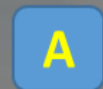


Cosmina Bondor

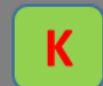
Review



ALWAYS



SEEK



KNOWLEDGE

Outline

About the theoretical test

About the final mark

30 Examples of questions for the theoretical test

- Recapitulation

period December - 15.January

The evaluation of the didactic activity - sem I

- student receives a **link** in the institutional e-mail box (...@elearn.umfcluj.ro),

Complete two forms to evaluate:

- (1) course
- (2) practical work

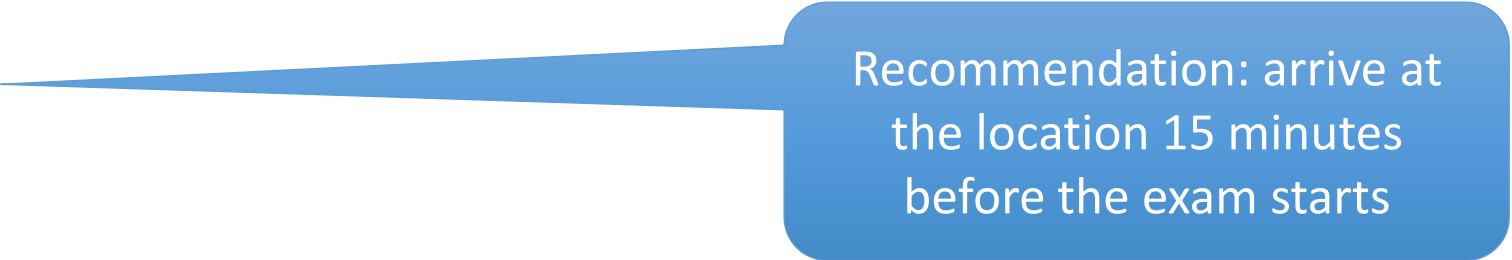
To access the institutional e-mail box (@elearn.umfcluj.ro) use the Outlook e-mail service (<https://outlook.office.com/mail/>) and enter the data used for Microsoft Teams.

- **The evaluations are confidential, the answers provided cannot be associated with the students' personal data.**
- If you encounter difficulties in carrying out the evaluations, or you want to send us aspects that need changes or additions, please send a message to prorectoratevaluate@umfcluj.ro

Theoretical exam

10.02.2025

8 o'clock



Recommendation: arrive at
the location 15 minutes
before the exam starts

Manta

Preda

Dermatologie

**English
language 2**

Must have at the theoretical/practical exam

- ID or Passport
- Student ID
- without phone (in the bag, close)
- without smart watch, tablet, laptop, etc. (in the bag, close)
- pen



Optional to have at the theoretical/practical exam

- office calculator
- white paste



Article 78. Students are allowed to attend the examination for a specific discipline only once per each examination session.

Article 79. (1) Students are allowed to attend the examination for a specific discipline no more than three times during an academic year.

- (5) Absence or non-admission to an examination until the end of the semester or module, results in the loss of one possibility of taking that exam.
- (8) Students who have an excuse for their absence from the examination, approved by the Council Board of the Faculty, benefit from all the possibilities to submit that examination.

- Article 108. (1) On leaving the classroom, students must hand the written papers and all their signed sheets.
- (2) Upon completion of the written exam, students must sign for the delivering of their written paper.
- (2) A discipline is promoted when the final mark, from both theoretical and practical exam, is at least 5.
- (5) If the student takes the first task of the exam but is absent from the second, he/she gets the minimum score and is declared failed.

- Students who didn't took the practical exam **cannot** enter to the theoretical exam.
- The students, who pass practical exams, are not obliged to take that part of the exam once again in the same academic year.
- The students who have **difference exams (transfer from another faculty)** should plan the exam with the teacher and also pay the university fee for the exam.

Varianta 1

1) The cholesterol level on a sample of 600 patients was tested for normality. Chi-Square goodness of fit test was used and a p-value of 0.0004 was obtained. The following is correct:

- a) The comparison of the sample mean among a population mean is proper
- b) Data are not normally distributed
- c) Kolmogorov-Smimov is the proper test
- d) The Chi-Square goodness of fit is not the proper test to be used
- e) Data are normally distributed

4) The values of systolic blood pressure measured in mmHg for a sample of 10 patients are as follows: 120, 100, 110, 120, 130, 160, 130, 120, 140, and 160. The arithmetic mean is equal to:

- a) 110
- b) 130
- c) 129
- d) 135
- e) 120

Catedra de Informatică Medicală

Facultate, an, grupă: _____

Numele: _____

Prenumele: _____

Data examenului: _____ (completați cu majuscule)

- | | A | B | C | D | E |
|----|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 2 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 3 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 4 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 5 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 6 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 7 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 8 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 9 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 10 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 11 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 12 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 13 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 14 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 15 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 16 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 17 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 18 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 19 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 20 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 21 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 22 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 23 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 24 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 25 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 26 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

- | | A | B | C | D | E |
|----|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 51 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 52 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 53 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 54 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 55 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 56 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 57 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 58 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 59 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 60 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 61 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 62 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 63 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 64 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 65 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 66 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 67 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 68 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 69 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 70 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 71 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 72 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 73 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 74 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 75 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 76 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

- | | A | B | C | D | E |
|-----|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 101 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 102 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 103 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 104 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 105 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 106 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 107 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 108 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 109 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 110 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 111 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 112 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 113 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 114 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 115 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 116 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 117 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 118 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 119 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 120 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 121 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 122 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 123 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 124 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 125 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 126 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Varianta 1

Pressure measured in mmHg
at four different times: 120, 100, 110, 120,
the arithmetic mean is equal

Varianta chestionar
de examinare

- 1
- 2
- 3
- 4



ABOUT theoretical test

- Theoretical test (90 minutes):
- 35 questions each worth 1 point
 - Multiple choice questions
 - Five possible response (A, B, C, D & E);

Question with * - only one correct answer

About Exam - Exam Policy

- Questions with one correct answer:

- 5 concordances = 1 point
- < 5 concordances = 0 point

- Questions with 2 correct answers :

- 5 concordances = 1 point
- 4 concordances = 0.8 point
- < 4 concordances = 0 point

17.* The 99% confidence interval ($Z_{\alpha} = 2.579$) associated to systolic blood pressure for a sample of 169 persons with an arithmetic mean of 135 mg/dl and a standard deviation of 20 mg/dl is:

A. [132 - 138]

B. [132 - 138]

C. [83 - 187]

D. [131 - 139]

E. Could not be determined based on provided data

Only one correct !

16) Let be a statistical series with the following data: 40, 60, 20, 20, 60, 80, 80, 40, 60, and 80. The relative frequency of 0.3 corresponds to:

a) 20

b) 40

c) 60

d) 80

e) None is correct

Two correct answers !

About Exam - Exam Policy

- Questions with > 2 correct answers:
 - 5 concordances = 1 point
 - 4 concordances = 0.8 point
 - 3 concordances = 0.3 point
 - < 3 concordances = 0 point

7) The following data represent the age of first episode of myocardial infarction on a series of male patients: 38, 50, 23, 45, 70, 33, 25, 40, 50, 62, and 59. The values of quartiles are as follows: $Q1 = 35.5$, $Q2 = 45$ and $Q3 = 54.5$. The following statements are true:

a) $Q2 - Q1 = 9.5$

b) $Q3 - Q2 = 9.5$

c) Data are asymmetrical distributed

d) Data are symmetrical distributed

e) Data are approximately symmetrical distributed

>2 answers correct !

Type of activity	Evaluation criteria	Type	Mark	Weight in final mark
Lecture	Theoretical exam	Compulsory	1-10	70%
Practical activity	Practical exam	Compulsory	1-10	30%
Class activity	Optional		0.6	

Exam – 35 questions with multiple choice

Maximum points 35 = Maximum mark 10.

If the obtained **points** ≥ 13.7 then

- Mark at theoretical test = $1 + \text{Obtained point} * 9/35$
- If the obtained points < 13.7 then
 - Mark at theoretical test = 4

If theoretical test **mark** ≥ 5 and practical test **mark** ≥ 5 than

- weighted average = $0.70 * \text{theoretical test mark} + 0.30 * \text{practical test mark}$
- Final mark = weighted average + $0.2 * \text{no. of Homework with mark higher than 5}$

Example

- 10 mark at practical test
- 18 points $\rightarrow 1 + 18 \cdot 9 / 35 = 5.63$ mark at theoretical test
- Weighted average = $0.7 * 5.63 + 10 * 0.3 = 6.94$
- 3 homeworks with marks > 5 meaning $+0.6$ at final mark

- Final mark $6.94 + 3 * 0.2 = 7.54 = 8$

Example

- 10 at practical test
- less than 13.7 points → 4 at theoretical test
- Final mark 4 (have to give only theoretical test in re-taken)

Example

- 4 at Practical test
- cannot enter to theoretical test
 - final mark 4 (have to give both tests in re-taken)

If you do not pass, no problem you can come

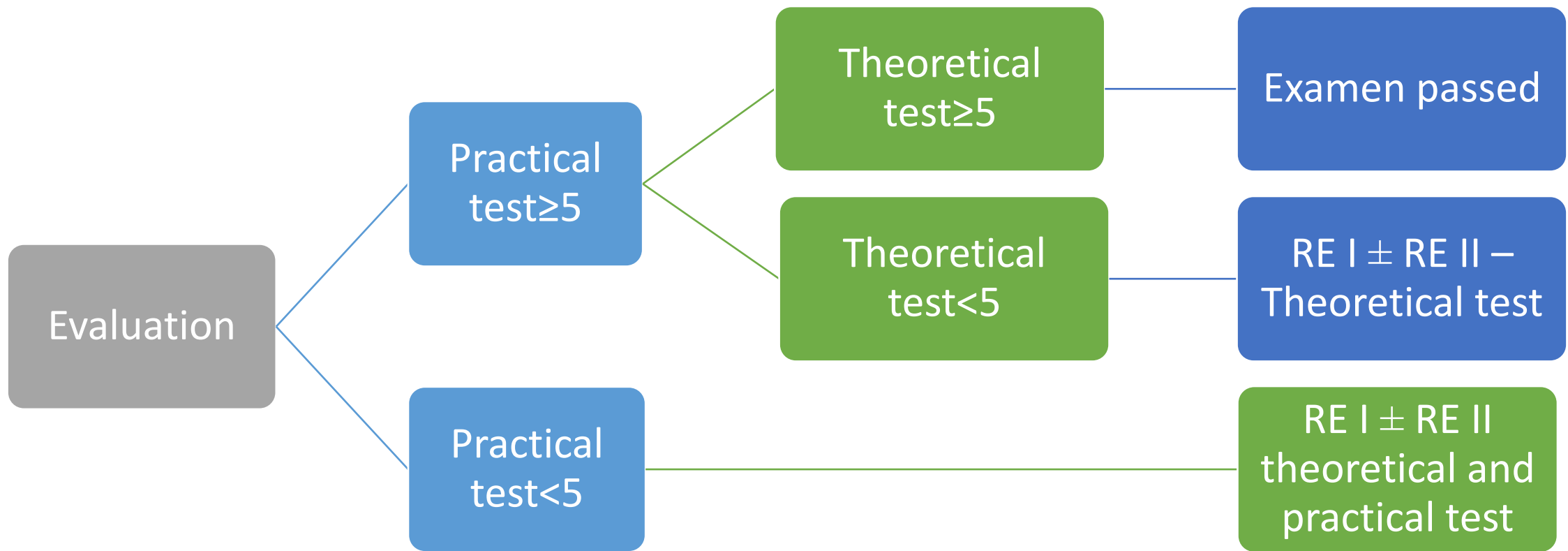
1st **Re-taken:**

2026 July 13th-17th

2nd **Re-taken:**

2026 September 14th -18th

Practical tests in the same day as theoretical test after the theoretical test



Results - Where?

- practical test – in the same day after the test
 - theoretical test – 2-3 days after the theoretical test
- final mark – 2-3 days after the theoretical test

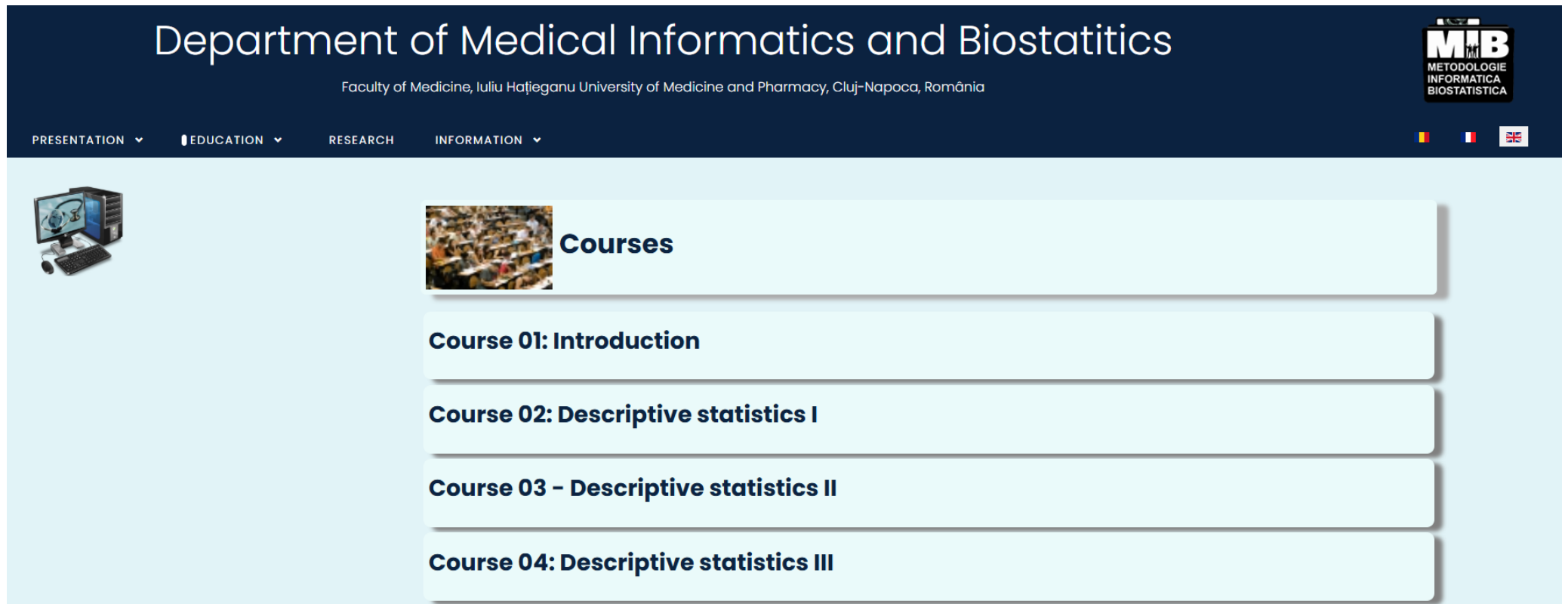
Consultations

- 9th February 2026
- Series 1: 10 o'clock
- Series 2: 12 o'clock
- online on Teams

Where are the presentations from courses?

Contact: cbondor@umfcluj.ro

www.info.umfcluj.ro





The screenshot shows the website for the Department of Medical Informatics and Biostatistics. The header includes the department name, faculty information, and a logo. The navigation menu has options for Presentation, Education, Research, and Information. The main content area features a 'Courses' section with a list of four courses: Course 01: Introduction, Course 02: Descriptive statistics I, Course 03 - Descriptive statistics II, and Course 04: Descriptive statistics III.

Department of Medical Informatics and Biostatistics

Faculty of Medicine, Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, România

PRESENTATION ▾ EDUCATION ▾ RESEARCH INFORMATION ▾

  **Courses**

- Course 01: Introduction**
- Course 02: Descriptive statistics I**
- Course 03 - Descriptive statistics II**
- Course 04: Descriptive statistics III**



Review



Variables
type

Qualitative
variable

```
graph TD; QV[Qualitative variable] --> QN[Qualitative nominal]; QV --> QO[Qualitative ordinal]; QN --> QD[Qualitative dichotomial]; QV2[Quantitative variable] --> D[Discrete]; QV2 --> C[Continuous];
```

Qualitative
nominal

Qualitative
ordinal

Qualitative
dichotomial

Quantitative
variable

Discrete

Continuous

Example Question: Variables type

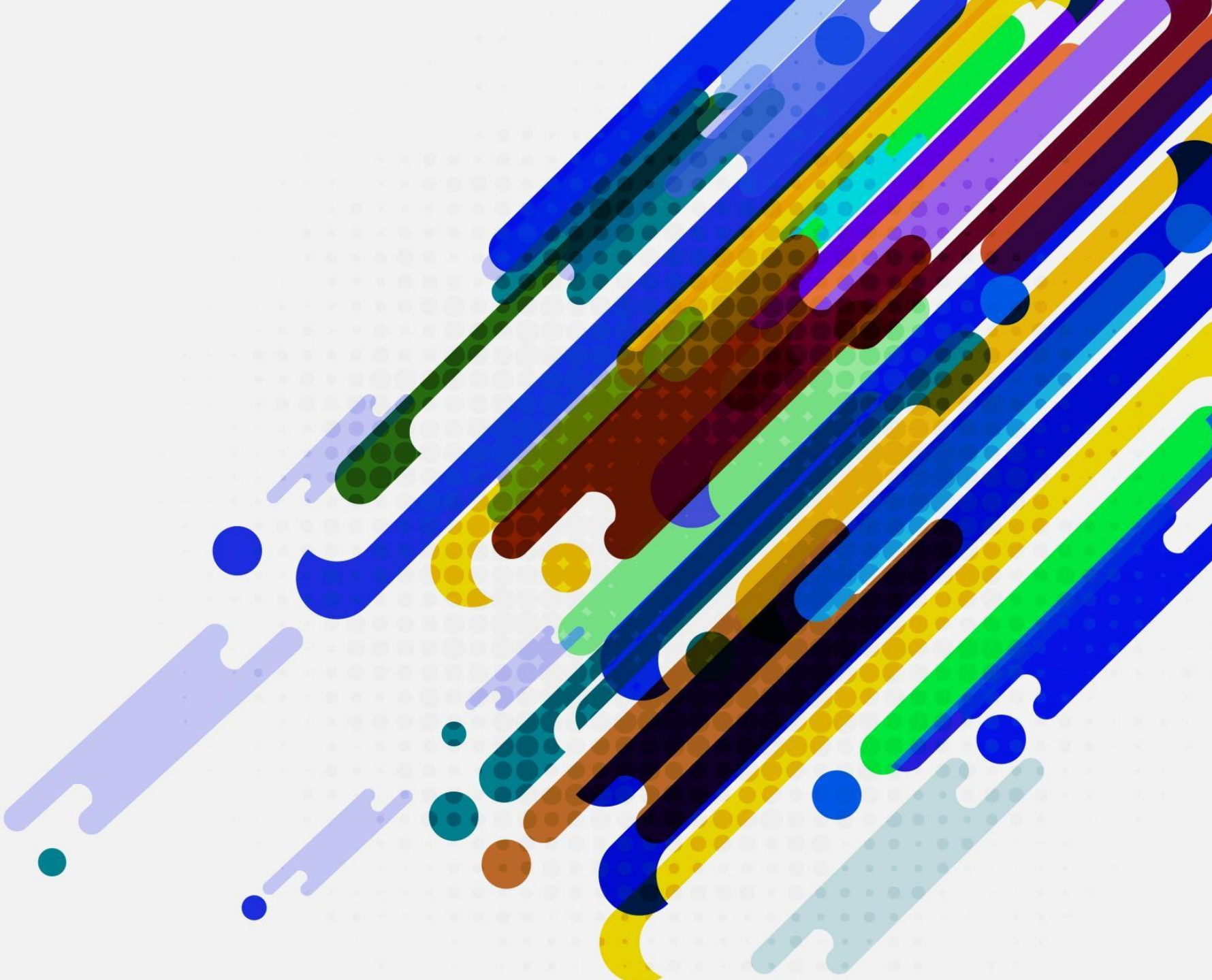
The variable Diabetes (Yes/No) is a

- A. Discrete quantitative variable
- B. Continuous quantitative variable
- C. Nominal qualitative variable
- D. Dichotomous qualitative variable
- E. Ordinal qualitative variable

Exemple Question: Variables type

The variable Diabetes (Yes/No) is a

- A. Discrete quantitative variable
- B. Continuous quantitative variable
- C. Nominal qualitative variable
- D. Dichotomous qualitative variable
- E. Ordinal qualitative variable



Descriptive
statistics

What can we do?

Count/calculate

Frequencies (absolute/relative)	Frequencies table
Contingency table	Relative risk, risk difference
Simple and compose probabilities	Conditional probabilities: Sensitivity, Specificity, Positive predictive value, Negative predictive value

Qualitative variables

Calculate

Arithmetic mean	Median	Mode
Standard deviation	Coefficient of variation	Range
Minimum /Minimum	Quartiles	Coefficient of determination

Quantitative variables

One qualitative variable

20 cases in intensive therapy

30 cases not in intensive therapy

We count the cases in each category

Frequencies

Frequencies table

	A	B	C
1	Id	Treatment	Intensive therapy
2	1	Methylprednisolone	Yes
3	2	Dexamethasone	No
4	3	Methylprednisolone	No
5	4	Dexamethasone	No
6	5	Methylprednisolone	No
7	6	Dexamethasone	No
8	7	Methylprednisolone	No
9	8	Dexamethasone	Yes
10	9	Methylprednisolone	No
11	10	Dexamethasone	No
12	11	Methylprednisolone	Yes
13	12	Dexamethasone	No
14	13	Methylprednisolone	No
15	14	Dexamethasone	No
16	15	Methylprednisolone	Yes

One qualitative variable

20 cases in intensive therapy = 40%

30 cases not in intensive therapy = 60%

Intensive therapy	Absolute frequency	Relative frequency (%)
Yes	20	40
No	30	60
Total	50	100

We count the cases in each category

Frequencies

Frequencies table

Example:

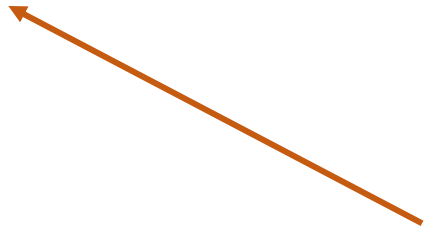
Into the study were included 10 patients. The following data were collected on the variable Diabetes: Yes, No, No, No, No, Yes, No, No, No, No, where Yes means that the patient has diabetes, No – the patient does not have diabetes. Which of the following statements are true?

- A. The relative frequency of people with diabetes in the studied sample is 20%
- B. The absolute frequency of people without diabetes in the studied sample is 8
- C. The empirical probability that a person suffers from diabetes is 0.2
- D. The empirical probability that a person does not suffer from diabetes is 0.8
- E. The relative frequency of people with diabetes in the studied sample is 2

Diabetes: Yes, No, No, No, No, Yes, No, No, No, No

B. The absolute frequency of people without diabetes in the studied sample is 8

2 patients with diabetes
8 patients without diabetes



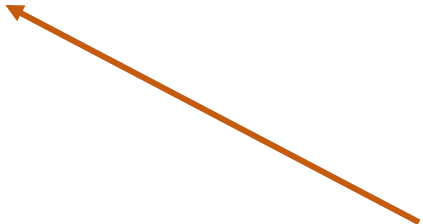
Absotute frequency F_a – count
cases on each category

Diabetes: Yes, No, No, No, No, Yes, No, No, No, No

A. The relative frequency of people with diabetes in the studied sample is 20%

$=2/10*100=20$, 20% from the patients have diabetes

$=8/10*100=80$, 80% from the patients did not have diabetes



Relative frequency – Divide the frequency with the total number of patients and multiply with 100
 $=F_a/n*100$

C. The empirical probability that a person suffers from diabetes is 0.2

D. The empirical probability that a person does not suffer from diabetes is 0.8

Diabetes: Yes, No, No, No, No, Yes, No, No, No, No

=2/10=0.2 the empirical probability that a person suffers from diabetes

=8/10=0.8 the empirical probability that a person does not suffer from diabetes



Empirical probability – Divide the frequency
with the total number of patients = $F_a/10$

Example:

Into the study were included 10 patients. The following data were collected on the variable Diabetes: Yes, No, No, No, No, Yes, No, No, No, No, where Yes means that the patient has diabetes, No – the patient does not have diabetes. Which of the following statements are true?

- A. The relative frequency of people with diabetes in the studied sample is 20%
- B. The absolute frequency of people without diabetes in the studied sample is 8
- C. The empirical probability that a person suffers from diabetes is 0.2
- D. The empirical probability that a person does not suffer from diabetes is 0.8
- E. The relative frequency of people with diabetes in the studied sample is 2

Two qualitative variables ex. Risk factor and disease

	With disease	Without disease	Total
Factor +	a	b	a+b
Factor -	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

	Intensive therapy = Yes	Intensive therapy = No	Total
Methylprednisolone	15	5	20
Dexamethazone	25	5	30
Total	40	10	50

cases in each category

Frequencies

Frequencies table

Contingency table



Example:

600 patients were included in the study. 190 had diabetes, of whom 150 were obese. Of those who did not have diabetes, 50 were obese. Which of the following statements is true?

- A. The risk of obese people having diabetes is 75%
- B. The risk of normal-weight people having diabetes is 10%
- C. The relative risk of obese people having diabetes compared to those of normal weight is 7.5
- D. The probability that a person does not have diabetes and is of normal weight is 0.6
- E. The probability that a person has diabetes and is of normal weight is 0.08

600 patients were included in the study. 190 had diabetes, of whom 150 were obese. Of those who did not have diabetes, 50 were obese.

190 had diabetes

Of those with diabetes
150 were obese

Of those who did not
have diabetes, 50 were
obese

	Diabetes= Yes	Diabetes= No	Total
Obese=Yes	150	50	
Obese=No			
Total	190		600

600 patients were included in the study.

$190 - 150 = 40$

$410 - 50 = 360$

$150 + 50 = 200$

	Diabetes= Yes	Diabetes= No	Total
Obese=Yes	150	50	200
Obese=No	40	360	400
Total	190	410	600

$600 - 190 = 410$

$600 + 200 = 400$

	Diabetes=Yes	Diabetes=No	Total
Obese=Yes	150	50	200
Obese=No	40	360	400
Total	190	410	600

A. The risk of obese people having diabetes is 75%

B. The risk of normal-weight people having diabetes is 10%

RIE=frequency of obese people with diabetes /frequency of obese = $a/(a+b)$

= $150/200=0.75$,
75% of obese people have diabetes

RIN=frequency of non-obese people with diabetes /frequency of non-obese = $c/(c+d)$

= $40/400=0.10$,
10% of non-obese people have diabetes

	Diabetes=Yes	Diabetes=No	Total
Obese=Yes	150	50	200
Obese=No	40	360	400
Total	190	410	600

- A. The risk of obese people having diabetes is 75%
- B. The risk of normal-weight people having diabetes is 10%
- C. The relative risk of obese people having diabetes compared to those of normal weight is 7.5

Relative risk
 $RR = R_{IE}/R_{IN}$

$=75/10=7.5$ or $0.75/0.10=7.5$

Obese people have a 7.5 times higher risk of diabetes than people who are not obese

	Diabetes=Yes	Diabetes=No	Total
Obese=Yes	150	50	200
Obese=No	40	360	400
Total	190	410	600

D. The probability that a person does not have diabetes and is of normal weight is 0.6

~~E. The probability that a person has diabetes and is of normal weight is 0.08~~

P (no diabetes/non obese)
 = divide frequency of the
 person without diabetes
 and non-obese to the total
 of the patients= $d/600$



= $360/600=0.6$

the empirical probability that a person from the sample does not have diabetes and is of normal weight

P (diabetes/non obese) =
 divide frequency of the
 person with diabetes and
 non-obese to the total of
 the patients= $c/600$



= $40/600=0.07$

the empirical probability that a person from the sample have diabetes and is of normal weight

Example:

600 patients were included in the study. 190 had diabetes, of whom 150 were obese. Of those who did not have diabetes, 50 were obese. Which of the following statements is true?

- A. The risk of obese people having diabetes is 75%
- B. The risk of normal-weight people having diabetes is 10%
- C. The relative risk of obese people having diabetes compared to those of normal weight is 7.5
- D. The probability that a person from the sample does not have diabetes and is of normal weight is 0.6
- E. The probability that a person has diabetes and is of normal weight is 0.08

What we can do?

Calculate

Arithmetic mean

Median

Mode

Standard deviation

Coefficient of variation

Range

Minimum /Maximum

Quartiles

Coefficient of determination

	A	B
	Id_pac	Age (years)
1		
2	1	75
3	2	50
4	3	60
5	4	75
6	5	74
7	6	81
8	7	44
9	8	34
10	9	54
11	10	63

- Arithmetic mean for age = $(75+50+60+75+74+81+44+34+54+63)/10 = 610/10 = 61$
- sort 34, 44, 50, 54, 60, 63, 74, 75, 75, 81
- median = $(60+63)/2 = 61.5$
- quartile 1 = 50
- quartile 3 = 75
- mode = 75
- minimum = 34
- maximum = 81
- range = $81-34 = 47$

• variation = $[(75-61)^2+(50-61)^2+(60-61)^2+(75-61)^2+(74-61)^2+(81-61)^2+(44-61)^2+(34-61)^2+(54-61)^2+(63-61)^2]/(10-1) = [14^2+(-11)^2+(-1)^2+14^2+13^2+20^2+(-17)^2+(-27)^2+(-7)^2+2^2]/9 = (196+121+1+192+163+411+289+729+49+4)/9 = 2155/9 = 239.47$

- standard deviation = $\sqrt{239.47} = 15.47$
- coefficient of variation = standard deviation/arithmetic mean = $15.47/61 = 0.25$

if we know the coefficient of correlation r , coefficient of determination is $d=r^2$, ex. $r=0.33$, $d=0.33^2=0.11$

Example:

10 patients were included in the study. The following data were collected on the variable Age: 100, 10, 90, 20, 40, 60, 50, 50, 50, 60. Which of the following statements are true?

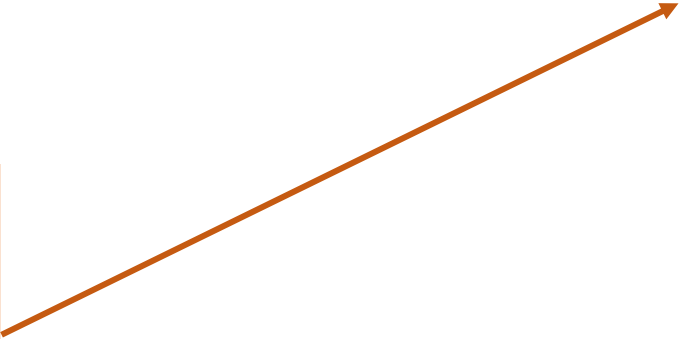
- A. The arithmetic mean of age on the studied sample is 53
- B. The mode of the age variable on the studied sample is 50
- C. The median of the age variable on the studied sample is 50
- D. The standard deviation of the age variable on the studied sample is 27.5
- E. The range of the age variable on the studied sample is 80

Age: 100, 10, 90, 20, 40, 60, 50, 50, 50, 60

A. The arithmetic mean of age on the studied sample is 53

$$=(100+10+90+20+40+60+50+50+50+60)/10=530/10=53$$

\bar{X} – sum of values divided by the total number of values



Age: 100, 10, 90, 20, 40, 60, 50, 50, 50, 60

B. The mode of the age variable on the studied sample is 50

100, 10, 90, 20, 40 once

60 twice

50 3 times



The most frequent value

Age: 100, 10, 90, 20, 40, 60, 50, 50, 50, 60

C. The median of the age variable on the studied sample is 50

10, 20, 40, 50, 50, 50, 60, 60, 90, 100

median = $(50+50)/2=50$

The value in the middle of the series

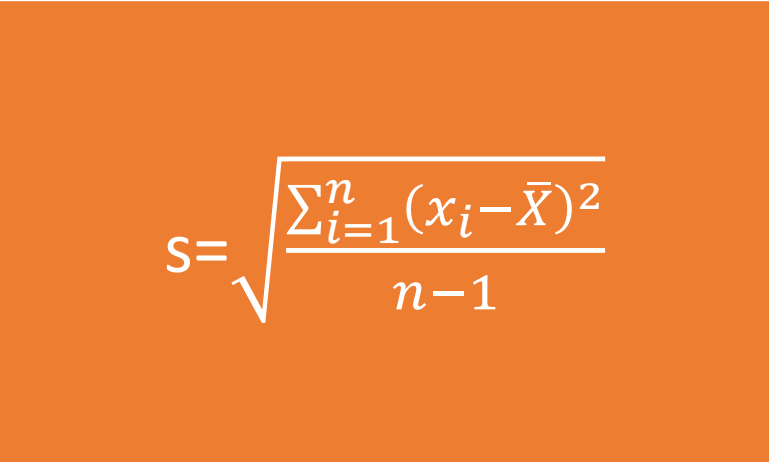
We sort the data in ascending order

We made the average of the two values in the middle or the median is the value in the middle

Age: 100, 10, 90, 20, 40, 60, 50, 50, 50, 60

A. D. The standard deviation of the age variable on the studied sample is 27.5

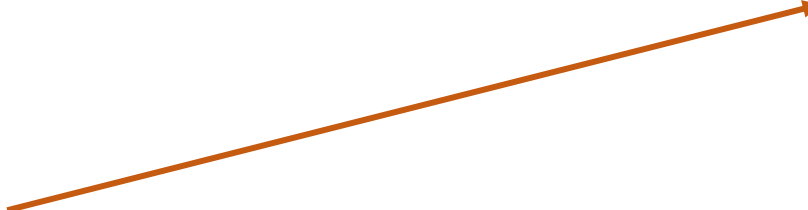
$$\begin{aligned} &= (100-53)^2+(10-53)^2+(90-53)^2+(20-53)^2+(40-53)^2+ (60-53)^2+ \\ &+(50-53)^2+(50-53)^2+(50-53)^2+(60-53)^2 = \\ &= (47)^2+(-43)^2 +(37)^2+(-33)^2+(-13)^2+(7)^2+(-3)^2+(-3)^2+(-3)^2+(7)^2 = \\ &=2209+1849+1369+1089+169+49+9+9+9+49 = 6810 \\ &s=\sqrt{6810/(10 - 1)}=\sqrt{756,667}=27,5 \end{aligned}$$


$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}}$$

Age: 100, 10, 90, 20, 40, 60, 50, 50, 50, 60

~~E. The range of the age variable on the studied sample is 80~~

$$= 100 - 10 = 90$$



Range = maximum-
minimum

Example:

10 patients were included in the study. The following data were collected on the variable Age: 100, 10, 90, 20, 40, 60, 50, 50, 50, 60. Which of the following statements is true?

- A. The arithmetic mean of age on the studied sample is 53
- B. The mode of the age variable on the studied sample is 50
- C. The median of the age variable on the studied sample is 50
- D. The standard deviation of the age variable on the studied sample is 27.5
- E. The range of the age variable on the studied sample is 80

Interpretations

Relative
risk

Risk
difference

Coefficient
of variation

Skewness

Normal
distribution

Kurtosis

Correlation
coefficient

The coefficient of variation ($CV = s/\bar{X} * 100$)

interpretation for positive values

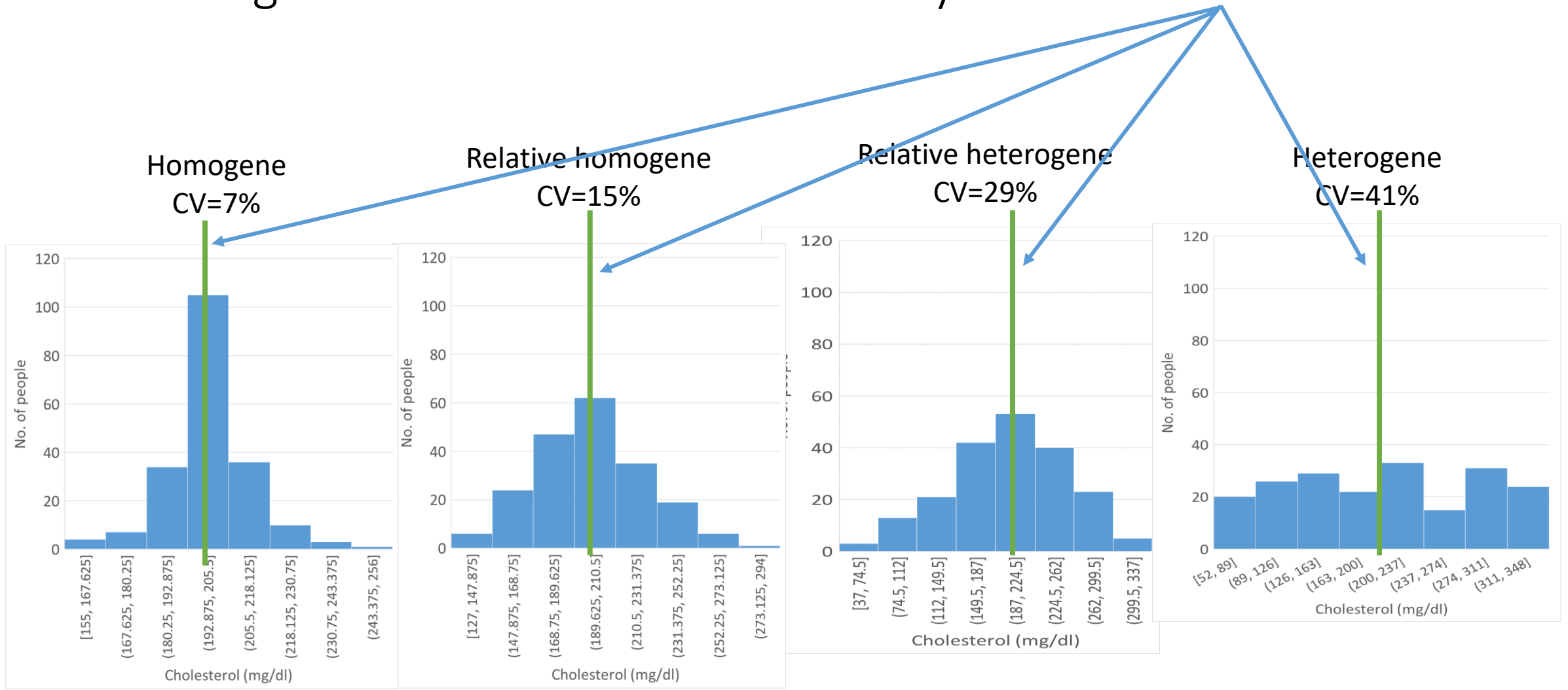
- $0\% \leq CV < 10\%$ data are homogeneous
- $10\% \leq CV < 20\%$ data are relatively homogeneous
- $20\% \leq CV < 30\%$ data are relatively heterogeneous
- $30\% \leq CV$ data are heterogeneous

- where s = standard deviation, \bar{X} = arithmetic mean

Evaluate which of the following statements are true

- A. A coefficient of variation of 0.96 shows that the data is heterogeneous
- B. A coefficient of variation of 0.26 shows that the data is relatively heterogeneous
- C. A coefficient of variation of 4 shows that the data is homogeneous
- D. A coefficient of variation of 24% shows that the data is relatively heterogeneous
- E. A coefficient of variation of 4% shows that the data is relatively homogeneous

homogeneous = values are in vicinity of the arithmetic mean
 heterogeneous = values are not in vicinity of the arithmetic mean



100 patients

100 patients

The coefficient of variation ($CV = s/\bar{X} * 100$)

interpretation for positive values

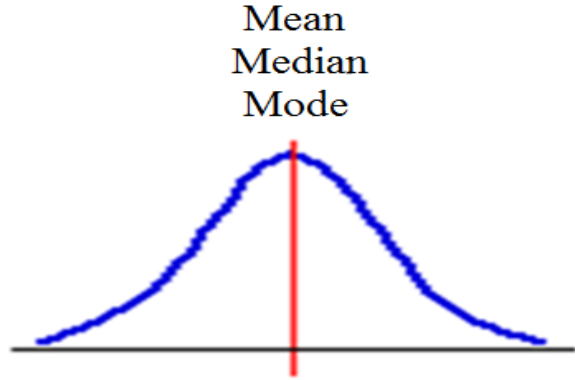
- $0\% \leq CV < 10\%$ data are homogeneous
- $10\% \leq CV < 20\%$ data are relatively homogeneous
- $20\% \leq CV < 30\%$ data are relatively heterogeneous
- $30\% \leq CV$ data are heterogeneous

- where s = standard deviation, \bar{X} = arithmetic mean

Evaluate which of the following statements are true

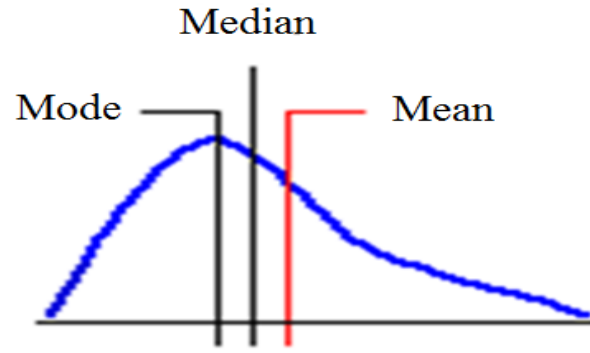
- A. A coefficient of variation of 0.96 shows that the data is heterogeneous
- B. A coefficient of variation of 0.26 shows that the data is relatively heterogeneous
- C. A coefficient of variation of 4 shows that the data is homogeneous
- D. A coefficient of variation of 24% shows that the data is relatively heterogeneous
- E. A coefficient of variation of 4% shows that the data is relatively homogeneous

Normal distribution



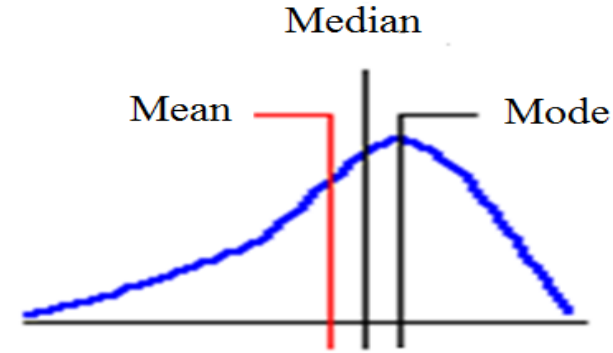
Symmetric
Skewness=0

symmetrical distribution



Positive asymmetry
Skewness>0

asymmetry to the right

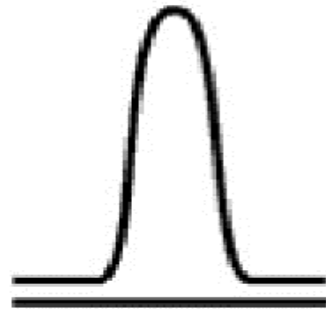


Negative asymmetry
Skewness<0

asymmetry to the left



Mesokurtic
kurtosis=0



Leptokurtic
kurtosis>0



Platykurtic
kurtosis<0

Rule: A serie of numbers is normal distributed if

- ✓ Arithmetic mean \approx Median \approx Mode
- ✓ Quartile 1, Quartile 3 are simetrical with the mean/median
- ✓ Skewness \approx 0 (between -1 to 1)
- ✓ Kurtosis \approx 0 (between -1 to 1)

Example:

Evaluate which of the following statements are true

- A. if the skewness is -3 then the data does not follow a normal distribution
- B. if the skewness is 5 then the data does not follow a normal distribution
- C. if the kurtosis is -2 then the data does not follow a normal distribution
- D. if the arithmetic mean, median and mode are equal then the data follow a normal distribution
- E. if the arithmetic mean is 10, quartile 1 is 5 and quartile 2 is 15 then the data does not follow a normal distribution

r/rho correlation coefficient

- r between -1 și 1
- $r < -0,25$
 - inversely proportional correlation
- $r > 0,25$
 - directly proportional correlation

COLTON'S EMPIRICAL RULES (1974)

a correlation coefficient

from -0.25 to 0.25

means a weak or null correlation

from 0.25 to 0.50 (or from -0.25 to -0.50)

an acceptable degree of association

from 0.5 to 0.75 (or from -0.5 to -0.75)

a moderate to good correlation

greater than 0.75 (or less than -0.75)

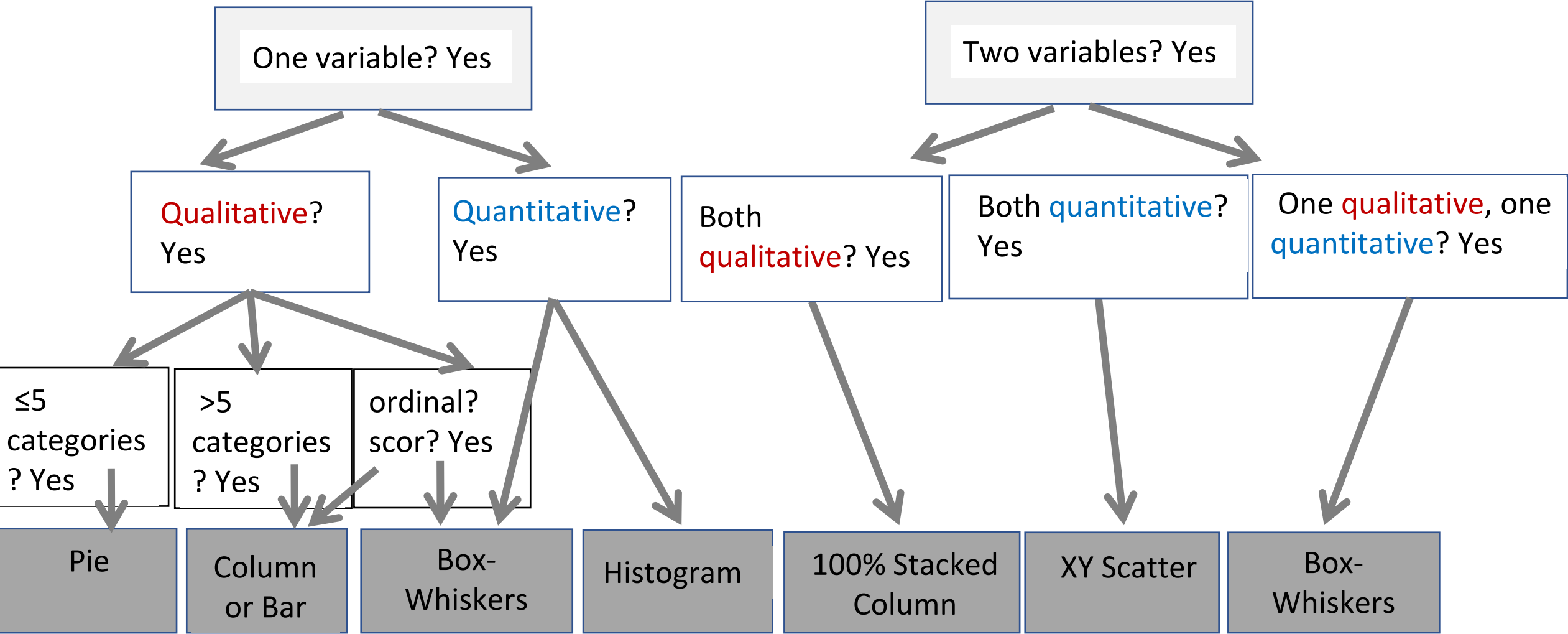
a very good association or correlation

Example:

Evaluate which of the following statements are true

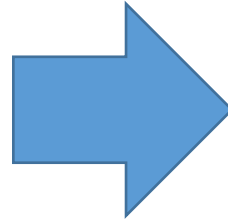
- A. A correlation coefficient of -0.96 shows an inversely proportional correlation between the considered series of values
- B. A correlation coefficient of 0.26 shows an acceptable correlation
- C. A correlation coefficient of -12 shows an inversely proportional correlation between the considered series of values
- D. The values of the correlation coefficient vary between -1 and 1
- E. A correlation coefficient of 1 shows an inversely proportional correlation between the considered series of values

How to choose the right chart?



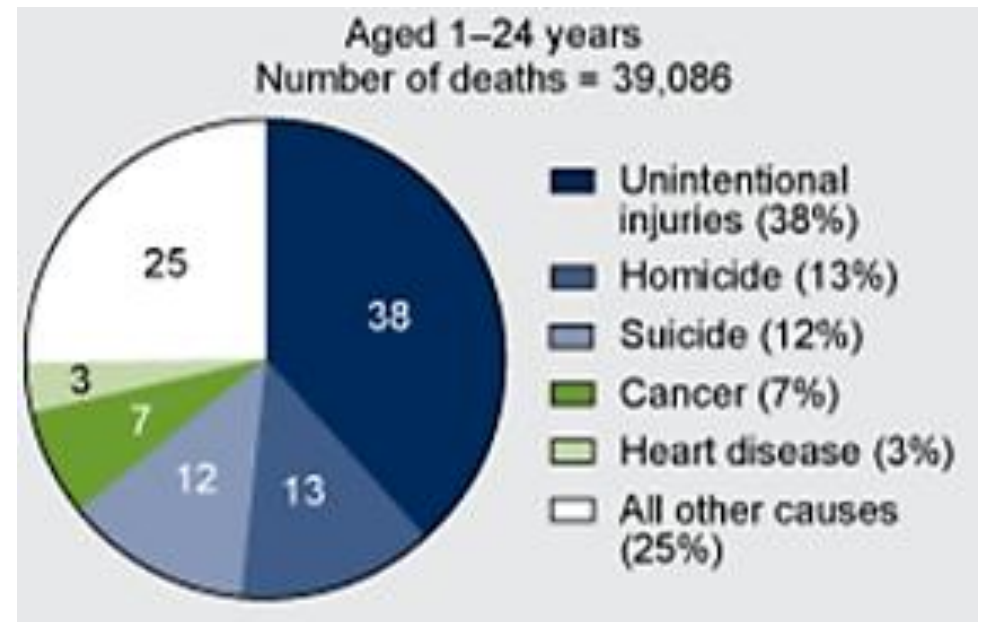
Choose the right chart

- One qualitative
 - nominal
 - dichotomial
 - ordinal



• Chart

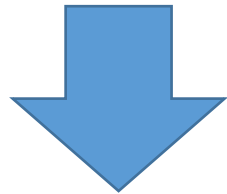
- **Pie (we prefer it when is possible)**
- Column
- Bar



One variables

- One qualitative

- nominal
- dichotomial
- ordinal



Column (when >6 categories or ordinal variable) or Bar chart

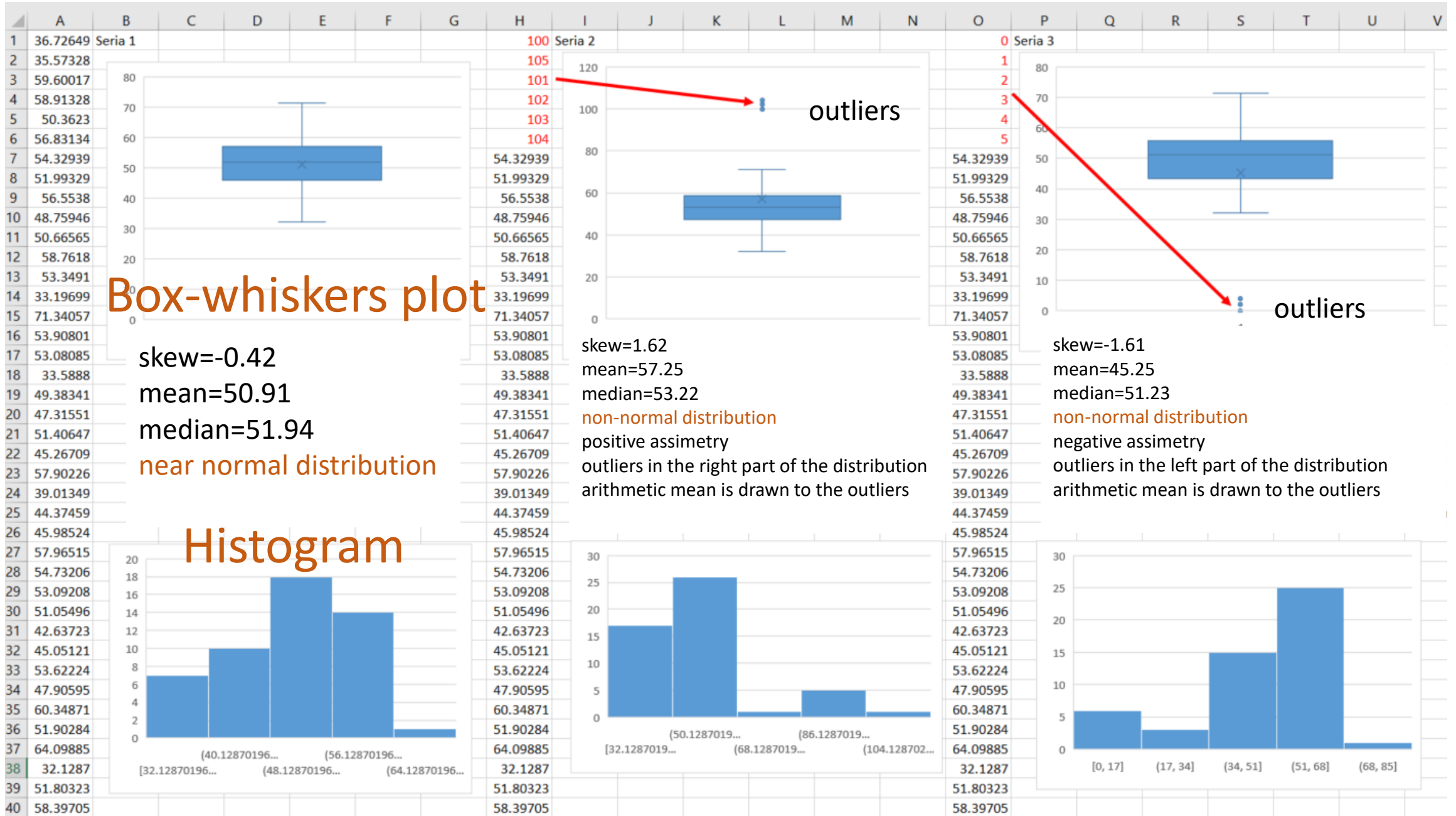


<https://www.dreamstime.com/>



<https://www.shutterstock.com/>

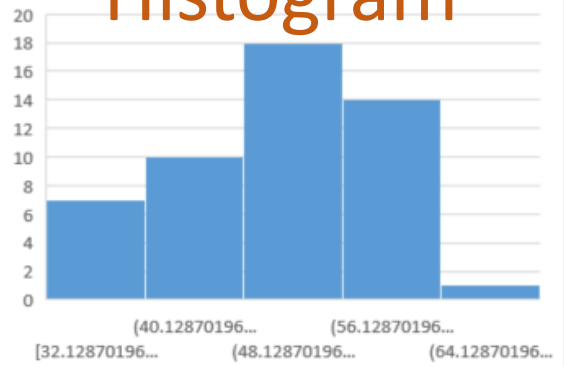
One quantitative variable



Box-whiskers plot

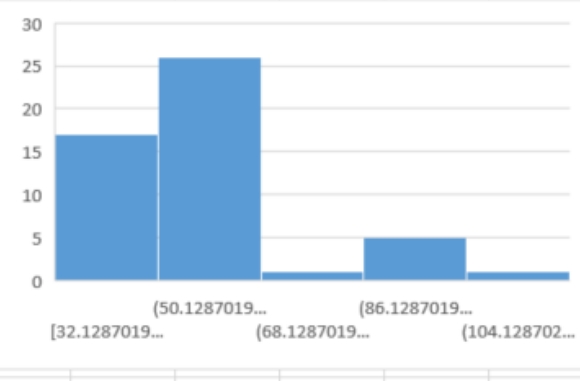
skew=-0.42
 mean=50.91
 median=51.94
 near normal distribution

Histogram



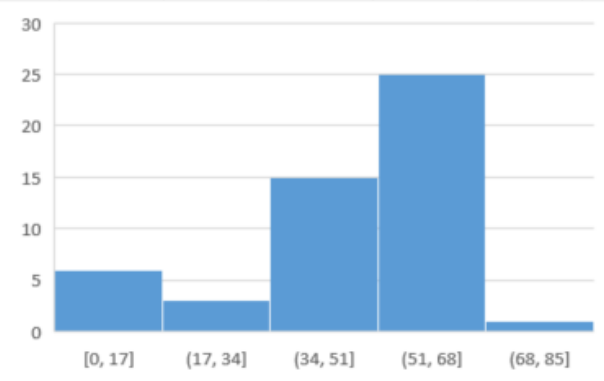
outliers

skew=1.62
 mean=57.25
 median=53.22
 non-normal distribution
 positive assimetry
 outliers in the right part of the distribution
 arithmetic mean is drawn to the outliers



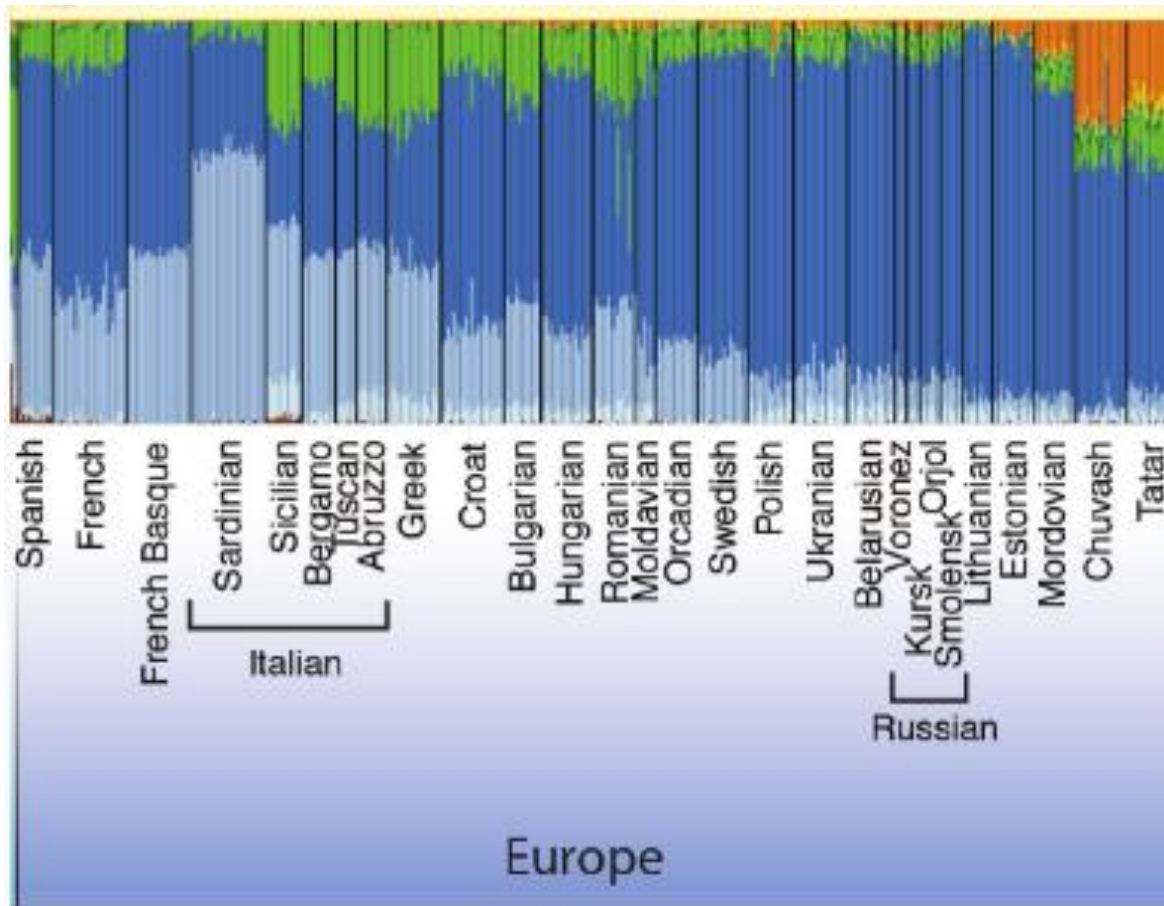
outliers

skew=-1.61
 mean=45.25
 median=51.23
 non-normal distribution
 negative assimetry
 outliers in the left part of the distribution
 arithmetic mean is drawn to the outliers

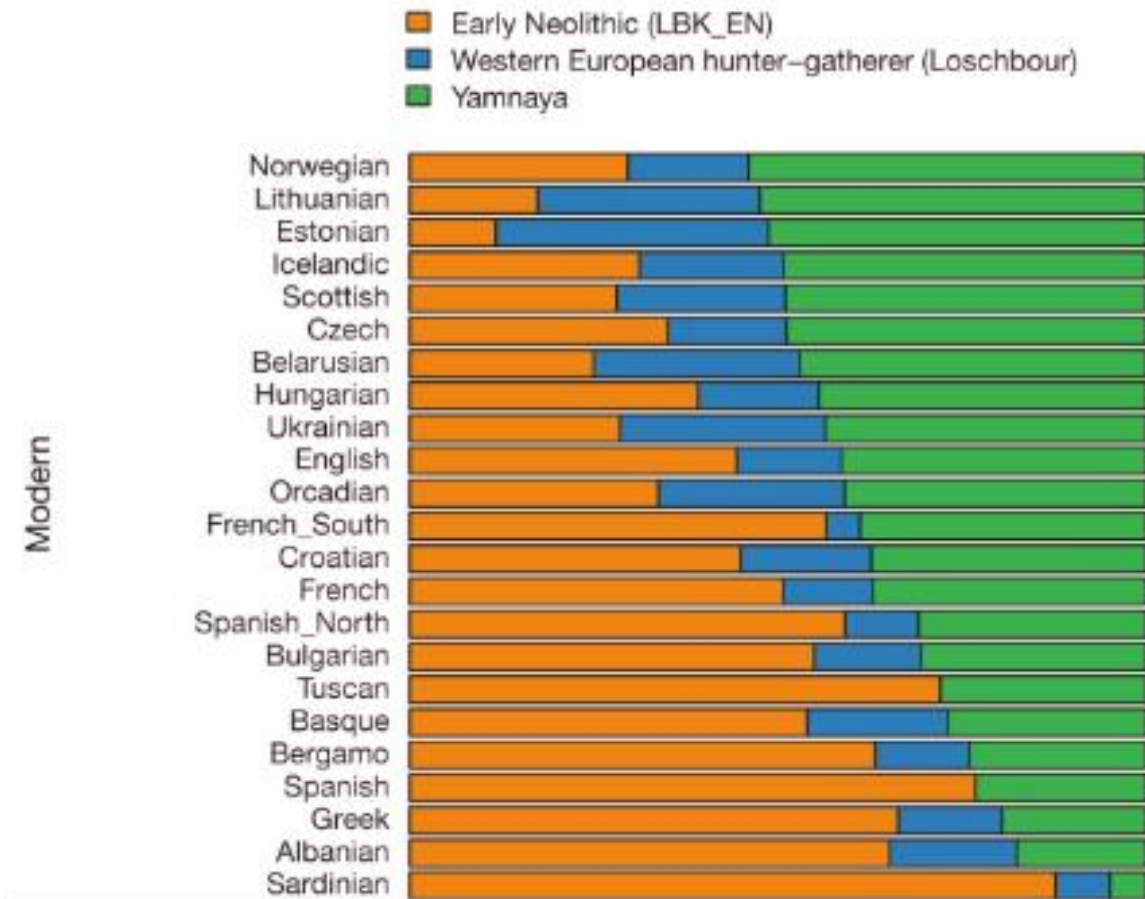


Two qualitative variables

Column or Bar chart – preferably 100% Stacked column/bar chart



Kovacevic L et al. Standing at the gateway to Europe--the genetic structure of Western balkan populations based on autosomal and haploid markers. PLoS One. 2014



Haak W et al . Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015.

- XY Scatter

- two quantitative variables
- or ordinal variables with many scores

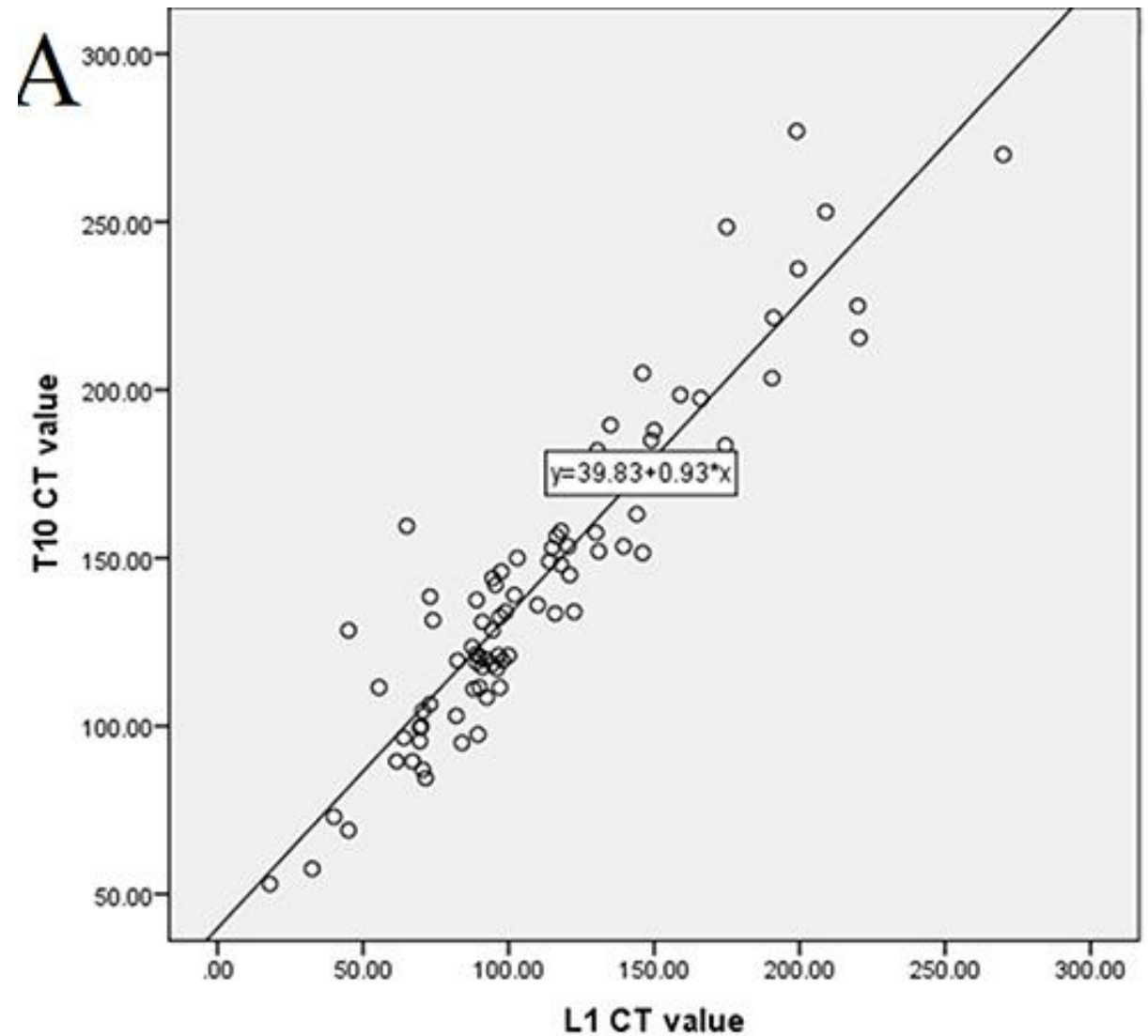
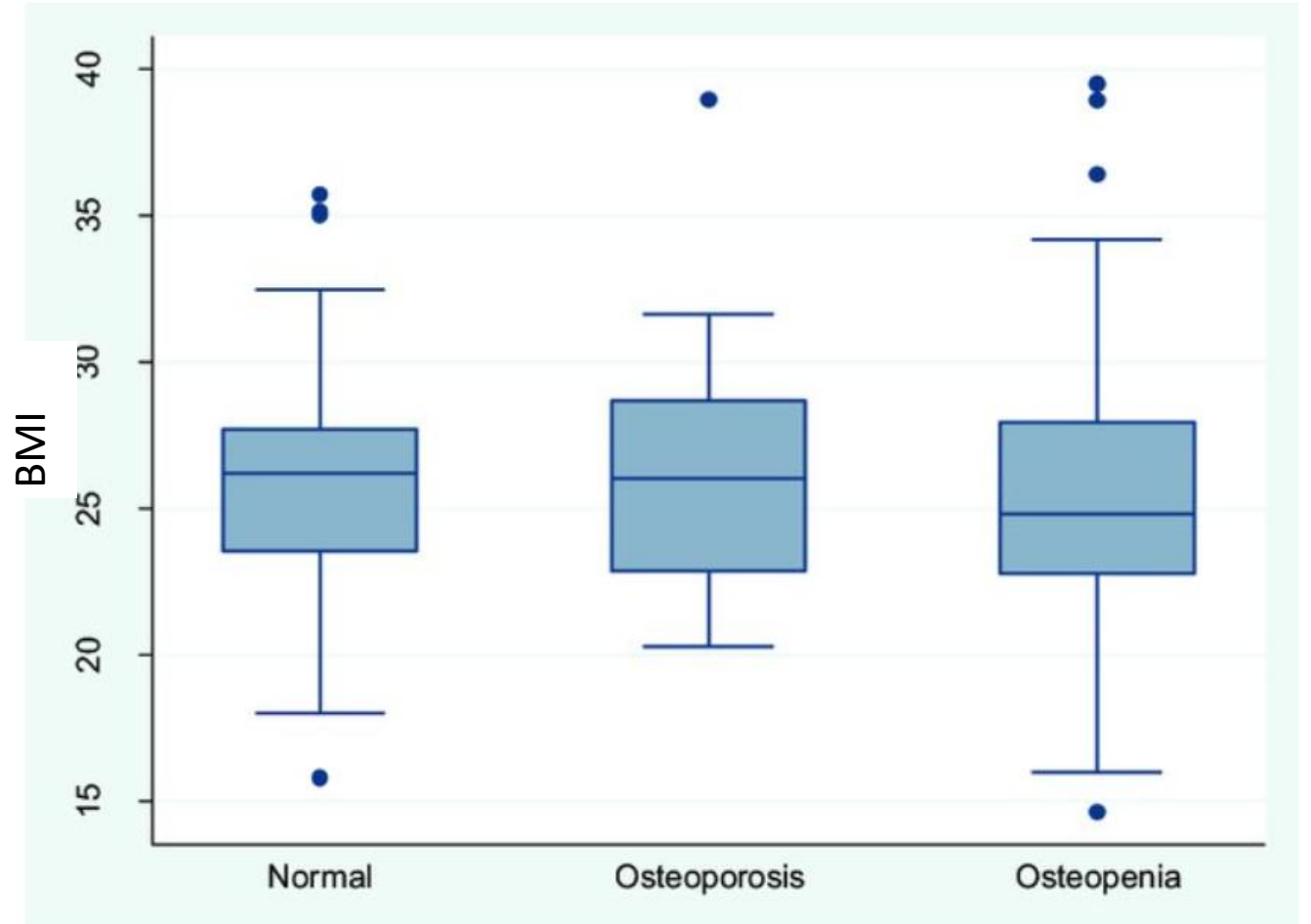


Figure 1 Correlation analysis of CT values of cancellous bone of vertebral bodies T10-T12 and L1. Statistics scatter diagrams and fitting lines of correlation analysis of CT values between **(A)** T10 vs L1

- Box-whiskers plot

- one quantitative variable (or ordinal with many scores) and one qualitative variable (groups, repeated measurements)



Hiremath RN, Yadav AK, Ghodke S, Yadav J, Latwal S, Kotwal A. Osteoporosis among household women: A growing but neglected phenomenon. Med J Armed Forces India. 2018 Jan;74(1):5-10.

Interpretations

Histogram

Box-whiskers
plot

100%
Column/Bar
chart

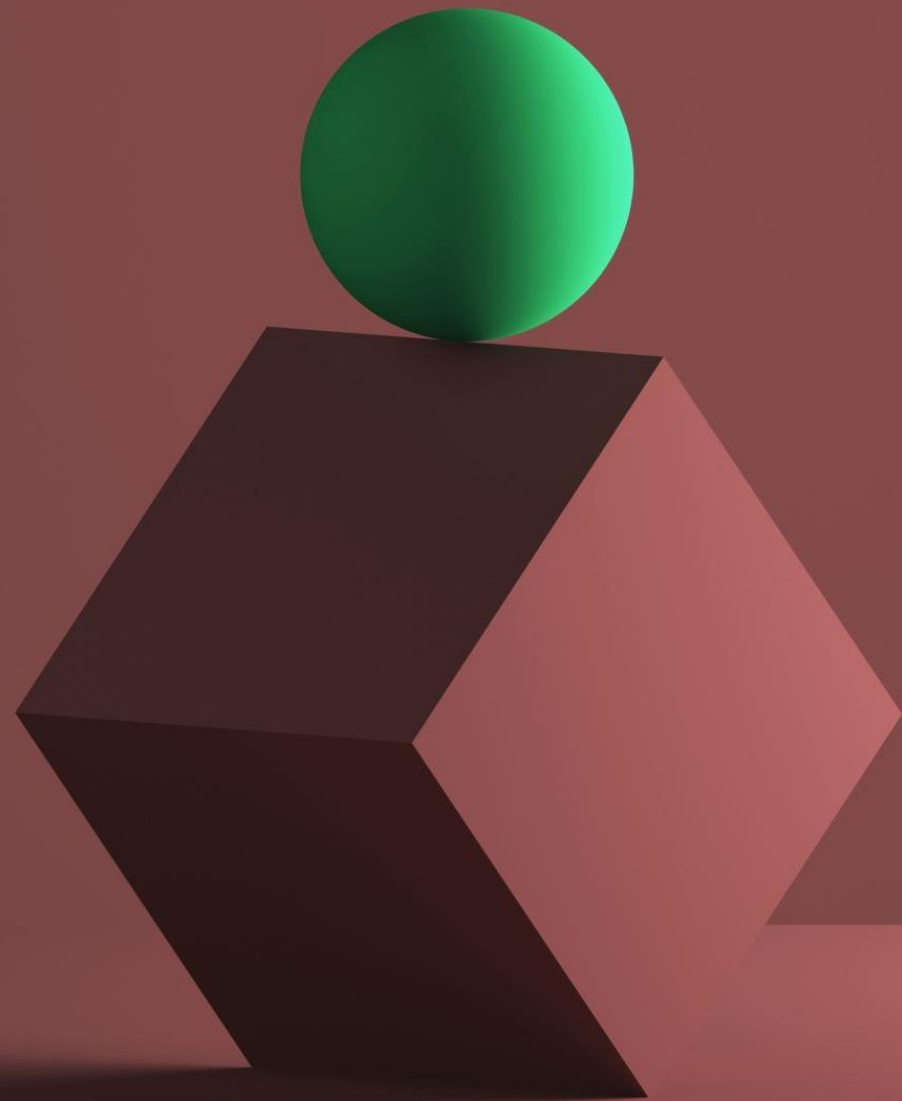
XY Scatter

- about
 - asymmetry
 - outliers cases
 - differences between groups
 - dependency
 - ascending/descending trend
 - graduality

Example

*A histogram is a suitable graphical representation for:

- A. Discrete quantitative
- B. Continuous quantitative**
- C. Nominal qualitative
- D. Dichotomous qualitative
- E. Ordinal qualitative



Statistical inference

Generalization to the population

Population P

Ex. Frequency of hypertension in the population ϕ



Objective:

- Frequency ϕ , proportion, relative risk, arithmetic mean μ  ...

How?

1.

- Extract randomly a representative sample

2.

- Calculate the statistics
 - Frequency f

3.

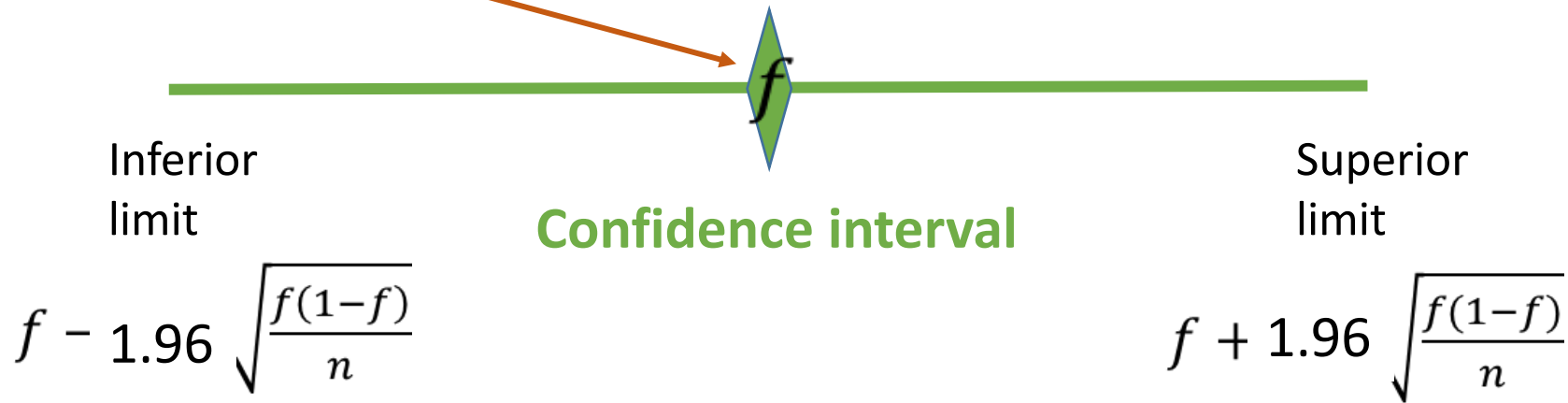
with inferencial statistics methods make a generalization to the population

frequency ϕ is the same as the sample frequency f

frequency ϕ can be found in the confidence interval

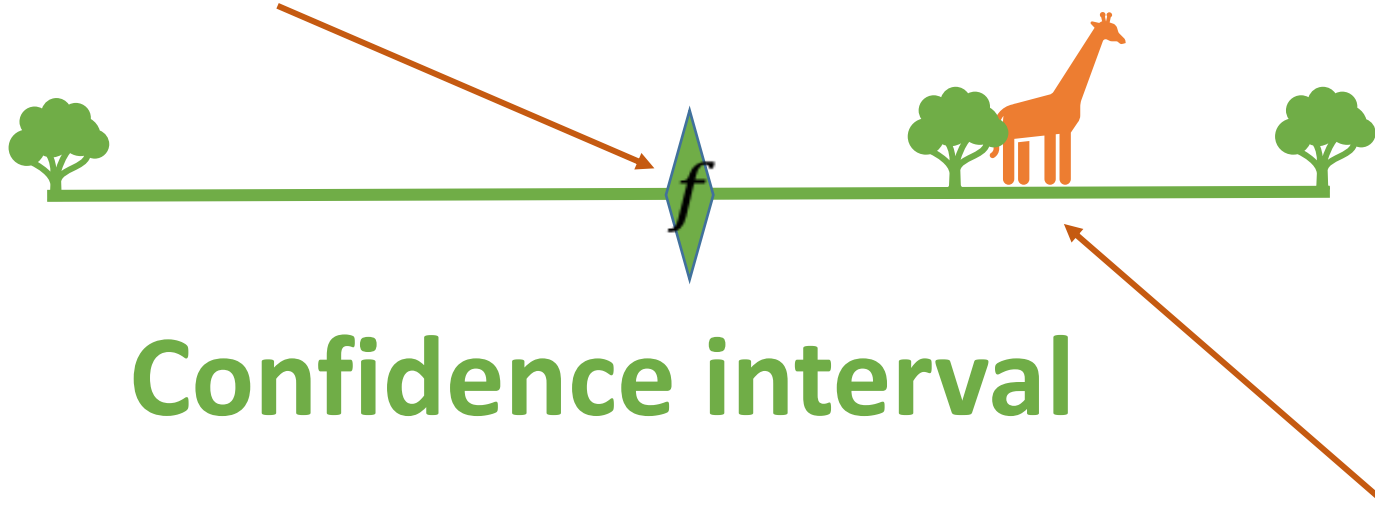
95% confidence interval for a frequency

Punctual estimation



95% confidence interval for a frequency

Punctual estimation = Frequency calculated on the sample



Confidence interval

Frequency of the disease/risk factor in the population ϕ

can be anywhere in the estimated interval

- We have a 95% chance of extracting a sample from where the estimated confidence interval will contain the population parameter

Ex. Arithmetic mean of cholesterol in population μ



Population P

Objective:

- arithmetic mean μ  ...

How?

1.

- Extract randomly a representative sample

2.

- Calculate the statistics
 - arithmetic mean \bar{X}

3.

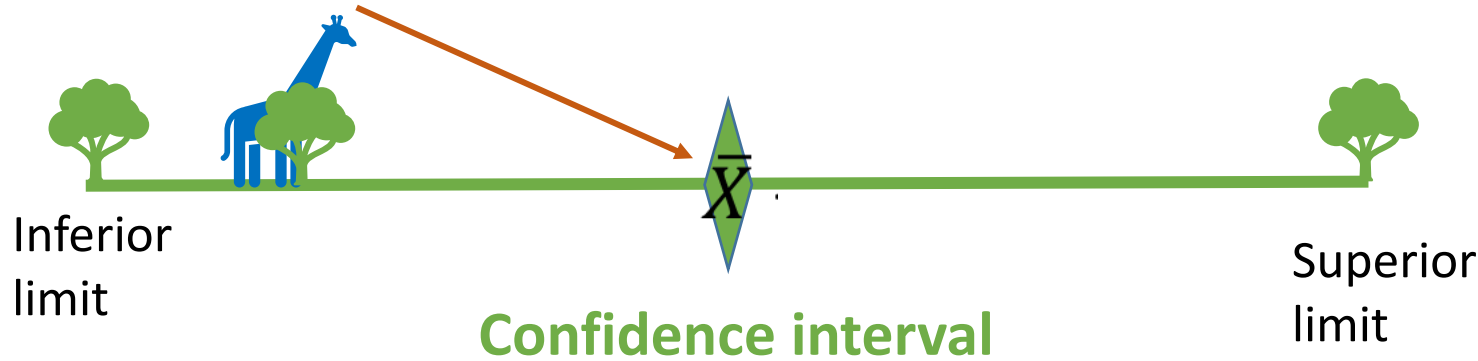
- with inferential statistics methods make a generalization to the population
 - **arithmetic mean μ** is the same as the sample arithmetic mean \bar{X}
 - **arithmetic mean μ** can be found in the confidence interval

95% confidence interval for an arithmetic mean

Punctual estimation

arithmetic mean of the population μ

- included in the confidence interval with a 95% probability



$$\bar{X} - 1.96 \frac{s}{\sqrt{n-1}}$$

$$\bar{X} + 1.96 \frac{s}{\sqrt{n-1}}$$

- arithmetic mean of the population μ  - can be anywhere in the estimated interval

- We have a 95% chance of extracting a sample from where the estimated confidence interval will contain the population parameter 

Example

*For a sample of 82 hospitalized patients, the mean length of stay for patients working in a toxic environment was found to be 25 days/year, with a standard deviation of 9 days. Determine the 95% confidence interval for the mean length of stay for patients working in a toxic environment. ($Z_{\alpha}=1.96$, $\alpha=0.05$).

A. (19; 31)

B. (23.04; 26.96)

C. (24.72; 29.51)

D. (23.23; 26.83)

E. (22.71; 28.51)

$$\left(\bar{X} - 1.96 \frac{s}{\sqrt{n-1}}, \bar{X} + 1.96 \frac{s}{\sqrt{n-1}}\right)$$

$$(25 - 1.96 * 9 / \sqrt{82 - 1}; 25 + 1.96 * 9 / \sqrt{82 - 1})$$

$$(25 - 1.96 * 9 / 9; 25 + 1.96 * 9 / 9)$$

$$(25 - 1.96 * 1; 25 + 1.96 * 1)$$

$$(25 - 1.96; 25 + 1.96)$$

$$(23.04; 26.96)$$

Example

For a sample of 5000 people, the frequency of HPV infection was 15.3%. Determine the 95% confidence interval for the frequency of HPV infection. ($Z_{\alpha}=1.96$, $\alpha=0.05$).

A. (13; 17)

B. (14.3%; 16.3%)

C. (0.143; 0.163)

D. (14.1; 16.5)

E. (10.3; 20.3)

$$\left(f - 1.96 \sqrt{\frac{f(1-f)}{n}}; f + 1.96 \sqrt{\frac{f(1-f)}{n}} \right)$$

$$\left(0.153 - 1.96 * \sqrt{\frac{0.153 * (1 - 0.153)}{5000}}; 0.153 + 1.96 * \sqrt{\frac{0.153 * (1 - 0.153)}{5000}} \right)$$

$$\left(0.153 - 1.96 * \sqrt{\frac{0.153 * 0.847}{5000}}; 0.153 + 1.96 * \sqrt{\frac{0.153 * 0.847}{5000}} \right)$$

$$\left(0.153 - 1.96 * \sqrt{\frac{0.129591}{5000}}; 0.153 + 1.96 * \sqrt{\frac{0.129591}{5000}} \right)$$

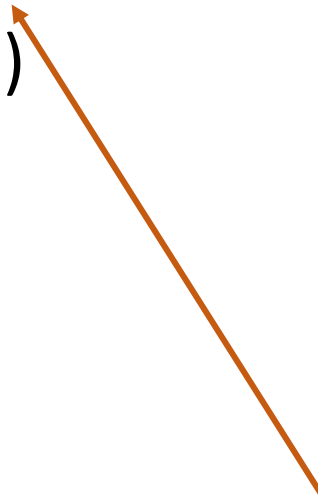
$$\left(0.153 - 1.96 * \sqrt{0.0000259182}; 0.153 + 1.96 * \sqrt{0.0000259182} \right)$$

$$\left(0.153 - 1.96 * 0.005; 0.153 + 1.96 * 0.005 \right)$$

$$\left(0.153 - 0.0099; 0.153 + 0.0099 \right)$$

$$\left(0.143; 0.163 \right)$$

$$\left(14.3\%; 16.3\% \right)$$



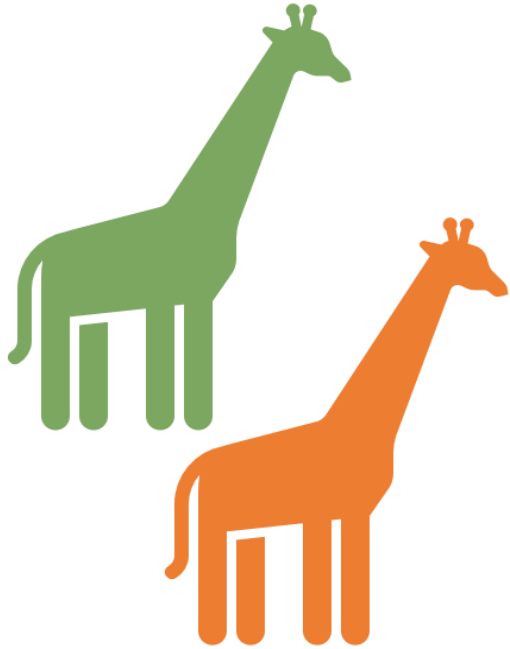


Statistical tests

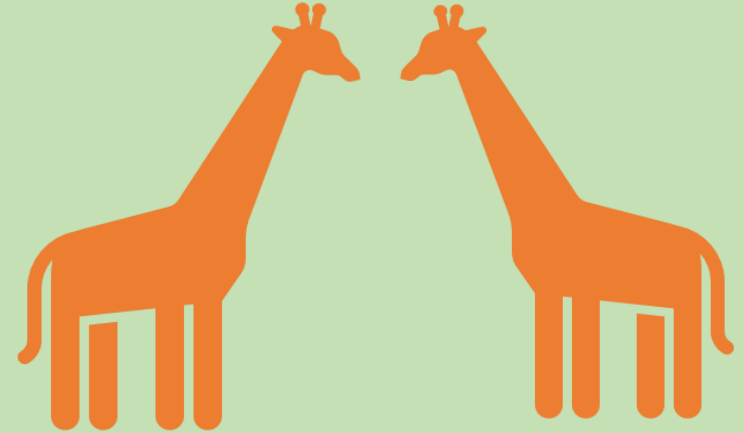
- Populations P1, P2,...
- Objective: study if there is differences between two frequencies in two/more populations (categories of the grouping variable)
- Frequencies ϕ_1  , ϕ_2  , ...

Study objective

Alternative hypothesis H1:
Different populations $\varphi_1 \neq \varphi_2$



Null hypothesis H0:
The same population: $\varphi_1 = \varphi_2$



we want to prove that this is false

Reductio ad absurdum - proof by contradiction, a proposition is proved true by proving that it is impossible for it to be false.

What is happening at the population level?

Test the **null hypothesis** that we do not have two different populations, but the same population from the p. of v. of the frequencies of the tested variable

Null hypothesis H₀ -implies the denial of the objective we want to investigate

Possible formulations:

- There is no statistically significant difference between groups in terms of frequency
- There is no association between 2 variables: Risk factor – disease
- The two variables are independent



Test the **null hypothesis** that we do not have two different populations, but the same population from the p. of v. of the frequencies of the tested variable

Null hypothesis H0 -implies the denial of the objective we want to investigate:

There is no statistically significant difference between groups in terms of frequency

There is no statistically significant association between 2 variables: Risk factor – disease

The two variables are independent

Alternative hypothesis H1 (negation of H0): refers to the objective we want to investigate

There is a statistically significant difference between groups in terms of frequency

There is a statistically significant association between 2 variables: Risk factor – disease

The two variables are dependent

Apply the statistical test --> find which of the hypotheses is true: H0 or H1

- Populations P1, P2,...

X, Y – qualitative variables

- Objective: study if there is differences between two frequencies in two/more populations (categories of the grouping variable)

- Frequencies ϕ_1 , ϕ_2 , ...

1.

- Extract randomly representative samples

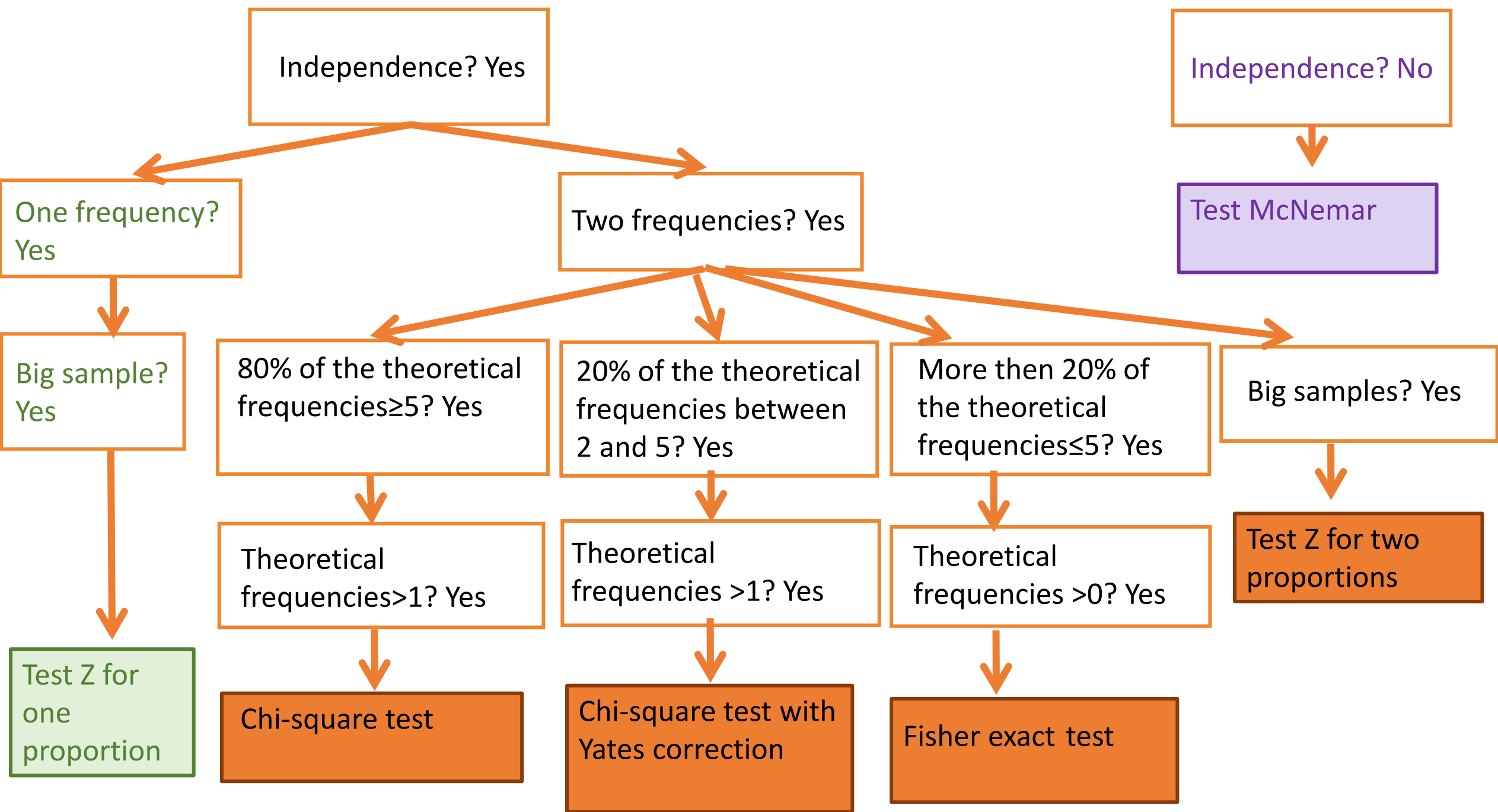
2.

- Calculate the statistics on the samples
 - Frequencies f1, f2

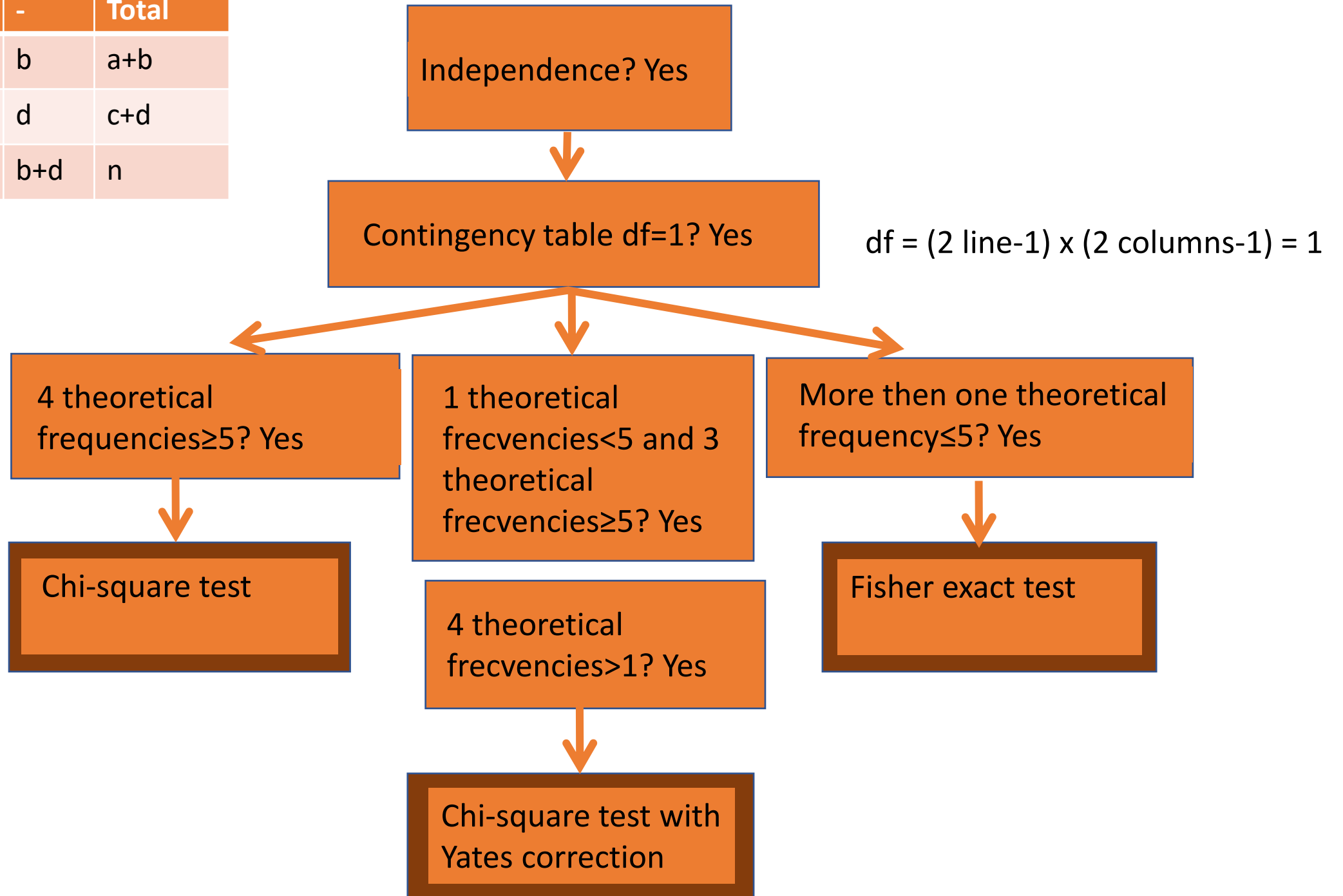
How?

3.

- Through a statistical test
 - test the hypothesis that we do not have two different populations, but the same population from the p. of v. of the frequencies of the tested variable
 - result: the probability p – the probability of finding differences equal to or greater than the one found on the sample if we repeat the study



	+	-	Total
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n



	With dental implant failure	Without dental implant failure	Total
Smokers	50	1350	1400
Non-smokers	75	1525	1600
Total	125	2875	3000

Observed contingency table

	With dental implant failure	Without dental implant failure	Total
Smokers	$=\frac{125 \cdot 1400}{3000} = 58$	$=\frac{2875 \cdot 1400}{3000} = 1342$	1400
Non-smokers	$=\frac{125 \cdot 1600}{3000} = 67$	$=\frac{2875 \cdot 1600}{3000} = 1533$	1600
Total	125	2875	3000

Theoretical contingency table (null table)

Example

Example - Theoretical contingency table (null table)

	With dental implant failure	Without dental implant failure	Total
Smokers	58	1342	
Non-smokers	67	1533	
Total			

4 theoretical frequencies ≥ 5 ? Yes

→ Chi square test

	With dental implant failure	Without dental implant failure	Total
Smokers	3	1397	
Non-smokers	67	1533	
Total			

- 1 theoretical frequencies < 5 and 3 theoretical frequencies ≥ 5 ? Yes

- 4 theoretical frequencies > 1 ? Yes

→ Chi-square test with Yates correction

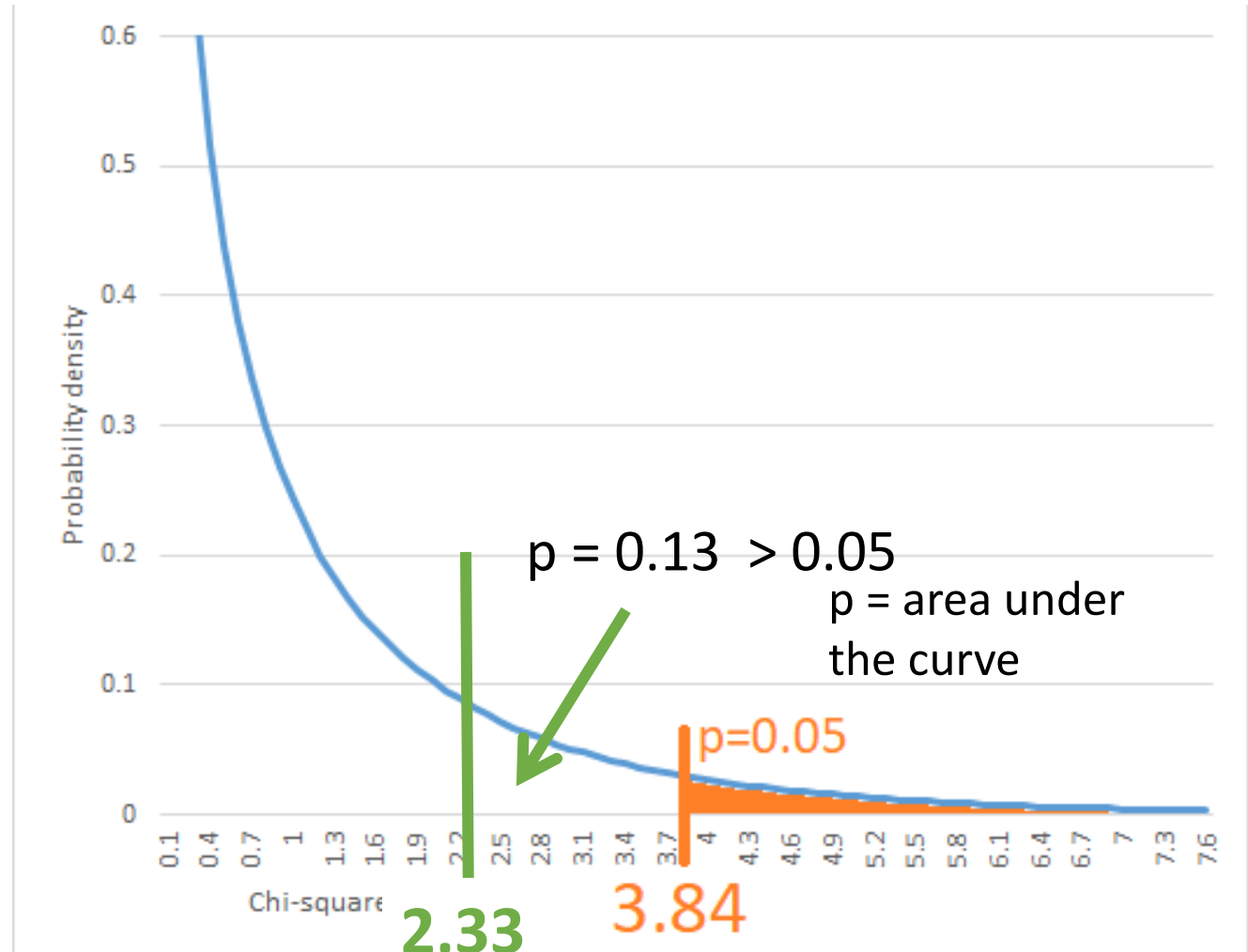
	With dental implant failure	Without dental implant failure	Total
Smokers	3	1397	
Non-smokers	4	1596	
Total			

- More than one theoretical frequency ≤ 5 ? Yes

→ Fisher-exact test

Test results – Chi-square test

$\chi^2 = 2.33$ is not $(3.84, +\infty)$



Results interpretations

$\chi^2 = 2.33 \notin [3.84, \infty)$ fail to reject null hypothesis: smoking and dental implant are independent

Or

$p=0.13 > 0.05$ fail to reject null hypothesis: smoking and dental implant are independent

A study was conducted to evaluate the association between the women age groups (20–30 and 31–40 years) and whether they have iron deficiency. Which of the following options would be the most appropriate method of analysing these data?

A. Correlation analysis

B. ANOVA test

C. Chi-square test

D. Student t test

E. none of the above

To apply a Chi-Square test

- A. one theoretical frequencies can be less than 5
- B. more than one theoretical frequencies can be less than 5
- C. no theoretical frequencies should be 0
- D. no theoretical frequencies should be less than 5
- E. all theoretical frequencies should be higher than 1

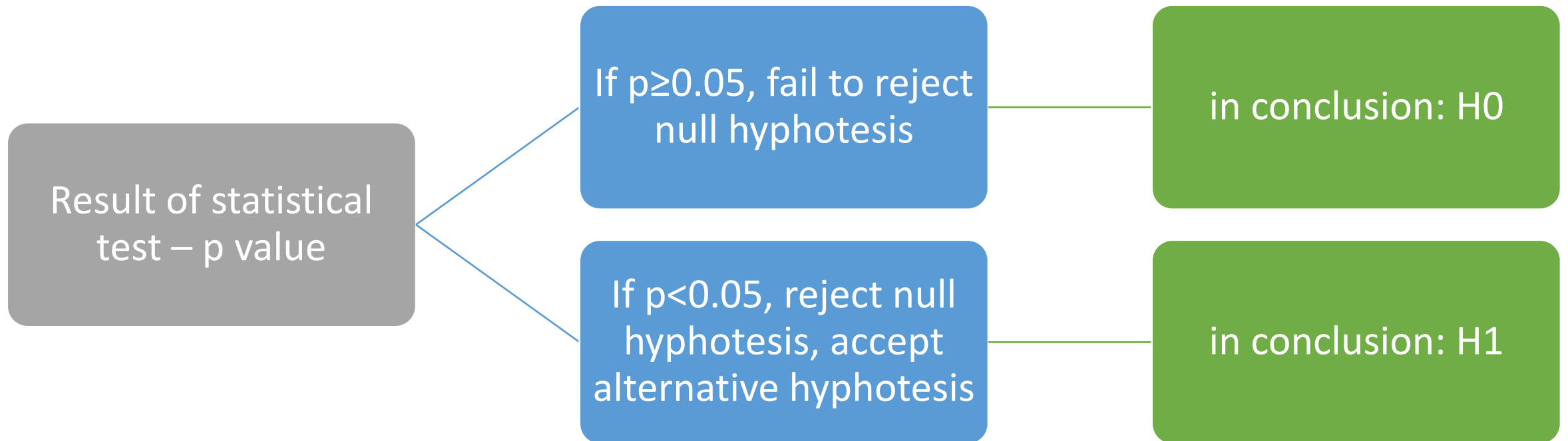
- From 3000 patients with dental implants 125 experienced early implant loss. From 125 with early implant loss 50 are smokers. From 3000 patients 1600 are non-smokers. Question: People who smoke are more likely to have dental implant failure?
- A. the parameter of the test is 1.33
- B. the parameter of the test is 2.33
- C. we reject null hypothesis and accept alternative hypothesis
- D. there was an association between smoking and dental implant failure
- E. fail to reject null hypothesis: smoking and dental implant are independent

3. Drinking coca cola (daily) is suspected to be associated with apparition of tooth caries. A sample of 500 persons was studied: 220 presented at least one caries and from these 100 reported to be coca-cola drinkers. A number of 210 patients without caries and no coca-cola drinkers were identified. The alternative hypothesis of the test could be:

- A. No answer is correct
- B. Daily coca-cola drinking and tooth caries are dependent
- C. There is an association between tooth caries and daily coca-cola drinking
- D. Daily coca-cola drinking and tooth caries are independent
- E. Tooth caries and daily coca-cola dinking are not associated

p value interpretation

- H0 – null hypothesis
- H1 – alternative hypothesis



24) *Suppose that the p-value in a hypothesis test is 0.08. If the significance level for the test is $\alpha = 0.05$, which of the following is the appropriate decision?

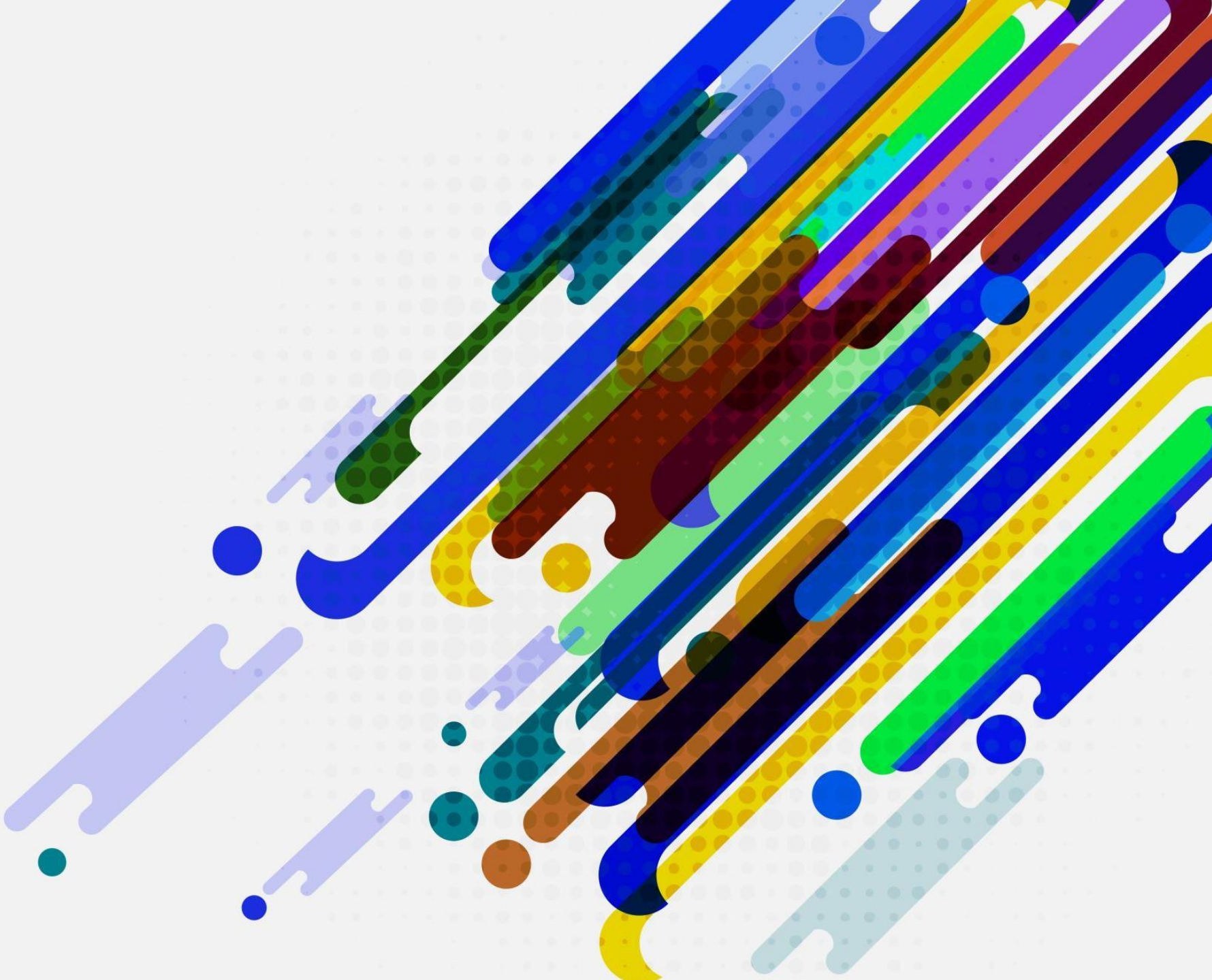
a) Fail to reject H_0

b) Reject H_0



c) There is not enough information given to know whether or not H_0 should be rejected

d) Fail to reject H_a

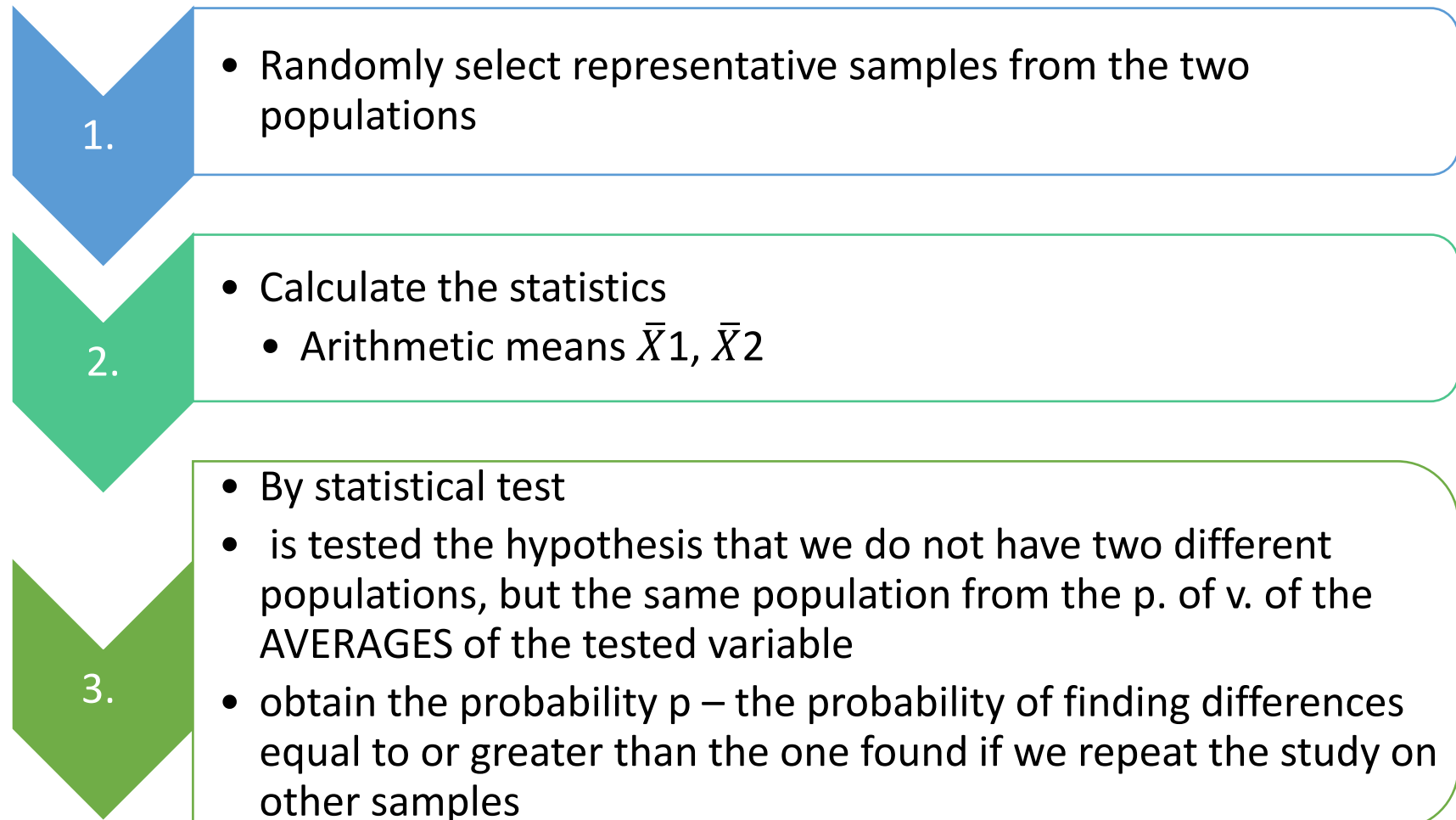
e) Reject H_a



Statistical test
for arithmetic
means

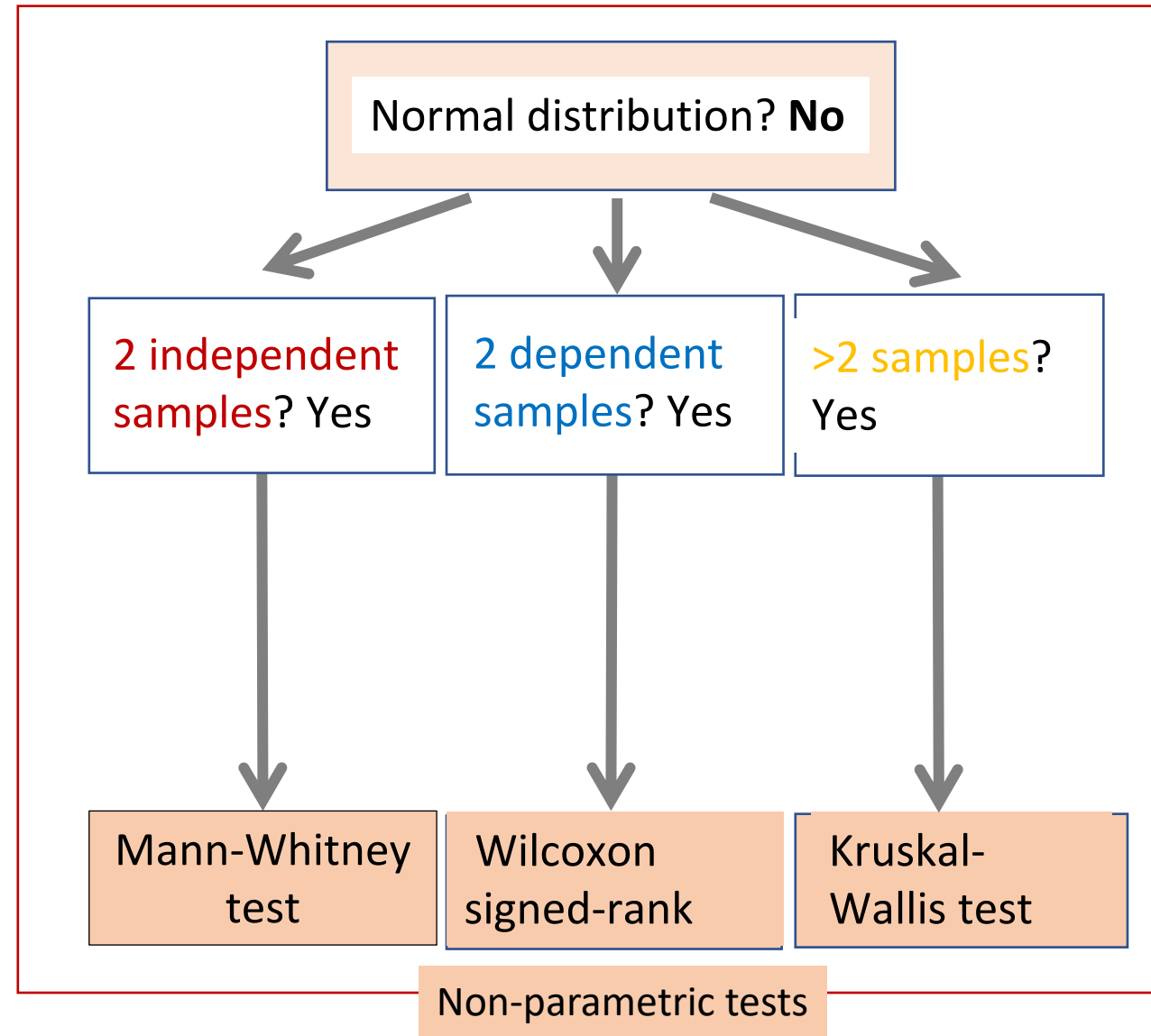
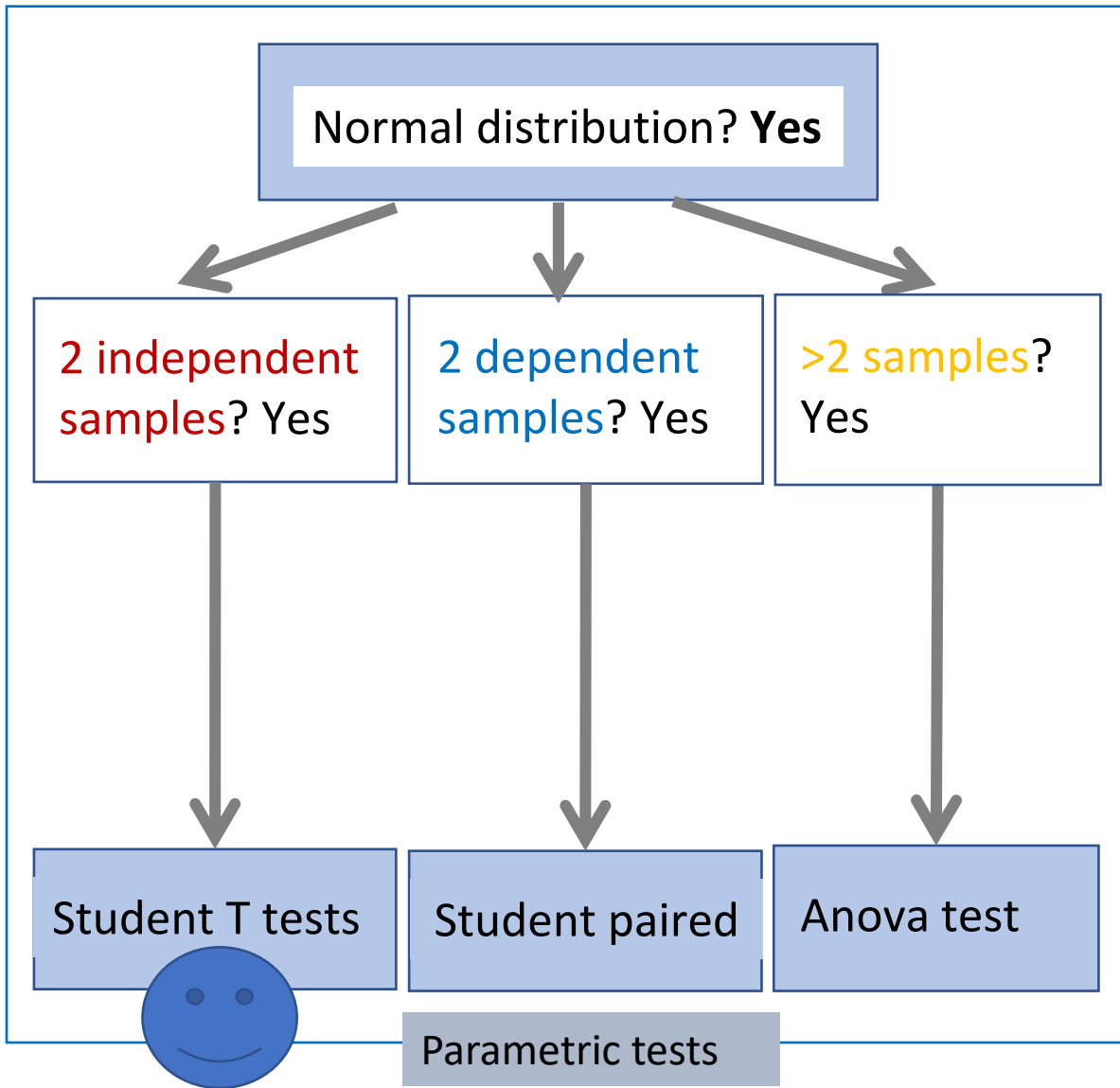
- Populations P1, P2,...
- Objective: study of the differences of the parameters of a QUANTITATIVE variable X in two/more populations (categories of the grouping variable Y)
- Arithmetic means μ_1  , μ_2  , ...

How?



simplified to:

Comparing quantitative variables grouped by a qualitative variable. How to choose the right statistical test?



2 groups

```
graph LR; A[2 groups] --- B[independent]; A --- C[dependent];
```

The diagram consists of a dark green rounded rectangle on the left containing the text '2 groups'. Two orange rounded rectangles are positioned to the right, one above the other. Two thin orange lines originate from the right side of the green box and connect to the left side of each orange box, forming a branching structure.

independent

dependent

Dependent or independent samples?

- **Dependent samples**

- the same patients tested two times
- after/ before treatments, procedure etc.
- data are in paires
- a pair – from the same patient

- **Ex. Compare the weight after and before the diet**

- each patient are tested after and before diet
- patient one: John
 - before 95kg
 - after 65kg
- etc.

} a pair of correspondent data

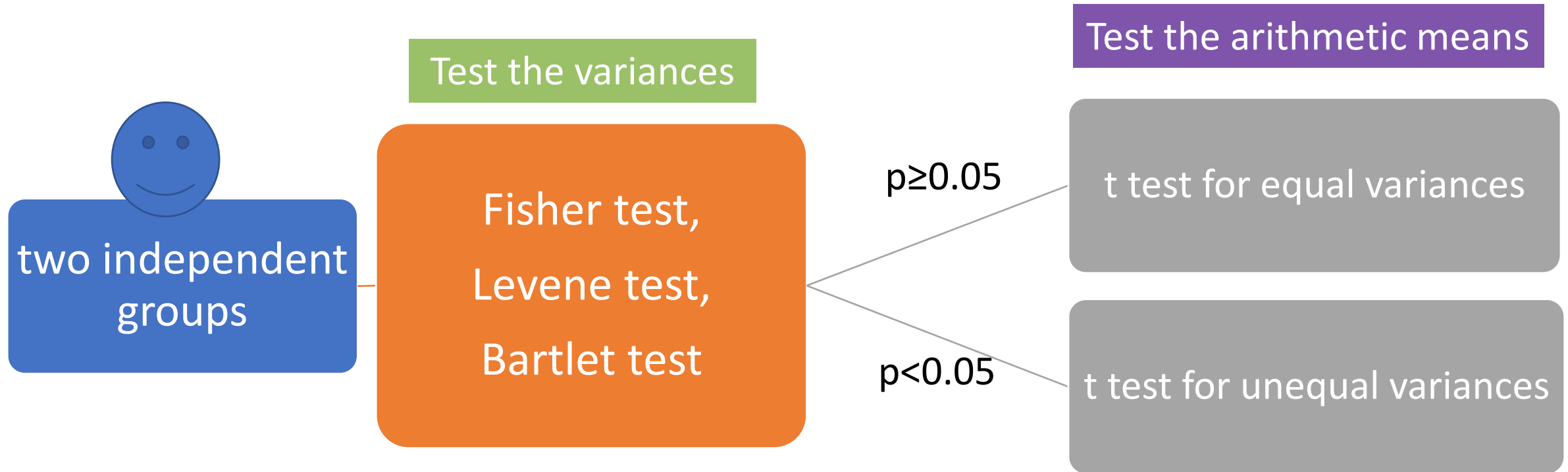
- **Independent samples**

- patients are not the same
- different groups
- data cannot be put in paires
- **Ex. Compare the weight on people who take diet with those without any diet**
 - we have 2 different groups

60 patients follow a diet }
80 patients do not follow a diet } no correspondence between data

- A study was conducted to compare 23 patients who drink Coca-Cola daily with 36 patients who do not. They were evaluated to assess the level of advanced glycation end products in their skin with autofluorescence (SAF) (%). Data were normally distributed. The following can be used to compare the SAF in these two groups:
 - A. a nonparametric test
 - B. a parametric test
 - C. the Wilcoxon test
 - D. the t-test for independent samples
 - E. the Mann-Whitney test

Comparing quantitative variables grouped by a qualitative variable – two independent groups with normal distributions



Variances = standard deviation square = s^2

The hypothesis that we do not have two different populations, but the same population in terms of the distribution of the means of the variable under study

Null hypothesis H_0 - assumes the denial of the objective we want to investigate

- There is no statistically significant difference between the groups in terms of the mean

Alternative hypothesis H_1 (denial of H_0): refers to the objective we want to investigate

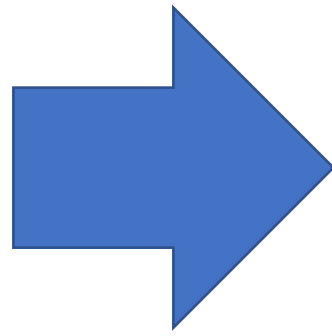
- There is a statistically significant difference between the groups in terms of the mean

Statistical test --> we choose between the two possibilities H_0 or H_1

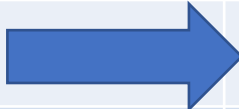
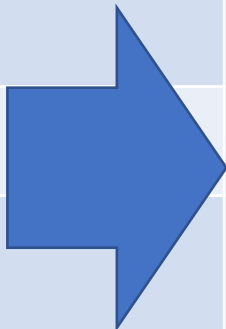
Example

	Obese (n=15)	Non-obese (n=15)
Glycemia (mg/dl) Arithmetic mean±Standard deviation	160±20	95±10

Normal distribution?



Shapiro-Wilk test for both samples taken separately

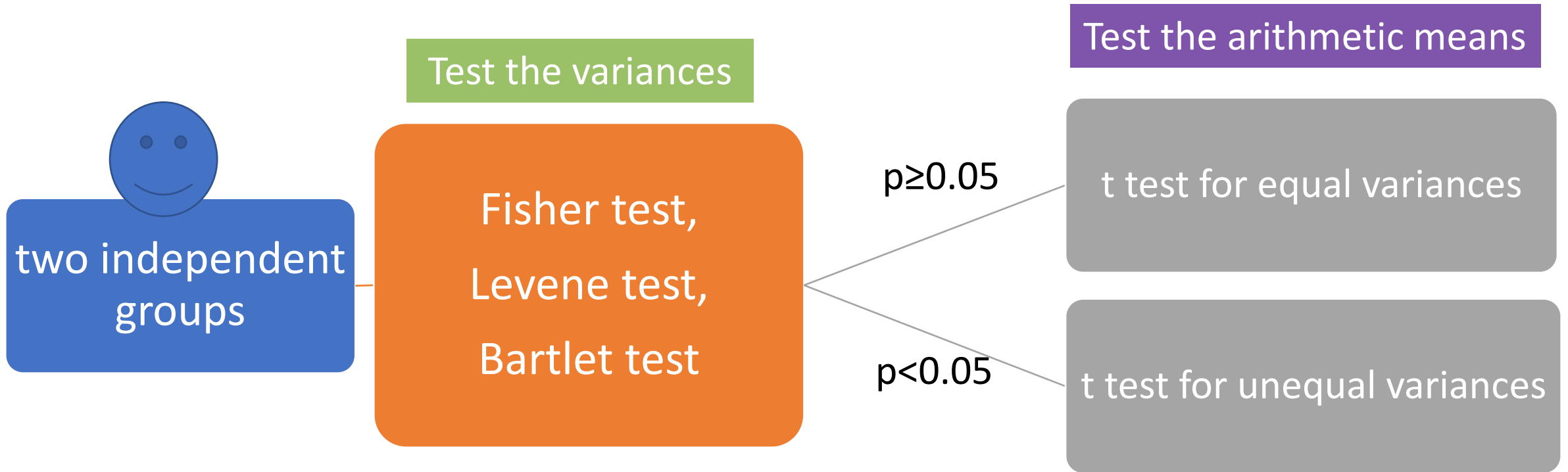
	p value for normal distribution of glycemia		Normal distribution Obese/Non-obese	Test the arithmetic mean differences with
	Obese (n=15)	Non-obese (n=15)		
Example 1	0.25	0.43	Yes/Yes 	Student t test
Example 2	0.03	0.06	No/Yes	Mann-Whitney test 
Example 3	0.15	0.001	Yes/No	
Example 4	0.02	0.001	No/No	

If data are normally distributed, how are the variances?

- **Null hypothesis H₀**: There is no significant statistical difference between the **variances** of glycemia in obese versus non-obese individuals
- **Alternative hypothesis H_a**: There is significant statistical difference between the **variances** of glycemia in obese versus non-obese individuals

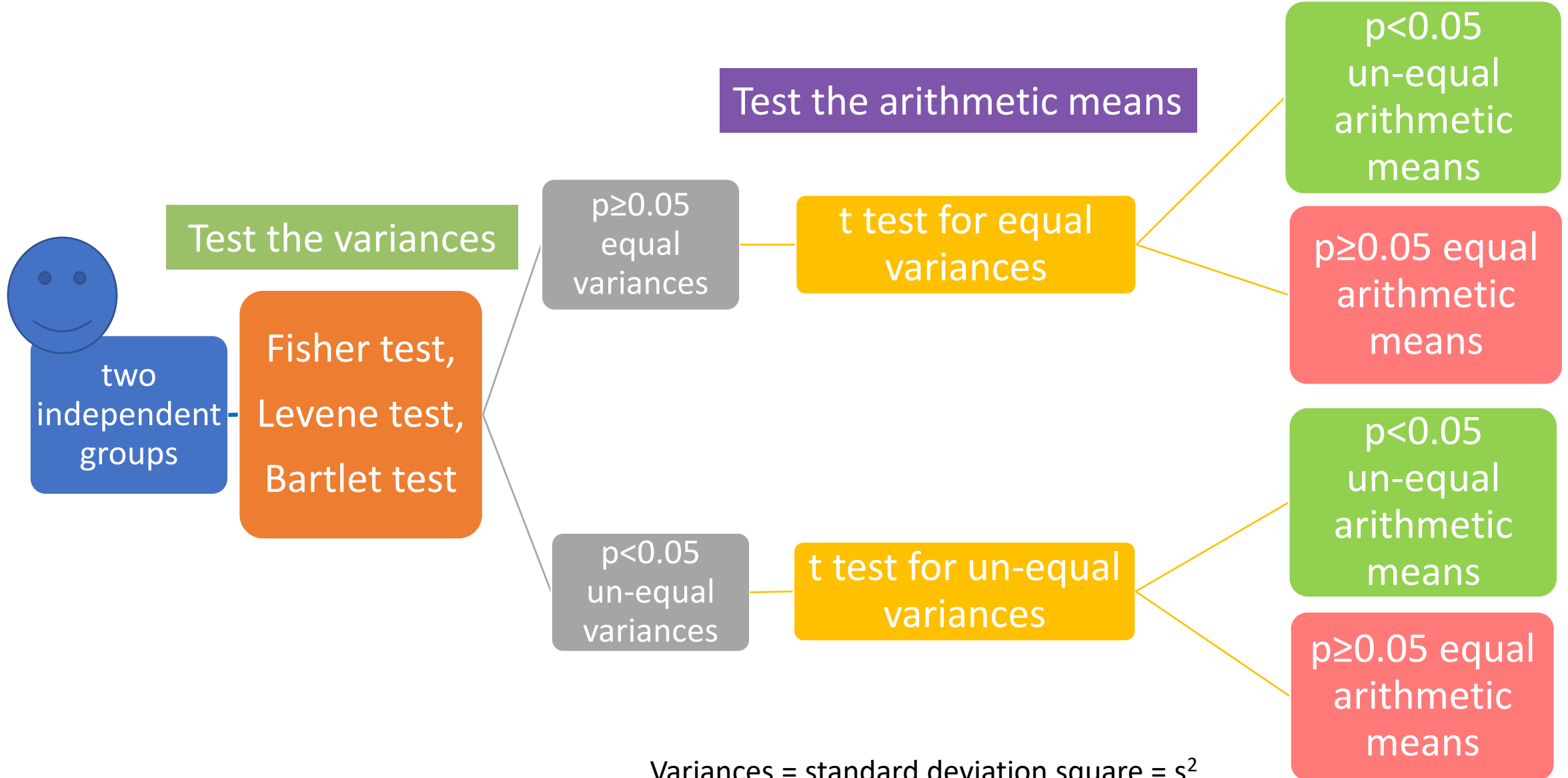
		Obese (n=15)	Non-obese (n=15)	How are the variances?
Example 1	Glycemia (mg/dl) Arith.mean±St.dev.	160±20	95±10	20 ² compare with 10 ²

Comparing quantitative variables grouped by a qualitative variable – two independent groups with normal distributions



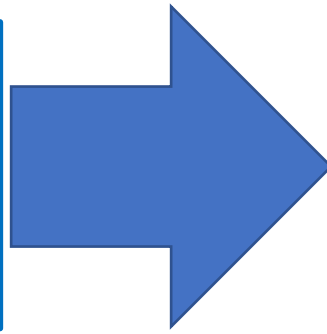
Variances = standard deviation square = s^2

Comparing quantitative variables grouped by a qualitative variable – two independent groups with normal distributions



Variances = standard deviation square = s^2

- Two independent samples
- Two **variances** to compare
- Data are **normally distributed**



Levene Test for variances

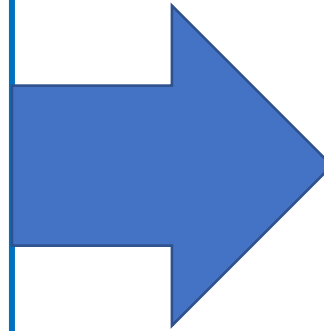
		Obese (n=15)	Non-obese (n=15)	p Levene test for variances	
Example 1	Glycemia (mg/dl) Arith.mean±St.dev.	20 ²	10 ²	0.04	un-equal variances

		Obese (n=15)	Non-obese (n=15)	p Levene test for variances	
Example 1	Glycemia (mg/dl) Arith.mean±St.dev.	20 ²	suppose is 15 ²	0.12	equal variances

- *A study was conducted to compare 23 patients who drink Coca-Cola daily with 36 patients who do not. They were evaluated to assess the level of advanced glycation end products in their skin with autofluorescence (SAF) (%). Data were normally distributed. The p value for the Levene test was $p=0.09$. The following test can be used to compare the SAF in these two groups:
 - A. the t-test for dependent samples
 - B. **the t-test for independent samples for equal variances**
 - C. the Wilcoxon test
 - D. the t-test for independent samples for un-equal variances
 - E. the Mann-Whitney test

If variances are un-equal, how are the means?

- Two independent samples
- Two means to compare
- **Un-equal** variances
- Normal distributions (n_1 or $n_2 < 30$)



T test (Student) for independent samples for **un-equal** variances

H₀: There is no statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

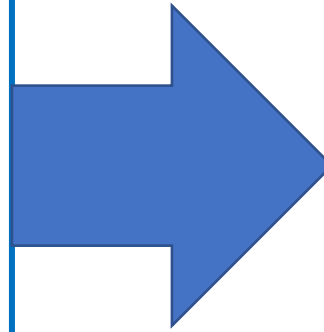
H_a: There is statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

		Obese (n=15)	Non-obese (n=15)	p T test (Student) for un-equal variances
Example 1	Glycemia (mg/dl) Arith.mean±St.dev	160 ±20	95 ±10	0.01

$p=0.01 < 0.05$ we reject null hypothesis, accept alternative hypothesis: There is statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

If variances are un-equal, how are the means?

- Two independent samples
- Two means to compare
- **Un-equal** variances
- Normal distributions (n_1 or $n_2 < 30$)



T test (Student) for independent samples for **un-equal** variances (**Welch test**)

H0: There is no statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

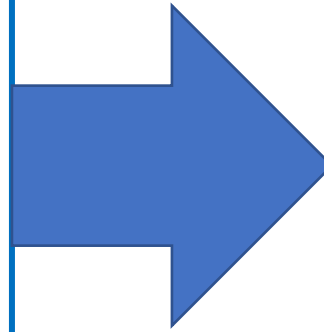
Ha: There is statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

		Obese (n=15)	Non-obese (n=15)	p T test (Student) for un-equal variances
Example 1	Glycemia (mg/dl) Arith.mean±St.dev	160 ±20	suppose 120 ±10	0.08

$p=0.08 > 0.05$ we fail to reject null hypothesis: There is no statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

If variances are **equal**, how are the means?

- Two independent samples
- Two means to compare
- **Equal** variances
- Normal distributions (n_1 or $n_2 < 30$)



T test (Student) for independent samples for **equal** variances

H0: There is no statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

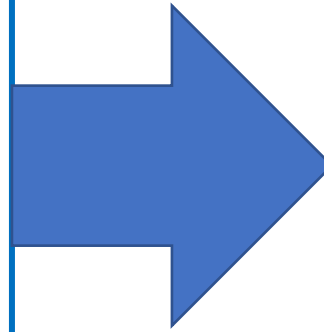
Ha: There is statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

		Obese (n=15)	Non-obese (n=15)	p T test (Student) for equal variances
Example 1	Glycemia (mg/dl) Arith.mean±St.dev	160 ±20	95 ±15	0.02

$p=0.02 < 0.05$ we reject null hypothesis, accept alternative hypothesis: There is statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

If variances are **equal**, how are the means?

- Two independent samples
- Two means to compare
- **Equal** variances
- Normal distributions (n_1 or $n_2 < 30$)



T test (Student) for independent samples for **equal** variances

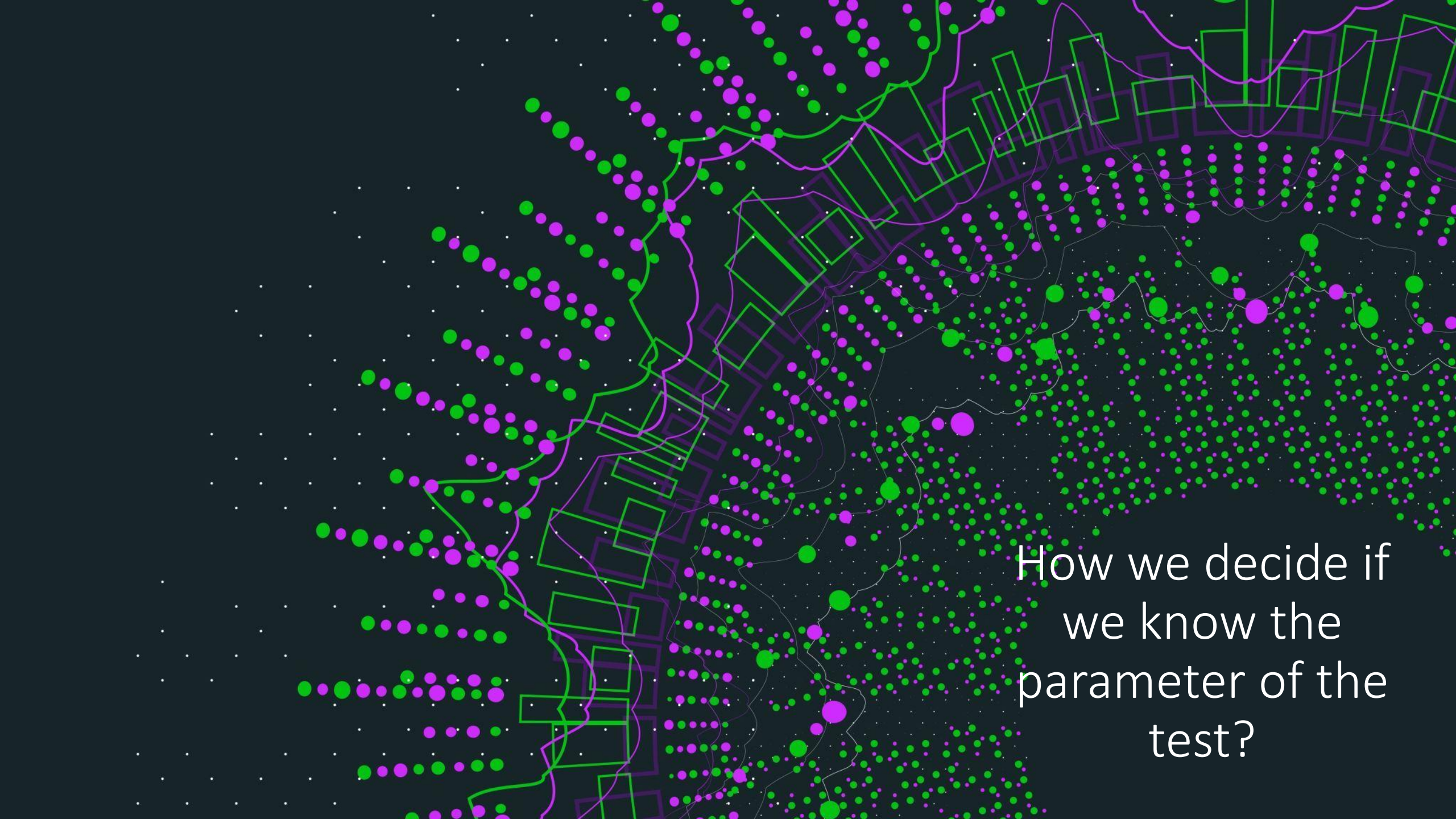
H0: There is no statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

Ha: There is statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

		Obese (n=15)	Non-obese (n=15)	p T test (Student) for equal variances
Example 1	Glycemia (mg/dl) Arith.mean±St.dev	160 ±20	suppose 150 ±15	0.32

$p=0.32 > 0.05$ we fail to reject null hypothesis: There is no statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

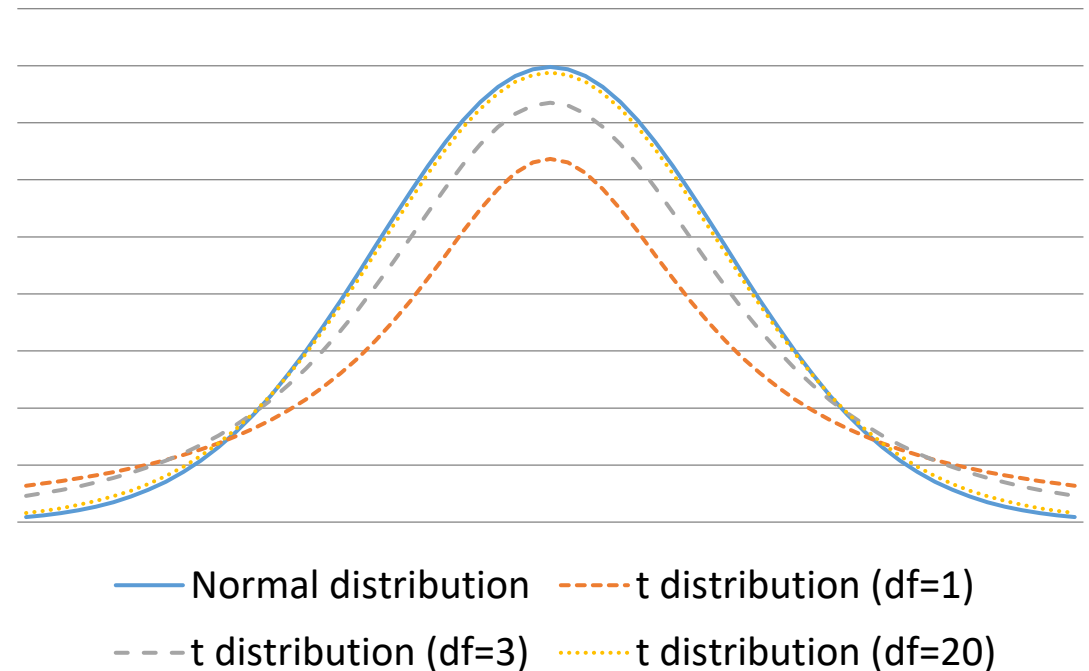
- A study was conducted to compare 23 patients who drink Coca-Cola daily with 36 patients who do not. They were evaluated to assess the level of advanced glycation end products in their skin with autofluorescence (SAF) (%). Data were normally distributed. The p value for the Levene test was $p=0.04$. The p value for the test for the means was $p=0.003$. The following are true:
 - A. We reject the null hypothesis and accept the alternative hypothesis; there is a statistically significant difference in SAF between the group that drinks Coca-Cola daily and the group that does not.
 - B. the t-test for independent samples for equal variances was used to compare the arithmetic means
 - C. We fail to reject the null hypothesis; there is no statistically significant difference in SAF between the group that drinks Coca-Cola daily and the group that does not.
 - D. the t-test for independent samples for un-equal variances was used to compare the arithmetic means
 - E. none of the above



How we decide if
we know the
parameter of the
test?

Student t distribution depend on the degree of freedom df

- $df = n_1 + n_2 - 2$
- ex. $df = 15 + 15 - 2 = 28$
- If $n > 30$ student distribution is near the same with normal distribution



Choosing the significance level and establishing the critical region

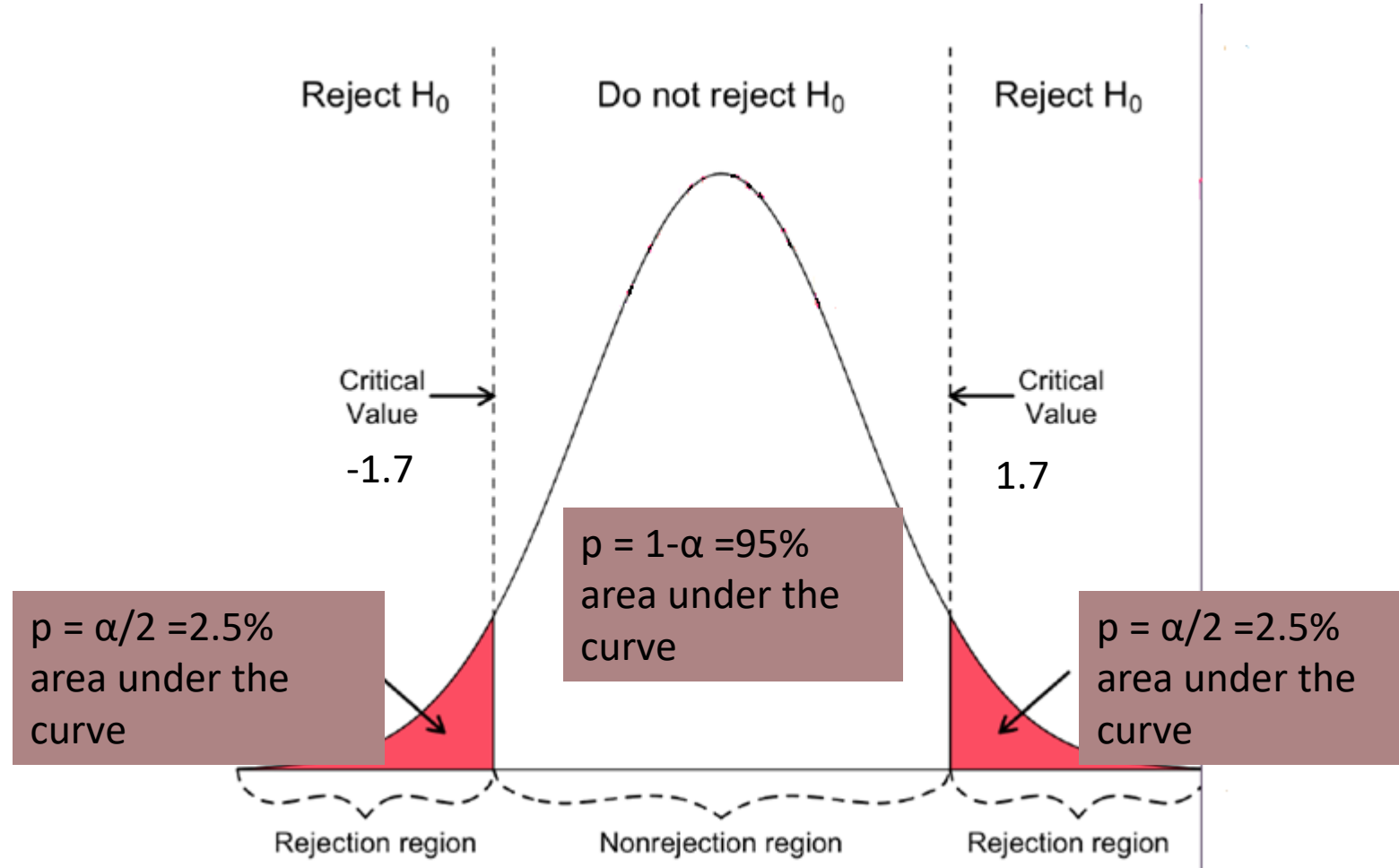
Two tail test

Rejection area

$$(-\infty, -1.7] \cup [1.7; +\infty)$$

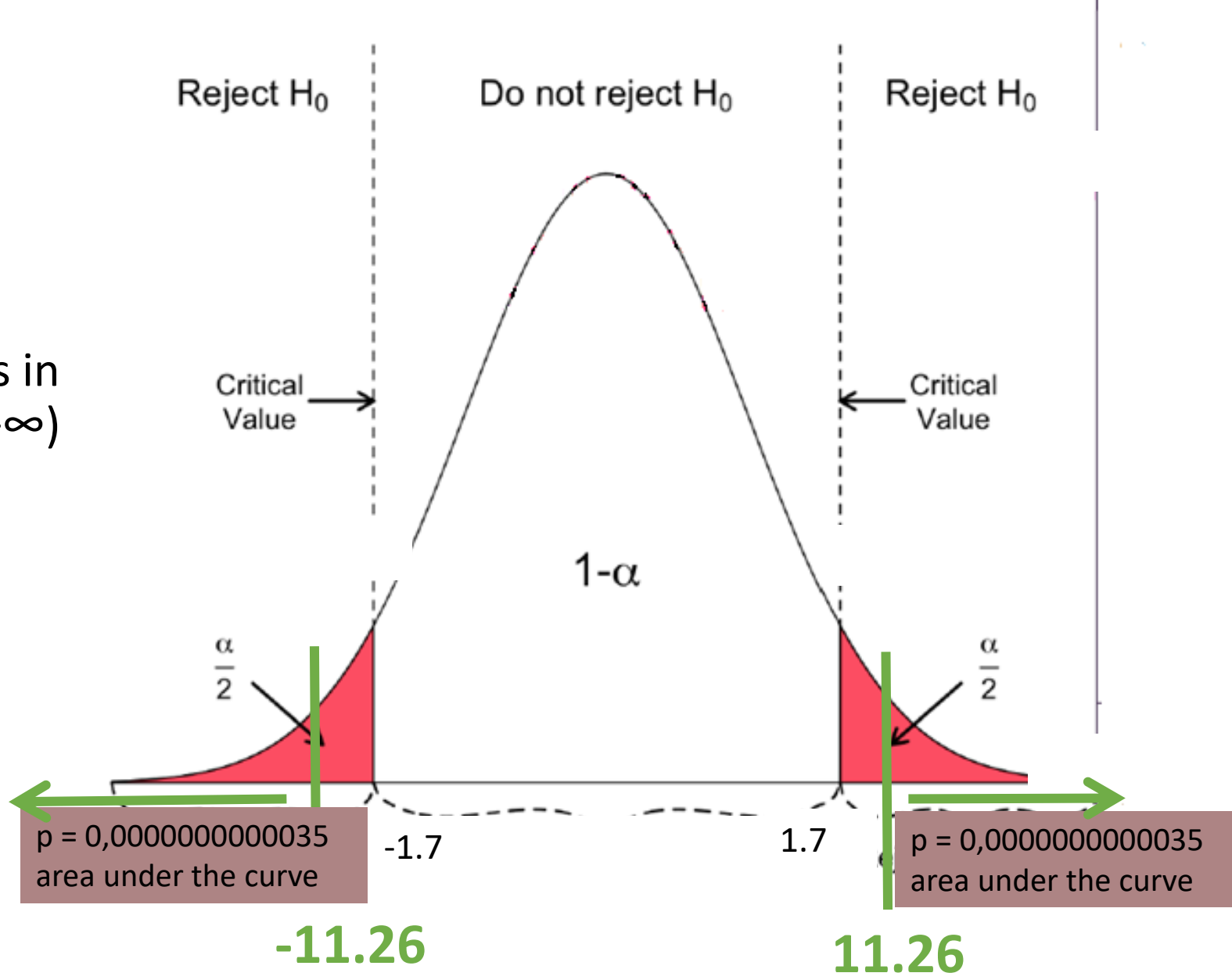
Acceptance area

$$(-1.7; 1.7)$$



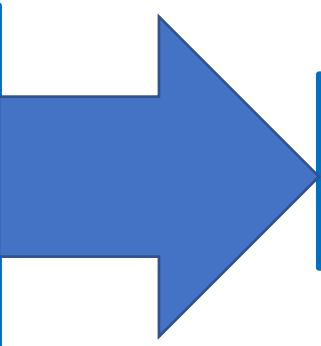
Two-tail test = we sum right and left area

the parameter of the test $t = 11.26$ is in the rejection area $(-\infty, -1.7] \cup [1.7; +\infty)$
we reject H_0 , accept H_1



If variances are **equal**, how are the means?

- Two independent samples
- Two means to compare
- **Equal** variances
- Normal distributions (n_1 or $n_2 < 30$)



T test (Student) for independent samples for **equal** variances

H0: There is no statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

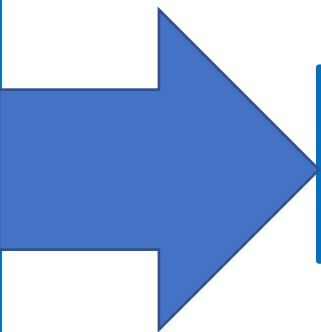
Ha: There is statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

		Obese (n=15)	Non-obese (n=15)	parameter of the T-test (Student) for equal variances
Example 1	Glycemia (mg/dl) Arith.mean±St.dev	160±20	95±15	11.26

$t=11.26$ is in the rejection area $(-\infty, -1.7] \cup [1.7; +\infty)$ we reject null hypothesis, accept alternative hypothesis: There is statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

If variances are **equal**, how are the means?

- Two independent samples
- Two means to compare
- **Equal** variances
- Normal distributions (n_1 or $n_2 < 30$)



T test (Student) for independent samples for **equal** variances

H0: There is no statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

Ha: There is statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

		Obese (n=15)	Non-obese (n=15)	parameter of the T-test (Student) for equal variances
Example 1	Glycemia (mg/dl) Arith.mean±St.dev	160 ±20	suppose 150 ±15	1.3

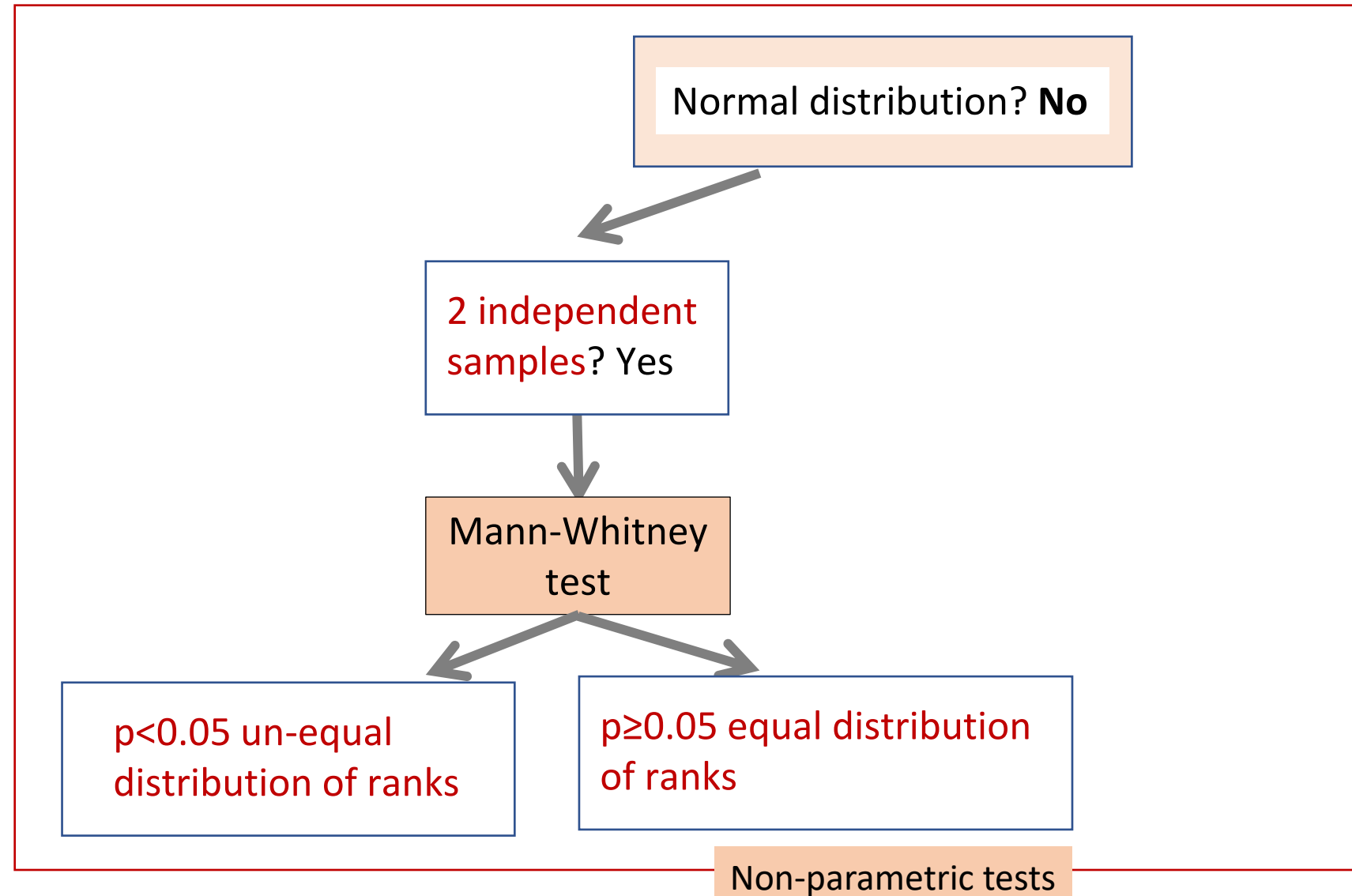
$t=1.3$ is not in the rejection area $(-\infty, -1.7] \cup [1.7; +\infty)$ we fail to reject null hypothesis: There is no statistical significant difference between the **arithmetic means** of glycemia in obese versus non-obese individuals

- A study was conducted to compare 23 patients who drink Coca-Cola daily with 36 patients who do not. They were evaluated to assess the level of advanced glycation end products in their skin with autofluorescence (SAF) (%). Data were normally distributed. The p value for the Levene test was $p=0.46$. The parameter of the test was 0.13 (t critical 1.91). The following are true:
- A. We reject the null hypothesis and accept the alternative hypothesis; there is a statistically significant difference in SAF between the group that drinks Coca-Cola daily and the group that does not.
- B. the t-test for independent samples for equal variances was used to compare the arithmetic means
- C. We fail to reject the null hypothesis; there is no statistically significant difference in SAF between the group that drinks Coca-Cola daily and the group that does not.
- D. the t-test for independent samples for un-equal variances was used to compare the arithmetic means
- E. the Mann-Whitney test was used to compare the ranks

The background features a vibrant, abstract design. A horizontal line divides the image into two halves. The top half is a solid cyan color. The bottom half is filled with a complex pattern of overlapping, semi-transparent geometric shapes in various colors including purple, blue, green, yellow, and orange. The shapes include triangles, squares, and circles, creating a layered, crystalline effect. The overall aesthetic is modern and digital.

non normal distribution

Comparing quantitative variables grouped by a qualitative variable.



- **Null hypothesis H0:** There is no significant statistical difference between the distributions **of ranks** of glycemia in obese versus non-obese individuals
- **Alternative hypothesis Ha:** There is significant statistical difference between the distributions **of ranks** of glycemia in obese versus non-obese individuals

		Obese (n=15)	Non-obese (n=15)	p (Mann-Whitney test)
Example 2, 3, 4 no normal distribution	Glycemia (mg/dl) Arith.mean±St.dev	160±20	95±10	0.02

$p=0.02 < 0.05$ reject H0, accept Ha: There is significant statistical difference between the distributions **of ranks** of glycemia in obese versus non-obese individuals

- **Null hypothesis H0:** There is no significant statistical difference between the distributions **of ranks** of glycemia in obese versus non-obese individuals
- **Alternative hypothesis Ha:** There is significant statistical difference between the distributions **of ranks** of glycemia in obese versus non-obese individuals

		Obese (n=15)	Non-obese (n=15)	p (Mann-Whitney test)
Example 2, 3, 4 no normal distribution	Glycemia (mg/dl) Arith.mean±St.dev	160±20	95±10	0.15

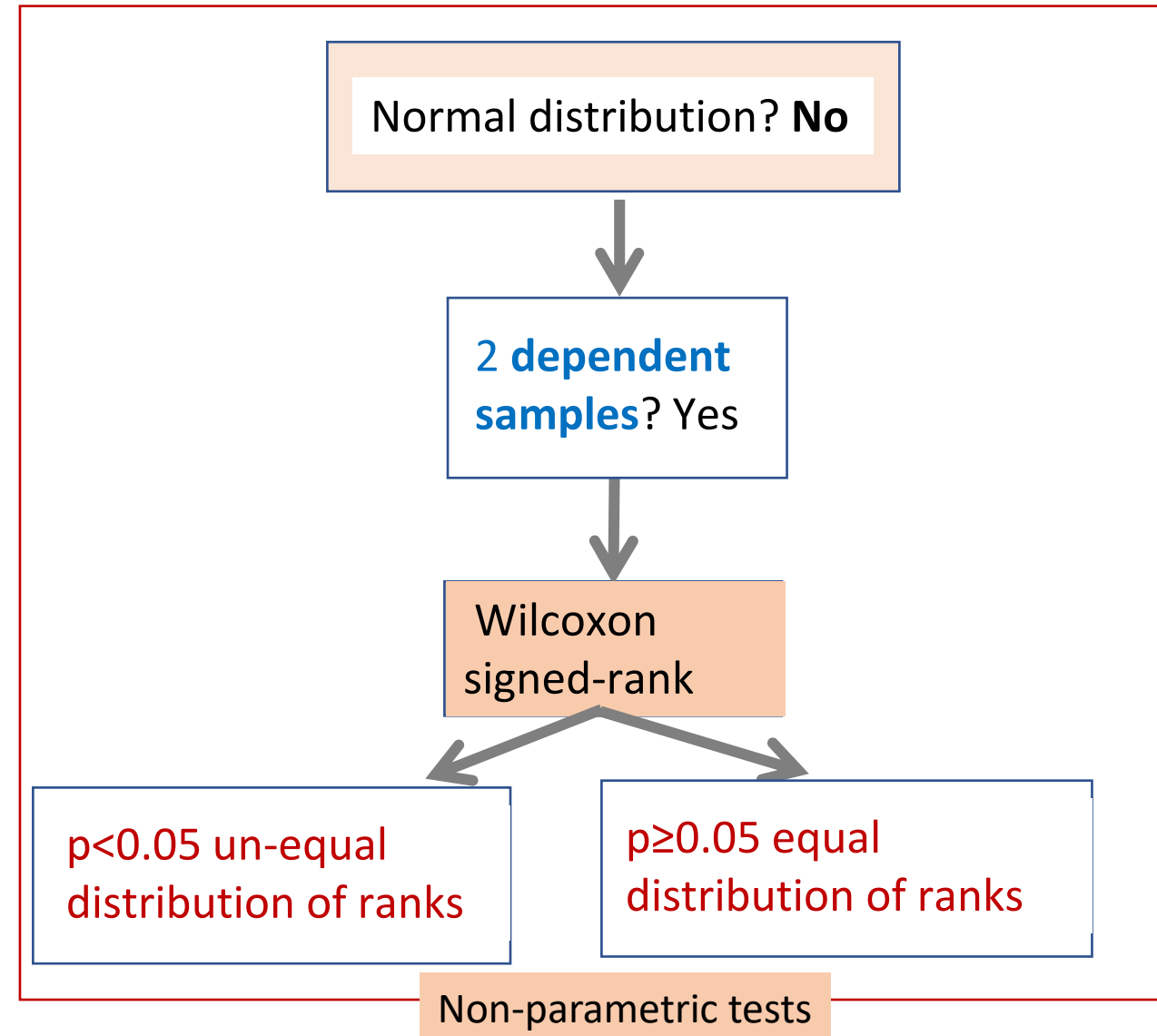
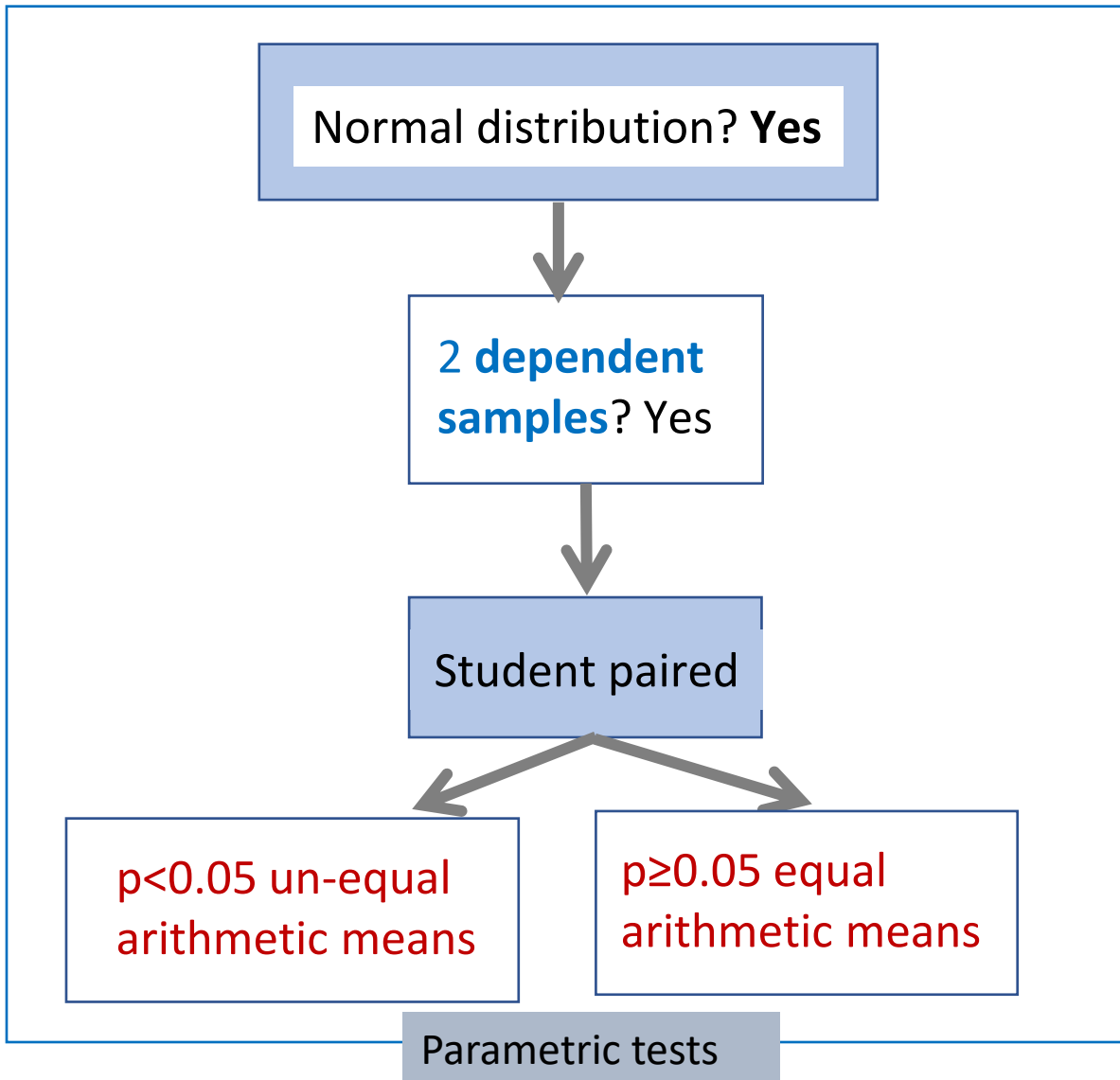
$p=0.15 > 0.05$ fail to reject H0: There is no significant statistical difference between the distributions **of ranks** of glycemia in obese versus non-obese individuals

- A study was conducted to compare 65 students taking a neurology course. They were tested to assess their ability to recognize Bell's palsy by looking at 10 images of patients with varying degrees of the condition. The students were tested before and after the course. A **score** was created with grades from 1 to 10 based on how well they recognized the 10 images. To test the scores obtained before and after the course, the following can be used:
 - **A. a nonparametric test**
 - B. a parametric test
 - **C. the Wilcoxon test**
 - D. the t-test for dependent samples
 - E. the Mann-Whitney test

The background features a vibrant, multi-colored geometric pattern. It consists of overlapping, semi-transparent shapes in shades of teal, green, yellow, orange, and purple. A horizontal rainbow gradient line runs across the middle of the image. The overall aesthetic is modern and abstract.

Dependent samples

Comparing quantitative variables grouped by a qualitative variable.

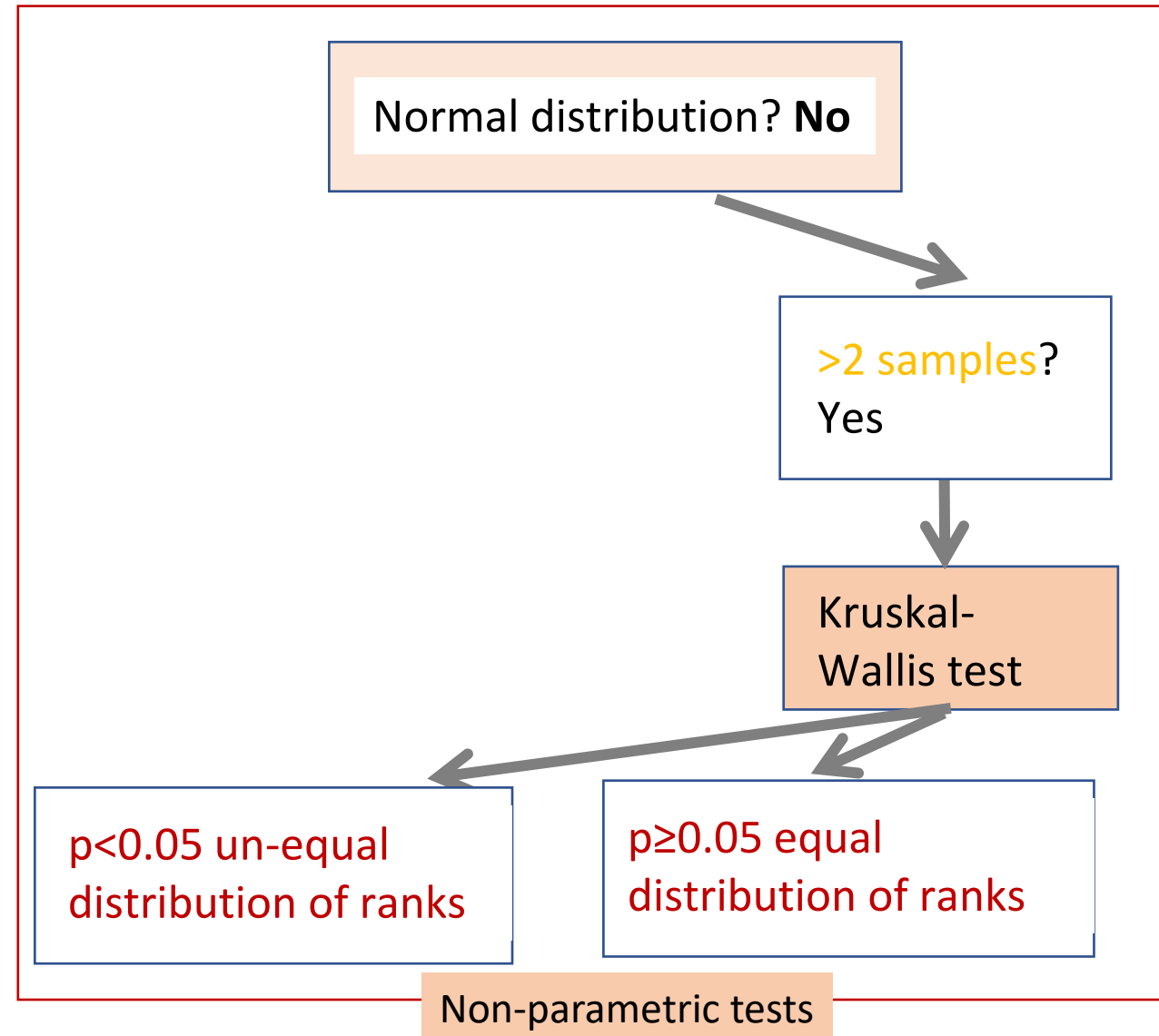
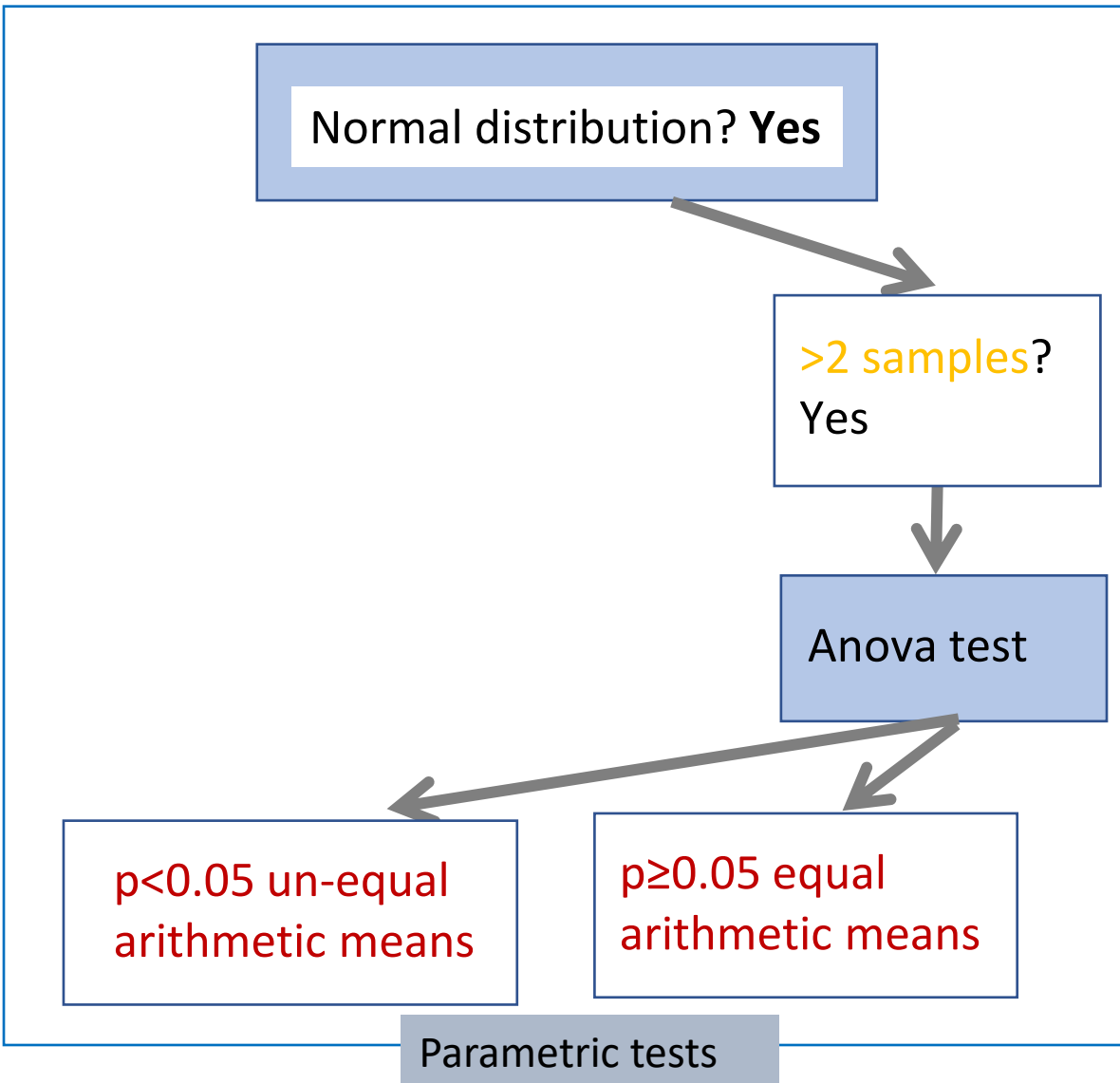


- A study was conducted to compare 55 patients with diabetes before and after a new treatment. They were evaluated to assess the level of glycemia. Data were normally distributed. The p value for the test was $p=0.67$. The following are true:
- A. We reject the null hypothesis and accept the alternative hypothesis; there is a statistically significant difference in glycemia before and after the treatment.
- B. the t-test for independent samples was used
- C. We fail to reject the null hypothesis; there is no statistically significant difference in glycemia before and after the treatment.
- D. the t-test for dependent samples was used
- E. the Mann-Whitney test was used

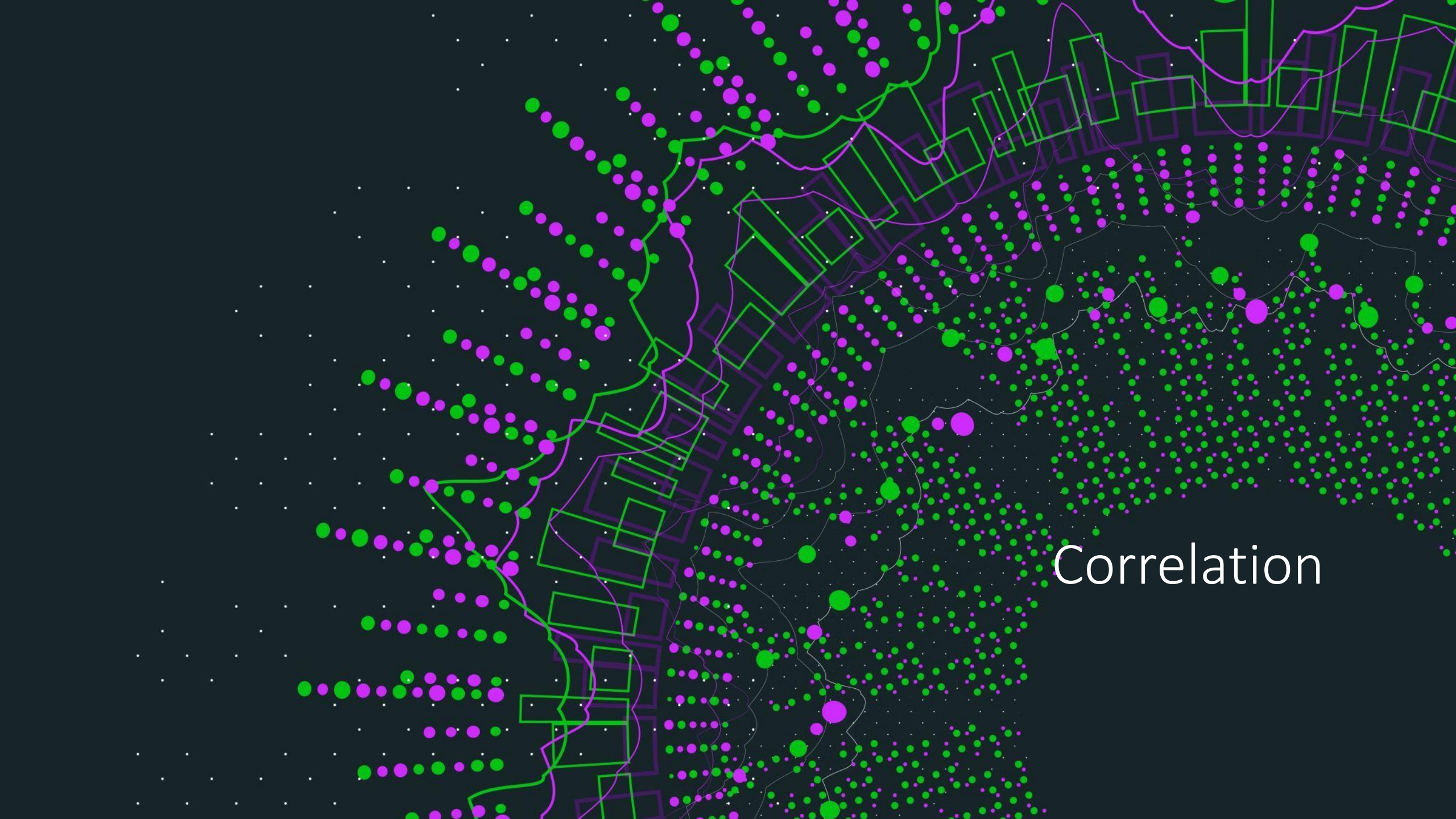
The background features a vibrant rainbow gradient from top to bottom. Overlaid on this are various semi-transparent geometric shapes, including triangles, squares, and circles, in shades of green, yellow, orange, and purple. Some shapes have a stippled or dotted texture. The overall effect is a modern, abstract digital design.

Multiple arithmetic means

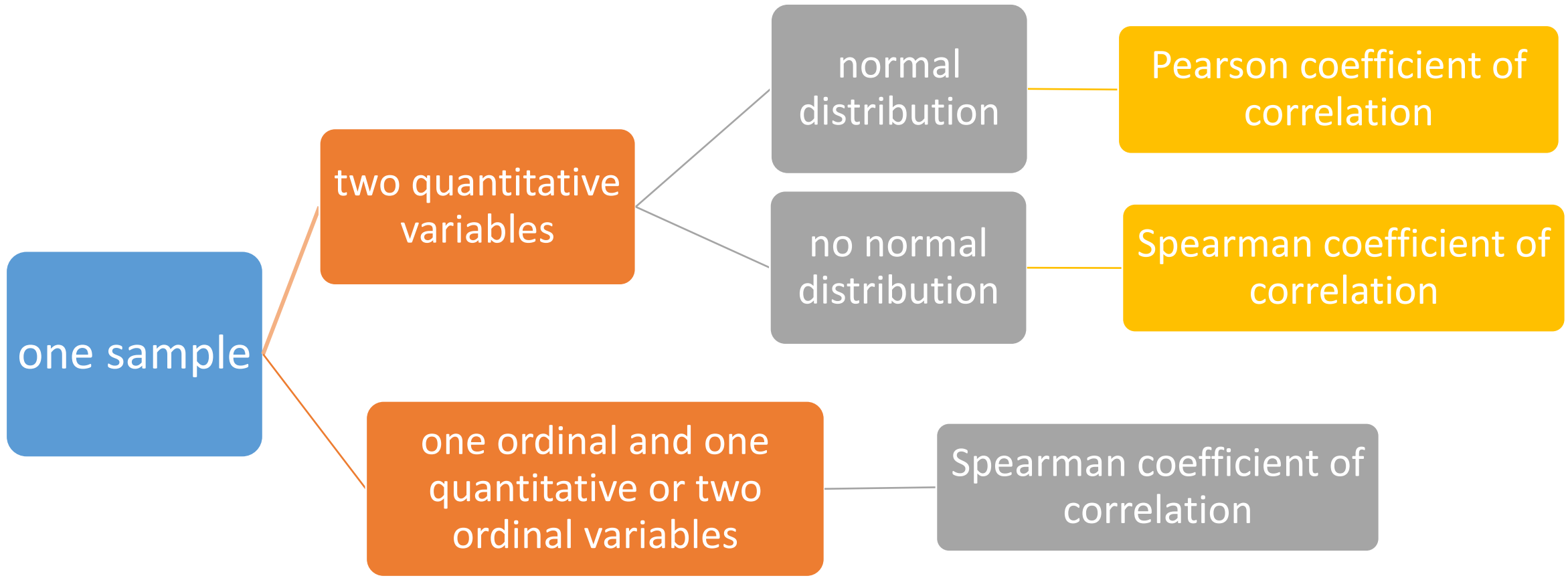
Comparing quantitative variables grouped by a qualitative variable.



- A study was conducted to compare 23 patients who drink Coca-Cola daily with, 33 patients who drink Coca-Cola occasionally and with 36 patients who do not drink Coca-Cola. They were evaluated to assess the level of advanced glycation end products in their skin with autofluorescence (SAF) (%). Data were normally distributed. The p value for the Levene test was $p=0.46$. The p value for the test for the means was $p=0.009$. The following are true:
 - A. We reject the null hypothesis and accept the alternative hypothesis; there is a statistically significant difference in SAF between the tested groups.
 - B. the t-test for independent samples for equal variances was used to compare the arithmetic means
 - C. We fail to reject the null hypothesis; there is no statistically significant difference in SAF between the tested groups.
 - D. the t-test for independent samples for un-equal variances was used to compare the arithmetic means
 - E. the Anova test was used to compare the arithmetic means



Correlation



Statistical significance test for the correlation coefficient

- Significance of the correlation coefficient - did the observed value in the sample occur by chance?
- if the statistical test is significant, the probability $p < 0.05$
- Interpretation at the population level - given by the values of r/ρ according to Colton's rules
- We can calculate a confidence interval for r/ρ
- if $p \geq 0.05$
 - interpretation - the experimental data do not allow us to assert the existence of a relationship at the population level between the variables under study
- if $p < 0.05$
 - interpretation - the experimental data allow us to assert the existence of a relationship at the population level between the variables under study

r/rho	p \geq 0.05	p<0.05
-0.25 to 0.25	without statistical significance	no correlation, significant statistically
0.25 to 0.5 or -0.5 to -0.25	without statistical significance	acceptable correlation, significant statistically
0.5 to 0.75 or -0.75 to -0.5	without statistical significance	good correlation, significant statistically
0.75 to 1 or -1 to -0.75	without statistical significance	very good correlation, significant statistically

Interpretation r/rho and p

Important

- $p \geq 0.05$ the coefficient only describes the sample from which it was drawn
- The correlated variables must be quantitative or ordinal
 - Contra example. Height and gender
- There must be a causal relationship between the variables for which correlation is calculated (the association must make sense)
- Contra example. Environmental temperature and IQ

We compute Pearson coefficient of correlation between stomach carcinoma surface of cancer and alcohol intake $r=0.49$, $p=0.03$. The following statements are true:

A. We reject alternative hypothesis and we conclude: there were no correlation between stomach carcinoma surface and alcohol intake

B. Quantity of intake alcohol significantly correlate with stomach carcinoma surface

C. We reject null hypothesis: there were correlation between stomach carcinoma surface and alcohol intake

D. Colton rules indicate a weak correlation between stomach carcinoma surface and alcohol intake

E. The correlation was direct

We compute Pearson coefficient of correlation between the BMI (body mass index) (kg/m^2) of the patients and the duration of long Covid-19 in days $r=0.14$, $p=0.03$. The following statements are true:

- A. We reject null hypothesis and accept alternative hypothesis; we conclude: there were significant statistically correlation between BMI and the duration of long Covid-19
- B. This association cannot be calculated because it make no sense
- C. We reject null hypothesis: there were no significant statistically correlation between BMI and the duration of long Covid-19
- D. Colton rules indicate no correlation between BMI and the duration of long Covid-19
- E. This association cannot be calculated because the assumptions are not met

*We compute Pearson coefficient of correlation between the height (m) of the patients and the duration of long Covid-19 in days $r=0.14$, $p=0.03$. The following statements are true:

- A. We reject null hypothesis and accept alternative hypothesis; we conclude: there were significant statistically correlation between BMI and the duration of long Covid-19
- B. This association cannot be calculated because it make no sense
- C. We reject null hypothesis: there were no significant statistically correlation between BMI and the duration of long Covid-19
- D. Colton rules indicate no correlation between BMI and the duration of long Covid-19
- E. This association cannot be calculated because the assumptions are not met

*We compute Pearson coefficient of correlation between the gender and the duration of long Covid-19 in days $r=0.14$, $p=0.03$. The following statements are true:

- A. We reject null hypothesis and accept alternative hypothesis; we conclude: there were significant statistically correlation between BMI and the duration of long Covid-19
- B. This association cannot be calculated because it make no sense
- C. We reject null hypothesis: there were no significant statistically correlation between BMI and the duration of long Covid-19
- D. Colton rules indicate no correlation between BMI and the duration of long Covid-19
- E. This association cannot be calculated because the assumptions are not met

RESEARCH SUMMARY

Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

F.P. Polack, et al. DOI: 10.1056/NEJMoa2034577

CLINICAL PROBLEM

Safe and effective vaccines to prevent severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and Covid-19 are urgently needed. No vaccines that protect against betacoronaviruses are currently available, and mRNA-based vaccines have not been widely tested.

CLINICAL TRIAL

A randomized, double-blind study of an mRNA vaccine encoding the SARS-CoV-2 spike protein.

43,548 participants ≥16 years old were assigned to receive the vaccine or placebo by intramuscular injection on day 0 and day 21. Participants were followed for safety and for the development of symptomatic Covid-19 for a median of 2 months.

RESULTS

Safety:

Vaccine recipients had local reactions (pain, erythema, swelling) and systemic reactions (e.g., fever, headache, myalgias) at higher rates than placebo recipients, with more reactions following the second dose. Most were mild to moderate and resolved rapidly.

Efficacy:

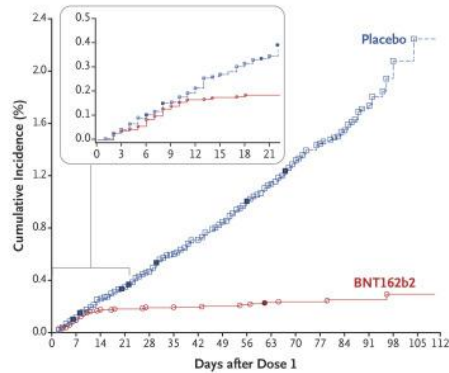
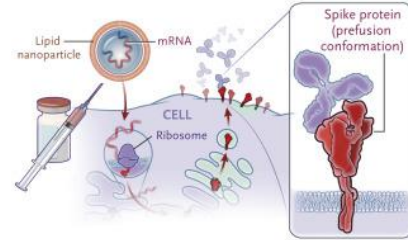
The vaccine showed some early protection 12 days after the first dose; 7 days after the second dose, 95% efficacy was observed.

LIMITATIONS AND REMAINING QUESTIONS

Further study is required to understand the following:

- Safety and efficacy beyond 2 months and in groups not included in this trial (e.g., children, pregnant women, and immunocompromised persons).
- Whether the vaccine protects against asymptomatic infection and transmission to unvaccinated persons.
- How to deal with those who miss the second vaccine dose.

Links: Full article | NEJM QuickTake | Editorial



	BNT162b2 Vaccine	Placebo
Symptomatic Covid-19	8	162
	N=18198	N=18325
Severe Covid-19	1	9
	N=21669	N=21686

Vaccine efficacy of 95% (95% credible interval, 90.3–97.6%)

CONCLUSIONS

Two doses of an mRNA-based vaccine were safe over a median of two months and provided 95% protection against symptomatic Covid-19 in persons 16 years of age or older.

Vaccinul Pfizer anti – Covid-19

- First published article
- Clinical trial on 43,548 participants
- Conducted before approval and marketing of the vaccine

RESEARCH SUMMARY

Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

F.P. Polack, et al. DOI: 10.1056/NEJMoa2034577

CLINICAL PROBLEM

Safe and effective vaccines to prevent severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and Covid-19 are urgently needed. No vaccines that protect against betacoronaviruses are currently available, and mRNA-based vaccines have not been widely tested.

CLINICAL TRIAL

A randomized, double-blind study of an mRNA vaccine encoding the SARS-CoV-2 spike protein.

43,548 participants ≥16 years old were assigned to receive the vaccine or placebo by intramuscular injection on day 0 and day 21. Participants were followed for safety and for the development of symptomatic Covid-19 for a median of 2 months.

RESULTS

Safety: Vaccine recipients had local reactions (pain, erythema, swelling) and systemic reactions (e.g., fever, headache, myalgias) at higher rates than placebo recipients, with more reactions following the second dose. Most were mild to moderate and resolved rapidly.

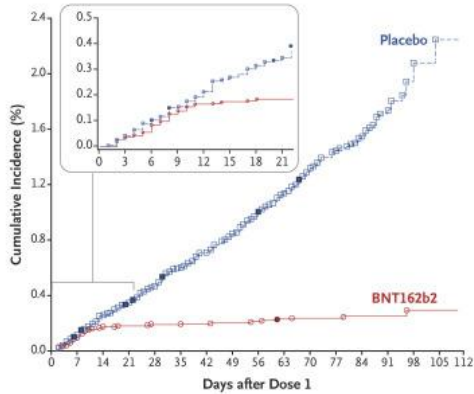
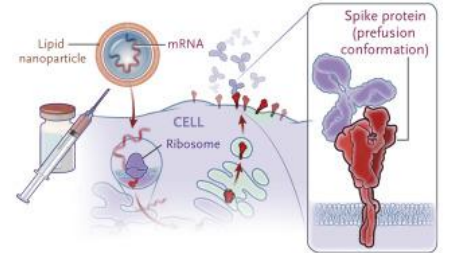
Efficacy: The vaccine showed some early protection 12 days after the first dose; 7 days after the second dose, 95% efficacy was observed.

LIMITATIONS AND REMAINING QUESTIONS

Further study is required to understand the following:

- Safety and efficacy beyond 2 months and in groups not included in this trial (e.g., children, pregnant women, and immunocompromised persons).
- Whether the vaccine protects against asymptomatic infection and transmission to unvaccinated persons.
- How to deal with those who miss the second vaccine dose.

Links: Full article | NEJM QuickTake | Editorial



	BNT162b2 Vaccine	Placebo
Symptomatic Covid-19	8	162
	N=18198	N=18325
Severe Covid-19	1	9
	N=21669	N=21686

Vaccine efficacy of 95% (95% credible interval, 90.3–97.6%)

CONCLUSIONS
Two doses of an mRNA-based vaccine were safe over a median of two months and provided 95% protection against symptomatic Covid-19 in persons 16 years of age or older.

Vaccine Pfizer anti – Covid-19

	BNT162b2 Vaccine	Placebo
Symptomatic Covid-19	8	162
	N=18198	N=18325
Severe Covid-19	1	9
	N=21669	N=21686

Vaccine efficacy of 95% (95% credible interval, 90.3–97.6%)

• What is the meaning?

Now you know!!!

	BNT162b2 Vaccine	Placebo
Symptomatic Covid-19	8	162
	N=18198	N=18325
Severe Covid-19	1	9
	N=21669	N=21686

Vaccine efficacy of 95% (95% credible interval, 90.3–97.6%)

- They have COVID-19:
 - 8 from 18. 198 on vaccinated sample = 0.0437%
 - 162 from 18. 325 on non-vaccinated sample = 0.88%
 - RR = 20.13

Efficacy

(162-8=154) from 162

- 152/162=95%
- 95% confidence interval 90.3 % – 97.6%

Vaccine efficacy at population level between 90.3% and 97.6 with a 5% error

If you still not understand the learning objectives please



ALWAYS



SEEK



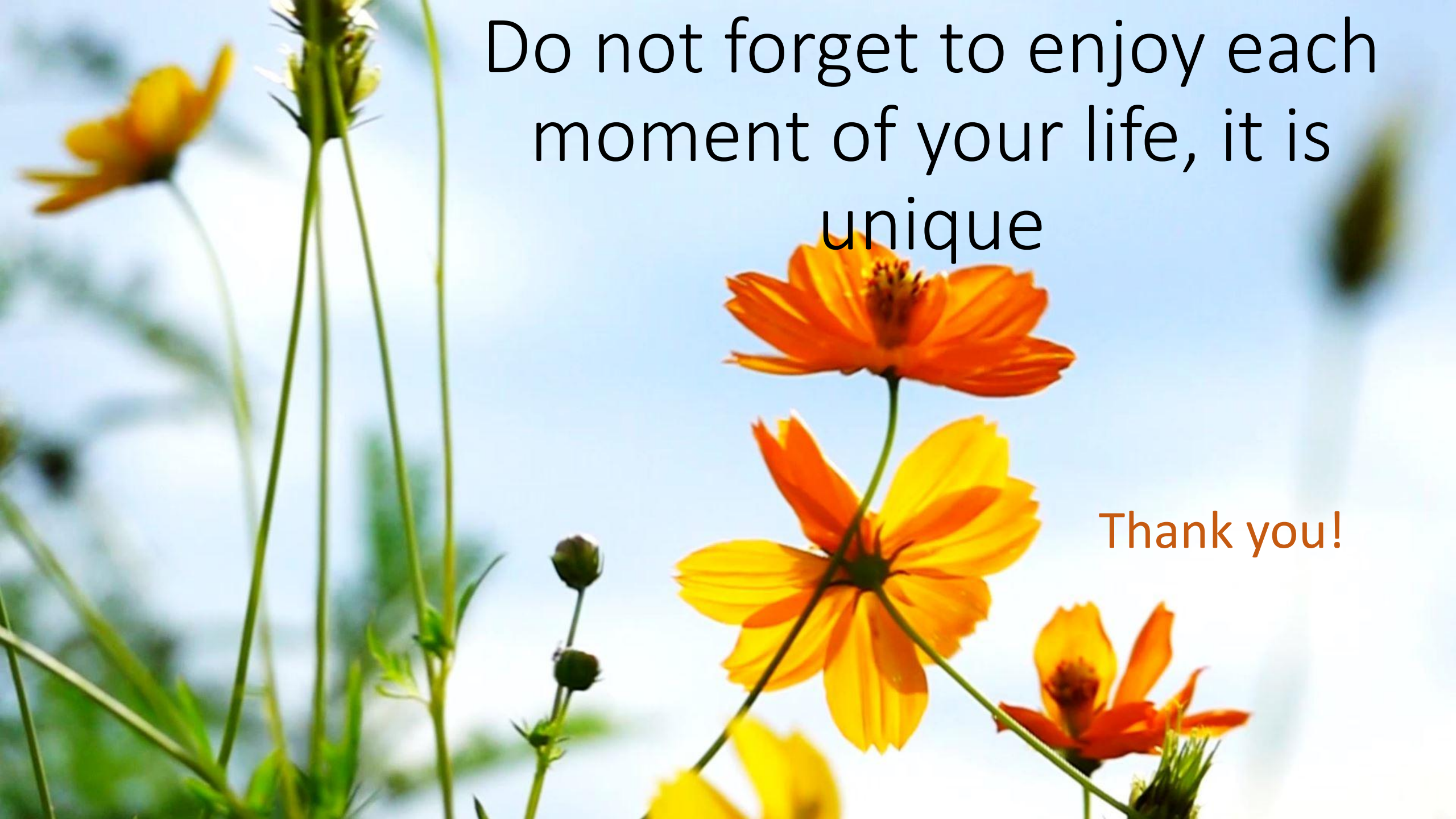
KNOWLEDGE

TEAMS - CHAT

Bondor Cosmina Ioana

- Contact:
 - email cbondor@umfcluj.ro





Do not forget to enjoy each
moment of your life, it is
unique

Thank you!