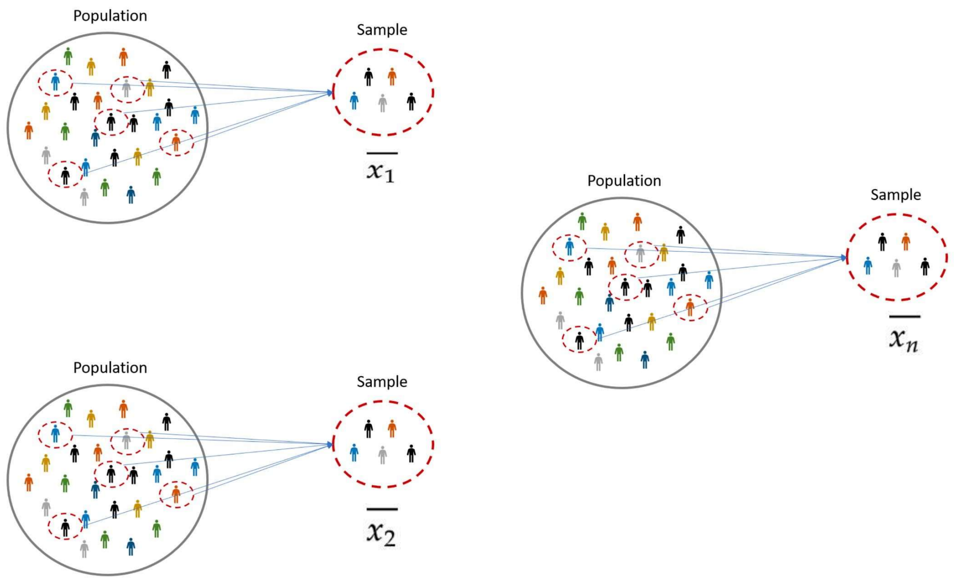
DATA | MODULE 2

Hypothesis Testifig

Week 5

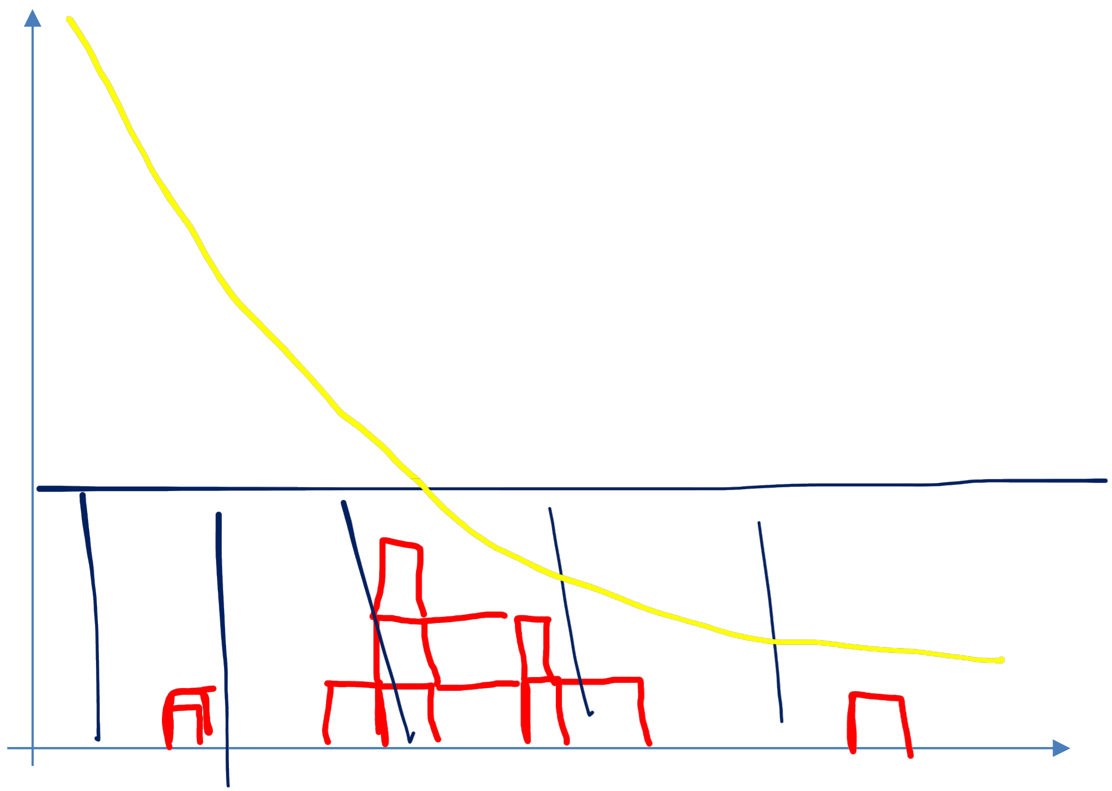
Table of Contents

Central Limit Theorem



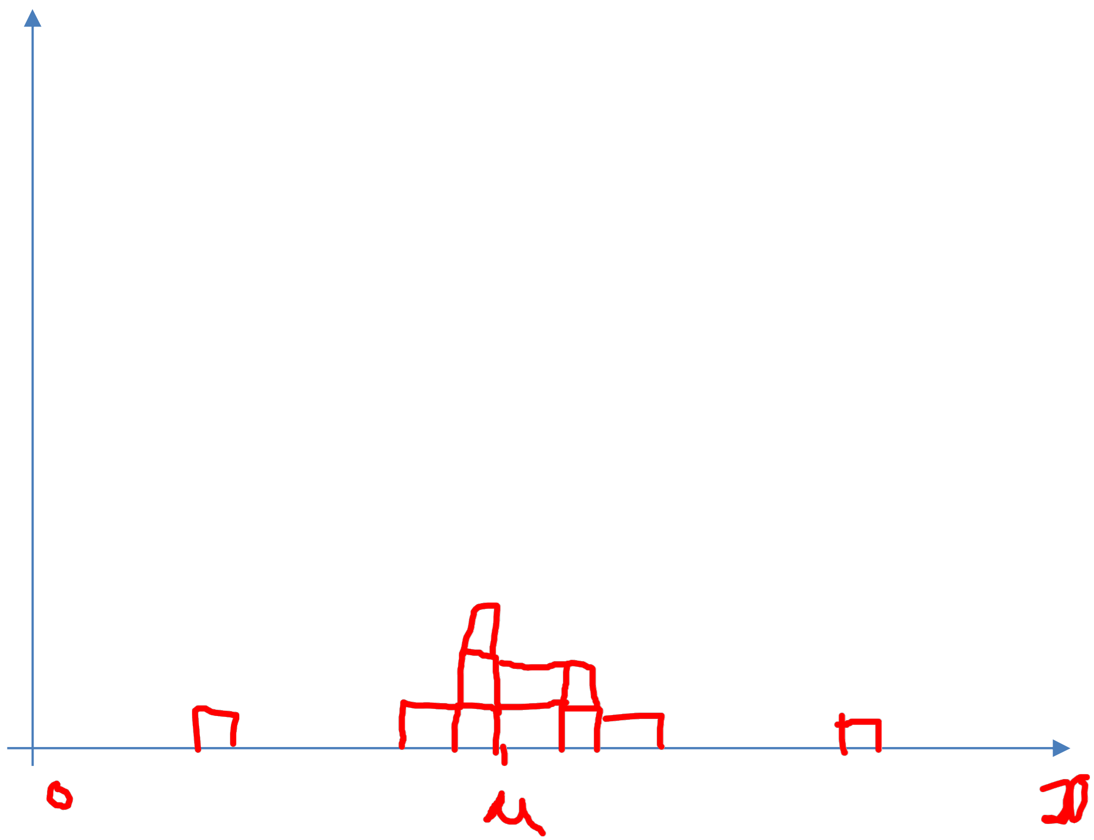
Imagine that you take samples from a population of people's heights, and for each sample you calculate their mean.

Central Limit Theorem



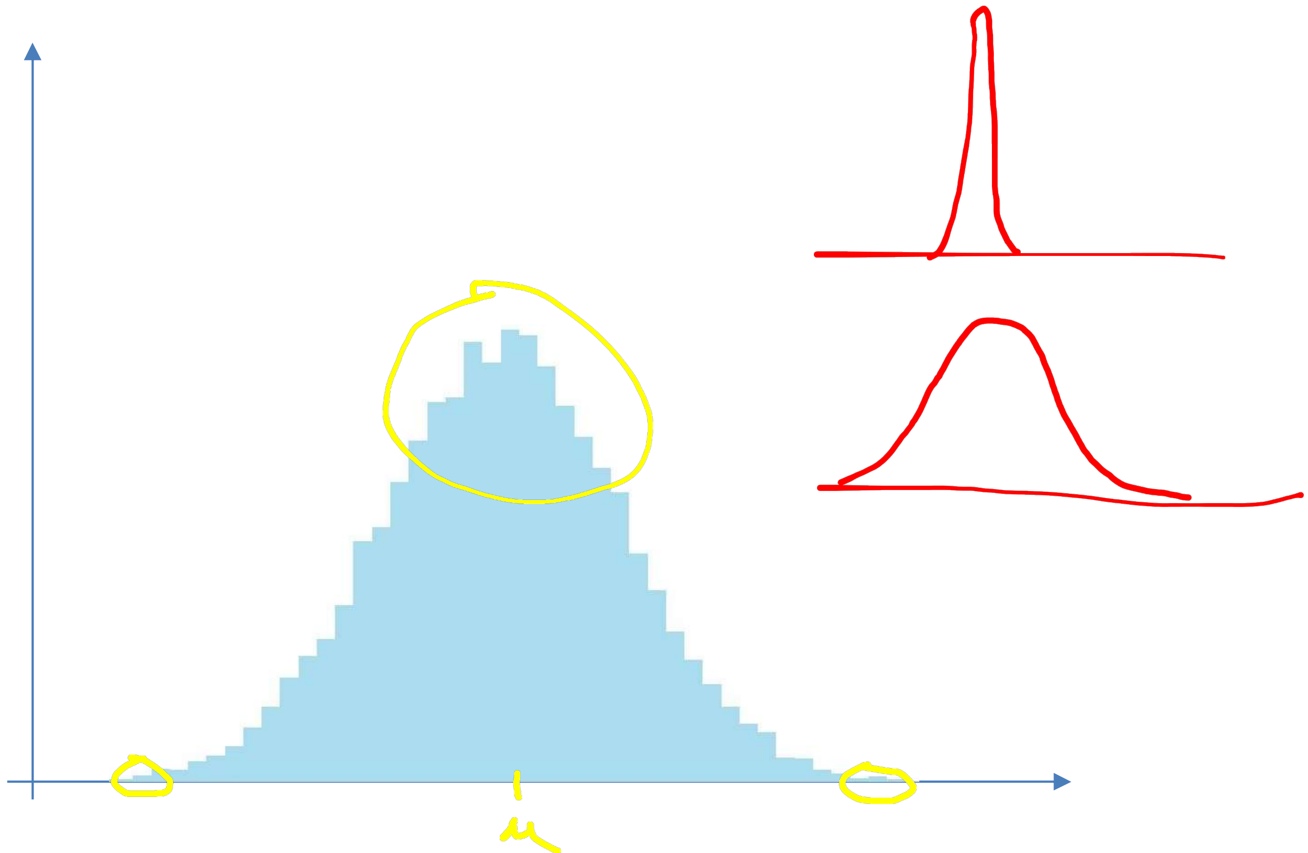
μ1 = 10
μ1 = 12
μ1 = 9
μ1 = 10
μ1 = 6
μ1 = 12
μ1 = 10
μ1 = 11
μ1 = 11
μ1 = 13
μ1 = 19

Central Limit Theorem



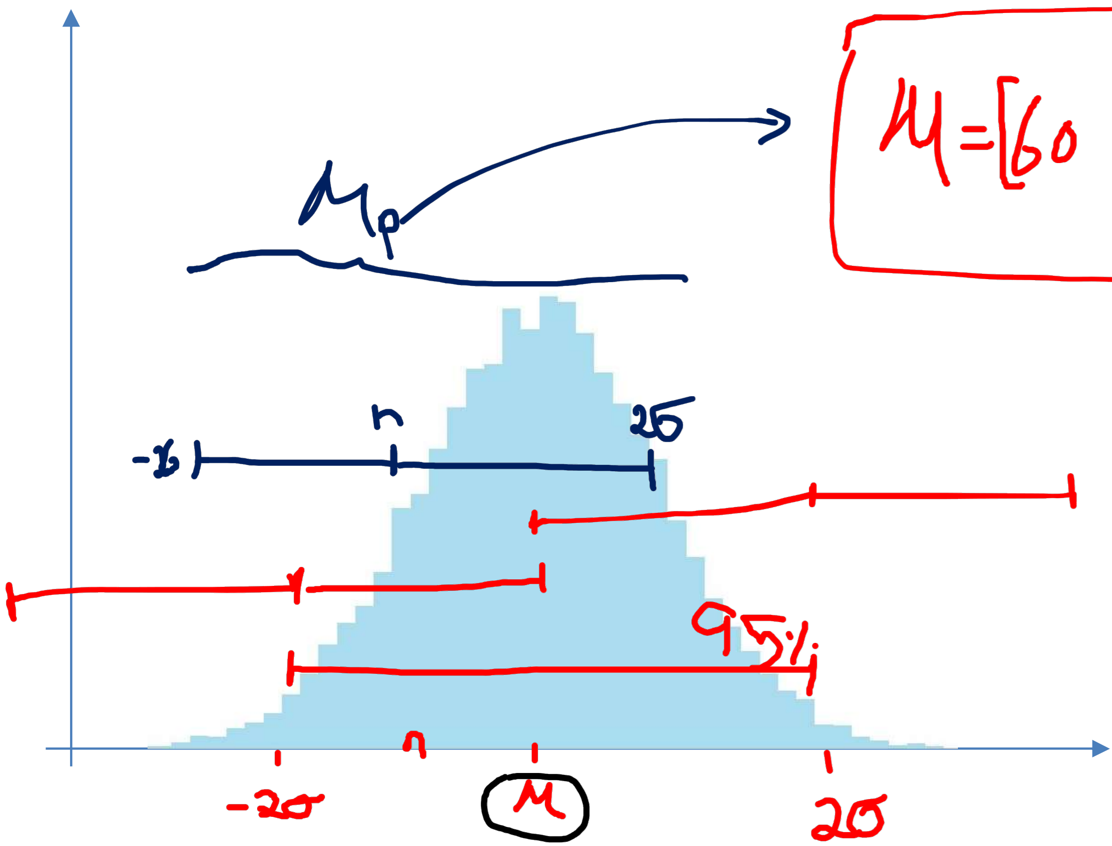
μ1 = 10
μ1 = 12
μ1 = 9
μ1 = 10
μ1 = 6
μ1 = 12
μ1 = 10
μ1 = 11
μ1 = 11
μ1 = 13
μ1 = 19

Central Limit Theorem



μ1 = 10
μ1 = 12
μ1 = 9
μ1 = 10
...
...
μ1 = 13
μ1 = 18
μ1 = 19
μ1 = 5
μ1 = 20

Central Limit Theorem



$\boxed{\mu = [60 : 84]}$

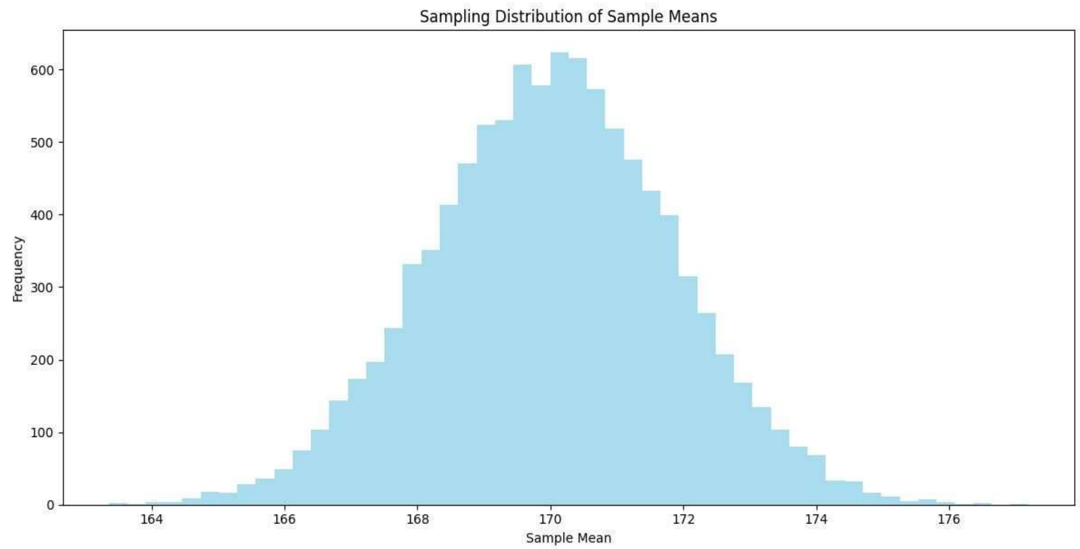
μ1 = 10
μ1 = 12
μ1 = 9
μ1 = 10
...
...
μ1 = 13
μ1 = 18
μ1 = 19
μ1 = 5
μ1 = 20

$\sigma = 6kg$

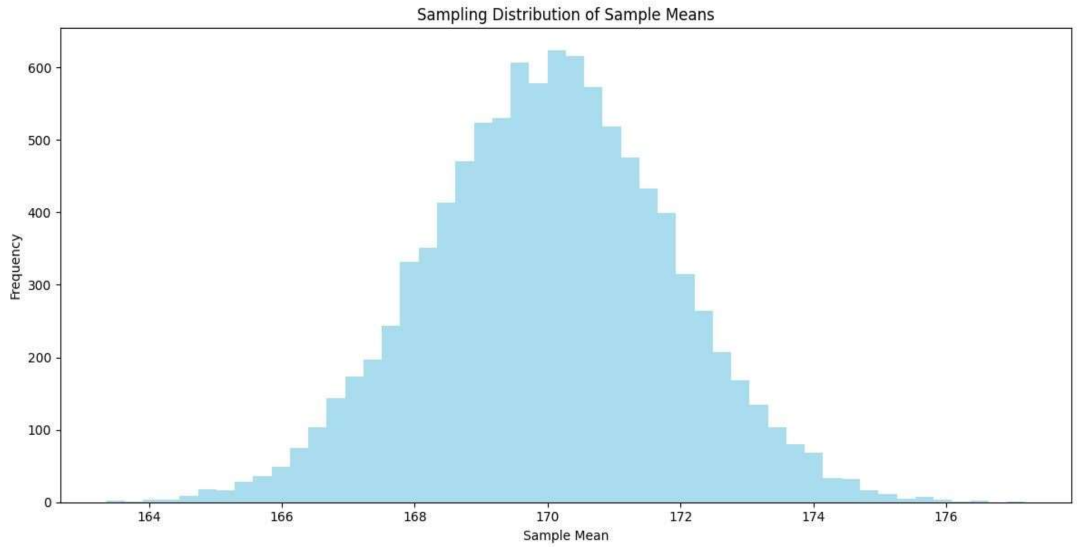
$\mu_s = 72$

Central Limit Theorem

- Now, if we plot all **sample means**, we get a familiar shape, a normal distribution. In other words, the distribution of the means of the sampling distribution tends towards a normal distribution.



Central Limit Theorem



- The sampling distribution will consistently tend toward a normal distribution, having its own mean and standard deviation. An essential takeaway is **that the mean of the sampling distribution equals the population mean.** This is happens regardless of the shape of the population distribution.

Central Limit Theorem

The Central Limit Theorem (CLT) states that, irrespective of the shape of the underlying population, the sampling distribution of the mean will approximate a normal distribution as the sample size grows larger ($n > 30$), assuming all samples are identical in size and randomly sampled.

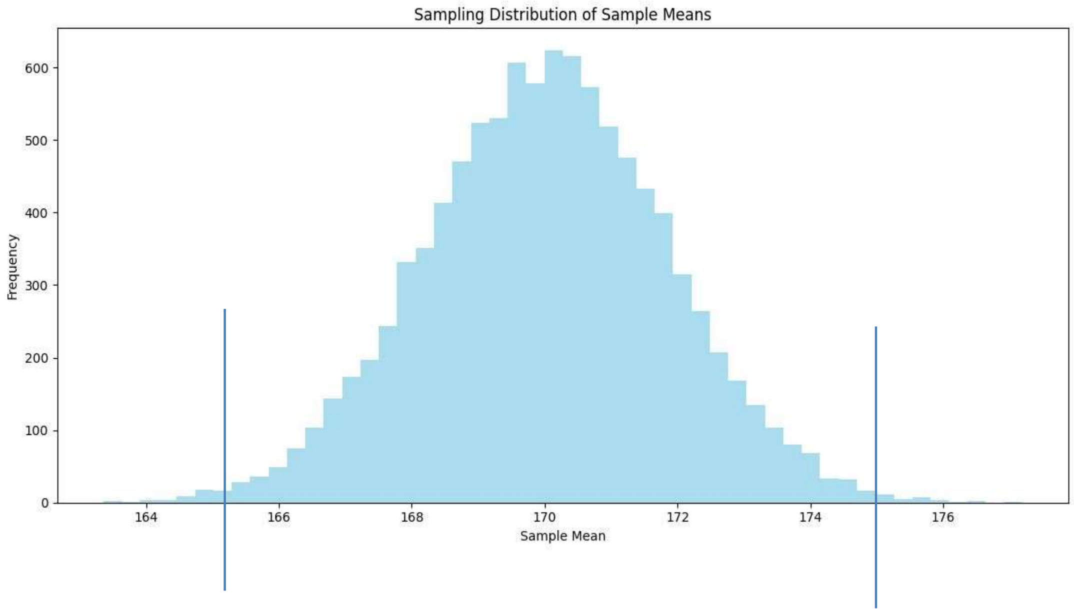
- Large Sample Size & Individual Data Points: Even with a large sample, the distribution of individual data points could still be non-normal. For instance, a dataset with millions of data points could exhibit significant skewness or extreme kurtosis.

- Large Sample Size & Averages of Samples: If you are taking multiple samples from a population and calculating their averages, the distribution of those averages tends to be normal due to the CLT, even if the underlying population is not normal.

- Practical Implications: While the CLT is powerful, it's important to remember that many statistical tests and methods assume that the individual data points (not their means) follow a normal distribution. Therefore, having a large dataset doesn't allow you to bypass these assumptions.

Hypothesis Testing

Hypothesis Testing



- What if we assume that being inside the lines would be good enough to consider something “valid”?

Hypothesis Testing

- The first step in hypothesis testing is formulating a hypothesis, which is a premise or claim that we want to test about a population parameter. The objective of a hypothesis test is to decide, based on a sample from the population, which of two complementary hypotheses is true:
    - **Null Hypothesis** H_0 - Currently "*accepted*" value for a parameter.
    - **Alternative Hypothesis** H_1- The alternative hypothesis is a statement asserting an alternative condition or outcome compared to the null hypothesis.

Hypothesis Testing

- In a hypothesis testing problem, after observing the sample the experimenter must decide either to accept H_0 as true or to reject H_0.
  - <u>Let's see the following example:</u>
    - If μ denotes a population parameter, let's say the change in a patient's blood pressure after taking a drug, we write:

$$H_0 = 0 \quad \text{versus} \quad H_1 \neq 0$$

    - The null hypothesis states that, on average, the drug has no effect on blood pressure, and the alternative hypothesis states that there is some effect.

Hypothesis Testing

Some definitions:

- Statistical hypothesis testing operates on the principle that while establishing universal truth is challenging, it is possible to demonstrate falsity through the presentation of counterexamples.

- The null hypothesis should assert the **absence of an effect** and explicitly state **equalities** ($=$, $\geq$ or $\leq$).

Hypothesis Testing

Some definitions:

- **Statistical Hypothesis Testing:** The practice of making inferences about population parameters by disproving a null hypothesis using sample data.

- **Null Hypothesis (H_0):** A statistical assertion indicating no effect or relationship, represented by equalities ($=$, $\geq$ or $\leq$).

Hypothesis Testing

Step by step:

After formulating the Null and Alternative Hypothesis we must:

1. Choose significance level
2. Collect Data
3. Calculate Test Statistic
4. Determine P-value
5. Decision-making

Hypothesis Testing: Step by Step

Step 1: Pick significance level

- The significance level, denoted by α, is the probability of incorrectly rejecting the null hypothesis when it is actually true. In other words, how likely is our test findings to yield a wrong conclusion.
    - Common choices for α are 1%, 5% or 10%. This will heavily depend on the nature of the problem studied.

Hypothesis Testing: Step by Step

Step 1: Choose significance level

- The significance level, denoted by α, is the probability of **incorrectly rejecting the null hypothesis when it is, in fact, true**. It represents how likely our test is to **produce a false positive result**.
    - Common choices for α include 1%, 5%, or 10%, and the selection **depends on the specific context of the study**.

Hypothesis Testing: Step by Step

Step 3: Calculate Test Statistic

- The test statistic used to assess the population mean depends on the specific type of test being conducted. In general, when checking for a population mean (μ), the common formula for the test statistic is

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

sample mean: $\overline{X}$ — hypothesized population mean: μ

Std of the population: σ — sample size: n

The test statistic quantifies **how many standard deviations the sample mean is from the hypothesized population mean**, aiding in the determination of whether to reject the null hypothesis based on the observed sample data.

Hypothesis Testing: Step by Step

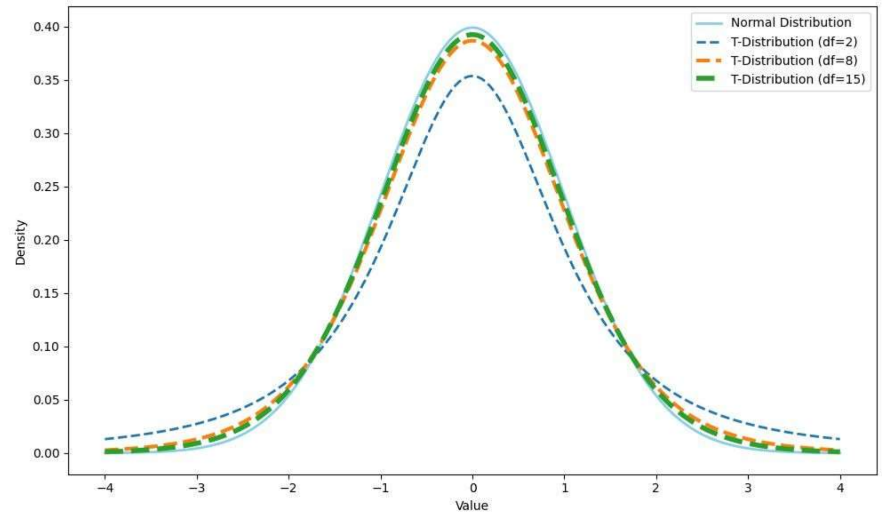
Step 3: Calculate Test Statistic

- When σ **is known**, the appropriate test statistic is the z-score. This statistic conforms to a normal distribution.

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

- When σ **is unknown or the number of samples is <30**, the appropriate test statistic is the t-statistic. This statistic conforms to a t-distribution.

$$\frac{\overline{X} - \mu}{s / \sqrt{n}}$$



Hypothesis Testing

Step 4: Determine P-value

- Having obtained a test statistic and considered the distribution of our samples (whether normal or T), we can now evaluate the probability of getting a sample mean as extreme as the one we observed. This probability is known as the p-value.

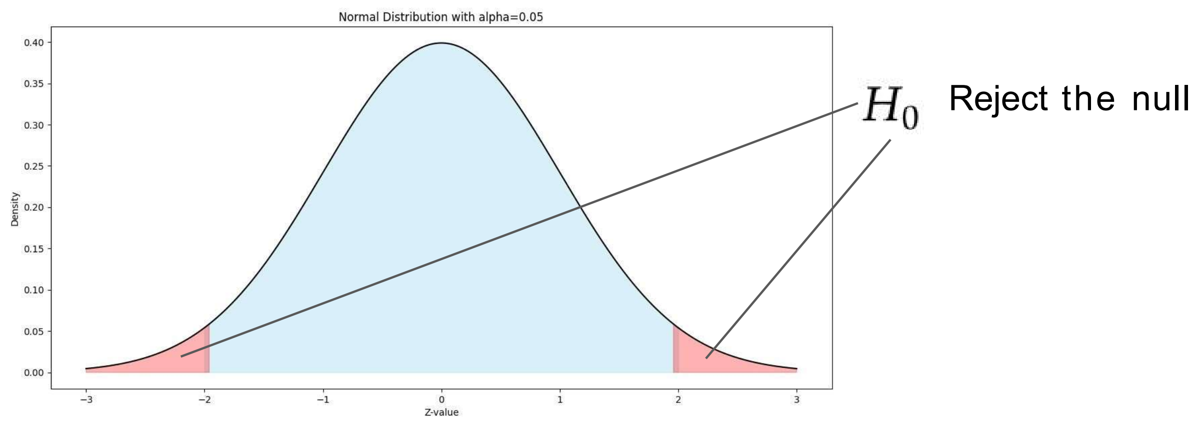
- If the p-value is low, it means that the likelihood of obtaining the observed results, or more extreme results, under H_0 is unlikely. This often leads to rejecting the null hypothesis in favor of an alternative hypothesis.

Hypothesis Testing

Step 5: Decision-making

- Finally, we need to decide whether to reject H_0 or fail to reject it.
  - In **two-tailed tests** (when we test for equalities):
    - If p_value < α we reject the null hypothesis.
    - If p_value > α we fail to reject the null hypothesis.

$$H_0 : k = \mu$$
$$H_1 : k \neq \mu$$



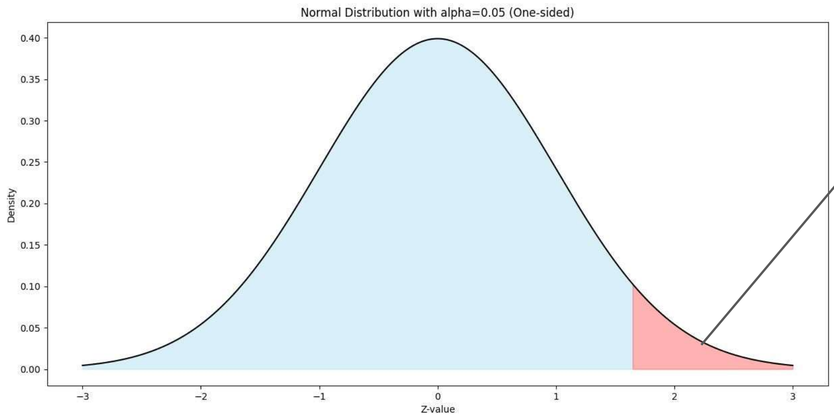
Hypothesis Testing

Step 5: Decision-making

- In **one-tailed tests** (when we test for inequalities):
  - If p_value < α **and** the test statistic is in the same direction as H_1 we reject the null hypothesis.
  - If p_value > α **or** the test statistic is in opposite direction as H_1 we fail to reject the null hypothesis.

H_0: $\mu \leq k$

H_1: $\mu > k$



H_0 Reject the null

Hypothesis Testing

Different Types of Tests

- One Sample T-test
  - $H_0 : k = \mu$
    $H_1 : k \neq \mu$
    in python: ttest_1samp(sample, k, alternative = “two-sided”)
  - $H_0 : k \leqslant \mu$
    $H_1 : k > \mu$
    in python: ttest_1samp(sample, k, alternative = “greater”)
  - $H_0 : k \geqq \mu$
    $H_1 : k < \mu$
    in python: ttest_1samp(sample, k, alternative = “less”)

Hypothesis Testing

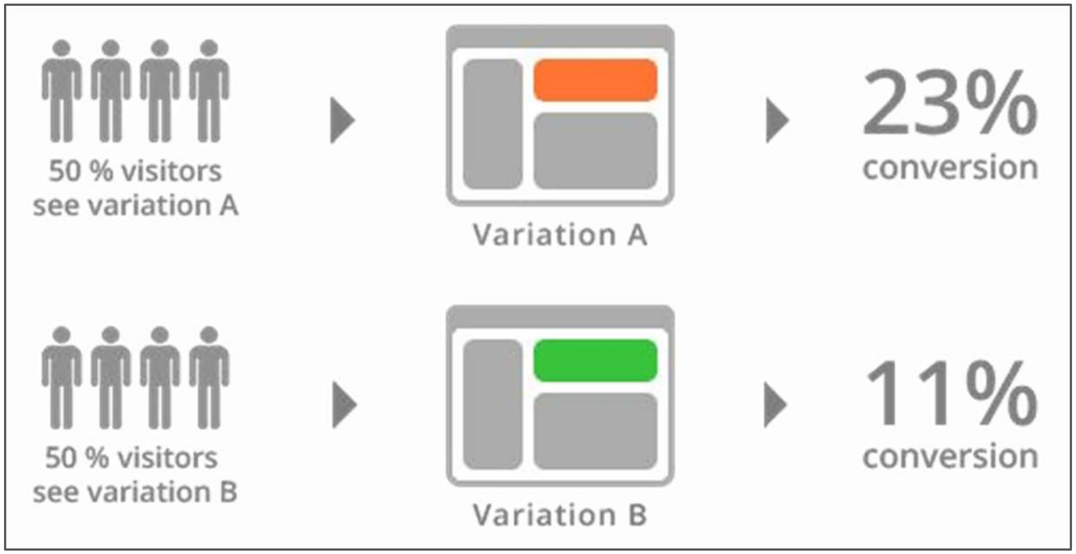
Different Types of Tests

- Two Sample T-test (compares the means of two independent samples)
  - $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$ **in python**: ttest_ind(sample1, sample2, alternative = “two-sided”)
  - $H_0: \mu_1 \leqslant \mu_2$ $H_1: \mu_1 > \mu_2$ **in python**: ttest_ind(sample1, sample2, alternative = “greater”)
  - $H_0: \mu_1 \geqq \mu_2$ $H_1: \mu_1 < \mu_2$ **in python**: ttest_ind(sample1, sample2, alternative = “less”)

Hypothesis Testing

Different Types of Tests

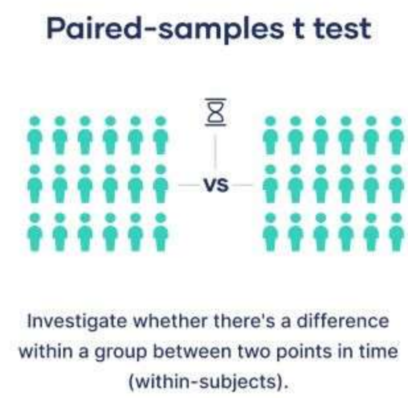
Two Sample T-test (compares the means of two independent samples)



Hypothesis Testing

Different Types of Tests

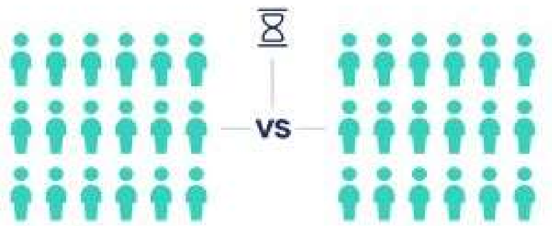
- Paired Sample T-test (compares the means of two dependent samples)
  - $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$ **in python**: ttest_rel(sample1, sample2, alternative = “two-sided”)
  - $H_0: \mu_1 \leqslant \mu_2$ $H_1: \mu_1 > \mu_2$ **in python**: ttest_rel(sample1, sample2, alternative = “greater”)
  - $H_0: \mu_1 \geqq \mu_2$ $H_1: \mu_1 < \mu_2$ **in python**: ttest_rel(sample1, sample2, alternative = “less”)



Hypothesis Testing

Different Types of Tests

Paired Sample T-test (compares the means of two dependent samples)



Investigate whether there's a difference within a group between two points in time (within-subjects)

Hypothesis Testing

Different Types of Tests

ANOVA (compares the means of multiple independent samples)

in python: f_oneway(sample_1, sample_2, sample_3, ..,sample_n)

$$H_0: \mu_1 = \mu_2 = \mu_3 = ... = \mu_n$$
$$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq ... \neq \mu_n$$

Summary

Hypothesis Testing:

- Create H0 and H1
- Pick a significance level: how confortable are we in being wrong
- Test Statics + get the pvalue
- If the pvalue is too smal, I reject H0

Lets see this in Python, while following slides?

RUNNING EXAMPLE

- Let's take our Titanic dataset. You have seen that the prices in first class were on average 85 dollars and someone told you that prices in 3rd class were usually a fifth of prices in first class. You are skeptical. Set up the hypotheses to test this.

- Now, you think the prices in third class are even cheaper than that. Set up the hypotheses to test this.

THE 5 STEPS OF HYPOTHESIS TESTING

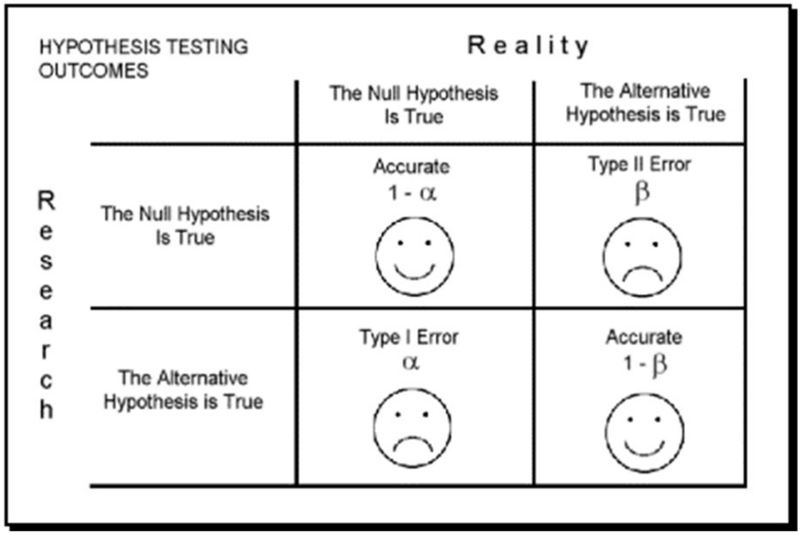
1. Set the hypothesis
2. **Choose significance / confidence level**
3. Sample
4. Compute statistic + Get p-value
5. Decide

SIGNIFICANCE AND TYPES OF ERRORS

We said that we would make our decision to reject the null when it strains credibility yo maintain it. But how do we measure this? We set a significance level **α** *à priori*.

Significance is the probability of rejecting the null if it happens to be true. This is the most damaging type of error so we should be demanding with our significance (**α** chosen at most 5%, usually)

We have seen that significance is the converse of confidence (1-**α**).



RUNNING EXAMPLE

- Let’s pick a significance level. Are you feeling confident about our claim or not?

THE 5 STEPS OF HYPOTHESIS TESTING

1. Set the hypothesis
2. Choose significance / confidence level
3. **Sample**
4. Compute statistic + Get p-value
5. Decide

RUNNING EXAMPLE

- Open the class Jupyter.
- Let's assume we have access to only 30 3rd class passengers (say, to get the price paid for the ticket you had to track down their families to send you a copy of the receipt).
- Sample your dataset for 30 entries

THE 5 STEPS OF HYPOTHESIS TESTING

1. Set the hypothesis
2. Choose significance / confidence level
3. Sample
4. **Compute statistic + Get p-value**
5. Decide

COMPUTE YOUR TEST STATISTIC

Your test statistic depends on what kind of test you are trying to make.

For each test, there is a formula that takes the circumstances of your observation and returns a number that is as high as it is unlikely that your observation came from H_0.

In the case of trying to infer the average of a population the relevant circumstances to consider are:

- The observed mean
- How much the samples vary from each other
- The number of observations sampled

GET THE P-VALUE

We said the statistic is higher the more unlikely it is for your sample to be sampled under H_0, but how can we quantify that?

We use the statistic to check **how likely is it to get a sample mean as extreme as the one we actually sampled.**

This quantity is called the **p-value**

If the p-value is very small, it means that it is extremely unlikely, under H_0, to get a result like the one we did in our sample and thus we should reject H_0, because staying with it strains credibility.

We compute the p-value from a table and the value of our statistic or, because it's 2020, from a scipy call.

one-tailed

```
st.ttest_1samp(sample,H0)
```

two-tailed

```
st.ttest_1samp(sample,
H0,alternative='less')
```

RUNNING EXAMPLE

- Get the p-value for your test statistic
- Watch out, the way to compute p-values for each of our hypotheses is actually different

THE 5 STEPS OF HYPOTHESIS TESTING

1. Set the hypothesis
2. Choose significance / confidence level
3. Sample
4. Compute statistic + Get p-value
5. **Decide**

DECISION CRITERIA

We now compare our obtained p-value (chance to see an observation at least as extreme as the one we saw) with our significance level .

- If $p<\alpha$, we have just witnessed an event that, if H_0 is true, happens less than a fraction α of the times. This strains credibility and we therefore reject H_0
- If $p\geq\alpha$, we have witnessed an event that, if H_0 is true, happens more than a fraction α of the times. This is not enough to convince us to change our minds about H_0 and thus we do not reject it

Word of warning: in single sided tests, your test statistic needs to "go against" H_0 for you to reject it. E.g. If H_0 posits that average weights are lower than 50 and your observation is below 50kg, you can't reject, even if the p value is smaller than α

RUNNING EXAMPLE

Make a decision on whether you reject that

- prices in 3rd class were usually a fifth of prices in first class
- prices in third class are a fifth of prices in first class or more expensive