HYPOTHESIS TESTING

HYPOTHESIS TESTING

Statistical Inference: Two statistical data can be numerically compared

Used to test whether results were obtained by luck or not.

HYPOTHESIS TESTING

Hypothesis testing is used to test the validity of a claim (null hypothesis) that is made about a population using sample data.

HYPOTHESIS TESTING

Null Hypothesis H_0

True until proven false

Usually posits no relationship

HYPOTHESIS TESTING

Null Hypothesis H_0

True until proven false

Usually posits no relationship

Alternative Hypothesis

Negation of null hypothesis

Usually asserts specific relationship

HYPOTHESIS TESTING

Null Hypothesis H_0

True until proven false

Usually posits no relationship

Select Test

Pick from vast library

Know which one to choose

Alternative Hypothesis

Negation of null hypothesis

Usually asserts specific relationship

HYPOTHESIS TESTING

Null Hypothesis H_0

True until proven false

Usually posits no relationship

Alternative Hypothesis

Negation of null hypothesis

Usually asserts specific relationship

Select Test

Pick from vast library

Know which one to choose

Test Statistic

Convert to p-value

How likely it was just luck?

HYPOTHESIS TESTING

Null Hypothesis H_0

True until proven false

Usually posits no relationship

Alternative Hypothesis

Negation of null hypothesis

Usually asserts specific relationship

Select Test

Pick from vast library

Know which one to choose

Test Statistic

Convert to p-value

How likely it was just luck?

Significance Level

Usually 1% or 5%

What threshold for luck?

HYPOTHESIS TESTING

Null Hypothesis H_0

True until proven false

Usually posits no relationship

Alternative Hypothesis

Negation of null hypothesis

Usually asserts specific relationship

Select Test

Pick from vast library

Know which one to choose

Test Statistic

Convert to p-value

How likely it was just luck?

Significance Level

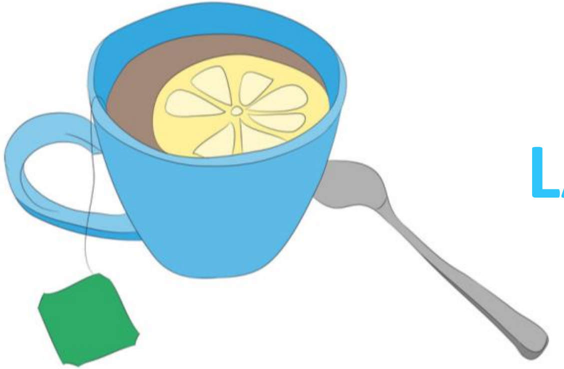
Usually 1% or 5%

What threshold for luck?

Accept or Reject

Small p-value? Reject H_0

Small: Below significance level



LADY TESTING TEA

Famous experiment:

Was tea added before or after milk?

LADY TESTING TEA

Null Hypothesis (H_0)

The lady **cannot** tell if milk was poured first

Alternate Hypothesis (H_1)

The lady **can** tell if milk was poured first

LADY TESTING TEA

Null Hypothesis (H_0)

The lady **cannot** tell if milk was poured first

Alternate Hypothesis (H_1)

The lady **can** tell if milk was poured first

It is good practice to assume that the null hypothesis is correct unless proven otherwise

LADY TESTING TEA

Null Hypothesis H_0

“Lady cannot tell difference”

Can’t tell if milk poured first

LADY TESTING TEA

Null Hypothesis H_0

"Lady cannot tell difference"

Can't tell if milk poured first

Alternative Hypothesis

"Lady can tell difference"

Can indeed discern if milk poured first

LADY TESTING TEA

Null Hypothesis H_0

“Lady cannot tell difference”

Can’t tell if milk poured first

Select Test

8 cups, 4 of each type

Lady got all 8 correct

Alternative Hypothesis

“Lady can tell difference”

Can indeed discern if milk poured first

LADY TESTING TEA

Null Hypothesis H_0

"Lady cannot tell difference"

Can't tell if milk poured first

Select Test

8 cups, 4 of each type

Lady got all 8 correct

Alternative Hypothesis

"Lady can tell difference"

Can indeed discern if milk poured first

Test Statistic

p-value = 1/70 = 1.4%

LADY TESTING TEA

Null Hypothesis H_0

“Lady cannot tell difference”

Can’t tell if milk poured first

Alternative Hypothesis

“Lady can tell difference”

Can indeed discern if milk poured first

Select Test

8 cups, 4 of each type

Lady got all 8 correct

Test Statistic

p-value = 1/70 = 1.4%

Significance Level

Choose 5% significance level

Part of design of experiment

LADY TESTING TEA

Null Hypothesis H_0

"Lady cannot tell difference"

Can't tell if milk poured first

Alternative Hypothesis

"Lady can tell difference"

Can indeed discern if milk poured first

Select Test

8 cups, 4 of each type

Lady got all 8 correct

Test Statistic

p-value = 1/70 = 1.4%

Significance Level

Choose 5% significance level

Part of design of experiment

Accept or Reject

1.4% < 5% → Reject H_0

Lady can indeed tell difference

LADY TESTING TEA

Experiment proved that she could

		Decision about Null Hypothesis	
		REJECT	DON'T REJECT
Null Hypothesis is actually	TRUE	Type I error	Correct Inference
	FALSE	Correct Inference	Type II error

		Decision about Null Hypothesis	
		REJECT	DON'T REJECT
Null Hypothesis is actually	TRUE	Type I error	
	FALSE		

Claim the lady can tell the difference based on spurious test results which are not statistically significant – False Positive

		Decision about Null Hypothesis	
		REJECT	DON'T REJECT
Null Hypothesis is actually	TRUE		
	FALSE		Type II error

Fail to realize that the test for the alternative hypothesis was statistically significant – False Negative

POWER of a Statistical Test

- Probability of rejecting H_0 when H_1 is true (high is good)

High statistical power implies low
probability of Type-II error

Alpha of a Statistical Test – Statistical Significance

- **α** is probability of rejecting H_0 when H_0 is true (low is good)

p-value of a Statistical Test

- p-value is compared to α to decide whether to accept H0
- p-value should be as small as possible (i.e. below α-threshold)
- Typical thresholds are: reject null hypothesis if $p < 1\%$ or $p < 5\%$

T-test

Used to learn about differences in averages across two categories.

Example:
Average male baby birth weight =
Average female baby birth weight?

Outputs:
t-statistic: score which indicate how different the means are;

p-value: whether the t-statistic is significant or not. Low values of p means that the result cannot have happened by chance.

One-sample t-test

Example:

Imagine your provider says that a specific product has an average size of 30mm.

Your procedure would be:

Claim that the null-hypothesis is: **The mean size of the population is equal to 30mm. (H0: $\mu = 30mm$)**

You run the t-test and if the p-value is less that 0.05 (5%), you would have 95% confidence that the null-hypothesis could be rejected. That is: 95% confidence that the product DOES NOT have an average of 30mm. (H1: $\mu != 30mm$)

Two sample location test

Example:

AB-Testing.

Imagine you have an information of a group and then you perform an **intervention.**

You want to test whether the means of your **control group** is statistically significantly different from your **treated group**.

Basically, you are testing a one-sample t-test for the difference on their means.

ANOVA
Analysis of Variance

Example:

You have more than two categories to check.

In the case that you have more than two categories to compare the means, the way to it is to use the analysis of variance.

This will only tell you whether there is some difference in some group. It will not tell you where the difference is.

Then you'd have to run multiple paired tests for each pair of categories.

IRON
HACK