties
   - Example

4. **Interval Estimation**
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5. **Plug-In Principle**
Where We’ve Been and Where We’re Going...
Where We’ve Been and Where We’re Going...

Data generating process

Probability

Inference

Observed data
Primary Goal for This Week

Racial Prejudice and Attitudes Toward Affirmative Action*

James H. Kuklinski, University of Illinois at Urbana-Champaign
Paul M. Sniderman, Stanford University
Kathleen Knight, University of Houston
Thomas Piazza, University of California-Berkeley
Philip E. Tetlock, Ohio State University
Gordon R. Lawrence, Williams College
Barbara Mellers, Ohio State University

https://www.jstor.org/stable/2111770
Primary Goal for This Week

Theory: We examine the relationship between blatant racial prejudice and anger toward affirmative action.

Hypotheses: (1) Blatantly prejudiced attitudes continue to pervade the white population in the United States. (2) Resistance to affirmative action is more than an extension of this prejudice. (3) White resistance to affirmative action is not unyielding and unalterably fixed.

Methods: Analysis of experiments embedded in a national survey of racial attitudes. Some of these experiments are designed to measure racial prejudice unobtrusively.

Results: Racial prejudice remains a major problem in the United States, but this prejudice alone cannot explain all of the anger toward affirmative action among whites. Although many whites strongly resist affirmative action, they express support for making extra efforts to help African-Americans.

https://www.jstor.org/stable/2111770
Primary Goal for This Week

We want to be able to interpret the numbers in this table (and a couple of numbers that can be derived from these numbers).

<table>
<thead>
<tr>
<th>Region</th>
<th>Experimental Condition</th>
<th>Estimated Percent Angry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Black Family</td>
</tr>
<tr>
<td>Non-South</td>
<td>2.28(^a)</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td>(.07)</td>
<td>(.05)</td>
</tr>
<tr>
<td></td>
<td>425(^b)</td>
<td>461</td>
</tr>
<tr>
<td>South</td>
<td>1.95</td>
<td>2.37</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
<td>(.08)</td>
</tr>
<tr>
<td></td>
<td>139</td>
<td>136</td>
</tr>
</tbody>
</table>

\(^a\)Standard error of the estimate.

\(^b\)Number of cases.
An Overview

| Stewart  (Princeton) | Week 3: Learning From Random Samples | September 14-18, 2020 | 7 / 142 |
An Overview
An Overview
An Overview
An Overview
An Overview
An Overview

Population Distribution
\[ Y \sim \mathcal{N}(\mu, \sigma^2) \]

Estimand / Parameter
\[ \mu, \sigma^2 \]

Sample
\( (Y_1, Y_2, \ldots, Y_N) \)

Estimator/Statistic
\( \hat{g}(Y_1, Y_2, \ldots, Y_N) \)

Estimate
\( \hat{g}(Y_1 = y_1, Y_2 = y_2, \ldots, Y_N = y_N) \)
Populations

Typically, we want to learn about the distribution of random variables for a population of interest. We will sometimes call the population distribution the data generating process and represent it with a pmf or pdf, $f_X(x; \theta)$. The population can be:

- finite: as in all residents of a country
- or infinite: as in all possible television ads.

With either a finite or infinite population our main goal in inference is to learn about the population distribution $f_X$ via summaries, like $\mathbb{E}[X]$ or $\mathbb{V}[X]$, which we call a population parameter (or just parameter).

Ideally we assume as little as possible about the form of $f_X$. 
Typically, we want to learn about the distribution of random variables for a population of interest.
Populations

- Typically, we want to learn about the distribution of random variables for a population of interest.
- We will sometimes call the population distribution the data generating process and represent it with a pmf or pdf, $f_X(x; \theta)$.
Typically, we want to learn about the distribution of random variables for a population of interest.

We will sometimes call the population distribution the data generating process and represent it with a pmf or pdf, $f_X(x; \theta)$.

The population can be:
- **finite**: as in all residents of a country
- **infinite**: as in all possible television ads.
Populations

- Typically, we want to learn about the distribution of random variables for a population of interest.
- We will sometimes call the population distribution the data generating process and represent it with a pmf or pdf, $f_X(x; \theta)$.
- The population can be:
  - finite: as in all residents of a country
  - or infinite: as in all possible television ads.
- With either a finite or infinite population our main goal in inference is to learn about the population distribution $f_X$ via summaries, like $E[X]$ or $V[X]$, which we call a population parameter (or just parameter).
Typically, we want to learn about the distribution of random variables for a population of interest.

We will sometimes call the population distribution the data generating process and represent it with a pmf or pdf, \( f_X(x; \theta) \).

The population can be:

- finite: as in all residents of a country
- or infinite: as in all possible television ads.

With either a finite or infinite population our main goal in inference is to learn about the population distribution \( f_X \) via summaries, like \( E[X] \) or \( V[X] \), which we call a population parameter (or just parameter).

Ideally we assume as little as possible about the form of \( f_X \).
The goal of statistical inference is to learn about the unobserved population distribution, which can be characterized by parameters. Estimands are the parameters to estimate. Often written with greek letters (e.g. $\mu$).

Estimators are functions which map our data to guesses about the estimand. Often denoted with a “hat” (e.g. $\hat{\mu}$).

Estimates are particular values of estimators that are realized in a given sample (e.g. 12).
Nomenclature: Estimands, Estimators, and Estimates

The goal of statistical inference is to learn about the unobserved population distribution, which can be characterized by parameters.
Nomenclature: Estimands, Estimators, and Estimates

The goal of statistical inference is to learn about the unobserved population distribution, which can be characterized by parameters.

- **Estimands** are the parameters to estimate.
  Often written with greek letters (e.g. \( \mu \)).
Nomenclature: Estimands, Estimators, and Estimates

The goal of statistical inference is to learn about the unobserved population distribution, which can be characterized by parameters.

- **Estimands** are the parameters to estimate. Often written with greek letters (e.g. \( \mu \)).

- **Estimators** are functions which map our data to guesses about the estimand. Often denoted with a “hat” (e.g. \( \hat{\mu} \)).
Nomenclature: Estimands, Estimators, and Estimates

The goal of statistical inference is to learn about the unobserved population distribution, which can be characterized by parameters.

- **Estimands** are the parameters to estimate. Often written with greek letters (e.g. $\mu$).
- **Estimators** are functions which map our data to guesses about the estimand. Often denoted with a “hat” (e.g. $\hat{\mu}$)
- **Estimates** are particular values of estimators that are realized in a given sample (e.g. 12)
Independent and Identically Distributed Samples

Statistical inference is learning about features of some population through a sampling mechanism. We will base most of our inferential machinery on the idea of random sampling. We will leverage the powerful assumption that we are observing IID—independent and identically distributed—samples of the random variable of interest.

Plain language: Data are sampled IID when each observation is drawn from the same distribution, and the way an observation is drawn does not depend on the values of any other draw.
Independent and Identically Distributed Samples

Statistical inference is learning about features of some population through a sampling mechanism.

- We will base most of our inferential machinery on the idea of random sampling.
Independent and Identically Distributed Samples

Statistical inference is learning about features of some population through a sampling mechanism.

- We will base most of our inferential machinery on the idea of random sampling.
- We will leverage the powerful assumption that we are observing IID—indeed independent and identically distributed—samples of the random variable of interest.
Independent and Identically Distributed Samples

Statistical inference is learning about features of some population through a sampling mechanism.

- We will base most of our inferential machinery on the idea of random sampling.
- We will leverage the powerful assumption that we are observing IID—indeed independent and identically distributed—samples of the random variable of interest.

Plain language:
Data are sampled IID when each observation is drawn from the same distribution, and the way an observation is drawn does not depend on the values of any other draw.
IID Formal Definition

Definition (Independent and Identically Distributed)

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be random variables with CDFs $F_1, F_2, \ldots, F_n$, respectively. Let $F_A$ denote the joint CDF of the random variables with indices in the set $A$. Then $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are independently and identically distributed if they satisfy the following:

- **Mutually independent:**
  \[
  \forall A \subseteq \{1, 2, \ldots, n\}, \forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n, F_A ((x_i)_{i \in A}) = \prod_{i \in A} F_i(x_i)
  \]

- **Identically distributed:**
  \[
  \forall i, j \in \{1, 2, \ldots, n\} \text{ and } \forall x \in \mathbb{R}, F_i(x) = F_j(x)
  \]

(Aronow and Miller Definition 3.1.1)
Sample Notation

Under IID, we take a draw from a random variable $X$ and then take another draw such that the outcome doesn't depend on the first. For a collection of samples of size $n$, we collect each of these units into a vector $X = (X_1, X_2, ..., X_n)$. IID tells us that each one is produced under the same random process. This is how we get leverage to do estimation!

We will usually use unsubscripted capital letters, $X$, to refer to properties that all these draws share. e.g. $E[X] = E[X_1] = E[X_2] = ... = E[X_n]$. 

Stewart (Princeton)  
Week 3: Learning From Random Samples  
September 14-18, 2020 13 / 142
Sample Notation

- Under IID, we take a draw from a random variable $X$ and then take another draw such that the outcome doesn’t depend on the first.
Sample Notation

- Under IID, we take a draw from a random variable $X$ and then take another draw such that the outcome doesn’t depend on the first.
- For a collection of samples of size $n$, we collect each of these units into a vector $X = (X_1, X_2, \ldots, X_n)$. 
Sample Notation

- Under IID, we take a draw from a random variable $X$ and then take another draw such that the outcome doesn’t depend on the first.
- For a collection of samples of size $n$, we collect each of these units into a vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.
- IID tells us that each one is produced under the same random process. This is how we get leverage to do estimation!
Sample Notation

- Under IID, we take a draw from a random variable $X$ and then take another draw such that the outcome doesn’t depend on the first.
- For a collection of samples of size $n$, we collect each of these units into a vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.
- IID tells us that each one is produced under the same random process. This is how we get leverage to do estimation!
- We will usually use unsubscripted capital letters, $X$, to refer to properties that all these draws share.
  e.g. $E[X] = E[X_1] = E[X_2] = \cdots = E[X_n]$
Independent and Identically Distributed

IID is an assumption that only \textit{approximates} the truth.
Independent and Identically Distributed

IID is an assumption that only *approximates* the truth. What happens if it doesn’t hold?

1) Observations may be dependent.
   - e.g. students in the same class who share a teacher.
   - when dependence is weak in the sense that we are still adding more information as we add more units then results generally carry through but likely to understate uncertainty.

2) Observations may not be identically distributed.
   - e.g. observations may be changing over time.
   - here the consequence is primarily for interpretability, limiting us to a pseudopopulation which aggregates over distinct distributions.

We want to avoid assumptions where we can to maintain credible inferences, but this is a relatively mild bedrock assumption. We will return to these issues more in later videos and in future weeks.
Independent and Identically Distributed

IID is an assumption that only approximates the truth. What happens if it doesn’t hold?

1) Observations may be dependent.
   - e.g. students in the same class who share a teacher.
   - when dependence is weak in the sense that we are still adding more information as we add more units then results generally carry through but likely to understate uncertainty.

2) Observations may not be identically distributed.
   - e.g. observations may be changing over time.
   - here the consequence is primarily for interpretability, limiting us to a pseudopopulation which aggregates over distinct distributions.

We want to avoid assumptions where we can to maintain credible inferences, but this is a relatively mild bedrock assumption. We will return to these issues more in later videos and in future weeks.
Independent and Identically Distributed

IID is an assumption that only approximates the truth. What happens if it doesn’t hold?

1) Observations may be dependent.
   - e.g. students in the same class who share a teacher.

2) Observations may not be identically distributed.
   - e.g. observations may be changing over time.
   - here the consequence is primarily for interpretability, limiting us to a pseudopopulation which aggregates over distinct distributions.

We want to avoid assumptions where we can to maintain credible inferences, but this is a relatively mild bedrock assumption. We will return to these issues more in later videos and in future weeks.
Independent and Identically Distributed

IID is an assumption that only approximates the truth. What happens if it doesn't hold?

1) Observations may be dependent.
   ▶ e.g. students in the same class who share a teacher.
   ▶ when dependence is weak in the sense that we are still adding more information as we add more units then results generally carry through but likely to understate uncertainty.
Independent and Identically Distributed

IID is an assumption that only **approximates** the truth. What happens if it doesn’t hold?

1) Observations may be **dependent**.
   - e.g. students in the same class who share a teacher.
   - when dependence is **weak** in the sense that we are still adding more information as we add more units then results generally carry through but likely to **understate uncertainty**.

2) Observations may not be identically distributed.
Independent and Identically Distributed

IID is an assumption that only approximates the truth. What happens if it doesn’t hold?

1) Observations may be dependent.
   ▶ e.g. students in the same class who share a teacher.
   ▶ when dependence is weak in the sense that we are still adding more information as we add more units then results generally carry through but likely to understate uncertainty.

2) Observations may not be identically distributed.
   ▶ e.g. observations may be changing over time.
Independent and Identically Distributed

IID is an assumption that only **approximates** the truth. What happens if it doesn't hold?

1) Observations may be dependent.
   - e.g. students in the same class who share a teacher.
   - when dependence is **weak** in the sense that we are still adding more information as we add more units then results generally carry through but likely to **understate uncertainty**.

2) Observations may not be identically distributed.
   - e.g. observations may be changing over time.
   - here the consequence is primarily for **interpretability**, limiting us to a **pseudopopulation** which aggregates over distinct distributions.
Independent and Identically Distributed

IID is an assumption that only **approximates** the truth. What happens if it doesn’t hold?

1) Observations may be dependent.
   - e.g. students in the same class who share a teacher.
   - when dependence is **weak** in the sense that we are still adding more information as we add more units then results generally carry through but likely to **understate uncertainty**.

2) Observations may not be identically distributed.
   - e.g. observations may be changing over time.
   - here the consequence is primarily for **interpretability**, limiting us to a **pseudopopulation** which aggregates over distinct distributions.

We want to avoid assumptions where we can to maintain **credible** inferences, but this is a relatively mild bedrock assumption.
Independent and Identically Distributed

IID is an assumption that only approximates the truth. What happens if it doesn’t hold?

1) Observations may be dependent.
   - e.g. students in the same class who share a teacher.
   - when dependence is weak in the sense that we are still adding more information as we add more units then results generally carry through but likely to understate uncertainty.

2) Observations may not be identically distributed.
   - e.g. observations may be changing over time.
   - here the consequence is primarily for interpretability, limiting us to a pseudopopulation which aggregates over distinct distributions.

We want to avoid assumptions where we can to maintain credible inferences, but this is a relatively mild bedrock assumption.

We will return to these issues more in later videos and in future weeks.
Sampling from Finite Populations

When we take a sample from a population it can be done with or without replacement.

If the unit is replaced such that it can be sampled again, each draw is taken from the same population governed by the $f_X$.

If the unit is not replaced, then each subsequent draw depends on the units previously sampled.

If the population is very large relative to the sample, this won’t turn out to matter much (because removing each unit doesn’t change the overall distribution $f_X$).

If the population is small relative to the sample size, it will be necessary to think carefully through the implications (see e.g. the challenge problem in problem set 3).
Sampling from Finite Populations

- When we take a sample from a population it can be done with or without replacement.
Sampling from Finite Populations

- When we take a sample from a population it can be done with or without replacement.
- If the unit is replaced such that it can be sampled again, each draw is taken from the same population governed by the $f_X$. 

If the population is very large relative to the sample, this won't turn out to matter much (because removing each unit doesn't change the overall distribution $f_X$).

If the population is small relative to the sample size, it will be necessary to think carefully through the implications (see e.g. the challenge problem in problem set 3).
Sampling from Finite Populations

- When we take a sample from a population it can be done with or without replacement.
- If the unit is replaced such that it can be sampled again, each draw is taken from the same population governed by the $f_X$.
- If the unit is not replaced, then each subsequent draw depends on the units previously sampled.
Sampling from Finite Populations

- When we take a sample from a population it can be done with or without replacement.
- If the unit is replaced such that it can be sampled again, each draw is taken from the same population governed by the $f_X$.
- If the unit is not replaced, then each subsequent draw depends on the units previously sampled.
- If the population is very large relative to the sample, this won’t turn out to matter much (because removing each unit doesn’t change the overall distribution $f_X$).
Sampling from Finite Populations

- When we take a sample from a population it can be done with or without replacement.
- If the unit is replaced such that it can be sampled again, each draw is taken from the same population governed by the $f_X$.
- If the unit is not replaced, then each subsequent draw depends on the units previously sampled.
- If the population is very large relative to the sample, this won’t turn out to matter much (because removing each unit doesn’t change the overall distribution $f_X$).
- If the population is small relative to the sample size, it will be necessary to think carefully through the implications (see e.g. the challenge problem in problem set 3).
Sampling in R

```r
## draw a sample of size 10 from our population
## drawn without replacement
my_sample <- dplyr::sample_n(my_data, size = 10,
                          replace = FALSE)
## this is a wrapper around sample.int()
my_sample <- my_data[sample.int(nrow(my_data),
                      size = 10, replace = FALSE), ]
```
What is An Estimator?

An estimator \( \hat{\theta} \) for some parameter \( \theta \), is a function of the sample \( \hat{\theta} = h(Y_1, \ldots, Y_n) \). Because it is a function of the sample, the estimator is a random variable.

We will study the properties of the estimator towards two goals:

1. Inference: How much uncertainty do we have in this estimate?
2. Evaluate Estimators: How do we choose which estimator to use?

We study estimators by considering their behavior across an infinite number of hypothetical samples of size \( n \) that could be drawn. The resulting distribution of estimates is the sampling distribution. In real applications, we cannot draw repeated samples, so we approximate the sampling distribution.
What is An Estimator?

- An estimator $\hat{\theta}$ for some parameter $\theta$, is a function of the sample $\hat{\theta} = h(Y_1, \ldots, Y_n)$.
What is An Estimator?

- An estimator $\hat{\theta}$ for some parameter $\theta$, is a function of the sample $\hat{\theta} = h(Y_1, \ldots, Y_n)$.
- Because it is a function of the sample, the estimator is a random variable.
What is An Estimator?

- An estimator $\hat{\theta}$ for some parameter $\theta$, is a function of the sample $\hat{\theta} = h(Y_1, \ldots, Y_n)$.
- Because it is a function of the sample, the estimator is a random variable.
- We will study the properties of the estimator towards two goals:

  1. Inference: How much uncertainty do we have in this estimate?
  2. Evaluate Estimators: How do we choose which estimator to use?

We study estimators by considering their behavior across an infinite number of hypothetical samples of size $n$ that could be drawn. The resulting distribution of estimates is the sampling distribution. In real applications, we cannot draw repeated samples, so we approximate the sampling distribution.
What is An Estimator?

- An estimator $\hat{\theta}$ for some parameter $\theta$, is a function of the sample $\hat{\theta} = h(Y_1, \ldots, Y_n)$.
- Because it is a function of the sample, the estimator is a random variable.
- We will study the properties of the estimator towards two goals:
  - **Inference**: How much uncertainty do we have in this estimate?
An estimator $\hat{\theta}$ for some parameter $\theta$, is a function of the sample $\hat{\theta} = h(Y_1, \ldots, Y_n)$.

Because it is a function of the sample, the estimator is a random variable.

We will study the properties of the estimator towards two goals:

1. Inference: How much uncertainty do we have in this estimate?
2. Evaluate Estimators: How do we choose which estimator to use?
What is An Estimator?

- An estimator $\hat{\theta}$ for some parameter $\theta$, is a function of the sample $\hat{\theta} = h(Y_1, \ldots, Y_n)$.
- Because it is a function of the sample, the estimator is a random variable.
- We will study the properties of the estimator towards two goals:
  1. **Inference**: How much uncertainty do we have in this estimate?
  2. **Evaluate Estimators**: How do we choose which estimator to use?
- We study estimators by considering their behavior across an infinite number of hypothetical samples of size $n$ that could be drawn. The resulting distribution of estimates is the **sampling distribution**.
What is An Estimator?

- An estimator $\hat{\theta}$ for some parameter $\theta$, is a function of the sample $\hat{\theta} = h(Y_1, \ldots, Y_n)$.
- Because it is a function of the sample, the estimator is a random variable.
- We will study the properties of the estimator towards two goals:
  1. Inference: How much uncertainty do we have in this estimate?
  2. Evaluate Estimators: How do we choose which estimator to use?
- We study estimators by considering their behavior across an infinite number of hypothetical samples of size $n$ that could be drawn. The resulting distribution of estimates is the sampling distribution.
- In real applications, we cannot draw repeated samples, so we approximate the sampling distribution.
The Sampling Distribution of the Sample Mean

Say we have the following population:

```r
c(4, 2, 3, 6, 9, 2, 3, 6, 8, 5, 2, 9, 6, 3, 4, 7, 6, 1, 2, 6, 9, 3, 1, 1, 1, 5, 7, 9)
```

We are going to take samples of size 10. How many possible samples are there?

```r
choose(length(pop), 10)
## [1] 13123110
```
The Sampling Distribution of the Sample Mean

Say we have the following population:

```r
pop <- c(4, 2, 3, 6, 9, 2, 3, 6, 8, 5, 2, 9, 6, 3, 4, 7, 6, 1, 2, 6, 9, 3, 1, 1, 1, 5, 7, 9)
```

```r
choose(length(pop), 10)
## [1] 13123110
```
The Sampling Distribution of the Sample Mean

Say we have the following population:

```r
pop <- c(4, 2, 3, 6, 9, 2, 3, 6, 8, 5, 2, 9, 6, 3,
        4, 7, 6, 1, 2, 6, 9, 3, 1, 1, 1, 5, 7, 9)
```

We are going to take samples of size 10. How many possible samples are there?
The Sampling Distribution of the Sample Mean

Say we have the following population:

```r
pop <- c(4, 2, 3, 6, 9, 2, 3, 6, 8, 5, 2, 9, 6, 3,
        4, 7, 6, 1, 2, 6, 9, 3, 1, 1, 1, 5, 7, 9)
```

We are going to take samples of size 10. How many possible samples are there?

```r
choose(length(pop), 10)
```

```r
## [1] 13123110
```
The Sampling Distribution

If we could draw each possible sample, we could calculate the sample mean in each one. This would form the full sampling distribution. We will simulate this by drawing 10,000 samples.

```r
sim_res <- replicate(10000, {
  mean(pop[sample.int(length(pop), 10)])
}) %>%
tibble(sample_mean = .)
rownames_to_column(sim_res, var = "replicate")
```

<table>
<thead>
<tr>
<th>replicate</th>
<th>sample_mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.4</td>
</tr>
<tr>
<td>2</td>
<td>4.9</td>
</tr>
<tr>
<td>3</td>
<td>3.7</td>
</tr>
<tr>
<td>4</td>
<td>3.6</td>
</tr>
<tr>
<td>5</td>
<td>5.3</td>
</tr>
</tbody>
</table>

Stewart (Princeton)
Week 3: Learning From Random Samples
September 14-18, 2020
The Sampling Distribution

If we could draw each possible sample, we could calculate the sample mean in each one. This would form the full sampling distribution. We will simulate this by drawing 10,000 samples.
The Sampling Distribution

If we could draw each possible sample, we could calculate the sample mean in each one. This would form the full sampling distribution. We will simulate this by drawing 10,000 samples.

```
sim_res <- replicate(10000, {
  mean(pop[sample.int(length(pop), 10)])
}) %>%
tibble(sample_mean = .) %>%
  rownames_to_column(var = "replicate")
```

```
sim_res[1:5, ]
```

````
# A tibble: 5 x 2
#  replicate sample_mean
#   <chr>      <dbl>
# 1     1        5.4
# 2     2        4.9
# 3     3        3.7
# 4     4        3.6
# 5     5        5.3
```
The Sampling Distribution

And we can plot this sampling distribution

```r
true_pop_mean = mean(pop)
ggplot(sim_res, aes(x = sample_mean)) +
  geom_histogram(fill = blue) +
  geom_vline(xintercept = true_pop_mean) +
  ggtitle("Sampling Distribution of Sample Mean") +
  xlab("Sample Mean") + theme_bw()
```

![Sampling Distribution of Sample Mean](image)
An Analytical Approach to the Sampling Distributions

The sampling distribution tells us how the estimator performs over many hypothetical samples. Unfortunately in real-world analysis we don't get to see the whole distribution, just one draw! Because the estimator is a random variable (remember it is a function of the statistic!), we can characterize the sampling distribution using the same tools from last week.

We will start with a common estimator, the sample mean, $X_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Under the identically and independently distributed assumption we can characterize properties of the distribution like the expectation and variance.
The sampling distribution tells us how the estimator performs over many hypothetical samples.
An Analytical Approach to the Sampling Distributions

- The sampling distribution tells us how the estimator performs over many hypothetical samples.
- Unfortunately in real-world analysis we don’t get to see the whole distribution, just one draw!
An Analytical Approach to the Sampling Distributions

- The sampling distribution tells us how the estimator performs over many hypothetical samples.
- Unfortunately in real-world analysis we don’t get to see the whole distribution, just one draw!
- Because the estimator is a random variable (remember it is a function of the statistic!) we can characterize the sampling distribution using the same tools from last week.
An Analytical Approach to the Sampling Distributions

- The sampling distribution tells us how the estimator performs over many hypothetical samples.
- Unfortunately in real-world analysis we don’t get to see the whole distribution, just one draw!
- Because the estimator is a random variable (remember it is a function of the statistic!) we can characterize the sampling distribution using the same tools from last week.
- We will start with a common estimator, the sample mean, \( \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \).
The sampling distribution tells us how the estimator performs over many hypothetical samples.

Unfortunately in real-world analysis we don’t get to see the whole distribution, just one draw!

Because the estimator is a random variable (remember it is a function of the statistic!) we can characterize the sampling distribution using the same tools from last week.

We will start with a common estimator, the sample mean, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Under the identically and independently distributed assumption we can characterize properties of the distribution like the expectation and variance.
Describing the Sampling Distribution for the Sample Mean

We would like a full description of the sampling distribution for the sample mean estimator, but it will be useful to separate this description into three parts.
Describing the Sampling Distribution for the Sample Mean

We would like a full description of the sampling distribution for the sample mean estimator, but it will be useful to separate this description into three parts.

If we assume that $X_1, \ldots, X_n \sim_{i.i.d.} ?(\mu, \sigma^2)$, then we would like to identify the following things about $\overline{X}_n$. 

...
Describing the Sampling Distribution for the Sample Mean

We would like a full description of the sampling distribution for the sample mean estimator, but it will be useful to separate this description into three parts.

If we assume that $X_1, \ldots, X_n \sim \text{i.i.d. } ?(\mu, \sigma^2)$, then we would like to identify the following things about $\bar{X}_n$.

- $E[\bar{X}_n]$
- $V[\bar{X}_n]$
- $f_{\bar{X}_n}$
- $f_{\bar{X}_n} \sim ?$
Expectation of $\overline{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

$$E[\overline{X}_n] =$$
Expectation of $\bar{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

$$E[\bar{X}_n] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]$$
Expectation of $\bar{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

\[
E[\bar{X}_n] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] \\
= \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right]
\]
Expectation of \( \overline{X}_n \)

Let \( X_1, X_2, \ldots X_n \) be identically and independently distributed from a population distribution with mean \( (E[X_i] = \mu) \) and variance \( (V[X_i] = \sigma^2) \).

\[
E[\overline{X}_n] = E\left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] \\
= \frac{1}{n} E\left[ \sum_{i=1}^{n} X_i \right] \\
= \frac{1}{n} \sum_{i=1}^{n} E[X_i]
\]
Expectation of $\bar{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

\[
E[\bar{X}_n] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu
\]
Expectation of $\bar{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

$$E[\bar{X}_n] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} E[X_i]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mu$$

$$= \frac{1}{n} n \times \mu$$

$$= \mu$$
Expectation of $\bar{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

\[
E[\bar{X}_n] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] \\
= \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] \\
= \frac{1}{n} \sum_{i=1}^{n} E[X_i] \\
= \frac{1}{n} \sum_{i=1}^{n} \mu \\
= \frac{1}{n} n \times \mu \\
= \mu
\]
Variance of $\overline{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean $(E[X_i] = \mu)$ and variance $(V[X_i] = \sigma^2)$.

$$V[\overline{X}_n] = V\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]$$
Variance of $\overline{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

\[
V[\overline{X}_n] = V\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]
\]

\[
= \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right]
\]

Note the $n$ in the denominator: as we have more observations, the variance of the sampling distribution will shrink.
**Variance of $\overline{X}_n$**

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

$$V[\overline{X}_n] = V\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} V[X_i] \text{(because i.i.d)}$$
Variance of $\overline{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

$$V[\overline{X}_n] = V\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} V[X_i] \text{(because i.i.d)}$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2$$
Variance of $\bar{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

$$V[\bar{X}_n] = V\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} V[X_i] \text{(because i.i.d)}$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2$$

$$= \frac{1}{n^2} n \times \sigma^2$$

Note the $n$ in the denominator: as we have more observations, the variance of the sampling distribution will shrink.
Variance of $\bar{X}_n$

Let $X_1, X_2, \ldots, X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

\[
V[\bar{X}_n] = V\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] \\
= \frac{1}{n^2} \sum_{i=1}^{n} V[X_i] \text{(because i.i.d)} \\
= \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 \\
= \frac{1}{n^2} n \times \sigma^2 \\
= \frac{\sigma^2}{n}
\]
Variance of $\bar{X}_n$

Let $X_1, X_2, \ldots X_n$ be identically and independently distributed from a population distribution with mean ($E[X_i] = \mu$) and variance ($V[X_i] = \sigma^2$).

\[
V[\bar{X}_n] = V\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]
\]

\[
= \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right]
\]

\[
= \frac{1}{n^2} \sum_{i=1}^{n} V[X_i] \text{(because i.i.d)}
\]

\[
= \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2
\]

\[
= \frac{1}{n^2} n \times \sigma^2
\]

\[
= \frac{\sigma^2}{n}
\]

Note the $n$ in the denominator: as we have more observations, the variance of the sampling distribution will shrink.
What about the “?”

If $X_1, \ldots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2)$, then

$$\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$$
What about the “?”

If \( X_1, \ldots, X_n \sim_{i.i.d.} N(\mu, \sigma^2) \), then

\[
\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})
\]

What if \( X_1, \ldots, X_n \) are not normally distributed?
Bernoulli (Coin Flip) Distribution
Bernoulli (Coin Flip) Distribution

Population Distribution

Sampling Distribution of the Mean

Sampling Distribution of the Mean

Sampling Distribution of the Mean
Poisson (Count) Distribution

Population Distribution

Sampling Distribution of the Mean

Sampling Distribution of the Mean

Sampling Distribution of the Mean

Sampling Distribution of the Mean
Uniform Distribution

Population Distribution

Sampling Distribution of the Mean

Sampling Distribution of the Mean

Sampling Distribution of the Mean

Sampling Distribution of the Mean
Why would this be true?

Images from *Hyperbole and a Half* by Allie Brosh.
Why would this be true?

Images from *Hyperbole and a Half* by Allie Brosh.
We Covered...
We Covered...

- Populations and samples.
We Covered...

- Populations and samples.
- Estimators, estimands and estimates.
We Covered . . .

- Populations and samples.
- Estimators, estimands and estimates.
- Sampling distributions.

Next time: the answer to 'why happening?' and the most important theorem in statistics.
We Covered...

- Populations and samples.
- Estimators, estimands and estimates.
- Sampling distributions.

Next time: the answer to ’why happening?’ and the most important theorem in statistics.
Where We’ve Been and Where We’re Going...

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
  - hypothesis testing
  - what is regression?

- **Long Run**
  - probability $\rightarrow$ inference $\rightarrow$ regression $\rightarrow$ causal inference
1. Estimation
   - Populations and Samples
   - Estimators
   - Analytical

2. Weak Law of Large Numbers and the Central Limit Theorem
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3. Properties of Estimators
   - Four Desirable Properties
   - Example

4. Interval Estimation
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5. Plug-In Principle
1 Estimation
   - Populations and Samples
   - Estimators
   - Analytical

2 Weak Law of Large Numbers and the Central Limit Theorem
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3 Properties of Estimators
   - Four Desirable Properties
   - Example

4 Interval Estimation
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5 Plug-In Principle
Cliffhanger

We started last time on a statistical cliffhanger: why did the sampling distribution of the mean look the same as the number of observations increased regardless of the population distribution?
Cliffhanger

We started last time on a statistical cliffhanger: why did the sampling distribution of the mean look the same as the number of observations increased regardless of the population distribution?

The answer has to do with the asymptotic behavior of the estimator—what happens as the sample size $n$ increases.
Cliffhanger

We started last time on a statistical cliffhanger: why did the sampling distribution of the mean look the same as the number of observations increased regardless of the population distribution?

The answer has to do with the asymptotic behavior of the estimator—what happens as the sample size $n$ increases.

In order to characterize this formally though, we are going to have to set up some more probabilistic infrastructure.
Cliffhanger

We started last time on a statistical cliffhanger: why did the sampling distribution of the mean look the same as the number of observations increased regardless of the population distribution?

The answer has to do with the asymptotic behavior of the estimator—what happens as the sample size $n$ increases.

In order to characterize this formally though, we are going to have to set up some more probabilistic infrastructure.

The key thing to remember is that the sample mean is itself a random variable.
Cliffhanger

We started last time on a statistical cliffhanger: why did the sampling distribution of the mean look the same as the number of observations increased regardless of the population distribution?

The answer has to do with the asymptotic behavior of the estimator—what happens as the sample size \( n \) increases.

In order to characterize this formally though, we are going to have to set up some more probabilistic infrastructure.

The key thing to remember is that the sample mean is itself a random variable.

**Warning**: This video is a little bit mathier. At the end I’ll wrap up with things you need to know so don’t stress out your first watch through.
Bounding a Random Variable

Let's start by seeing how much we can learn about an estimator using only the things we've calculated so far (the expectation and variance).

Theorem (Chebychev's Inequality)

Let $X$ be a random variable with finite $\sigma_X > 0$. Then $\forall \epsilon > 0$, $P[|X - E[X]| \geq \epsilon \sigma_X] \leq \frac{1}{\epsilon^2}$ (Aronow and Miller Theorem 2.1.18)

This allows us to put an upper bound on the probability that a draw from the distribution will be more than a given number of standard deviations from the mean.

This lets us bound the behavior of a random variable knowing only the expectation and variance (regardless of distributional shape!).
Bounding a Random Variable

Let’s start by seeing how much we can learn about an estimator using only the things we’ve calculated so far (the expectation and variance).
Bounding a Random Variable

Let’s start by seeing how much we can learn about an estimator using only the things we’ve calculated so far (the expectation and variance).

Theorem (Chebychev’s Inequality)

Let $X$ be a random variable with finite $\sigma[X] > 0$. Then $\forall \epsilon > 0$,

$$P[|X - E[X]| \geq \epsilon \sigma[X]] \leq \frac{1}{\epsilon^2}$$

(Aronow and Miller Theorem 2.1.18)
Bounding a Random Variable

Let’s start by seeing how much we can learn about an estimator using only the things we’ve calculated so far (the expectation and variance).

Theorem (Chebychev’s Inequality)

Let $X$ be a random variable with finite $\sigma[X] > 0$. Then $\forall \epsilon > 0$,

$$P[|X - E[X]| \geq \epsilon \sigma[X]] \leq \frac{1}{\epsilon^2}$$

(Aronow and Miller Theorem 2.1.18)

This allows us to put an upper bound on the probability that a draw from the distribution will be more than a given number of standard deviations from the mean.
Bounding a Random Variable

Let’s start by seeing how much we can learn about an estimator using only the things we’ve calculated so far (the expectation and variance).

**Theorem (Chebychev’s Inequality)**

Let $X$ be a random variable with finite $\sigma[X] > 0$. Then $\forall \epsilon > 0$,

$$P[|X - E[X]| \geq \epsilon \sigma[X]] \leq \frac{1}{\epsilon^2}$$

(Aronow and Miller Theorem 2.1.18)

This allows us to put an upper bound on the probability that a draw from the distribution will be more than a given number of standard deviations from the mean.

This let’s us bound the behavior of a random variable knowing only the expectation and variance (regardless of distributional shape!).
Chebychev’s Inequality for the Sample Mean

To apply this to the sample mean, we plug in the expectation and variance to the Chebychev’s inequality and re-arranging terms, we get the following handy result.

\[
\text{Theorem (Chebychev’s Inequality for the Sample Mean)}
\]

Let \(X_1, X_2, \ldots, X_n\) be i.i.d. random variables with finite variance \(\text{Var}[X] > 0\). Then,

\[
P[|X_n - \mathbb{E}[X]| \geq \epsilon] \leq \frac{\text{Var}[X]}{\epsilon^2 n}
\]

(Aronow and Miller Theorem 3.2.5)

This allows us to put an upper bound on the probability that the sample mean for a given sample size and known variance will be some arbitrary distance from the true mean.
Chebychev’s Inequality for the Sample Mean

To apply this to the sample mean, we plug in the expectation and variance to the Chebychev’s inequality and re-arranging terms, we get the following handy result.

**Theorem (Chebychev’s Inequality for the Sample Mean)**

Let \( X_1, X_2, \ldots, X_n \) be i.i.d. random variables with finite variance \( V[X] > 0 \). Then, \( \forall \epsilon > 0 \),

\[
P \left[ \left| \overline{X}_n - E[X] \right| \geq \epsilon \right] \leq \frac{V[X]}{\epsilon^2 n}
\]

*(Aronow and Miller Theorem 3.2.5)*

This allows us to put an upper bound on the probability that the sample mean for a given sample size and known variance will be some arbitrary distance from the true mean.
Planning a Survey

- Problem: planning a survey to estimate Biden support in the 2020 election.

- How many people should we survey so that our estimator has no less than a \( 0.95 \) probability of being within \( 0.02 \) of the true population proportion?

- i.e. how do we get a margin of ±2 percentage points?

- Notation

  - \( \pi \) is proportion of voters expressing support for Biden (the estimand).

  - \( X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(\pi) \) be the iid random variables for each respondent.

  - \( \hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} X_i \) is our sample mean estimator.

- What do we know?

  - \( E[X_n] = E[X] \) and \( V[X_n] = V[X] \)

  - We know Bernoulli has variance of \( \pi(1-\pi) \) which is maximized at \( \pi = 0.5 \).
Planning a Survey

Problem: planning a survey to estimate Biden support in the 2020 election.

- How many people should we survey so that our estimator has no less than a .95 probability of being within .02 of the true population proportion?

Notation

- $\pi$ is proportion of voters expressing support for Biden (the estimand).
- $X_1, X_2, ..., X_n \sim \text{Bernoulli}(\pi)$ be the iid random variables for each respondent.
- $\hat{\pi} = \frac{X_n}{n}$ is our sample mean estimator.

What do we know?

- $E[X_n] = E[X]$ and $V[X_n] = V[X]$.
- We know Bernoulli has variance of $\pi(1-\pi)$ which is maximized at $\pi = .5$. 
Planning a Survey

- Problem: planning a survey to estimate Biden support in the 2020 election.
  - How many people should we survey so that our estimator has no less than a .95 probability of being within .02 of the true population proportion?
  - i.e. how do we get a margin of ± 2 percentage points?

- Notation
  - $\pi$ is proportion of voters expressing support for Biden (the estimand).
  - $X_1, X_2, ..., X_n \sim \text{Bernoulli}(\pi)$ be the iid random variables for each respondent.
  - $\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is our sample mean estimator.

- What do we know?
  - $E[X_n] = E[X]$ and $V[X_n] = V[X]$.
  - We know Bernoulli has variance of $\pi(1-\pi)$ which is maximized at $\pi = 0.5$. 
Planning a Survey

Problem: planning a survey to estimate Biden support in the 2020 election.

- How many people should we survey so that our estimator has no less than a .95 probability of being within .02 of the true population proportion?
- i.e. how do we get a margin of ± 2 percentage points?

Notation

- $\pi$ is proportion of voters expressing support for Biden (the estimand).
- $X_1, X_2, \ldots X_n \sim \text{Bernoulli}(\pi)$ be the iid random variables for each respondent.
- $\hat{\pi} = \bar{X}_n$ is our sample mean estimator.
Planning a Survey

- Problem: planning a survey to estimate Biden support in the 2020 election.
  - How many people should we survey so that our estimator has no less than a .95 probability of being within .02 of the true population proportion?
  - i.e. how do we get a margin of ± 2 percentage points?
- Notation
  - $\pi$ is proportion of voters expressing support for Biden (the estimand).
  - $X_1, X_2, \ldots X_n \sim \text{Bernoulli}(\pi)$ be the iid random variables for each respondent.
  - $\hat{\pi} = \bar{X}_n$ is our sample mean estimator.
- What do we know?
  - $E[\bar{X}_n] = E[X]$ and $V[\bar{X}_n] = \frac{V[X]}{n}$
Planning a Survey

- Problem: planning a survey to estimate Biden support in the 2020 election.
  - How many people should we survey so that our estimator has no less than a .95 probability of being within .02 of the true population proportion?
  - i.e. how do we get a margin of ± 2 percentage points?

- Notation
  - \( \pi \) is proportion of voters expressing support for Biden (the estimand).
  - \( X_1, X_2, \ldots X_n \sim \text{Bernoulli}(\pi) \) be the iid random variables for each respondent.
  - \( \hat{\pi} = \bar{X}_n \) is our sample mean estimator.

- What do we know?
  - \( E[\bar{X}_n] = E[X] \) and \( V[\bar{X}_n] = \frac{V[X]}{n} \)
  - We know Bernoulli has variance of \( \pi(1 - \pi) \) which is maximized at \( \pi = .5 \).
Planning a Survey

What does $n$ have to be to maintain $P(|\bar{X}_n - E[X]| \geq 0.02) \leq .05$?
Planning a Survey

What does $n$ have to be to maintain $P(|\bar{X}_n - E[X]| \geq 0.02) \leq 0.05$?

$$P \left[ |\bar{X}_n - E[X]| \geq \epsilon \right] \leq \frac{V[X]}{\epsilon^2 n}$$
Planning a Survey

What does \( n \) have to be to maintain \( P(|\bar{X}_n - E[X]| \geq 0.02) \leq 0.05 \)?

\[
P \left[ |\bar{X}_n - E[X]| \geq \epsilon \right] \leq \frac{V[X]}{\epsilon^2 n}
\]

\[
P \left[ |\bar{X}_n - E[X]| \geq 0.02 \right] \leq \frac{V[X]}{(0.02^2 n}
\]
Planning a Survey

What does $n$ have to be to maintain $P(|\bar{X}_n - E[X]| \geq 0.02) \leq .05$?

\[
P \left[ |\bar{X}_n - E[X]| \geq \varepsilon \right] \leq \frac{V[X]}{\varepsilon^2 n}
\]

\[
P \left[ |\bar{X}_n - E[X]| \geq 0.02 \right] \leq \frac{V[X]}{(0.02^2 n)} \leq \frac{\pi (1 - \pi)}{0.0004 n}
\]
Planning a Survey

What does $n$ have to be to maintain $P(|\bar{X}_n - E[X]| \geq 0.02) \leq .05$?

\[
P \left[ |\bar{X}_n - E[X]| \geq \epsilon \right] \leq \frac{V[X]}{\epsilon^2 n}
\]

\[
P \left[ |\bar{X}_n - E[X]| \geq 0.02 \right] \leq \frac{V[X]}{(0.02)^2 n}
\]

\[
\leq \frac{\pi (1 - \pi)}{0.0004 n}
\]

\[
\leq \frac{0.25}{0.0004 n}
\]
Planning a Survey

What does $n$ have to be to maintain $P(|\bar{X}_n - E[X]| \geq 0.02) \leq 0.05$?

\[
P \left[ |\bar{X}_n - E[X]| \geq \epsilon \right] \leq \frac{V[X]}{\epsilon^2 n}
\]

\[
P \left[ |\bar{X}_n - E[X]| \geq 0.02 \right] \leq \frac{V[X]}{(0.02^2 n)}
\]

\[
< \frac{\pi (1 - \pi)}{0.0004 n}
\]

\[
< \frac{0.25}{0.0004 n}
\]

\[
< \frac{1}{0.0016 n}
\]
Planning a Survey

What does $n$ have to be to maintain $P(|X_n - E[X]| \geq 0.02) \leq .05$?

$$P [|X_n - E[X]| \geq \epsilon] \leq \frac{V[X]}{\epsilon^2 n}$$

$$P [|X_n - E[X]| \geq 0.02] \leq \frac{V[X]}{(0.02^2 n)}$$

$$\leq \frac{\pi (1 - \pi)}{0.0004n}$$

$$\leq \frac{0.25}{0.0004n}$$

$$\leq \frac{1}{0.0016n}$$

We want to bound the probability by 0.05 which requires $\frac{1}{0.0016n} \leq 0.05$ which means...
Planning a Survey

What does $n$ have to be to maintain $P(|\bar{X}_n - E[X]| \geq 0.02) \leq .05$?

\[
P \left[ |\bar{X}_n - E[X]| \geq \epsilon \right] \leq \frac{V[X]}{\epsilon^2 n}
\]

\[
P \left[ |\bar{X}_n - E[X]| \geq 0.02 \right] \leq \frac{V[X]}{0.02^2 n}
\]

\[
\leq \frac{\pi(1 - \pi)}{0.0004 n} \leq 0.25
\]

\[
\leq \frac{0.0004 n}{0.0016 n} \leq \frac{1}{0.0016 n}
\]

We want to bound the probability by 0.05 which requires $\frac{1}{0.0016 n} \leq 0.05$ which means...we need $n \geq 12,500$ respondents!
Planning a Survey

That's an expensive survey! Do we really need that many people?

No! Chebyshev provides a bound that is guaranteed to hold (and in the worst case variance), but actual probabilities are much smaller.

We can use simulation to assess. Let's simulate a survey with $\frac{\pi}{12} = \frac{55}{100}$ and 12,500 respondents and see how many are far away?

```r
nsims <- 10000
holder <- vector(mode="numeric", length=nsims)
for (i in 1:nsims) {
  my.sample <- rbinom(n=12500, size=1, prob=.55)
  holder[i] <- mean(my.sample)
}
mean(abs(holder - .55) > 0.02)
```

None were outside the range!
Planning a Survey

That’s an expensive survey! Do we really need that many people?

No! Chebyshev provides a bound that is guaranteed to hold (and in the worst case variance), but actual probabilities are much smaller.

We can use simulation to assess. Let’s simulate a survey with $\pi = 0.55$ and 12,500 respondents and see how many are far away?

```r
nsims <- 10000
holder <- vector(mode="numeric", length=nsims)
for (i in 1:nsims) {
  my.sample <- rbinom(n=12500, size=1, prob=.55)
  holder[i] <- mean(my.sample)
}
mean(abs(holder - .55) > 0.02)

None were outside the range!
Planning a Survey

That’s an expensive survey! Do we really need that many people?

No! Chebyshev provides a bound that is guaranteed to hold (and in the worst case variance), but actual probabilities are much smaller.

```r
nsims <- 10000
holder <- vector(mode="numeric", length=nsims)
for (i in 1:nsims) {
  my.sample <- rbinom(n=12500, size=1, prob=.55)
  holder[i] <- mean(my.sample)
}
mean(abs(holder - .55) > 0.02)

None were outside the range!
```
Planning a Survey

That’s an expensive survey! Do we really need that many people?

No! Chebyshev provides a bound that is guaranteed to hold (and in the worst case variance), but actual probabilities are much smaller.

We can use simulation to assess. Let’s simulate a survey with $\pi = 0.55$ and 12,500 respondents and see how many are far away?

```r
cpy <- 0.55
nsims <- 10000
holder <- vector(mode="numeric", length=nsims)
for (i in 1:nsims) {
  my.sample <- rbinom(n=12500, size=1, prob=0.55)
  holder[i] <- mean(my.sample)
}
mean(abs(holder - 0.55) > 0.02)

None were outside the range!
Planning a Survey

That’s an expensive survey! Do we really need that many people?

No! Chebyshev provides a bound that is guaranteed to hold (and in the worst case variance), but actual probabilities are much smaller.

We can use simulation to assess. Let’s simulate a survey with $\pi = .55$ and 12,500 respondents and see how many are far away?

```r
nsims <- 10000
holder <- vector(mode="numeric", length=nsims)
for (i in 1:nsims) {
  my.sample <- rbinom(n=12500, size=1, prob=.55)
  holder[i] <- mean(my.sample)
}
mean(abs(holder - .55) > 0.02)
```
Planning a Survey

That's an expensive survey! Do we really need that many people?

No! Chebyshev provides a bound that is guaranteed to hold (and in the worst case variance), but actual probabilities are much smaller.

We can use simulation to assess. Let's simulate a survey with $\pi = .55$ and 12,500 respondents and see how many are far away?

```r
nsims <- 10000
holder <- vector(mode="numeric", length = nsims)
for (i in 1:nsims) {
  my.sample <- rbinom(n=12500, size=1, prob=.55)
  holder[i] <- mean(my.sample)
}
mean(abs(holder - .55) > 0.02)
```

None were outside the range!
We have IID random variables $X_1, \ldots, X_n$ with unknown distribution. Current knowledge about distribution of $X_n$ is:

- **Expectation** is $E[X_n] = E[X]$.
- **Variance** is $V[X_n] = V[X]$.

We still want to know more about the distribution of $X_n$. We can think about behavior of $X_n$ as $n$ gets large by thinking of the estimators with increasing sample sizes as a sequence of random variables $X_1, X_2, \ldots, X_n = X_1 + X_2 + \cdots + X_n$.

What does this sequence converge to?
Taking Stock

- We have IID random variables $X_1, \ldots, X_n$ with unknown distribution.
We have IID random variables $X_1, \ldots, X_n$ with unknown distribution.

Current knowledge about distribution of $\bar{X}_n$

- Expectation is $E[\bar{X}_n] = E[X]$
- Variance is $V[\bar{X}_n] = \frac{V[X]}{n}$
- Tail probabilities using the above and Chebyshev
Taking Stock

- We have IID random variables $X_1, \ldots, X_n$ with unknown distribution.
- Current knowledge about distribution of $\bar{X}_n$
  - Expectation is $E[\bar{X}_n] = E[X]$
  - Variance is $V[\bar{X}_n] = \frac{V[X]}{n}$
  - Tail probabilities using the above and Chebyshev
- We still want to know more about the distribution of $\bar{X}_n$. 

Stewart (Princeton)
We have IID random variables $X_1, \ldots, X_n$ with unknown distribution.

Current knowledge about distribution of $\overline{X}_n$

- Expectation is $E[\overline{X}_n] = E[X]$
- Variance is $V[\overline{X}_n] = \frac{V[X]}{n}$
- Tail probabilities using the above and Chebyshev

We still want to know more about the distribution of $\overline{X}_n$.

We can think about behavior of $\overline{X}_n$ as $n$ gets large by thinking of the estimators with increasing sample sizes as a sequence of random variables

$$\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_n = X_1, \frac{X_1 + X_2}{2}, \ldots, \frac{X_1 + \cdots + X_n}{n}$$
Taking Stock

- We have IID random variables $X_1, \ldots, X_n$ with unknown distribution.
- Current knowledge about distribution of $\bar{X}_n$
  - Expectation is $E[\bar{X}_n] = E[X]$
  - Variance is $V[\bar{X}_n] = \frac{V[X]}{n}$
  - Tail probabilities using the above and Chebyshev

- We still want to know more about the distribution of $\bar{X}_n$.
- We can think about behavior of $\bar{X}_n$ as $n$ gets large by thinking of the estimators with increasing sample sizes as a sequence of random variables

\[
\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n = X_1, \frac{X_1 + X_2}{2}, \ldots, \frac{X_1 + \cdots + X_n}{n}
\]

- What does this sequence converge to?
What do we mean by 'converge'?

Definition (Convergence in Probability)

Let \((T_1, T_2, T_3, \ldots)\) be a sequence of random variables and let \(c \in \mathbb{R}\). Then \(T_n\) converges in probability to \(c\) if for all accuracy levels satisfying \(\epsilon > 0\),

\[
\lim_{n \to \infty} P\left[ |T_n - c| \geq \epsilon \right] = 0
\]

We will write this as \(T_n \overset{p}{\to} c\) or \(\text{plim}_{n \to \infty} T_n = c\).

(Aronow and Miller Theorem 3.2.6)

Intuition: the probability that the random variable \(T_n\) lies outside a super tiny interval around \(c\) approaches zero as \(n\) approaches infinity.

NB: Any continuous function of the sequence itself convergence to the value of the function at the probability limit by the Continuous Mapping Theorem (Aronow and Miller Theorem 3.2.7).
What do we mean by 'converge'?

Definition (Convergence in Probability)

Let \(( T_1, T_2, T_3, \ldots )\) be a sequence of random variables and let \(c \in \mathbb{R}\). Then \( T_n \) converges in probability to \(c\) if for all accuracy levels satisfying \(\epsilon > 0\),

\[
\lim_{n \to \infty} P[|T_n - c| \geq \epsilon] = 0
\]

We will write this as

\[
T_n \xrightarrow{p} c \quad \text{or} \quad \lim_{n \to \infty} T_n = c.
\]

(Aronow and Miller Theorem 3.2.6)
What do we mean by 'converge'? 

**Definition (Convergence in Probability)**

Let \((T_1, T_2, T_3, \ldots)\) be a sequence of random variables and let \(c \in \mathbb{R}\). Then \(T_n\) converges in probability to \(c\) if for all accuracy levels satisfying \(\epsilon > 0\),

\[
\lim_{{n \to \infty}} P\left[ |T_n - c| \geq \epsilon \right] = 0
\]

We will write this as

\[
T_n \xrightarrow{p} c \quad \text{or} \quad \lim_{{n \to \infty}} T_n = c.
\]

(Aronow and Miller Theorem 3.2.6)

Intuition: the probability that the random variable \(T_n\) lies outside a super tiny interval around \(c\) approaches zero as \(n\) approaches infinity.
What do we mean by 'converge'?

**Definition (Convergence in Probability)**

Let \((T_1, T_2, T_3, \ldots)\) be a sequence of random variables and let \(c \in \mathbb{R}\). Then \(T_n\) converges in probability to \(c\) if for all accuracy levels satisfying \(\epsilon > 0\),

\[
\lim_{n \to \infty} P\left[|T_n - c| \geq \epsilon\right] = 0
\]

We will write this as

\[
T_n \xrightarrow{p} c \quad \text{or} \quad \operatorname{plim}_{n \to \infty} T_n = c.
\]

(Aronow and Miller Theorem 3.2.6)

**Intuition:** the probability that the random variable \(T_n\) lies outside a super tiny interval around \(c\) approaches zero as \(n\) approaches infinity.

**NB:** Any continuous function of the sequence itself convergence to the value of the function at the probability limit by the Continuous Mapping Theorem (Aronow and Miller Theorem 3.2.7 )
(Weak) Law of Large Numbers (WLLN)

Definition (Weak Law of Large Numbers)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with finite variance $\mathbb{V}[X] > 0$, and let $X_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then $X_n \overset{p}{\rightarrow} E[X]$. (Aronow and Miller Theorem 3.2.8)

Proof is based on Chebychev’s inequality for the mean plus a result called the Squeeze Theorem for Sequences.

Intuition: The probability of the sample mean being far away from the expectation of $X$ goes to zero as the sample size gets big. The distribution of $X_n$ collapses on $E[X]$. No assumptions necessary about the distribution of $X$ beyond i.i.d. sampling and a finite variance!
(Weak) Law of Large Numbers (WLLN)

Definition (Weak Law of Large Numbers)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with finite variance $V[X] > 0$, and let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$\bar{X}_n \xrightarrow{p} E[X]$$

(Aronow and Miller Theorem 3.2.8) Proof is based on Chebychev’s inequality for the mean plus a result called the Squeeze Theorem for Sequences.
(Weak) Law of Large Numbers (WLLN)

**Definition (Weak Law of Large Numbers)**

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with finite variance $V[X] > 0$, and let $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$\overline{X}_n \xrightarrow{p} E[X]$$

(Aronow and Miller Theorem 3.2.8) Proof is based on Chebychev’s inequality for the mean plus a result called the Squeeze Theorem for Sequences.

- **Intuition:** The probability of the sample mean being far away from the expectation of $X$ goes to zero as the sample size gets big.
(Weak) Law of Large Numbers (WLLN)

**Definition (Weak Law of Large Numbers)**

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with finite variance $V[X] > 0$, and let $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$\overline{X}_n \xrightarrow{p} E[X]$$

(Aronow and Miller Theorem 3.2.8) Proof is based on Chebychev’s inequality for the mean plus a result called the Squeeze Theorem for Sequences.

- Intuition: The probability of the sample mean being far away from the expectation of $X$ goes to zero as the sample size gets big.
- The distribution of $\overline{X}_n$ collapses on $E[X]$. 


(Weak) Law of Large Numbers (WLLN)

**Definition (Weak Law of Large Numbers)**

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with finite variance $\text{Var}[X] > 0$, and let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$\bar{X}_n \xrightarrow{p} E[X]$$

(Aronow and Miller Theorem 3.2.8) Proof is based on Chebychev’s inequality for the mean plus a result called the Squeeze Theorem for Sequences.

- **Intuition:** The probability of the sample mean being far away from the expectation of $X$ goes to zero as the sample size gets big.
- The distribution of $\bar{X}_n$ collapses on $E[X]$.
- No assumptions necessary about the distribution of $X$ beyond i.i.d. sampling and a finite variance!
Weak Law of Large Numbers

Proportion of Heads vs. Coin Flips

Stewart (Princeton) Week 3: Learning From Random Samples September 14-18, 2020 42 / 142
Weak Law of Large Numbers

- This is an incredibly useful result!
Weak Law of Large Numbers

- This is an incredibly useful result!
- As the sample mean gets large it approximates the expectation to any arbitrary degree of precision.
Weak Law of Large Numbers

- This is an incredibly useful result!
- As the sample mean gets large it approximates the expectation to any arbitrary degree of precision.
- An implication of the Weak Law of Large Numbers is that the CDF of $X$ can be estimated to arbitrary precision with random iid samples from $X$. We will return to this result in two videos.
Weak Law of Large Numbers

- This is an incredibly useful result!
- As the sample mean gets large it approximates the expectation to any arbitrary degree of precision.
- An implication of the Weak Law of Large Numbers is that the CDF of $X$ can be estimated to arbitrary precision with random iid samples from $X$. We will return to this result in two videos.

Okay that’s pretty cool, but we are almost ready to state the coolest result in statistics.
Convergence in Distribution

We want to know what form the sampling distribution will have asymptotically. For this, we need a notion of what it means for a distribution to converge.

**Definition (Convergence in Distribution)**

Let $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, ...)\) be a sequence of random variables with CDFs $(F_1, F_2, F_3, ...)\) and let $\mathcal{X}$ be a random variable with CDF $F_T$. Then $\mathcal{X}_n$ converges in distribution to $\mathcal{X}$ if for all $t \in \mathbb{R}$ at which $F_T$ is continuous

$$\lim_{n \to \infty} F_n(t) = F_T(t)$$

We write this as $\mathcal{X}_n \xrightarrow{d} \mathcal{X}$.

Intuition: when $n$ is big, the distribution of $\mathcal{X}_n$ is very similar to $F_T$, the distribution of $\mathcal{X}$.

We will call this the asymptotic distribution or the limit distribution.

NB: convergence in probability is a special case of convergence in distribution with a degenerate distribution.
Convergence in Distribution

We want to know what form the sampling distribution will have asymptotically. For this we need a notion of what it means for a distribution to converge.

Definition (Convergence in Distribution)

Let \((T(1), T(2), T(3), \ldots)\) be a sequence of random variables with CDFs \((F(1), F(2), F(3), \ldots)\) and let \(T\) be a random variable with CDF \(F_T\). Then \(T(n)\) converges in distribution to \(T\) if for all \(t \in \mathbb{R}\) at which \(F_T\) is continuous

\[
\lim_{n \to \infty} F(n)(t) = F_T(t).
\]

We write this as \(T(n) \xrightarrow{d} T\).

Intuition: when \(n\) is big, the distribution of \(T(n)\) is very similar to \(F_T\), the distribution of \(T\).

We will call this the asymptotic distribution or the limit distribution.

NB: convergence in probability is a special case of convergence in distribution with a degenerate distribution.
Convergence in Distribution

We want to know what form the sampling distribution will have asymptotically. For this we need a notion of what it means for a distribution to converge.

**Definition (Convergence in Distribution)**

Let \((T_1, T_2, T_3, \ldots)\) be a sequence of random variables with CDFs \((F_1, F_2, F_3, \ldots)\) and let \(T\) be a random variable with CDF \(F_T\). Then \(T_n\) converges in distribution to \(T\) if for all \(t \in \mathbb{R}\) at which \(F_T\) is continuous

\[
\lim_{n \to \infty} F_n(t) = F_T(t).
\]

We write this as

\[
T_n \xrightarrow{d} T.
\]
Convergence in Distribution

We want to know what form the sampling distribution will have asymptotically. For this we need a notion of what it means for a distribution to converge.

Definition (Convergence in Distribution)

Let \((T_1, T_2, T_3, \ldots)\) be a sequence of random variables with CDFs \((F_1, F_2, F_3, \ldots)\) and let \(T\) be a random variable with CDF \(F_T\). Then \(T_n\) converges in distribution to \(T\) if for all \(t \in \mathbb{R}\) at which \(F_T\) is continuous

\[
\lim_{n \to \infty} F_{(n)}(t) = F_T(t).
\]

We write this as

\[
T_n \xrightarrow{d} T.
\]

Intuition: when \(n\) is big, the distribution of \(T_n\) is very similar to \(F_T\), the distribution of \(T\).
Convergence in Distribution

We want to know what form the sampling distribution will have asymptotically. For this we need a notion of what it means for a distribution to converge.

**Definition (Convergence in Distribution)**

Let \((T_1, T_2, T_3, \ldots)\) be a sequence of random variables with CDFs \((F_1, F_2, F_3, \ldots)\) and let \(T\) be a random variable with CDF \(F_T\). Then \(T(n)\) converges in distribution to \(T\) if for all \(t \in \mathbb{R}\) at which \(F_T\) is continuous

\[
\lim_{n \to \infty} F(n)(t) = F_T(t).
\]

We write this as

\[
T(n) \xrightarrow{d} T.
\]

- **Intuition:** when \(n\) is big, the distribution of \(T(n)\) is very similar to \(F_T\), the distribution of \(T\).
- We will call this the asymptotic distribution or the limit distribution.
Convergence in Distribution

We want to know what form the sampling distribution will have asymptotically. For this we need a notion of what it means for a distribution to converge.

### Definition (Convergence in Distribution)

Let \((T_1, T_2, T_3, \ldots)\) be a sequence of random variables with CDFs \((F_1, F_2, F_3, \ldots)\) and let \(T\) be a random variable with CDF \(F_T\). Then \(T_n\) converges in distribution to \(T\) if for all \(t \in \mathbb{R}\) at which \(F_T\) is continuous

\[
\lim_{n \to \infty} F_n(t) = F_T(t).
\]

We write this as

\[
T_n \xrightarrow{d} T.
\]

- Intuition: when \(n\) is big, the distribution of \(T_n\) is very similar to \(F_T\), the distribution of \(T\).
- We will call this the asymptotic distribution or the limit distribution.
- NB: convergence in probability is a special case of convergence in distribution with a degenerate distribution.
Standardizing the Sample Mean

Last prerequisite!

Definition (Standardizing a Random Variable)
For i.i.d. random variables $X_1, X_2, \ldots, X_n$ with finite $E[X] = \mu$ and finite $V[X] = \sigma^2 > 0$, the standardized sample mean is

$$Z = \frac{(X - E[X])}{\sigma} = \sqrt{n} \left( \frac{X - \mu}{\sigma} \right)$$

(Aronow and Miller Definition 3.2.23)
For any $X$ this will have $E[Z] = 0$ and $V[Z] = 1$. This is often called the Z-score.
Standardizing the Sample Mean

Last prerequisite!

**Definition (Standardizing a Random Variable)**

For i.i.d. random variables $X_1, X_2, \ldots, X_n$ with finite $E[X] = \mu$ and finite $V[X] = \sigma^2 > 0$, the standardized sample mean is

$$Z = \frac{\overline{X} - E[\overline{X}]}{\sigma[\overline{X}]} = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma}$$

(Aronow and Miller Definition 3.2.23)
Standardizing the Sample Mean

Last prerequisite!

**Definition (Standardizing a Random Variable)**

For i.i.d. random variables $X_1, X_2, \ldots, X_n$ with finite $E[X] = \mu$ and finite $V[X] = \sigma^2 > 0$, the standardized sample mean is

$$Z = \frac{(\overline{X} - E[\overline{X}])}{\sigma[\overline{X}]} = \frac{\sqrt{n} (\overline{X} - \mu)}{\sigma}$$

(Aronow and Miller Definition 3.2.23)

- For any $X$ this will have $E[Z] = 0$ and $V[Z] = 1.$
Standardizing the Sample Mean

Last prerequisite!

**Definition (Standardizing a Random Variable)**

For i.i.d. random variables $X_1, X_2, \ldots, X_n$ with finite $E[X] = \mu$ and finite $V[X] = \sigma^2 > 0$, the standardized sample mean is

$$Z = \frac{(X - E[X])}{\sigma[X]} = \frac{\sqrt{n}}{\sigma} (X - \mu)$$

(Aronow and Miller Definition 3.2.23)

- For any $X$ this will have $E[Z] = 0$ and $V[Z] = 1$.
- This is often called the $Z$-score.
The Puzzle

Population Distribution

Sampling Distribution of the Mean

Sampling Distribution of the Mean

Sampling Distribution of the Mean

n=2

n=5

n=30
The Central Limit Theorem

Definition (Lindeberg-Lévy Central Limit Theorem)
Let $X_1, \ldots, X_n$ be i.i.d. random variables each with (finite) $E[X] = \mu$ and finite variance $\sigma^2 > 0$. Then, for any population distribution of $X$,

$$\sqrt{n}(X_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

CLT also implies that the standardized sample mean converges to a standard normal random variable:

$$Z_n \equiv \frac{X_n - E[X_n]}{\sqrt{V[X_n]}} = \frac{X_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1).$$

This is free of distribution assumptions on $X$!

This makes it easy to characterize the sampling distribution of the sample mean for large $n$.

NB: the equivalence of the two forms is due to Slutsky's Theorem (see e.g. Aronow and Miller Theorem 3.2.25).
The Central Limit Theorem

Definition (Lindeberg-Lévy Central Limit Theorem)

Let $X_1, \ldots, X_n$ be i.i.d. random variables each with (finite) $E[X] = \mu$ and finite variance $\sigma^2 > 0$. Then, for any population distribution of $X$,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$
The Central Limit Theorem

Definition (Lindeberg-Lévy Central Limit Theorem)

Let \( X_1, \ldots, X_n \) be i.i.d. random variables each with (finite) \( E[X] = \mu \) and finite variance \( \sigma^2 > 0 \). Then, for any population distribution of \( X \),

\[
\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).
\]

CLT also implies that the standardized sample mean converges to a standard normal random variable:

\[
Z_n \equiv \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{V[\bar{X}_n]}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).
\]
The Central Limit Theorem

Definition (Lindeberg-Lévy Central Limit Theorem)

Let $X_1, \ldots, X_n$ be i.i.d. random variables each with (finite) $E[X] = \mu$ and finite variance $\sigma^2 > 0$. Then, for any population distribution of $X$, 

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- CLT also implies that the standardized sample mean converges to a standard normal random variable:

$$Z_n \equiv \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{V[\bar{X}_n]}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

- This is free of distribution assumptions on $X$!
The Central Limit Theorem

### Definition (Lindeberg-Lévy Central Limit Theorem)

Let $X_1, \ldots, X_n$ be i.i.d. random variables each with (finite) $E[X] = \mu$ and finite variance $\sigma^2 > 0$. Then, for any population distribution of $X$,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- CLT also implies that the standardized sample mean converges to a standard normal random variable:

$$Z_n \equiv \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{V[\bar{X}_n]}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

- This is free of distribution assumptions on $X$!
- This makes it easy to characterize the sampling distribution of the sample mean for large $n$. 

---

**NB:** the equivalence of the two forms is due to Slutsky's Theorem (see e.g. Aronow and Miller Theorem 3.2.25).
The Central Limit Theorem

**Definition (Lindeberg-Lévy Central Limit Theorem)**

Let $X_1, ..., X_n$ be i.i.d. random variables each with (finite) $E[X] = \mu$ and finite variance $\sigma^2 > 0$. Then, for *any* population distribution of $X$,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

• CLT also implies that the standardized sample mean converges to a standard normal random variable:

$$Z_n \equiv \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{V[\bar{X}_n]}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

• This is *free of distribution assumptions on $X$!*
• This makes it easy to characterize the sampling distribution of the sample mean for large $n$.

• NB: the equivalence of the two forms is due to Slutsky’s Theorem (see e.g. Aronow and Miller Theorem 3.2.25).
Question:
Question:

As the number of observations in a dataset increases, which of the following is true?

A) The distribution of $X$ becomes more normally distributed.

B) The distribution of $X$ becomes more normally distributed.

C) Both statements are true.
Question:

As the number of observations in a dataset increases, which of the following is true?

A) The distribution of $X$ becomes more normally distributed.
Question:

As the number of observations in a dataset increases, which of the following is true?

A) The distribution of $X$ becomes more normally distributed.

B) The distribution of $\bar{X}$ becomes more normally distributed.
Question:

As the number of observations in a dataset increases, which of the following is true?

A) The distribution of $X$ becomes more normally distributed.
B) The distribution of $\bar{X}$ becomes more normally distributed.
C) Both statements are true.
Replanning that Survey

Recall we wanted to find $n$ such that,

$$P\left(\left|X_n - \pi\right| > 0.02\right) \leq 0.05.$$

By the CLT, for large $n$, then $X_n - \pi \sim N(0, \sigma^2 / n)$.

Plugging in our conservative variance of 0.25 we get $X_n - \pi \sim N(0, 1.4n)$.

Standardizing, we get

$$Z = \frac{X_n - \pi}{\sqrt{\frac{1}{4n}}} = 2\sqrt{n}(X_n - \pi) \sim N(0, 1).$$

It is easier to work with this standardized variable so:

$$P\left(\left|Z\right| > 0.02(2\sqrt{n})\right) \leq 0.05.$$
Replanning that Survey

Recall we wanted to find $n$ such that,

$$P(|\bar{X}_n - \pi| > 0.02) \leq 0.05$$
Replanning that Survey

Recall we wanted to find $n$ such that,

$$P(|\bar{X}_n - \pi| > 0.02) \leq 0.05$$

By the CLT, for large $n$, then

$$\bar{X}_n - \pi \sim \mathcal{N}(0, \sigma^2/n)$$
Replanning that Survey

Recall we wanted to find $n$ such that,

$$P(|\bar{X}_n - \pi| > 0.02) \leq 0.05$$

By the CLT, for large $n$, then

$$\bar{X}_n - \pi \sim \mathcal{N}(0, \sigma^2 / n)$$

Plugging in our conservative variance of 0.25 we get

$$\bar{X}_n - \pi \sim \mathcal{N}(0, \frac{1}{4n})$$
Replanning that Survey

Recall we wanted to find $n$ such that,

$$P(|\bar{X}_n - \pi| > 0.02) \leq 0.05$$

By the CLT, for large $n$, then

$$\bar{X}_n - \pi \sim \mathcal{N}(0, \sigma^2/n)$$

Plugging in our conservative variance of 0.25 we get

$$\bar{X}_n - \pi \sim \mathcal{N}(0, \frac{1}{4n})$$

Standardizing, we get

$$Z = \frac{\bar{X}_n - \pi}{1/\sqrt{4n}} = 2\sqrt{n}(\bar{X}_n - \pi) \sim \mathcal{N}(0, 1)$$
Replanning that Survey

Recall we wanted to find $n$ such that,

$$P(\left| \bar{X}_n - \pi \right| > 0.02) \leq 0.05$$

By the CLT, for large $n$, then

$$\bar{X}_n - \pi \sim \mathcal{N}(0, \sigma^2 / n)$$

Plugging in our conservative variance of 0.25 we get

$$\bar{X}_n - \pi \sim \mathcal{N}(0, \frac{1}{4n})$$

Standardizing, we get

$$Z = \frac{\bar{X}_n - \pi}{1/\sqrt{4n}} = 2\sqrt{n}(\bar{X}_n - \pi) \sim \mathcal{N}(0, 1)$$

It is easier to work with this standardized variable so:

$$P(|Z| > 0.02(2\sqrt{n})) \leq 0.05$$
Replanning that Survey

20.04 \sqrt{n} \leq 0.05

\[ P(Z < -0.04 \sqrt{n}) + P(Z > 0.04 \sqrt{n}) \leq 0.05 \]

The standard normal is symmetric around 0, so we can equivalently say,

\[ 2P(Z < -0.04 \sqrt{n}) \leq 0.025 \]

To solve for \( n \) we plug in the quantile \( P(Z \leq q) = 0.025 \) which we can get from the inverse CDF of the standard Normal.

Typing \( \text{qnorm}(0.025, \text{mean}=0, \text{sd}=1) \) in R gets us -1.96.

We need 

\[-0.04 \sqrt{n} \leq -1.96 \]

which is 

\[ n > 2401 \] respondents.

This is much lower than the 12,500 from Chebyshev, but that makes sense here because we used more information.
Replanning that Survey

\[
P(|Z| > 0.04\sqrt{n}) \leq 0.05
\]
Replanning that Survey

\[
P(\vert Z \vert > 0.04\sqrt{n}) \leq 0.05
\]

\[
P(Z < -0.04\sqrt{n}) + P(Z > 0.04\sqrt{n}) \leq 0.05
\]
Replanning that Survey

\[
P(|Z| > 0.04\sqrt{n}) \leq 0.05
\]

\[
P(Z < -0.04\sqrt{n}) + P(Z > 0.04\sqrt{n}) \leq 0.05
\]

The standard normal is symmetric around 0, so we can equivalently say,

\[
2P(Z < -0.04\sqrt{n}) \leq 0.05
\]
Replanning that Survey

\[
P(\left|Z\right| > 0.04\sqrt{n}) \leq 0.05
\]

\[
P(Z < -0.04\sqrt{n}) + P(Z > 0.04\sqrt{n}) \leq 0.05
\]

The standard normal is symmetric around 0, so we can equivalently say,

\[
2P(Z < -0.04\sqrt{n}) \leq 0.05
\]

\[
P(Z < -0.04\sqrt{n}) \leq 0.025
\]
Replanning that Survey

\[ P(|Z| > 0.04\sqrt{n}) \leq 0.05 \]
\[ P(Z < -0.04\sqrt{n}) + P(Z > 0.04\sqrt{n}) \leq 0.05 \]

The standard normal is symmetric around 0, so we can equivalently say,

\[ 2P(Z < -0.04\sqrt{n}) \leq 0.05 \]
\[ P(Z < -0.04\sqrt{n}) \leq 0.025 \]

To solve for \( n \) we plug in the quantile \( P(Z \leq q) = 0.025 \) which we can get from the inverse CDF of the standard Normal.
Replanning that Survey

\[ P(|Z| > 0.04\sqrt{n}) \leq 0.05 \]
\[ P(Z < -0.04\sqrt{n}) + P(Z > 0.04\sqrt{n}) \leq 0.05 \]

The standard normal is symmetric around 0, so we can equivalently say,

\[ 2P(Z < -0.04\sqrt{n}) \leq 0.05 \]
\[ P(Z < -0.04\sqrt{n}) \leq 0.025 \]

To solve for \( n \) we plug in the quantile \( P(Z \leq q) = 0.025 \) which we can get from the inverse CDF of the standard Normal. Typing `qnorm(0.025, mean=0, sd=1)` in R gets us -1.96.
Replanning that Survey

\[
P(|Z| > 0.04\sqrt{n}) \leq 0.05
\]
\[
P(Z < -0.04\sqrt{n}) + P(Z > 0.04\sqrt{n}) \leq 0.05
\]

The standard normal is symmetric around 0, so we can equivalently say,

\[
2P(Z < -0.04\sqrt{n}) \leq 0.05
\]
\[
P(Z < -0.04\sqrt{n}) \leq 0.025
\]

To solve for \( n \) we plug in the quantile \( P(Z \leq q) = 0.025 \) which we can get from the inverse CDF of the standard Normal.

Typing \texttt{qnorm(0.025, mean=0, sd=1)} in R gets us -1.96.

We need \(-0.04\sqrt{n} \leq -1.96\) which is \( n > 2401 \) respondents.
Replanning that Survey

\[ P(|Z| > 0.04\sqrt{n}) \leq 0.05 \]
\[ P(Z < -0.04\sqrt{n}) + P(Z > 0.04\sqrt{n}) \leq 0.05 \]

The standard normal is symmetric around 0, so we can equivalently say,

\[ 2P(Z < -0.04\sqrt{n}) \leq 0.05 \]
\[ P(Z < -0.04\sqrt{n}) \leq 0.025 \]

To solve for \( n \) we plug in the quantile \( P(Z \leq q) = 0.025 \) which we can get from the inverse CDF of the standard Normal. Typing `qnorm(0.025, mean=0, sd=1)` in R gets us -1.96. We need \(-0.04\sqrt{n} \leq -1.96\) which is \( n > 2401 \) respondents. This is much lower than the 12,500 from Chebyshev, but that makes sense here because we used more information.
Planning a Survey

We can use simulation to assess again. Let’s simulate a survey with $\pi = .55$ and 2,401 respondents and see how many our outside our prescribed margin of error?

```r
nsims <- 10000
holder <- vector(mode="numeric", length=nsims)
for (i in 1:nsims) {
  my.sample <- rbinom(n=2401, size=1, prob=.55)
  holder[i] <- mean(my.sample)
}
mean(abs(holder - .55) > 0.02)
We get 0.0485!
```
Planning a Survey

We can use simulation to assess again. Let’s simulate a survey with $\pi = .55$ and 2,401 respondents and see how many our outside our prescribed margin of error?

```r
nsims <- 10000
holder <- vector(mode="numeric", length=nsims)
for (i in 1:nsims) {
  my.sample <- rbinom(n=2401, size=1, prob=.55)
  holder[i] <- mean(my.sample)
}
mean(abs(holder - .55) > 0.02)
```

We get 0.0485!
Planning a Survey

We can use simulation to assess again. Let’s simulate a survey with \( \pi = .55 \) and 2,401 respondents and see how many are outside our prescribed margin of error?

```r
nsims <- 10000
holder <- vector(mode="numeric", length=nsims)
for (i in 1:nsims) {
    my.sample <- rbinom(n=2401, size=1, prob=.55)
    holder[i] <- mean(my.sample)
}
mean(abs(holder - .55) > 0.02)
```

We get 0.0485!
Real Talk: this has been a mathy video.
Real Talk: this has been a mathy video.

What you **need** to know:
Real Talk: this has been a mathy video.

What you need to know:

- two types of stochastic convergence

  ▶ Convergence in probability: values in the sequence eventually take a constant value (i.e. the limiting distribution is a point mass)
  ▶ Convergence in distribution: values in the sequence continue to vary, but the variation eventually comes to follow an unchanging distribution (i.e. the limiting distribution is a well characterized distribution)

intuition for the weak law of large numbers means will asymptotically have normal sampling distributions due to the central limit theorem

It is okay if you didn't follow all the math here. We will keep coming back to these ideas.
Real Talk: this has been a mathy video.

What you need to know:
- two types of stochastic convergence
  - Convergence in probability: values in the sequence eventually take a constant value (i.e. the limiting distribution is a point mass)
Real Talk: this has been a mathy video.

What you need to know:

- two types of stochastic convergence
  - **Convergence in probability**: values in the sequence eventually take a constant value (i.e. the limiting distribution is a point mass)
  - **Convergence in distribution**: values in the sequence continue to vary, but the variation eventually comes to follow an unchanging distribution (i.e. the limiting distribution is a well characterized distribution)
Real Talk: this has been a mathy video.

What you need to know:

- two types of stochastic convergence
  - Convergence in probability: values in the sequence eventually take a constant value (i.e. the limiting distribution is a point mass)
  - Convergence in distribution: values in the sequence continue to vary, but the variation eventually comes to follow an unchanging distribution (i.e. the limiting distribution is a well characterized distribution)
- intuition for the weak law of large numbers
Real Talk: this has been a mathy video.

What you need to know:

- two types of stochastic convergence
  - Convergence in probability: values in the sequence eventually take a constant value (i.e. the limiting distribution is a point mass)
  - Convergence in distribution: values in the sequence continue to vary, but the variation eventually comes to follow an unchanging distribution (i.e. the limiting distribution is a well characterized distribution)

- intuition for the weak law of large numbers
- means will asymptotically have normal sampling distributions due to the central limit theorem
Real Talk: this has been a mathy video.

What you need to know:

- two types of stochastic convergence
  - Convergence in probability: values in the sequence eventually take a constant value (i.e. the limiting distribution is a point mass)
  - Convergence in distribution: values in the sequence continue to vary, but the variation eventually comes to follow an unchanging distribution (i.e. the limiting distribution is a well characterized distribution)

- intuition for the weak law of large numbers

- means will asymptotically have normal sampling distributions due to the central limit theorem

- what asymptotic properties are
Real Talk: this has been a mathy video.

What you need to know:

- two types of stochastic convergence
  - **Convergence in probability**: values in the sequence eventually take a constant value (i.e. the limiting distribution is a point mass)
  - **Convergence in distribution**: values in the sequence continue to vary, but the variation eventually comes to follow an unchanging distribution (i.e. the limiting distribution is a well characterized distribution)

- intuition for the weak law of large numbers
- means will asymptotically have normal sampling distributions due to the central limit theorem
- what asymptotic properties are

It is okay if you didn’t follow all the math here. We will keep coming back to these ideas.
The Central Limit Theorem is deep and amazing. Want to learn more about Central Limit Theorem?

Watch this video (Joe Blitzstein):
https://www.youtube.com/watch?v=OprNqnHsVIA&list=PLLVp1P80IVc8EktkrD3Q8td0GmId7DjW0&index=31&t=0s

There are many CLT variants that deal with non-iid random variables as well!
We Covered...
We Covered . . .

- Chebychev’s Inequality
We Covered...

- Chebychev’s Inequality
- Weak Law of Large Numbers
We Covered...

- Chebychev’s Inequality
- Weak Law of Large Numbers
- Central Limit Theorem
We Covered...

- Chebychev’s Inequality
- Weak Law of Large Numbers
- Central Limit Theorem

Next time: properties of estimators.
Where We’ve Been and Where We’re Going…

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
  - hypothesis testing
  - what is regression?

- **Long Run**
  - probability $\rightarrow$ inference $\rightarrow$ regression $\rightarrow$ causal inference
1. Estimation
   - Populations and Samples
   - Estimators
   - Analytical

2. Weak Law of Large Numbers and the Central Limit Theorem
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3. Properties of Estimators
   - Four Desirable Properties
   - Example

4. Interval Estimation
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5. Plug-In Principle
1. Estimation
   - Populations and Samples
   - Estimators
   - Analytical

2. Weak Law of Large Numbers and the Central Limit Theorem
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3. Properties of Estimators
   - Four Desirable Properties
   - Example

4. Interval Estimation
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5. Plug-In Principle
Example: The scariest pieces of mail ever!

Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment

ALAN S. GERBER  Yale University
DONALD P. GREEN  Yale University
CHRISTOPHER W. LARIMER  University of Northern Iowa

https://doi.org/10.1017/S000305540808009X
Voter turnout theories based on rational self-interested behavior generally fail to predict significant turnout unless they account for the utility that citizens receive from performing their civic duty. We distinguish between two aspects of this type of utility, intrinsic satisfaction from behaving in accordance with a norm and extrinsic incentives to comply, and test the effects of priming intrinsic motives and applying varying degrees of extrinsic pressure. A large-scale field experiment involving several hundred thousand registered voters used a series of mailings to gauge these effects. Substantially higher turnout was observed among those who received mailings promising to publicize their turnout to their household or their neighbors. These findings demonstrate the profound importance of social pressure as an inducement to political participation.

https://doi.org/10.1017/S000305540808009X
Example: The scariest pieces of mail ever!

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We’ve been talking about the problem for years, but it only seems to get worse. This year, we’re taking a new approach. We’re sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

<table>
<thead>
<tr>
<th>MAPLE DR</th>
<th>Aug 04</th>
<th>Nov 04</th>
<th>Aug 06</th>
</tr>
</thead>
<tbody>
<tr>
<td>9995 JOSEPH JAMES SMITH</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9995 JENNIFER KAY SMITH</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9997 RICHARD B JACKSON</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9999 KATHY MARIE JACKSON</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

https://doi.org/10.1017/S000305540808009X
Basic Analysis

They make their data available (https://isps.yale.edu/research/data/d001). We can analyze it.

```r
load("gerber_green_larimer.RData")

# turn turnout variable into a numeric
social$voted <- 1 * (social$voted == "Yes")

neigh.mean <- mean(social$voted[social$treatment == "Neighbors"])

contr.mean <- mean(social$voted[social$treatment == "Civic Duty"])

neigh.mean - contr.mean
```

$$0.378 - 0.315 = 0.063$$

Is this a “real” effect? Is it big?
They make their data available (https://isps.yale.edu/research/data/d001). We can analyze it.

```r
load("gerber_green_larimer.RData")

## turn turnout variable into a numeric
social$voted <- 1 * (social$voted == "Yes")

neigh.mean <- mean(social$voted[social$treatment == "Neighbors"])
neigh.mean

contr.mean <- mean(social$voted[social$treatment == "Civic Duty"])
contr.mean

neigh.mean - contr.mean
```

Is this a “real” effect? Is it big?
Basic Analysis

They make their data available (https://isps.yale.edu/research/data/d001). We can analyze it.

load("gerber_green_larimer.RData")
## turn turnout variable into a numeric
social$voted <- 1 * (social$voted == "Yes")
neigh.mean <- mean(social$voted[social$treatment == "Neighbors"])
neigh.mean
contr.mean <- mean(social$voted[social$treatment == "Civic Duty"])
contr.mean
neigh.mean - contr.mean

.378 – .315 = .063
Basic Analysis

They make their data available (https://isps.yale.edu/research/data/d001). We can analyze it.

```
load("gerber_green_larimer.RData")
## turn turnout variable into a numeric
social$voted <- 1 * (social$voted == "Yes")
neigh.mean <- mean(social$voted[social$treatment == "Neighbors"])
neigh.mean
contr.mean <- mean(social$voted[social$treatment == "Civic Duty"])
contr.mean
neigh.mean - contr.mean

.378 - .315 = .063
```

Is this a “real” effect? Is it big?
Desirable Properties of Estimators

Sometimes there are many possible estimators for a given parameter. Which one should we choose? We'd like an estimator that gets the right answer on average. We'd like an estimator that doesn't change much from sample to sample. We'd like an estimator that gets closer to the right answer (probabilistically) as the sample size increases. We'd like an estimator that has a known sampling distribution (approximately) when the sample size is large.
Desirable Properties of Estimators

Sometimes there are many possible estimators for a given parameter. Which one should we choose?
Desirable Properties of Estimators

Sometimes there are many possible estimators for a given parameter. Which one should we choose?

- We’d like an estimator that gets the right answer on average.
Desirable Properties of Estimators

Sometimes there are many possible estimators for a given parameter. Which one should we choose?

- We’d like an estimator that gets the right answer on average.
- We’d like an estimator that doesn’t change much from sample to sample.
Desirable Properties of Estimators

Sometimes there are many possible estimators for a given parameter. Which one should we choose?

- We’d like an estimator that gets the right answer on average.
- We’d like an estimator that doesn’t change much from sample to sample.
- We’d like an estimator that gets closer to the right answer (probabilistically) as the sample size increases.
Desirable Properties of Estimators

Sometimes there are many possible estimators for a given parameter. Which one should we choose?

- We’d like an estimator that gets the right answer on average.
- We’d like an estimator that doesn’t change much from sample to sample.
- We’d like an estimator that gets closer to the right answer (probabilistically) as the sample size increases.
- We’d like an estimator that has a known sampling distribution (approximately) when the sample size is large.
Properties of Estimators

Estimators are random variables, for which randomness comes from repeated sampling from the population. The distribution of an estimator due to repeated sampling is called the sampling distribution. The properties of an estimator refer to the characteristics of its sampling distribution.

Finite-sample Properties (apply for any sample size):
- **Unbiasedness**: Is the sampling distribution of our estimator centered at the true parameter value?
  \[ \mathbb{E}[\hat{\mu}] = \mu \]

- **Efficiency**: Is the variance of the sampling distribution of our estimator reasonably small?
  \[ \text{Var}(\hat{\mu}_1) < \text{Var}(\hat{\mu}_2) \]

Asymptotic Properties (kick in when \( n \) is large):
- **Consistency**: As our sample size grows to infinity, does the sampling distribution of our estimator converge to the true parameter value?

- **Asymptotic Normality**: As our sample size grows large, does the sampling distribution of our estimator approach a normal distribution?
Properties of Estimators

Estimators are random variables, for which randomness comes from repeated sampling from the population.
Properties of Estimators

Estimators are random variables, for which randomness comes from repeated sampling from the population.

The distribution of an estimator due to repeated sampling is called the sampling distribution.
Properties of Estimators

Estimators are random variables, for which randomness comes from repeated sampling from the population.

The distribution of an estimator due to repeated sampling is called the sampling distribution.

The properties of an estimator refer to the characteristics of its sampling distribution.
Properties of Estimators

Estimators are random variables, for which randomness comes from repeated sampling from the population.

The distribution of an estimator due to repeated sampling is called the sampling distribution.

The properties of an estimator refer to the characteristics of its sampling distribution.

Finite-sample Properties (apply for any sample size):

Unbiasedness: Is the sampling distribution of our estimator centered at the true parameter value?

\[ \mathbb{E}[\hat{\mu}] = \mu \]

Efficiency: Is the variance of the sampling distribution of our estimator reasonably small?

\[ \text{Var}[\hat{\mu}_1] < \text{Var}[\hat{\mu}_2] \]

Asymptotic Properties (kick in when \( n \) is large):

Consistency: As our sample size grows to infinity, does the sampling distribution of our estimator converge to the true parameter value?

Asymptotic Normality: As our sample size grows large, does the sampling distribution of our estimator approach a normal distribution?

Stewart (Princeton)
Properties of Estimators

Estimators are random variables, for which randomness comes from repeated sampling from the population.

The distribution of an estimator due to repeated sampling is called the sampling distribution.

The properties of an estimator refer to the characteristics of its sampling distribution.

Finite-sample Properties (apply for any sample size):

- **Unbiasedness**: Is the sampling distribution of our estimator centered at the true parameter value? $E[\hat{\mu}] = \mu$

Asymptotic Properties (kick in when $n$ is large):

- **Consistency**: As our sample size grows to infinity, does the sampling distribution of our estimator converge to the true parameter value?
- **Asymptotic Normality**: As our sample size grows large, does the sampling distribution of our estimator approach a normal distribution?
Properties of Estimators

Estimators are random variables, for which randomness comes from repeated sampling from the population.

The distribution of an estimator due to repeated sampling is called the sampling distribution.

The properties of an estimator refer to the characteristics of its sampling distribution.

Finite-sample Properties (apply for any sample size):

- **Unbiasedness**: Is the sampling distribution of our estimator centered at the true parameter value? \( E[\hat{\mu}] = \mu \)

- **Efficiency**: Is the variance of the sampling distribution of our estimator reasonably small? \( V[\hat{\mu}_1] < V[\hat{\mu}_2] \)

Asymptotic Properties (kick in when \( n \) is large):

- **Consistency**: As our sample size grows to infinity, does the sampling distribution of our estimator converge to the true parameter value?

- **Asymptotic Normality**: As our sample size grows large, does the sampling distribution of our estimator approach a normal distribution?
Properties of Estimators

Estimators are random variables, for which randomness comes from repeated sampling from the population.

The distribution of an estimator due to repeated sampling is called the sampling distribution.

The properties of an estimator refer to the characteristics of its sampling distribution.

Finite-sample Properties (apply for any sample size):

- **Unbiasedness**: Is the sampling distribution of our estimator centered at the true parameter value? $E[\hat{\mu}] = \mu$
- **Efficiency**: Is the variance of the sampling distribution of our estimator reasonably small? $V[\hat{\mu}_1] < V[\hat{\mu}_2]$

Asymptotic Properties (kick in when $n$ is large):


Properties of Estimators

Estimators are random variables, for which randomness comes from **repeated sampling** from the population.

The distribution of an estimator due to repeated sampling is called the **sampling distribution**.

The properties of an estimator refer to the characteristics of its sampling distribution.

Finite-sample Properties (apply for any sample size):

- **Unbiasedness**: Is the sampling distribution of our estimator centered at the true parameter value? \( E[\hat{\mu}] = \mu \)

- **Efficiency**: Is the variance of the sampling distribution of our estimator reasonably small? \( V[\hat{\mu}_1] < V[\hat{\mu}_2] \)

Asymptotic Properties (kick in when \( n \) is large):

- **Consistency**: As our sample size grows to infinity, does the sampling distribution of our estimator converge to the true parameter value?
Properties of Estimators

Estimators are random variables, for which randomness comes from repeated sampling from the population.

The distribution of an estimator due to repeated sampling is called the sampling distribution.

The properties of an estimator refer to the characteristics of its sampling distribution.

Finite-sample Properties (apply for any sample size):

- **Unbiasedness**: Is the sampling distribution of our estimator centered at the true parameter value? $E[\hat{\mu}] = \mu$

- **Efficiency**: Is the variance of the sampling distribution of our estimator reasonably small? $V[\hat{\mu}_1] < V[\hat{\mu}_2]$

Asymptotic Properties (kick in when $n$ is large):

- **Consistency**: As our sample size grows to infinity, does the sampling distribution of our estimator converge to the true parameter value?

- **Asymptotic Normality**: As our sample size grows large, does the sampling distribution of our estimator approach a normal distribution?
1: Bias
(not getting the right answer on average)

Definition
Bias is the expected difference between the estimator and the parameter.

Over repeated samples, an unbiased estimator is right on average.

Bias(\hat{\mu}) = E[\hat{\mu} - E[X]] = E[\hat{\mu}] - \mu

Bias is not the difference between a particular estimate and the parameter.

For example, Bias(X_n) \neq E[x_n - E[X]]

An estimator is unbiased if and only if:

Bias(\hat{\mu}) = 0
1: Bias
(not getting the right answer on average)

**Definition**

Bias is the expected difference between the estimator and the parameter. Over repeated samples, an unbiased estimator is right on average.

Bias(\(\hat{\mu}\)) = E[\(\hat{\mu} - E[X]\)] = E[\(\hat{\mu}\)] - \(\mu\)

Bias is not the difference between a particular estimate and the parameter. For example, \(\text{Bias}(X_n) \neq E[x_n - E[X]]\)

An estimator is unbiased if and only if:

\(\text{Bias}(\hat{\mu}) = 0\)
1: Bias
(not getting the right answer on average)

Definition

Bias is the expected difference between the estimator and the parameter. Over repeated samples, an unbiased estimator is right on average.

\[
\text{Bias}(\hat{\mu}) = E[\hat{\mu} - E[X]] = E[\hat{\mu}] - \mu
\]
1: Bias
(not getting the right answer on average)

**Definition**

Bias is the expected difference between the estimator and the parameter. Over repeated samples, an unbiased estimator is right on average.

\[
\text{Bias}(\hat{\mu}) = E[\hat{\mu} - E[X]]
\]

\[
= E[\hat{\mu}] - \mu
\]

Bias is **not** the difference between a particular estimate and the parameter. For example,

\[
\text{Bias}(\overline{X}_n) \neq E[\overline{x}_n - E[X]]
\]
1: Bias
(not getting the right answer on average)

**Definition**

Bias is the expected difference between the **estimator** and the parameter. Over repeated samples, an unbiased estimator is right on average.

\[
\text{Bias}(\hat{\mu}) = E[\hat{\mu} - E[X]] = E[\hat{\mu}] - \mu
\]

Bias is **not** the difference between a particular estimate and the parameter. For example,

\[
\text{Bias}(\bar{X}_n) \neq E[\bar{x}_n - E[X]]
\]

An estimator is **unbiased** if and only if:

\[
\text{Bias}(\hat{\mu}) = 0
\]
Example: Estimators for Population Mean

Candidate estimators:

1. \( \hat{\mu}_1 = Y_1 \) (the first observation)
2. \( \hat{\mu}_2 = \frac{1}{2}(Y_1 + Y_n) \) (average of the first and last observation)
3. \( \hat{\mu}_3 = \frac{42}{4} \)
4. \( \hat{\mu}_4 = \bar{Y}_n \) (the sample average)

How do we choose between these estimators?
Example: Estimators for Population Mean

Candidate estimators:

1. $\hat{\mu}_1 = Y_1$ (the first observation)
2. $\hat{\mu}_2 = \frac{1}{2}(Y_1 + Y_n)$ (average of the first and last observation)
3. $\hat{\mu}_3 = 42$
4. $\hat{\mu}_4 = \overline{Y}_n$ (the sample average)

How do we choose between these estimators?
Bias of Example Estimators

Which of these estimators are unbiased?

1. \(E[Y_1 - \mu] = \)
2. \(E\left[\frac{1}{2}(Y_1 + Y_n) - \mu\right] = \)
3. \(E[42 - \mu] = \)
4. \(E[\bar{Y}_n - \mu] = \)

Estimators 1, 2, and 4 are unbiased because they get the right answer on average. Estimator 3 is biased.
Bias of Example Estimators

Which of these estimators are unbiased?

1. \( E[Y_1 - \mu] = \mu - \mu = 0 \)
2. \( E[\frac{1}{2}(Y_1 + Y_n) - \mu] = \)
3. \( E[42 - \mu] = \)
4. \( E[\bar{Y}_n - \mu] = \)
Bias of Example Estimators

Which of these estimators are unbiased?

1. \( E[Y_1 - \mu] = \mu - \mu = 0 \)
2. \( E[\frac{1}{2}(Y_1 + Y_n) - \mu] = \frac{1}{2}(E[Y_1] + E[Y_n]) - \mu \)
3. \( E[42 - \mu] = \)
4. \( E[Y_n - \mu] = \)
Bias of Example Estimators

Which of these estimators are unbiased?

1. \( E[Y_1 - \mu] = \mu - \mu = 0 \)
2. \( E[\frac{1}{2}(Y_1 + Y_n) - \mu] = \frac{1}{2}(E[Y_1] + E[Y_n]) - \mu = \frac{1}{2}(\mu + \mu) - \mu = 0 \)
3. \( E[42 - \mu] = \)
4. \( E[\bar{Y}_n - \mu] = \)
Bias of Example Estimators

Which of these estimators are unbiased?

1. \( E[Y_1 - \mu] = \mu - \mu = 0 \)
2. \( E\left[ \frac{1}{2} (Y_1 + Y_n) - \mu \right] = \frac{1}{2} (E[Y_1] + E[Y_n]) - \mu = \frac{1}{2} (\mu + \mu) - \mu = 0 \)
3. \( E[42 - \mu] = 42 - \mu \)
4. \( E[\bar{Y}_n - \mu] = \)
Bias of Example Estimators

Which of these estimators are unbiased?

1. \[ E[Y_1 - \mu] = \mu - \mu = 0 \]
2. \[ E\left[\frac{1}{2} (Y_1 + Y_n) - \mu\right] = \frac{1}{2} (E[Y_1] + E[Y_n]) - \mu = \frac{1}{2} (\mu + \mu) - \mu = 0 \]
3. \[ E[42 - \mu] = 42 - \mu \]
4. \[ E[\bar{Y}_n - \mu] = \frac{1}{n} \sum_{1}^{n} E[Y_i] - \mu \]
Bias of Example Estimators

Which of these estimators are unbiased?

1. $E[Y_1 - \mu] = \mu - \mu = 0$
2. $E\left[\frac{1}{2}(Y_1 + Y_n) - \mu\right] = \frac{1}{2}(E[Y_1] + E[Y_n]) - \mu = \frac{1}{2}(\mu + \mu) - \mu = 0$
3. $E[42 - \mu] = 42 - \mu$
4. $E[\bar{Y}_n - \mu] = \frac{1}{n} \sum_1^n E[Y_i] - \mu = \mu - \mu = 0$

Estimators 1, 2, and 4 are unbiased because they get the right answer on average. Estimator 3 is biased.
Bias of Example Estimators

Which of these estimators are unbiased?

1. \[ E[Y_1 - \mu] = \mu - \mu = 0 \]
2. \[ E[\frac{1}{2}(Y_1 + Y_n) - \mu] = \frac{1}{2}(E[Y_1] + E[Y_n]) - \mu = \frac{1}{2}(\mu + \mu) - \mu = 0 \]
3. \[ E[42 - \mu] = 42 - \mu \]
4. \[ E[\bar{Y}_n - \mu] = \frac{1}{n} \sum_{i=1}^{n} E[Y_i] - \mu = \mu - \mu = 0 \]

- Estimators 1, 2, and 4 are unbiased because they get the right answer on average.
Bias of Example Estimators

Which of these estimators are unbiased?

1. \( E[Y_1 - \mu] = \mu - \mu = 0 \)
2. \( E\left[\frac{1}{2}(Y_1 + Y_n) - \mu\right] = \frac{1}{2}(E[Y_1] + E[Y_n]) - \mu = \frac{1}{2} (\mu + \mu) - \mu = 0 \)
3. \( E[42 - \mu] = 42 - \mu \)
4. \( E[\bar{Y}_n - \mu] = \frac{1}{n} \sum_1^n E[Y_i] - \mu = \mu - \mu = 0 \)

- Estimators 1, 2, and 4 are unbiased because they get the right answer on average.
- Estimator 3 is biased.
Age population distribution in blue

Sampling Distribution for $\bar{X}_4$

Sampling Distribution for $\tilde{X}_4$
Age population distribution in blue, sampling distributions in red

Sampling Distribution for $\bar{X}_4$

Sampling Distribution for $\tilde{X}_4$
Age population distribution in blue, sampling distributions in red
2: Efficiency

(doesn’t change much sample to sample)

- All else equal, we prefer estimators that have a sampling distribution with smaller variance.
2: Efficiency

(doesn’t change much sample to sample)

● All else equal, we prefer estimators that have a sampling distribution with smaller variance.

Definition (Efficiency)

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of $\theta$, then $\hat{\theta}_1$ is more efficient relative to $\hat{\theta}_2$ iff

$$V[\hat{\theta}_1] < V[\hat{\theta}_2]$$

● Under repeated sampling, estimates based on $\hat{\theta}_1$ are likely to be closer to $\theta$
2: Efficiency

doesn’t change much sample to sample

- All else equal, we prefer estimators that have a sampling distribution with smaller variance.

Definition (Efficiency)

If \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) are unbiased estimators of \( \theta \), then \( \hat{\theta}_1 \) is more efficient relative to \( \hat{\theta}_2 \) iff

\[
V[\hat{\theta}_1] < V[\hat{\theta}_2]
\]

- Under repeated sampling, estimates based on \( \hat{\theta}_1 \) are likely to be closer to \( \theta \)
- Note that this does **not** imply that a particular estimate is always close to the true parameter value
2: Efficiency

(doesn’t change much sample to sample)

- All else equal, we prefer estimators that have a sampling distribution with smaller variance.

**Definition (Efficiency)**

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of $\theta$, then $\hat{\theta}_1$ is more efficient relative to $\hat{\theta}_2$ iff

$$V[\hat{\theta}_1] < V[\hat{\theta}_2]$$

- Under repeated sampling, estimates based on $\hat{\theta}_1$ are likely to be closer to $\theta$.
- Note that this does **not** imply that a particular estimate is always close to the true parameter value.
- The standard deviation of the sampling distribution of an estimator, $\sqrt{V[\hat{\theta}]}$, is often called the **standard error** of the estimator.
2: Efficiency

(doesn’t change much sample to sample)

- All else equal, we prefer estimators that have a sampling distribution with smaller variance.

**Definition (Efficiency)**

If \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) are unbiased estimators of \( \theta \), then \( \hat{\theta}_1 \) is more efficient relative to \( \hat{\theta}_2 \) iff

\[
V[\hat{\theta}_1] < V[\hat{\theta}_2]
\]

- Under repeated sampling, estimates based on \( \hat{\theta}_1 \) are likely to be closer to \( \theta \)
- Note that this does **not** imply that a particular estimate is always close to the true parameter value
- The standard deviation of the sampling distribution of an estimator, \( \sqrt{V[\hat{\theta}]} \), is often called the **standard error** of the estimator

Aronow and Miller discuss efficiency in terms of MSE (more on this in a second).
Variance of Example Estimators

What is the variance of our estimators?

1. $V[Y_1] = \sigma^2$
2. $V\left[\frac{1}{2}(Y_1 + Y_n)\right] = \frac{1}{4}V[Y_1 + Y_n] = \frac{1}{2}\sigma^2$
3. $V[42] = 0$
4. $V[Y_n] = \frac{1}{n}n\sigma^2 = \sigma^2$

Among the unbiased estimators, the sample average has the smallest variance. This means that Estimator 4 (the sample average) is likely to be closer to the true value $\mu$, than Estimators 1 and 2.
Variance of Example Estimators

What is the variance of our estimators?

1. $V[Y_1] = \sigma^2$
2. $V[\frac{1}{2}(Y_1 + Y_n)] = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{1}{2}\sigma^2$
3. $V[42] = 0$
4. $V[\bar{Y}_n] = \frac{1}{n}\sigma^2$

Among the unbiased estimators, the sample average has the smallest variance. This means that Estimator 4 (the sample average) is likely to be closer to the true value $\mu$, than Estimators 1 and 2.
Variance of Example Estimators

What is the variance of our estimators?

1. \( V[Y_1] = \sigma^2 \)

2. \( V[\frac{1}{2}(Y_1 + Y_n)] = \frac{1}{4} V[Y_1 + Y_n] \)

3. \( V[42] = \)

4. \( V[\bar{Y}_n] = \)
Variance of Example Estimators

What is the variance of our estimators?

1. \( V[Y_1] = \sigma^2 \)
2. \( V[\frac{1}{2}(Y_1 + Y_n)] = \frac{1}{4} V[Y_1 + Y_n] = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{1}{2}\sigma^2 \)
3. \( V[42] = \)
4. \( V[\bar{Y}_n] = \)
Variance of Example Estimators

What is the variance of our estimators?

1. \( V[Y_1] = \sigma^2 \)
2. \( V\left[\frac{1}{2}(Y_1 + Y_n)\right] = \frac{1}{4} V[Y_1 + Y_n] = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{1}{2}\sigma^2 \)
3. \( V[42] = 0 \)
4. \( V[\bar{Y}_n] = \)

What is the variance of our estimators?

1. $V[Y_1] = \sigma^2$
2. $V[\frac{1}{2}(Y_1 + Y_n)] = \frac{1}{4} V[Y_1 + Y_n] = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{1}{2}\sigma^2$
3. $V[42] = 0$
4. $V[Y_n] = \frac{1}{n^2} \sum_{1}^{n} V[Y_i]$
Variance of Example Estimators

What is the variance of our estimators?

1. $V[Y_1] = \sigma^2$
2. $V[\frac{1}{2}(Y_1 + Y_n)] = \frac{1}{4} V[Y_1 + Y_n] = \frac{1}{4} (\sigma^2 + \sigma^2) = \frac{1}{2} \sigma^2$
3. $V[42] = 0$
4. $V[\bar{Y}_n] = \frac{1}{n^2} \sum_{1}^{n} V[Y_i] = \frac{1}{n^2} n\sigma^2 = \frac{1}{n} \sigma^2$

Among the unbiased estimators, the sample average has the smallest variance. This means that Estimator 4 (the sample average) is likely to be closer to the true value $\mu$, than Estimators 1 and 2.
What is the variance of our estimators?

1. $V[Y_1] = \sigma^2$

2. $V[\frac{1}{2}(Y_1 + Y_n)] = \frac{1}{4} V[Y_1 + Y_n] = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{1}{2}\sigma^2$

3. $V[42] = 0$

4. $V[\bar{Y}_n] = \frac{1}{n^2} \sum_1^n V[Y_i] = \frac{1}{n^2} n\sigma^2 = \frac{1}{n}\sigma^2$

Among the unbiased estimators, the sample average has the smallest variance. This means that Estimator 4 (the sample average) is likely to be closer to the true value $\mu$, than Estimators 1 and 2.
Age population distribution in blue

Sampling Distribution for $\bar{X}_4$

Sampling Distribution for $\tilde{X}_4$
Age population distribution in blue, sampling distributions in red

Sampling Distribution for $\bar{X}_4$

Sampling Distribution for $\tilde{X}_4$
Age population distribution in blue, sampling distributions in red

Sampling Distribution for $\bar{X}_4$

Sampling Distribution for $\tilde{X}_4$
Trading Off Bias and Variance

Salganik (2018), Figure 3.1
Mean Squared Error

How can we choose between an unbiased estimator and a biased, but lower variance estimator?
Mean Squared Error

How can we choose between an unbiased estimator and a biased, but lower variance estimator?

**Definition (Mean Squared Error)**

To compare estimators in terms of both efficiency and unbiasedness we can use the Mean Squared Error (MSE), the expected squared difference between $\hat{\theta}$ and $\theta$:

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Bias(\hat{\theta})^2 + V(\hat{\theta}) = \left[ E[\hat{\theta}] - \theta \right]^2 + V(\hat{\theta})$$
Mean Squared Error

How can we choose between an unbiased estimator and a biased, but lower variance estimator?

**Definition (Mean Squared Error)**

To compare estimators in terms of both efficiency and unbiasedness we can use the Mean Squared Error (MSE), the expected squared difference between \( \hat{\theta} \) and \( \theta \):

\[
MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Bias(\hat{\theta})^2 + V(\hat{\theta}) = \left[ E[\hat{\theta}] - \theta \right]^2 + V(\hat{\theta})
\]

Sometimes (as in Aronow and Miller Definition 3.2.16) efficiency is defined as having lower MSE.
3-4: Asymptotic Evaluations: A Brief Review

(what happens as sample size increases)

Unbiasedness and efficiency are finite-sample properties of estimators, which hold regardless of sample size. Estimators also have asymptotic properties, i.e., the characteristics of sampling distributions when sample size becomes infinitely large. To define asymptotic properties, consider a sequence of estimators at increasing sample sizes: \( \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n \). For example, the sequence of sample means \( \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n \) is defined as:

\[
\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n = \frac{X_1}{n}, \frac{X_1 + X_2}{n}, \ldots, \frac{X_1 + \cdots + X_n}{n}
\]

Asymptotic properties of an estimator are defined by the behavior of \( \hat{\theta}_1, \ldots, \hat{\theta}_n \) when \( n \) goes to infinity.
Unbiasedness and efficiency are finite-sample properties of estimators, which hold regardless of sample size.
Unbiasedness and efficiency are finite-sample properties of estimators, which hold regardless of sample size.

Estimators also have asymptotic properties, i.e., the characteristics of sampling distributions when sample size becomes infinitely large.
3-4: Asymptotic Evaluations: A Brief Review

(what happens as sample size increases)

- Unbiasedness and efficiency are **finite-sample properties** of estimators, which hold regardless of sample size.
- Estimators also have **asymptotic properties**, i.e., the characteristics of sampling distributions when sample size becomes infinitely large.
- To define asymptotic properties, consider a **sequence** of estimators at increasing sample sizes:

\[ \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n \]
3-4: Asymptotic Evaluations: A Brief Review

(what happens as sample size increases)

- Unbiasedness and efficiency are **finite-sample properties** of estimators, which hold regardless of sample size.
- Estimators also have **asymptotic properties**, i.e., the characteristics of sampling distributions when sample size becomes infinitely large.
- To define asymptotic properties, consider a **sequence** of estimators at increasing sample sizes:

$$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$$

- For example, the sequence of sample means ($\bar{X}_n$) is defined as:

$$\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n = X_1, \frac{X_1 + X_2}{2}, \ldots, \frac{X_1 + \cdots X_n}{n}$$
3-4: Asymptotic Evaluations: A Brief Review
(what happens as sample size increases)

- Unbiasedness and efficiency are **finite-sample properties** of estimators, which hold regardless of sample size.
- Estimators also have **asymptotic properties**, i.e., the characteristics of sampling distributions when sample size becomes infinitely large.
- To define asymptotic properties, consider a **sequence** of estimators at increasing sample sizes:
  \[ \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n \]

- For example, the sequence of sample means \( (\bar{X}_n) \) is defined as:
  \[
  \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n = X_1, \frac{X_1 + X_2}{2}, \ldots, \frac{X_1 + \cdots X_n}{n}
  \]

- Asymptotic properties of an estimator are defined by the behavior of \( \hat{\theta}_1, \ldots, \hat{\theta}_n \) when \( n \) goes to infinity.
3: Consistency

(does it get closer to the right answer as sample size increases)

**Definition**

An estimator \( \hat{\theta}_n \) is **consistent** if the sequence \( \hat{\theta}_1, ..., \hat{\theta}_n \) converges in probability to the true parameter value \( \theta \) as sample size \( n \) grows to infinity:

\[
\hat{\theta}_n \xrightarrow{p} \theta \quad \text{or} \quad \lim_{n \to \infty} \hat{\theta}_n = \theta
\]
3: Consistency

(does it get closer to the right answer as sample size increases)

**Definition**

An estimator $\hat{\theta}_n$ is **consistent** if the sequence $\hat{\theta}_1, \ldots, \hat{\theta}_n$ converges in probability to the true parameter value $\theta$ as sample size $n$ grows to infinity:

$$\hat{\theta}_n \xrightarrow{p} \theta \quad \text{or} \quad \text{plim}_{n \to \infty} \hat{\theta}_n = \theta$$

- Often seen as a minimal requirement for estimators
3: Consistency

(does it get closer to the right answer as sample size increases)

**Definition**

An estimator $\hat{\theta}_n$ is **consistent** if the sequence $\hat{\theta}_1, ..., \hat{\theta}_n$ converges in probability to the true parameter value $\theta$ as sample size $n$ grows to infinity:

$$\hat{\theta}_n \xrightarrow{p} \theta \quad \text{or} \quad \lim_{n \to \infty} \hat{\theta}_n = \theta$$

- Often seen as a minimal requirement for estimators
- A consistent estimator may still perform badly in small samples
3: Consistency

(does it get closer to the right answer as sample size increases)

Definition

An estimator $\hat{\theta}_n$ is consistent if the sequence $\hat{\theta}_1, \ldots, \hat{\theta}_n$ converges in probability to the true parameter value $\theta$ as sample size $n$ grows to infinity:

$$\hat{\theta}_n \xrightarrow{p} \theta \quad \text{or} \quad \lim_{n \to \infty} \hat{\theta}_n = \theta$$

- Often seen as a minimal requirement for estimators
- A consistent estimator may still perform badly in small samples
- Two ways to verify consistency:
  1. Analytic: Often easier to check if $E[\hat{\theta}_n] \to \theta$ and $V[\hat{\theta}_n] \to 0$
  2. Simulation: Increase $n$ and see how the sampling distribution changes
3: Consistency
(does it get closer to the right answer as sample size increases)

Definition

An estimator \( \hat{\theta}_n \) is consistent if the sequence \( \hat{\theta}_1, \ldots, \hat{\theta}_n \) converges in probability to the true parameter value \( \theta \) as sample size \( n \) grows to infinity:

\[
\hat{\theta}_n \xrightarrow{p} \theta \quad \text{or} \quad \lim_{n \to \infty} \hat{\theta}_n = \theta
\]

- Often seen as a minimal requirement for estimators
- A consistent estimator may still perform badly in small samples

Two ways to verify consistency:

1. **Analytic**: Often easier to check if \( E[\hat{\theta}_n] \to \theta \) and \( V[\hat{\theta}_n] \to 0 \)
2. **Simulation**: Increase \( n \) and see how the sampling distribution changes

- Does unbiasedness imply consistency?
- Does consistency imply unbiasedness?
Deriving Consistency of Estimators

Our candidate estimators:

1. \( \hat{\mu}_1 = Y_1 \)
2. \( \hat{\mu}_2 = 4 \)
3. \( \hat{\mu}_3 = \bar{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n) \)
4. \( \hat{\mu}_4 = \tilde{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n) \)

Which of these estimators are consistent for \( \mu \)?
Deriving Consistency of Estimators

Our candidate estimators:

1. $\hat{\mu}_1 = Y_1$
2. $\hat{\mu}_2 = 4$
3. $\hat{\mu}_3 = \bar{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n)$
4. $\hat{\mu}_4 = \tilde{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n)$

Which of these estimators are consistent for $\mu$?

1. $E[\hat{\mu}_1] = \mu$ and $V[\hat{\mu}_1] = \sigma^2$
Deriving Consistency of Estimators

Our candidate estimators:

1. $\hat{\mu}_1 = Y_1$
2. $\hat{\mu}_2 = 4$
3. $\hat{\mu}_3 = \bar{Y}_n \equiv \frac{1}{n} (Y_1 + \cdots + Y_n)$
4. $\hat{\mu}_4 = \tilde{Y}_n \equiv \frac{1}{n+5} (Y_1 + \cdots + Y_n)$

Which of these estimators are consistent for $\mu$?

1. $E[\hat{\mu}_1] = \mu$ and $V[\hat{\mu}_1] = \sigma^2$
2. $E[\hat{\mu}_2] = 4$ and $V[\hat{\mu}_2] = 0$
Deriving Consistency of Estimators

Our candidate estimators:

1. \( \hat{\mu}_1 = Y_1 \)
2. \( \hat{\mu}_2 = 4 \)
3. \( \hat{\mu}_3 = \overline{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n) \)
4. \( \hat{\mu}_4 = \tilde{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n) \)

Which of these estimators are consistent for \( \mu \)?

1. \( E[\hat{\mu}_1] = \mu \) and \( V[\hat{\mu}_1] = \sigma^2 \)
2. \( E[\hat{\mu}_2] = 4 \) and \( V[\hat{\mu}_2] = 0 \)
3. \( E[\hat{\mu}_3] = \mu \) and \( V[\hat{\mu}_3] = \frac{1}{n} \sigma^2 \)
Deriving Consistency of Estimators

Our candidate estimators:

1. \( \hat{\mu}_1 = Y_1 \)
2. \( \hat{\mu}_2 = 4 \)
3. \( \hat{\mu}_3 = \bar{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n) \)
4. \( \hat{\mu}_4 = \bar{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n) \)

Which of these estimators are consistent for \( \mu \)?

1. \( E[\hat{\mu}_1] = \mu \) and \( V[\hat{\mu}_1] = \sigma^2 \)
2. \( E[\hat{\mu}_2] = 4 \) and \( V[\hat{\mu}_2] = 0 \)
3. \( E[\hat{\mu}_3] = \mu \) and \( V[\hat{\mu}_3] = \frac{1}{n} \sigma^2 \)
4. \( E[\hat{\mu}_4] = \frac{n}{n+5} \mu \) and \( V[\hat{\mu}_4] = \frac{n}{(n+5)^2} \sigma^2 \)
Consistency

The sample mean is a consistent estimator for \( \mu \).

\[
\bar{X}_n \sim_{approx} N \left( \mu, \frac{\sigma^2}{n} \right)
\]
The sample mean is a consistent estimator for $\mu$.

$$X_n \sim \text{approx } N\left(\mu, \frac{\sigma^2}{n}\right)$$

As $n$ increases, $\frac{\sigma^2}{n}$ approaches 0.

$n =$
Consistency

The sample mean is a consistent estimator for $\mu$.

$$\bar{X}_n \sim_{\text{approx}} N \left( \mu, \frac{\sigma^2}{n} \right)$$

As $n$ increases, $\frac{\sigma^2}{n}$ approaches 0.

$$n = 1$$
Consistency

The sample mean is a consistent estimator for $\mu$.

$$\overline{X}_n \sim \text{approx } N\left(\mu, \frac{\sigma^2}{n}\right)$$

As $n$ increases, $\frac{\sigma^2}{n}$ approaches 0.

$n = 25$
The sample mean is a consistent estimator for $\mu$.

$$\overline{X}_n \sim_{approx} N \left( \mu, \frac{\sigma^2}{n} \right)$$

As $n$ increases, $\frac{\sigma^2}{n}$ approaches 0.

$$n = 100$$
Inconsistency

An estimator can be inconsistent in several ways:

- The sampling distribution collapses around the wrong value
- The sampling distribution never collapses around anything
An estimator can be inconsistent in several ways:
Inconsistency

An estimator can be inconsistent in several ways:

- The sampling distribution collapses around the wrong value
Inconsistency

An estimator can be inconsistent in several ways:
- The sampling distribution collapses around the wrong value
- The sampling distribution never collapses around anything
Inconsistency

Consider the median estimator: 
\[ \tilde{X}_n = \text{median}(Y_1, \ldots, Y_n) \]
Inconsistency

Consider the median estimator: \( \tilde{X}_n = \text{median}(Y_1, \ldots, Y_n) \) Is this estimator consistent for the expectation? 

\[ n = \]
Consider the median estimator: $\tilde{X}_n = \text{median}(Y_1, \ldots, Y_n)$. Is this estimator consistent for the expectation? $n = 1$
Inconsistency

Consider the median estimator: \( \tilde{X}_n = \text{median}(Y_1, \ldots, Y_n) \) Is this estimator consistent for the expectation? 

\( n = 25 \)
Consider the median estimator: $\tilde{X}_n = \text{median}(Y_1, \ldots, Y_n)$ Is this estimator consistent for the expectation?

$n = 100$
4: Asymptotic Distribution
(known sampling distribution for large sample size)

We are also interested in the shape of the sampling distribution of an estimator as the sample size increases. Due to the central limit theorem, the sampling distributions of many estimators converge towards a normal distribution such that,

$$\hat{\theta}_n - \theta \sqrt{V[\hat{\theta}_n]} \xrightarrow{d} N(0,1)$$

This will play a crucial role in our ability to form confidence intervals.
4: Asymptotic Distribution
(known sampling distribution for large sample size)

We are also interested in the shape of the sampling distribution of an estimator as the sample size increases.

\[ \hat{\theta}_n - \theta \sqrt{V[\hat{\theta}_n]} \overset{d}{\rightarrow} N(0, 1) \]

This will play a crucial role in our ability to form confidence intervals.
4: Asymptotic Distribution
(known sampling distribution for large sample size)

We are also interested in the shape of the sampling distribution of an estimator as the sample size increases.

Due to the central limit theorem, the sampling distributions of many estimators converge towards a normal distribution such that,

\[
\frac{\hat{\theta}_n - \theta}{\sqrt{V[\hat{\theta}_n]}} \xrightarrow{d} \mathcal{N}(0, 1)
\]
4: Asymptotic Distribution
(known sampling distribution for large sample size)

We are also interested in the shape of the sampling distribution of an estimator as the sample size increases.

Due to the central limit theorem, the sampling distributions of many estimators converge towards a normal distribution such that,

\[
\frac{\hat{\theta}_n - \theta}{\sqrt{V[\hat{\theta}_n]}} \xrightarrow{d} \mathcal{N}(0, 1)
\]

This will play a crucial role in our ability to form confidence intervals.
## Summary of Properties

<table>
<thead>
<tr>
<th>Concept</th>
<th>Criteria</th>
<th>Intuition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unbiasedness</td>
<td>$E[\hat{\mu}] = \mu$</td>
<td>Right on average</td>
</tr>
<tr>
<td>Efficiency</td>
<td>$\text{Var}[\hat{\mu}_1] &lt; \text{Var}[\hat{\mu}_2]$</td>
<td>Low variance</td>
</tr>
<tr>
<td>Consistency</td>
<td>$\hat{\mu}_n \xrightarrow{p} \mu$</td>
<td>Converge to estimand as $n \to \infty$</td>
</tr>
<tr>
<td>Asymptotic Normality</td>
<td>$\hat{\mu}_n \overset{\text{approx.}}{\sim} N(\mu, \frac{\sigma^2}{n})$</td>
<td>Approximately normal in large $n$</td>
</tr>
</tbody>
</table>
How do we learn about an estimator?

Repeating theme in this class—how to characterize an estimator.
How do we learn about an estimator?

Repeating theme in this class—how to characterize an estimator.

1. Define estimand of interest (causal quantity, survey outcome, model parameter)
How do we learn about an estimator?

Repeating theme in this class—how to characterize an estimator.

1. Define estimand of interest (causal quantity, survey outcome, model parameter)
2. Find an estimator for the quantity of interest
How do we learn about an estimator?

Repeating theme in this class—how to characterize an estimator.

1. Define estimand of interest (causal quantity, survey outcome, model parameter)
2. Find an estimator for the quantity of interest
3. Is this estimator consistent?
4. Is this estimator unbiased?
5. What is the variance of the estimator?
6. Can we find an unbiased estimator for the variance (and is it consistent)?
7. What are the finite sample properties of the estimator?
8. What are the asymptotic properties of the estimator?
How do we learn about an estimator?

Repeating theme in this class—how to characterize an estimator.

1. Define estimand of interest (causal quantity, survey outcome, model parameter)
2. Find an estimator for the quantity of interest
3. Is this estimator consistent?
4. Is this estimator unbiased?
How do we learn about an estimator?

Repeating theme in this class—how to characterize an estimator.

1. Define estimand of interest (causal quantity, survey outcome, model parameter)
2. Find an estimator for the quantity of interest
3. Is this estimator consistent?
4. Is this estimator unbiased?
5. What is the variance of the estimator?
How do we learn about an estimator?

Repeating theme in this class—how to characterize an estimator.

1. Define estimand of interest (causal quantity, survey outcome, model parameter)
2. Find an estimator for the quantity of interest
3. Is this estimator consistent?
4. Is this estimator unbiased?
5. What is the variance of the estimator?
6. Can we find an unbiased estimator for the variance (and is it consistent)?
How do we learn about an estimator?

Repeating theme in this class—how to characterize an estimator.

1. Define estimand of interest (causal quantity, survey outcome, model parameter)
2. Find an estimator for the quantity of interest
3. Is this estimator consistent?
4. Is this estimator unbiased?
5. What is the variance of the estimator?
6. Can we find an unbiased estimator for the variance (and is it consistent)?
7. What are the finite sample properties of the estimator?
How do we learn about an estimator?

Repeating theme in this class—how to characterize an estimator.

1. Define estimand of interest (causal quantity, survey outcome, model parameter)
2. Find an estimator for the quantity of interest
3. Is this estimator consistent?
4. Is this estimator unbiased?
5. What is the variance of the estimator?
6. Can we find an unbiased estimator for the variance (and is it consistent)?
7. What are the finite sample properties of the estimator?
8. What are the asymptotic properties of the estimator?
Back to the Example

Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment

ALAN S. GERBER  Yale University
DONALD P. GREEN  Yale University
CHRISTOPHER W. LARIMER  University of Northern Iowa
Population vs. Sampling Distribution

We want to think about the sampling distribution of the estimator.

But remember that we only get to see one draw from the sampling distribution. Thus ideally we want an estimator with good properties.
Population vs. Sampling Distribution

We want to think about the sampling distribution of the estimator.
Population vs. Sampling Distribution

We want to think about the sampling distribution of the estimator.

But remember that we only get to see one draw from the sampling distribution. Thus ideally we want an estimator with good properties.
Population vs. Sampling Distribution

We want to think about the sampling distribution of the estimator.

But remember that we only get to see one draw from the sampling distribution. Thus ideally we want an estimator with good properties.
Asymptotic Normality

Going back to the Gerber, Green, and Larimer result. The estimator is difference in means.

The estimate is 0.063.

Suppose we have an estimate of the estimator's standard error: \( \hat{SE}(\hat{\theta}) = 0.02. \)

What if there was no difference in means in the population (\( \mu_y - \mu_x = 0 \))? By asymptotic Normality, \( (\hat{\theta} - 0) / \hat{SE}(\hat{\theta}) \) \( \sim N(0, 1) \).

By the properties of Normals, we know that this implies that \( \hat{\theta} \sim N(0, \hat{SE}(\hat{\theta})) \).
Asymptotic Normality

Going back to the Gerber, Green, and Larimer result...
Asymptotic Normality

Going back to the Gerber, Green, and Larimer result...

- The estimator is difference in means
Asymptotic Normality

Going back to the Gerber, Green, and Larimer result...

- The estimator is **difference in means**
- The estimate is 0.063
Asymptotic Normality

Going back to the Gerber, Green, and Larimer result...

- The estimator is **difference in means**
- The estimate is 0.063
- Suppose we have an estimate of the estimator’s **standard error** $\hat{SE}(\hat{\theta}) = 0.02$. 

Asymptotic Normality

Going back to the Gerber, Green, and Larimer result... 

- The estimator is **difference in means**
- The estimate is 0.063
- Suppose we have an estimate of the estimator’s **standard error**
  \( \hat{SE}(\hat{\theta}) = 0.02 \).
- What if there was **no** difference in means in the population
  \( (\mu_y - \mu_x = 0) \)?
Asymptotic Normality

Going back to the Gerber, Green, and Larimer result . . .

- The estimator is difference in means
- The estimate is 0.063
- Suppose we have an estimate of the estimator’s standard error $\hat{SE}(\hat{\theta}) = 0.02$.
- What if there was no difference in means in the population $(\mu_y - \mu_x = 0)$?
- By asymptotic Normality $(\hat{\theta} - 0)/SE(\hat{\theta}) \sim N(0, 1)$
Asymptotic Normality

Going back to the Gerber, Green, and Larimer result...

- The estimator is difference in means
- The estimate is 0.063
- Suppose we have an estimate of the estimator’s standard error \( \hat{SE}(\hat{\theta}) = 0.02 \).
- What if there was no difference in means in the population \( \mu_y - \mu_x = 0 \)?
- By asymptotic Normality \( (\hat{\theta} - 0)/SE(\hat{\theta}) \sim N(0, 1) \)
- By the properties of Normals, we know that this implies that \( \hat{\theta} \sim N(0, SE(\hat{\theta})) \)
Asymptotic Normality

We can plot this to get a feel for it.

Does the observed difference in means seem plausible if there really were no difference between the two groups in the population?
Asymptotic Normality

We can plot this to get a feel for it.
Asymptotic Normality

We can plot this to get a feel for it.

Does the observed difference in means seem plausible if there really were no difference between the two groups in the population?
Asymptotic Normality

We can plot this to get a feel for it.

Does the observed difference in means seem plausible if there really were no difference between the two groups in the population?
The scariest pieces of mail ever! continued

Summarizes the relationships between political science research and campaigns. Also, attempts to weaponize the results of Gerber et al (2008).
We Covered...
We Covered...

- Four properties of estimators: bias, efficiency, consistency and asymptotic normality.
We Covered...

- Four properties of estimators: bias, efficiency, consistency and asymptotic normality.
- A brief example of how we can use asymptotic normality in an example that will return!
We Covered. . .

- Four properties of estimators: bias, efficiency, consistency and asymptotic normality.
- A brief example of how we can use asymptotic normality in an example that will return!

Next Time:
We Covered...

- Four properties of estimators: bias, efficiency, consistency and asymptotic normality.
- A brief example of how we can use asymptotic normality in an example that will return!

Next Time: interval estimation
Where We’ve Been and Where We’re Going...

- Last Week
Where We’ve Been and Where We’re Going...

- Last Week
  - random variables
Where We’ve Been and Where We’re Going…

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)

- **Next Week**
  - hypothesis testing
  - what is regression?

- **Long Run**
  - probability → inference → regression → causal inference
Where We’ve Been and Where We’re Going…

- Last Week
  - random variables
  - joint distributions
- This Week
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals
- Next Week
  - hypothesis testing
  - what is regression?
- Long Run
  - probability → inference → regression → causal inference

Stewart (Princeton)
Where We’ve Been and Where We’re Going...

- Last Week
  - random variables
  - joint distributions
- This Week
  - estimators and sampling distributions

Long Run
- probability → inference → regression → causal inference
Where We’ve Been and Where We’re Going...

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
Where We’ve Been and Where We’re Going...

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals
Where We’ve Been and Where We’re Going...

- Last Week
  - random variables
  - joint distributions
- This Week
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals
- Next Week
Where We’ve Been and Where We’re Going...

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
  - hypothesis testing
Where We’ve Been and Where We’re Going…

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
  - hypothesis testing
  - what is regression?
Where We’ve Been and Where We’re Going...

- Last Week
  - random variables
  - joint distributions

- This Week
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- Next Week
  - hypothesis testing
  - what is regression?

- Long Run
Where We’ve Been and Where We’re Going...

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
  - hypothesis testing
  - what is regression?

- **Long Run**
  - probability → inference → regression → causal inference
1. Estimation
   - Populations and Samples
   - Estimators
   - Analytical

2. Weak Law of Large Numbers and the Central Limit Theorem
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3. Properties of Estimators
   - Four Desirable Properties
   - Example

4. Interval Estimation
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5. Plug-In Principle
1. Estimation
   - Populations and Samples
   - Estimators
   - Analytical

2. Weak Law of Large Numbers and the Central Limit Theorem
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3. Properties of Estimators
   - Four Desirable Properties
   - Example

4. Interval Estimation
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5. Plug-In Principle
What is Interval Estimation?

- A point estimator \( \hat{\theta} \) estimates a scalar population parameter \( \theta \) with a single number.
What is Interval Estimation?

- A point estimator $\hat{\theta}$ estimates a scalar population parameter $\theta$ with a single number.
- However, because we are dealing with a random sample, we might also want to report uncertainty in our estimate.
What is Interval Estimation?

- A point estimator $\hat{\theta}$ estimates a scalar population parameter $\theta$ with a single number.
- However, because we are dealing with a random sample, we might also want to report uncertainty in our estimate.
- An interval estimator for $\theta$ takes the following form:

$$[\hat{\theta}_{\text{lower}}, \hat{\theta}_{\text{upper}}]$$

where $\hat{\theta}_{\text{lower}}$ and $\hat{\theta}_{\text{upper}}$ are random quantities that vary from sample to sample.
What is Interval Estimation?

- A point estimator $\hat{\theta}$ estimates a scalar population parameter $\theta$ with a single number.
- However, because we are dealing with a random sample, we might also want to report uncertainty in our estimate.
- An interval estimator for $\theta$ takes the following form:
  
  $[\hat{\theta}_{lower}, \hat{\theta}_{upper}]$

  where $\hat{\theta}_{lower}$ and $\hat{\theta}_{upper}$ are random quantities that vary from sample to sample.

- The interval represents the range of possible values within which we estimate the true value of $\theta$ to fall.
What is Interval Estimation?

- A point estimator $\hat{\theta}$ estimates a scalar population parameter $\theta$ with a single number.
- However, because we are dealing with a random sample, we might also want to report uncertainty in our estimate.
- An interval estimator for $\theta$ takes the following form:
  
  $$[\hat{\theta}_{\text{lower}}, \hat{\theta}_{\text{upper}}]$$

  where $\hat{\theta}_{\text{lower}}$ and $\hat{\theta}_{\text{upper}}$ are random quantities that vary from sample to sample.
- The interval represents the range of possible values within which we estimate the true value of $\theta$ to fall.
- An interval estimate is a realized value from an interval estimator. The estimated interval typically forms what we call a confidence interval, which we will define shortly.
Population with Known $\sigma^2$

Suppose we have an i.i.d. random sample of size $n$, $X_1, \ldots, X_n$, from with $E[X] = \mu$, $V[X] = \sigma^2$. From previous lecture, we know that the sampling distribution of the sample average in large samples is:

$$X_n \sim N(\mu, \sigma^2/n) = N(\mu, 1/n)$$

Therefore, the standardized sample average is distributed as follows:

$$X_n - \mu \sim N(0, 1)$$

This implies

$$P\left(-1.96 < \frac{X_n - \mu}{\sqrt{n}} < 1.96\right) = 0.95$$
Population with Known $\sigma^2$

Suppose we have an i.i.d. random sample of size $n$, $X_1, \ldots, X_n$, from with $E[X] = \mu$, $V[X] = 1$. 

From previous lecture, we know that the sampling distribution of the sample average in large samples is:

$X_n \sim \mathcal{N}(\mu, \sigma^2/n) = \mathcal{N}(\mu, 1/n)$

Therefore, the standardized sample average is distributed as follows:

$X_n - \mu \sim \mathcal{N}(0, 1)$

This implies

$P(-1.96 < X_n - \mu / \sqrt{n} < 1.96) = 0.95$
Population with Known $\sigma^2$

Suppose we have an i.i.d. random sample of size $n$, $X_1, \ldots, X_n$, from with $E[X] = \mu$, $V[X] = 1$.

From previous lecture, we know that the sampling distribution of the sample average in large samples is:

$$\overline{X}_n \sim N(\mu, \frac{1}{n})$$

This implies

$$P\left(-1.96 < \frac{\overline{X}_n - \mu}{\sqrt{n}} < 1.96\right) = 0.95$$
Population with Known $\sigma^2$

Suppose we have an i.i.d. random sample of size $n$, $X_1, \ldots, X_n$, from with $E[X] = \mu$, $V[X] = 1$.

From previous lecture, we know that the sampling distribution of the sample average in large samples is:

$$X_n \sim N(\mu, \sigma^2/n) = N(\mu, 1/n)$$
Population with Known $\sigma^2$

Suppose we have an i.i.d. random sample of size $n$, $X_1, \ldots, X_n$, from with $E[X] = \mu$, $V[X] = 1$.

From previous lecture, we know that the sampling distribution of the sample average in large samples is:

$$\overline{X}_n \sim N(\mu, \sigma^2/n) = N(\mu, 1/n)$$

Therefore, the standardized sample average is distributed as follows:

$$\frac{\overline{X}_n - \mu}{1/\sqrt{n}} \sim \text{standard normal}$$
Population with Known $\sigma^2$

Suppose we have an i.i.d. random sample of size $n$, $X_1, \ldots, X_n$, from with $E[X] = \mu$, $V[X] = 1$.

From previous lecture, we know that the sampling distribution of the sample average in large samples is:

$$\overline{X}_n \sim N(\mu, \sigma^2 / n) = N(\mu, 1/n)$$

Therefore, the standardized sample average is distributed as follows:

$$\frac{\overline{X}_n - \mu}{1/\sqrt{n}} \sim N(0, 1)$$
Population with Known $\sigma^2$

Suppose we have an i.i.d. random sample of size $n$, $X_1, \ldots, X_n$, from with $E[X] = \mu$, $V[X] = 1$.

From previous lecture, we know that the sampling distribution of the sample average in large samples is:

$$X_n \sim N(\mu, \sigma^2/n) = N(\mu, 1/n)$$

Therefore, the standardized sample average is distributed as follows:

$$\frac{X_n - \mu}{1/\sqrt{n}} \sim N(0, 1)$$

This implies

$$P \left( -1.96 < \frac{X_n - \mu}{1/\sqrt{n}} < 1.96 \right) = .95$$
CDF of the Standard Normal Distribution

CDF(1.96) = 0.975
CDF(−1.96) = 0.025
CDF of the Standard Normal Distribution

CDF(1.96) = .975
CDF(-1.96) = .025
Constructing a Confidence Interval with Known $\sigma^2$

So we know that:

$$P\left(-1.96 < \frac{\bar{X}_n - \mu}{1/\sqrt{n}} < 1.96\right) = .95$$
Constructing a Confidence Interval with Known $\sigma^2$

So we know that:

$$P\left(-1.96 < \frac{\overline{X}_n - \mu}{1/\sqrt{n}} < 1.96\right) = .95$$

Rearranging yields:

$$P\left(\overline{X}_n - 1.96/\sqrt{n} < \mu < \overline{X}_n + 1.96/\sqrt{n}\right) = .95$$
Constructing a Confidence Interval with Known $\sigma^2$

So we know that:

$$P \left( -1.96 < \frac{\bar{X}_n - \mu}{1/\sqrt{n}} < 1.96 \right) = .95$$

Rearranging yields:

$$P \left( \bar{X}_n - 1.96/\sqrt{n} < \mu < \bar{X}_n + 1.96/\sqrt{n} \right) = .95$$

This implies that the following interval estimator

$$[ \bar{X}_n - 1.96/\sqrt{n}, \bar{X}_n + 1.96/\sqrt{n} ]$$

contains the true population mean $\mu$ with probability 0.95.
Constructing a Confidence Interval with Known $\sigma^2$

So we know that:

$$ P \left( -1.96 < \frac{\bar{X}_n - \mu}{1/\sqrt{n}} < 1.96 \right) = .95 $$

Rearranging yields:

$$ P \left( \frac{\bar{X}_n - 1.96}{\sqrt{n}} < \mu < \frac{\bar{X}_n + 1.96}{\sqrt{n}} \right) = .95 $$

This implies that the following interval estimator

$$ \left[ \frac{\bar{X}_n - 1.96}{\sqrt{n}}, \frac{\bar{X}_n + 1.96}{\sqrt{n}} \right] $$

contains the true population mean $\mu$ with probability 0.95.

We call this estimator a 95% confidence interval for $\mu$. 
Kuklinski Example

\[ \bar{Y} \sim_{approx} (? , ?) \]
Kuklinski Example

\[ \bar{Y} \sim_{approx} ?(\mu, ?) \]
Kuklinski Example

\[
\bar{Y} \sim_{\text{approx}} ?(\mu, \sigma^2 / n)
\]
Kuklinski Example

\[ \bar{Y} \sim_{\text{approx}} N(\mu, \sigma^2/n) \]
Kuklinski Example

\[
\bar{Y} \sim_{approx} N(\mu, \sigma^2 / n)
\]

Suppose the 1,161 respondents in the Kuklinski data set were the population, with \( \mu = 42.7 \) and \( \sigma^2 = 257.9 \).
Kuklinski Example

\[ \overline{Y} \sim_{\text{approx}} N(\mu, \sigma^2 / n) \]

Suppose the 1,161 respondents in the Kuklinski data set were the population, with \( \mu = 42.7 \) and \( \sigma^2 = 257.9 \).

If we sampled 100 respondents, the sampling distribution of \( \overline{Y}_{100} \) is:

\[ \overline{Y}_{100} \sim_{\text{approx}} ?(?, ?) \]
Kuklinski Example

\[ \bar{Y} \sim_{approx} N(\mu, \sigma^2 / n) \]

Suppose the 1,161 respondents in the Kuklinski data set were the population, with \( \mu = 42.7 \) and \( \sigma^2 = 257.9 \).

If we sampled 100 respondents, the sampling distribution of \( \bar{Y}_{100} \) is:

\[ \bar{Y}_{100} \sim_{approx} ?(42.7, ?) \]
Kuklinski Example

$$\bar{Y} \sim_{approx} N(\mu, \sigma^2/n)$$

Suppose the 1,161 respondents in the Kuklinski data set were the population, with $$\mu = 42.7$$ and $$\sigma^2 = 257.9$$.

If we sampled 100 respondents, the sampling distribution of $$\bar{Y}_{100}$$ is:

$$\bar{Y}_{100} \sim_{approx} ?(42.7, 257.9/100)$$
Kuklinski Example

\[ \bar{Y} \sim_{approx} N(\mu, \sigma^2 / n) \]

Suppose the 1,161 respondents in the Kuklinski data set were the population, with \( \mu = 42.7 \) and \( \sigma^2 = 257.9 \).

If we sampled 100 respondents, the sampling distribution of \( \bar{Y}_{100} \) is:

\[ \bar{Y}_{100} \sim_{approx} N(42.7, 2.579) \]
The standard error of \( \bar{Y} \)

The standard error of the sample mean is the standard deviation of the sampling distribution for \( \bar{Y} \):

\[
SE(\bar{Y}) = \sqrt{\frac{\sigma^2}{n}}
\]
The standard error of \( \bar{Y} \)

The **standard error** of the sample mean is the standard deviation of the sampling distribution for \( \bar{Y} \):

\[
SE(\bar{Y}) = \sqrt{V(\bar{Y})} = \frac{\sigma}{\sqrt{n}}
\]
The standard error of \( \bar{Y} \)

The standard error of the sample mean is the standard deviation of the sampling distribution for \( \bar{Y} \):

\[
SE(\bar{Y}) = \sqrt{V(\bar{Y})} = \frac{\sigma}{\sqrt{n}}
\]

What is the probability that \( \bar{Y} \) falls within 1.96 SEs of \( \mu \)?
The standard error of $\bar{Y}$

The standard error of the sample mean is the standard deviation of the sampling distribution for $\bar{Y}$:

$$SE(\bar{Y}) = \sqrt{V(\bar{Y})} = \frac{\sigma}{\sqrt{n}}$$

What is the probability that $\bar{Y}$ falls within 1.96 SEs of $\mu$?
The standard error of $\overline{Y}$

The standard error of the sample mean is the standard deviation of the sampling distribution for $\overline{Y}$:

$$SE(\overline{Y}) = \sqrt{V(\overline{Y})} = \frac{\sigma}{\sqrt{n}}$$

What is the probability that $\overline{Y}$ falls within 1.96 SEs of $\mu$?
Normal Population with Unknown $\sigma^2$

In practice, it is rarely the case that we somehow know the true value of $\sigma^2$ and our previous example relied on that knowledge.
Normal Population with Unknown $\sigma^2$

In practice, it is rarely the case that we somehow know the true value of $\sigma^2$ and our previous example relied on that knowledge.

Suppose now that we have an i.i.d. random sample of size $n \ X_1, \ldots, X_n$ where $\sigma^2$ is unknown.
Normal Population with Unknown $\sigma^2$

In practice, it is rarely the case that we somehow know the true value of $\sigma^2$ and our previous example relied on that knowledge.

Suppose now that we have an i.i.d. random sample of size $n$ $X_1, \ldots, X_n$ where $\sigma^2$ is unknown. Then, as before,

$$\bar{X}_n \sim N(\mu, \sigma^2/n)$$

and so

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$
Normal Population with Unknown $\sigma^2$

In practice, it is rarely the case that we somehow know the true value of $\sigma^2$ and our previous example relied on that knowledge.

Suppose now that we have an i.i.d. random sample of size $n$ $X_1, \ldots, X_n$ where $\sigma^2$ is unknown. Then, as before,

$$\bar{X}_n \sim N(\mu, \sigma^2/n)$$

and so

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

Previously, we then constructed the interval:

$$[\bar{X}_n - z_{\alpha/2} \sigma/\sqrt{n}, \bar{X}_n + z_{\alpha/2} \sigma/\sqrt{n}]$$
Normal Population with Unknown $\sigma^2$

In practice, it is rarely the case that we somehow know the true value of $\sigma^2$ and our previous example relied on that knowledge.

Suppose now that we have an i.i.d. random sample of size $n$ $X_1, \ldots, X_n$ where $\sigma^2$ is unknown. Then, as before,

$$\overline{X}_n \sim N(\mu, \sigma^2 / n)$$

and so

$$\frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \sim N(0, 1).$$

Previously, we then constructed the interval:

$$[\overline{X}_n - z_{\alpha/2} \sigma / \sqrt{n}, \overline{X}_n + z_{\alpha/2} \sigma / \sqrt{n}]$$

But we can not directly use this now because
Normal Population with Unknown $\sigma^2$

In practice, it is rarely the case that we somehow know the true value of $\sigma^2$ and our previous example relied on that knowledge.

Suppose now that we have an i.i.d. random sample of size $n X_1, ..., X_n$ where $\sigma^2$ is unknown. Then, as before,

$$\overline{X}_n \sim N(\mu, \sigma^2 / n) \quad \text{and so} \quad \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \sim N(0, 1).$$

Previously, we then constructed the interval:

$$[\overline{X}_n - z_{\alpha/2} \sigma / \sqrt{n}, \overline{X}_n + z_{\alpha/2} \sigma / \sqrt{n}]$$

But we can not directly use this now because $\sigma^2$ is unknown.
Normal Population with Unknown $\sigma^2$

In practice, it is rarely the case that we somehow know the true value of $\sigma^2$ and our previous example relied on that knowledge.

Suppose now that we have an i.i.d. random sample of size $n X_1, \ldots, X_n$ where $\sigma^2$ is unknown. Then, as before,

$$\overline{X}_n \sim N(\mu, \sigma^2/n)$$ and so

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Previously, we then constructed the interval:

$$[\overline{X}_n - z_{\alpha/2}\sigma/\sqrt{n}, \overline{X}_n + z_{\alpha/2}\sigma/\sqrt{n}]$$

But we can not directly use this now because $\sigma^2$ is unknown.

Instead, we need an estimator of $\sigma^2$, $\hat{\sigma}^2$. 
Estimators for the Population Variance

Two possible estimators of population variance:

\[ S^2_0 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

\[ S^2_1 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

Which do we prefer?

Let's check properties of these estimators.

1. **Unbiasedness:** We can show (after some algebra) that

\[ E[S^2_0] = \frac{n}{n-1} \sigma^2 \]

\[ E[S^2_1] = \sigma^2 \]

2. **Consistency:** We can show that

\[ S^2_0 \xrightarrow{p} \sigma^2 \]

\[ S^2_1 \xrightarrow{p} \sigma^2 \]

\( S^2_1 \) (unbiased and consistent) is commonly called the sample variance.
Estimators for the Population Variance

Two possible estimators of population variance:

\[ S_{0n}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \]

\[ S_{1n}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \]

Which do we prefer?
Estimators for the Population Variance

Two possible estimators of population variance:

\[ S_{0n}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \]

\[ S_{1n}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \]

Which do we prefer? Let’s check properties of these estimators.
Estimators for the Population Variance

Two possible estimators of population variance:

\[ S_{0n}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \]

\[ S_{1n}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \]

Which do we prefer? Let’s check properties of these estimators.

1. Unbiasedness: We can show (after some algebra) that

\[ E[S_{0n}^2] = \frac{n-1}{n} \sigma^2 \quad \text{and} \quad E[S_{1n}^2] = \sigma^2 \]
Estimators for the Population Variance

Two possible estimators of population variance:

\[
S_{0n}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2
\]

\[
S_{1n}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2
\]

Which do we prefer? Let’s check properties of these estimators.

1. **Unbiasedness:** We can show (after some algebra) that

\[
E[S_{0n}^2] = \frac{n-1}{n} \sigma^2 \quad \text{and} \quad E[S_{1n}^2] = \sigma^2
\]

2. **Consistency:** We can show that

\[
S_{0n}^2 \xrightarrow{p} \sigma^2 \quad \text{and} \quad S_{1n}^2 \xrightarrow{p} \sigma^2
\]
Estimators for the Population Variance

Two possible estimators of population variance:

\[ S_{0n}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \]

\[ S_{1n}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \]

Which do we prefer? Let’s check properties of these estimators.

1. **Unbiasedness:** We can show (after some algebra) that

\[ E[S_{0n}^2] = \frac{n-1}{n} \sigma^2 \quad \text{and} \quad E[S_{1n}^2] = \sigma^2 \]

2. **Consistency:** We can show that

\[ S_{0n}^2 \xRightarrow{p} \sigma^2 \quad \text{and} \quad S_{1n}^2 \xRightarrow{p} \sigma^2 \]

\( S_{1n}^2 \) (unbiased and consistent) is commonly called the sample variance.
Estimating $\sigma$ and the SE

Returning to Kulinski et. al...

We will use the sample variance:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

and thus the sample standard deviation can be written as

$$S = \sqrt{S^2}$$

We will plug in $S$ for $\sigma$ and our estimated standard error will be

$$\hat{SE}[^{\hat{\mu}}] = \frac{S}{\sqrt{n}}$$
Estimating $\sigma$ and the SE

Returning to Kulinski et. al...

We will use the sample variance:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

and thus the sample standard deviation can be written as

$$S = \sqrt{S^2}$$
Estimating $\sigma$ and the SE

Returning to Kulinski et. al...

We will use the sample variance:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

and thus the sample standard deviation can be written as

$$S = \sqrt{S^2}$$

We will plug in $S$ for $\sigma$ and our estimated standard error will be

$$\hat{SE}[^{\hat{\mu}}] = \frac{S}{\sqrt{n}}$$
95% Confidence Intervals

If $X_1, \ldots, X_n$ are i.i.d. and $n$ is large, then

$$\hat{\mu} \sim N(\mu, (SE[\hat{\mu}])^2)$$
95% Confidence Intervals

If $X_1, \ldots, X_n$ are i.i.d. and $n$ is large, then

$$\hat{\mu} \sim N(\mu, (\hat{SE}[\hat{\mu}])^2)$$

$$\hat{\mu} - \mu \sim N(0, (\hat{SE}[\hat{\mu}])^2)$$

Sampling distribution of $\hat{\mu} - \mu$
95% Confidence Intervals

If $X_1, \ldots, X_n$ are i.i.d. and $n$ is large, then

\[
\hat{\mu} \sim N(\mu, (\hat{SE}[\hat{\mu}])^2) \\
\hat{\mu} - \mu \sim N(0, (\hat{SE}[\hat{\mu}])^2) \\
\hat{\mu} - \mu \over \hat{SE}[\hat{\mu}] \sim N(0, 1)
\]
95% Confidence Intervals

If $X_1, ..., X_n$ are i.i.d. and $n$ is large, then

$$\hat{\mu} \sim N(\mu, (\text{SE}[\hat{\mu}])^2)$$

$$\hat{\mu} - \mu \sim N(0, (\text{SE}[\hat{\mu}])^2)$$

$$\frac{\hat{\mu} - \mu}{\text{SE}[\hat{\mu}]} \sim N(0, 1)$$

We know that

$$P \left( -1.96 \leq \frac{\hat{\mu} - \mu}{\text{SE}[\hat{\mu}]} \leq 1.96 \right) = 95\%$$
95% Confidence Intervals

We can work backwards from this:

\[
P \left( -1.96 \leq \frac{\hat{\mu} - \mu}{SE[\hat{\mu}]} \leq 1.96 \right) = 95\% 
\]

\[
P = 95\% 
\]

\[
P = 95\% 
\]
95% Confidence Intervals

We can work backwards from this:

\[
P \left( -1.96 \leq \frac{\hat{\mu} - \mu}{\text{SE}[\hat{\mu}]} \leq 1.96 \right) = 95\%
\]

\[
P \left( -1.96 \text{SE}[\hat{\mu}] \leq \hat{\mu} - \mu \leq 1.96 \text{SE}[\hat{\mu}] \right) = 95\%
\]

= 95%
95% Confidence Intervals

We can work backwards from this:

\[
P \left( -1.96 \leq \frac{\hat{\mu} - \mu}{SE[\hat{\mu}]} \leq 1.96 \right) = 95% \\
P \left( -1.96\hat{SE}[\hat{\mu}] \leq \hat{\mu} - \mu \leq 1.96\hat{SE}[\hat{\mu}] \right) = 95% \\
P \left( \hat{\mu} - 1.96\hat{SE}[\hat{\mu}] \leq \mu \leq \hat{\mu} + 1.96\hat{SE}[\hat{\mu}] \right) = 95%
\]
95% Confidence Intervals

We can work backwards from this:

\[ P \left( -1.96 \leq \frac{\hat{\mu} - \mu}{SE[\hat{\mu}]} \leq 1.96 \right) = 95\% \]

\[ P \left( -1.96SE[\hat{\mu}] \leq \hat{\mu} - \mu \leq 1.96SE[\hat{\mu}] \right) = 95\% \]

\[ P \left( \hat{\mu} - 1.96SE[\hat{\mu}] \leq \mu \leq \hat{\mu} + 1.96SE[\hat{\mu}] \right) = 95\% \]

The random quantities in this statement are \( \hat{\mu} \) and \( SE[\hat{\mu}] \).
95% Confidence Intervals

We can work backwards from this:

\[
P \left( -1.96 \leq \frac{\hat{\mu} - \mu}{\hat{SE}[\hat{\mu}]} \leq 1.96 \right) = 95% \\
P \left( -1.96\hat{SE}[\hat{\mu}] \leq \hat{\mu} - \mu \leq 1.96\hat{SE}[\hat{\mu}] \right) = 95% \\
P \left( \hat{\mu} - 1.96\hat{SE}[\hat{\mu}] \leq \mu \leq \hat{\mu} + 1.96\hat{SE}[\hat{\mu}] \right) = 95% 
\]

The random quantities in this statement are $\hat{\mu}$ and $\hat{SE}[\hat{\mu}]$. Once the data are observed, nothing is random!
What does this mean?

We can simulate this process using the Kuklinski data:

1) Draw a sample of size 100:

2) Calculate $\hat{\mu}$ and $\hat{SE}[\hat{\mu}]$:

3) Construct the 95% CI:
What does this mean?

We can simulate this process using the Kuklinski data:

1) Draw a sample of size 100:

2) Calculate $\hat{\mu}$ and $\hat{SE}[\hat{\mu}]$:

   $\hat{\mu} = 42.32 \quad \hat{SE}[\hat{\mu}] = 1.498$

3) Construct the 95% CI:
What does this mean?

We can simulate this process using the Kuklinski data:

1) Draw a sample of size 100:

2) Calculate $\hat{\mu}$ and $SE[\hat{\mu}]$:

$$\hat{\mu} = 42.32 \quad SE[\hat{\mu}] = 1.498$$

3) Construct the 95% CI:

$$(39.4, 45.3)$$
What does this mean?

We can simulate this process using the Kuklinski data:

1) Draw a sample of size 100:

2) Calculate $\hat{\mu}$ and $\hat{SE}[\hat{\mu}]$:

3) Construct the 95% CI:
What does this mean?

We can simulate this process using the Kuklinski data:

1) Draw a sample of size 100:

2) Calculate $\hat{\mu}$ and $\hat{SE}[\hat{\mu}]$:

$$\hat{\mu} = 41.93 \quad \hat{SE}[\hat{\mu}] = 1.604$$

3) Construct the 95% CI:
What does this mean?

We can simulate this process using the Kuklinski data:

1) Draw a sample of size 100:

\[ \hat{\mu} = 41.93 \quad \text{SE}[\hat{\mu}] = 1.604 \]

2) Calculate \( \hat{\mu} \) and \( \text{SE}[\hat{\mu}] \):

3) Construct the 95% CI:

(38.8, 45.1)
What does this mean?

We can simulate this process using the Kuklinski data:

1) Draw a sample of size 100:

2) Calculate $\hat{\mu}$ and $\hat{SE}[\hat{\mu}]$:

3) Construct the 95% CI:
What does this mean?

We can simulate this process using the Kuklinski data:

1) Draw a sample of size 100:

2) Calculate $\hat{\mu}$ and $\hat{SE}[\hat{\mu}]$:

   $\hat{\mu} = 43.53 \quad \hat{SE}[\hat{\mu}] = 1.555$

3) Construct the 95% CI:
What does this mean?

We can simulate this process using the Kuklinski data:

1) Draw a sample of size 100:

\[
\begin{align*}
\hat{\mu} &= 43.53 \\
\hat{SE}[\hat{\mu}] &= 1.555
\end{align*}
\]

3) Construct the 95% CI:

\((40.5, 46.6)\)
What does this mean?

By repeating this process, we generate the sampling distribution of the 95% CIs.

Most of the CIs cover the true $\mu$; some do not. In the long run, we expect 95% of the CIs generated to contain the true value.
What does this mean?

By repeating this process, we generate the sampling distribution of the 95% CIs.

Most of the CIs cover the true $\mu$;

![Age distribution graph]
What does this mean?

By repeating this process, we generate the sampling distribution of the 95% CIs.

Most of the CIs cover the true $\mu$; some do not.
What does this mean?

By repeating this process, we generate the sampling distribution of the 95% CIs.

Most of the CIs cover the true $\mu$; some do not.

In the long run, we expect 95% of the CIs generated to contain the true value.
Interpreting a Confidence Interval

This can be tricky, so let's break it down.

Imagine we implement the interval estimator $X_n \pm 1.96/\sqrt{n}$ for a particular sample and obtain the estimate of $[2.5, 4].$

Does this mean that there is a .95 probability that the true parameter value $\mu$ lies between these two particular numbers?

No!

Confidence intervals are easy to construct, but difficult to interpret:

▶ Each confidence interval estimate from a particular sample either contains $\mu$ or not.
▶ The probability statement is a property of the procedure. If we were to repeatedly calculate the interval estimator over many random samples from the same population, 95% of the time the constructed confidence intervals would cover $\mu.$
▶ Therefore, we refer to .95 as the coverage probability.
Interpreting a Confidence Interval

This can be tricky, so let’s break it down.
Interpreting a Confidence Interval

This can be tricky, so let’s break it down.

- Imagine we implement the interval estimator $\bar{X}_n \pm 1.96/\sqrt{n}$ for a particular sample and obtain the estimate of $[2.5, 4]$.
Interpreting a Confidence Interval

This can be tricky, so let’s break it down.

- Imagine we implement the interval estimator $\bar{X}_n \pm 1.96/\sqrt{n}$ for a particular sample and obtain the estimate of [2.5, 4].
- Does this mean that there is a .95 probability that the true parameter value $\mu$ lies between these two particular numbers?
Interpreting a Confidence Interval

This can be tricky, so let’s break it down.

- Imagine we implement the interval estimator \( \bar{X}_n \pm 1.96/\sqrt{n} \) for a particular sample and obtain the estimate of \([2.5, 4]\).

- Does this mean that there is a .95 probability that the true parameter value \( \mu \) lies between these two particular numbers? **No!**

Confidence intervals are easy to construct, but difficult to interpret:

- Each confidence interval estimate from a particular sample either contains \( \mu \) or not.

- The probability statement is a property of the procedure. If we were to repeatedly calculate the interval estimator over many random samples from the same population, 95% of the time the constructed confidence intervals would cover \( \mu \).

- Therefore, we refer to .95 as the coverage probability.
Interpreting a Confidence Interval

This can be tricky, so let’s break it down.

- Imagine we implement the interval estimator $\bar{X}_n \pm 1.96 / \sqrt{n}$ for a particular sample and obtain the estimate of $[2.5, 4]$.
- Does this mean that there is a .95 probability that the true parameter value $\mu$ lies between these two particular numbers? No!
- Confidence intervals are easy to construct, but difficult to interpret:
Interpreting a Confidence Interval

This can be tricky, so let’s break it down.

- Imagine we implement the interval estimator $\bar{X}_n \pm 1.96/\sqrt{n}$ for a particular sample and obtain the estimate of [2.5, 4].
- Does this mean that there is a .95 probability that the true parameter value $\mu$ lies between these two particular numbers? No!
- Confidence intervals are easy to construct, but difficult to interpret:
  - Each confidence interval estimate from a particular sample either contains $\mu$ or not.
Interpreting a Confidence Interval

This can be tricky, so let’s break it down.

- Imagine we implement the interval estimator $\bar{X}_n \pm 1.96/\sqrt{n}$ for a particular sample and obtain the estimate of $[2.5, 4]$.

- Does this mean that there is a .95 probability that the true parameter value $\mu$ lies between these two particular numbers? **No!**

- Confidence intervals are easy to construct, but difficult to interpret:
  
  ▶ Each confidence interval estimate from a particular sample either contains $\mu$ or not.
  
  ▶ The probability statement is a property of the procedure. If we were to repeatedly calculate the interval estimator over many random samples from the same population, 95% of the time the constructed confidence intervals would cover $\mu$. 
Interpreting a Confidence Interval

This can be tricky, so let’s break it down.

- Imagine we implement the interval estimator $\overline{X}_n \pm 1.96/\sqrt{n}$ for a particular sample and obtain the estimate of [2.5, 4].
- Does this mean that there is a .95 probability that the true parameter value $\mu$ lies between these two particular numbers? **No!**
- Confidence intervals are easy to construct, but difficult to interpret:
  - Each confidence interval estimate from a particular sample either contains $\mu$ or not.
  - The probability statement is a property of the procedure. If we were to repeatedly calculate the interval estimator over many random samples from the same population, 95% of the time the constructed confidence intervals would cover $\mu$.
  - Therefore, we refer to .95 as the **coverage probability**.
What makes a good confidence interval?

1. The **coverage probability**: how likely it is that the interval covers the truth.
What makes a good confidence interval?

1. The **coverage probability**: how likely it is that the interval covers the truth.
2. The **length** of the confidence interval:
What makes a good confidence interval?

1. The **coverage probability**: how likely it is that the interval covers the truth.
2. The **length** of the confidence interval:
   - Infinite intervals \((-\infty, \infty)\) have coverage probability 1
   - Zero-length intervals, like \([\bar{Y}, \bar{Y}]\), have coverage probability 0

You want the shortest confidence interval with the desired coverage probability.
What makes a good confidence interval?

1. **The coverage probability**: how likely it is that the interval covers the truth.

2. **The length** of the confidence interval:
   - Infinite intervals \((-\infty, \infty)\) have coverage probability 1
   - For a probability, a confidence interval of \([0, 1]\) also have coverage probability 1
What makes a good confidence interval?

1. The **coverage probability**: how likely it is that the interval covers the truth.

2. The **length** of the confidence interval:
   - Infinite intervals \((-\infty, \infty)\) have coverage probability 1
   - For a probability, a confidence interval of \([0, 1]\) also have coverage probability 1
   - Zero-length intervals, like \([\bar{Y}, \bar{Y}]\), have coverage probability 0
What makes a good confidence interval?

1. The **coverage probability**: how likely it is that the interval covers the truth.

2. The **length** of the confidence interval:
   - Infinite intervals \((-\infty, \infty)\) have coverage probability 1
   - For a probability, a confidence interval of \([0, 1]\) also have coverage probability 1
   - Zero-length intervals, like \([\bar{Y}, \bar{Y}]\), have coverage probability 0

You want the shortest confidence interval with the desired coverage probability.
Is 95% all there is?

Our 95% CI had the following form:

\[
\hat{\mu} \pm 1.96 \hat{\text{SE}}[\hat{\mu}]
\]

Remember where 1.96 came from?

\[
P(-1.96 \leq \hat{\mu} - \mu \leq 1.96) = 95%
\]

What if we want a different percentage?

\[
P(-z \leq \hat{\mu} - \mu \leq z) = (1 - \alpha)\%
\]

How can we find \( z \)?
Is 95% all there is?

Our 95% CI had the following form: $$\hat{\mu} \pm 1.96 \hat{SE}[\hat{\mu}]$$

Remember where 1.96 came from?
Is 95% all there is?

Our 95% CI had the following form:  $\hat{\mu} \pm 1.96 \hat{SE}[\hat{\mu}]$

Remember where 1.96 came from?

$$P \left( -1.96 \leq \frac{\hat{\mu} - \mu}{\hat{SE}[\hat{\mu}]} \leq 1.96 \right) = 95\%$$
Is 95% all there is?

Our 95% CI had the following form: \( \hat{\mu} \pm 1.96 \hat{SE}[\hat{\mu}] \)

Remember where 1.96 came from?

\[
P \left( -1.96 \leq \frac{\hat{\mu} - \mu}{\hat{SE}[\hat{\mu}]} \leq 1.96 \right) = 95\% 
\]

What if we want a different percentage?
Is 95% all there is?

Our 95% CI had the following form: \( \hat{\mu} \pm 1.96 \hat{SE}[\hat{\mu}] \)

Remember where 1.96 came from?

\[
P \left( -1.96 \leq \frac{\hat{\mu} - \mu}{\hat{SE}[\hat{\mu}]} \leq 1.96 \right) = 95\%
\]

What if we want a different percentage?

\[
P \left( -z \leq \frac{\hat{\mu} - \mu}{\hat{SE}[\hat{\mu}]} \leq z \right) = (1 - \alpha)\%
\]

How can we find \( z \)?
We know that $z$ comes from the probability in the tails of the standard normal distribution.
We know that $z$ comes from the probability in the tails of the standard normal distribution.

When $(1 - \alpha) = 0.95$, we want to pick $z$ so that 2.5% of the probability is in each tail.
We know that $z$ comes from the probability in the tails of the standard normal distribution.

When $(1 - \alpha) = 0.95$, we want to pick $z$ so that 2.5% of the probability is in each tail.

This gives us a value of 1.96 for $z$. 

What if we want a 50% confidence interval?
What if we want a 50% confidence interval?

When \((1 - \alpha) = 0.50\), we want to pick \(z\) so that 25% of the probability is in each tail.
What if we want a 50% confidence interval?

When \((1 - \alpha) = 0.50\), we want to pick \(z\) so that 25% of the probability is in each tail.

This gives us a value of 0.67 for \(z\).
\[(1 - \alpha)\% \text{ Confidence Intervals}\]
(1 − \(\alpha\))% Confidence Intervals

In general, let \(z_{\alpha/2}\) be the value associated with \((1 − \alpha)\)% coverage:

\[
P(-z_{\alpha/2} \leq \hat{\mu} - \mu \leq z_{\alpha/2} \hat{SE} = (1 - \alpha)\)%
\]

\[
P(\hat{\mu} - z_{\alpha/2} \hat{SE} \leq \mu \leq \hat{\mu} + z_{\alpha/2} \hat{SE} = (1 - \alpha)\)%
\]

We usually construct the \((1 - \alpha)\)% confidence interval with the following formula.

\[\hat{\mu} \pm z_{\alpha/2} \hat{SE} \]
(1 − \(\alpha\))% Confidence Intervals

In general, let \(z_{\alpha/2}\) be the value associated with (1 − \(\alpha\))% coverage:

\[
P \left( -z_{\alpha/2} \leq \frac{\hat{\mu} - \mu}{SE[\hat{\mu}]} \leq z_{\alpha/2} \right) = (1 - \alpha)\%
\]

\[
P \left( \hat{\mu} - z_{\alpha/2}SE[\hat{\mu}] \leq \mu \leq \hat{\mu} + z_{\alpha/2}SE[\hat{\mu}] \right) = (1 - \alpha)\%
\]
(1 − α)% Confidence Intervals

In general, let \( z_{\alpha/2} \) be the value associated with \( (1 − \alpha)\% \) coverage:

\[
P \left( -z_{\alpha/2} \leq \frac{\hat{\mu} - \mu}{SE[\hat{\mu}]} \leq z_{\alpha/2} \right) = (1 − \alpha)\%
\]

\[
P \left( \hat{\mu} - z_{\alpha/2}SE[\hat{\mu}] \leq \mu \leq \hat{\mu} + z_{\alpha/2}SE[\hat{\mu}] \right) = (1 − \alpha)\%
\]

We usually construct the \( (1 − \alpha)\% \) confidence interval with the following formula.

\[\hat{\mu} \pm z_{\alpha/2}SE[\hat{\mu}]\]
Statistical problems emerge from real science

Comparing different methods of growing barley (Full history: https://www.jstor.org/stable/2245613)
The problem with small samples

Up to this point, we have relied on large sample sizes to construct confidence intervals. If the sample is large enough, then the sampling distribution of the sample mean follows a normal distribution. If the sample is large enough, then the sample standard deviation ($S$) is a good approximation for the population standard deviation ($\sigma$).

When the sample size is small, we need to know something about the distribution in order to construct confidence intervals with the correct coverage (because we can't appeal to the CLT or assume that $S$ is a good approximation of $\sigma$).
The problem with small samples

Up to this point, we have relied on large sample sizes to construct confidence intervals.
The problem with small samples

Up to this point, we have relied on large sample sizes to construct confidence intervals.
The problem with small samples

Up to this point, we have relied on large sample sizes to construct confidence intervals.

If the sample is large enough, then the sampling distribution of the sample mean follows a normal distribution.
The problem with small samples

Up to this point, we have relied on large sample sizes to construct confidence intervals.

If the sample is large enough, then the sampling distribution of the sample mean follows a normal distribution.
The problem with small samples

Up to this point, we have relied on large sample sizes to construct confidence intervals.

If the sample is large enough, then the sampling distribution of the sample mean follows a normal distribution.

If the sample is large enough, then the sample standard deviation ($S$) is a good approximation for the population standard deviation ($\sigma$).
The problem with small samples

Up to this point, we have relied on large sample sizes to construct confidence intervals.

If the sample is large enough, then the sampling distribution of the sample mean follows a normal distribution.

If the sample is large enough, then the sample standard deviation ($S$) is a good approximation for the population standard deviation ($\sigma$).
The problem with small samples

Up to this point, we have relied on large sample sizes to construct confidence intervals.

If the sample is large enough, then the sampling distribution of the sample mean follows a normal distribution.

If the sample is large enough, then the sample standard deviation ($S$) is a good approximation for the population standard deviation ($\sigma$).

When the sample size is small, we need to know something about the distribution in order to construct confidence intervals with the correct coverage (because we can’t appeal to the CLT or assume that $S$ is a good approximation of $\sigma$).
BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

By STUDENT.

https://www.jstor.org/stable/2331554
Canonical Small Sample Example

What happens if we use the large-sample formula?
What happens if we use the large-sample formula?

The percent alcohol in Guinness beer is distributed $N(4.2, 0.09)$. 
Canonical Small Sample Example

What happens if we use the large-sample formula?

The percent alcohol in Guinness beer is distributed $N(4.2, 0.09)$.

Take 100 six-packs of Guinness and construct CIs of the form

$$\hat{\mu} \pm 1.96 \hat{SE} [\hat{\mu}]$$
What happens if we use the large-sample formula?

The percent alcohol in Guinness beer is distributed $N(4.2, 0.09)$.

Take 100 six-packs of Guinness and construct CIs of the form

$$\hat{\mu} \pm 1.96 \hat{SE}[\hat{\mu}]$$

In this sample, only 88 of the 100 CIs cover the true value.
The $t$ distribution

If $X$ is normally distributed, then $X$ is normally distributed even in small samples. Assume $X \sim N(\mu, \sigma^2)$.

If we know $\sigma$, then $X - \mu \frac{\sigma}{\sqrt{n}} \sim N(0,1)$.

We rarely know $\sigma$ and have to use an estimate instead: $X - \mu \frac{s}{\sqrt{n}} \sim \text{Student's } t$.
The $t$ distribution

If $X$ is normally distributed, then $\bar{X}$ is normally distributed even in small samples. Assume

$$X \sim N(\mu, \sigma^2)$$
The $t$ distribution

If $X$ is normally distributed, then $\overline{X}$ is normally distributed even in small samples. Assume

$$X \sim N(\mu, \sigma^2)$$

If we know $\sigma$, then

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$
The $t$ distribution

If $X$ is normally distributed, then $\overline{X}$ is normally distributed even in small samples. Assume

$$X \sim N(\mu, \sigma^2)$$

If we know $\sigma$, then

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

We rarely know $\sigma$ and have to use an estimate instead:

$$\frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$
The $t$ distribution

Since we have to estimate $\sigma$, the distribution of $\frac{\bar{X} - \mu}{s/\sqrt{n}}$ is still bell-shaped but is more spread out.
The \( t \) distribution

Since we have to estimate \( \sigma \), the distribution of \( \frac{\bar{X} - \mu}{s / \sqrt{n}} \) is still bell-shaped but is more spread out.

As the sample size increases, our estimates of \( \sigma \) improve and extreme values of \( \frac{\bar{X} - \mu}{s / \sqrt{n}} \) become less likely.
Since we have to estimate $\sigma$, the distribution of $\frac{X - \mu}{s/\sqrt{n}}$ is still bell-shaped but is more spread out.

As the sample size increases, our estimates of $\sigma$ improve and extreme values of $\frac{X - \mu}{s/\sqrt{n}}$ become less likely.

Eventually the $t$ distribution converges to the standard normal.
The $t$ distribution

Since we have to estimate $\sigma$, the distribution of $\frac{X - \mu}{s / \sqrt{n}}$ is still bell-shaped but is more spread out.

As the sample size increases, our estimates of $\sigma$ improve and extreme values of $\frac{X - \mu}{s / \sqrt{n}}$ become less likely.

Eventually the $t$ distribution converges to the standard normal.
The $t$ distribution

Since we have to estimate $\sigma$, the distribution of $\frac{X - \mu}{s/\sqrt{n}}$ is still bell-shaped but is more spread out.

As the sample size increases, our estimates of $\sigma$ improve and extreme values of $\frac{X - \mu}{s/\sqrt{n}}$ become less likely.

Eventually the $t$ distribution converges to the standard normal.
(1 − α)% Confidence Intervals

In general, let $t_{\alpha/2}$ be the value associated with (1 − α)% coverage:

$$P(-t_{\alpha/2} \leq \hat{\mu} - \mu \leq t_{\alpha/2}) = (1 - \alpha)\%$$

We usually construct the (1 − α)% confidence interval with the following formula.

$$\hat{\mu} \pm t_{\alpha/2} \hat{SE}[\hat{\mu}]$$
(1 − α)% Confidence Intervals

In general, let $t_{\alpha/2}$ be the value associated with $(1 - \alpha)%$ coverage:

$$P \left( -t_{\alpha/2} \leq \frac{\hat{\mu} - \mu}{SE[\hat{\mu}]} \leq t_{\alpha/2} \right) = (1 - \alpha)%$$

$$P \left( \hat{\mu} - t_{\alpha/2} SE[\hat{\mu}] \leq \mu \leq \hat{\mu} + t_{\alpha/2} SE[\hat{\mu}] \right) = (1 - \alpha)%$$
(1 − α)% Confidence Intervals

In general, let $t_{\alpha/2}$ be the value associated with $(1 − \alpha)%$ coverage:

$$P \left( -t_{\alpha/2} \leq \frac{\hat{\mu} - \mu}{SE[\hat{\mu}]} \leq t_{\alpha/2} \right) = (1 - \alpha)%$$

$$P \left( \hat{\mu} - t_{\alpha/2}SE[\hat{\mu}] \leq \mu \leq \hat{\mu} + t_{\alpha/2}SE[\hat{\mu}] \right) = (1 - \alpha)%$$

We usually construct the $(1 − \alpha)%$ confidence interval with the following formula.

$$\hat{\mu} \pm t_{\alpha/2}SE[\hat{\mu}]$$
Small Sample Example

When we generated 95% CIs with the large sample formula

$$\hat{\mu} \pm 1.96 \hat{SE} [\hat{\mu}]$$

only 88 out of 100 intervals covered the true value.
Small Sample Example

When we generated 95% CIs with the large sample formula

$$\hat{\mu} \pm 1.96 \hat{SE}[\hat{\mu}]$$

only 88 out of 100 intervals covered the true value.

When we use the correct small-sample formula

$$\hat{\mu} \pm t_{\alpha/2} \hat{SE}[\hat{\mu}]$$

Percent alcohol

3.9 4.0 4.1 4.2 4.3 4.4 4.5

Stewart (Princeton)  Week 3: Learning From Random Samples  September 14-18, 2020  113 / 142
Small Sample Example

When we generated 95% CIs with the large sample formula

\[ \hat{\mu} \pm 1.96\hat{SE}[\hat{\mu}] \]

only 88 out of 100 intervals covered the true value.

When we use the correct small-sample formula

\[ \hat{\mu} \pm 2.57\hat{SE}[\hat{\mu}] \]
Small Sample Example

When we generated 95% CIs with the large sample formula

\[ \hat{\mu} \pm 1.96 \hat{SE}[\hat{\mu}] \]

only 88 out of 100 intervals covered the true value.

When we use the correct small-sample formula

\[ \hat{\mu} \pm 2.57 \hat{SE}[\hat{\mu}] \]

95 of the 100 CIs in this sample cover the truth.
Another Rationale for the \( t \)-Distribution

Does \( \bar{X}_n \sim N(\mu, S_n^2/n) \), which would imply \( \frac{\bar{X}_n-\mu}{S_n/\sqrt{n}} \sim N(0, 1) \)?

No, because \( S_n \) is a random variable instead of a parameter (like \( \sigma \)). Thus, we need to derive the sampling distribution of the new random variable.

It turns out that \( T_n \) follows Student’s \( t \)-distribution with \( n-1 \) degrees of freedom.

Theorem (Distribution of \( t \)-Value from a Normal Population)

Suppose we have an i.i.d. random sample of size \( n \) from \( N(\mu, \sigma^2) \). Then, the sample mean \( \bar{X}_n \) standardized with the estimated standard error \( S_n/\sqrt{n} \) satisfies,

\[
T_n = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim \tau_{n-1}
\]
Another Rationale for the $t$-Distribution

Does $\overline{X}_n \sim N(\mu, S_n^2/n)$, which would imply $\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim N(0, 1)$? No, because $S_n$ is a random variable instead of a parameter (like $\sigma$).
Another Rationale for the $t$-Distribution

Does $\overline{X}_n \sim N(\mu, S_n^2/n)$, which would imply $\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim N(0, 1)$?

No, because $S_n$ is a random variable instead of a parameter (like $\sigma$).

Thus, we need to derive the sampling distribution of the new random variable.
Another Rationale for the $t$-Distribution

Does $\bar{X}_n \sim N(\mu, S^2_n/n)$, which would imply $\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim N(0, 1)$? No, because $S_n$ is a random variable instead of a parameter (like $\sigma$).

Thus, we need to derive the sampling distribution of the new random variable. It turns out that $T_n$ follows Student’s $t$-distribution with $n - 1$ degrees of freedom.

**Theorem (Distribution of $t$-Value from a Normal Population)**

Suppose we have an i.i.d. random sample of size $n$ from $N(\mu, \sigma^2)$. Then, the sample mean $\bar{X}_n$ standardized with the estimated standard error $S_n/\sqrt{n}$ satisfies,

$$T_n \equiv \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim \tau_{n-1}$$
Kuklinski Example Returns

The Kuklinski et al. (1997) article compares responses to the baseline list with responses to the treatment list. How should we estimate the difference between the two groups? How should we obtain a confidence interval for our estimate?
The Kuklinski et al. (1997) article compares responses to the baseline list with responses to the treatment list.
Kuklinski Example Returns

The Kuklinski et al. (1997) article compares responses to the baseline list with responses to the treatment list.

- How should we estimate the difference between the two groups?
The Kuklinski et al. (1997) article compares responses to the baseline list with responses to the treatment list.

- How should we estimate the difference between the two groups?
- How should we obtain a confidence interval for our estimate?
Comparing Two Groups

We will often assume the following when comparing two groups, $X_{11}, X_{12}, \ldots, X_{1n} \sim i.i.d. \mu_1, \sigma_1^2$.

$X_{21}, X_{22}, \ldots, X_{2n} \sim i.i.d. \mu_2, \sigma_2^2$.

The two samples are independent of each other.

We will usually be interested in comparing $\mu_1$ to $\mu_2$, although we will sometimes need to compare $\sigma_1^2$ to $\sigma_2^2$ in order to make the first comparison.
Comparing Two Groups

We will often assume the following when comparing two groups,
Comparing Two Groups

We will often assume the following when comparing two groups,

- $X_{11}, X_{12}, \ldots, X_{1n_1} \sim i.i.d. ?(\mu_1, \sigma_1^2)$
- $X_{21}, X_{22}, \ldots, X_{2n_2} \sim i.i.d. ?(\mu_2, \sigma_2^2)$
- The two samples are independent of each other.
Comparing Two Groups

We will often assume the following when comparing two groups,

- $X_{11}, X_{12}, ..., X_{1n_1} \sim i.i.d.? \left( \mu_1, \sigma_1^2 \right)$
- $X_{21}, X_{22}, ..., X_{2n_2} \sim i.i.d.? \left( \mu_2, \sigma_2^2 \right)$
- The two samples are independent of each other.

We will usually be interested in comparing $\mu_1$ to $\mu_2$, although we will sometimes need to compare $\sigma_1^2$ to $\sigma_2^2$ in order to make the first comparison.
Sampling Distribution for $\bar{X}_1 - \bar{X}_2$

What is the expected value of $\bar{X}_1 - \bar{X}_2$?
Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the expected value of $\overline{X}_1 - \overline{X}_2$?

$$E[\overline{X}_1 - \overline{X}_2]$$
Sampling Distribution for $\bar{X}_1 - \bar{X}_2$

What is the expected value of $\bar{X}_1 - \bar{X}_2$?

$$E[\bar{X}_1 - \bar{X}_2] = E[\bar{X}_1] - E[\bar{X}_2]$$
Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the expected value of $\overline{X}_1 - \overline{X}_2$?

\[
E[\overline{X}_1 - \overline{X}_2] = E[\overline{X}_1] - E[\overline{X}_2] = \frac{1}{n_1} \sum E[X_{1i}] - \frac{1}{n_2} \sum E[X_{2j}]
\]
Sampling Distribution for $\bar{X}_1 - \bar{X}_2$

What is the expected value of $\bar{X}_1 - \bar{X}_2$?

$$E[\bar{X}_1 - \bar{X}_2] = E[\bar{X}_1] - E[\bar{X}_2]$$

$$= \frac{1}{n_1} \sum E[X_{1i}] - \frac{1}{n_2} \sum E[X_{2j}]$$

$$= \frac{1}{n_1} \sum \mu_1 - \frac{1}{n_2} \sum \mu_2$$
Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the expected value of $\overline{X}_1 - \overline{X}_2$?

\[
E[\overline{X}_1 - \overline{X}_2] = E[\overline{X}_1] - E[\overline{X}_2] \\
= \frac{1}{n_1} \sum E[X_{1i}] - \frac{1}{n_2} \sum E[X_{2j}] \\
= \frac{1}{n_1} \sum \mu_1 - \frac{1}{n_2} \sum \mu_2 \\
= \mu_1 - \mu_2
\]
Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the variance of $\overline{X}_1 - \overline{X}_2$?
Sampling Distribution for $\bar{X}_1 - \bar{X}_2$

What is the variance of $\bar{X}_1 - \bar{X}_2$?

$$\text{Var} [\bar{X}_1 - \bar{X}_2]$$
Sampling Distribution for $\bar{X}_1 - \bar{X}_2$

What is the variance of $\bar{X}_1 - \bar{X}_2$?

$$Var[\bar{X}_1 - \bar{X}_2] = Var[\bar{X}_1] + Var[\bar{X}_2]$$
What is the variance of $X_1 - X_2$?

\[
\text{Var}[X_1 - X_2] = \text{Var}[X_1] + \text{Var}[X_2]
\]

\[
= \frac{1}{n_1^2} \sum \text{Var}[X_{1i}] + \frac{1}{n_2^2} \sum \text{Var}[X_{2j}]
\]
Sampling Distribution for \( \overline{X}_1 - \overline{X}_2 \)

What is the variance of \( \overline{X}_1 - \overline{X}_2 \)?

\[
\begin{align*}
\text{Var}[\overline{X}_1 - \overline{X}_2] &= \text{Var}[\overline{X}_1] + \text{Var}[\overline{X}_2] \\
&= \frac{1}{n_1^2} \sum \text{Var}[X_{1i}] + \frac{1}{n_2^2} \sum \text{Var}[X_{2j}] \\
&= \frac{1}{n_1^2} \sum \sigma_1^2 + \frac{1}{n_2^2} \sum \sigma_2^2
\end{align*}
\]
Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the variance of $\overline{X}_1 - \overline{X}_2$?

\[
\text{Var}[\overline{X}_1 - \overline{X}_2] = \text{Var}[\overline{X}_1] + \text{Var}[\overline{X}_2]
\]
\[
= \frac{1}{n_1^2} \sum \text{Var}[X_{1i}] + \frac{1}{n_2^2} \sum \text{Var}[X_{2j}]
\]
\[
= \frac{1}{n_1^2} \sum \sigma_1^2 + \frac{1}{n_2^2} \sum \sigma_2^2
\]
\[
= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}
\]
Sampling Distribution for $\bar{X}_1 - \bar{X}_2$

What is the distributional form for $\bar{X}_1 - \bar{X}_2$?
Sampling Distribution for $\bar{X}_1 - \bar{X}_2$

What is the distributional form for $\bar{X}_1 - \bar{X}_2$?

- $\bar{X}_1$ is distributed $\sim N(\mu_1, \frac{\sigma^2_1}{n_1})$. 

- $\bar{X}_2$ is distributed $\sim N(\mu_2, \frac{\sigma^2_2}{n_2})$. 

- $\bar{X}_1 - \bar{X}_2$ is distributed $\sim N(\mu_1 - \mu_2, \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2})$. 

What is the distributional form for $\overline{X}_1 - \overline{X}_2$?

- $\overline{X}_1$ is distributed $\sim N(\mu_1, \frac{\sigma^2_1}{n_1})$.
- $\overline{X}_2$ is distributed $\sim N(\mu_2, \frac{\sigma^2_2}{n_2})$. 
What is the distributional form for \( \bar{X}_1 - \bar{X}_2 \)?

- \( \bar{X}_1 \) is distributed \( \sim N(\mu_1, \frac{\sigma_1^2}{n_1}) \).
- \( \bar{X}_2 \) is distributed \( \sim N(\mu_2, \frac{\sigma_2^2}{n_2}) \).
- \( \bar{X}_1 - \bar{X}_2 \) is distributed \( \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \).
CIs for $\mu_1 - \mu_2$

Using the same type of argument that we used for the univariate case, we write a $(1 - \alpha)\%$ CI as the following:

$$\bar{X}_1 - \bar{X}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
Interval estimation of the population proportion

Let's say that we have a sample of iid Bernoulli random variables, $Y_1, \ldots, Y_n$, where each takes $Y_i = 1$ with probability $\pi$. Note that this is also the population proportion of ones. We have shown in previous weeks that the expectation of one of these variables is just the probability of seeing a 1:

$$E[Y_i] = \pi.$$

The variance of a Bernoulli random variable is a simple function of its mean:

$$\text{Var}(Y_i) = \pi(1-\pi).$$

Problem

Show that the sample proportion, $\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} Y_i$, of the above iid Bernoulli sample, is unbiased for the true population proportion, $\pi$, and that the sampling variance is equal to $\pi(1-\pi) \frac{1}{n}$.

Note that if we have an estimate of the population proportion, $\hat{\pi}$, then we also have an estimate of the sampling variance:

$$\hat{\text{Var}}(\hat{\pi}) = \hat{\pi}(1-\hat{\pi}) \frac{1}{n}.$$

Given the facts from the previous problem, we just apply the same logic from the population mean to show the following confidence interval:

$$P(\hat{\pi} - z_{\alpha/2} \times \sqrt{\hat{\pi}(1-\hat{\pi}) \frac{1}{n}} \leq \pi \leq \hat{\pi} + z_{\alpha/2} \times \sqrt{\hat{\pi}(1-\hat{\pi}) \frac{1}{n}}) = 1 - \alpha.$$
Interval estimation of the population proportion

- Let’s say that we have a sample of iid Bernoulli random variables, $Y_1, \ldots, Y_n$, where each takes $Y_i = 1$ with probability $\pi$. Note that this is also the population proportion of ones. We have shown in previous weeks that the expectation of one of these variable is just the probability of seeing a 1: $E[Y_i] = \pi$. 
Interval estimation of the population proportion

Let’s say that we have a sample of iid Bernoulli random variables, $Y_1, \ldots, Y_n$, where each takes $Y_i = 1$ with probability $\pi$. Note that this is also the population proportion of ones. We have shown in previous weeks that the expectation of one of these variable is just the probability of seeing a 1: $E[Y_i] = \pi$.

The variance of a Bernoulli random variable is a simple function of its mean: $\text{Var}(Y_i) = \pi(1 - \pi)$. 

Problem: Show that the sample proportion, $\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} Y_i$, of the above iid Bernoulli sample, is unbiased for the true population proportion, $\pi$, and that the sampling variance is equal to $\pi(1 - \pi) n$.

Note that if we have an estimate of the population proportion, $\hat{\pi}$, then we also have an estimate of the sampling variance: $\hat{\pi}(1 - \hat{\pi}) n$.

Given the facts from the previous problem, we just apply the same logic from the population mean to show the following confidence interval:

$$P(\hat{\pi} - z_{\alpha/2} \times \sqrt{\hat{\pi}(1 - \hat{\pi})} n \leq \pi \leq \hat{\pi} + z_{\alpha/2} \times \sqrt{\hat{\pi}(1 - \hat{\pi})} n) = 1 - \alpha$$
Interval estimation of the population proportion

- Let’s say that we have a sample of iid Bernoulli random variables, \( Y_1, \ldots, Y_n \), where each takes \( Y_i = 1 \) with probability \( \pi \). Note that this is also the population proportion of ones. We have shown in previous weeks that the expectation of one of these variable is just the probability of seeing a 1: \( E[Y_i] = \pi \).

- The variance of a Bernoulli random variable is a simple function of its mean: \( \text{Var}(Y_i) = \pi(1 - \pi) \).

- **Problem** Show that the sample proportion, \( \hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} Y_i \), of the above iid Bernoulli sample, is unbiased for the true population proportion, \( \pi \), and that the sampling variance is equal to \( \frac{\pi(1 - \pi)}{n} \).
Interval estimation of the population proportion

- Let’s say that we have a sample of iid Bernoulli random variables, $Y_1, \ldots, Y_n$, where each takes $Y_i = 1$ with probability $\pi$. Note that this is also the population proportion of ones. We have shown in previous weeks that the expectation of one of these variable is just the probability of seeing a 1: $E[Y_i] = \pi$.

- The variance of a Bernoulli random variable is a simple function of its mean: $\text{Var}(Y_i) = \pi(1 - \pi)$.

- **Problem** Show that the sample proportion, $\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} Y_i$, of the above iid Bernoulli sample, is unbiased for the true population proportion, $\pi$, and that the sampling variance is equal to $\frac{\pi(1-\pi)}{n}$.

- Note that if we have an estimate of the population proportion, $\hat{\pi}$, then we also have an estimate of the sampling variance: $\frac{\hat{\pi}(1-\hat{\pi})}{n}$. 

Given the facts from the previous problem, we just apply the same logic from the population mean to show the following confidence interval:

$$P\left(\hat{\pi} - z_{\alpha/2} \times \frac{\sqrt{\hat{\pi}(1 - \hat{\pi})}}{n} \leq \pi \leq \hat{\pi} + z_{\alpha/2} \times \frac{\sqrt{\hat{\pi}(1 - \hat{\pi})}}{n}\right) = 1 - \alpha$$
Interval estimation of the population proportion

- Let’s say that we have a sample of iid Bernoulli random variables, $Y_1, \ldots, Y_n$, where each takes $Y_i = 1$ with probability $\pi$. Note that this is also the population proportion of ones. We have shown in previous weeks that the expectation of one of these variable is just the probability of seeing a 1: $E[Y_i] = \pi$.

- The variance of a Bernoulli random variable is a simple function of its mean: $\text{Var}(Y_i) = \pi(1 - \pi)$.

- **Problem** Show that the sample proportion, $\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} Y_i$, of the above iid Bernoulli sample, is unbiased for the true population proportion, $\pi$, and that the sampling variance is equal to $\frac{\pi(1-\pi)}{n}$.

- Note that if we have an estimate of the population proportion, $\hat{\pi}$, then we also have an estimate of the sampling variance: $\frac{\hat{\pi}(1-\hat{\pi})}{n}$.

- Given the facts from the previous problem, we just apply the same logic from the population mean to show the following confidence interval:

$$P \left( \hat{\pi} - z_{\alpha/2} \times \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \leq \pi \leq \hat{\pi} + z_{\alpha/2} \times \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \right) = (1 - \alpha)$$
Gerber, Green, and Larimer experiment

Let's go back to the Gerber, Green, and Larimer experiment from last class. Here are the results of their experiment:

Let's use what we have learned up until now and the information in the table to calculate a 95% confidence interval for the difference in proportions voting between the Neighbors group and the Civic Duty group.

You may assume that the samples within each group are iid and the two samples are independent.
Gerber, Green, and Larimer experiment

Let’s go back to the Gerber, Green, and Larimer experiment from last class. Here are the results of their experiment:
Gerber, Green, and Larimer experiment

Let’s go back to the Gerber, Green, and Larimer experiment from last class. Here are the results of their experiment:

<table>
<thead>
<tr>
<th>Experimental Group</th>
<th>Control</th>
<th>Civic Duty</th>
<th>Hawthorne</th>
<th>Self</th>
<th>Neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Voting</td>
<td>29.7%</td>
<td>31.5%</td>
<td>32.2%</td>
<td>34.5%</td>
<td>37.8%</td>
</tr>
<tr>
<td>N of Individuals</td>
<td>191,243</td>
<td>38,218</td>
<td>38,204</td>
<td>38,218</td>
<td>38,201</td>
</tr>
</tbody>
</table>

Let’s use what we have learned up until now and the information in the table to calculate a 95% confidence interval for the difference in proportions voting between the Neighbors group and the Civic Duty group.
Let’s go back to the Gerber, Green, and Larimer experiment from last class. Here are the results of their experiment:

<table>
<thead>
<tr>
<th>Experimental Group</th>
<th>Control</th>
<th>Civic Duty</th>
<th>Hawthorne</th>
<th>Self</th>
<th>Neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Voting</td>
<td>29.7%</td>
<td>31.5%</td>
<td>32.2%</td>
<td>34.5%</td>
<td>37.8%</td>
</tr>
<tr>
<td>N of Individuals</td>
<td>191,243</td>
<td>38,218</td>
<td>38,204</td>
<td>38,218</td>
<td>38,201</td>
</tr>
</tbody>
</table>

- Let’s use what we have learned up until now and the information in the table to calculate a 95% confidence interval for the difference in proportions voting between the Neighbors group and the Civic Duty group.
- You may assume that the samples with in each group are iid and the two samples are independent.
Calculating the CI for social pressure effect

- We know distribution of sample proportion turned among Civic Duty group $\hat{\pi}_C \sim N(\pi_C, (\pi_C(1 - \pi_C))/n_C)$
Calculating the CI for social pressure effect

- We know distribution of sample proportion turned among Civic Duty group $\hat{\pi}_C \sim N(\pi_C, (\pi_C(1 - \pi_C))/n_C)$
- Sample proportions are just sample means, so we can do difference in means:
  \[
  \hat{\pi}_N - \hat{\pi}_C \sim N(\pi_N - \pi_C, \sqrt{SE^2_N + SE^2_C})
  \]
Calculating the CI for social pressure effect

- We know distribution of sample proportion turned among Civic Duty group $\hat{\pi}_C \sim N(\pi_C, (\pi_C(1 - \pi_C))/n_C)$
- Sample proportions are just sample means, so we can do difference in means:
  $$\hat{\pi}_N - \hat{\pi}_C \sim N(\pi_N - \pi_C, \sqrt{SE_N^2 + SE_C^2})$$
- Replace the variances with our estimates:
  $$\hat{\pi}_N - \hat{\pi}_C \sim N(\pi_N - \pi_C, \sqrt{\hat{SE}_N^2 + \hat{SE}_C^2})$$
Calculating the CI for social pressure effect

- We know distribution of sample proportion turned among Civic Duty group $\hat{\pi}_C \sim N(\pi_C, (\pi_C(1 - \pi_C))/n_C)$

- Sample proportions are just sample means, so we can do difference in means:

$$\hat{\pi}_N - \hat{\pi}_C \sim N\left(\pi_N - \pi_C, \sqrt{SE^2_N + SE^2_C}\right)$$

- Replace the variances with our estimates:

$$\hat{\pi}_N - \hat{\pi}_C \sim N\left(\pi_N - \pi_C, \sqrt{\hat{SE}^2_N + \hat{SE}^2_C}\right)$$

- Apply usual formula to get 95% confidence interval:

$$(\hat{\pi}_N - \hat{\pi}_C) \pm 1.96 \times \sqrt{\hat{SE}^2_N + \hat{SE}^2_C}$$
Calculating the CI for social pressure effect

- We know distribution of sample proportion turned among Civic Duty group $\hat{\pi}_C \sim N(\pi_C, (\pi_C(1 - \pi_C))/n_C)$
- Sample proportions are just sample means, so we can do difference in means:
  \[
  \hat{\pi}_N - \hat{\pi}_C \sim N\left(\pi_N - \pi_C, \sqrt{SE_N^2 + SE_C^2}\right)
  \]
- Replace the variances with our estimates:
  \[
  \hat{\pi}_N - \hat{\pi}_C \sim N\left(\pi_N - \pi_C, \sqrt{\hat{SE}_N^2 + \hat{SE}_C^2}\right)
  \]
- Apply usual formula to get 95% confidence interval:
  \[
  (\hat{\pi}_N - \hat{\pi}_C) \pm 1.96 \times \sqrt{\hat{SE}_N^2 + \hat{SE}_C^2}
  \]
- Remember that we can calculate the sample variance for a sample proportion like so: $(\hat{\pi}_C(1 - \hat{\pi}_C))/n_C$
Now, we calculate the 95% confidence interval:

\[
\hat{\pi}_N - \hat{\pi}_C \pm 1.96 \times \sqrt{\hat{\pi}_N (1 - \hat{\pi}_N) / n_N + \hat{\pi}_C (1 - \hat{\pi}_C) / n_C}
\]

```r
n.n <- 38201
samp.var.n <- (0.378 * (1 - 0.378))/n.n
n.c <- 38218
samp.var.c <- (0.315 * (1 - 0.315))/n.c
se.diff <- sqrt(samp.var.n + samp.var.c)

## lower bound
(0.378 - 0.315) - 1.96 * se.diff
## [1] 0.05626701

## upper bound
(0.378 - 0.315) + 1.96 * se.diff
## [1] 0.06973299
```

Thus, the confidence interval for the effect is [0.056267, 0.069733].

### TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

<table>
<thead>
<tr>
<th>Experimental Group</th>
<th>Control</th>
<th>Civic Duty</th>
<th>Hawthorne</th>
<th>Self</th>
<th>Neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Voting</td>
<td>29.7%</td>
<td>31.5%</td>
<td>32.2%</td>
<td>34.5%</td>
<td>37.8%</td>
</tr>
<tr>
<td>N of Individuals</td>
<td>191,243</td>
<td>38,218</td>
<td>38,204</td>
<td>38,218</td>
<td>38,201</td>
</tr>
</tbody>
</table>
Gerber, Green, and Larimer experiment

Now, we calculate the 95% confidence interval:

\[
(\hat{\pi}_N - \hat{\pi}_C) \pm 1.96 \times \sqrt{\frac{\hat{\pi}_N(1 - \hat{\pi}_N)}{n_N} + \frac{\hat{\pi}_C(1 - \hat{\pi}_C)}{n_C}}
\]

<table>
<thead>
<tr>
<th>Experimental Group</th>
<th>Control</th>
<th>Civic Duty</th>
<th>Hawthorne</th>
<th>Self</th>
<th>Neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Voting (%)</td>
<td>29.7</td>
<td>31.5</td>
<td>32.2</td>
<td>34.5</td>
<td>37.8</td>
</tr>
<tr>
<td>N of Individuals</td>
<td>191,243</td>
<td>38,218</td>
<td>38,204</td>
<td>38,218</td>
<td>38,201</td>
</tr>
</tbody>
</table>

Thus, the confidence interval for the effect is [0.056267, 0.069733].
Gerber, Green, and Larimer experiment

Now, we calculate the 95% confidence interval:

\[
(\hat{\pi}_N - \hat{\pi}_C) \pm 1.96 \times \sqrt{\frac{\hat{\pi}_N(1-\hat{\pi}_N)}{n_N} + \frac{\hat{\pi}_C(1-\hat{\pi}_C)}{n_C}}
\]

\[
n.n <- 38201
samp.var.n <- (0.378 * (1 - 0.378))/n.n
n.c <- 38218
samp.var.c <- (0.315 * (1 - 0.315))/n.c
se.diff <- sqrt(samp.var.n + samp.var.c)
## lower bound
(0.378 - 0.315) - 1.96 * se.diff
## [1] 0.05626701
## upper bound
(0.378 - 0.315) + 1.96 * se.diff
## [1] 0.06973299

Thus, the confidence interval for the effect is \([0.056267, 0.069733]\).
We can use our analytic samples to find a confidence interval.
To use the confidence interval formula, we need to find:

1. The distribution
2. Confidence level • Alpha
3. Sidedness
4. Critical value(s)
5. Standard error of our estimate

For a proportion, the formula is:

\[ SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \]

```r
## Calculating our critical value
cv <- qnorm(.975)
cv

## [1] 1.959964
```

```r
## Finding the standard error of our estimate
se <- sqrt(red.sample*(1-red.sample)/n.samp)
se

## [1] 0.01966499
```
Calculating the confidence interval

\[ CI(\alpha) = [r - \frac{z_{\alpha/2} \times SE}{r} + \frac{z_{\alpha/2} \times SE}{r}] \]

Finding and printing the confidence interval

\[
c(\text{red.sample} - \text{cv*se}, \text{red.sample} + \text{cv*se})
\]

## [1] 0.2234573 0.3005427
Our results

26.2% red with a 95 percent confidence interval of $[22.3, 30.1]$
We hope our sample is in the 95%
We Covered...
We Covered...

- Interval estimates provide a means of assessing uncertainty.
We Covered...

- Interval estimates provide a means of assessing uncertainty.
- Interval estimators have sampling distributions.
We Covered...

- Interval estimates provide a means of assessing uncertainty.
- Interval estimators have sampling distributions.
- Interval estimates should be interpreted in terms of repeated sampling.

Next Time:
We Covered...

- Interval estimates provide a means of assessing uncertainty.
- Interval estimators have sampling distributions.
- Interval estimates should be interpreted in terms of repeated sampling.

Next Time: The plug-in principle!
Where We’ve Been and Where We’re Going…

- **Last Week**
  - random variables
  - joint distributions

- **This Week**
  - estimators and sampling distributions
  - estimator properties (bias, variance, consistency)
  - confidence intervals

- **Next Week**
  - hypothesis testing
  - what is regression?

- **Long Run**
  - probability $\rightarrow$ inference $\rightarrow$ regression $\rightarrow$ causal inference
1. **Estimation**
   - Populations and Samples
   - Estimators
   - Analytical

2. **Weak Law of Large Numbers and the Central Limit Theorem**
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3. **Properties of Estimators**
   - Four Desirable Properties
   - Example

4. **Interval Estimation**
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5. **Plug-In Principle**
1. Estimation
   - Populations and Samples
   - Estimators
   - Analytical

2. Weak Law of Large Numbers and the Central Limit Theorem
   - Chebychev’s Inequality
   - Weak Law of Large Numbers
   - The Central Limit Theorem

3. Properties of Estimators
   - Four Desirable Properties
   - Example

4. Interval Estimation
   - Intervals
   - Large Sample Intervals for a Mean
   - Small Sample Intervals for a Mean
   - Comparing Two Groups
   - Interval Estimation for a Proportion

5. Plug-In Principle
Coming up with Estimators

We now know how to study some properties of estimators, but how do we come up with candidate estimators?

The simplest way is to use the sample analog. Ex: If we're interested in the population mean, we use the sample mean. This is justified because of the plug-in principle. The Weak Law of Large Numbers tells us that the empirical CDF is a good sample analog of the true CDF (which fully describes a distribution).
Coming up with Estimators

We now know how to study some properties of estimators, but how do we come up with candidate estimators?

The simplest way is to use the sample analog. Ex: If we're interested in the population mean, we use the sample mean. This is justified because of the plug-in principle. The Weak Law of Large Numbers tells us that the empirical CDF is a good sample analog of the true CDF (which fully describes a distribution).
Coming up with Estimators

We now know how to study some properties of estimators, but how do we come up with candidate estimators?

- The simplest way is to use the sample analog.

Ex: If we're interested in the population mean, we use the sample mean. This is justified because of the plug-in principle. The Weak Law of Large Numbers tells us that the empirical CDF is a good sample analog of the true CDF (which fully describes a distribution).
Coming up with Estimators

We now know how to study some properties of estimators, but how do we come up with candidate estimators?

- The simplest way is to use the sample analog.
- Ex: If we’re interested in the population mean, we use the sample mean.
Coming up with Estimators

We now know how to study some properties of estimators, but how do we come up with candidate estimators?

- The **simplest** way is to use the **sample analog**.
- **Ex**: If we’re interested in the population mean, we use the sample mean
- This is justified because of the **plug-in principle**.
We now know how to study some properties of estimators, but how do we come up with candidate estimators?

- The **simplest** way is to use the **sample analog**.
- **Ex:** If we’re interested in the population mean, we use the sample mean.
- This is justified because of the **plug-in principle**.
- The Weak Law of Large Numbers tells us that the **empirical CDF** is a good sample analog of the true CDF (which fully describes a distribution).
The Plug-in Principle in Action

Say we have a $\mathcal{N}(5, 4)$ distribution

![Empirical CDF graph](image)
The Plug-in Principle

Note that the CDF is:
\[ F(x) = P(X \leq x) = E[I(X \leq x)] \]
we define the empirical CDF (eCDF) as:
\[ \hat{F}(x) = I(X \leq x), \quad \forall x \in \mathbb{R} \]
WLLN tells us that the eCDF will be unbiased and consistently estimated. Any given sample will, on average, look representative of the true distribution.

For iid random variables \( X_1, X_2, \ldots, X_n \) with common CDF \( F \), the plug-in estimator of \( \theta = T(F) \) is:
\[ \hat{\theta} = T(\hat{F}) \]
if \( T \) is well-behaved, then \( \hat{\theta} \) is also asymptotically normal.
The Plug-in Principle

Note that the CDF is:

\[ F(x) = P(X \leq x) = E[\mathbb{1}(X \leq x)] \]
The Plug-in Principle

Note that the CDF is:

\[ F(x) = P(X \leq x) = E[\mathbb{I}(X \leq x)] \]

we define the empirical CDF (eCDF) as:

\[ \hat{F}(x) = \mathbb{I}(X \leq x), \forall x \in \mathbb{R} \]
The Plug-in Principle

Note that the CDF is:

\[ F(x) = P(X \leq x) = E[\mathbb{1}(X \leq x)] \]

we define the empirical CDF (eCDF) as:

\[ \hat{F}(x) = \mathbb{1}(X \leq x), \forall x \in \mathbb{R} \]

WLLN tells us that the eCDF will be unbiased and consistently estimated. Any given sample will, on average, look representative of the true distribution.
The Plug-in Principle

Note that the CDF is:

\[ F(x) = P(X \leq x) = E[\mathbb{I}(X \leq x)] \]

we define the empirical CDF (eCDF) as:

\[ \hat{F}(x) = \mathbb{I}(X \leq x), \forall x \in \mathbb{R} \]

- WLLN tells us that the eCDF will be unbiased and consistently estimated. Any given sample will, on average, look representative of the true distribution.

For iid random variables \( X_1, X_2, \ldots, X_n \) with common CDF \( F \), the plug-in estimator of \( \theta = T(F) \) is:

\[ \hat{\theta} = T(\hat{F}) \]
The Plug-in Principle

Note that the CDF is:

\[ F(x) = P(X \leq x) = E[\mathbb{I}(X \leq x)] \]

we define the empirical CDF (eCDF) as:

\[ \hat{F}(x) = \mathbb{I}(X \leq x), \forall x \in \mathbb{R} \]

- WLLN tells us that the eCDF will be unbiased and consistently estimated. Any given sample will, on average, look representative of the true distribution.

For iid random variables \( X_1, X_2, \ldots, X_n \) with common CDF \( F \), the plug-in estimator of \( \theta = T(F) \) is:

\[ \hat{\theta} = T(\hat{F}) \]

if \( T \) is well-behaved, then \( \hat{\theta} \) is also asymptotically normal.
Bootstrapped Sampling Distributions

What if there was a way to replace thinking with computers?
What if there was a way to replace thinking with computers?

What if there was a way to replacing analytical derivations, which can be hard, with computer simulations which are easy?
Bootstrapped Sampling Distributions

What if there was a way to replace thinking with computers?

What if there was a way to replacing analytical derivations, which can be hard, with computer simulations which are easy?

The plug-in principle gives us a way forward.
Step 1: Resampling

Original sample

Step 2: Estimation

Replicate sample 1
$P_{(1)}$

Replicate sample 2
$P_{(2)}$

Replicate sample $R$
$P_{(R)}$

Step 3: Construction

Confidence Interval

Source: Salganik (2006)
This works for almost* any estimator

*basically it works when plug-in estimation works
Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy

B. Efron and R. Tibshirani

The Bootstrap

1. Take a random sample of size $n$ from our original sample.
2. Calculate our estimate using this bootstrap sample.
3. Repeat steps 1 and 2 many (B) times.
4. Using the resulting collection of bootstrap estimates, calculate the standard deviation of the bootstrap distribution of our estimator. This serves as our estimate of the standard deviation of the sampling distribution.
The Bootstrap

**Bootstrap**: Use the eCDF as a plug-in for the CDF, and resample from that. I.e. we are pretending our sample eCDF looks sufficiently close to our true CDF, and so we’re sampling from the eCDF as an approximation to repeated sampling from the true CDF. This is called a resampling method.
Bootstrap: Use the eCDF as a plug-in for the CDF, and resample from that. I.e. we are pretending our sample eCDF looks sufficiently close to our true CDF, and so we’re sampling from the eCDF as an approximation to repeated sampling from the true CDF. This is called a resampling method.

1. Take a with replacement sample of size $n$ from our sample.
The Bootstrap

Bootstrap: Use the eCDF as a plug-in for the CDF, and resample from that. I.e. we are pretending our sample eCDF looks sufficiently close to our true CDF, and so we’re sampling from the eCDF as an approximation to repeated sampling from the true CDF. This is called a resampling method.

1. Take a with replacement sample of size $n$ from our sample.
2. Calculate our would-be estimate using this bootstrap sample.
The Bootstrap

**Bootstrap:** Use the eCDF as a plug-in for the CDF, and resample from that. I.e. we are pretending our sample eCDF looks sufficiently close to our true CDF, and so we’re sampling from the eCDF as an approximation to repeated sampling from the true CDF. This is called a resampling method.

1. Take a with replacement sample of size $n$ from our sample.
2. Calculate our would-be estimate using this bootstrap sample.
3. Repeat steps 1 and 2 many (B) times.
The Bootstrap

Bootstrap: Use the eCDF as a plug-in for the CDF, and resample from that. I.e. we are pretending our sample eCDF looks sufficiently close to our true CDF, and so we’re sampling from the eCDF as an approximation to repeated sampling from the true CDF. This is called a resampling method.

1. Take a with replacement sample of size $n$ from our sample.
2. Calculate our would-be estimate using this bootstrap sample.
3. Repeat steps 1 and 2 many (B) times.
4. Using the resulting collection of bootstrap estimates, calculate the standard deviation of the bootstrap distribution of our estimator. This serves our estimate of the standard deviation of the sampling distribution.
Example of a Bootstrap

```r
samp <- c(9.7, 4.99, 5.9, 3.58, 8.15, 5.54, 4.77, 5.01, 4.89, 3.42, 8.63, 7.17, 8.93, 7.5, 4.93, 8.6, 6.26, 7.31, 8.96, 3.95)

obs_mean = mean(samp)

obs_mean
```

## [1] 6.4095
Example of a Bootstrap

```r
# resample WITH REPLACEMENT reps times
# recalculate the mean within each bootstrap replicate
boot_samp_dist <- replicate(2000, {
  mean(samp[sample.int(length(samp), replace = TRUE)])
})

ggplot(tibble(boot_samp_dist = boot_samp_dist),
   aes(x = boot_samp_dist)) +
  geom_density() +
  geom_vline(xintercept = obs_mean) +
  theme_bw() + ggtitle("Bootstrap Sampling Distribution For the Sample Mean") +
  xlab("mean")
```
Example of a Bootstrap

Bootstrap Sampling Distribution For the Sample Mean
Two ways to calculate bootstrap intervals

1) Using normal approximation intervals, use the estimates from step 4.

\[
X - \Phi^{-1}(1 - \alpha/2) \times \hat{\sigma}_{\text{boot}},
\]

\[
X + \Phi^{-1}(1 - \alpha/2) \times \hat{\sigma}_{\text{boot}},
\]

Note here that the standard error is just the standard deviation of the bootstrap replicates. There is no square root of \(n\). Why?

2) Percentile method for the CI: Sort \(B\) bootstrap estimates from smallest to largest. The \(\alpha\) interval is constructed as

\[
\text{CI}_{1 - \alpha} = \left[ \alpha/2 \times B_{\text{sample}}, (1 - \alpha/2) \times B_{\text{sample}} \right]
\]

Percentile method does not rely on normal approximation, and behaves better with small \(n\).
Two ways to calculate bootstrap intervals

1) Using normal approximation intervals, use the estimates from step 4.

\[
\left[ \bar{X} - \Phi^{-1}(1 - \alpha/2) \cdot \hat{\sigma}_{\text{boot}}, \bar{X} + \Phi^{-1}(1 - \alpha/2) \cdot \hat{\sigma}_{\text{boot}} \right]
\]

Note here that the standard error is just the standard deviation of the bootstrap replicates. There is no square root of \( n \).

2) Percentile method for the CI: Sort \( B \) bootstrap estimates from smallest to largest. The \( \alpha \) interval is constructed as

\[
\text{CI}_{1-\alpha} = [\alpha/2 \cdot B_{\text{sample}}, (1 - \alpha/2) \cdot B_{\text{sample}}]
\]

Percentile method does not rely on normal approximation, and behaves better with small \( n \).
Two ways to calculate bootstrap intervals

1) Using normal approximation intervals, use the estimates from step 4.

\[
\left[ \bar{X} - \Phi^{-1}(1 - \alpha/2) \times \hat{\sigma}_{\text{boot}}, \bar{X} + \Phi^{-1}(1 - \alpha/2) \times \hat{\sigma}_{\text{boot}} \right]
\]

- Note here that the standard error is just the standard deviation of the bootstrap replicates. There is no square root of \( n \). Why?
Two ways to calculate bootstrap intervals

1) Using normal approximation intervals, use the estimates from step 4.

\[
\left[ \bar{X} - \Phi^{-1}(1 - \alpha/2) \cdot \hat{\sigma}_{\text{boot}}, \bar{X} + \Phi^{-1}(1 - \alpha/2) \cdot \hat{\sigma}_{\text{boot}} \right]
\]

- Note here that the standard error is just the standard deviation of the bootstrap replicates. There is no square root of \( n \). Why?

2) Percentile method for the CI: Sort \( B \) bootstrap estimates from smallest to largest. \( \alpha \) interval is constructed as

\[
CI_{1-\alpha} = \left[ \alpha/2 \cdot B \text{ sample}, (1 - \alpha/2) \cdot B \text{ sample} \right]
\]

- Percentile method does not rely on normal approximation, and behaves better with small \( n \).
We covered

The plug-in principle.
The bootstrap.
We will return to both in future weeks.
We covered

- The plug-in principle.
We covered

- The plug-in principle.
- The bootstrap.
We covered

- The plug-in principle.
- The bootstrap.
- We will return to both in future weeks.
This Week in Review

- Estimation!
- Central Limit Theorem!
- Properties of Estimators!
- Intervals!
- Plug-In Principle!
This Week in Review

- Estimation!
- Central Limit Theorem!
- Properties of Estimators!
- Intervals!
- Plug-In Principle!

Going Deeper:

This Week in Review

- Estimation!
- Central Limit Theorem!
- Properties of Estimators!
- Intervals!
- Plug-In Principle!

Going Deeper:


Next week: hypothesis testing and regression!