LGN 5822 - Biometrical Genetics

L03 – Hypothesis Testing

Michele Jorge Silva Siqueira

2023

Hypothesis Testing

Motivation

- Hypothesis testing is one of the most important concepts in statistics

- You want to verify that your data is statistically significant and that the result did not just happen by chance (unlikely)

Hypothesis Testing

Motivation

- We will use data analysis to obtain information about probability distributions of particular relevance

How probabilities are distributed a set of possible events

Hypothesis Testing

The role of research in decision making

Understand your hypotheses!

Motivation: Importance of Hypothesis Testing

- **Evidence-Based Decision Making:** Hypothesis testing allows researchers and decision makers to base their conclusions on solid statistical evidence

Motivation: Importance of Hypothesis Testing

- **Validation of Scientific Results:** Hypothesis testing helps determine whether observed results are statistically significant or can be explained by chance

- This is fundamental!

Motivation: Importance of Hypothesis Testing

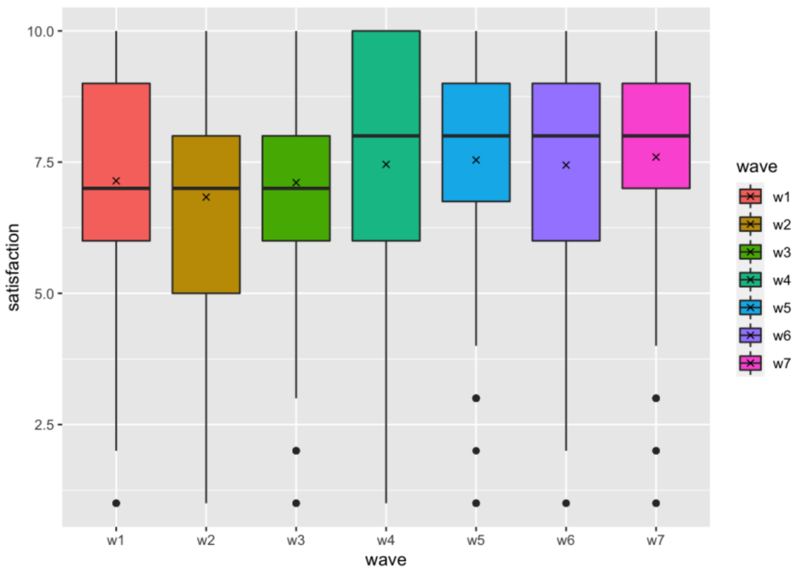
- **Quality Control and Processes:** Hypothesis tests are used to guarantee the quality of products and processes (industry, for example)

- Help identify significant deviations from quality standards and take corrective action

Hypothesis Testing

Motivation: Importance of Hypothesis Testing

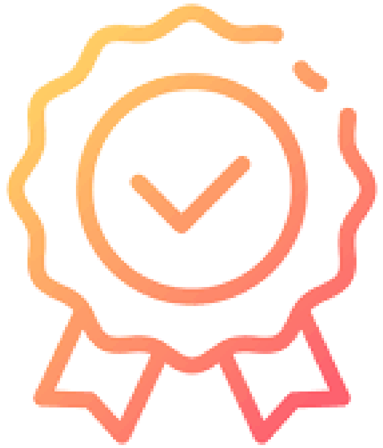
- **Comparison of Groups and Variables:** Hypothesis testing allows you to compare groups of data or variables to determine whether there are significant differences between them



Motivation: Importance of Hypothesis Testing

- **Research Quality Assurance:** Hypothesis testing helps ensure that research results are reliable, replicable and transparent

- It is important to follow a systematic and statistical procedure



Hypothesis Testing

Motivation

- In summary, hypothesis testing is fundamental in many areas, helping to support decisions, validate findings, identify statistical relationships and contribute to the advancement of knowledge and practice in research

Hypothesis Testing

Statistical Inference

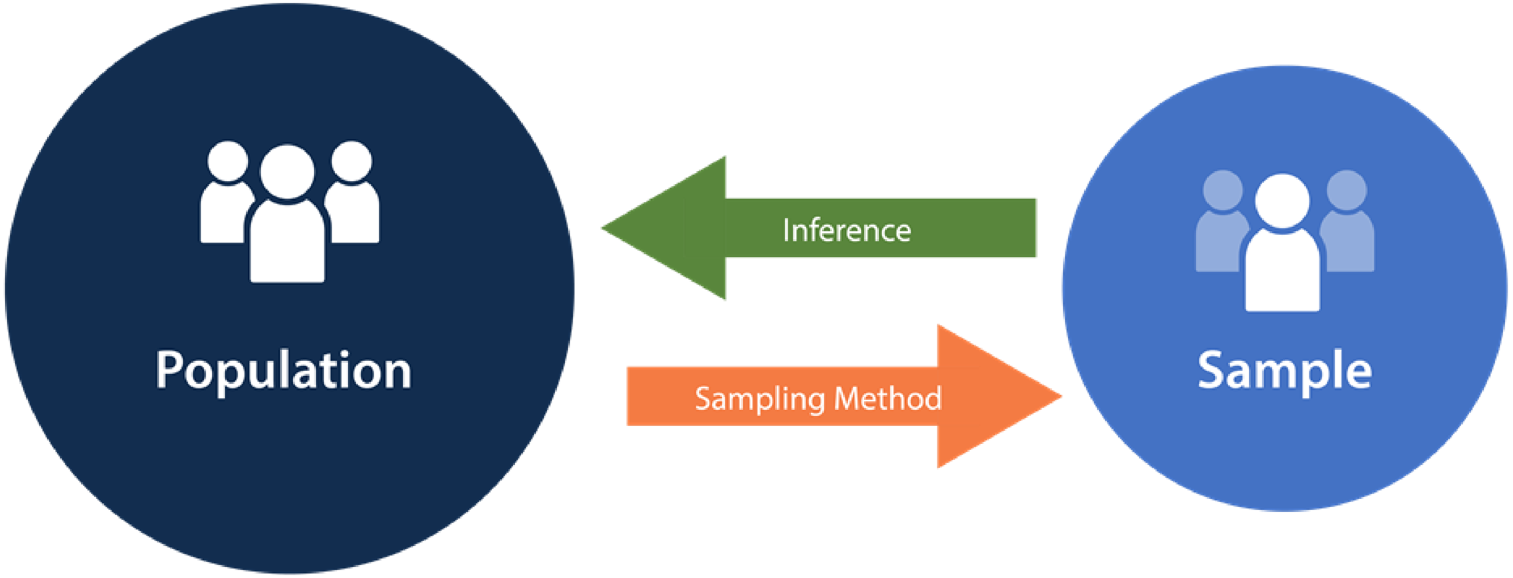
- Sampled data sets will be used to infer properties of the original populations

- This is the goal of **statistical inference**

Hypothesis Testing

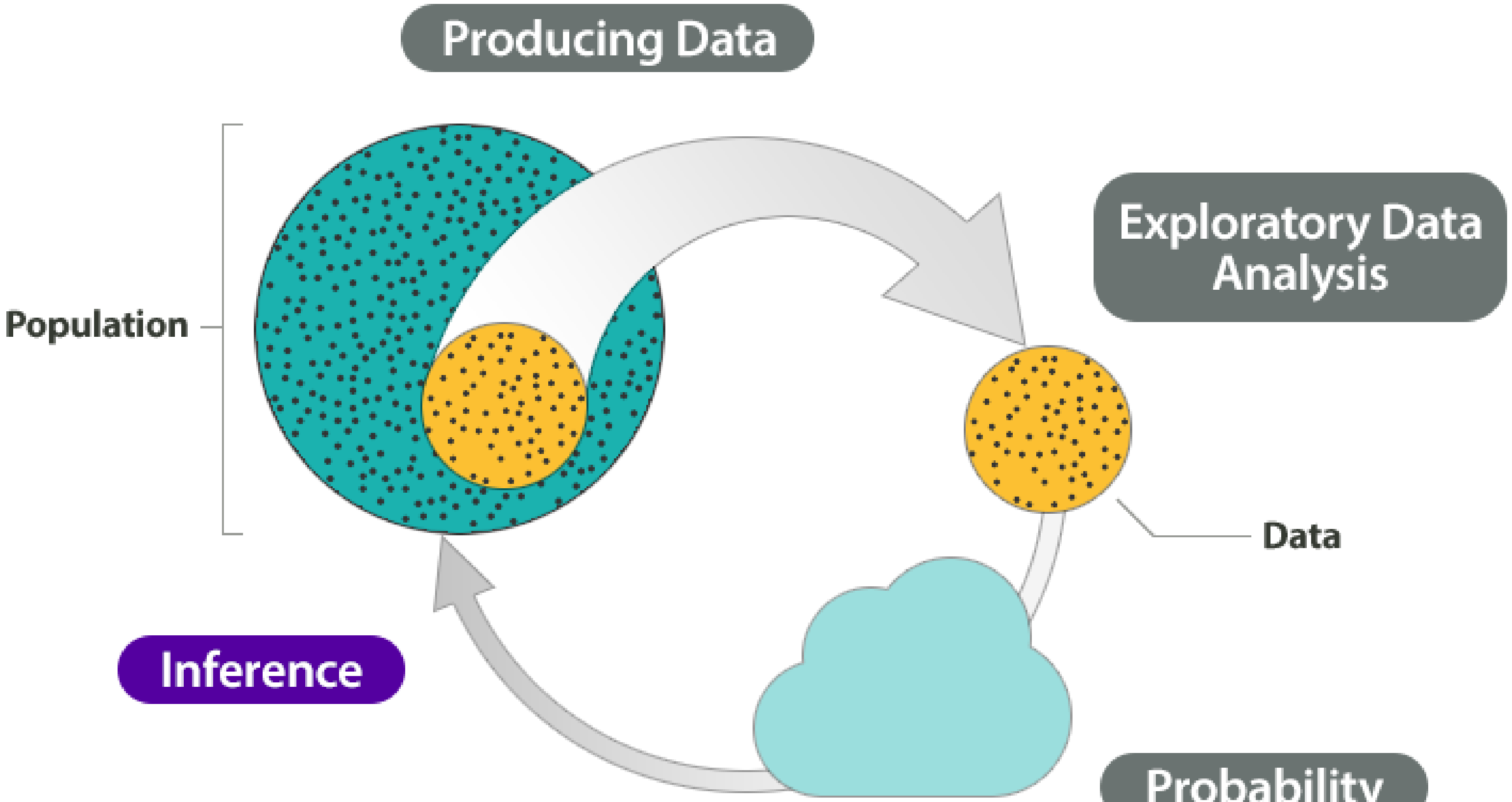
Statistical Inference

- Statistical Inference allows researchers to make conclusions about a population based on evidence from a sample



Hypothesis Testing

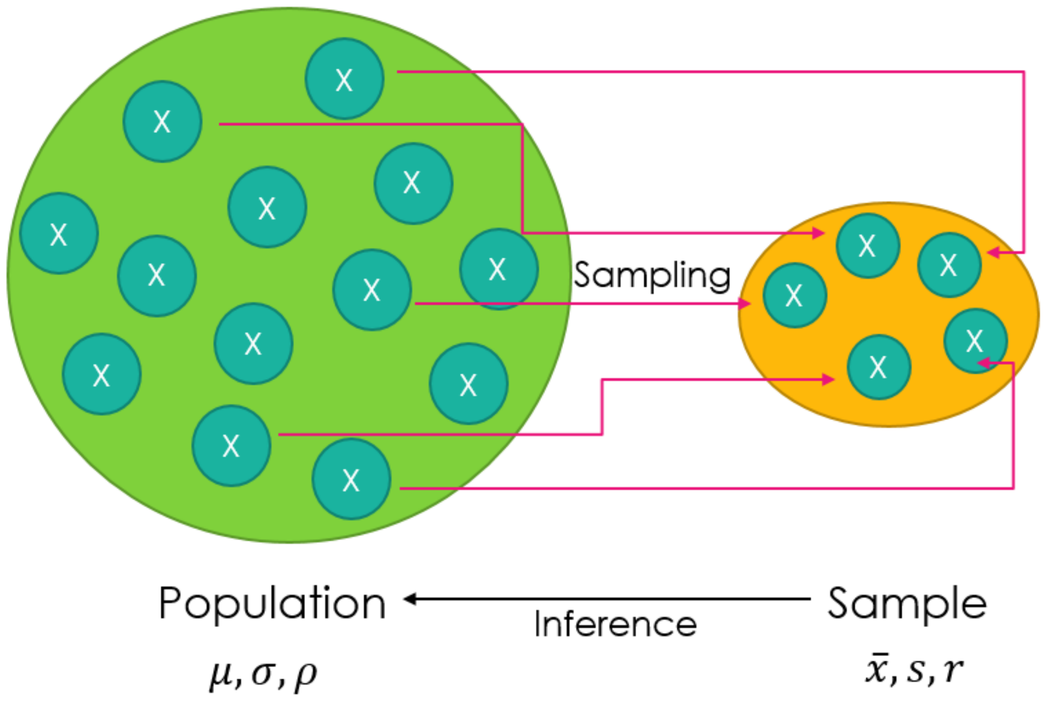
Statistical Inference



Hypothesis Testing

Statistical Inference

- Statistical inference involves two main types of inference:
  - Population Parameter Estimation
  - Hypothesis Testing



Hypothesis Testing

Statistical Inference

- **Population Parameter Estimation:** This is the process of estimating the unknown values (parameters) of a population based on a representative sample
    - Point Estimation: This involves calculating a single value, usually the mean or proportion, based on the sample data
    - Confidence Interval estimation: This involves creating an interval that contains the population parameter with a specified confidence level.

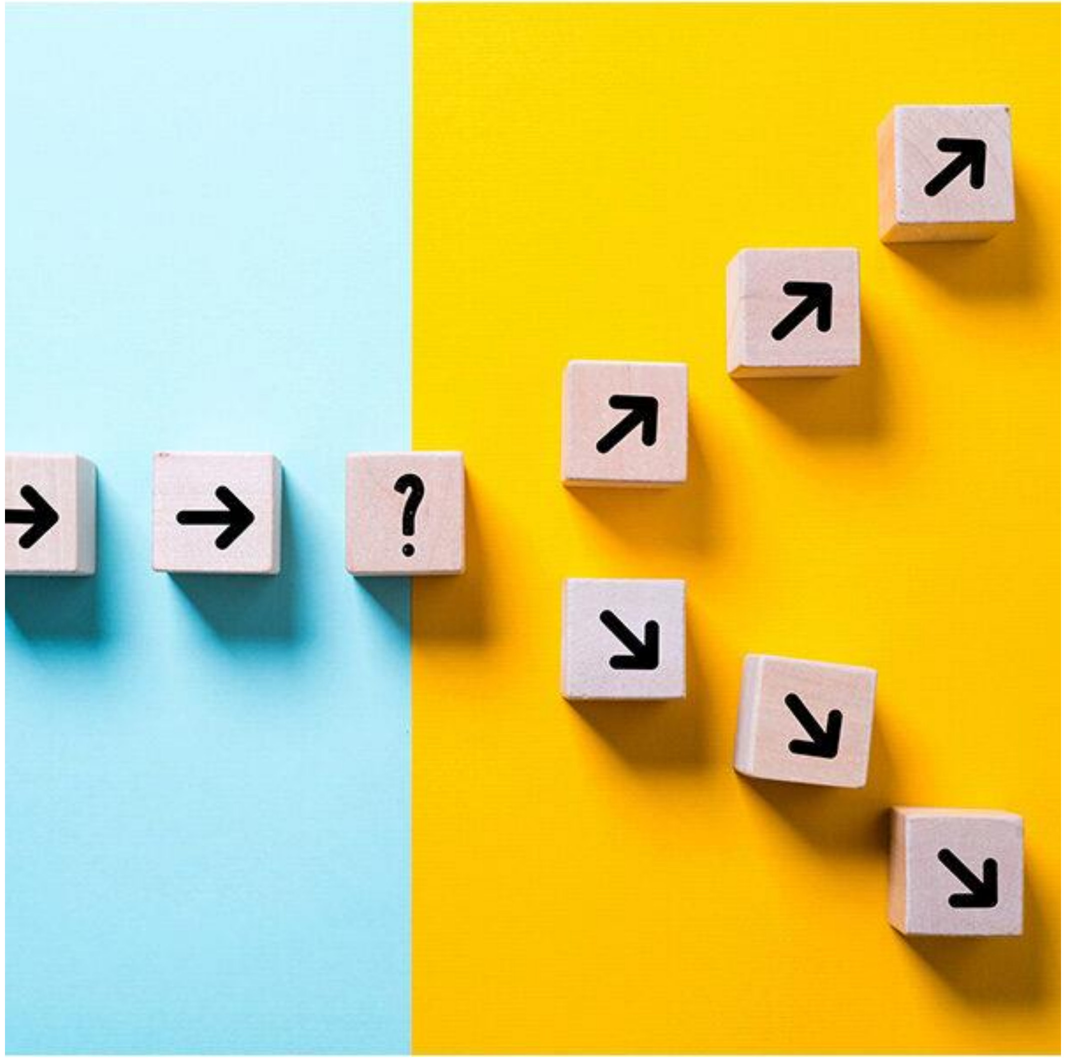
Hypothesis Testing

Statistical Inference

- **Hypothesis Testing**
  - This involves formulating null and alternative hypotheses, collecting data, calculating test statistics, and evaluating statistical evidence to determine whether the null hypothesis should be rejected

Hypothesis Testing

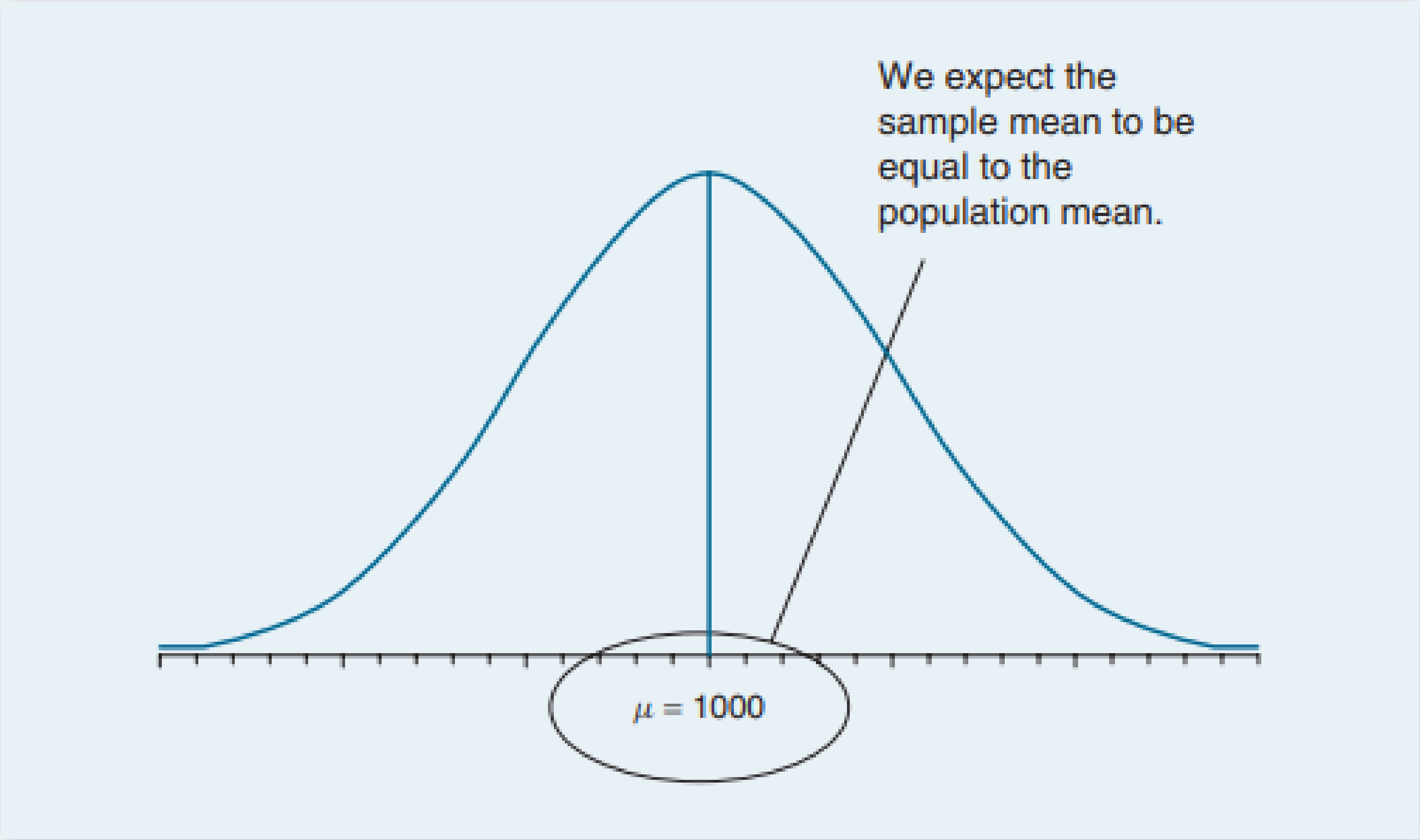
Decision Making



Hypothesis Testing

- Hypothesis Testing is a type of statistical analysis in which you put your assumptions about a population parameter to the test

Hypothesis Testing



Hypothesis Testing

What is your SCIENTIFIC HYPOTHESIS?

Hypothesis Testing

What is Scientific Hypothesis?

- It is an attempt to answer a scientific question or explain an observed phenomenon based on available evidence and prior knowledge
- A scientific hypothesis is one of the initial steps of the scientific method and is tested through experimentation and data analysis

Hypothesis Testing

Steps to Hypothesis Testing

Step 1: Defining the hypotheses

Step 2: Set the criteria for a decision (level of significance)

Step 3: Data collection (experimental trials)

Step 4: Statistic test

Step 5: Make a decision

Hypothesis Testing

Step 1: Defining Hypothesis

Null Hypothesis

- The **null hypothesis**, denoted by H_0, is the hypothesis that we want to test and states that there is no effect, difference or association between variables in the population ($H_0: \mu = 120$; $H_0: \mu_1 = \mu_2$)

- An experiment is frequently designed to assess whether a given null hypothesis of interest can be rejected (or nullified)

Hypothesis Testing

Step 1: Defining Hypothesis

Null Hypothesis

- This is a starting point so that **we can decision making**

- Initially, we state that the null hypothesis is true

Hypothesis Testing

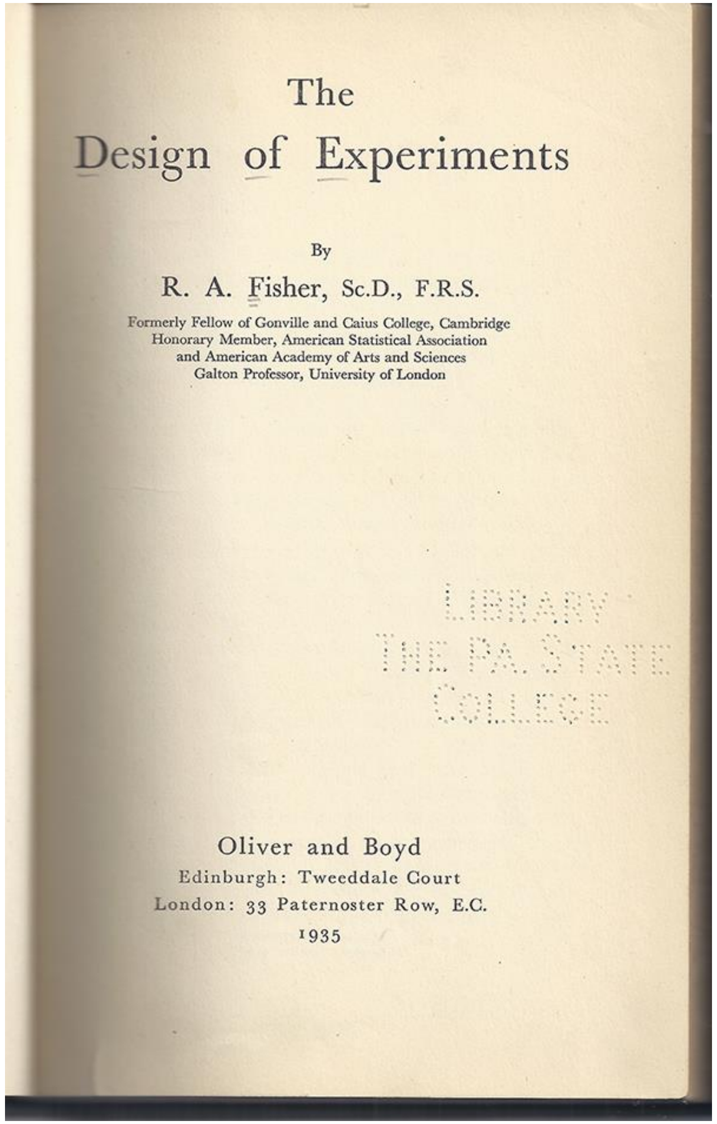
Step 1: Defining Hypothesis

Alternative Hypothesis

- We compare H_0 against a competing **alternative hypothesis** (H_1)
- Let's test whether the value of a population parameter is smaller, larger, or different from the value stated in the null hypothesis $(H_1: \mu \neq 120\ ; H_1: \mu_1 > \mu_2)$
- It is denoted by H_1

Hypothesis Testing

Lady Tasting Tea



The

Design of Experiments

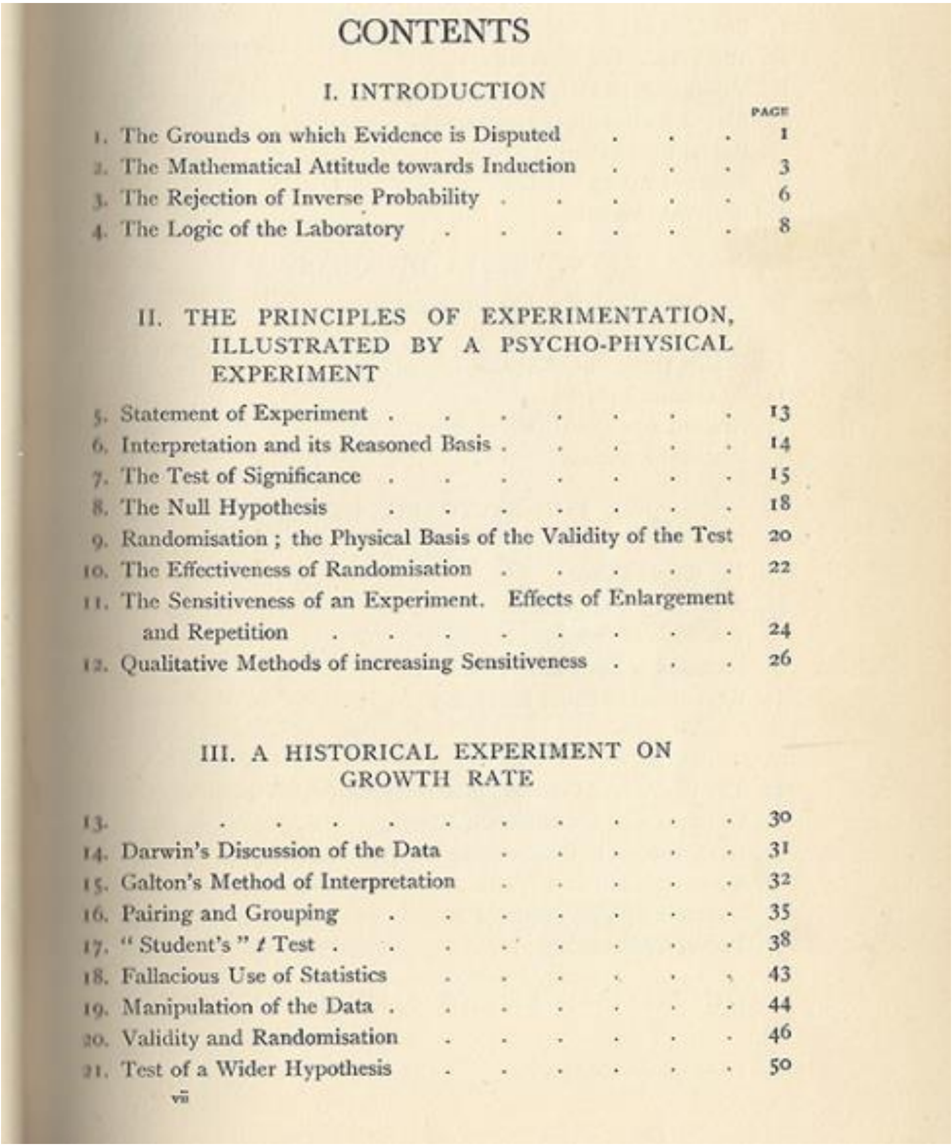
By

R. A. Fisher, Sc.D., F.R.S.

Formerly Fellow of Gonville and Caius College, Cambridge
Honorary Member, American Statistical Association
and American Academy of Arts and Sciences
Galton Professor, University of London

LIBRARY
THE PA. STATE
COLLEGE

Oliver and Boyd
Edinburgh: Tweeddale Court
London: 33 Paternoster Row, E.C.
1935



CONTENTS

I. INTRODUCTION

II. THE PRINCIPLES OF EXPERIMENTATION, ILLUSTRATED BY A PSYCHO-PHYSICAL EXPERIMENT

III. A HISTORICAL EXPERIMENT ON GROWTH RATE

vii

R. A. Fisher to publish *The Design of Experiments* in 1935.

Hypothesis Testing

Lady Tasting Tea

Rothamsted Agricultural Experiment Station

Box, J. F. (1980). RA Fisher and the design of experiments, 1922–1926. *The American Statistician*, *34*(1), 1-7.

Hypothesis Testing

Lady Tasting Tea

Hypothesis Testing

Lady Tasting Tea

Who were the people involved?

Hypothesis Testing

Lady Tasting Tea

The null hypothesis (H_0): The lady cannot distinguish between two types of cups of tea with milk

The alternative hypothesis (H_1): The lady can distinguish between two types of cups of tea with milk

Hypothesis Testing

Lady Tasting Tea

How to test?

4 cups with the tea added before the milk and **4 cups** with the tea added after the milk

Hypothesis Testing

Lady Tasting Tea

What is the probability of the lady getting all 8 cups right, considering the null hypothesis true?

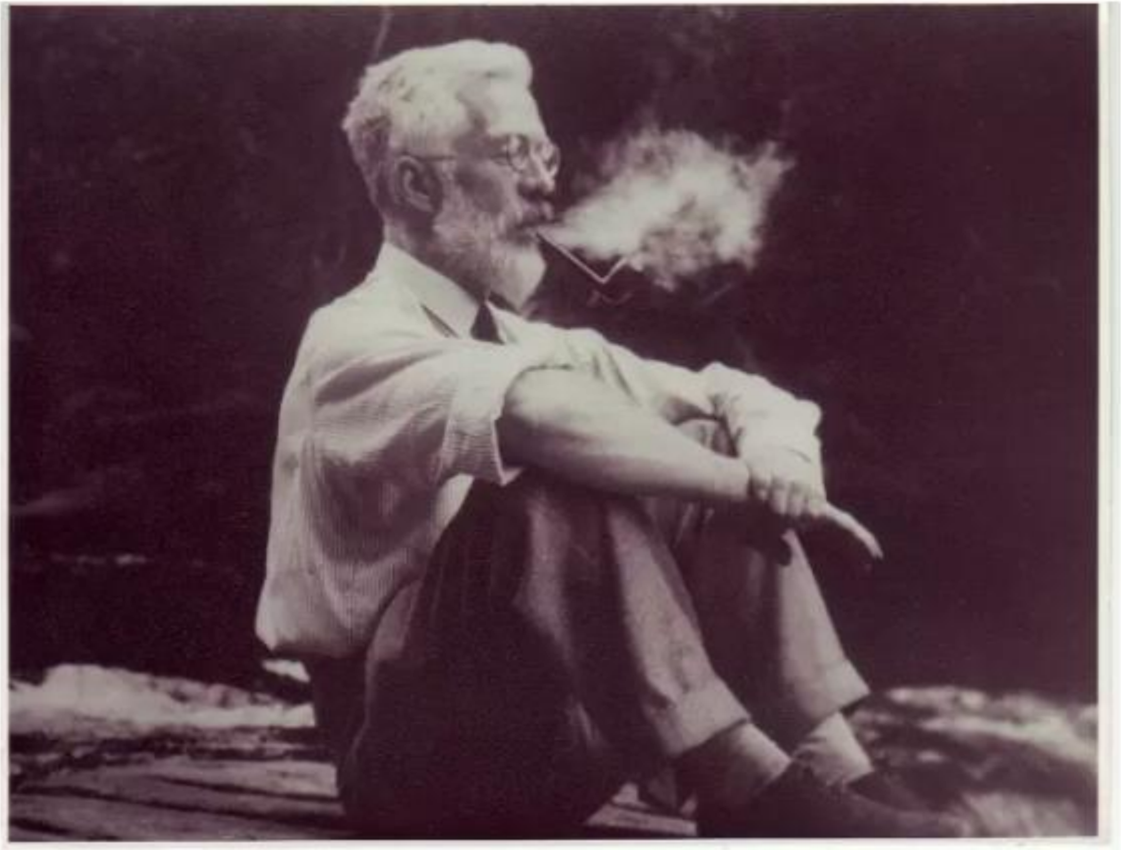
$$C_{(8,4)} = \binom{8}{4} = \frac{8!}{4! \times 4!} = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} = 70$$

$$\frac{1}{70} \times 100 = 1,43\%$$

The calculated probability value (1.43%) corresponds to p-value

Hypothesis Testing

H_0: Smoking does not cause lung cancer



Ronald Fisher (1890-1962)

Correlation is not causation!

Hypothesis Testing

Step 2: Set the criteria for a decision

Level of significance

- To set the criteria for a decision, we state the **level of significance** for a test
- The probability of rejecting the null hypothesis when it is true
- Remember that choosing the significance level is a decision that must be made before carrying out hypothesis testing

Hypothesis Testing

Step 3: Data collection (experimental trials)

Be careful when collecting data!

Hypothesis Testing

Step 4: Statistic test

- A statistical test is a procedure that uses sample data to make decisions about a population

- The value of the test statistic is used to make a decision regarding the null hypothesis

Hypothesis Testing

Step 4: Statistic test

- Choosing the appropriate statistical test depends on the nature of the data and the research question

- Furthermore, it is essential to understand the assumptions associated with each test and correctly interpret the statistical results

Hypothesis Testing

Step 4: Statistic test

Hypothesis Test	
Z - test	Z-score
T - test	T- score
ANOVA	F - statistic
Chi-Square Test	Chi-square statitsic

And others…

Hypothesis Testing

Step 5: Make decision

- We use the value of the test statistic to make a decision about the null hypothesis

- Based on the comparison between the test statistic and the critical value or p-value

Hypothesis Testing

True or False?

The p-value is the probability of rejecting the null hypothesis when it is true

- This is the definition of α (significance level)

Hypothesis Testing

What is a p-value?

Definition

- p-value or "probability value"

- The p-value is the probability that the test statistic has an extreme value relative to the observed value when the null hypothesis $(\mathbf{H_0})$ is true

- It is a function of the sample data (and also a random variable)

- The p-value is always obtained from a sample

Hypothesis Testing

What is a p-value?

H_0: The coin is biased

What is the probability of flipping a coin 5 times and all 5 times it comes up tails?

$$\left(\frac{1}{2}\right)^5 = 3.125\%$$

It's a very small chance of happening. But, it is possible.

This probability is the p-value

Hypothesis Testing

Don´t confuse!

- The p-value with the significance level !

- The significance level is the probability of rejecting the null hypothesis when it is true

- The significance level must be chosen before performing the statistical test

Hypothesis Testing

IN FOCUS NEWS

REPRODUCIBILITY

Statisticians issue warning on *P* values

Statement aims to halt missteps in the quest for certainty.

BY MONYA BAKER

Misuse of the *P* value — a common test for judging the strength of scientific evidence — is contributing to the number of research findings that cannot be reproduced, the American Statistical Association (ASA) warned on 8 March. The group has taken the unusual step of issuing principles to guide use of the *P* value, which it says cannot determine whether a hypothesis is true or whether results are important.

This is the first time that the 177-year-old ASA has made explicit recommendations on such a foundational matter, says executive director Ron Wasserstein. The society's members had become increasingly concerned that the *P* value was being misapplied, in ways that cast doubt on statistics generally, he adds.

cannot indicate the importance of a finding; for instance, a drug can have a statistically significant effect on patients' blood glucose levels without having a therapeutic effect.

Giovanni Parmigiani, a biostatistician at the Dana Farber Cancer Institute in Boston, Massachusetts, says that misunderstandings about what information a *P* value provides often crop up in textbooks and practice manuals. A course correction is long overdue, he adds. "Surely if this happened twenty years ago, biomedical research could be in a better place now."

FRUSTRATION ABOUNDS

Criticism of the *P* value is nothing new. In 2011, researchers trying to raise awareness about false positives gamed an analysis to reach a statistically significant finding: that listening to music by the Beatles makes undergraduates younger

Baker, M. Statisticians issue warning on P values. *Nature* 531, 151 (2016).

Hypothesis Testing

Hypothesis Testing

How do we use and interpret p-values?

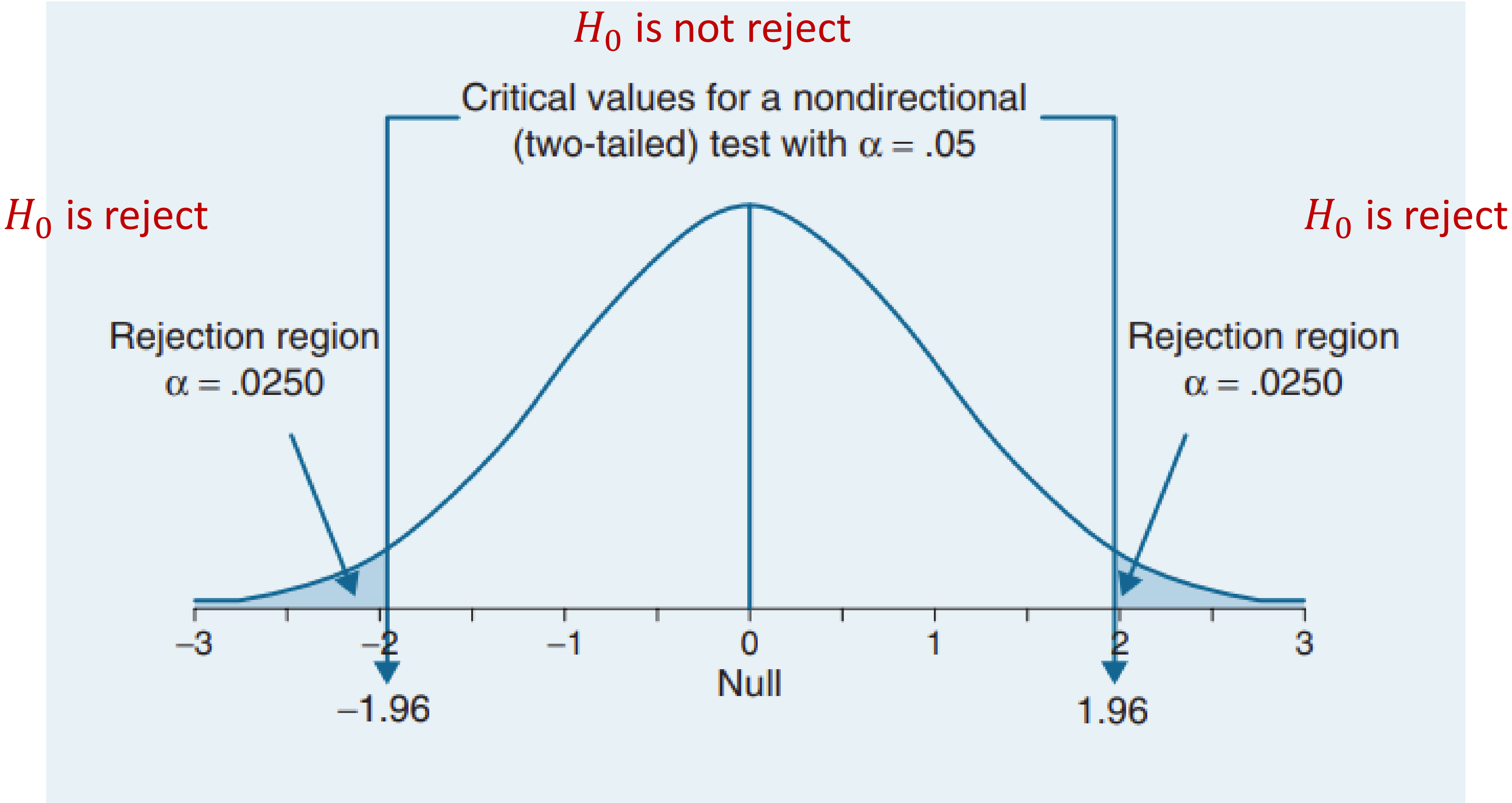
- We can compare a p-value with a predefined α

p-value interpretation

- If $\boldsymbol{p} \leq \boldsymbol{\alpha}$, we **reject** the null hypothesis H_0
- If $\boldsymbol{p} > \boldsymbol{\alpha}$, we **accept** the null hypothesis H_0

Hypothesis Testing

Z-test



Frequentist statistics: use of p-value and α as evidence to test a given hypothesis

The t statistic

- We will use the **t-test as an example** to understand several concepts of hypothesis testing
- This will help our discussion of values of p and various testing problems
- It will be useful when we deal with more complex linear models and other tests in future classes
- Also known as Student's t-test

Hypothesis Testing

The t statistic

- The t- test is primarily used to compare the means of two groups or samples and determine whether there is a statistically significant difference between them

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

Where $\bar{X}$ is the mean of the sample
μ is the mean of population
s is the standart deviation
n is the number of observations

One sample t-test

- One sample t-test is used to determine whether the mean of a single sample is statistically different from a reference value

Hypothesis Testing

One sample t-test

- You perform an experiment to evaluate the yield of the new genotype, with **n** replicates

- Let $y_1, y_2, y_3, \ldots, y_n$ denote the observed values

One sample t-test

- Assume that the observed values y_i are independently sampled from a distribution with **unknown** mean μ and variance σ^2

- We usually assume that $y_i \sim N\ (\mu, \sigma^2)$, but for large n the **sample mean** is close to a a Normal distribution even if the y_i are not normally distributed

Hypothesis Testing

One sample t-test

- What hypothesis do we want to test?

Hypothesis Testing

One sample t-test

- What hypothesis do we want to test?

Hypothesis

- Let μ_0 be a predefined value of interest
- We want to test $H_0: \mu = \mu_0$
- The alternative hypothesis (two-sided) is $H_1: \mu \neq \mu_0$

Hypothesis Testing

One sample t-test

- We can use the Student's t-test for that purpose

- From the observed values, calculate the sample mean $\bar{y}$ and sample variance s^2

- It can be shown that, under $\boldsymbol{H_0}$, the statistic

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

follows a t distribution with n - 1 degrees of freedom

Hypothesis Testing

- Let´s Practice 01!

\# Create a random data sample

\# Define a reference value and and level of significance (mean that will be tested)

\# Run the one-sample t-test

use the t.test( ) function

\#Interpret your results

Hypothesis Testing

- Let´s Practice 01!


```
> # Create a random data sample
> sample <- c(22, 25, 28, 32, 27, 30, 35, 38, 31, 29)
> # Define a reference value (average that will be tested)
> reference_value <- 30
> # Run the one-sample t-test
> result_test <- t.test(sample, mu = reference_value)
> result_test

        One Sample t-test

data:  sample
t = -0.20324, df = 9, p-value = 0.8435
alternative hypothesis: true mean is not equal to 30
95 percent confidence interval:
 26.36082 33.03918
sample estimates:
mean of x
     29.7
```

Hypothesis Testing

- Let´s Practice 02!

- For example, we want to evaluate whether a newly obtained genotype has the potential to be launched as a new cultivar on the seed market

- Suppose further that farmers only accept genotypes with an average yield of **4 t/ha**

Hypothesis Testing

- Let´s Practice 02!

Test $H_0: \mu = 4t/ha$ for the following observed yields (in t/ha):

3.5	4.3	3.3	5.7	4.4
3.3	4.6	4.8	4.7	3.8

- First, calculate mean, variance and standard deviation

- Next, calculate the t-statistic
  #use the t.test( ) function

Hypothesis Testing

- First, calculate mean and variance

```
> yields <- c(3.5, 4.3, 3.3, 5.7, 4.4, 3.3, 4.6, 4.8, 4.7, 3.8)
>
> mean(yields)
[1] 4.24
>
> var(yields)
[1] 0.5915556
>
> sd(yields)
[1] 0.7691265
```

Hypothesis Testing

- Next, calculate the t-statistic

```
> #by equation t-test
> numerator <- (mean(yields)-4)
> numerator
[1] 0.24
> denominator <- (sd(yields)/sqrt(10))
> denominator
[1] 0.2432192
> t_stat <- numerator/denominator
> t_stat
[1] 0.9867644
```

Hypothesis Testing

- The pt( ) function in R is used to calculate the cumulative distribution function of the Student's t-distribution

  #pt(q, df, lower.tail = TRUE)

- The abs( ) function is used to calculate the absolute value of a number, that is, to obtain the positive value of a number

  # abs( )

- The corresponding p-value

```
> 2*pt(abs(t_stat), df = 9, lower.tail = FALSE)
[1] 0.3495419
```

Hypothesis Testing

#or use the t.test( ) function

```
> t.test(yields, mu =4)

        One Sample t-test

data:  yields
t = 0.98676, df = 9, p-value = 0.3495
alternative hypothesis: true mean is not equal to 4
95 percent confidence interval:
 3.6898 4.7902
sample estimates:
mean of x
     4.24
```

Hypothesis Testing

Two sample t-test

- The two-sample t-test, is used to determine whether the means of two independent samples are statistically different from each other

Hypothesis Testing

Two sample t-test

- Assume that the observed values are independently sampled from distributions with **unknown** means μ_1 and μ_2 for the two groups

- Assume a **common unknown variance $\boldsymbol{\sigma^2}$**

- Again, we commonly assume normality: $y_{1i} \sim N(\mu_{1,} \sigma^2)$ and $y_{2i} \sim N(\mu_{2,} \sigma^2)$

Hypothesis Testing

- How can we define the **Null Hypothesis**?

Hypothesis Testing

Hypothesis

- In this case, we want to test H_0: $\mu_1 = \mu_2$
- Or, equivalently, H_0: $\mu_1 - \mu_2 = 0$
- The alternative hypothesis (two-sided) is H_1: $\mu_1 \neq \mu_2$

Hypothesis Testing

Considering equal variance

- We can now use the two-sample t-test
- From the observed values, calculate the sample means $\overline{y_1}$ and $\overline{y_2}$ for groups one and two, respectively
- Similarly, calculate sample variances s_1^2 and s_2^2
- Then, obtain a pooled variance estimate: $s_p^2 = \frac{s_1^2 + s_2^2}{2}$

Hypothesis Testing

- It can be shown that, **under $\boldsymbol{H_0}$**, the statistic

$$t = \frac{\overline{y_1} - \overline{y_2}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

follows a t distribution with $2n - 2$ degrees of freedom

Hypothesis Testing

Considering unequal variance

$$t = \frac{\overline{y_1} - \overline{y_2}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

Hypothesis Testing

- Let´s Practice 03!

Create two random data sample (Group A and Group B)

Run the two-sample t-test

use the t.test( ) function

#Interpret your results

Hypothesis Testing

- Let´s Practice 03!


```
> # Sample data for Group A and Group B
> group_a <- c(25, 30, 35, 40, 45)
> group_b <- c(20, 28, 32, 38, 42)
> # Perform a two-sample t-test
> t_test_result <- t.test(group_a, group_b)
> t_test_result

        Welch Two Sample t-test

data:  group_a and group_b
t = 0.57417, df = 7.9436, p-value = 0.5817
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.063636 15.063636
sample estimates:
mean of x mean of y 
       35        32 
```

Hypothesis Testing

- Let´s Practice 04!

- Suppose you want to compare the growth rate of two strains of a given bacteria

- You carry out a lab experiment and measure the colony sizes after two days of growth

Hypothesis Testing

- Let´s Practice 04!

  - For each group, you measure the size of n independent colonies

  - Let $y_{11}, y_{12}, \dots, y_{1n}$ denote the observed values for the first group, while values $y_{21}, y_{22}, \dots, y_{2n}$ represent the measured colony sizes for the second group

Hypothesis Testing

- Let´s Practice 04!

  - Test $H_0: \mu_1 = \mu_2$ for the following observed colony sizes (cm)

Group 1		Group 2	
0.4	0.7	1.2	1.0
0.9	0.3	0.7	0.8
0.9	0.8	1.6	1.2
0.8	1.2	0.9	0.7
0.3	1.3	1.0	0.4
0.5	0.3	0.9	0.8
0.5	0.9	1.6	1.5

- Obtain the sample means
- Calculate sample variances and the pooled estimate
- Calculate the t statistic
- And assess its p −value

Hypothesis Testing

- Obtain the sample means

```
> colonies_1 <- c(0.4, 0.7, 0.9, 0.3, 0.9, 0.8,
+                 + 0.8, 1.2, 0.3, 1.3, 0.5, 0.3, 0.5, 0.9)
>
> colonies_2 <- c(1.2, 1, 0.7, 0.8, 1.6, 1.2, 0.9,
+                   + 0.7, 1, 0.4, 0.9, 0.8, 1.6, 1.5)
>
> mean(colonies_1)
[1] 0.7
>
> mean(colonies_2)
[1] 1.021429
```

Hypothesis Testing

- Calculate sample variances and the pooled estimate

```
> var(colonies_1)
[1] 0.1076923
>
> var(colonies_2)
[1] 0.1295055
>
> mediavariancia <- (var(colonies_1) + var(colonies_2))/2
> mediavariancia
[1] 0.1185989
>
> (pooled_var <- (var(colonies_1) + var(colonies_2))/2)
[1] 0.1185989
```

Hypothesis Testing

- Calculate the t statistic:

```
> (mean_diff <- mean(colonies_1) - mean(colonies_2))
[1] -0.3214286
>
> (t_stat <- mean_diff/sqrt(pooled_var * 2/14))
[1] -2.46941
```

Hypothesis Testing

- And assess its p –value

```
> 2 * pt(abs(t_stat), df = 26, lower.tail = FALSE)
[1] 0.02041979
```

- If $\boldsymbol{p} \leq \boldsymbol{\alpha}$, we **reject** the null hypothesis H_0
- If $\boldsymbol{p} > \boldsymbol{\alpha}$, we **accept** the null hypothesis H_0

Hypothesis Testing

#or use the t.test( ) function

```
> t.test(colonies_1, colonies_2, var.equal = TRUE)

        Two Sample t-test

data:  colonies_1 and colonies_2
t = -2.4694, df = 26, p-value = 0.02042
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.58898477 -0.05387237
sample estimates:
mean of x mean of y 
 0.700000  1.021429 
```

Hypothesis Testing

Considerations about the t-test

#Notes

- We can test hypotheses such as $H_1: \mu > \mu_0$ (one-tailed tests)
- There are modifications to the t-test when sample sizes and/or variances are different
- There are also appropriate t-tests for paired samples
  - Paired samples are obtained from the same group of individuals or objects in two different conditions

Hypothesis Testing

Paired Samples t-test: Repeated Experiments

- The paired samples t-test is used to compare the means of two related measurements that have been collected in pairs
- To assess the change in a continuous outcome across time or within-subjects across two observations

Hypothesis Testing

- Let´s Practice!

- Consider evaluating plant growth (plant height) before and after nitrogen addition (n=10)

    - Create the dataframe
    - Visualize the data
    - Perform the paired sample t-test

      #use the argument paired = TRUE

    - Demonstrate the histogram

Hypothesis Testing

- Let´s Practice!

  - Create the dataframe

```
# Create dataframe
data <- data.frame(
  Number = 1:10,
  height_plants1 = c(120, 122, 118, 125, 130, 128, 123, 126, 119, 121),
  height_plants2 = c(115, 121, 117, 124, 129, 126, 122, 125, 118, 120)
)
```

Hypothesis Testing

Paired Samples t-test: Repeated Experiments

- Visualize the data

```
> # Visualize the data
> print(data)
   Number height_plants1 height_plants2
1       1            120            115
2       2            122            121
3       3            118            117
4       4            125            124
5       5            130            129
6       6            128            126
7       7            123            122
8       8            126            125
9       9            119            118
10     10            121            120
```

Paired Samples t-test: Repeated Experiments

- Perform the paired sample t-test

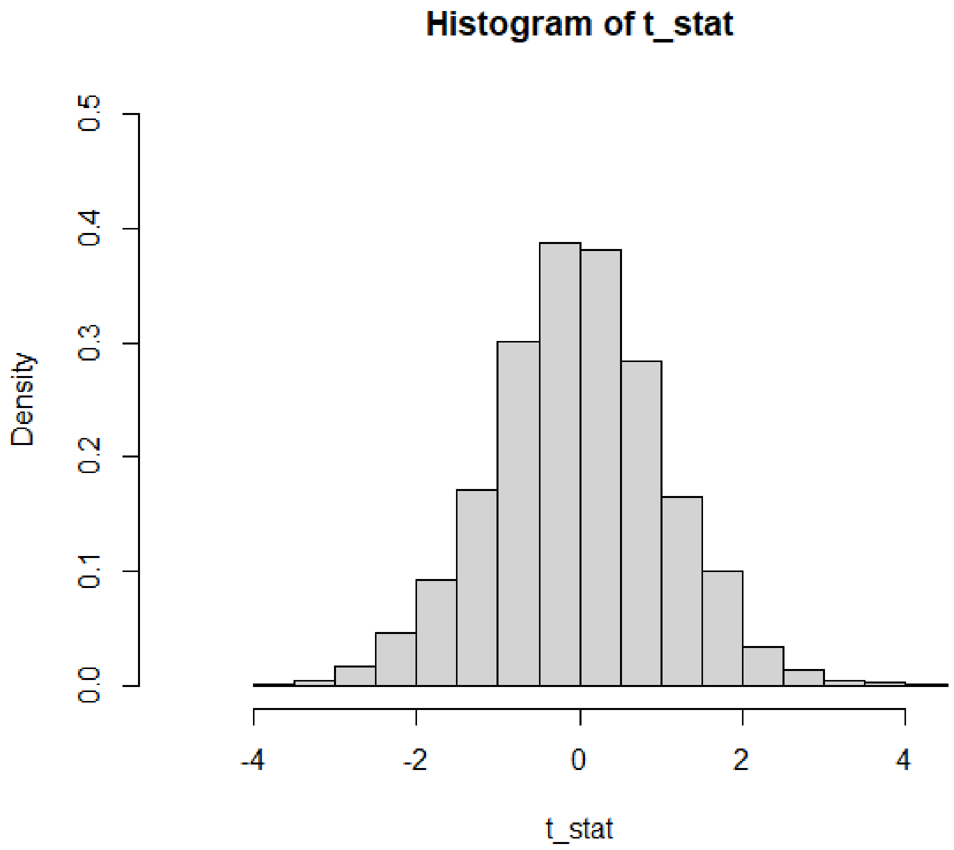
```
> # Perform the t-test for paired samples
> result_test <- t.test(data$height_plants1, data$height_plants2, paired = TRUE)
> print(result_teste)

        Paired t-test

data:  data$height_plants1 and data$height_plants2
t = 3.737, df = 9, p-value = 0.004647
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 0.5920007 2.4079993
sample estimates:
mean difference 
            1.5 
```

Hypothesis Testing

Paired Samples t-test: Repeated Experiments

- Demonstrate the histogram

```
> hist(t_stat)
```



Hypothesis Testing

No hypothesis test is 100% certain. Because the test is based on probabilities, there is always a chance of making an incorrect conclusion

Hypothesis Testing

- Possible errors associated with hypothesis testing

	Declared non-significant (H_0 not rejected)	Declared significant (H_0 rejected)
H_0 is true	Correct Decision	
H_0 is non-true		

Hypothesis Testing

	Declared non-significant (H_0 not rejected)	Declared significant (H_0 rejected)
H_0 is true	Correct Decision	
H_0 is non-true		Correct Decision

Hypothesis Testing

Types of Error

	Declared non-significant (H_0 not rejected)	Declared significant (H_0 rejected)
H_0 is true	Correct Decision	**Type I error**
H_0 is non-true		Correct Decision

Type I error: When the null hypothesis (H_0) is true and you reject it

- Called a "false positive" or "false discovery"
- The probability of type I error is denoted as α
  - For example, if you choose a significance level of 0.05 for a test, that means there is a 5% chance of making a Type I error by rejecting H_0 when it is true

Hypothesis Testing

Types of Error

	Declared non-significant (H_0 not rejected)	Declared significant (H_0 rejected)
H_0 is true	Correct Decision	**Type I error**
H_0 is non-true	**Type II error**	Correct Decision

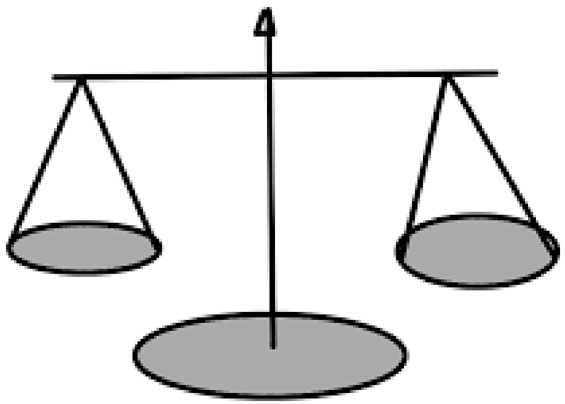
Type II error: When you accept the null hypothesis (H_0) when it is false

- Called a "false negative"
- The probability of type II error is denoted as β
- Power of the test: 1 - β
  - The probability of correctly rejecting H_0 when it is false

Hypothesis Testing

Types of Error

Generally, you can reduce type I error by increasing (up) the significance level, but this will increase (up) type II error



Hypothesis Testing

Types of Error

- Resume

	Declared non-significant (H_0 not rejected)	Declared significant (H_0 rejected)
H_0 is true	Correct Decision	Correct Decision
H_0 is non-true	**Type II error**	True positive

Type I error: **false positive**
Type II error: **false negative**

Hypothesis Testing

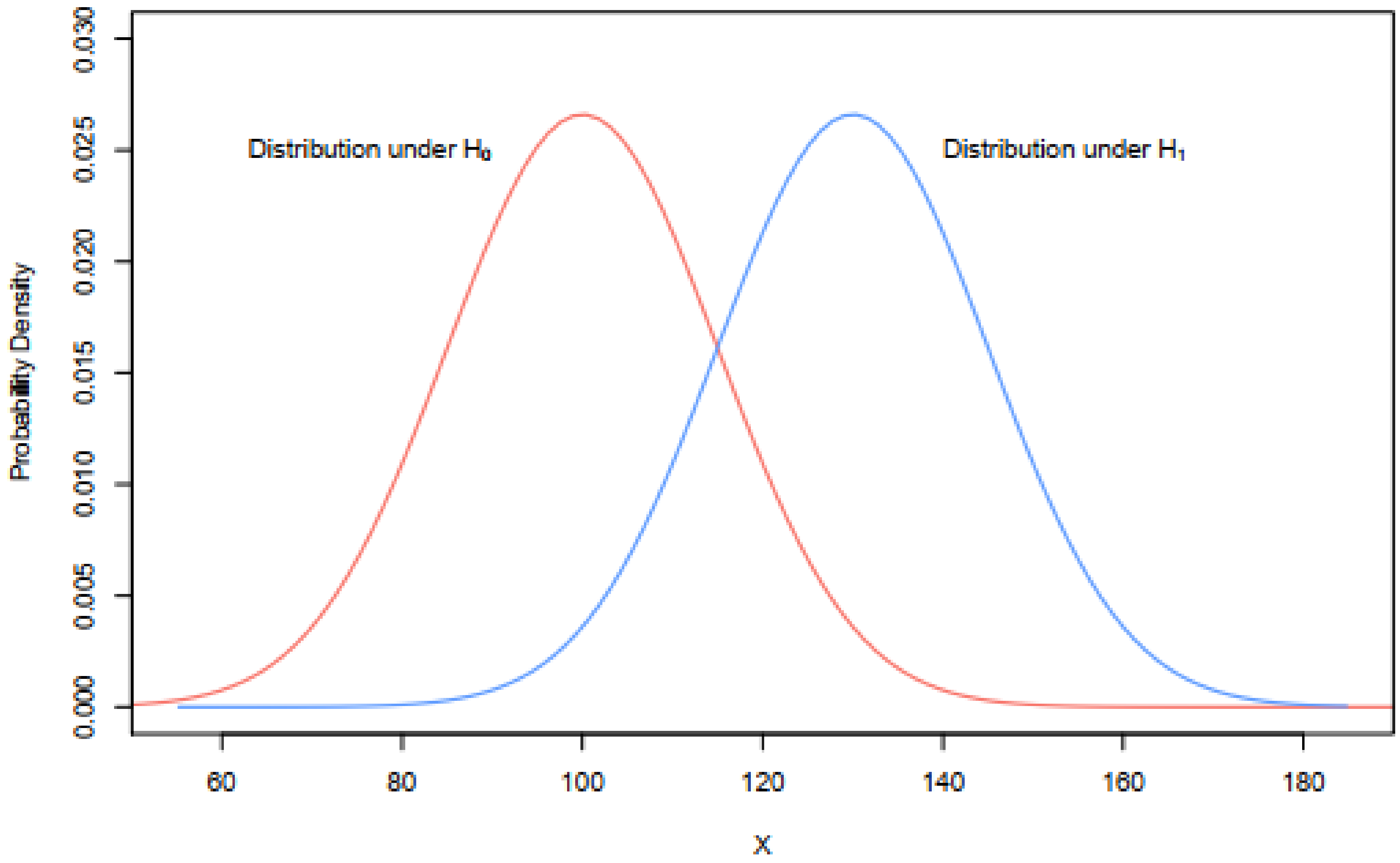
Types of Error

When the null hypothesis (H_0) is true and you reject it

When you accept the null hypothesis (H_0) when it is false

Hypothesis Testing

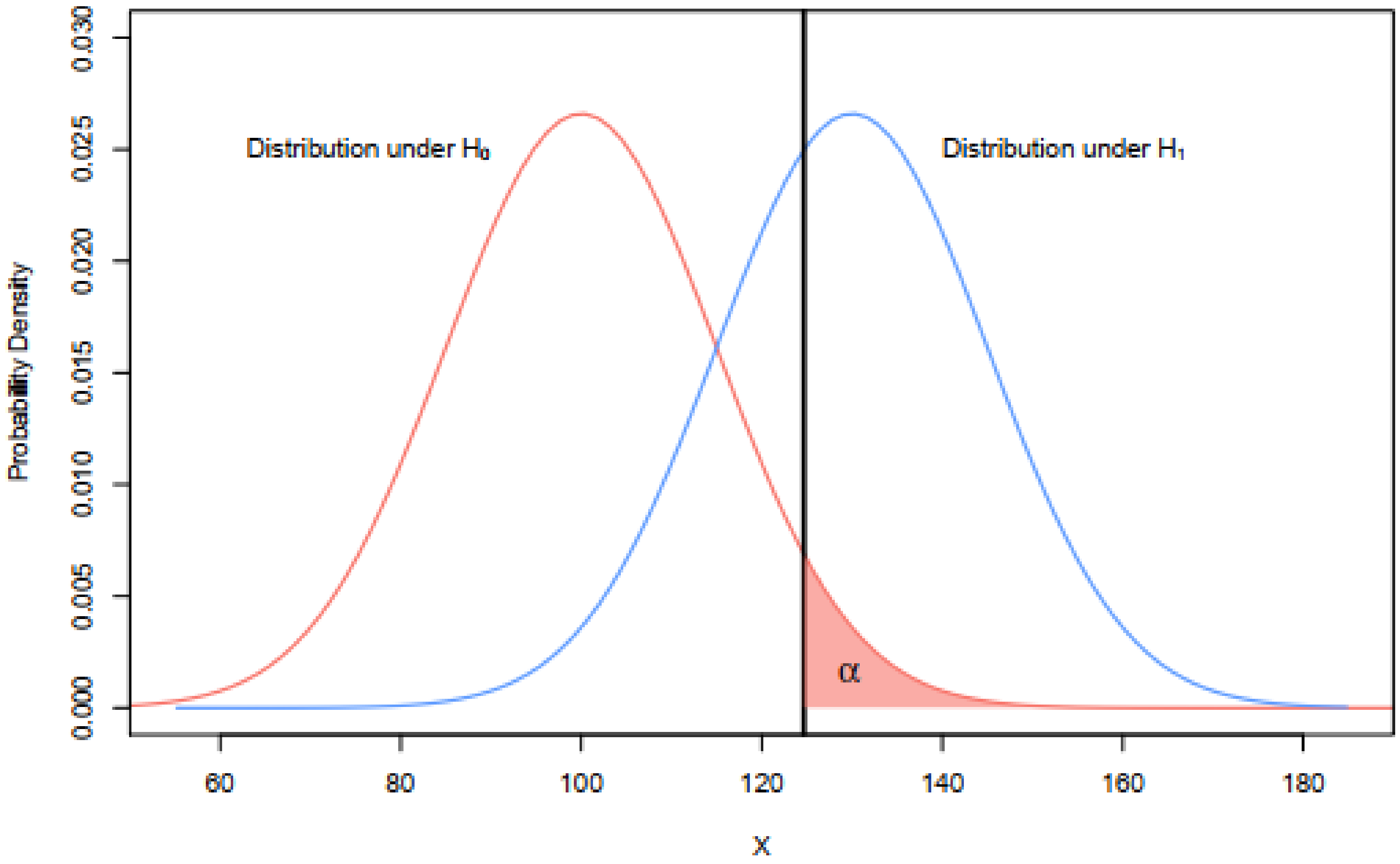
Types of Error



Adapted from http://stats.stackexchange.com/questions/7402/

Hypothesis Testing

Types of Error

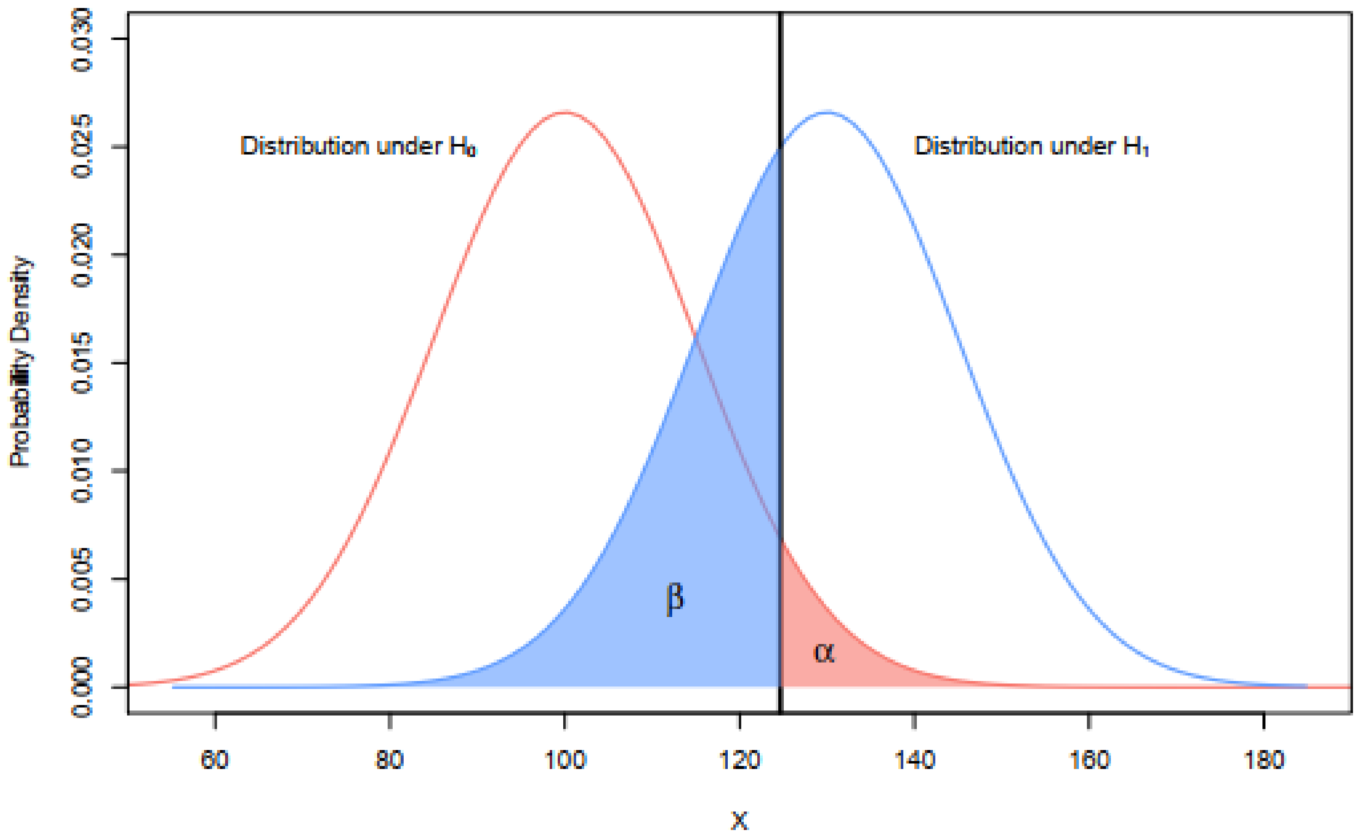


The probability of type I error is denoted as α

Adapted from http://stats.stackexchange.com/questions/7402/

Hypothesis Testing

Types of Error

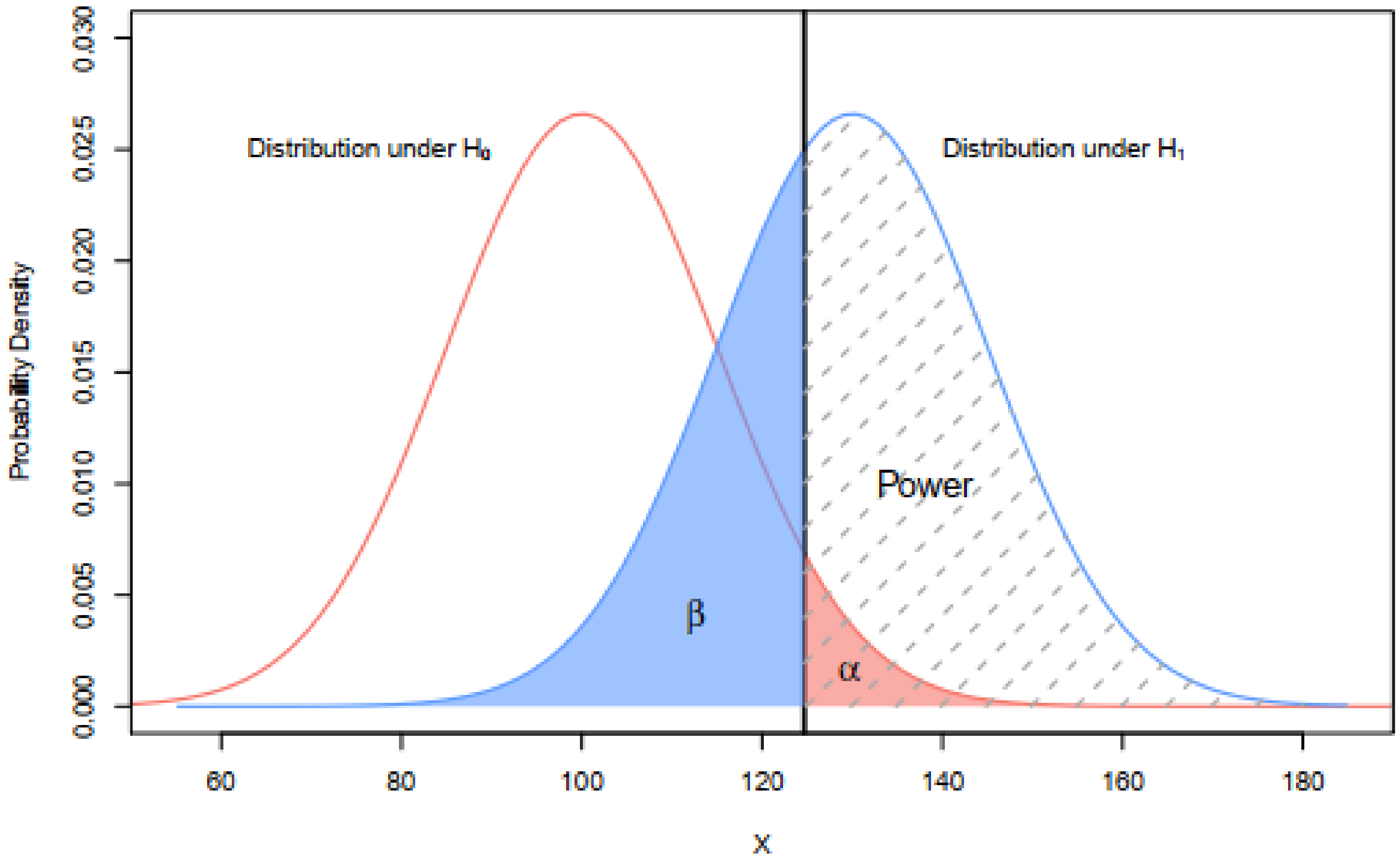


The probability of type II error is denoted as **β**

Adapted from http://stats.stackexchange.com/questions/7402/

Hypothesis Testing

Types of Error



The probability of correctly rejecting H_0 when it is false

Adapted from http://stats.stackexchange.com/questions/7402/

Hypothesis Testing

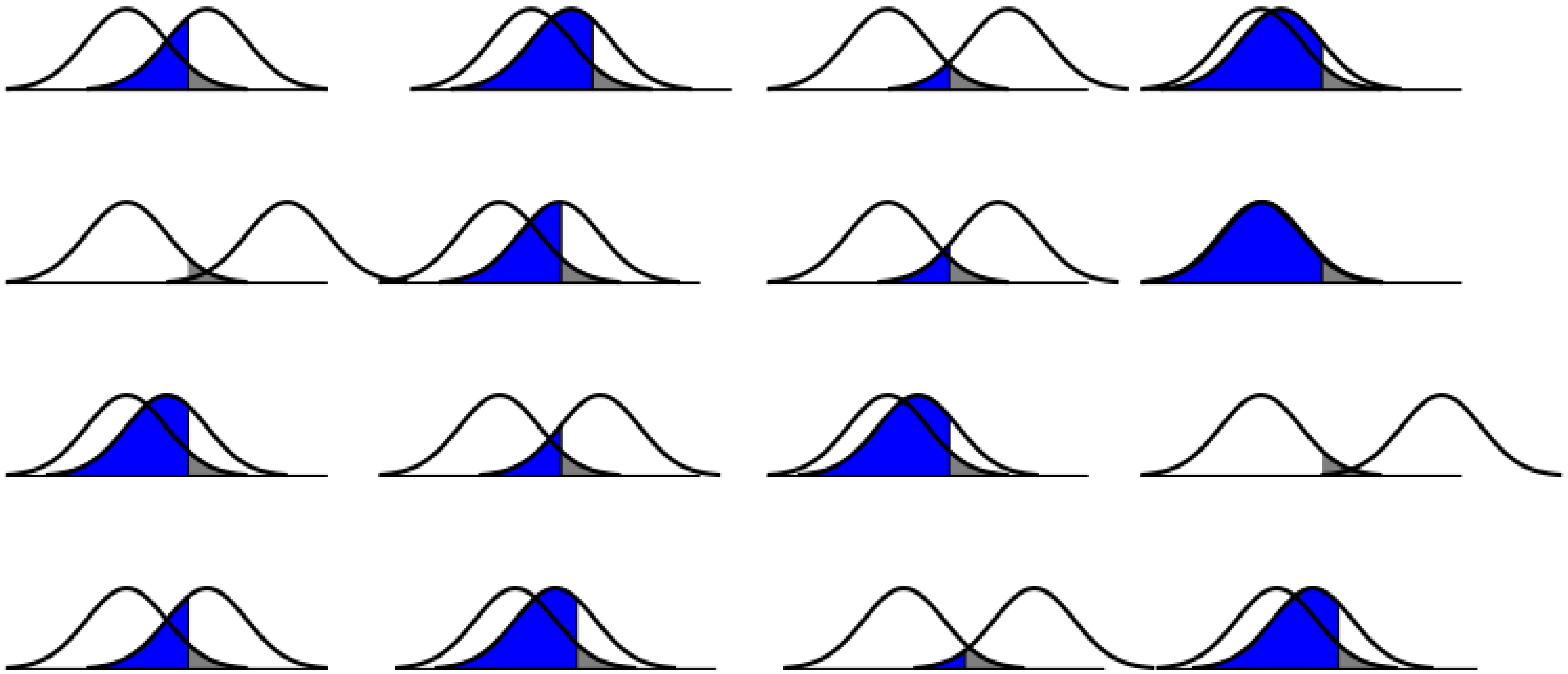
Multiple Comparison Test

- When we have multiple **alternative hypotheses**, we refer to it as Multiple Testing

- It is important to test multiple hypotheses is to control Type I error

Hypothesis Testing

Multiple Comparison Test



Each test has possible **Type I** and **Type II** errors, and there are many possible ways to combine them. The probability of a **Type I** error grows with the number of tests

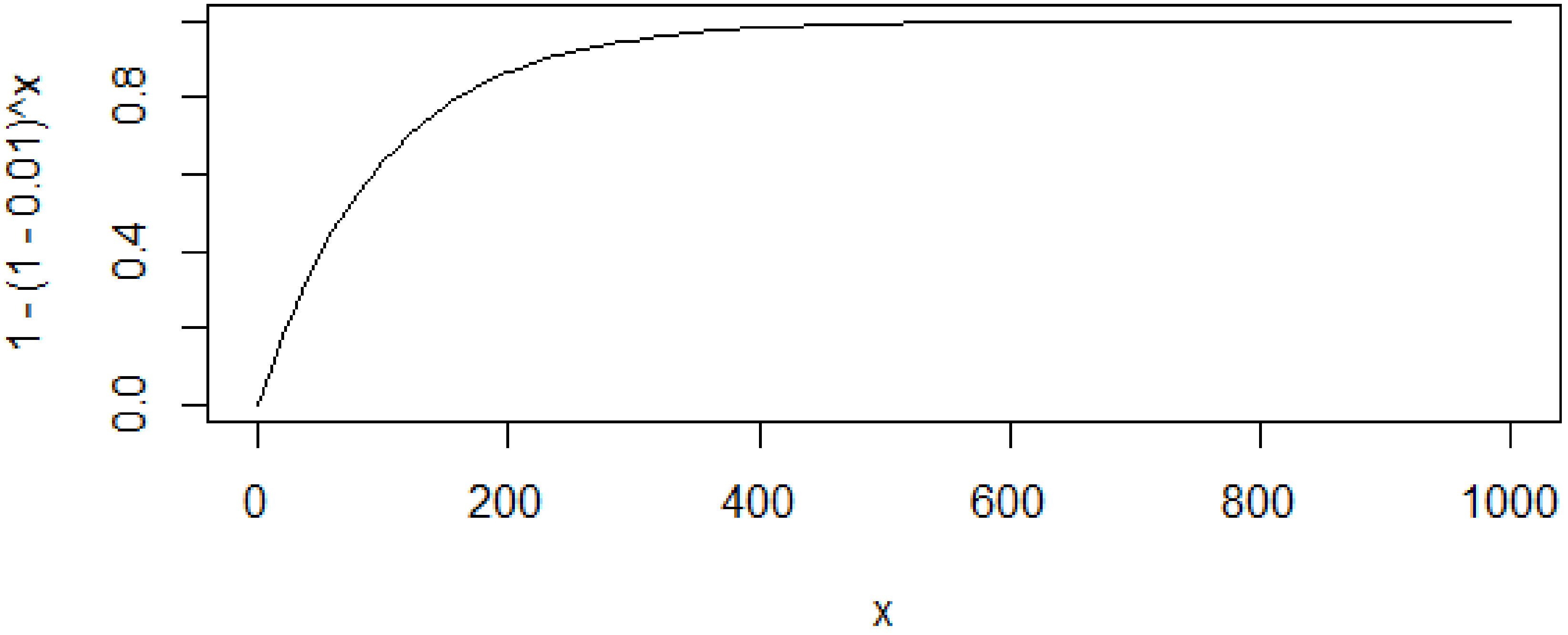
Multiple Comparison Test

- Now suppose we perform m independent hypothesis tests, each at a significance level α
- What is the probability of at least one false positive (Type I error) ?

Hypothesis Testing

Multiple Comparison Test

```
curve(1-(1-0.01)^x, xlim = c(0,1000))
```

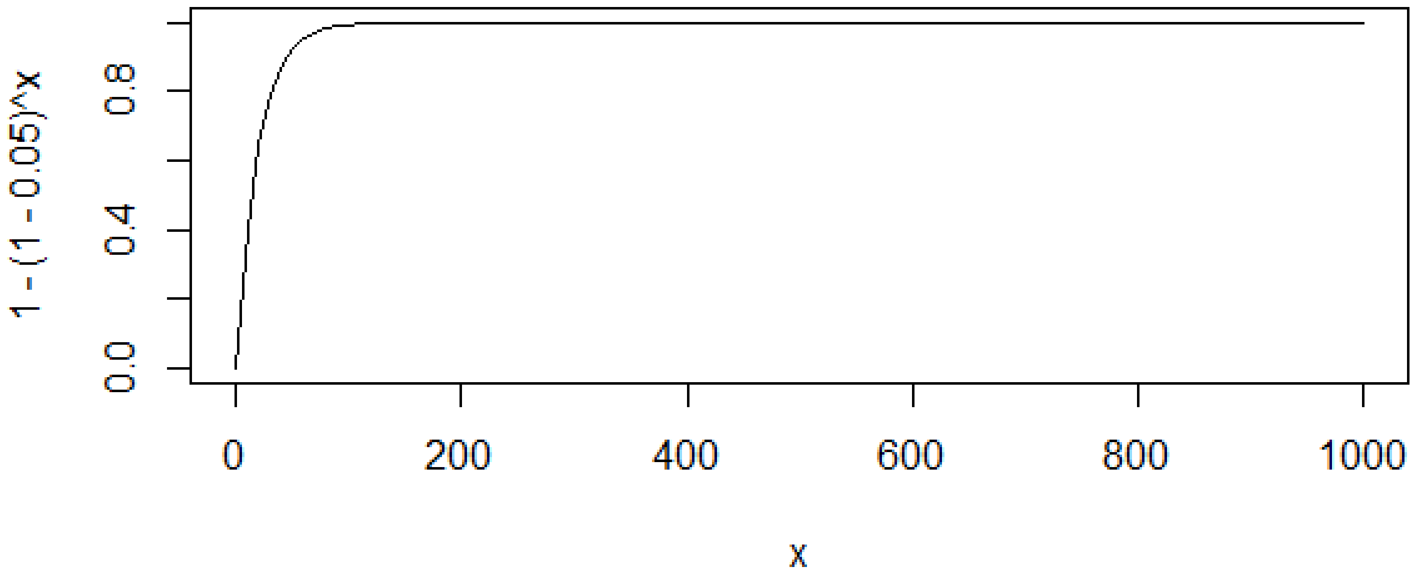


The false positive curve is smaller for 1% than for 5%

Hypothesis Testing

Multiple Comparison Test

```
curve(1-(1-0.05)^x, xlim = c(0,1000))
```



The false positive curve

Hypothesis Testing

Multiple Comparison Test

- The probability of no false positives in m independent tests is given by

$$(1 - \alpha)$$

$$\alpha = 5\%$$
$$1 - \alpha = 95\%$$

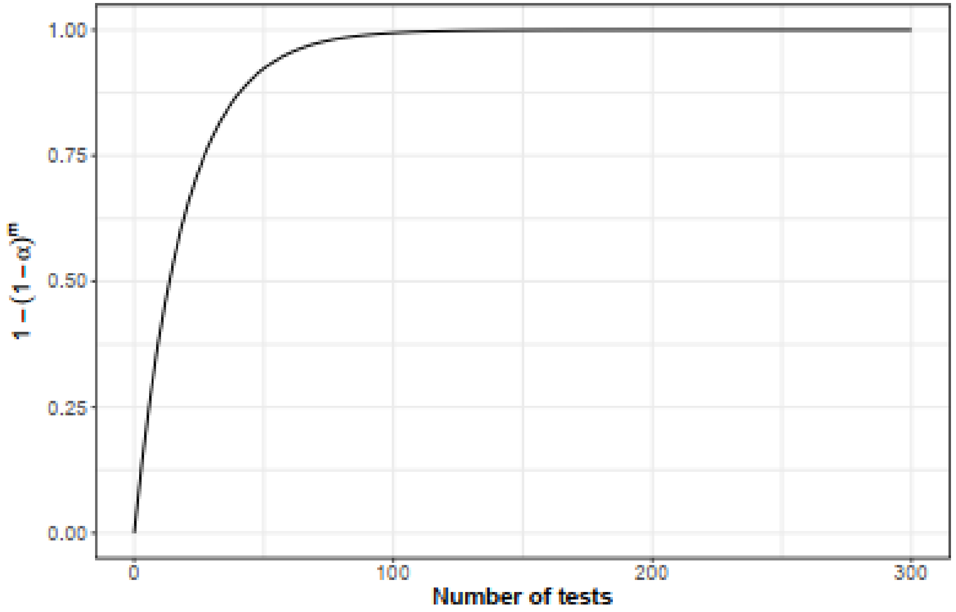
This means that there is a 95% probability of not making a type I error (false positive) when rejecting the null hypothesis

Hypothesis Testing

Multiple Comparison Test

- **Familywise Error Rate - FWER** is a statistical that controls the global type I error in a family of tests $1-(1-\alpha)^m$

m is the total number hypotheses tested



Hypothesis Testing

Multiple Comparison Test

- It is important to consider the **Familywise Error Rate (FWER)** when performing **multiple comparisons** because the chances of making a Type I error for a series of comparisons is greater than the error rate for a **separate comparison.**

Multiple Comparison Test

Controlling False Positives

- We need to use some strategy to control the occurrence of false positives (type I error) in multiple tests
- There are some ways to control false positives in a statistical test

Bonferroni Correction

- The Bonferroni test is a statistical procedure used to control false positives (type I error) when you perform multiple comparisons tests

Hypothesis Testing

Bonferroni Correction

The Bonferroni Method

- For n hypothesis tests with a significance level (α)

- It is possible to control the global type I error rate (known as Familywise Error Rate - FWER)

- Each test is individually compared with a value of $\alpha^* = \frac{\alpha}{m}$

    - α^* is the adjusted significance level for each individual test
    - α is the global significance level chosen to control type I error
    - m is the total number of tests performed

Hypothesis Testing

Bonferroni Correction

- $1 - \alpha$: probability that a type I error will not occur on a test
- $(1 - \alpha)^m$ probability of not having a type I error in m tests
- Note that we are assuming that the m tests are independent
- Then, $1 - \alpha^* = (1 - \alpha)^m$

  $$\alpha^* = 1 - (1 - \alpha)^m$$

Hypothesis Testing

Bonferroni Correction

- Example:

$m = 14$
$\alpha = 0.05$

What is the probability of at least one false positive in the 14 tests?

$$\alpha^* = 1 - (1 - \alpha)^m$$

Response: $\alpha^* = 0.51$

Bonferroni Correction

Terms

- When m is large, the Bonferroni correction may be overly conservative
- For example, for m = 30000 and α= 0.05, the value of α^* for individual tests is:

$$\alpha^* = \frac{\alpha}{m} = \frac{0.05}{30000} = 1.67e^{-6} = 0.00000167$$

False Discovery Rate

False Discovery Rate (FDR)

- Proposed as an alternative to control type I error
- An alternative to the Bonferroni correction is the **False Discovery Rate**
- Allows an acceptable false positive ratio
- More flexible and adaptable way than Bonferroni correction
- The Benjamini-Hochberg method is the most common method for controlling FDR

Hypothesis Testing

False Discovery Rate

False Discovery Rate (FDR): The Benjamini-Hochberg method

- Calculate the p-values for each hypothesis test
- Sort the p-values in ascending order
- For each p-value, calculate the FDR

False Discovery Rate

False Discovery Rate (FDR): The Benjamini-Hochberg method

$$FDR = \frac{(m/N)\ *\ p}{i}$$

m is the number of rejected tests with p-value less than or equal to p
N is the total number of tests
p is the current p-value being considered
i is the position of the p-value in the ordered list

Hypothesis Testing

False Discovery Rate

False Discovery Rate (FDR): The Benjamini-Hochberg method

- Define a FDR criterion, usually a value between 0.05 (5%) and 0.10 (10%)
- Identify all p-values that have an FDR less than or equal to the FDR criterion

Hypothesis Testing

Multiple Testing: False Discovery Rate

False Discovery Rate (FDR): The Benjamini-Hochberg method

- The Benjamini-Hochberg Procedure is a powerful tool that decreases the False Discovery Rate

- The Benjamini-Hochberg Procedure also helps you to avoid **Type I** errors (false positives)!

Example

False Discovery Rate (FDR): The Benjamini-Hochberg method

- Perform your hypothesis tests and calculate the p-values
- Suppose you have a list of p-values in a vector called p-values
- Order the p-values in ascending order

```
#use p.adjust function
```

Hypothesis Testing

Example

False Discovery Rate (FDR): The Benjamini-Hochberg method

```
# p_values list
p_values <- c(0.02, 0.03, 0.05, 0.07, 0.1, 0.01)
p_values

# Order the  p -values in ascending order
p_values_order <- sort(p_values)
p_values_order

# Calculate adjusted FDR using the BH method
fdr_asjusted <- p.adjust(p_values_order, method = "BH")
fdr_asjusted
```

Hypothesis Testing

Example

False Discovery Rate (FDR): The Benjamini-Hochberg method

```
> # p_values list
> p_values <- c(0.02, 0.03, 0.05, 0.07, 0.1, 0.01)
> p_values
[1] 0.02 0.03 0.05 0.07 0.10 0.01
> # Order the p-values in ascending order
> p_values_order <- sort(p_values)
> p_values_order
[1] 0.01 0.02 0.03 0.05 0.07 0.10
> # Calculate adjusted FDR using the BH method
> fdr_asjusted <- p.adjust(p_values_order, method = "BH")
> fdr_asjusted
[1] 0.060 0.060 0.060 0.075 0.084 0.100
```

- That is, they are the FDR values adjusted for each p-value

References

1. Baker, M. Statisticians issue warning on P values. *Nature* 531, 151 (2016).

2. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 57, 289–300 (1995).

3. Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. Nature methods, 12 (3), 179-185.