Central Limit Theorem

By Anish Mukherjee, Ashwani Rajan, Yanan Cao

Outline

- What is the Central Limit Theorem?
- Interactive demonstration
- Application in Linear regression (hypothesis testing)

What Central Limit Theorem is?

- Given a large sample size the distribution of sample averages approaches a normal distribution irrespective of the population.

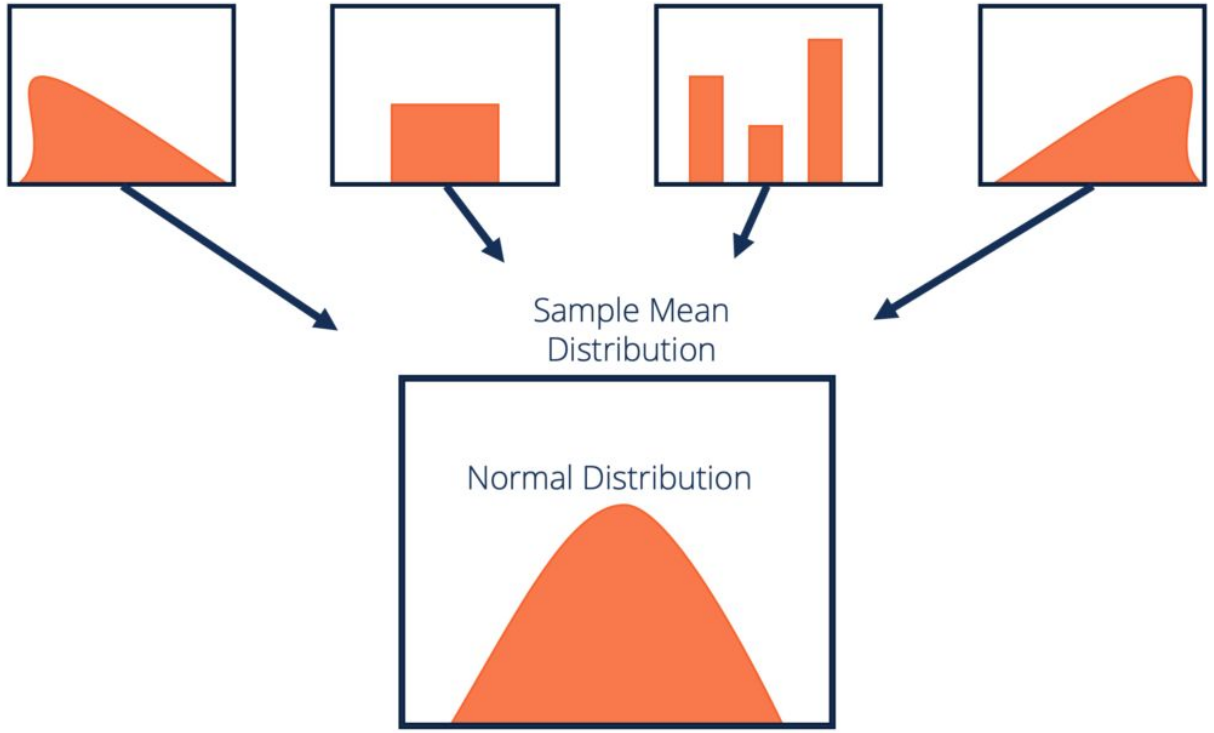
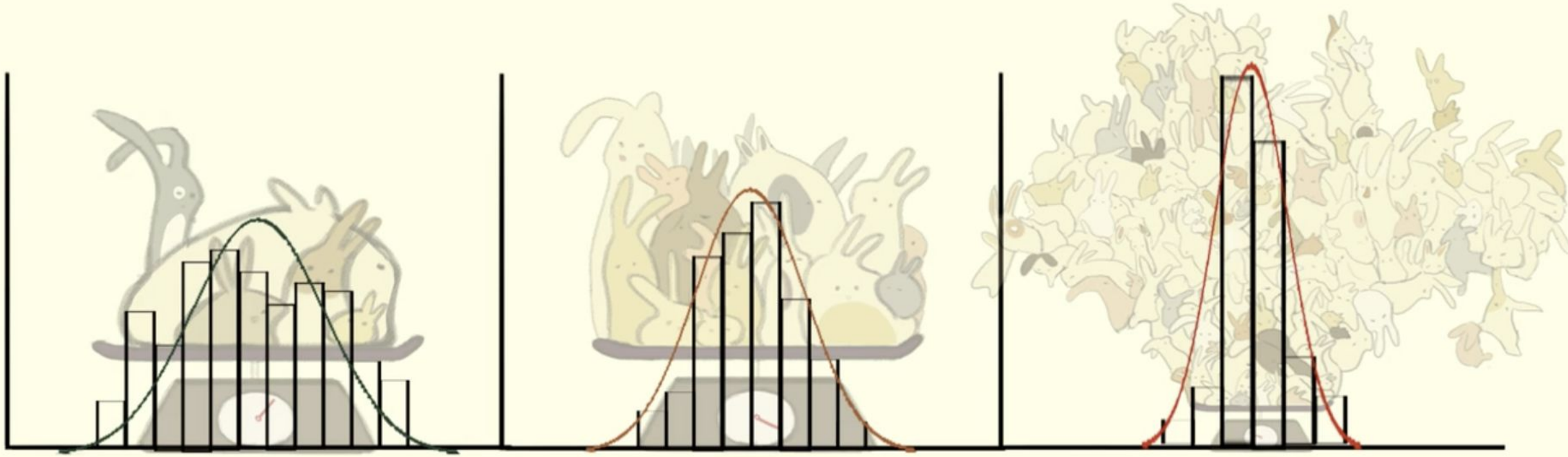


Image Credits: https://corporatefinanceinstitute.com

Central Limit Theorem



The averages of samples have approximately normal distributions

Sample size ⟶ Bigger

Distribution of Averages ⟶ More normal and narrower

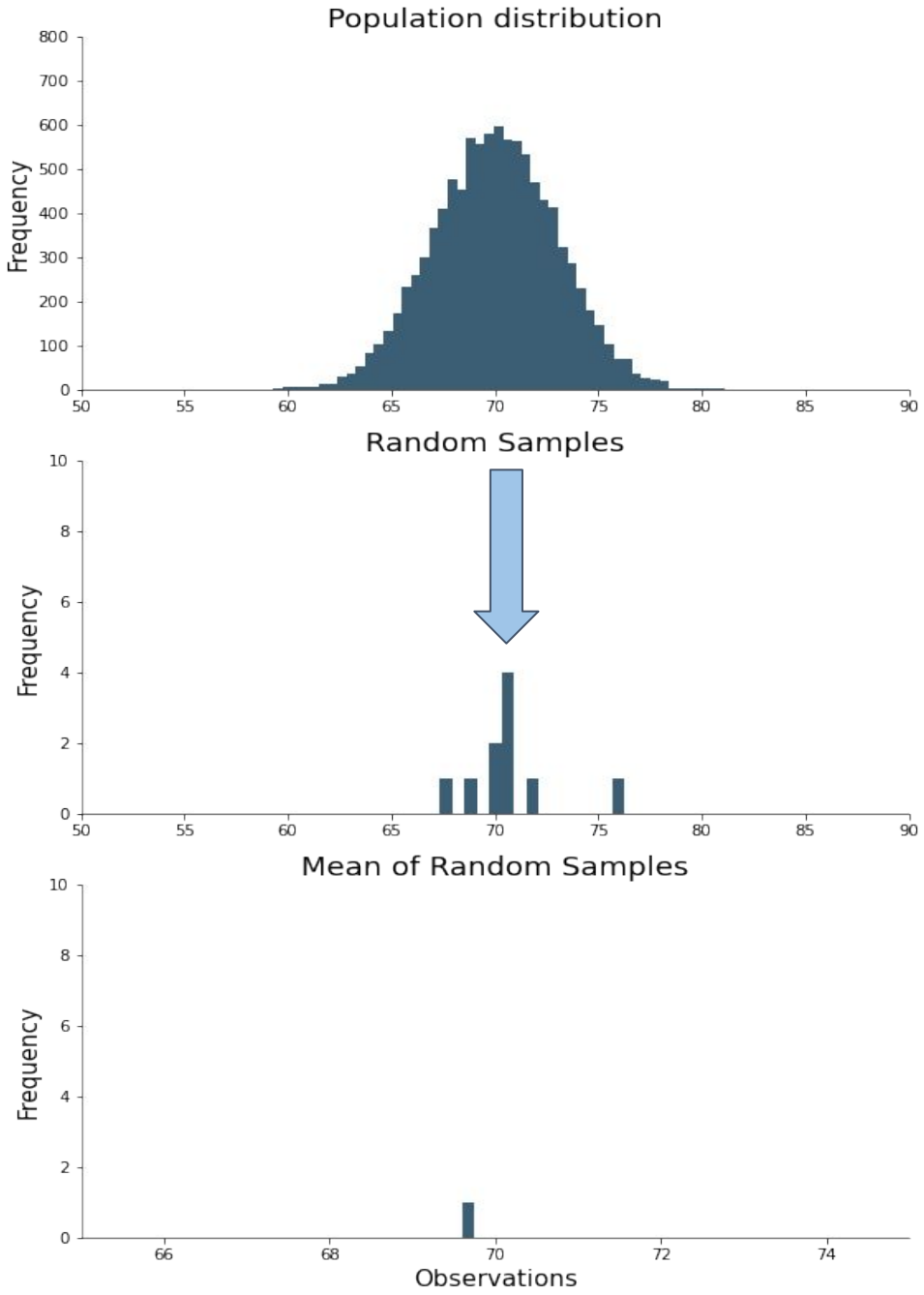
Image Credits: Casey Dunn & Creature Cast on Vimeo

Central Limit Theorem conditions

- The parent distribution should have a finite mean and variance
- The individual observations are randomly selected and the sample is representative of the population
- The sample size should be at least 30 and less than 10 % of the population

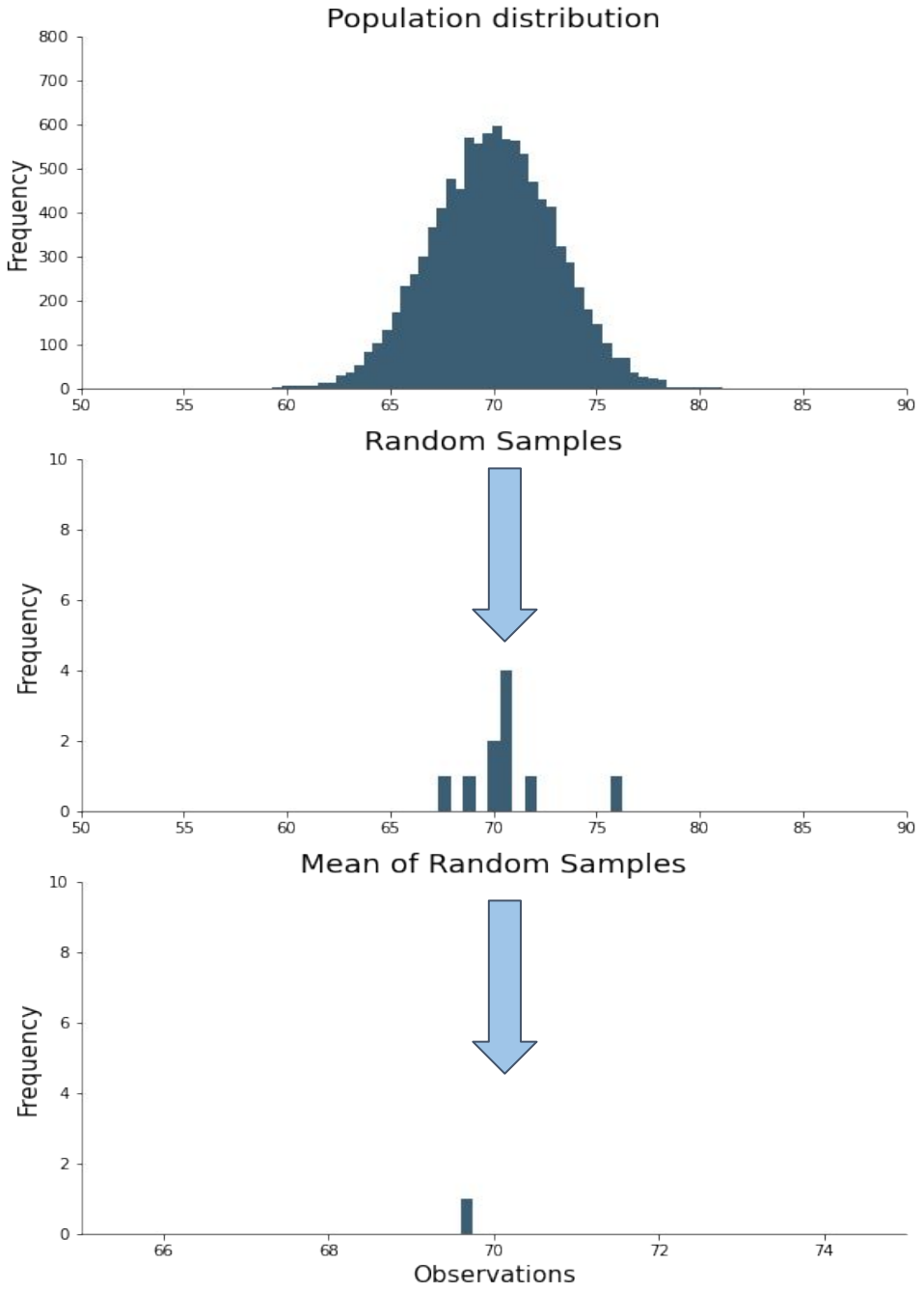
Iterative Process

- Height of all people in the US: Normal Distribution!
- Samples: Randomly select N data points from the population
- Calculate the average of the points in a sample
- Repeat the process a few thousand times



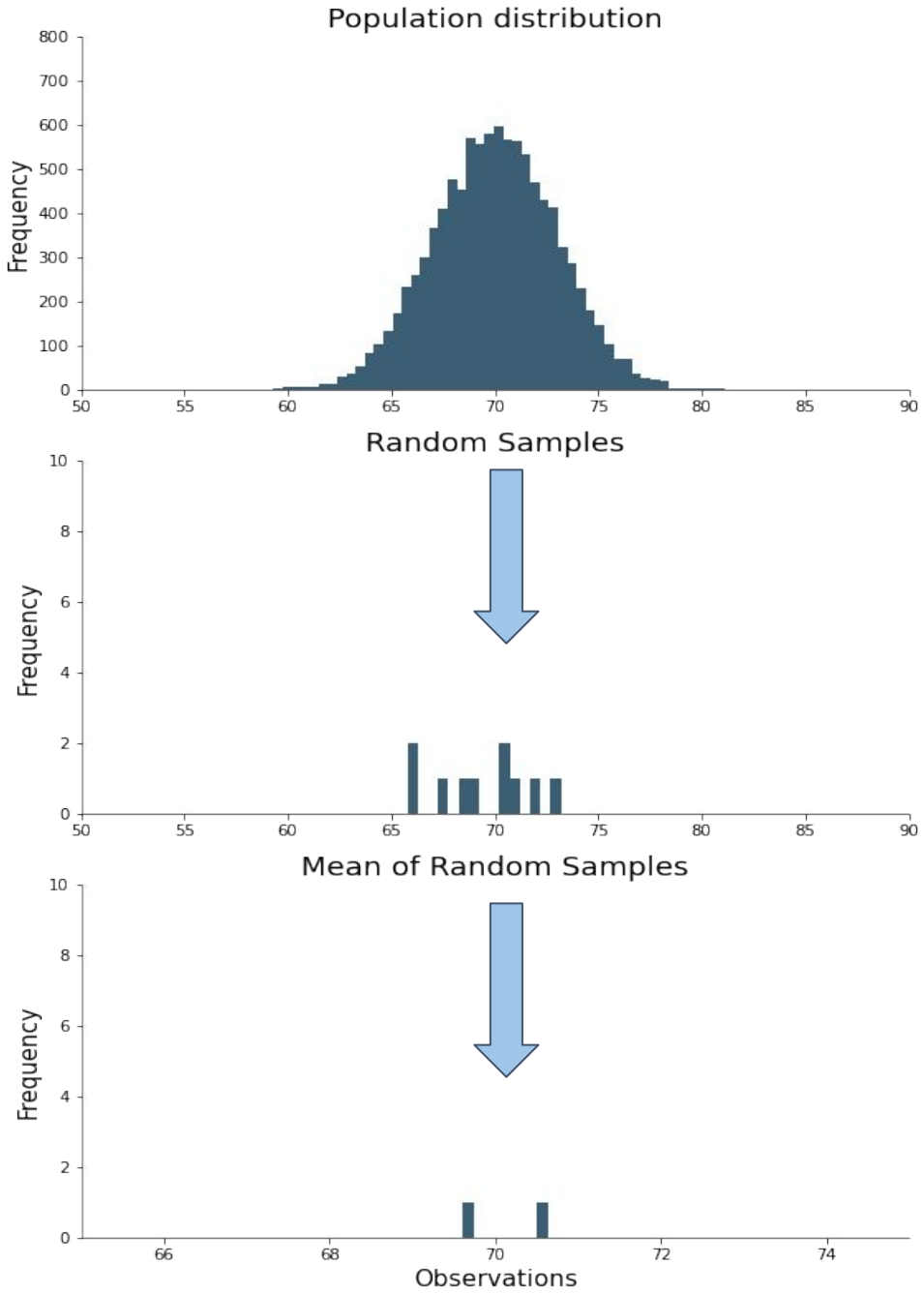
Iterative Process

- Height of all people in the US: Normal Distribution!
- Samples: Randomly select N data points from the population
- Calculate the average of the points in a sample
- Repeat the process a few thousand times



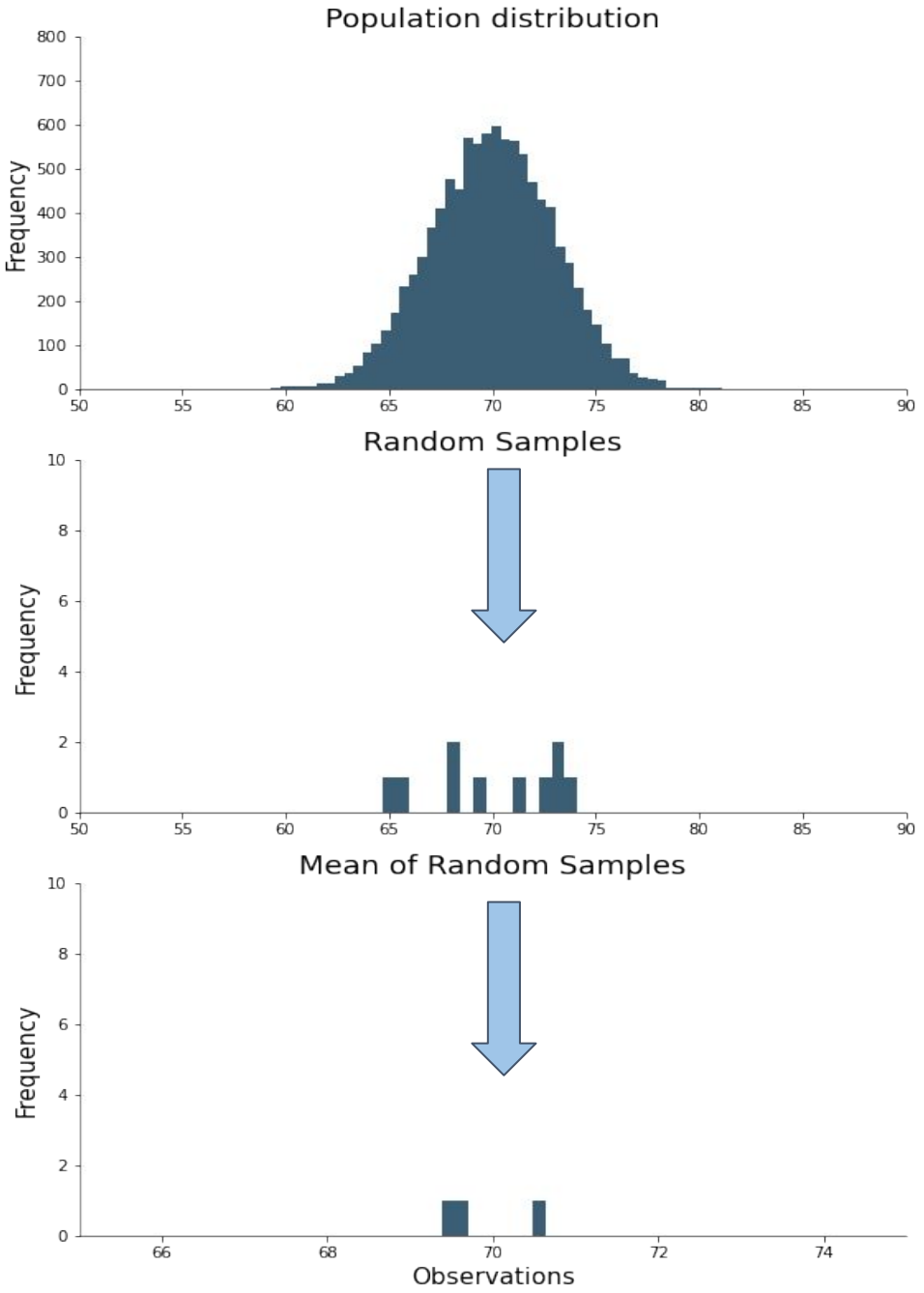
Iterative Process

- Height of all people in the US: Normal Distribution!
- Samples: Randomly select N data points from the population
- Calculate the average of the points in a sample
- Repeat the process a few thousand times



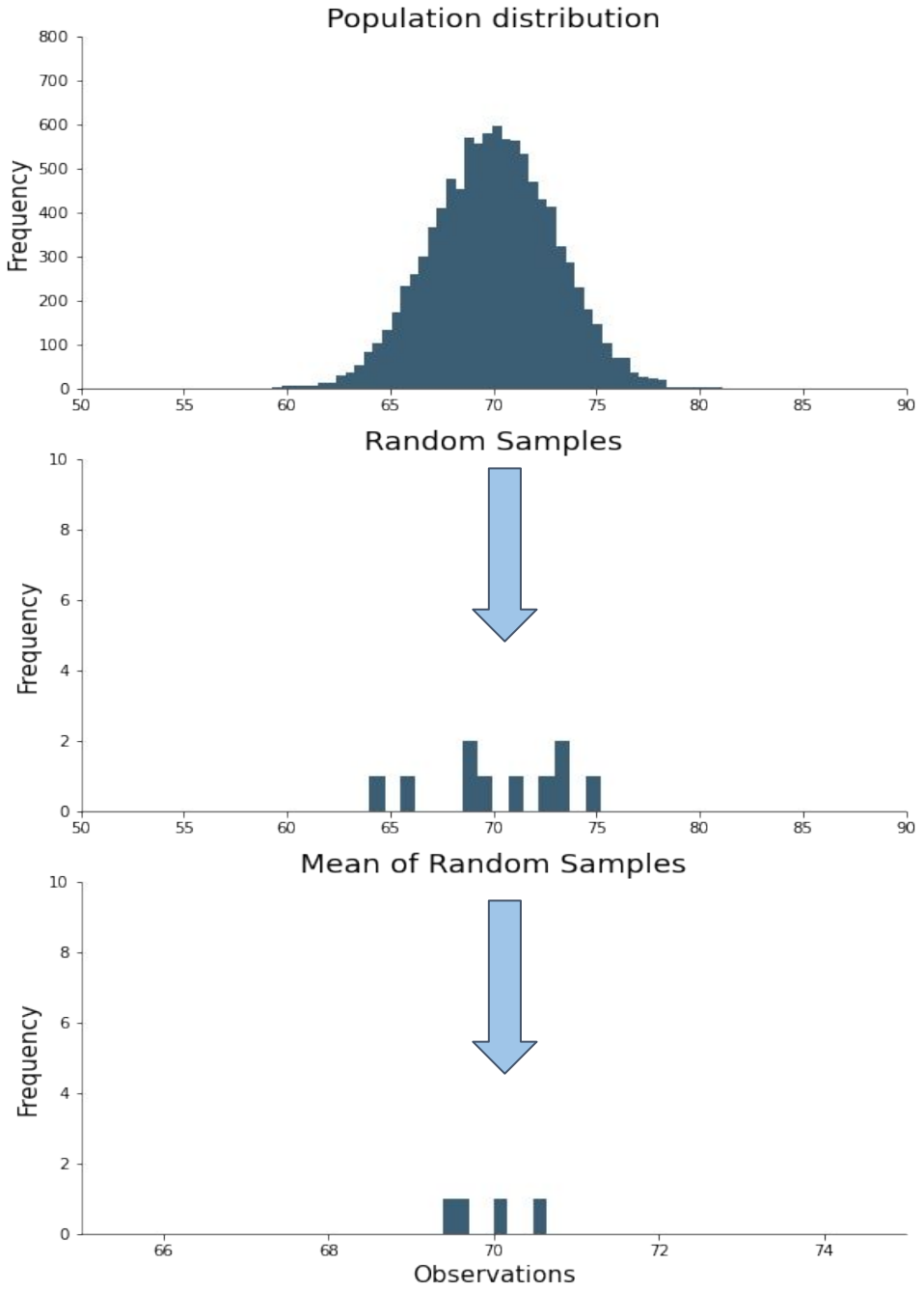
Iterative Process

- Height of all people in the US: Normal Distribution!
- Samples: Randomly select N data points from the population
- Calculate the average of the points in a sample
- Repeat the process a few thousand times



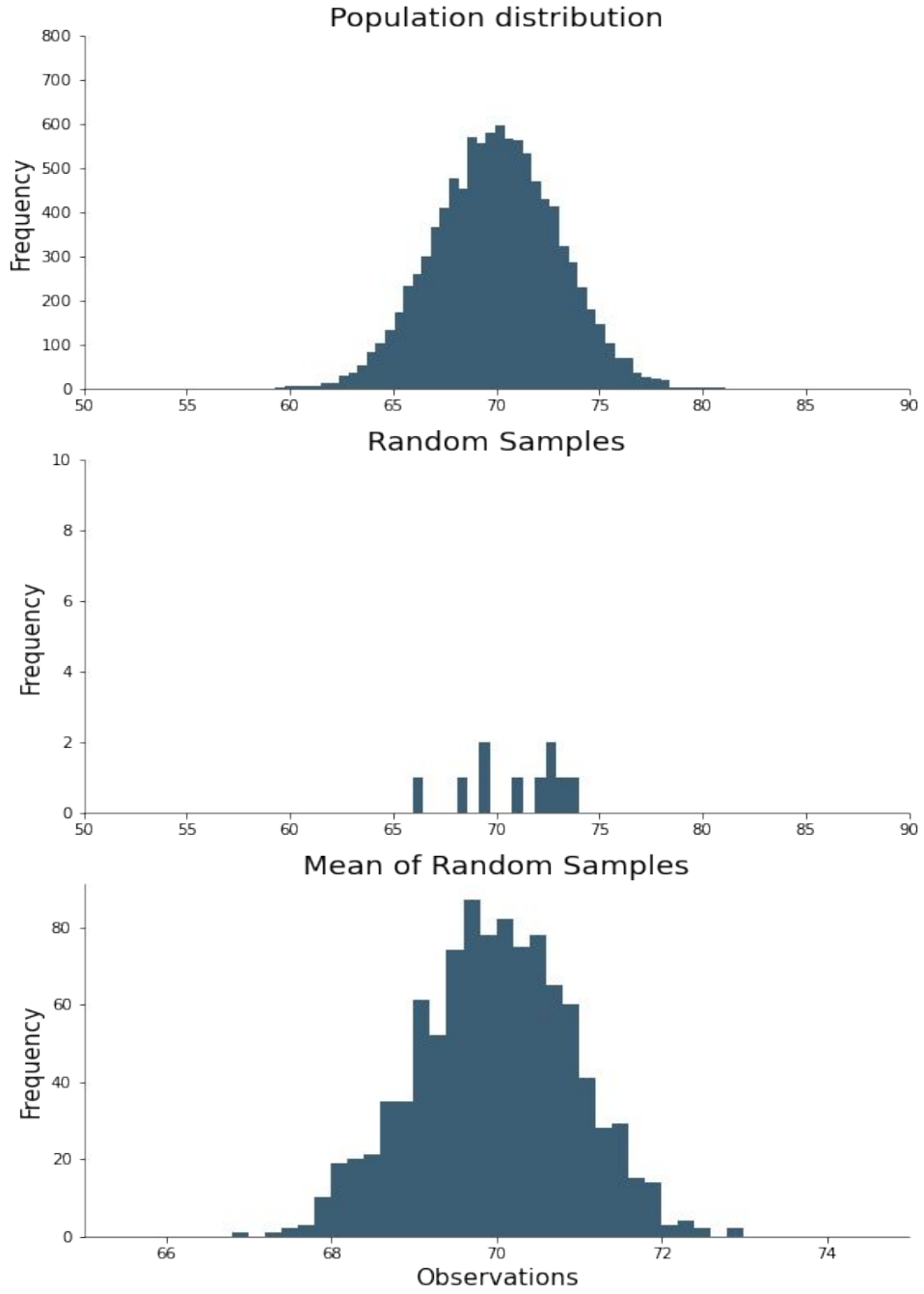
Iterative Process

- Height of all people in the US: Normal Distribution!
- Samples: Randomly select N data points from the population
- Calculate the average of the points in a sample
- Repeat the process a few thousand times



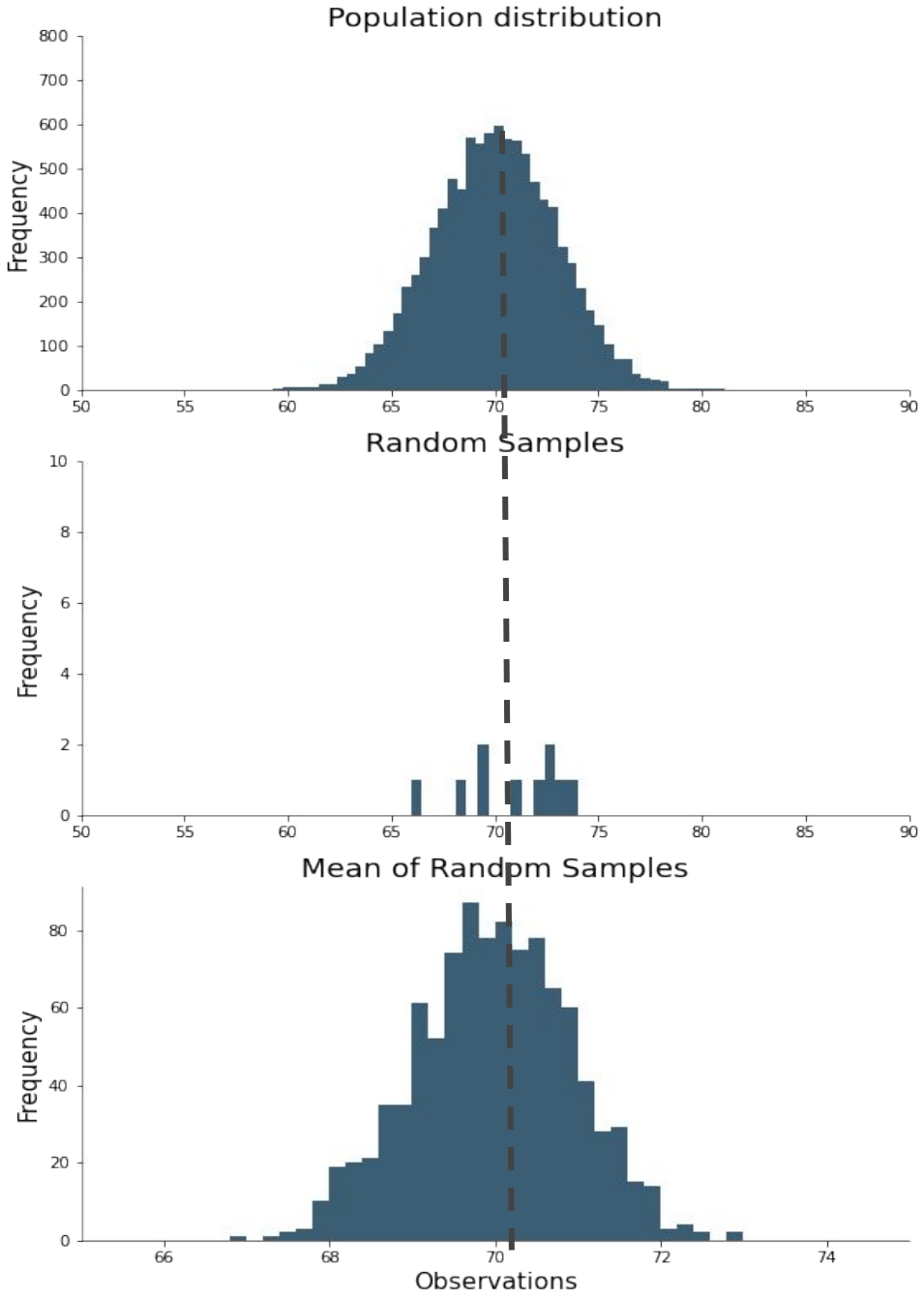
Iterative Process

- Height of all people in the US: Normal Distribution!
- After 1000 iterations: Sample means are normally distributed.
  - E[Sample Means] = Population Mean
  - Sampling distribution is narrower: Lower variance



Iterative Process

- Height of all people in the US: Normal Distribution!
- After 1000 iterations: Sample means are normally distributed.
  - E[Sample Means] = Population Mean
  - Sampling distribution is narrower: Lower variance



Interactive Demonstration

Further application

"Only extreme departure of distribution of Y from normality yield suspicious results."

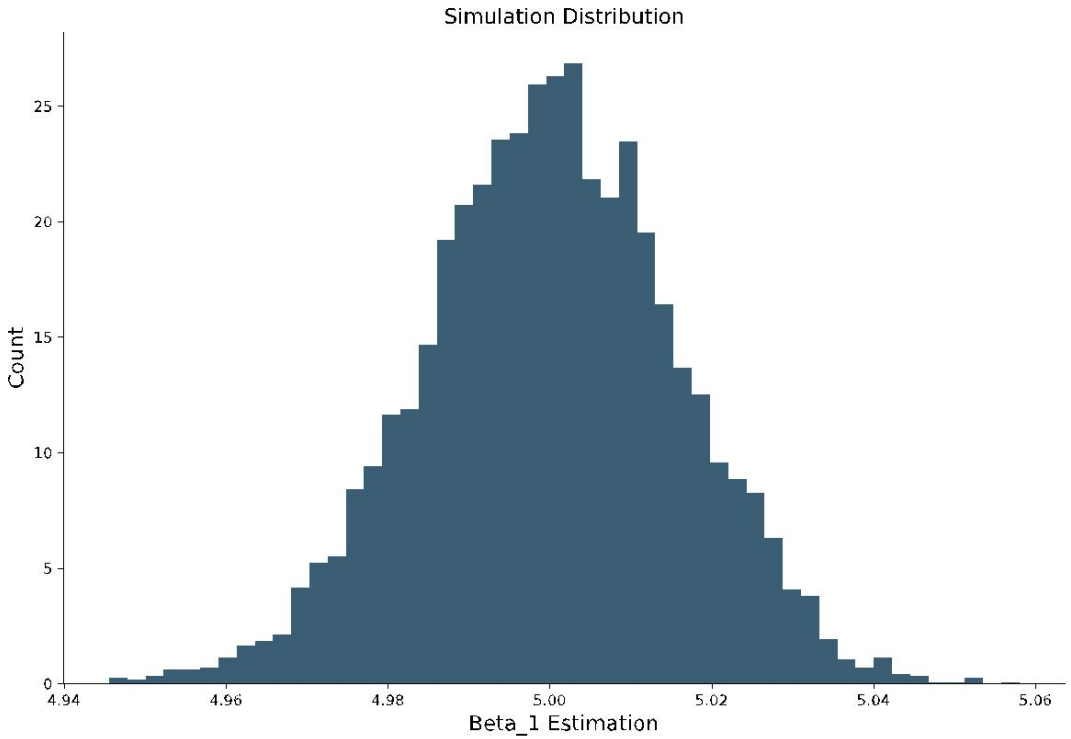
— Dr. David Kleinbaum

Further application

A simulation:

$$y = 3 + 5 \cdot x + \varepsilon$$

- Where ε is independent and identically exponentially distributed.
- Steps:
  - Generate a sample set with 100 observations and do regression
  - Generate 5000 sample sets
  - Do histogram of 5000 $\widehat{\beta_1}$

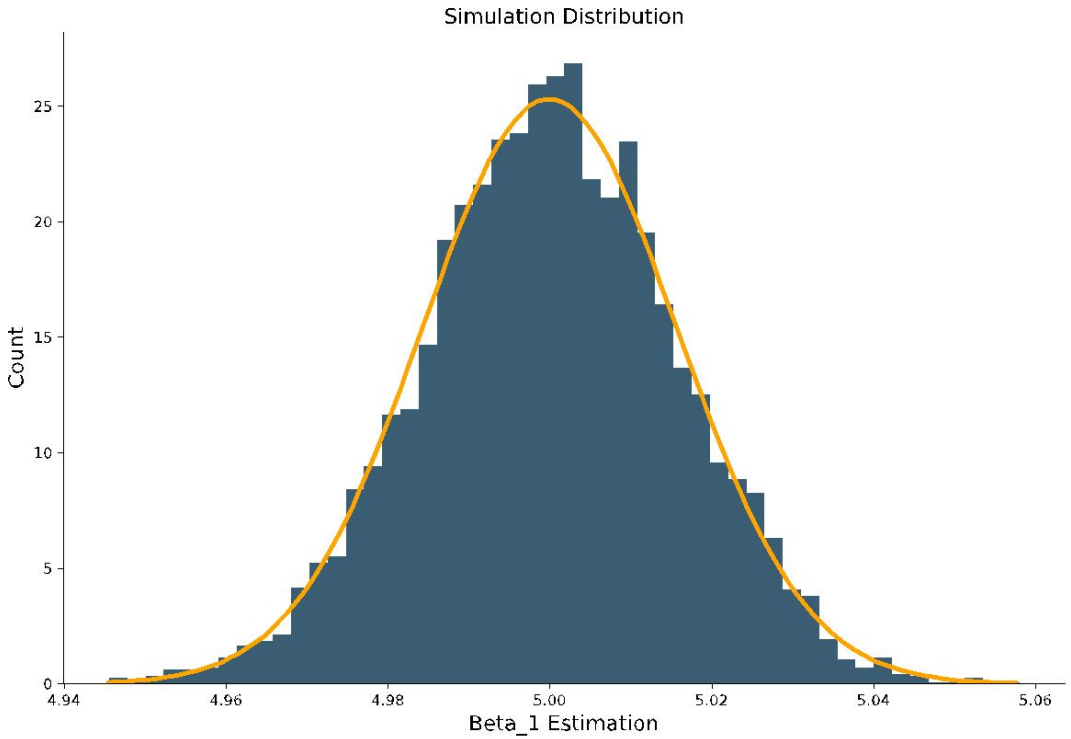


Further application

A simulation:

$$y = 3 + 5 \cdot x + \varepsilon$$

- Where ε is independent and identically exponentially distributed.
- Steps:
  - Generate a sample set with 100 observations and do regression
  - Generate 5000 sample sets
  - Do histogram of 5000 $\widehat{\beta_1}$

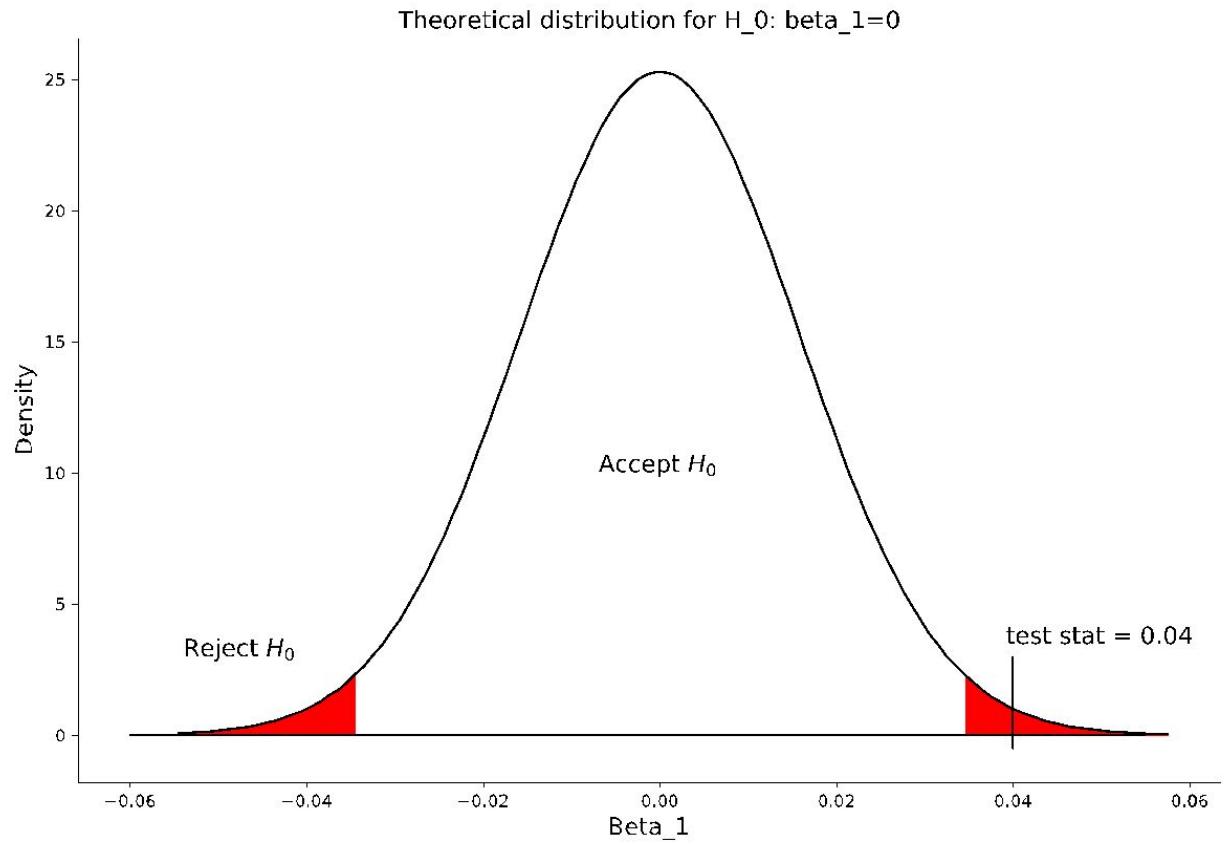


Further application

Hypothesis test with known variance:

- H_0: Beta_1 = 0
- H_1: Beta_1 ≠ 0

- $t_{0.975} = 0.0335$
- test statistics = 0.04 > 0.0335
- Reject H_0 and accept H_1 at a 5% level of significance



Overview

- What is Central Limit Theorem?
  - Parent distribution with finite mean and variance
  - Randomly selected samples
  - Large enough sample size at least larger than 30
- Interactive demonstration
- Application in hypothesis test

Q&A

Thanks for listening!