TUT206 Oct 04

Recap: bootstrapping

sample = **np.random.choice**(original_data, size=len(original_data), **replace=True)**

Bootstrapping is a resampling technique used to estimate the sampling distribution of a statistic by taking repeated samples with replacement from the original sample.

Recap: sample mean vs mean

Aspect	Population Mean (Mean)	Sample Mean
Denotation	μ	$\bar{x}$
What it Represents	The average of **all members** of a population	The average of a **subset** (sample) of the population
Nature	A **fixed** value	A **variable** value that depends on the sample chosen
Purpose	Describes the **true central value** of the population	**Estimates** the population mean (μ)

Recap: sample mean vs mean

Population Mean	Sample Mean
$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$
N = number of items in the population	n = number of items in the sample

Recap: variance and standard deviation

Variance and Standard Deviation Formula

	Population	Sample
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
Standard Deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$	$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$

```
sd = np.std(original_data, ddof=1)
```

Recap: Standard deviation vs standard error

Aspect	Standard Deviation (SD)	Standard Error (SE)
What it measures	The variability of **individual data points** in a sample/population	The variability of a **sample statistic** (e.g., sample mean)
Used for	Describing the spread of a dataset	Describing the accuracy of a sample statistic as an estimate of a population parameter
Formula	Measures the deviation of data points from the mean	Measures the deviation of sample means from the population mean
Effect of sample size	Unaffected by sample size	Decreases as sample size increases

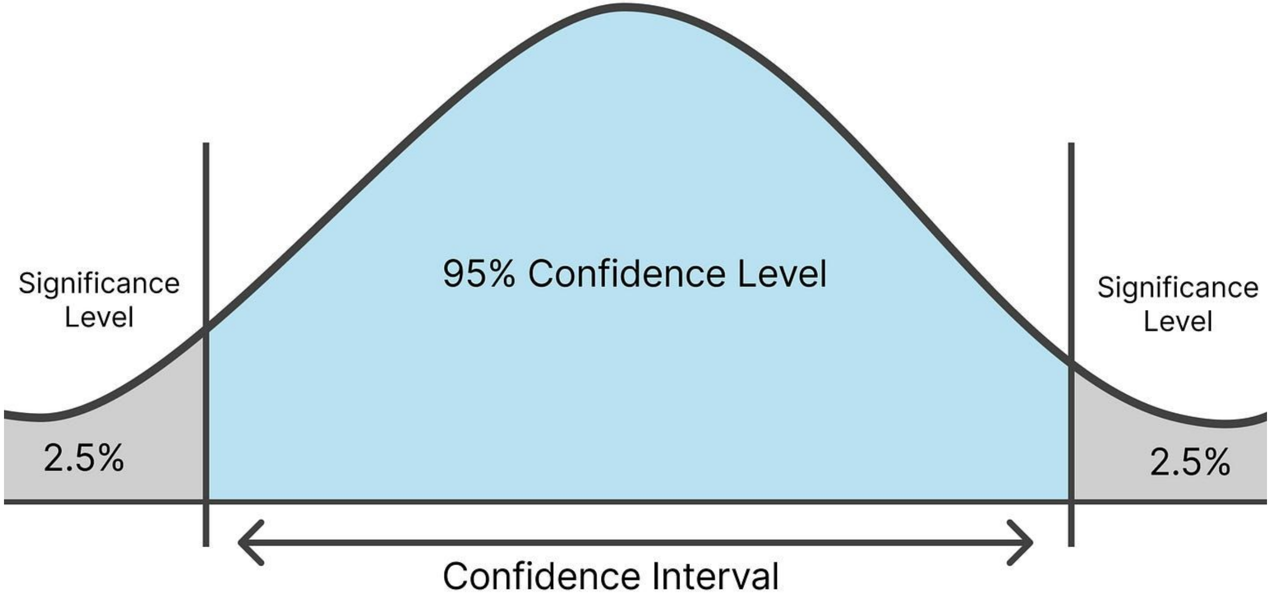
Recap: Standard deviation vs standard error

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$SE = \frac{SD}{\sqrt{n}}$$

sd = np.std(original_data, ddof=1)

Recap: confidence interval



A **confidence interval** gives us a range of values within which we are confident the true population parameter lies. For example, a **95% confidence interval** means that we are 95% confident that the interval contains the population mean.

Recap: confidence interval

```
lower_bound, upper_bound = np.percentile(bootstrapped_means, [2.5, 97.5])
```

$$CI = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

CI = confidence interval

$\bar{x}$ = sample mean

z = confidence level value

s = sample standard deviation

n = sample size

Why is "Single Sample" in Quotes in the TUT Title?

The phrase **"single sample"** is in quotes because, although we are using only **one sample**, the idea of generating a **sampling distribution** (e.g., via bootstrapping) involves simulating many repeated resamples to understand the variability of our sample statistic.

Demo I: Introducing Formal Hypothesis Testing

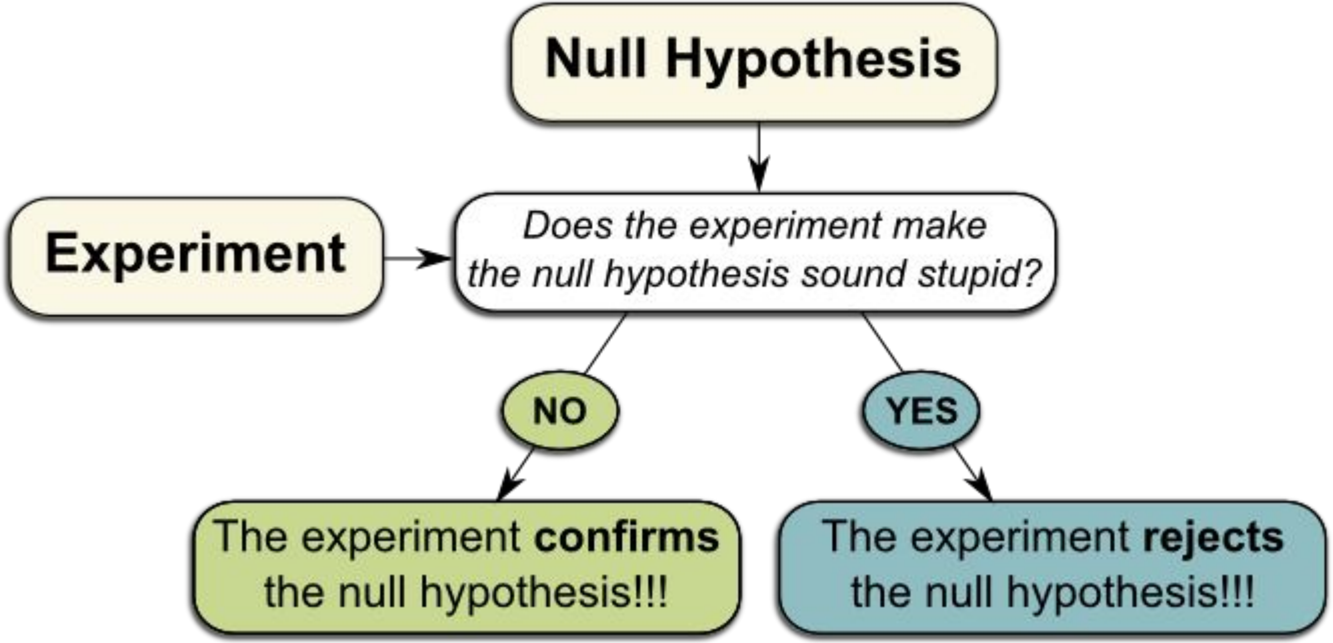
- **Null Hypothesis** (H_0): Represents the **current belief** or "status quo". It is what we assume to be true unless there is sufficient evidence to prove otherwise.

X bar

  - Example: $H_0 : \mu = \mu_0$, where μ_0 is the hypothesized value of the population mean.
- **Alternative Hypothesis** (H_1): Represents the **new claim** or alternative to the null hypothesis.
  - Example: $H_1 : \mu \neq \mu_0$ (Two-tailed test) or $H_1 : \mu > \mu_0$ (One-tailed test).

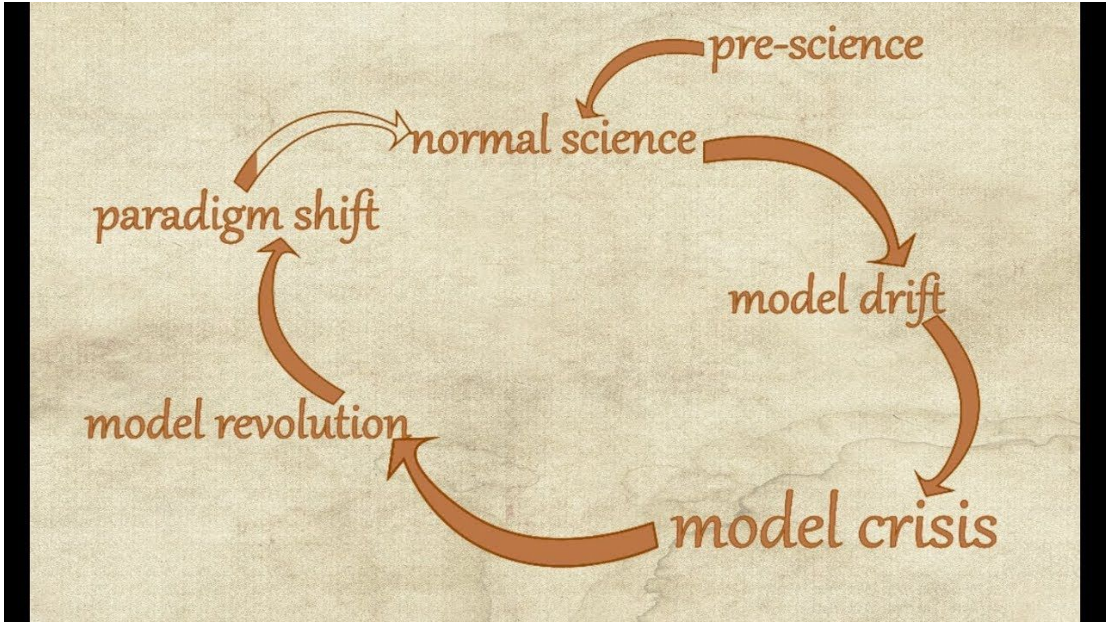
Keep in mind, we can either “reject null hypothesis” or “doesn’t reject null hypothesis”

Demo I: Introducing Formal Hypothesis Testing



Demo I: Introducing Formal Hypothesis Testing

Thomas Kuhn's model of scientific progress, commonly known as the **Kuhnian Cycle**. This model describes how science evolves over time through different stages



https://colab.research.google.com/drive/1JDlyGaVOKthKzl6Xk7TlpbfDNq70JZp
4?usp=sharing

Demo I: Introducing Formal Hypothesis Testing

- **InitialHealthScore**: The health score of the patient before receiving the vaccine or treatment.
- **FinalHealthScore**: The health score of the patient after receiving the vaccine or treatment.
- **HealthScoreChange**: Calculated as `FinalHealthScore - InitialHealthScore`, this value indicates whether the patient's health has **improved, remained the same, or worsened** after the treatment.

Demo I: Introducing Formal Hypothesis Testing

The Null Hypothesis [and Alternative Hypothesis]

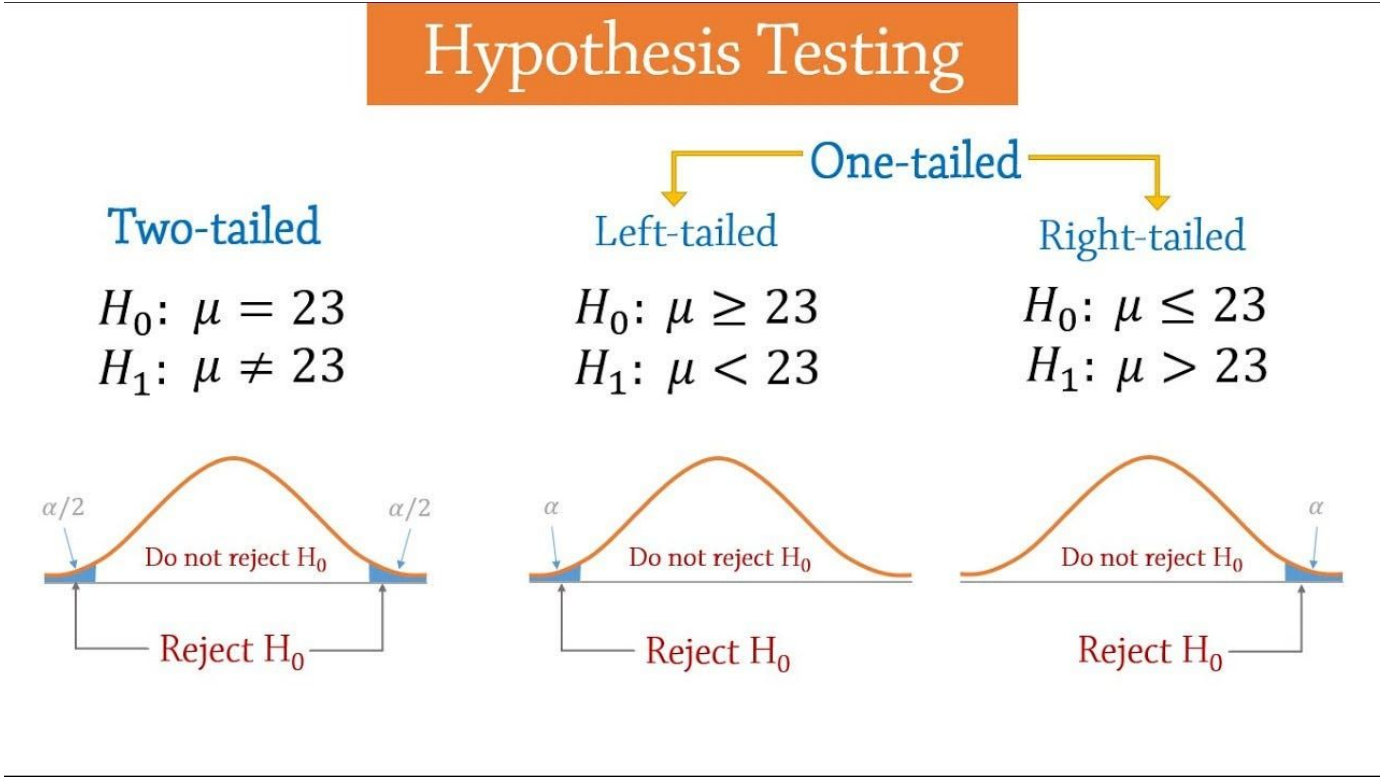
The **null hypothesis** usually simply states the "no effect" (on average) assumption

H_0 : The vaccine has no effect **(on average)** on patient health$H_1 : H_0$ is false

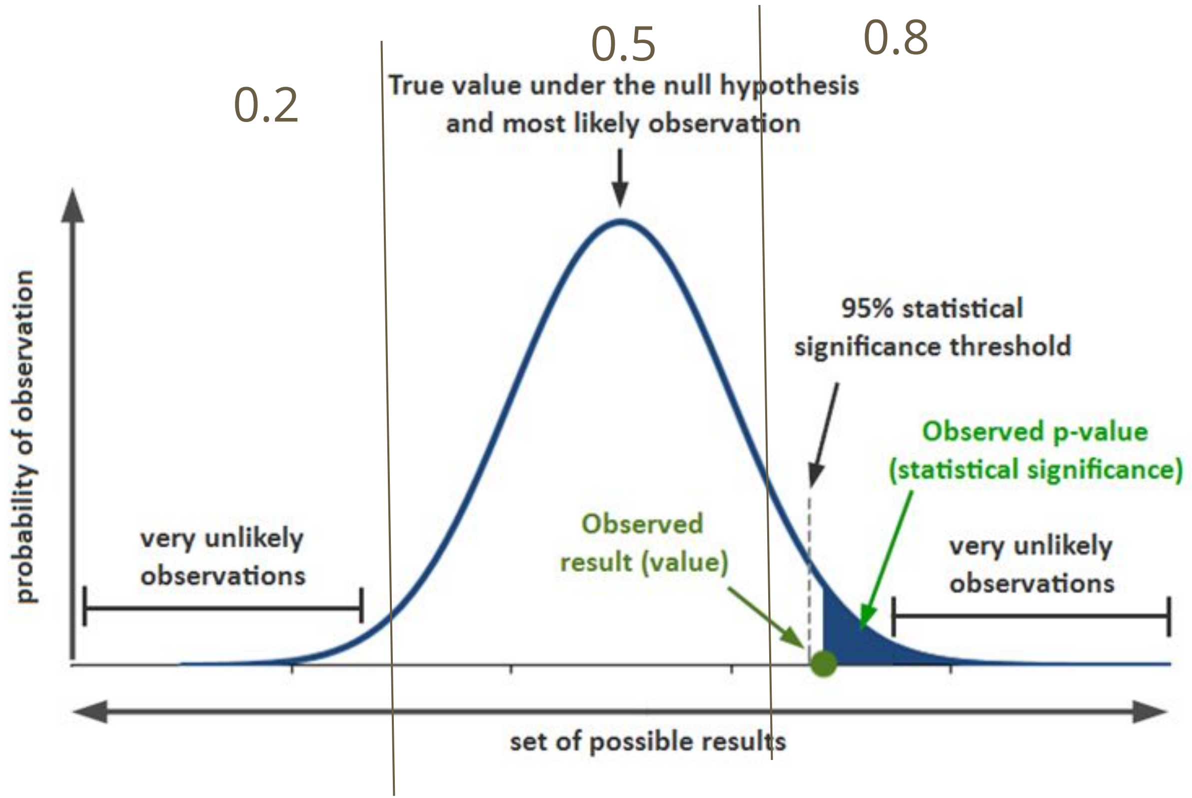
To empasize that "**(on average)**" refers to the pupulation parameter μ (the average effect), it is helpful to more formally (and concisely) express this equivalently as

$$H_0 : \mu = 0 \quad \text{ and } \quad H_A : H_0 \text{ is false}$$

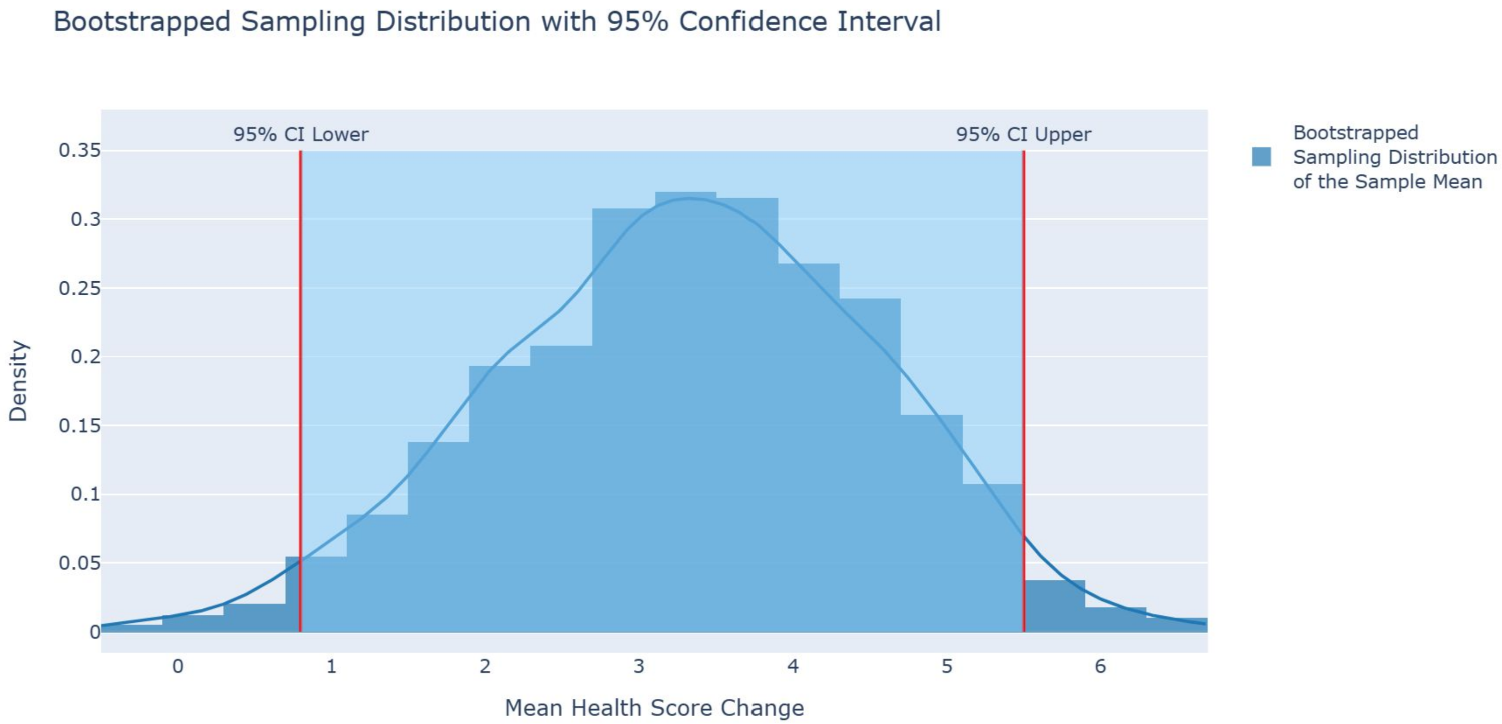
Demo I: Introducing Formal Hypothesis Testing



Demo I: Introducing Formal Hypothesis Testing



Demo I: Introducing Formal Hypothesis Testing



Recap