

Trainer: PhD. MsC. Bondor Cosmina-Ioana

Hypothesis Testing II

A ALWAYS

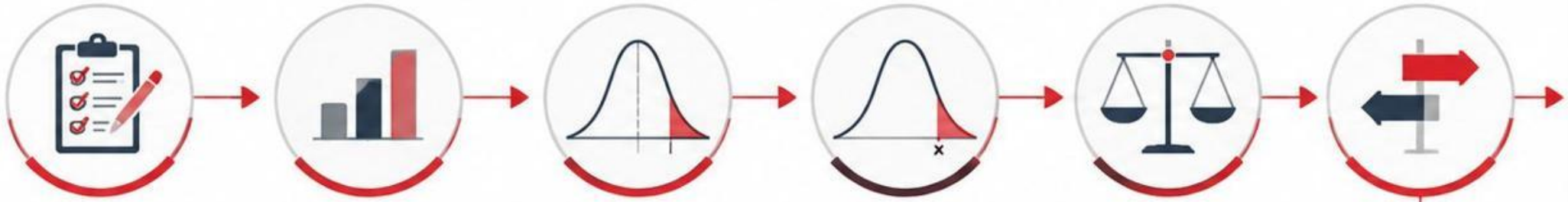
S SEEK

K KNOWLEDGE

Objectives

- Compare arithmetic means
    - Parametric test
- Finding the right statistical test

UNDERSTANDING HYPOTHESIS TESTING AND THE P-VALUE



1

STATE THE HYPOTHESES

Define the null hypothesis (H_0) and the alternative hypothesis (H_1).

2

CHOOSE THE SIGNIFICANCE LEVEL (α)

Set the threshold for decision making (commonly 0.05).

3

COLLECT DATA & CALCULATE THE TEST STATISTIC

Use your sample data to compute the test statistic.

4

CALCULATE THE P-VALUE

The p-value is the probability of observing data as extreme as (or more extreme than) what you observed, assuming H_0 is true.

5

COMPARE THE P-VALUE TO α

Determine whether the p-value is less than or equal to α.

6

IF P-VALUE ≤ α

REJECT THE NULL HYPOTHESIS (H_0)

There is sufficient evidence to support the alternative hypothesis (H_1).

IF P-VALUE > α

FAIL TO REJECT THE NULL HYPOTHESIS (H_0)

There is not sufficient evidence to support the alternative hypothesis (H_1).

KEY TAKEAWAY

The p-value helps you assess how strong the evidence is against the null hypothesis. A smaller p-value provides stronger evidence to reject the null hypothesis.

Example 2

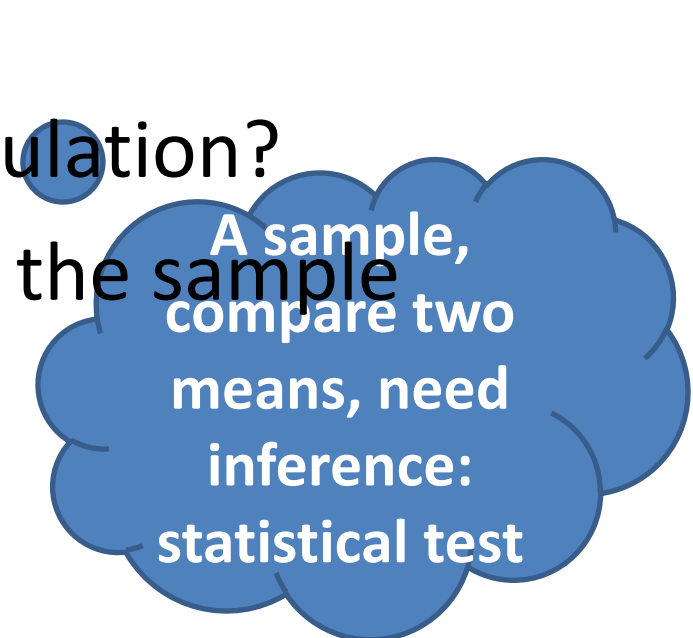
Hypothesis testing – Z test

Scenario: Average systolic blood pressure (SBP) in a representative sample of 81 people from company A is $\bar{X} = 125$ with a standard deviation of s=5.

In the general population average SBP μ=144 and standard deviation σ=6.

Is the sample a representative sample for the population?

There are significant differences between the SBP the sample population and the population?



- Average systolic blood pressure (SBP) in a representative sample of 81 people from company A is $\bar{X} = 125$ with a standard deviation of s=5. A healthcare researcher believes that due to decreased stress levels over the past year, the current average SBP is increased.
- In the general population average SBP μ=144 and standard deviation σ=6.

Is the sample a representative sample for the population?

i.e. Is the sample population the same as the population?

i.e. Is the difference between the sample mean and the historical population mean statistically significant at the 5% level, or could it be due to random sampling variation?

- **Population Mean (μ_o):** 144 mmHg
- **Known Population Standard Deviation (σ):** 6 mmHg (Meets the Z test condition for standard deviation)
- **Sample Size (n):** 81 (Meets the the Z test condition about large sample $n \geq 30$)
- **Sample Mean ($\bar{x}$):** 125 mmHg (The observed data).
- **Significance Level (α):** 0.05.

Step 1: Formulating Hypothesis

H0 - null hypothesis: there is no statistically significant difference between the sample average and population average (difference = 0, mean are equal)

H1 - alternative hypothesis: there is statistically significant difference between the sample average and population average (the differences are not due random variation)

Step 2 – Finding the right statistical test

- What we compare? (means, variances, percentages….?)
- How is the distribution? (normal or non-normal, binomial…?)
- How are the samples? (independent or dependent?)
- How is the sample size? (small or large?)
- How are the variances (equal or unequal?)
- Is the population standard deviation know? (Yes/No)
- How many samples (one, two, three or many?)

Step 2: Finding the right statistical test

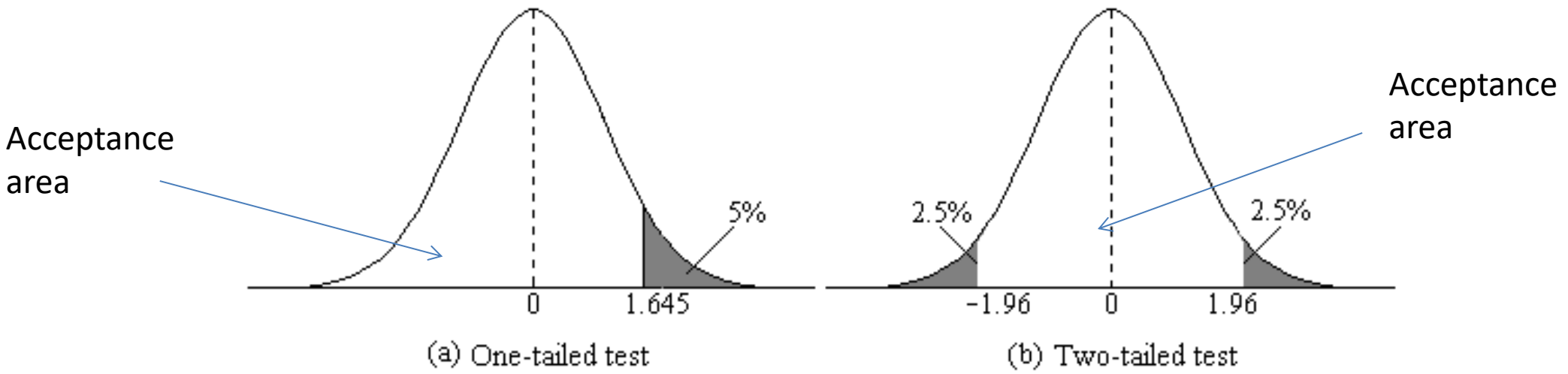
We choose a statistical test based on the normal distribution, we know the population standard deviation and, the sample is large.

Z test for one mean:

$$z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

Step 3. Determine the value the test parameter must attain to be declared significant

- Z_α =1.96 = **critical value** for a statistical two-tailed test
- Rejection area – critical interval – area where H_0 should be rejected



Step 4. Perform the calculations

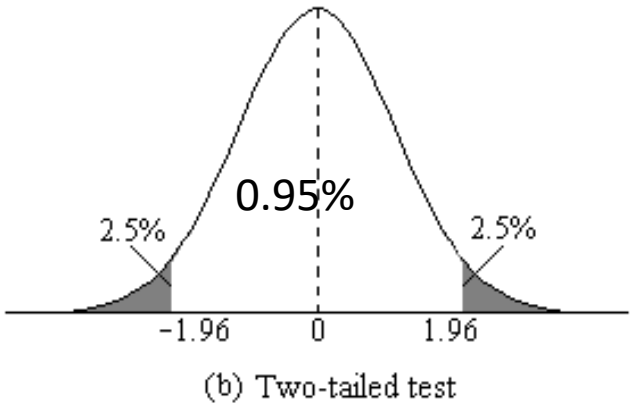
$$z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\frac{125-144}{6/\sqrt{81}}=-0,35$$

Step 5. Draw and state the conclusion

$(-Z_\alpha, Z_\alpha) = (-1.96, 1.96)$ ($\alpha=0.05$) – acceptance area (AUC = 95%)
The p is higher than $\alpha=0.05$ – fail to reject H_0

$(-\infty, -Z_\alpha] \cup [Z_\alpha, \infty)$ – rejection area (AUC = 5%)
The p is less than $\alpha=0.05$ – accept H_1

AUC – area under the curve

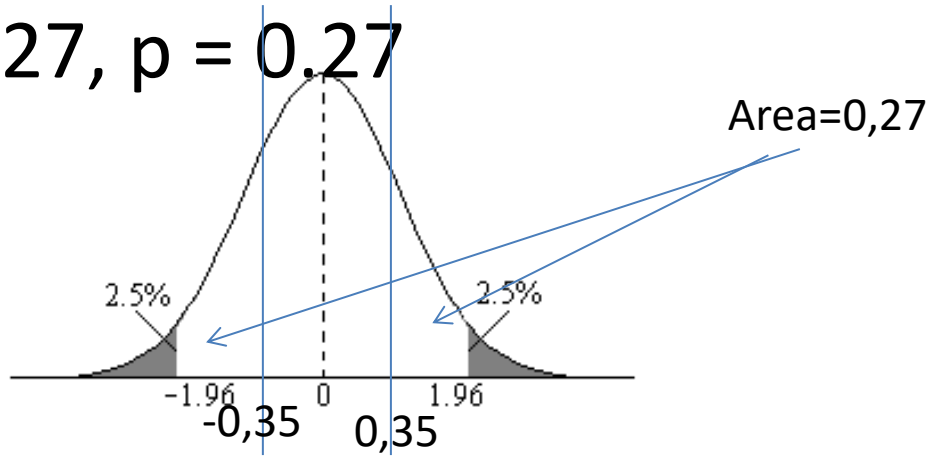


Step 5. Draw and state the conclusion

-1.96 <Z = -0.35 <1.96 we fail to reject null hypothesis

The average SBP from the sample did not differ significantly from the population average SBP i.e. The sample is a representative sample of the population.

Area under the curve for Z = -0.35 is 0.27, p = 0.27

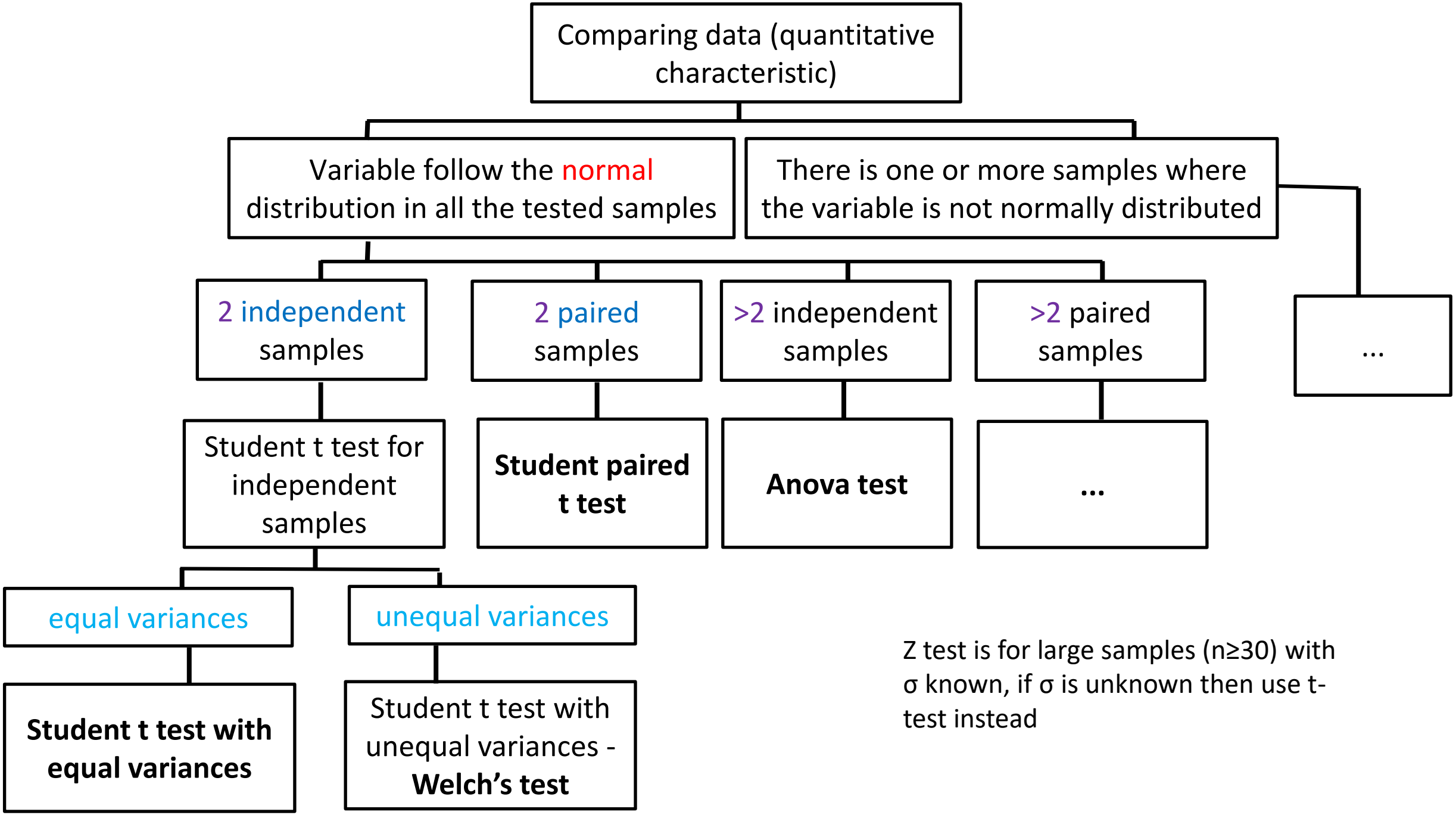


~~We accept the null hypothesis~~

We fail to reject the null hypothesis

Comparing arithmetic means with statistical tests

- Parametric test
  - the math of these tests relies on normal distribution properties
  - ex. Z-test, T-test, Anova test
  - they test means of quantitative data
- Non-parametric test



Z test is for large samples (n≥30) with σ known, if σ is unknown then use t-test instead

Test for normality

Checking the condition of normality of data: methods

Why do we check if the data follows the Normal distribution?

- to apply parametric tests, with normality condition:
  - Student-t test
  - ANOVA test

Verifying the existence of the Normal law

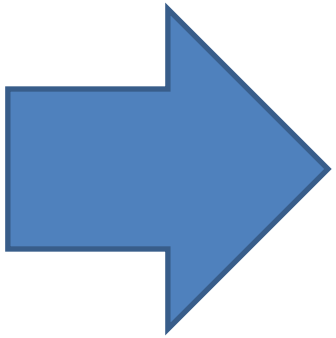
- graphs modalities
  - Histogram (symmetrical, bell shape)
  - Box and whisker plot (symmetrical around the median, no extreme cases)
  - Quantile graph

Verifying the existence of the Normal law

- descriptive statistics (not very reliable)
  - If the mean ≈ median ≈ mode
  - If the coefficient of kurtosis ≈ 0 / belongs to [-1, 1]
  - If the coefficient of symmetry ≈0 / belongs to [-1, 1]
- normality tests (are not recommended)
  - Kolmogorov-Smirnov test
  - Lilliefors test
  - Shapiro-Wilk test
    - if $p < 0.05 \rightarrow$ the distribution of the data is not Normal distribution,
    - if $p \geq 0.05 \rightarrow$ Normal distribution

Normal distribution?

Normal distribution?



- Shapiro-Wilk test (small and large samples),

or

- Lilliefors test (large samples),

or

- Kolmogorov-Smirnov test (small samples)

Example

Scenario	**An inflammation marker C reactive protein (CRP) was measured in 300 subjects with periodontitis. We want to know if CRP is normally distributed**
Statistical test	Shapiro-Wilk test
Null hypothesys H0	There is no statistically significant difference between the distribution of CRP and the normal distribution
Alternative hypothesis H1	There is a statistically significant difference between the distribution of CRP and the normal distribution
Test result	p=0.04
Decision	p=0.04<0.05, we reject the null hypothesis H0 and accept alternative hypothesis H1
Conclusion	There is **a statistically significant difference** between the distribution of CRP and normal distribution, i.e. CRP did not follow normal distribution

Example

Scenario	**A marker from blood used to evaluate kidney function serum creatinine was tested in 1000 subjects with diabetes mellitus. We want to know if serum creatinine is normally distributed**
Statistical test	Shapiro-Wilk test
Null hypothesys H0	There is no statistically significant difference between the distribution of serum creatinine and the normal distribution
Alternative hypothesis H1	There is a statistically significant difference between the distribution of serum creatinine and the normal distribution
Test result	p=0.15
Decision	p=0.15>0.05 **we fail to reject the null hypothesis H0**
Conclusion	There is **no statistically significant difference** between the distribution of serum creatinine and normal distribution, i.e. serum creatinine follow normal distribution

Test for variances

Checking the variances (square sample standard deviation): methods

Why do we check if the variances are equal or unequal?

- to choose between the parametric tests:
  - Student-t test for independent samples with equal variances.
  - Student-t test for independent samples with unequal variances – Welch's test
- to verify the condition for the parametric test
  - ANOVA test

Verifying the homogeneity of variances

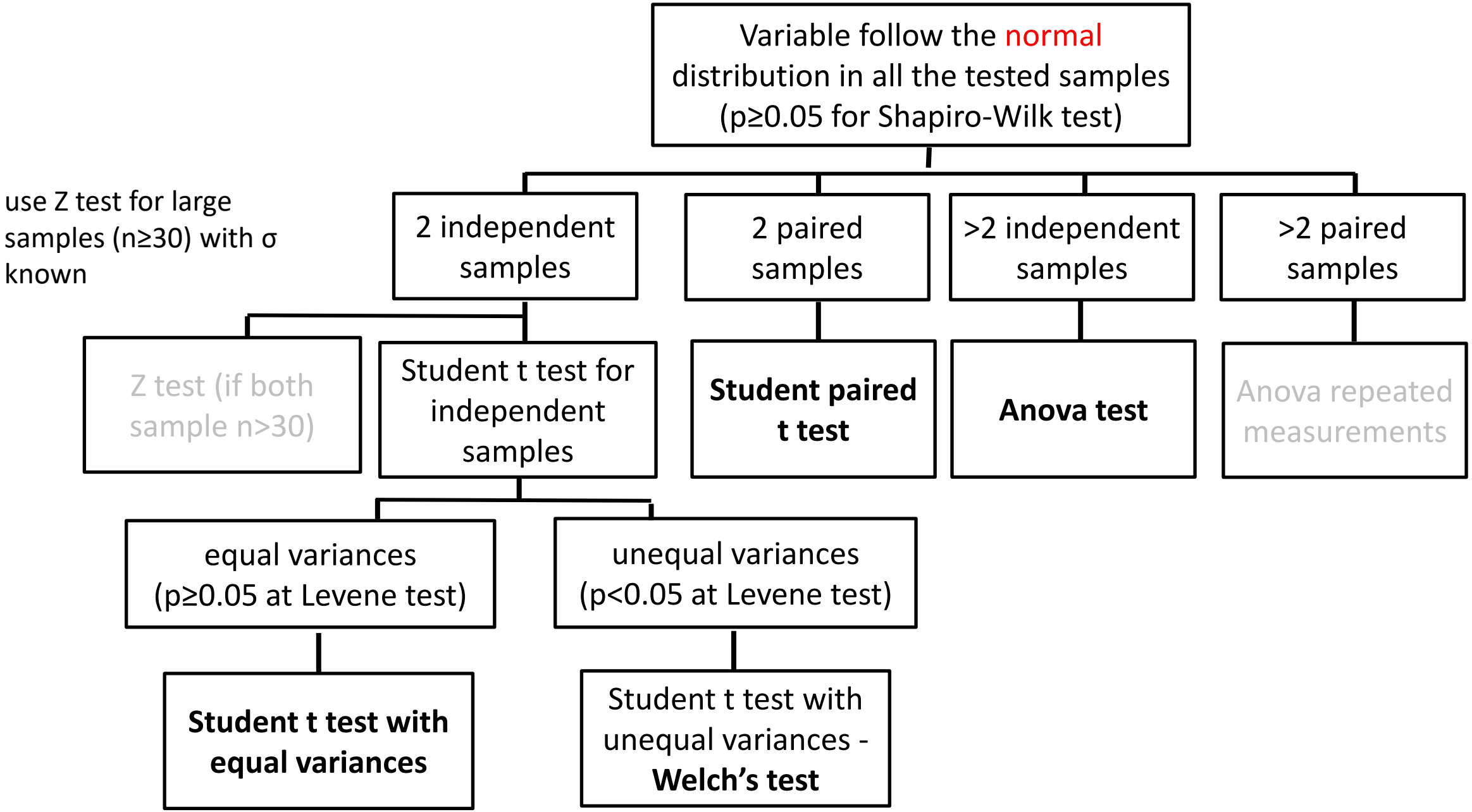
- with statistical test
  - Fisher test (available in Excel)
  - Levene test (available in Jamovi)
  - Bartlet test (available in EpiInfo)
    - if $p < 0.05 \rightarrow$ the variances are unequal
    - if $p \geq 0.05 \rightarrow$ the variances are equal

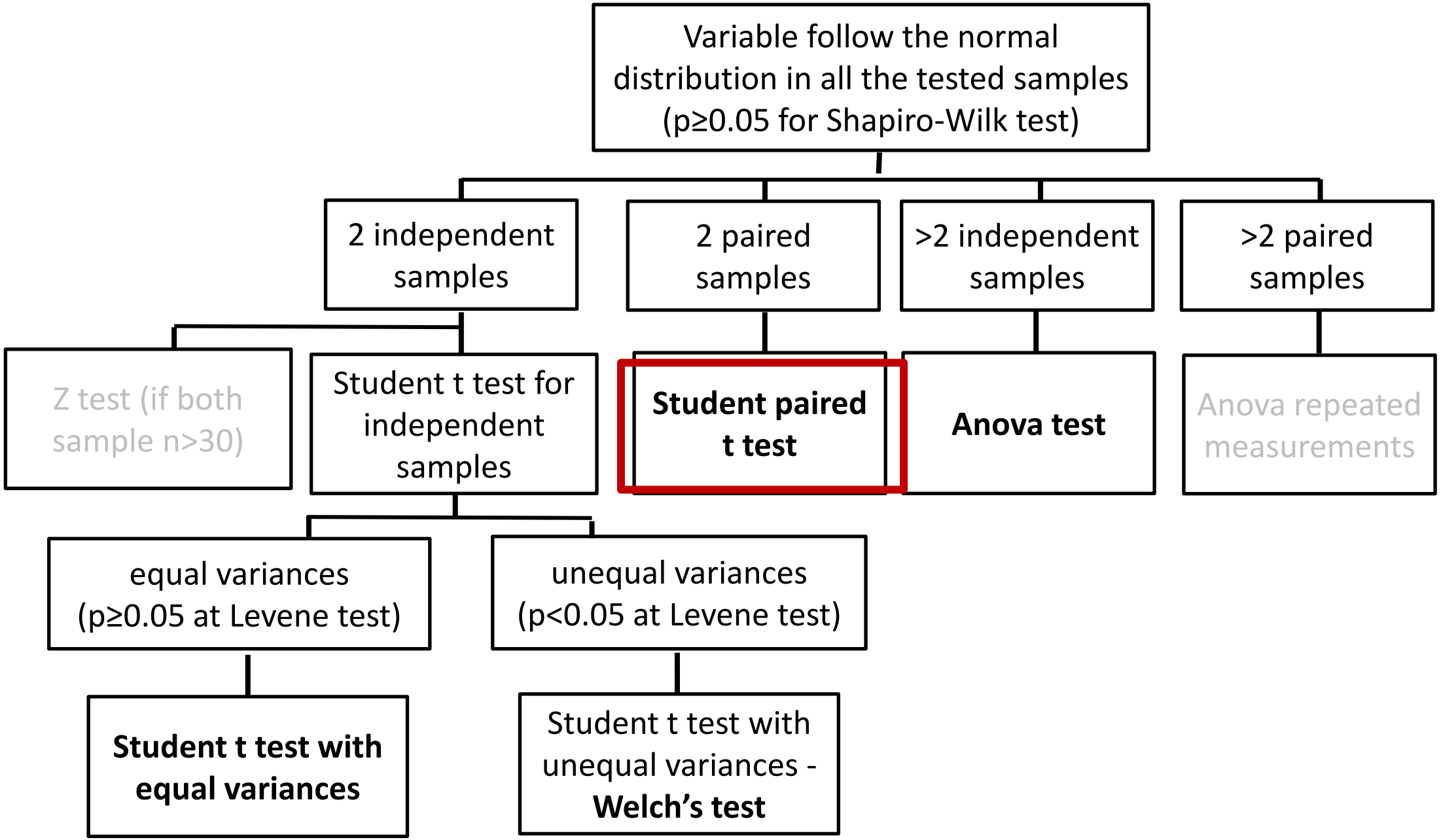
Example

Scenario	**The mean weight loss in an intervention sample (n=29) in 6 months was 2.95 kg +/- 1.59, control sample (n=29) was 0.63 kg +/- 1.48. Is the difference between the variances of weight loss in 6 months in the intervention group and in the control group stistically significant?**
Statistical test	Levene test
Null hypothesis H0	There is no statistically significant difference between the variances of weight loss in the intervention group and in the control group
Alternative hypothesis H1	There is a statistically significant difference between the variances of weight loss in the intervention group and in the control group
Test result	p=0.35
Decision	p=0.35>0.05, **we fail to reject the null hypothesis H0**
Conclusion	There is **no statistically significant difference** between the the variances of weight loss in the intervention group and in the control group, i.e. variances are equal

Test for means

with Jamovi

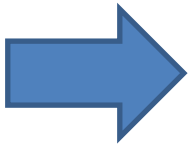

Variable follow the normal distribution in all the tested samples (p≥0.05 for Shapiro-Wilk test)
use Z test for large samples (n≥30) with σ known
2 independent samples
2 paired samples
>2 independent samples
>2 paired samples
Z test (if both sample n>30)
Student t test for independent samples
Student paired t test
Anova test
Anova repeated measurements
equal variances (p≥0.05 at Levene test)
unequal variances (p<0.05 at Levene test)
Student t test with equal variances
Student t test with unequal variances - Welch's test


Variable follow the normal distribution in all the tested samples (p≥0.05 for Shapiro-Wilk test)
2 independent samples
2 paired samples
>2 independent samples
>2 paired samples
Z test (if both sample n>30)
Student t test for independent samples
Student paired t test
Anova test
Anova repeated measurements
equal variances (p≥0.05 at Levene test)
unequal variances (p<0.05 at Levene test)
Student t test with equal variances
Student t test with unequal variances - Welch's test

Paired samples design

Twins	monozygotic twins
Patients follow in time	the evolution of various parameters followed for a period of time ex. changing of a characteristics as a consequence of age, smoking, hipertension or diabetes. Each patient is tested several times
Match samples	samples are match for some characteristics such as gender, age, social situation, level of education, etc.
Two observers	study model in which each patient is evaluated by two different medical doctors

- **Characteristics of the dependent samples (pairs):**
  - Have the same number of individuals.
  - Each value in the first group has a corresponding value in the second group.



choose a statistical test that compares paired samples

Example

Scenario	**To investigate that the use of metamizol sodic monohidrat contributes to the occurrence of agranulocytosis. A random sample of individuals follow a treatement with metamizol sodic monohidrat in high dose. Were measured the number of neutrophils at the patient's admission 3.6+/-1.85 and at the patient's discharge 2.9+/-1.85.**
Assumptions	
the observations are independent	Yes
2 dependent samples	Yes
the data are normally distributed	Yes
population variances are unkows	Yes
Name of the test	Student t test for dependent (paired) samples

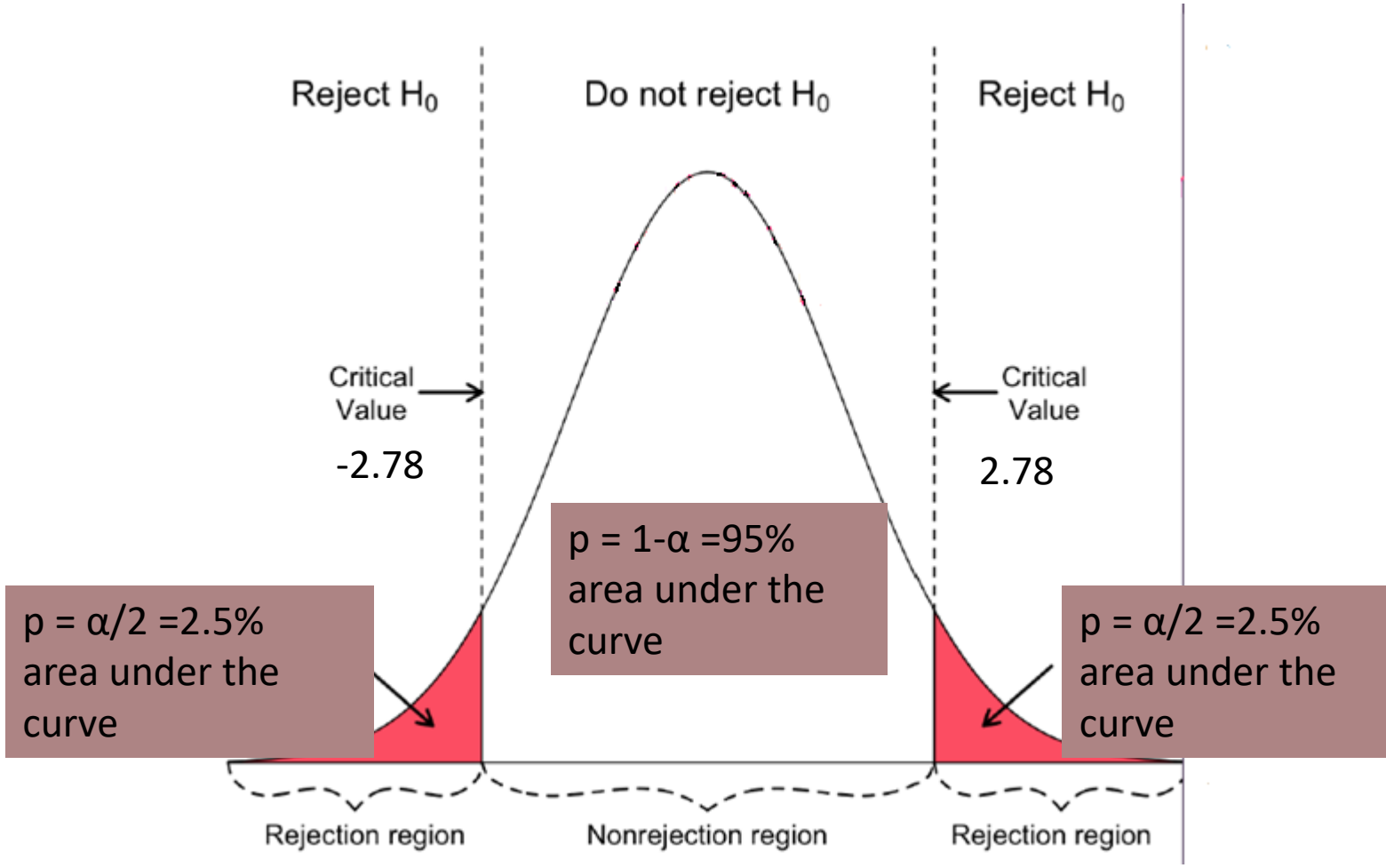
The same patient

Data

	Neutrofils (x1000/mm³)		
	Admission	Discharge	d=Diference (Admission-Discharge)
Patient 1	3	0.5	2.5
Patient 2	1.5	2	-0.5
Patient 3	5	3.5	1.5
Patient 4	2.5	3	-0.5
Patient 5	6	5.5	0.5
Arithmetic mean	**3.60**	**2.90**	**0.70**
Standard deviation	**1.85**	**1.85**	**1.30**

2.

- **Null hypothesis H_0:** The average of the differences between baseline and final neutrofiles values does not differ significantly from 0.
- **Alternative hypothesis H_1** (two tail test): The average of the differences between baseline and final neutrofiles values differs significantly from 0.



5. Test parameter

$$t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}} = \frac{0.7}{\frac{1.3}{\sqrt{5}}} = 1.2$$

p=0.30

- n = sample size
- $\bar{d}$ = mean of the differences;
- s = standard deviation of the differences

	Admission	Discharge
Arithmetic mean	3.60	2.90
Standard deviation	3.43	3.43
Number of observation	5	5
Degree of freedom	4	
t – test parameter	1.20	
	0.15	
	2.13	
P (T≤t) two tail	0.30	
t critic two tail	2.78	

6. Decision making

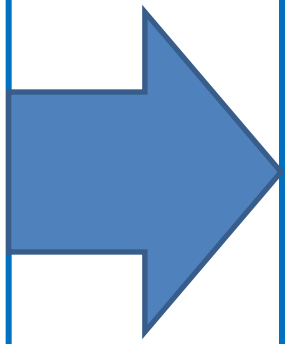
- Because t=1.2 ∈ (-2.78,2.78) **we fail to reject the null hypothesis**: the average of the differences between baseline and final neutrofiles values does **not differ** statistically significantly from 0.
- Because p=0.30>0.05 **we fail to reject the null hypothesis**: the average of the differences between baseline and final neutrofiles values does **not differ** statistically significantly from 0.

Independent samples

If variances are unequal, how are the means?

Assumptions

- The observations are independent
- Two independent samples
- Two means to compare
- **Unequal** variances
- Normal distributions (n_1 or n_2<30)



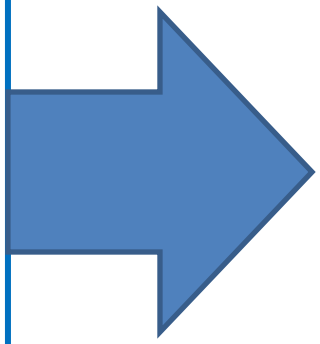
T test (Student) for unequal variances (Welch test)

parametric test

If variances are equal, how are the means?

Assumptions

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



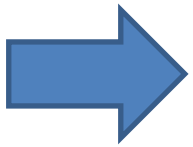
T test (Student) for equal variances

parametric test

Independent samples design

Samples	different subjects in different groups
Patients follow in time	we compare the evolution of various parameters followed for a period of time, but we compare between different groups ex. changing of a characteristics as a consequence of age, smoking, hipertension or diabetes. Each patient is tested two times. We make the difference between initial and final values. We compare the difference between urban area with rural area of residence (different subjects).
Match samples	samples are not match

- **Characteristics of the independent samples:**
  - Can have different or equal number of individuals.
  - There is no correspondence between the individuals



choose a statistical test that compares independent samples

Example

Scenario	**We want to see if diet B is different than diet C in lowering the cholesterol. 100 subjects follow diet B, 100 subjects follow diet C The final cholesterol for diet B group 290±100mg/dl, for diet C 250±50 mg/dl group (average±standard deviation)**
Assumptions	
the observations are independent	Yes
2 independent samples	Yes
the data are normally distributed	Yes
population variances are unkows	Yes
variances are equal	No
Name of the test	Student t test for independent samples with unequal variances (Welch test)

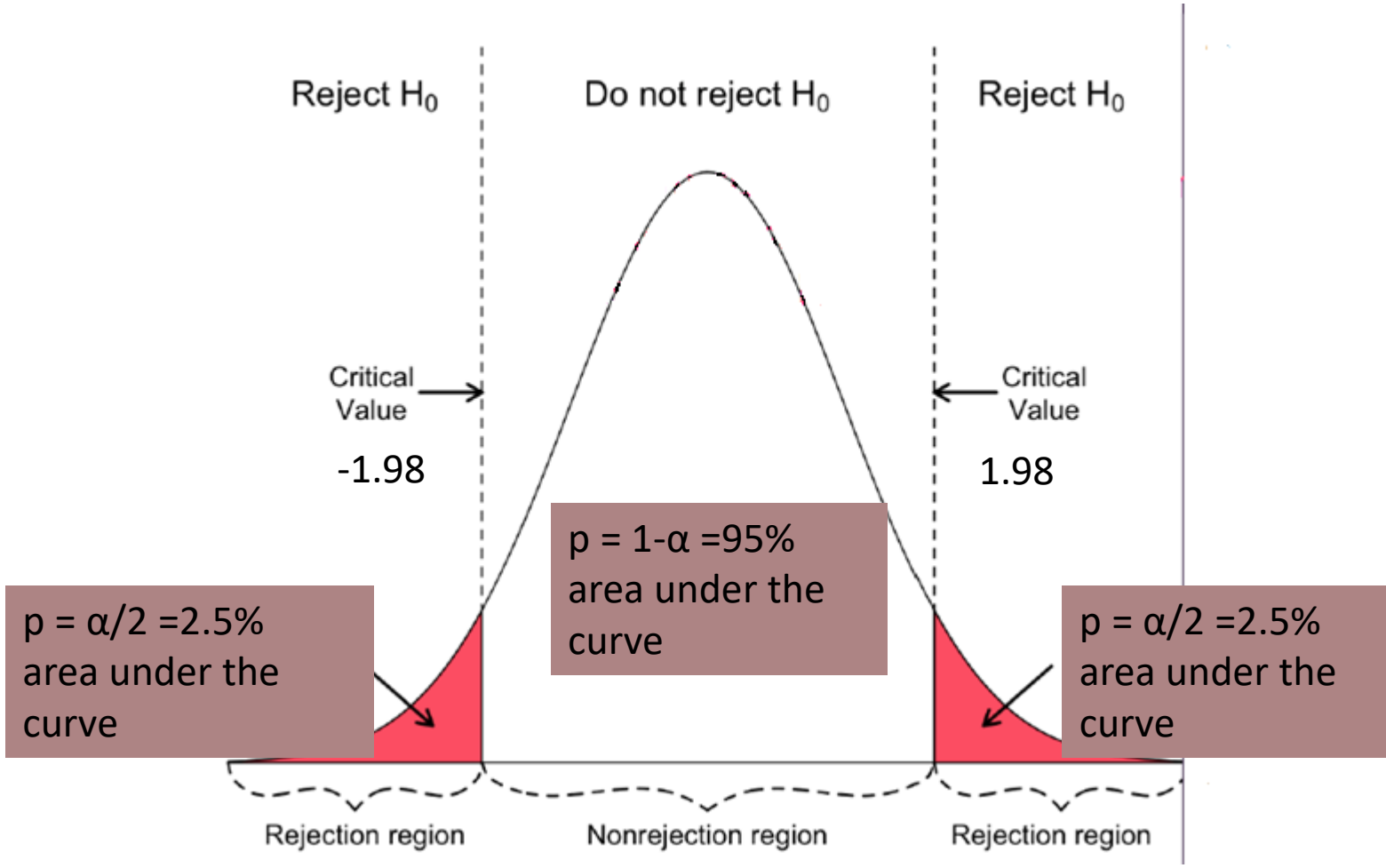
Different patients

Data

	Cholesterol (mg/dl)	
	Final cholesterol for group with diet B	Final cholesterol for group with diet C
Patient 1	300	50
Patient 2	150	200
Patient ...	...	...
Patient 99	250	300
Patient 100	60	55
Arithmetic mean	**290**	**250**
Standard deviation	**100**	**50**

2.

- **H0 - null hypothesis:** there is no statistically significant difference between the averages of cholesterol for subjects with diet C versus diet B
- **Alternative hypothesis H_1** (two tail test): there is statistically significant difference between the averages of cholesterol for subjects with diet C versus diet B



5. Test parameter

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{290 - 250}{\sqrt{\frac{100^2}{100} + \frac{50^2}{100}}} = 3.58$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{10000}{100} + \frac{2500}{100}\right)^2}{\frac{\left(\frac{10000}{100}\right)^2}{99} + \frac{\left(\frac{2500}{100}\right)^2}{99}} = 146$$

p=0.0005

- n = sample size
- $\bar{X}$ = mean
- s = standard deviations

	Final cholesterol for group with diet B	Final cholesterol for group with diet C
Arithmetic mean	290	250
Standard deviations	100	50
Number of observation	100	100
Degree of freedom	146	
t – test parameter	3.58	
	0.15	
P (T≤t) two tail	0.0005	
t critic two tail	1.98	

6. Decision making

- Because t=3.58 $\notin$ (-∞,−1.97)∪(1.97,∞) **we reject the null hypothesis, accept the alternative hypothesis**: there is statistically significant difference between the averages of cholesterol for subjects with diet B versus diet C.
- Because p=0.0004<0.05 **we reject the null hypothesis, accept the alternative hypothesis**: there is statistically significant difference between the averages of cholesterol for subjects with diet B versus diet C.

Example

Scenario	**We compare the recovery times of patients undergoing two different surgical procedures for the same condition. 25 subjects with standard procedure, 30 subjects with new procedure The recovery time for standard procedure group: 15±2days, for new procedure group: 12±3days (average±standard deviation)**
Assumptions	
the observations are independent	Yes (we have 25 different urelated subject in the first group and 30 different urelated subjects in the second group)
2 independent samples	Yes
the data are normally distributed	Yes
population variances are unkows	Yes
variances are equal	Yes
Name of the test	Student t test for independent samples with equal variances

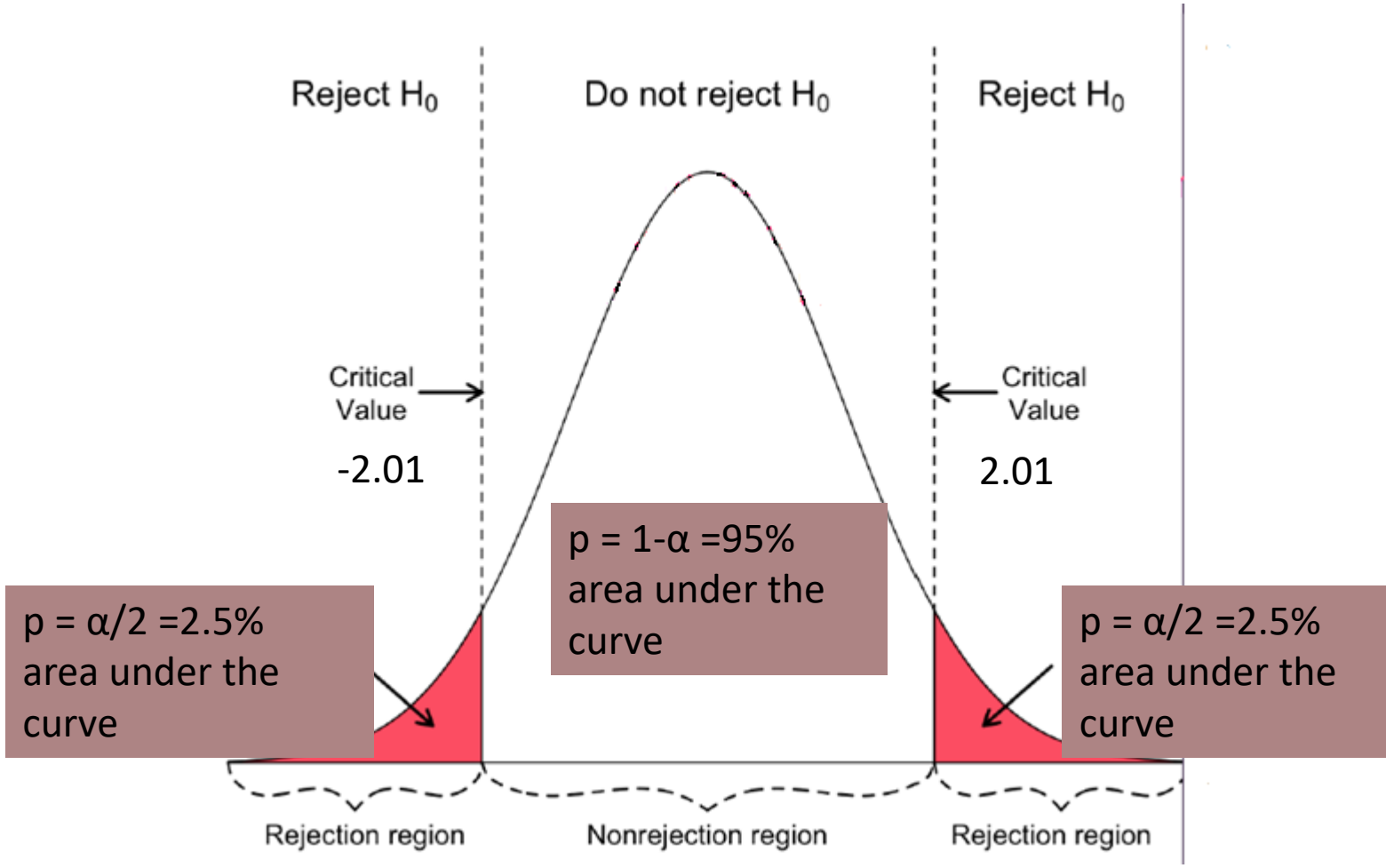
Different patients

Data

	Recovery time (days)	
	Standard procedure	New procedure
Patient 1	10	9
Patient 2	15	12
Patient ...	...	...
Patient 25	11	10
Patient ...		...
Patient 30		14
Arithmetic mean	**15**	**12**
Standard deviation	**2**	**3**

2.

- **H0 - null hypothesis:** there is no statistically significant difference between the averages of recovery time for subjects with new procedure versus standard procedure
- **Alternative hypothesis H_1** (two tail test): there is statistically significant difference between the averages of recovery time for subjects with new procedure versus standard procedure



5. Test parameter

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{15 - 12}{\sqrt{6.73^2\left(\frac{1}{25} + \frac{1}{30}\right)}} = 1.64$$

where $s_p = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{24\cdot 4+29\cdot 9}{53}$=6.73

p=0.11

- n = sample size
- $\bar{X}$ = mean
- s = standard deviations

	Standard procedure	New procedure
Arithmetic mean	15	12
Standard deviations	2	3
Number of observation	25	30
Degree of freedom	53	
t – test parameter	1.64	
	0.15	
P (T≤t) two tail	0.11	
t critic two tail	2.01	

6. Decision making

- Because t=1.64 ∈ (-∞,−2.01)∪(2.01,∞) **we fail to reject the null hypothesis**: there is no statistically significant difference between the averages of recovery time for subjects with new procedure versus standard procedure
- Because p=0.11>0.05 **we fail to reject the null hypothesis**: there is no statistically significant difference between the averages of recovery time for subjects with new procedure versus standard procedure

*Suppose that the result of a statistical test is p-value = 0.001. If the significance level for the test is α =0.05, which of the following is the appropriate decision?

A. Reject alternative hypothesis H_a

B. Reject null hypothesis H_0, accept null hypothesis H_0

C. Reject null hypothesis H_0, accept alternative hypothesis H_a

D. Fail to reject null hypothesis H_0, accept alternative hypothesis H_a

E. Fail to reject alternative hypothesis H_a

*Suppose that the result of a statistical test is p-value = 0.001. If the significance level for the test is $\alpha = 0.05$, which of the following is the appropriate decision?

A. Reject alternative hypothesis H_a

B. Reject null hypothesis H_0, accept null hypothesis H_0

C. Reject null hypothesis H_0, accept alternative hypothesis H_a

D. Fail to reject null hypothesis H_0, accept alternative hypothesis H_a

E. Fail to reject alternative hypothesis H_a

- Forty-three adults with hypertension were enrolled in a 12-week swimming program and the effect of swimming on systolic blood pressure (SBP) at baseline and at the end of the 12 weeks was monitored. Select all correct statements.
- A. If SBP at baseline has p = 0.026 (Shapiro-Wilk), it is correct to apply the Student's t-test for independent groups.
- B. A p = 0.026 of the Shapiro-Wilk test for baseline SBP indicates that the data are normally distributed.
- C. A p = 0.026 of the Shapiro-Wilk test for baseline SBP indicates that the SBP distribution is statistically significantly different from the theoretical normal distribution for $\alpha = 0.05$.
- D. A p-value of 0.126 and 0.148 for the initial and final SBP test indicates a normal distribution of SBP values at the two determinations.
- E. A p-value of 0.026 and 0.048 for the initial and final SBP test indicates a normal distribution of SBP values at the two determinations

- Forty-three adults with hypertension were enrolled in a 12-week swimming program and the effect of swimming on systolic blood pressure (SBP) at baseline and at the end of the 12 weeks was monitored. Select all correct statements.
- A. If SBP at baseline has p = 0.026 (Shapiro-Wilk), it is correct to apply the Student's t-test for independent groups.
- B. A p = 0.026 of the Shapiro-Wilk test for baseline SBP indicates that the data are normally distributed.
- C. A p = 0.026 of the Shapiro-Wilk test for baseline SBP indicates that the SBP distribution is statistically significantly different from the theoretical normal distribution for $\alpha = 0.05$.
- D. A p-value of 0.126 and 0.148 for the initial and final SBP test indicates a normal distribution of SBP values at the two determinations.
- E. A p-value of 0.026 and 0.048 for the initial and final SBP test indicates a normal distribution of SBP values at the two determinations

- We have reported the following results in patients with acute coronary syndrome in the literature: Women (n=248), Men (n=541), BMI (30.53 ± 7.91, 29.95 ± 5.64 p=0.2991), where BMI = body mass index, Levene test $p<0.001$, Shapiro-Wilk test $p>0.05$. Select all correct answers for $\alpha = 0.05$.
- A. Since p = 0.2291 we have sufficient evidence to reject H0.
- B. Student's t-test was applied for independent groups and equal variances.
- C. H1: μBMI-F = μBMI-B
- D. Student's t-test was applied for independent groups and unequal variances.
- E. Since p = 0.2291 we do not have sufficient evidence to reject H0.

- We have reported the following results in patients with acute coronary syndrome in the literature: Women (n=248), Men (n=541), BMI (30.53 ± 7.91, 29.95 ± 5.64 p=0.2991), where BMI = body mass index, Levene test $p<0.001$, Shapiro-Wilk test $p>0.05$. Select all correct answers for $\alpha = 0.05$.
- A. Since p = 0.2291 we have sufficient evidence to reject H0.
- B. Student's t-test was applied for independent groups and equal variances.
- C. H1: μBMI-F = μBMI-B
- D. Student's t-test was applied for independent groups and unequal variances.
- E. Since p = 0.2291 we do not have sufficient evidence to reject H0.

- Thank you!