# APPLIED STATISTICS: Data Analysis

## VOLUME III: ADVANCED

Master data analysis with a simple and effective method that allows for fast results and autonomy in your studies

1. Simplified Essential Concepts
2. Illustrated step-by-step data analysis
3. The best free software for your analysis

LEARN STATISTICS EASILY

APPLIED STATISTICS: Data Analysis
3 VOLUMES
Master data analysis with a simple and effective method that allows for fast results and autonomy in your studies
★★★★★
LEARN STATISTICS EASILY
$27
3 VOLUMES INCLUDED
ADD TO CART
VISA
~~$47~~ $27

## MASTER DATA ANALYSIS QUICKLY, EFFORTLESSLY, AND WITH UNDENIABLE RESULTS

Discover our SIMPLE teaching method that will empower you to analyze your data on your own in no time.

We cover ALL the essential steps and only what's truly necessary for analyzing your data.

Built on the proven principle that it's entirely possible to accurately analyze data without complex concepts or formulas.

This book will serve you well, even if you have no prior knowledge of statistics. *All 3 volumes are included in this offer.*

**GET YOURS BY CLICKING HERE!**

## MORE INFORMATION ABOUT VOLUME III

The **Volume III** initiative is one-of-a-kind.

It allows us to incorporate topics suggested by our followers in each new edition.

And those who have already purchased the package will have lifetime access to these updates.

Volume III covers more advanced topics than those covered in the first two volumes.

As we plan to regularly update Volume III with new topics, we encourage you to send us your suggestions via our Instagram profile.

**@learnstatisticseasily**

# FORGET EVERYTHING YOU HAVE EVER LEARNED ABOUT STATISTICS TO ANALYZE YOUR DATA

Applied Statistics: Data Analysis (Volume III - Advanced) / Alves, Ana. - Learn Statistics Easily, 2023. 133 p. 2nd ed.

1. Statistics. 2. Biostatistics. 3. Learn Statistics Easily. 4. Applied statistics. 5. Data analysis. 6. Inferential statistics. 7. Descriptive statistics. 8. Graphs. 9. Sample size.

Learn Statistics Easily
statisticseasily.com

# PREFACE

Welcome to your ultimate guide to unlocking the power of data analysis – quickly, easily, and confidently.

This book presents a groundbreaking teaching method that empowers you to independently analyze your data with speed and precision.

We've distilled the essentials, providing only the necessary information to conquer data analysis without getting lost in complexities.

Say goodbye to intimidating concepts, formulas, and tables. This guide is designed to benefit you, even if your knowledge of statistics is limited.

Our innovative approach to "*learning data analysis quickly, easily, independently, and with confidence*" sets this book apart from the rest.

Let this guide be your invaluable companion as you embark on the exciting data analysis journey.

# MASTERING OUR METHODOLOGY

**(a)** We distill only the most vital concepts, making them effortlessly understandable.

**(b)** Crystal-clear examples and diagrams bring each concept to life.

**(c)** Our algorithm for selecting statistical analyses and graphs is streamlined and straightforward.

**(d)** We tackle the most prevalent statistical analyses, covering 99% of real-world scenarios.

**(e)** Our detailed, step-by-step instructions, paired with vivid illustrations, make data analysis a breeze to grasp.

**(f)** Experience the ultimate in user-friendly, comprehensive, and intuitive free statistical software.

# SUMMARY

1. **GETTING STARTED: ESSENTIAL KNOWLEDGE**
Grasp the key concepts in a simplified and accessible manner.

2. **TOP FREE STATISTICAL SOFTWARE**
Discover premier tools for data analysis, graphing, spreadsheets, and sample size calculations.

3. **DESCRIPTIVE STATISTICS: SUMMARY MEASURES**
Dive into the most vital measures for summarizing and showcasing your data.

4. **INFERENTIAL STATISTICS: UNLEASHING DATA ANALYSIS**
Learn to choose the right analysis and apply it with precision.

5. **PICKING THE PERFECT GRAPH (VOL. II)**
Follow a step-by-step guide to selecting and creating the ideal graph for your data.

6. **BONUS CONTENT & ADVANCED TOPICS (VOL. III)**
Delve into extra tips and explore slightly more sophisticated subjects.

# TABLE OF CONTENTS - VOLUME III

## 1 CALCULATING SAMPLE SIZE

## 2 HOW TO PERFORM A SAMPLE

## 3 EXTRA ADVANCED TOPICS

## 4 ADVANCED INFERENTIAL ANALYSIS

# TABLE OF CONTENTS - VOLUME III

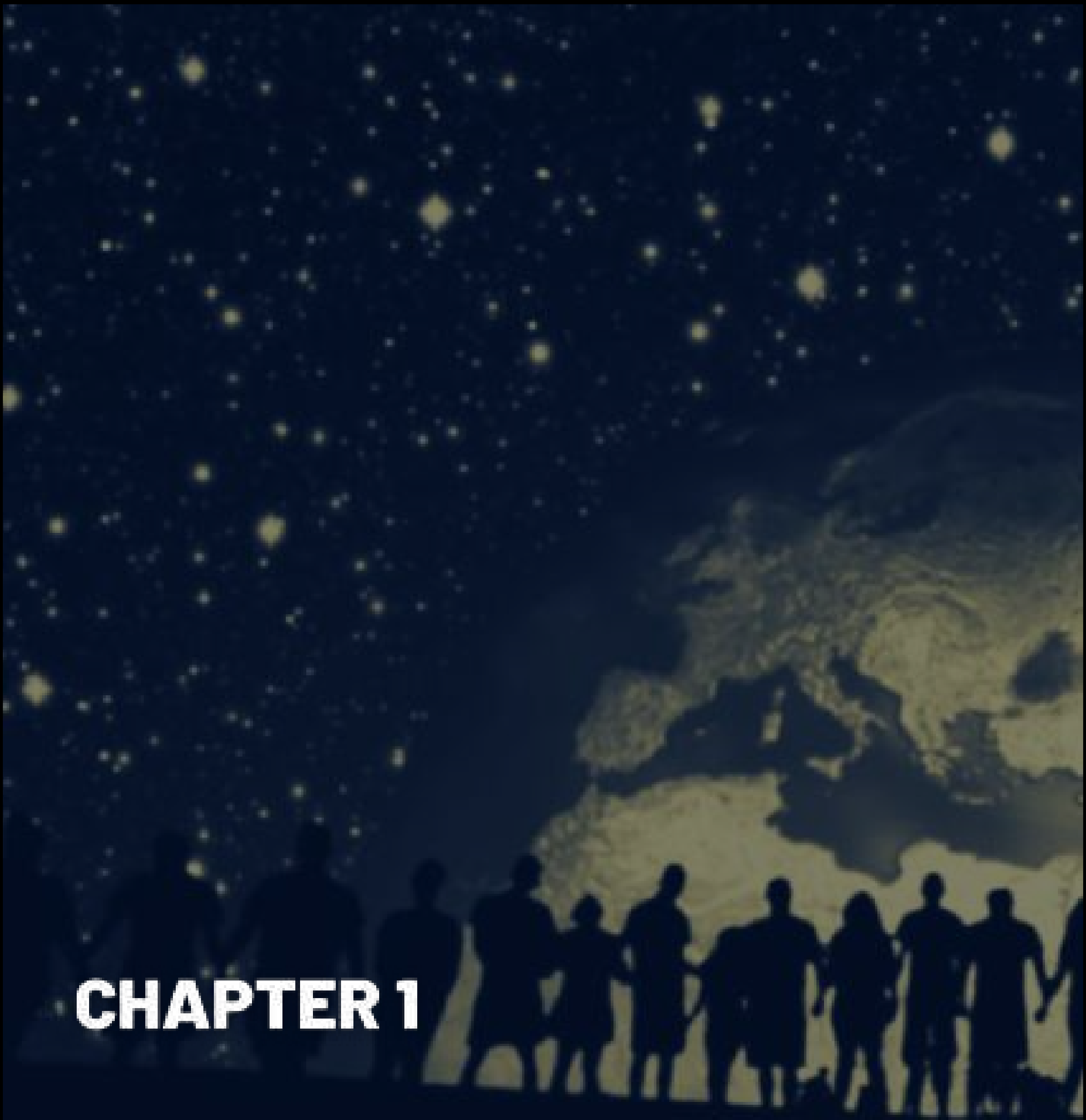
## CHAPTER 1: CALCULATING SAMPLE SIZE

### LOOKING FOR DIFFERENCES BETWEEN UNPAIRED GROUPS

### LOOKING FOR DIFFERENCES BETWEEN PAIRED GROUPS

### LOOKING FOR RELATIONSHIPS BETWEEN VARIABLES

> “MUCH TO LEARN, YOU STILL HAVE
>
> YODA



## CHAPTER 1

# CALCULATING SAMPLE SIZE

# 1. WHY CALCULATE SAMPLE SIZE?

When conducting a **sample**, we collect data from only a portion of the elements that make up the population.

This is because collecting data from all elements in the population would be **time-consuming** and **expensive**.

Moreover, collecting all this data would be **unnecessary** if we follow the steps outlined here.

**Why is defining an appropriate sample size important?**

1. **An appropriate sample size results in a reduction of the sampling error.**
2. **Unduly large samples result in the wastage of time and money.**
3. **Excessively small samples yield unreliable results.**

Due to the **sampling error** — represented by the difference between the true population value and the sample result — a sample will never perfectly represent the population.

**SAMPLE ERROR = TRUE POPULATION VALUE - SAMPLE RESULT**

This error **hinders** our ability to infer population characteristics from the sample data.

Depending on its magnitude, this error can lead to **false** conclusions.

Therefore, in a more superficial analysis, the **smaller** the sampling error, the better.

How can we **reduce** the sampling error?

Firstly, we should always consider the best **cost-benefit** balance. Then, using an adequate sample size **calculated** beforehand, we should collect the data using an appropriate sampling **method** (**Volume III: Chapter 2**).

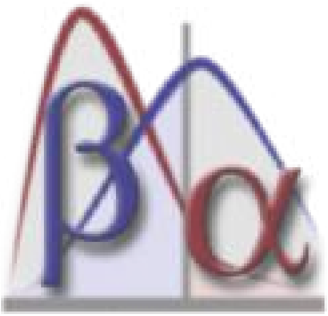
# 2. SAMPLE SIZE CALCULATION SOFTWARE

**G*Power:** Statistical Power Analyses is a powerful and free statistical tool for sample size calculation.

Aside from other functions, such as those related to test power and effect size, G*Power is the ideal software for computing the **sample size** of the tests we cover in this book, including t-tests, ANOVA, linear regression, and so on.

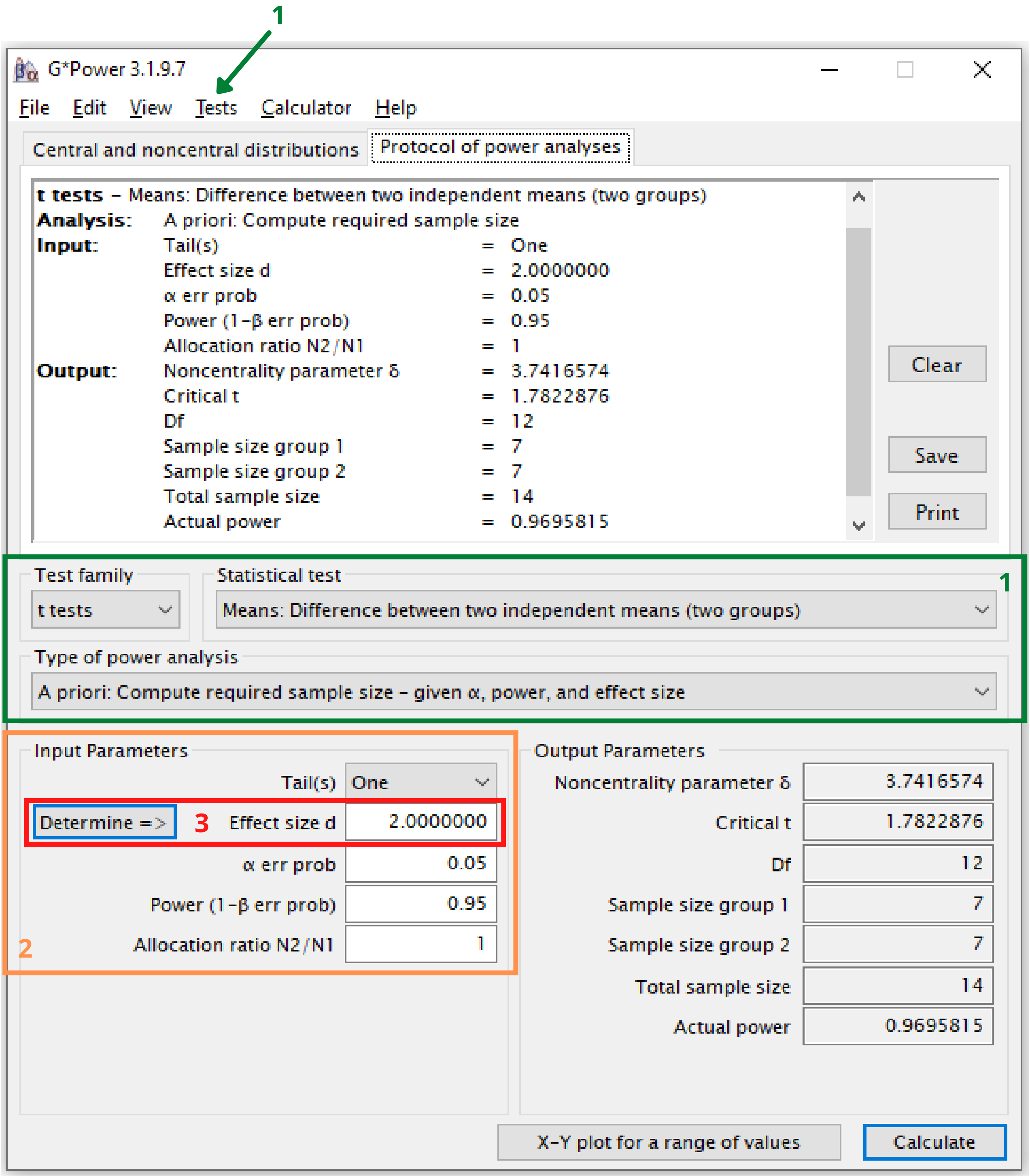
You can download G*Power from its official website.

On the website, scroll down until you find the **Download** section, choose the desired version, either Windows or macOS, and download and install G*Power.



Each sample size calculation in G*Power can be divided into three simple steps:

- 🟩 **1.** Selection of the appropriate test
- 🟧 **2.** Inputting the parameters
- 🟥 **3.** Estimating the effect size


1
G*Power 3.1.9.7
File
Edit
View
Tests
Calculator
Help
Central and noncentral distributions
Protocol of power analyses
Clear
Save
Print
1
2
Determine =>
3
X-Y plot for a range of values
Calculate

```
t tests – Means: Difference between two independent means (two groups)
Analysis:   A priori: Compute required sample size
Input:      Tail(s)                          =  One
            Effect size d                    =  2.0000000
            α err prob                       =  0.05
            Power (1–β err prob)             =  0.95
            Allocation ratio N2/N1           =  1
Output:     Noncentrality parameter δ        =  3.7416574
            Critical t                       =  1.7822876
            Df                               =  12
            Sample size group 1              =  7
            Sample size group 2              =  7
            Total sample size                =  14
            Actual power                     =  0.9695815
```

Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

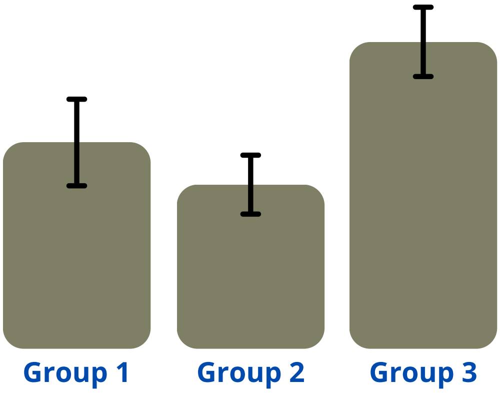
Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

| | |
|---|---|
| Tail(s) | One |
| Effect size d | 2.0000000 |
| α err prob | 0.05 |
| Power (1–β err prob) | 0.95 |
| Allocation ratio N2/N1 | 1 |

Output Parameters

| | |
|---|---|
| Noncentrality parameter δ | 3.7416574 |
| Critical t | 1.7822876 |
| Df | 12 |
| Sample size group 1 | 7 |
| Sample size group 2 | 7 |
| Total sample size | 14 |
| Actual power | 0.9695815 |



**CALCULATING THE SAMPLE SIZE**

(ARE THESE UNPAIRED GROUPS DIFFERENT?)

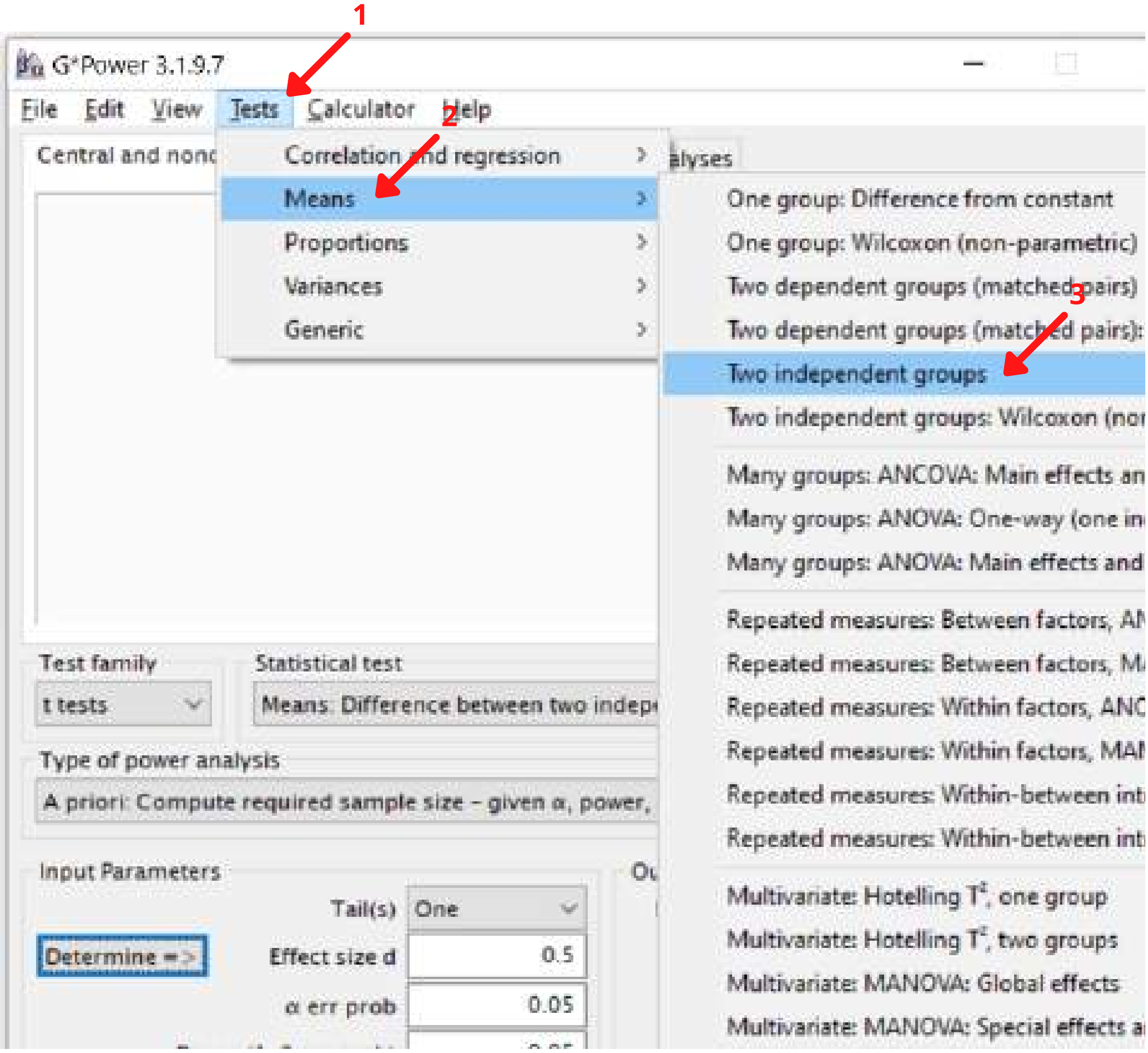
# 5 COMPARING UNPAIRED GROUPS

## 5.1 INDEPENDENT SAMPLES T-TEST

**Example:**
Test if there is a difference in height between male and female individuals in an indigenous tribe.

**Sample Size Calculation:**
From a pilot study, we obtained 4 observations from each of the 2 groups and then estimated their mean and standard deviation.

🟩 **1.** When opening G*Power, click on **Tests: Means: Two independent groups**.


1
2
3
G*Power 3.1.9.7
File
Edit
View
Tests
Calculator
Help
Central and nonc
alyses
Correlation and regression
Means
Proportions
Variances
Generic
One group: Difference from constant
One group: Wilcoxon (non-parametric)
Two dependent groups (matched pairs)
Two dependent groups (matched pairs):
Two independent groups
Two independent groups: Wilcoxon (no
Many groups: ANCOVA: Main effects an
Many groups: ANOVA: One-way (one in
Many groups: ANOVA: Main effects and
Repeated measures: Between factors, AN
Repeated measures: Between factors, MA
Repeated measures: Within factors, ANC
Repeated measures: Within factors, MAN
Repeated measures: Within-between int
Repeated measures: Within-between int
Multivariate: Hotelling T², one group
Multivariate: Hotelling T², two groups
Multivariate: MANOVA: Global effects
Multivariate: MANOVA: Special effects a
Determine =>
Ou

Test family: t tests

Statistical test: Means: Difference between two indep

Type of power analysis: A priori: Compute required sample size – given α, power,

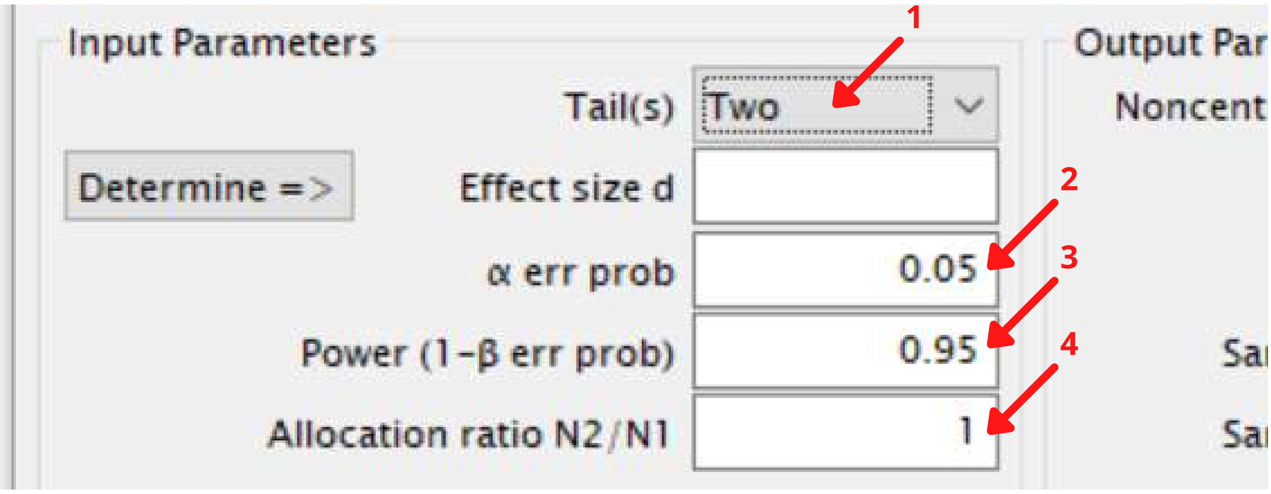
Input Parameters

Tail(s): One

Effect size d: 0.5

α err prob: 0.05

**2.** Now we will fill in the **Input Parameters** according to the following information:



Input Parameters

Tail(s) Two

Determine => Effect size d

α err prob 0.05

Power (1-β err prob) 0.95

Allocation ratio N2/N1 1

- **Tail(s):** Mark **One** if the test is one-tailed or **Two** if the test is two-tailed. But when will the test be one-tailed or two-tailed?

  If the alternative hypothesis (H1) is specific, for example, "*the mean of group 1 is greater than the mean of group 2,*" we use the one-tailed test.

  If the alternative hypothesis is general — without the initial distinction of greater or smaller — for example, "*the means are different between the groups,*" we use the two-tailed test.

  This hypothesis should be based on existing knowledge in the study field. If you are unsure, keeping the **Two**-tailed option is best.

- **α err prob (significance level):** The probability of rejecting the null hypothesis when it is true (Type I error).

  Usually, the values used are **0.05** or **0.01**. A significance level of 0.05, for example, indicates a 5% risk of concluding that there is a difference when there is no real difference.

- **Power (1 – β err prob) (test power):** The probability of rejecting the null hypothesis if false is how much the test controls Type II error.

  Usually, an acceptable value is between **0.80** and **0.99**.

  The higher the test power, the better, but as we increase it, the required sample size also increases.

- **Allocation ratio N2/N1:** If you want the sample size of the two groups to be the same, enter the value **1**.

  Enter the required value if you wish to have another allocation ratio between the groups.

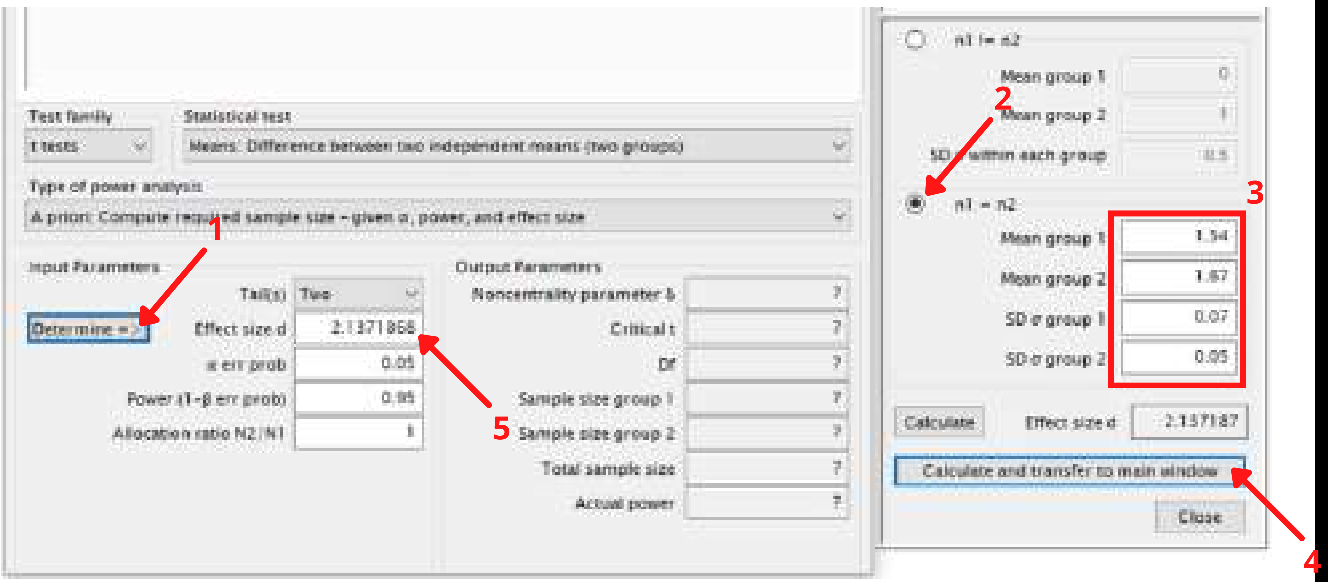
🟥 **3. Effect size d:** This measure represents the strength of the difference detected between the tested groups.

Conventionally, d = 0.20 is considered weak, d = 0.50 is considered medium, and d = 0.80 is considered strong.

To estimate d for our sample size calculation, click on **Determine =>**, and a new tab will open. In this new tab, we need to enter values for the **Mean** and **SD σ** (standard deviation), which we obtained from the pilot study, for each of the two groups.

These summary measures can be easily calculated according to the instructions in **Volume I: Chapter 3: Topic 5**.

After filling in, click on **Calculate and transfer to the main window**. The d value appropriate for your data will be calculated, and the **Effect size d** field will be filled in.



Test family: t tests

Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters

| | |
|---|---|
| Tail(s) | Two |
| Effect size d | 2.1371868 |
| α err prob | 0.05 |
| Power (1-β err prob) | 0.95 |
| Allocation ratio N2/N1 | 1 |

Output Parameters

| | |
|---|---|
| Noncentrality parameter δ | ? |
| Critical t | ? |
| Df | ? |
| Sample size group 1 | ? |
| Sample size group 2 | ? |
| Total sample size | ? |
| Actual power | ? |

○ n1 != n2

| | |
|---|---|
| Mean group 1 | 0 |
| Mean group 2 | 1 |
| SD σ within each group | 0.5 |

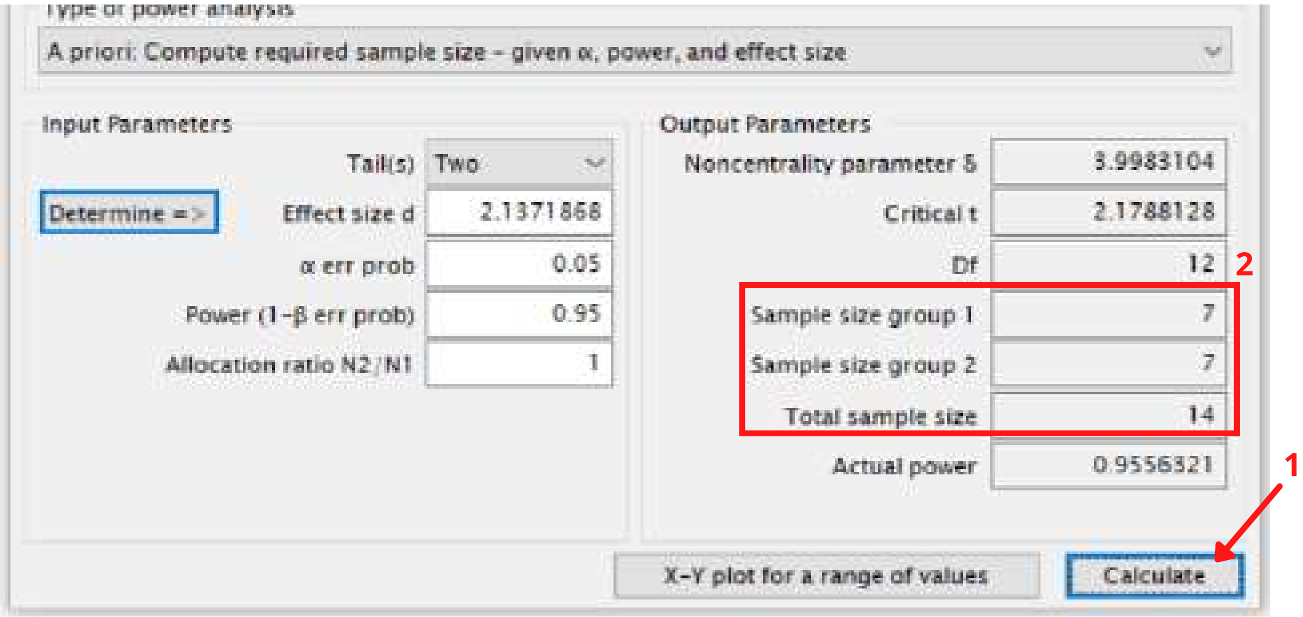
● n1 = n2

| | |
|---|---|
| Mean group 1 | 1.54 |
| Mean group 2 | 1.67 |
| SD σ group 1 | 0.07 |
| SD σ group 2 | 0.05 |
| Effect size d | 2.157187 |

**4.** With all the **Input Parameters** filled in, we need to click **Calculate** in the lower right corner to obtain the number of samples we need to collect.

In this case, our calculated total sample size was 14 elements, 7 in group 1 and 7 in group 2, which we must obtain randomly from the population.

Thus, we will need this sample size to detect this effect size; and, assuming it is correct, with a 5% probability of committing a Type I error and 95% probability of not committing a Type II error.



Type of power analysis: A priori: Compute required sample size - given α, power, and effect size

Input Parameters

| | |
|---|---|
| Tail(s) | Two |
| Effect size d | 2.1371868 |
| α err prob | 0.05 |
| Power (1-β err prob) | 0.95 |
| Allocation ratio N2/N1 | 1 |

Output Parameters

| | |
|---|---|
| Noncentrality parameter δ | 3.9983104 |
| Critical t | 2.1788128 |
| Df | 12 |
| Sample size group 1 | 7 |
| Sample size group 2 | 7 |
| Total sample size | 14 |
| Actual power | 0.9556321 |

# NOTES

**1.** How can we obtain the two groups' **mean** and **standard deviation** values to estimate the effect size?

Commonly, these values are **estimated** in two ways:

**(1)** through a pilot collection with a few samples or

**(2)** through other studies with similar populations, species, or conditions.

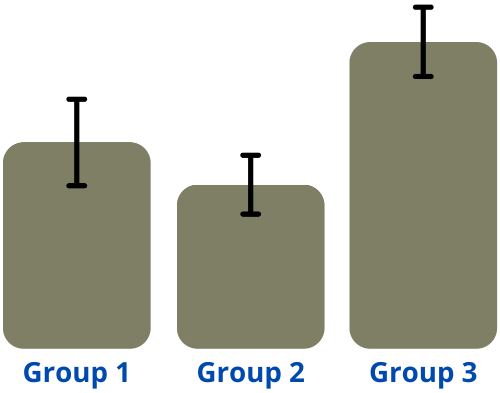
Sometimes, we may have reason to directly **assume** the expected effect size without **estimating** the mean and standard deviation summary measures.

**2.** To calculate the sample size for the **Mann-Whitney U test**, a non-parametric equivalent to the t-test:

Click on **Tests: Means: Two independent groups: Wilcoxon (non-parametric)** in the first step.

Keep the **Parent distribution** option as **Normal**.

The rest of the steps are the same.



CALCULATING THE SAMPLE SIZE

(ARE THESE PAIRED GROUPS DIFFERENT?)

# 6

# COMPARING PAIRED GROUPS

6.1 PAIRED SAMPLES T-TEST

6.2 ONE-WAY REPEATED MEASURES ANOVA

6.3 TWO-WAY REPEATED MEASURES ANOVA

6.4 MIXED-DESIGN ANOVA

## 6.1 PAIRED SAMPLES T-TEST

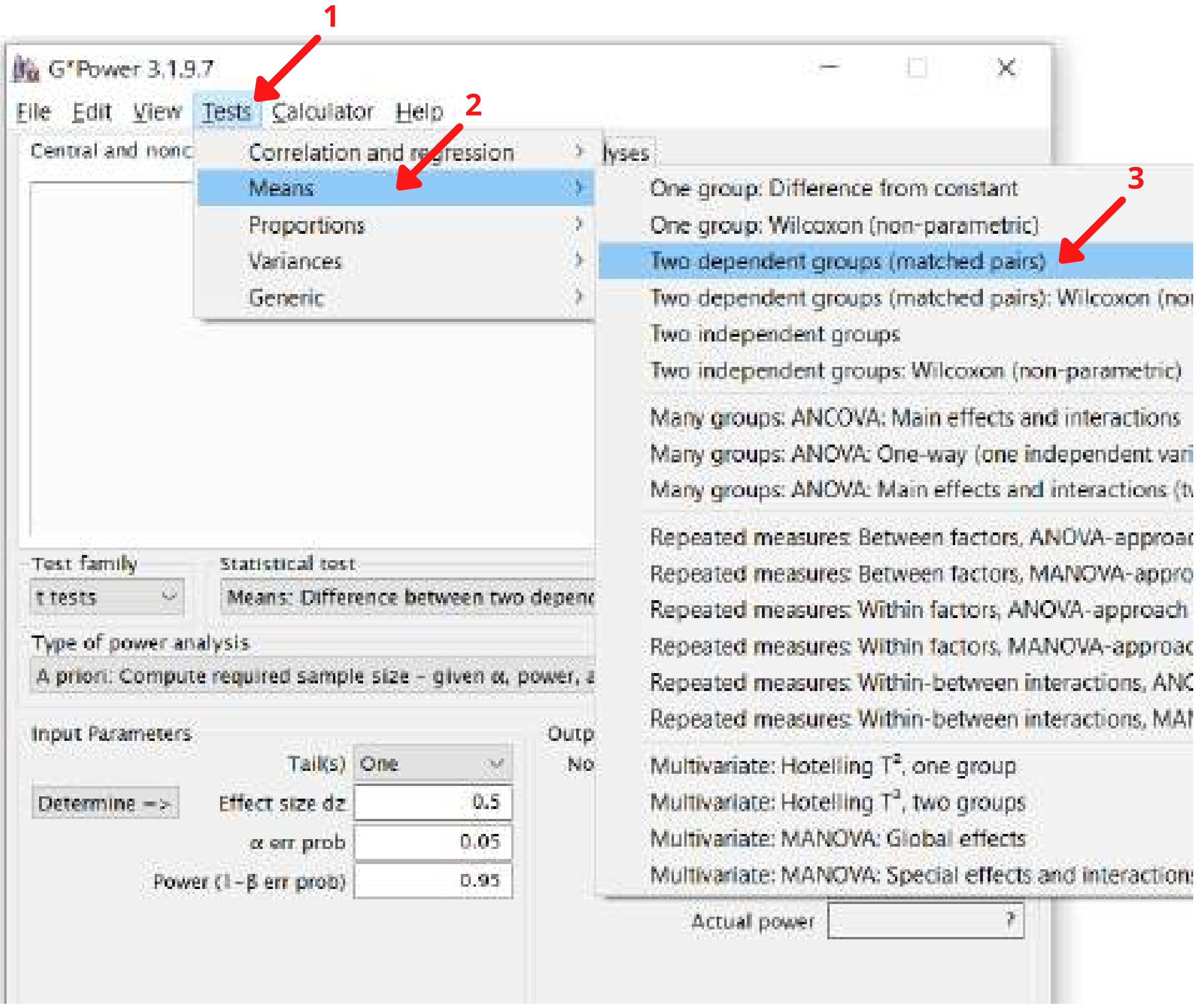
**Example:**

Does medication use affect older women's blood pressure (BP)? The BP of each participant will be measured twice, once before and once after the intervention (medication).

**Sample Size Calculation:**

From a pilot study, we obtained BP measurements of 4 older women before and after the intervention. We then calculated the data's means, standard deviations, and correlation coefficient.

**1.** To start, open G*Power and click on **Tests: Means: Two dependent groups (matched pairs)**.


1
2
3
G*Power 3.1.9.7
File
Edit
View
Tests
Calculator
Help
Central and nonc
lyses
Correlation and regression
Means
Proportions
Variances
Generic
One group: Difference from constant
One group: Wilcoxon (non-parametric)
Two dependent groups (matched pairs)
Two dependent groups (matched pairs): Wilcoxon (no
Two independent groups
Two independent groups: Wilcoxon (non-parametric)
Many groups: ANCOVA: Main effects and interactions
Many groups: ANOVA: One-way (one independent vari
Many groups: ANOVA: Main effects and interactions (t
Repeated measures: Between factors, ANOVA-approac
Repeated measures: Between factors, MANOVA-appro
Repeated measures: Within factors, ANOVA-approach
Repeated measures: Within factors, MANOVA-approac
Repeated measures: Within-between interactions, ANC
Repeated measures: Within-between interactions, MAI
Multivariate: Hotelling T², one group
Multivariate: Hotelling T², two groups
Multivariate: MANOVA: Global effects
Multivariate: MANOVA: Special effects and interaction
Input Parameters
Outp
No
Determine ->
Actual power
?

Test family: t tests

Statistical test: Means: Difference between two depend

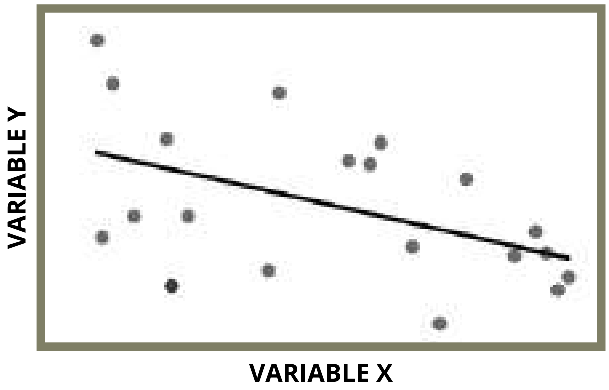
Type of power analysis: A priori: Compute required sample size - given α, power, a

Tail(s): One

Effect size dz: 0.5

α err prob: 0.05

Power (1-β err prob): 0.95



## CALCULATING THE SAMPLE SIZE

(ARE THESE VARIABLES RELATED?)

# 7

# LOOKING FOR RELATIONSHIP BETWEEN VARIABLES

7.1 PEARSON CORRELATION

7.2 SIMPLE & 7.3 MULTIPLE LINEAR REGRESSION

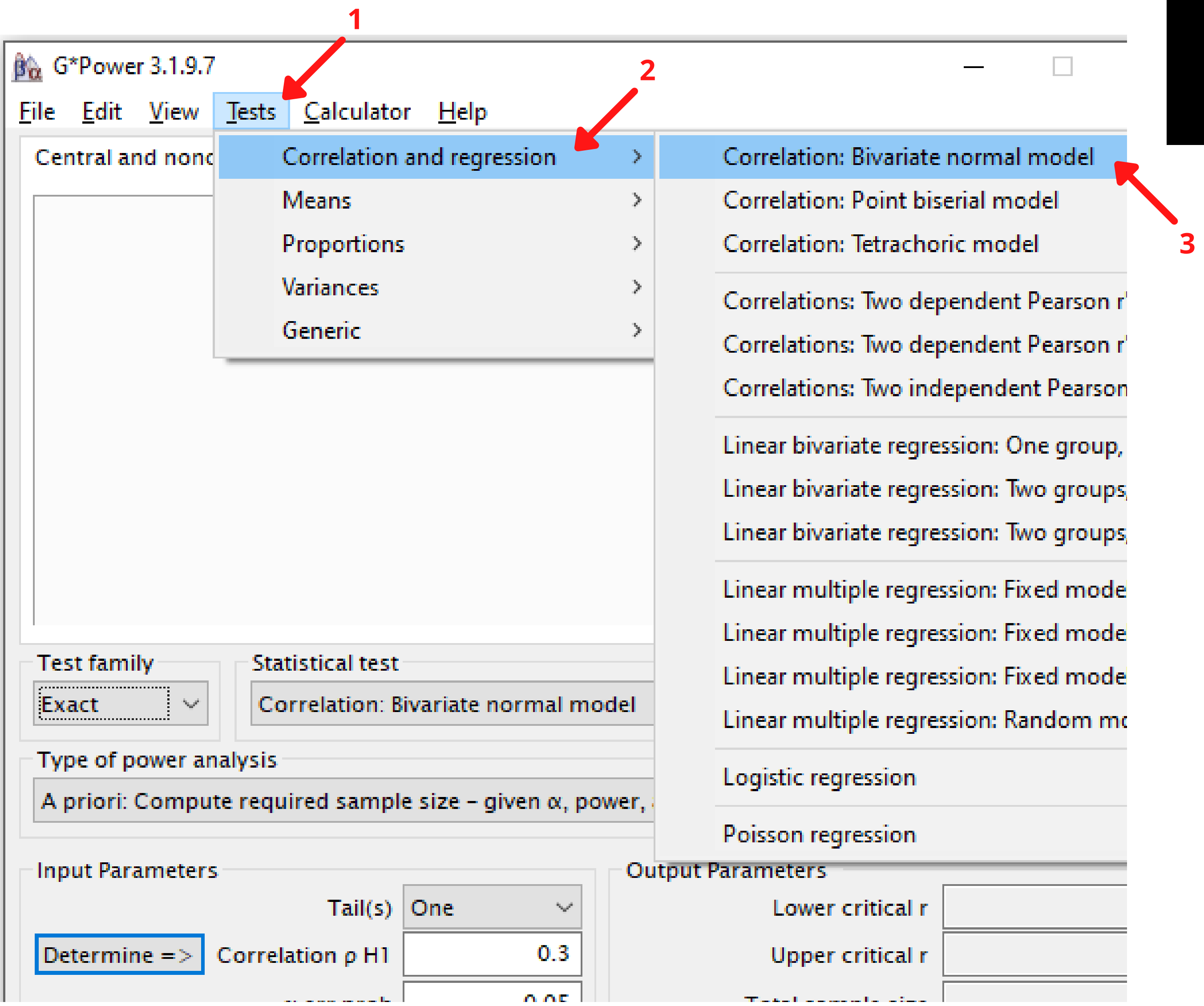
7.4 SIMPLE & 7.5 MULTIPLE LOGISTIC REGRESSION

## 7.1 PEARSON CORRELATION

**Example:**
Testing whether there is a correlation between the weight and height of members of an indigenous tribe.

**Sample Size Calculation:**
We obtained the weight and height of 6 individuals from a pilot study and estimated the correlation coefficient between these two variables.

🟩 **1.** Open G*Power and click on **Tests: Correlation and regression: Correlation: Bivariate normal model**.


1
2
3
G*Power 3.1.9.7
File Edit View Tests Calculator Help
Central and nonc
Correlation and regression
Means
Proportions
Variances
Generic
Correlation: Bivariate normal model
Correlation: Point biserial model
Correlation: Tetrachoric model
Correlations: Two dependent Pearson r'
Correlations: Two dependent Pearson r'
Correlations: Two independent Pearson
Linear bivariate regression: One group,
Linear bivariate regression: Two groups
Linear bivariate regression: Two groups
Linear multiple regression: Fixed mode
Linear multiple regression: Fixed mode
Linear multiple regression: Fixed mode
Linear multiple regression: Random mo
Logistic regression
Poisson regression
Input Parameters
Output Parameters
Determine =>
Lower critical r
Upper critical r
Total sample size

Test family: Exact

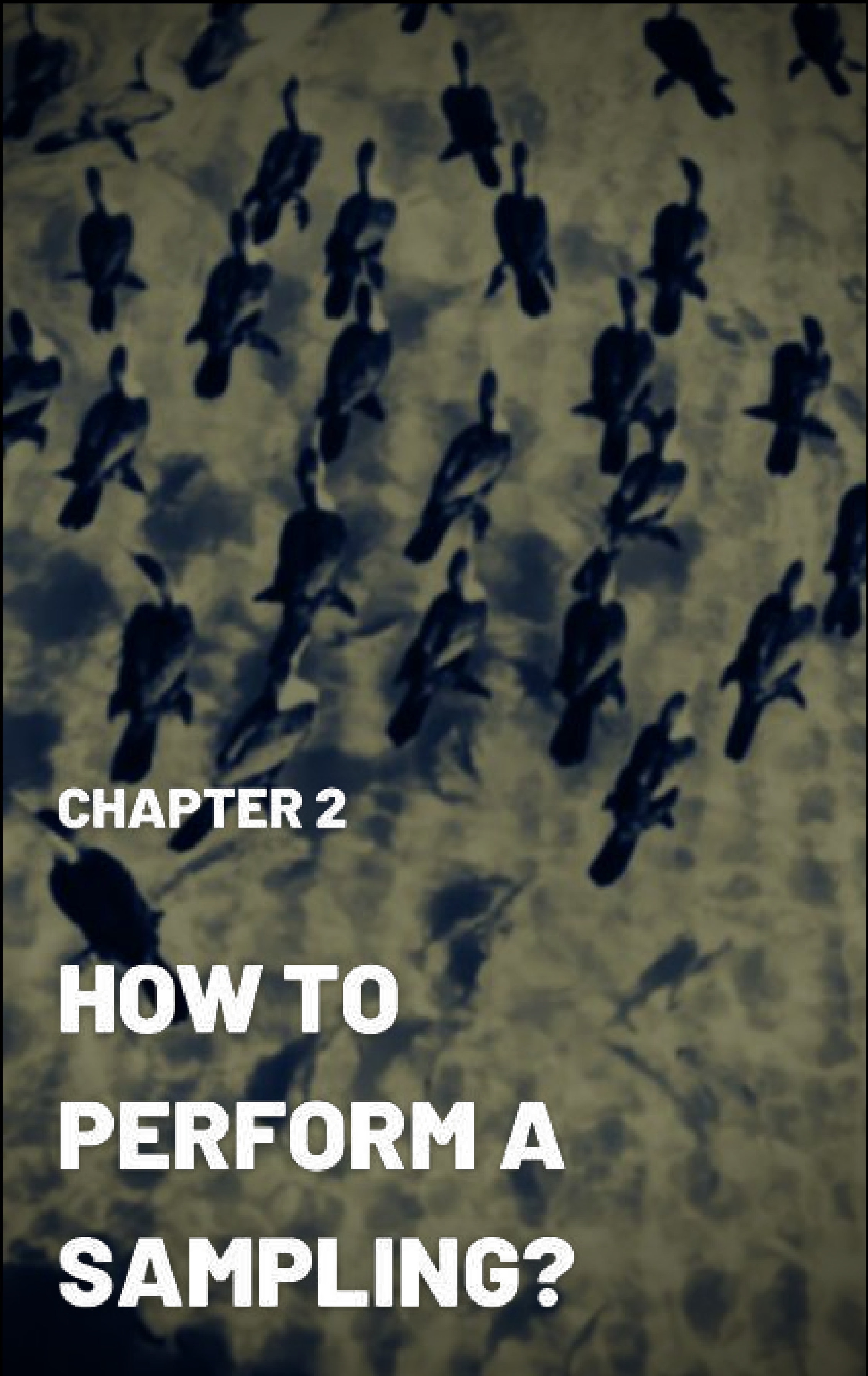
Statistical test: Correlation: Bivariate normal model

Type of power analysis: A priori: Compute required sample size – given α, power,

Tail(s): One

Correlation ρ H1: 0.3

α err prob: 0.05



## CHAPTER 2

# HOW TO PERFORM A SAMPLING?

## 1. WHY SHOULD WE USE AN APPROPRIATE SAMPLING METHOD?

We saw in the previous chapter that calculating the sample size is essential for us to have more **reliable** results in our analyses at the best possible cost-benefit.

However, suppose the sampling is **inadequately** done — in that case, it is not enough to calculate the sample size and obtain an adequate number of elements.

A **poorly** done sampling introduces biases in the sample, increases the sampling error, and leads us to draw false conclusions.

Sampling represents a process in which we **select** a part of the elements that make up a population to characterize it.

This **process** can be done in numerous ways.

Still, there are more **appropriate** ways to do it according to the situation.

Thus, we will discuss **sampling methods** and how we should conduct them correctly.

## 3. TYPES OF SAMPLING

There are two general **types of sampling**, non-probabilistic and probabilistic.

⚠ In **non-probabilistic sampling**, there is a deliberate and arbitrary selection of elements from the population.

It depends on subjective criteria of accessibility and researcher judgment.

They are called, for example, convenience sampling, and judgment sampling, among others.

Thus, for obvious reasons, we will not discuss these types of sampling here.

⚠ In **probabilistic sampling**, the selection of elements is entirely random, so each individual in the population has a known probability of being part of the sample.

It includes the most rigorous and widely used selection methods in scientific research.

"*We must not forget that sampling must always be **PRECISE** and **REPRESENTATIVE** of the population.*"

## 3.3 STRATIFIED RANDOM SAMPLING

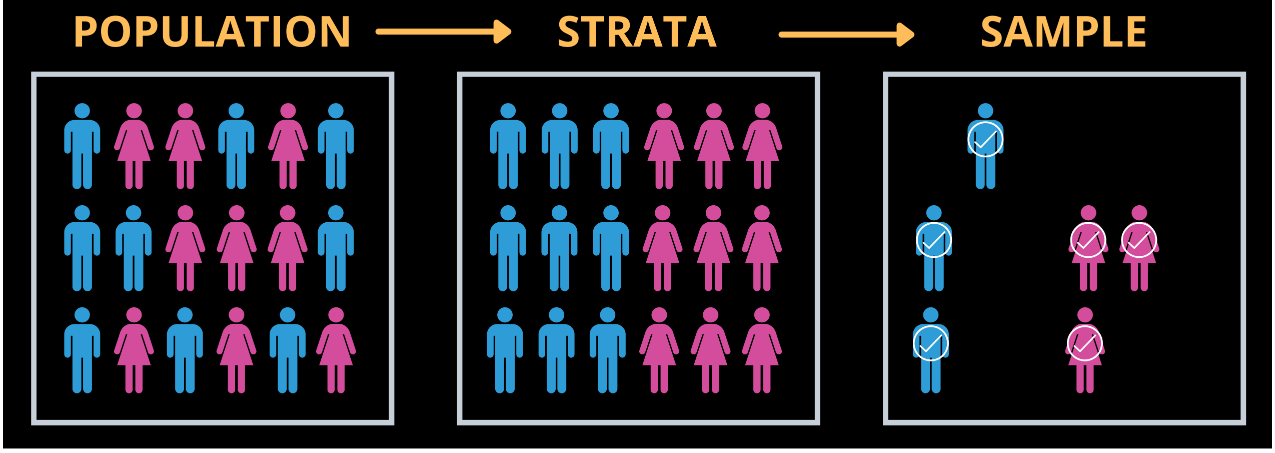
**Stratified sampling** is a probabilistic sampling method used when one or more criteria must be considered in the population before selecting elements. This method involves two steps:

**1.** Dividing the population into two or more groups (strata) based on the characteristic of interest. **2.** Randomly select elements from each stratum proportionally to its size.

This method allows for better control of the representativeness of **different** elements within the population.

## STRATIFIED
### RANDOM SAMPLING

**EXAMPLE:** To select 6 people from a population of 18, dividing the population into 2 strata based on sex. Then, randomly choose 3 men and 3 women from their respective strata, proportionally to their size.



## AND NEVER FORGET ABOUT

# SAMPLING BIAS

DRAWING CONCLUSIONS FROM NON-REPRESENTATIVE DATA OF THE TARGET POPULATION.

DO YOU PREFER DOGS OR CATS?
THE COOLEST DOG CONTEST

In 1948, a Chicago newspaper erroneously projected the next American president based on a phone survey. They did not consider that, at that time, only the upper class could have a telephone. The golden lesson is to always ensure that your sample represents the population.

## 5. HOW TO DO IT IN PRACTICE

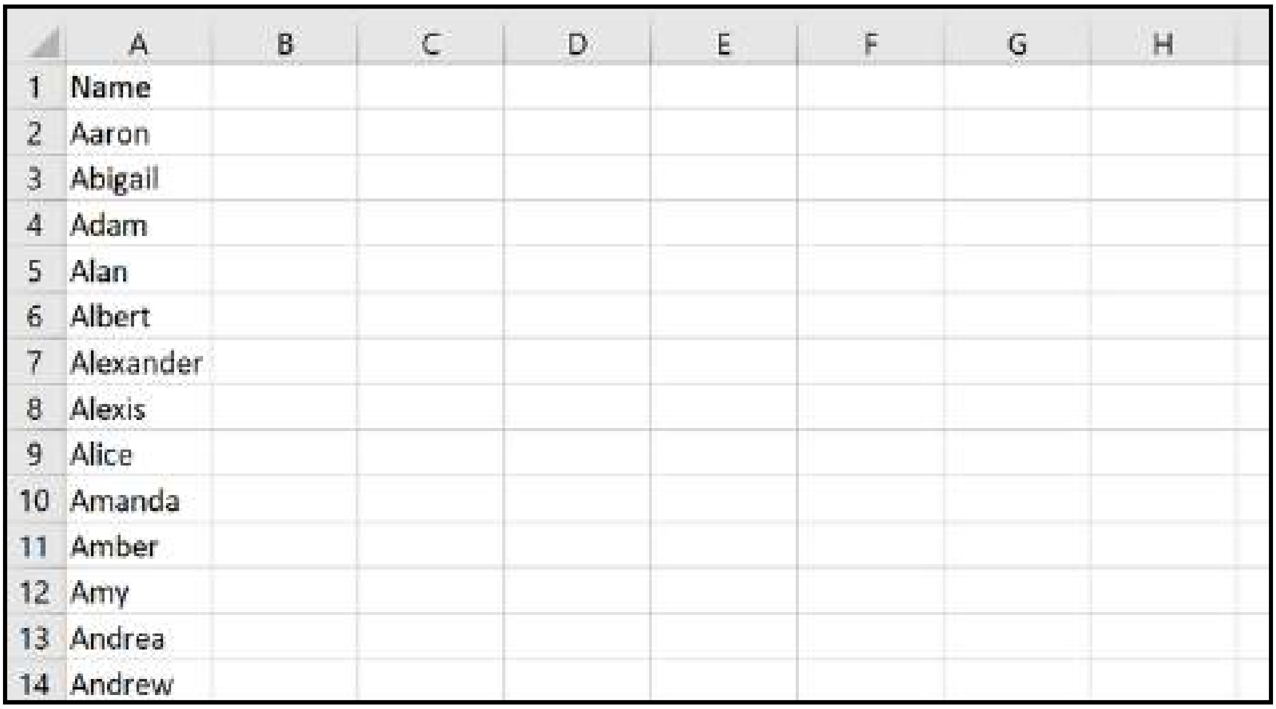
We could do the **drawings** the old-fashioned way, like using a bag or an urn in a bingo game. But obviously, we have more practical and faster possibilities for this nowadays.

So, first, we need each population element to have a unique **identification** and to be arranged in a spreadsheet. This identification can be any number, name, code, social security number, etc.

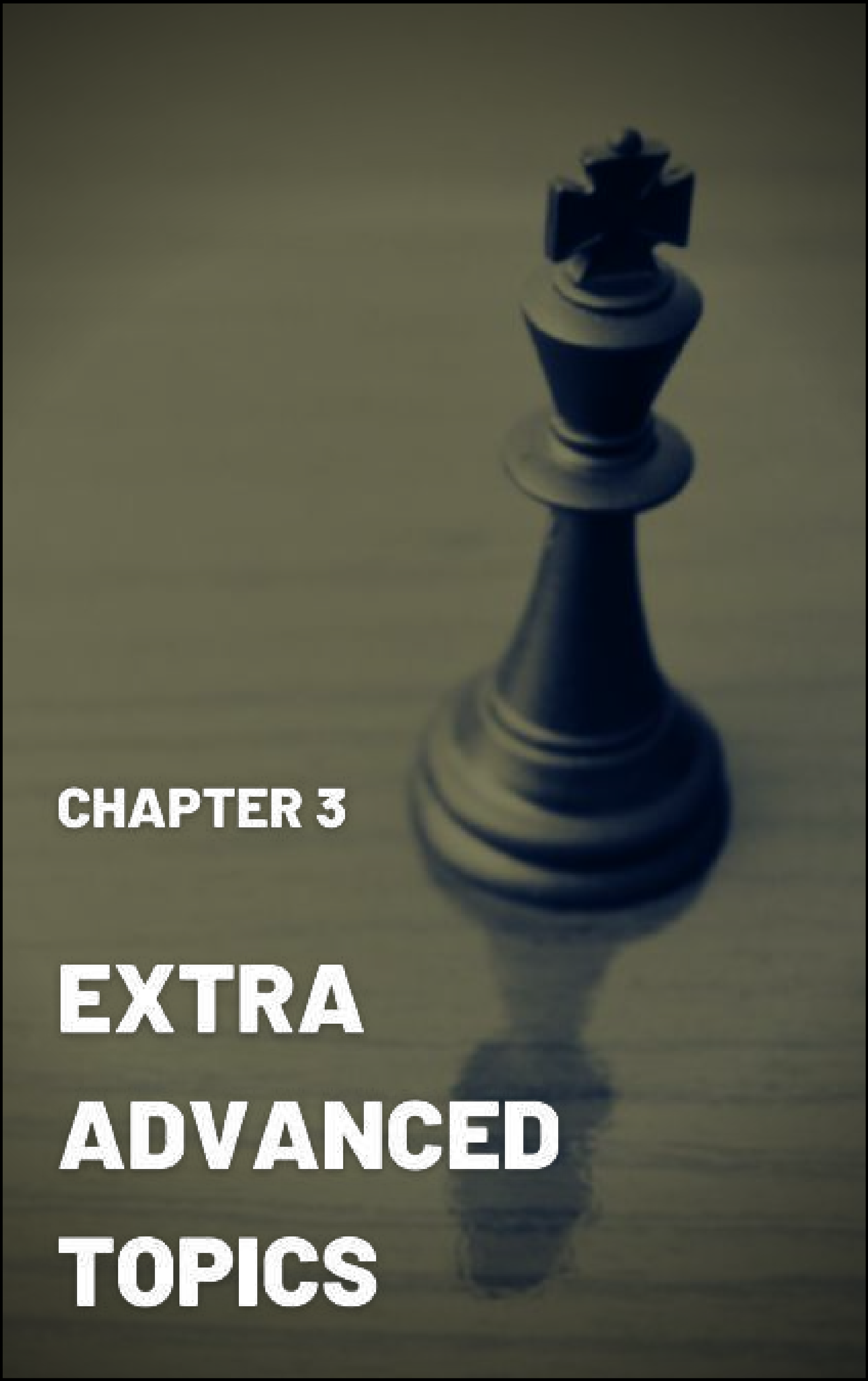
Remember that each **element** can be a person, an animal, a medical record, a quadrant, a country, a city, a forest, a lagoon, a Petri dish, a house, a purchase, etc.

For our **example**, let's suppose we have 200 names of patients treated in a clinic, and we want to sample 20 for research.

**1.** Here, we have this example with the patients' names in alphabetical order.



| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Name | | | | | | | |
| 2 | Aaron | | | | | | | |
| 3 | Abigail | | | | | | | |
| 4 | Adam | | | | | | | |
| 5 | Alan | | | | | | | |
| 6 | Albert | | | | | | | |
| 7 | Alexander | | | | | | | |
| 8 | Alexis | | | | | | | |
| 9 | Alice | | | | | | | |
| 10 | Amanda | | | | | | | |
| 11 | Amber | | | | | | | |
| 12 | Amy | | | | | | | |
| 13 | Andrea | | | | | | | |
| 14 | Andrew | | | | | | | |



## CHAPTER 3

# EXTRA ADVANCED TOPICS

# 1. P-VALUE: THE RETURN

In **Volume I**, we presented a didactic definition of p-value, but it does not accurately reflect its meaning.

Therefore, for initial contact with the topic, this **simplification** contributes to its understanding.

Thus, as promised, we present a **precise definition** of the p-value, which simultaneously requires greater attention and abstraction.

---

When we perform an inferential hypothesis test, such as chi-square, t-test, ANOVA, correlation, regression, etc., we basically have two hypotheses:

| **NULL HYPOTHESIS (H0)** | **ALTERNATIVE HYPOTHESIS (H1)** |
|---|---|
| **The standard, simplest, that there is '*no difference between the groups*' or '*no relationship between the variables*'.** | **An alternative, complementary state to H0, that there are '*differences between groups*' or '*relationships between the variables*'.** |

Thus, the fundamental objective of any hypothesis test is to define whether we will reject or not the null hypothesis (H0) — and this definition will depend on **two fundamental factors**:

Thus, in technical terms, the **p-value** can be precisely defined as:

> ***The p-value represents the probability of obtaining a result equal to (or more extreme than) the one obtained from our data, assuming that the null hypothesis is true.***

**If my test returned, for example, a p-value of 2%, what does this mean?**

If we consider H0 true, the probability of obtaining results equal to (or more extreme than) ours would be only 2%.

However, since it is lower than $\alpha = 5\%$, we reject H0.

See next for a more detailed explanation.

**$p < 0.05$**

***"We Are The Champions"***

It was a summer afternoon in Harpenden, **England**, in the early 1920s.

A group of **scientists** — a statistician, a phycologist, and a biochemist — had sat at a table for afternoon tea at the Rothamsted Experimental Station.

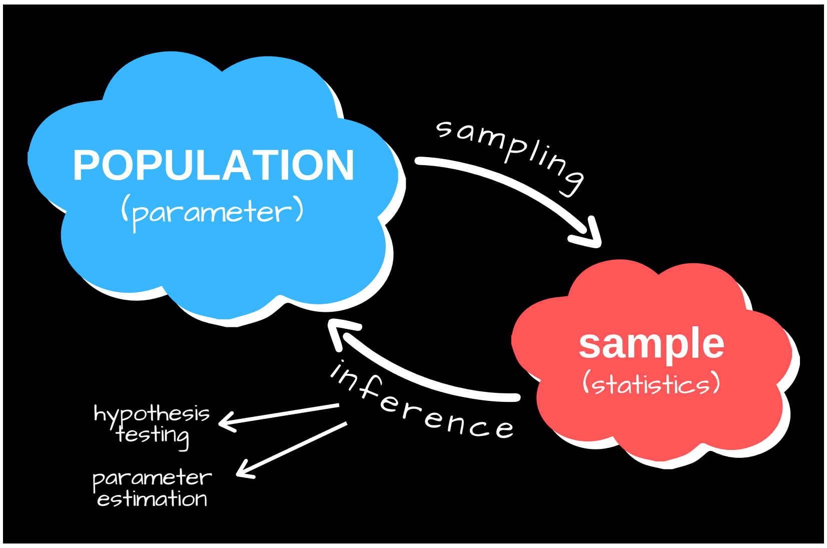
The phycologist, **Muriel Bristol**, insisted that tea poured over milk tasted different than when milk was poured over tea.

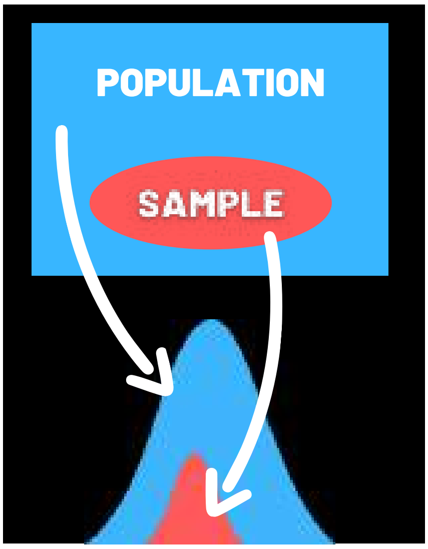
Everyone **questioned** her — *What could be the difference?*

They couldn't conceive that there could indeed be any **difference** in taste.

## 3. CONFIDENCE INTERVAL

Some concepts already discussed in **Volume I** are essential for understanding the confidence interval, such as population, sample, and sampling error.





A **population** encompasses all elements.

On the other hand, a **sample** includes only a portion of these elements.

**Inferential statistics** allow us to draw conclusions about the population through the sample.

However, since a sample represents only a part of the population, it will **never** be perfect.

Thus, the **sampling error** is born, represented by the difference between the true population value and the value obtained in the sampling.

Imagine we want to **estimate** the average weight of all oranges in an orchard.

The best way to do this is by **collecting** a random sample with an appropriate sample size and presenting the mean and the confidence interval.

What **factors** can influence the size of the confidence interval?

**(1)** The larger the **sample size**, the smaller the confidence interval.

**(2)** The lower the **variability** within the population, the smaller the confidence interval.

**(3)** The lower the **confidence level** (1 - α) — for example, 90, 95, or 99% — the smaller the confidence interval.

Now, back to the **oranges**.

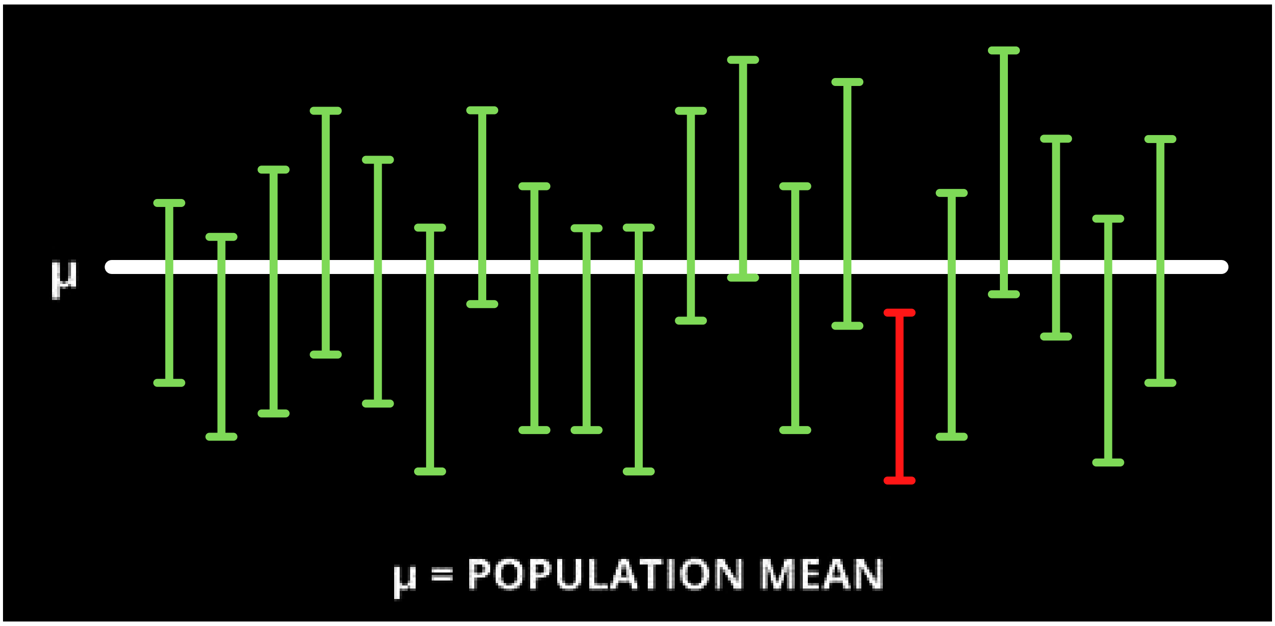
In our **sampling** of 10 oranges, we obtained a mean weight of 150 g with a confidence interval (95%) of 125 to 175 g.

***But what does this mean?***

To explain, we will perform 19 **other random samples**, each with 10 other oranges randomly selected from the orchard.

**Thus, 95% of the 20 random samples we made will contain the population mean of the weight of the oranges in the orchard.**

See the **graph** with an example of 20 samples, where only one (5%) does not contain the population mean, as expected.



# WAIT!

# IT`S NOT OVER YET!

MORE CONTENT WILL BE INCLUDED IN THIS VOLUME. IT WILL BE UPDATED MORE TIMES! VISIT OUR INSTAGRAM PROFILE AND MAKE TOPIC SUGGESTIONS.

@learnstatisticseasily

APPLIED STATISTICS: Data Analysis
3 VOLUMES
Master data analysis with a simple and effective method that allows for fast results and autonomy in your studies
$27
3 VOLUMES INCLUDED
ADD TO CART
~~$47~~ $27

**MASTER DATA ANALYSIS QUICKLY, EFFORTLESSLY, AND WITH UNDENIABLE RESULTS**

Discover our SIMPLE teaching method that will empower you to analyze your data on your own in no time.

We cover ALL the essential steps and only what's truly necessary for analyzing your data.

Built on the proven principle that it's entirely possible to accurately analyze data without complex concepts or formulas.

This book will serve you well, even if you have no prior knowledge of statistics. *All 3 volumes are included in this offer.*

**GET YOURS BY CLICKING HERE!**

**Unlock the secrets to analyzing your data swiftly, effortlessly, and confidently.**

**Our SIMPLE approach focuses on teaching you precisely what you need to know to conquer data analysis.**

**Leave behind the complexities of concepts, formulas, and tables – this course proves that accurate data analysis is achievable for everyone.**

**This accessible resource is tailored for those with little or no prior knowledge of statistics.**

**Discover our unparalleled method for "*fast, easy, and confident data analysis*" – a game-changer you won't find anywhere else.**

LEARN STATISTICS EASILY