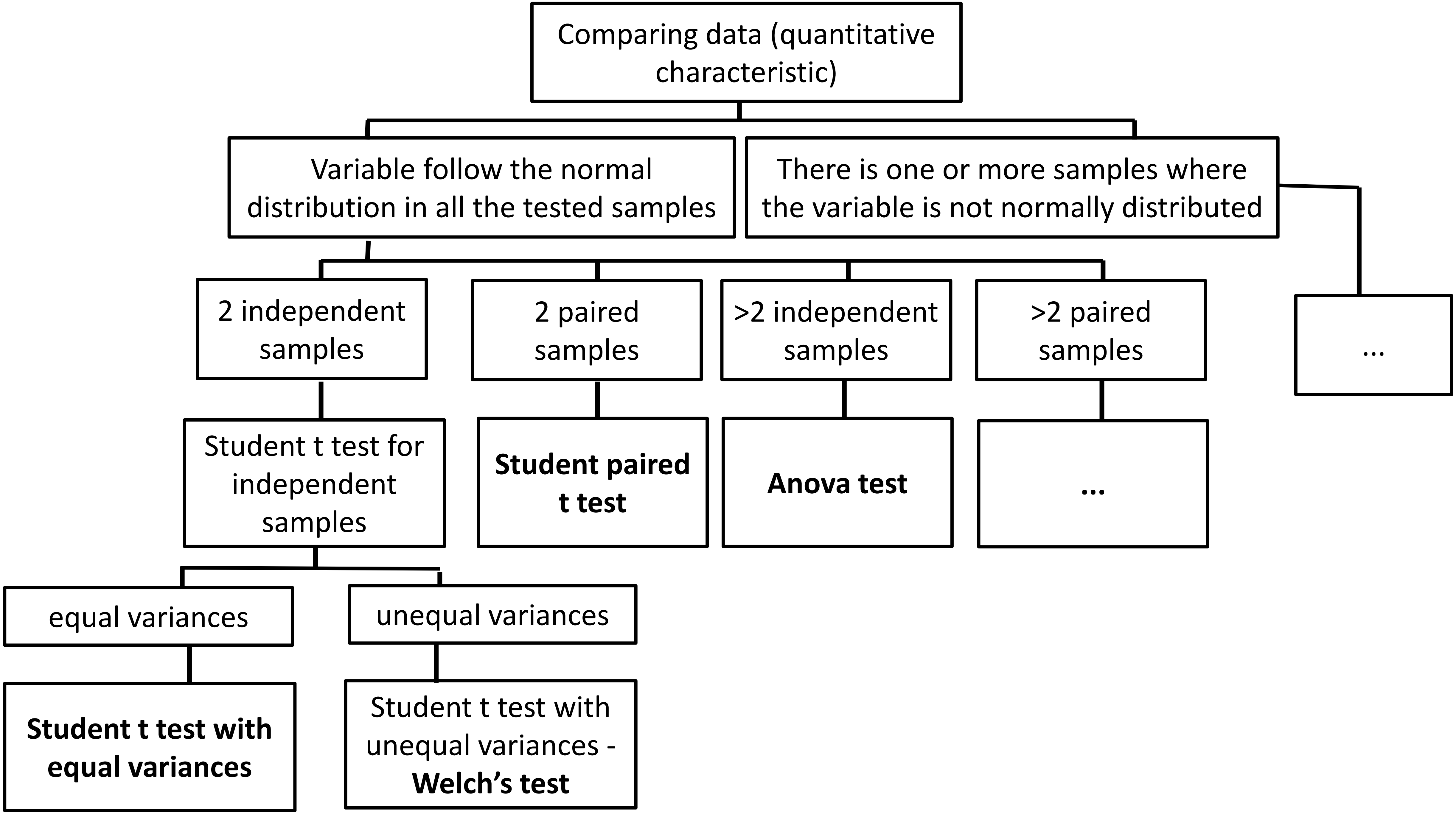
Trainer: PhD. MsC. Bondor Cosmina-Ioana

Hypothesis Testing II

A ALWAYS
S SEEK
K KNOWLEDGE

Objectives

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


Comparing data (quantitative characteristic)
Variable follow the normal distribution in all the tested samples
There is one or more samples where the variable is not normally distributed
2 independent samples
2 paired samples
>2 independent samples
>2 paired samples
...
Student t test for independent samples
Student paired t test
Anova test
...
equal variances
unequal variances
Student t test with equal variances
Student t test with unequal variances - Welch's test

We can answer medical questions

- We perform data analysis

1. We formulate the null/alternative hypothesis
2. We verify that the test assumptions are met, we choose the statistical test
3. We establish the confidence level
4. We establish the critical region for rejecting / accepting the null hypothesis
5. We calculate the test parameter
6. We take the statistical decision

Objective: Alternative hypothesis (H1) - There are statistically significant differences in the populations between two arithmetic means

Method:

absurd assumption

suppose that null hypothesis H_0 is true

compute the probability p of obtaining a diference equal with the observed difference (on the samples) or a more extreme one

choose the level of error α, usually, 0.05

if $p < \alpha$, the observed data are unlikely under the H0 and suggests rejecting the H0, accepting the H1

if $p \geq \alpha$, the data are consistent with the H0, we cannot reject H0

Test for normality

Checking the condition of normality of data: methods

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

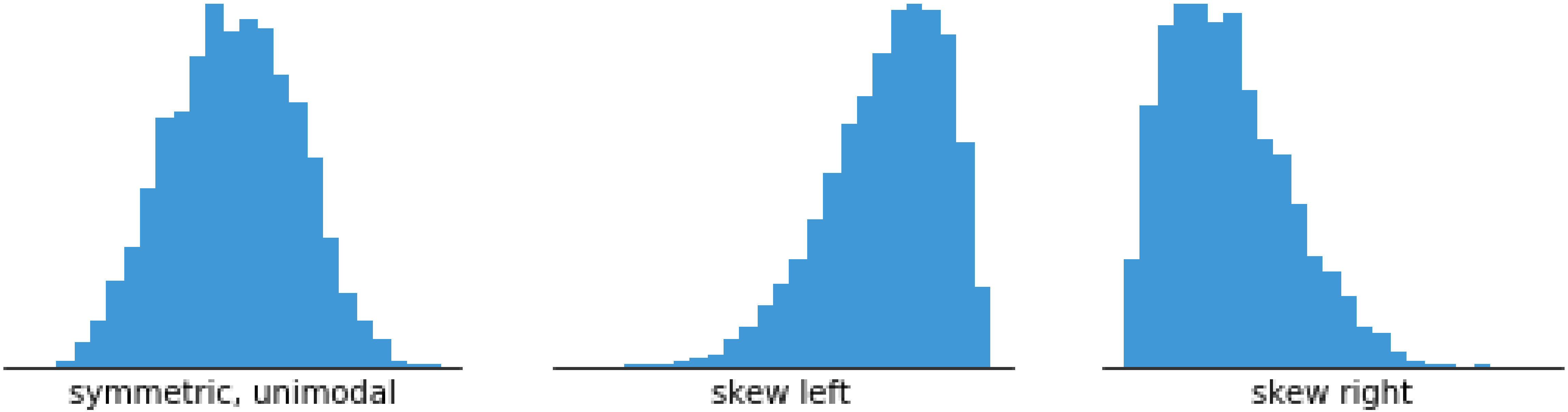
- to apply parametric tests, with normality condition:
  - Student-t test
  - ANOVA test

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

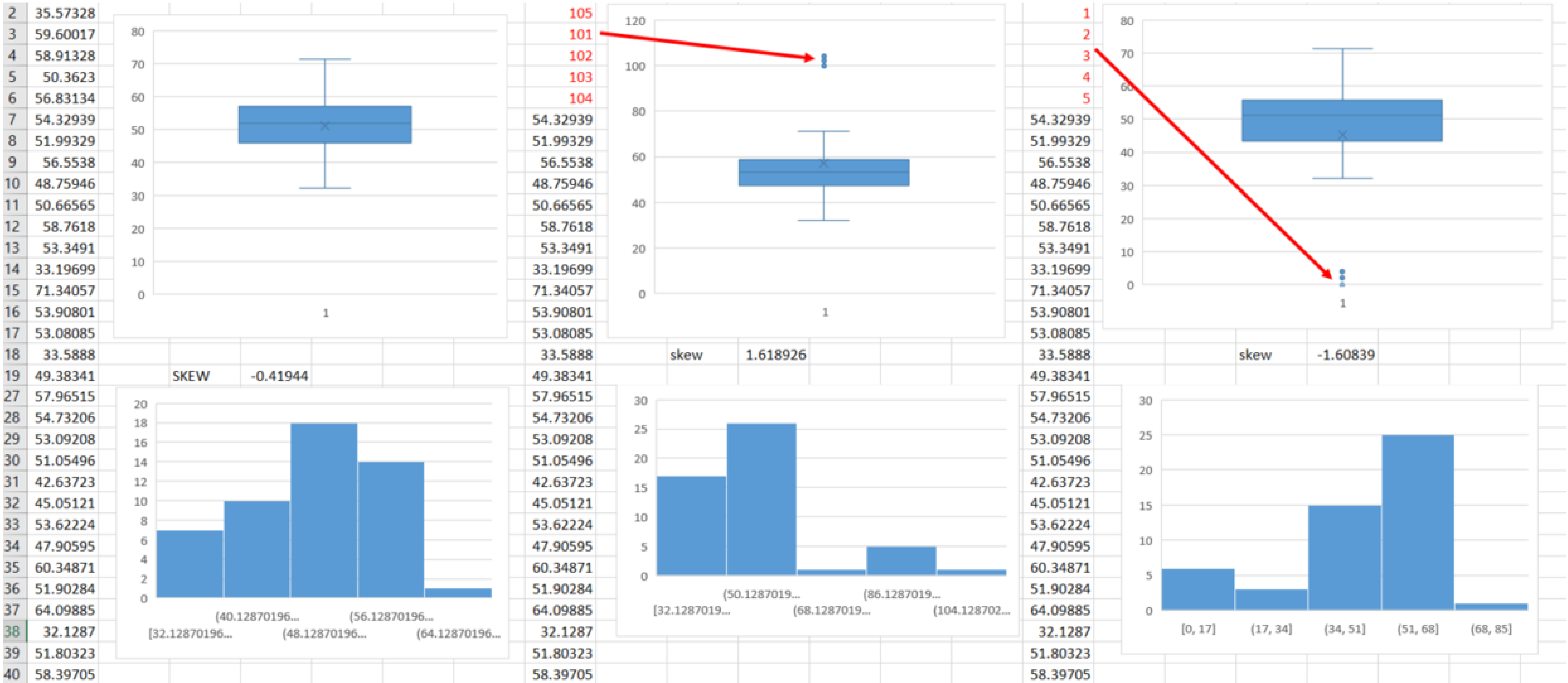
- descriptive statistics (not very reliable)
  - If the mean ≈ median ≈ mode
  - If the coefficient of kurtosis ≈ 0 / belongs to [-1, 1]
  - If the coefficient of symmetry ≈0 / belongs to [-1, 1]
- normality tests (are not recommended)
  - Kolmogorov-Smirnov test
  - Lilliefors test
  - Shapiro-Wilk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



Mike Yi. Data tutorial. https://chartio.com/learn/charts/histogram-complete-guide/

~ normal

not normal



2	35.57328	105	1
3	59.60017	101	2
4	58.91328	102	3
5	50.3623	103	4
6	56.83134	104	5
7	54.32939	54.32939	54.32939
8	51.99329	51.99329	51.99329
9	56.5538	56.5538	56.5538
10	48.75946	48.75946	48.75946
11	50.66565	50.66565	50.66565
12	58.7618	58.7618	58.7618
13	53.3491	53.3491	53.3491
14	33.19699	33.19699	33.19699
15	71.34057	71.34057	71.34057
16	53.90801	53.90801	53.90801
17	53.08085	53.08085	53.08085
18	33.5888	33.5888	33.5888
19	49.38341	49.38341	49.38341
27	57.96515	57.96515	57.96515
28	54.73206	54.73206	54.73206
29	53.09208	53.09208	53.09208
30	51.05496	51.05496	51.05496
31	42.63723	42.63723	42.63723
32	45.05121	45.05121	45.05121
33	53.62224	53.62224	53.62224
34	47.90595	47.90595	47.90595
35	60.34871	60.34871	60.34871
36	51.90284	51.90284	51.90284
37	64.09885	64.09885	64.09885
38	32.1287	32.1287	32.1287
39	51.80323	51.80323	51.80323
40	58.39705	58.39705	58.39705

~ normal

not normal

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

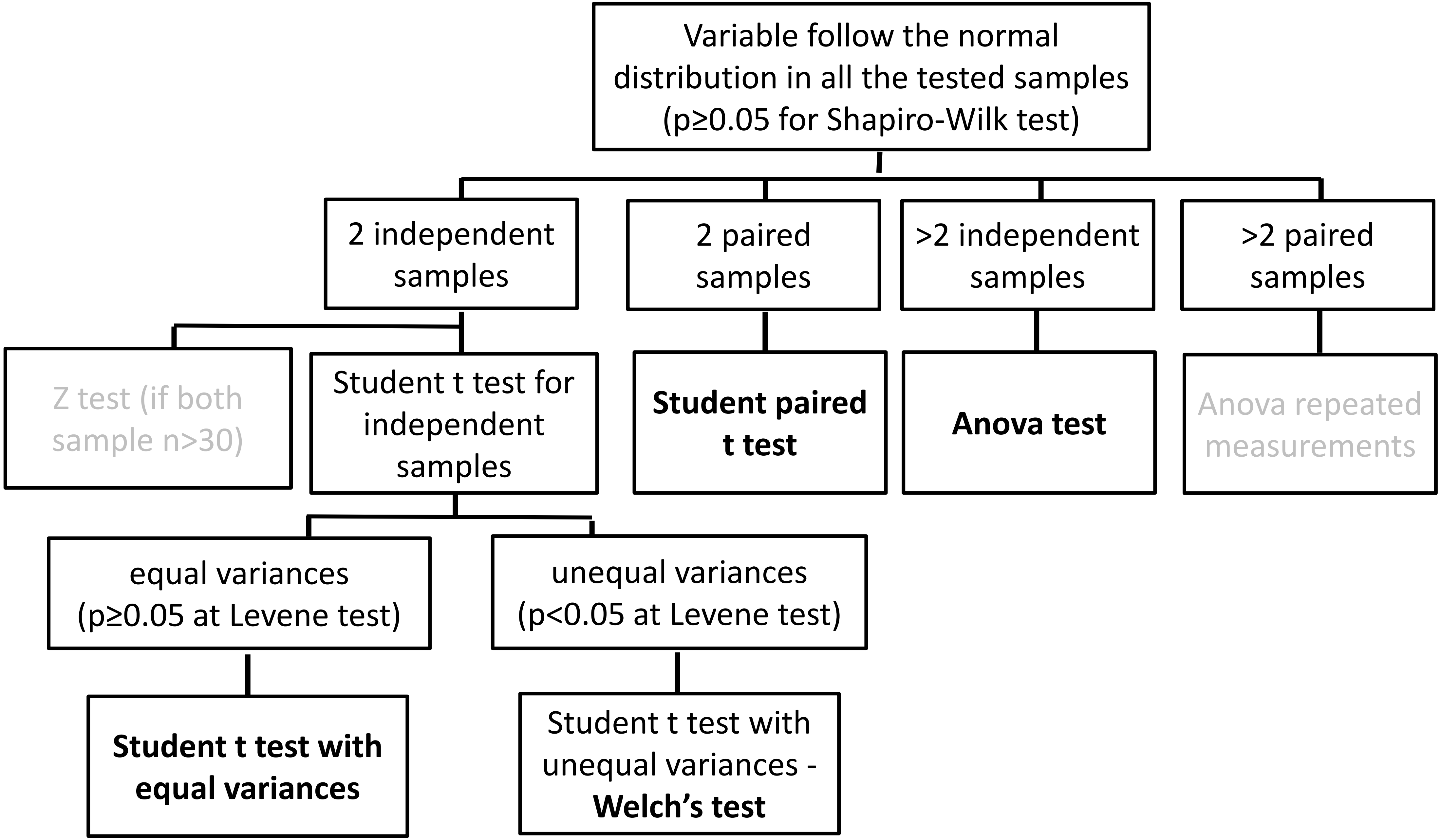
Verifying the homogeneity of variances

- statistical test
  - Fisher test (available in Excel)
  - Levene test (available in Jamovi)
  - Bartlet test (available in EpiInfo)
    - if $p < 0.05$ → the variances are unequal
    - if $p \geq 0.05$ → 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. There are differences between the variances of weight loss in 6 months in the intervention group and in the control group?**
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



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

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

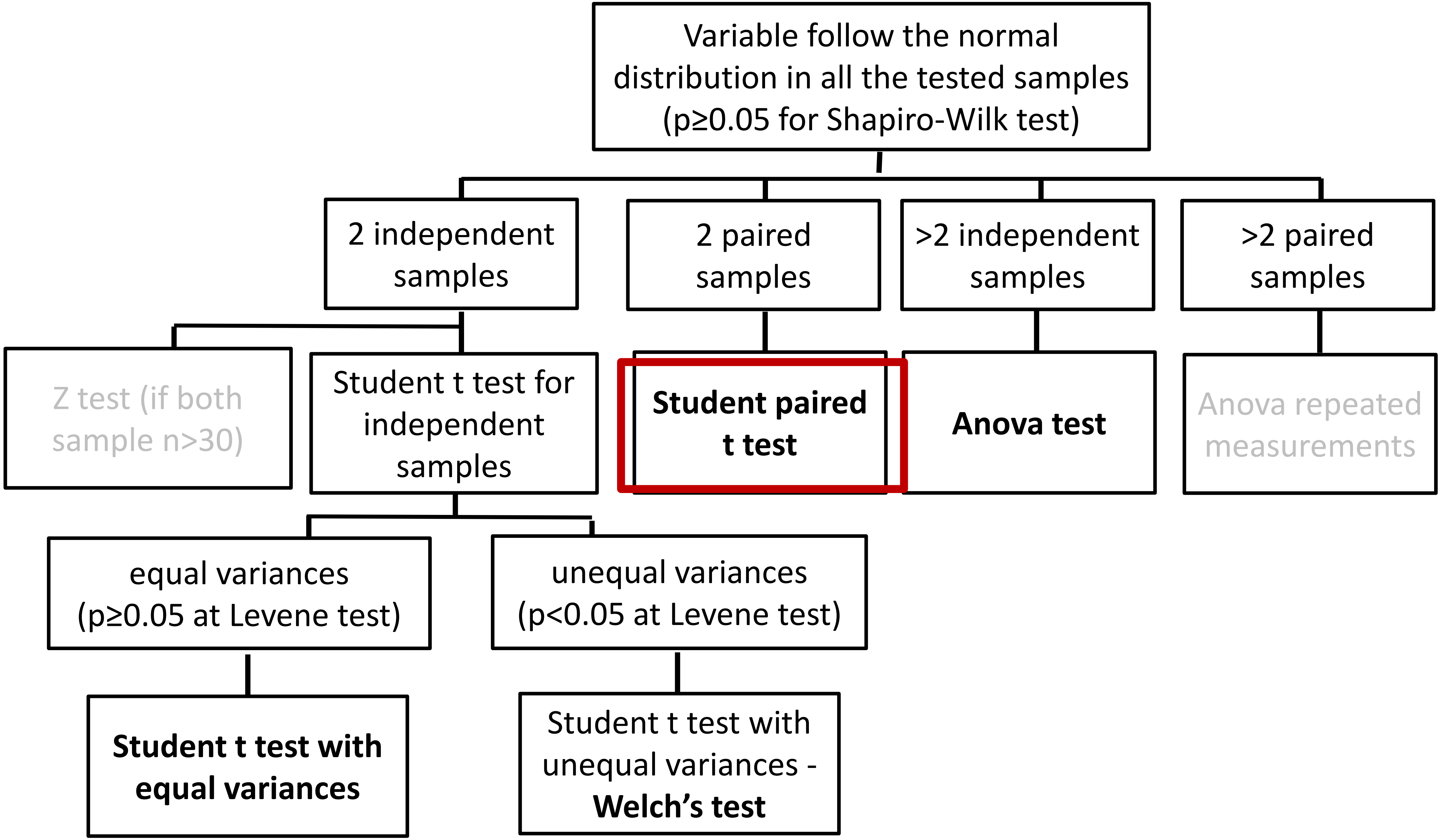
If variances are unequal, how are the means?

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

parametric 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

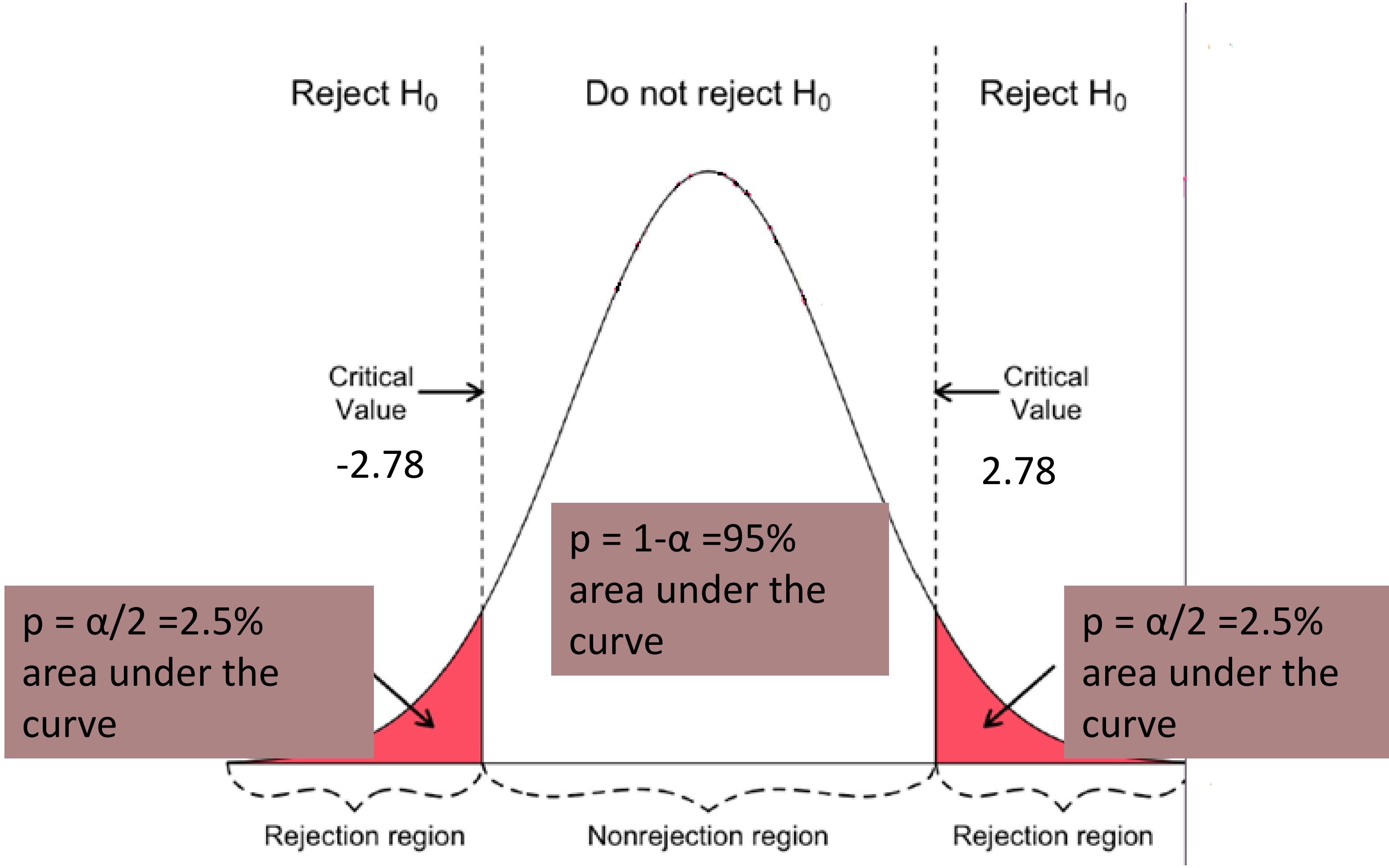
Scenario	**To investigate that the use of metamizol sodic monohidrat contributes to the occurrence of agranulocytosis. A random sample of individuals follow a treatement with metamizol sodic monohidrat in high dose. Were measured the number of neutrophils at the patient's admission 3.6+/-1.85 and at 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

Data

	Neutrofils (x1000/mm^3)		
	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

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

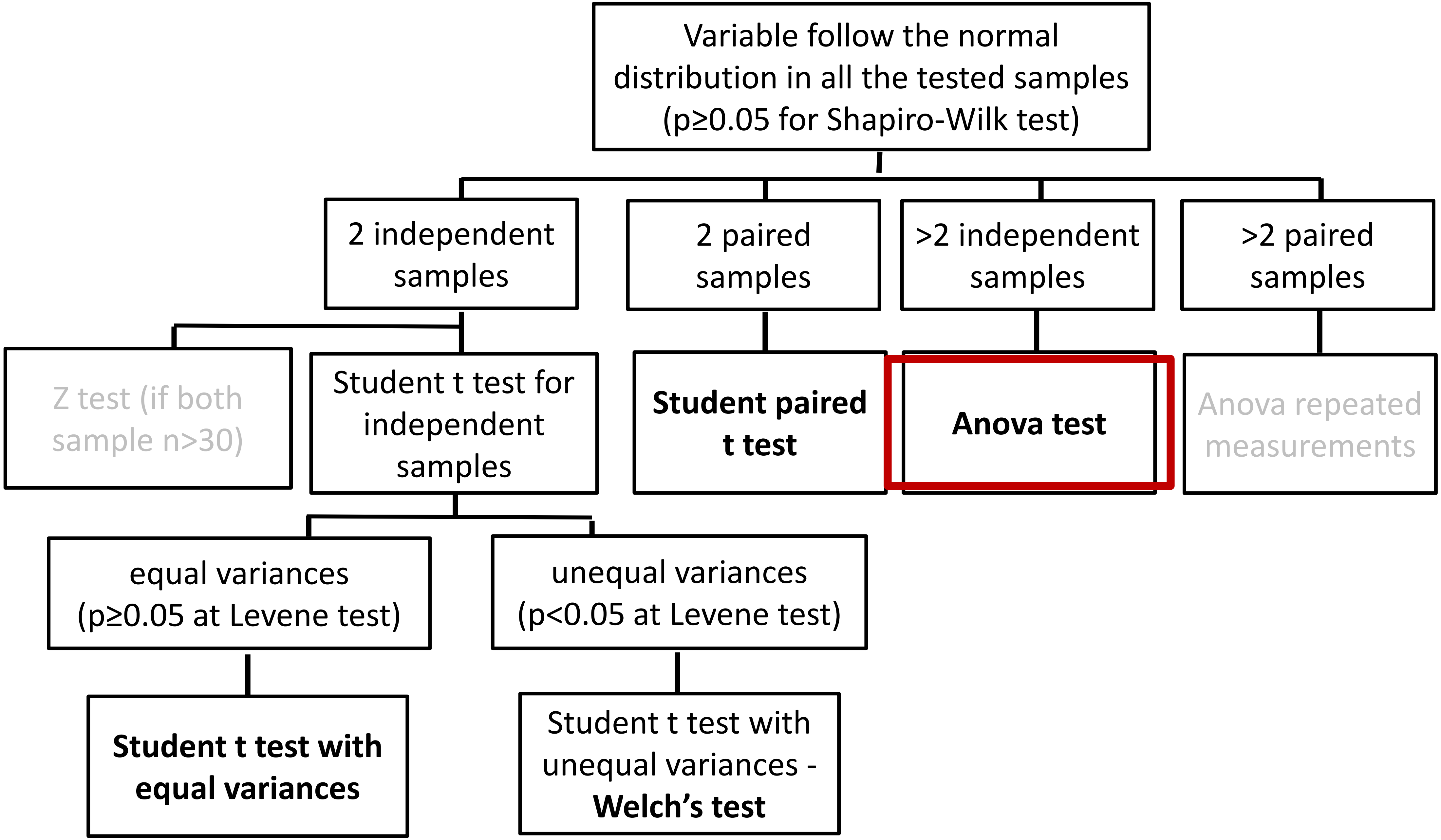
p=0.30

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

	Admission	Discharge
Arithmetic mean	3.60	2.90
Variances	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.


Variable follow the normal distribution in all the tested samples ($p \geq 0.05$ 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 ($p \geq 0.05$ 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

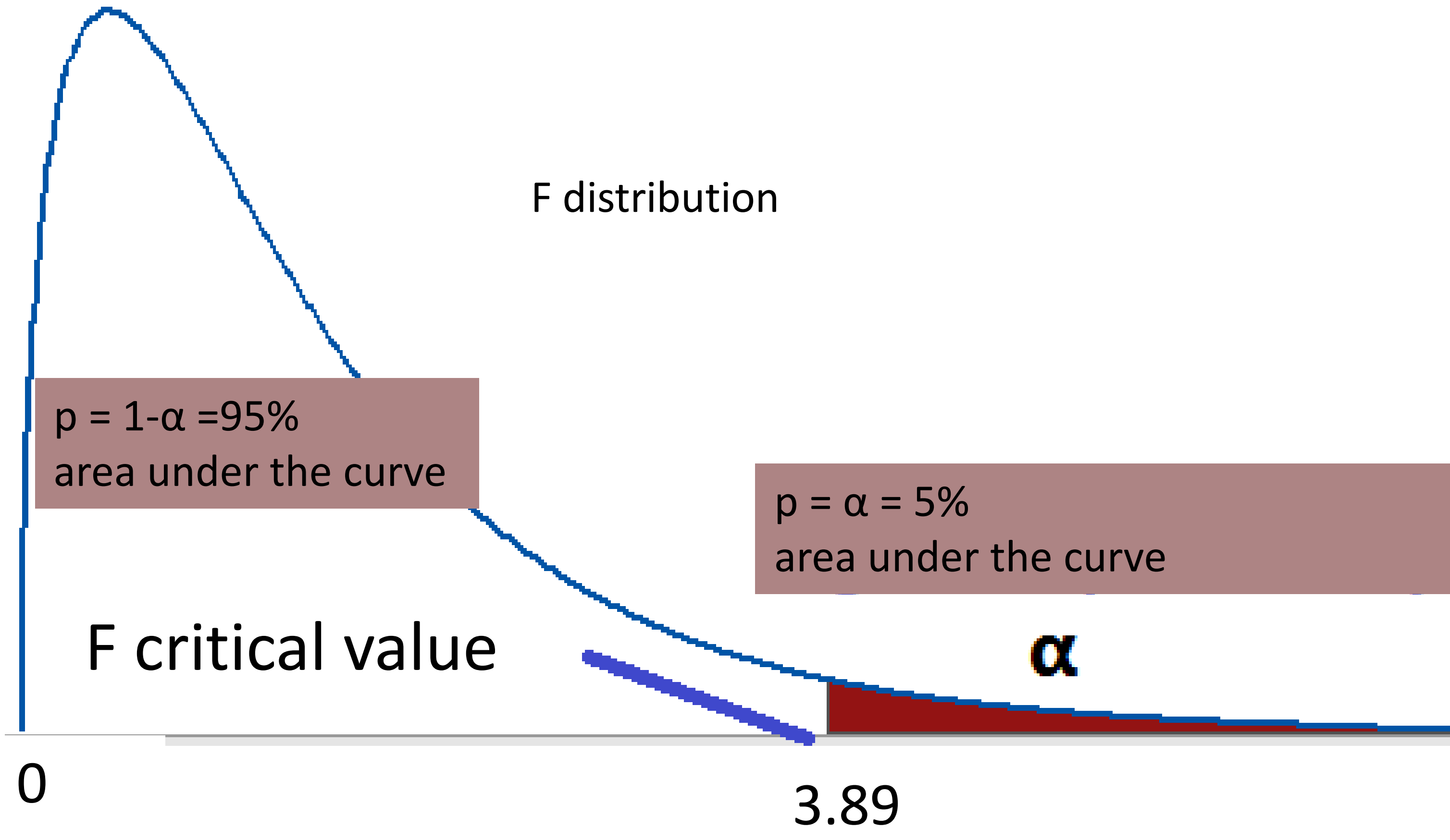
Multiple comparisons - ANOVA test

- We consider p independent samples with means $m_1,\ldots,m_p$
- Objective: to evaluate if there is a difference at the populations level between p means of a quantitative variable measured on p samples randomly extracted from the populations.

Example

Scenario	**To evaluate the association between the polymorphism of the gene encoding the vitamin D receptor (RVD) and obesity. The RVD polymorphism has three genotypes FF, ff and Ff. We select an equal number of subjects (5) with three different genotype. The BMI of the subjects was measured.**
Assumptions	
the observations are independent	Yes
independent samples>2	Yes, 3
the data are normally distributed	Yes
variances	equal
Name of the test	Anova test

- **The null hypothesis H_0:** There are no statistically significant **differences** in the mean BMI among the three groups of polymorphisms in the gene encoding the vitamin D receptor.
- **The alternative hypothesis H_1:** There are statistically significant **differences** in the mean BMI among the three groups of polymorphisms in the gene encoding the vitamin D receptor.



Calculate the parameter of the test F

- Calculation of the mean variation between groups

$MVBG = (n_1*(m_1-m_t)^2 + n_2*(m_2-m_t)^2 + n_3*(m_3-m_t)^2 + \ldots)/(p – 1)$

- Calculation of the mean variation within groups

$MVWG = ((n_1-1)*s_1^2 + (n_2-1)*s_2^2 + (n_3-1)*s_3^2 + \ldots)/(n-p)$

- The F statistic of the test:

$$F = MVBG / MVWG$$

- n=total number of observations,
- n_1,n_2,n_3 – the number of observations per group,
- p – number of groups

- m_t – the mean of all observations,
- m_1, m_2, m_3… the means per group,
- s_1, s_2, s_3, … sampling standard deviations of the groups

Calculate the parameter of the test F

- Calculation of the mean variation between groups

$MVBG = (n_1*(m_1-m_t)^2 + n_2*(m_2-m_t)^2 + n_3*(m_3-m_t)^2 + \ldots)/(p - 1)= (5*(40-34)^2+5*(25-34)^2+5*(30-34)^2)/2=332.5$

- Calculation of the mean variation within groups

$MVWG = ((n_1-1)*s_1^2+ (n_2-1)*s_2^2+ (n_3-1)*s_3^2+ \ldots)/(n-p)=(4*10^2+4*5^2+4*10^2)/(15-3)=75$

- The F statistic of the test:

$F = MVBG / MVWG=332.5/75=4.43$

$p=0.02$

- $n= 15$ total number of observations,
- $n_1=5$,$n_2=5$,$n_3=5$ – the number of observations per group,
- $p=3$ – number of groups

- $m_t=34$ – the mean of all observations,
- $m_1=40$, $m_2=25$, $m_3=30$… the means per group,
- $s_1=10$, $s_2=5$, $s_3=10$, … sampling standard deviations of the groups

6. Decision making

If F ∈ Rejection region, we reject H0 → accept H1 (otherwise, we FAIL to reject H0)

- If $p \geq 0.05$ fail to reject the null hypothesis: we cannot say that there are statistically significant differences between the samples mean, so all the samples belong to the same population
- If $p < 0.05$ reject the null hypothesis and validate the alternative hypothesis: there is a statistically significant difference between the samples mean. There are differences, but we do not know which means are different.

$F = 4.43 \in [3.89; +\infty)$, we reject H0, accept H1: the **means** of BMI **differs** statistically significantly between the three groups of RVD polymorphisms

$p = 0.02 < 0.05$, we reject H0, accept H1: the **means** of BMI **differs** statistically significantly between the three groups of RVD polymorphisms

if $p<0.05$ at Anova test we do not know which means are different

3 groups – 3 comparisons: 2^3 possible situations-1

post-hoc analysis

if we want to know which of the means are different

Comparision two by two:

FF with ff

FF with Ff

ff with Ff

Possible situations comparing two by two groups

Comparisons between			
FF group with the ff group	FF group with the Ff group	ff group with the Ff group	p for Anova
$p<0.05$	$p<0.05$	$p<0.05$	$p<0.05$, significant differences between groups
$p<0.05$	$p<0.05$	$p\geq0.05$	$p<0.05$, significant differences between groups
$p<0.05$	$p\geq0.05$	$p\geq0.05$	$p<0.05$, significant differences between groups
$p<0.05$	$p\geq0.05$	$p<0.05$	$p<0.05$, significant differences between groups
$p\geq0.05$	$p<0.05$	$p<0.05$	$p<0.05$, significant differences between groups
$p\geq0.05$	$p<0.05$	$p\geq0.05$	$p<0.05$, significant differences between groups
$p\geq0.05$	$p\geq0.05$	$p\geq0.05$	**$p\geq0.05$ no significant differences between groups**
$p\geq0.05$	$p\geq0.05$	$p<0.05$	$p<0.05$, significant differences between groups

3 groups – 3 comparisons: 2^3 = 8 possible situations-1

4 groups – 6 comparisons: 2^4 =16 possible situations-1

…

10 groups – 45 comparisions: 2^{10} = 1024 possible situations-1

…

big complexity

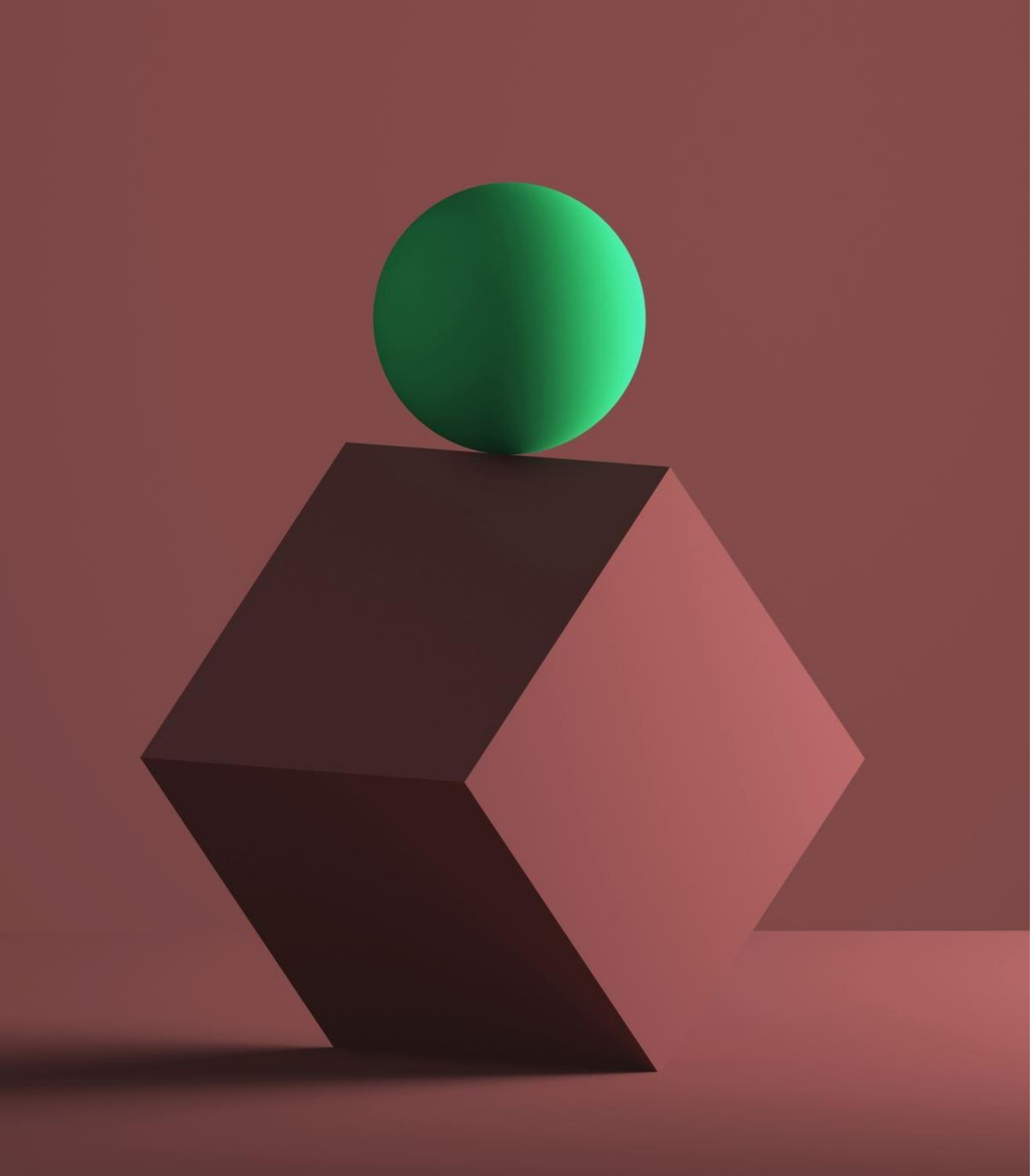
Post-hoc analysis = comparison two by two groups

- compare
  - sample 1 with sample 2 - error 5%
  - sample 1 with sample 3 - error 5%
  - sample 2 with sample 3 - error 5%
    - total α error = 15%

we nee to correct the error to have a significance level α = 0.05 for post-hoc tests

→ correction methods:

- Bonferroni correction:
  - we must compare the p of each T test not to 0.05, but to 0.05 / 3 (the number of tests performed) = 0.0166 as the significance level for the post hoc T test.
  - Or we can multiply the p of each test with 3, and compare them with 0.05 (the simplest technique).
- other correction Scheffe, Tukey etc.



Statistical tests – qualitative variables

Compare two frequencies

- Populations P_1, P_2,...
- Objective: the study of differences of a qualitative variable X frequency in two/more populations
- Frecvencies ϕ_1, ϕ_2, ...

How?

1.
- Extract representative samples from the populations

2.
- Calculate the indicators
  - Frecvencies f1, f2

3.
- **With statistical test**
  - verify the hypothesis: we do not have two different populations, but one population
  - obtain the probability p – the probability of finding differences equal to or smaller (extreme) than the one found on the samples under study if we repeat the study on other samples with random selection (to be due to chance)

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

Null hypothesis H0 - assumes 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

Alternative hypothesis H1 (denial 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

Statistical test --> we choose between the two possibilities H0 or H1

Important

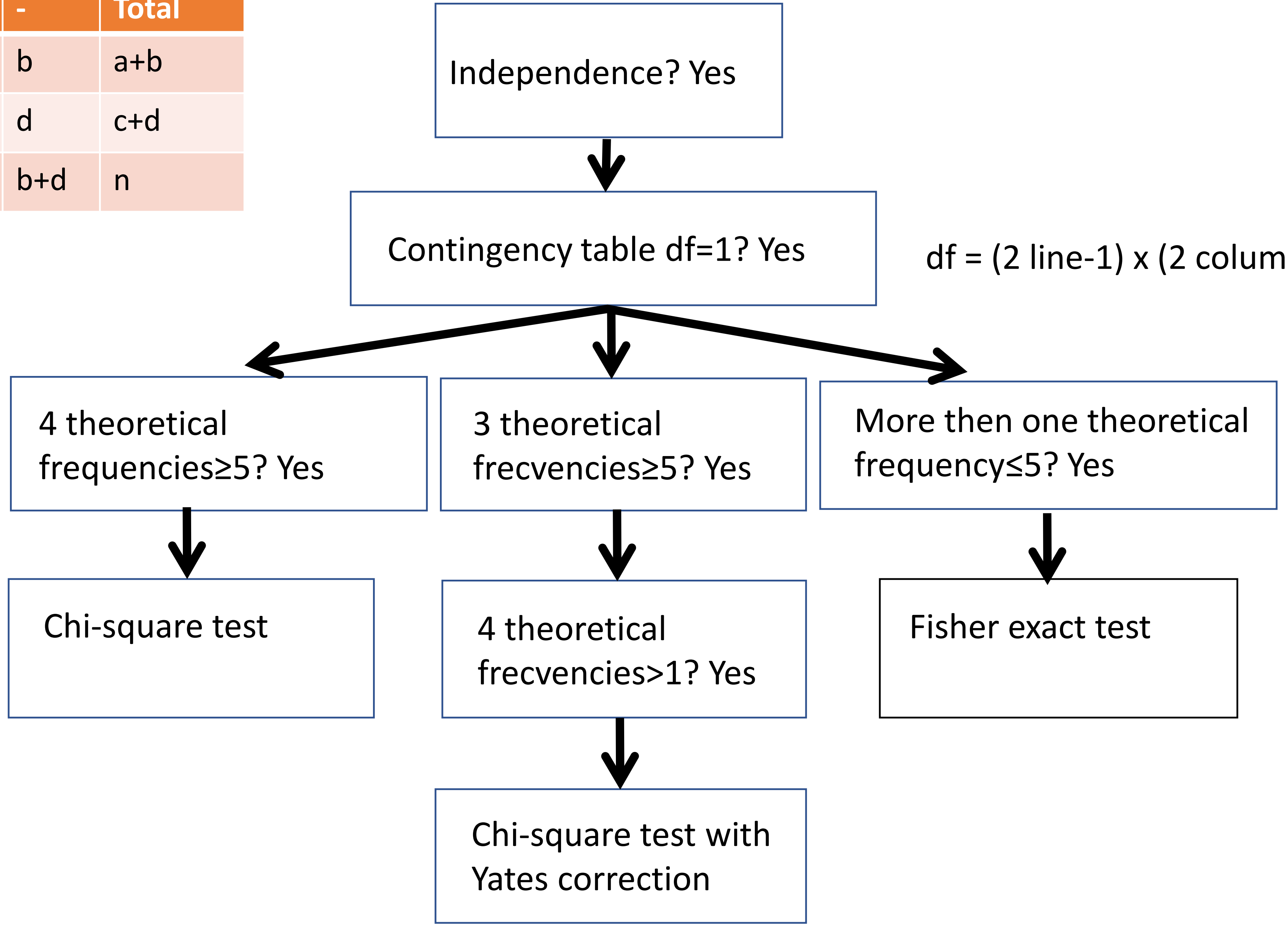
Specify the statistic you are comparing: e.g. frequencies

- in the statement of statistical hypotheses
- in the interpretation of the result of a statistical test

Use the expression

- "statistically significant difference"
- "significantly different"

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



df = (2 line-1) x (2 columns-1) = 1

Test for frequencies

Chi-square test

Scenario: In some cases, graft rejection occurs immediately after kidney transplantation.

Question: Obese people are more likely to experience graft rejection immediately after kidney transplantation?

Conducting the study

- We randomly select from among those who are going to have kidney transplants
  - 1500 obese people
  - 1500 non-obese people
- We track the occurrence of graft rejection over a period of 1 month after the transplant

Observed contingency table

	graft rejection $^+$	graft rejection $^-$	Total
Obese$^+$	50	1450	1500
Obese$^-$	62	1438	1500
Total	112	2888	3000

	(%) graft rejection $^+$
Obese	=50/1500*100= 3.33%
Non-obese	=62/1500*100= 4.13%

We found a difference between samples frequencies 3.33% - 4.13% = -0.8%

Formulating hypotheses

Null hypothesis H0

- There is **no** statistically significant difference between obese and non-obese patients in the frequency of graft rejection after (up to one month) kidney transplantation
- There is **no** statistically significant association between obesity and graft rejection after (up to one month) kidney transplantation

Alternative hypothesis H1

- There is a statistically significant difference between obese and non-obese individuals in the frequency of graft rejection after (up to one month) kidney transplantation
- There is a statistically significant association between obesity and graft rejection after (up to one month) kidney transplantation

both are good

Observed frequency table

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

Theoretical frequency table

	+	-	Total
+	$\frac{(a+c)*(a+b)}{n}$	$\frac{(b+d)*(a+b)}{n}$	a+b
-	$\frac{(a+c)*(c+d)}{n}$	$\frac{(b+d)*(c+d)}{n}$	c+d
Total	a+c	b+d	n

- We assume by absurdity that the null hypothesis is true
- We calculate the theoretical table in which obesity is not a risk factor

Calculation of the theoretical table (null)

- We assume by absurdity that the null hypothesis is true
- We calculate the theoretical table in which obesity is not a risk factor

	graft rejection $^{+}$	graft rejection $^{-}$	Total
Obese^{+}	50	1450	1500
Obese^{-}	62	1438	1500
Total	112	2888	3000

	graft rejection $^{+}$	graft rejection $^{-}$	Total
Obese^{+}	=(112·1500)/3000=56	=(2888·1500)/3000=1444	1500
Obese^{-}	=(112·1500)/3000=56	=(2888·1500)/3000=1444	1500
Total	112	2888	3000

Calculation of the theoretical table (null)

- Obesity is not a risk factor in this table
- Obese and non-obese have graft rejection in the same proportion 56/1500=3.73%

	graft rejection +	graft rejection -	Total
Obese+	56	1444	1500
Obese-	56	1444	1500
Total	112	2888	3000

	graft rejection + (%)
Obese	=56/1500=3.73%
Non-Obese	=56/1500=3.73%

- We compare the observed table with the theoretical one
- If we find a small difference (below the critical threshold), then obesity is not a risk factor
- If we find a large difference (above the critical threshold), then obesity is a risk factor

- Observed contingency table

	graft rejection $^+$	graft rejection $^-$	Total
Obese$^+$	50	1450	1500
Obese$^-$	62	1438	1500
Total	112	2888	3000

?=

- Theoretic (Null) contingency table

	graft rejection $^+$	graft rejection $^-$	Total
Obese$^+$	56	1444	1500
Obese$^-$	56	1444	1500
Total	112	2888	3000

We calculate the difference between the tables

Observed frequency table

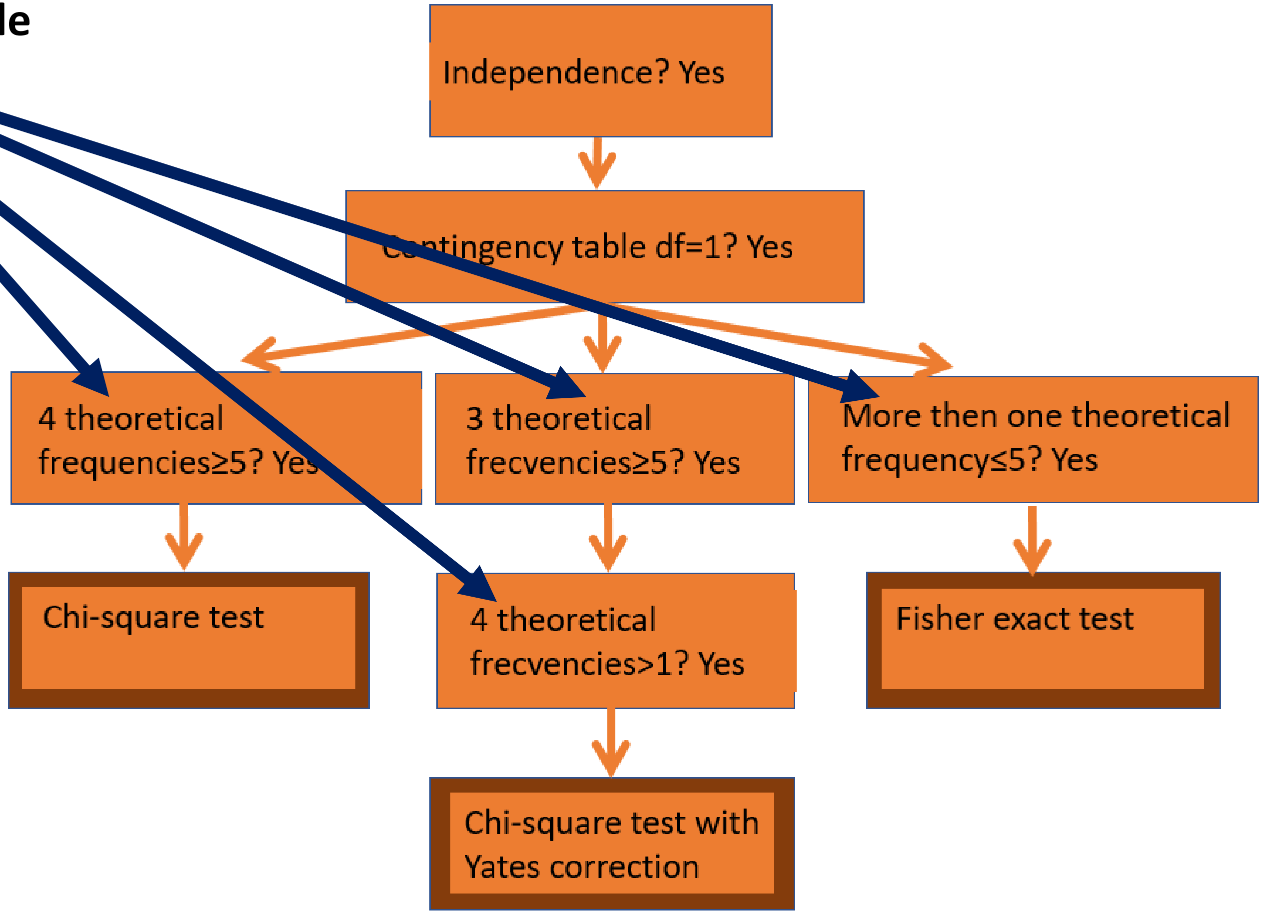
	+	-	Total
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n

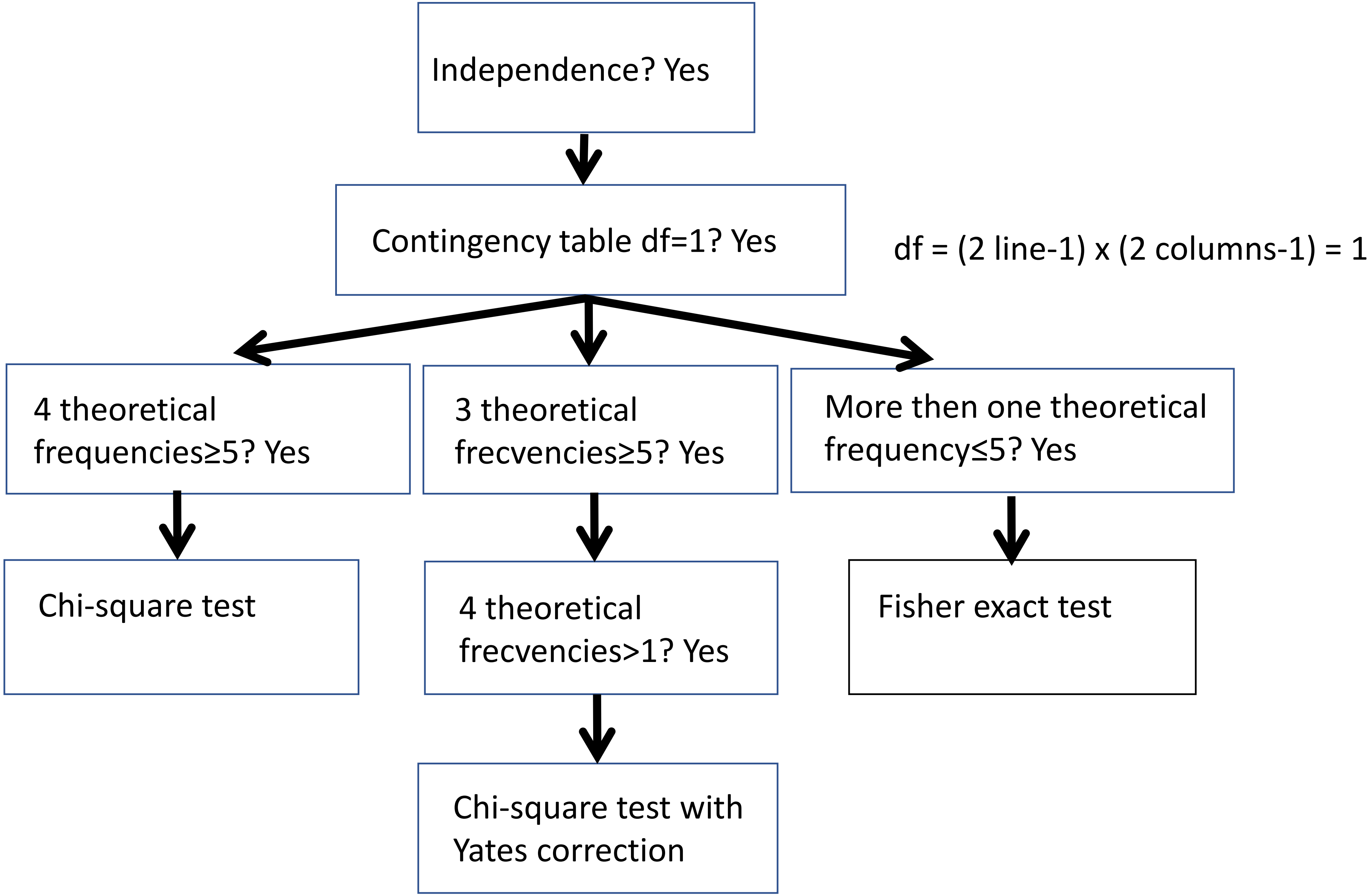
Choosing the test

Theoretical frequency table

	+	-	Total
+	$\frac{(a+c)*(a+b)}{n}$	$\frac{(b+d)*(a+b)}{n}$	a+b
-	$\frac{(a+c)*(c+d)}{n}$	$\frac{(b+d)*(c+d)}{n}$	c+d
Total	a+c	b+d	n

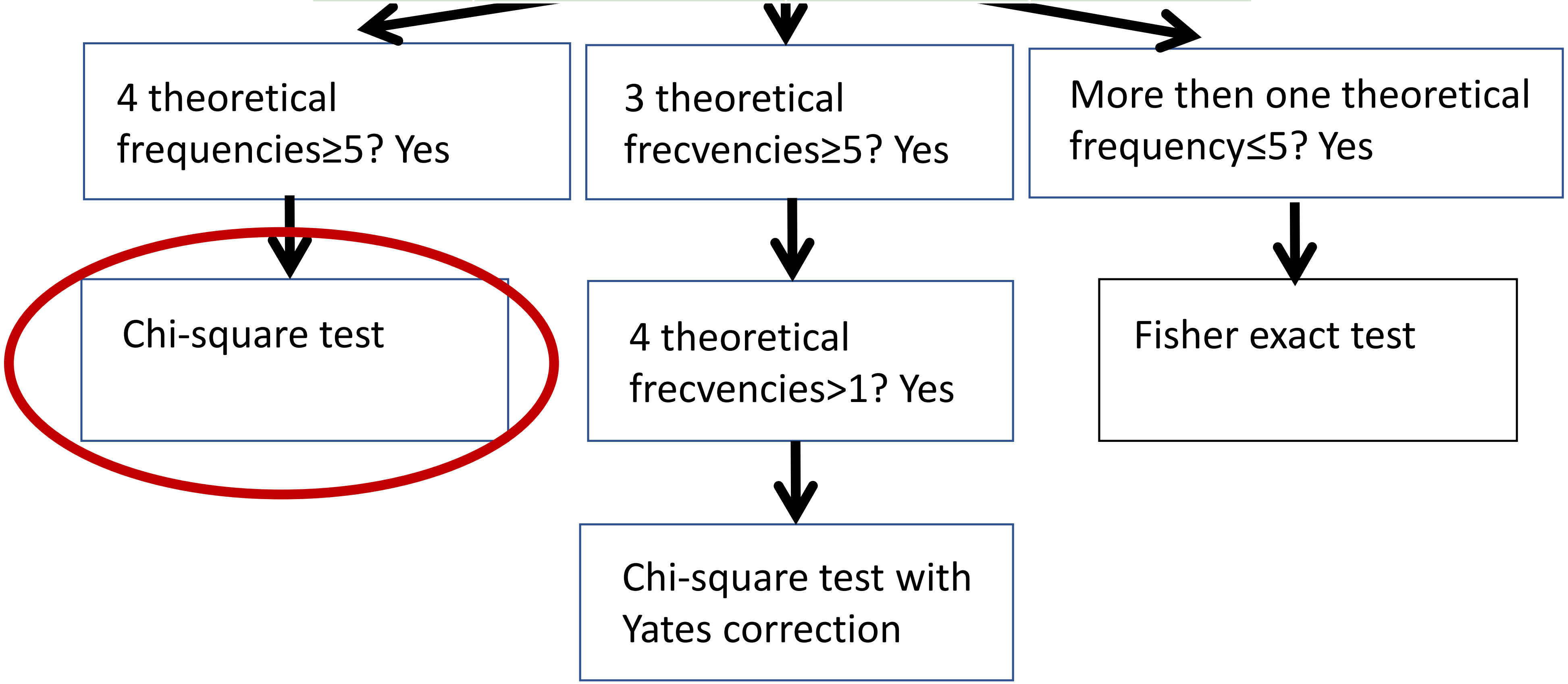
df = (2 line-1) x (2 columns-1) = 1




Independence? Yes
Contingency table df=1? Yes
df = (2 line-1) x (2 columns-1) = 1
4 theoretical frequencies≥5? Yes
3 theoretical frecvencies≥5? Yes
More then one theoretical frequency≤5? Yes
Chi-square test
4 theoretical frecvencies>1? Yes
Fisher exact test
Chi-square test with Yates correction

	graft rejection $^{+}$	graft rejection $^{-}$	Total
Obese^{+}	56	1444	1500
Obese^{-}	56	1444	1500
Total	112	2888	3000

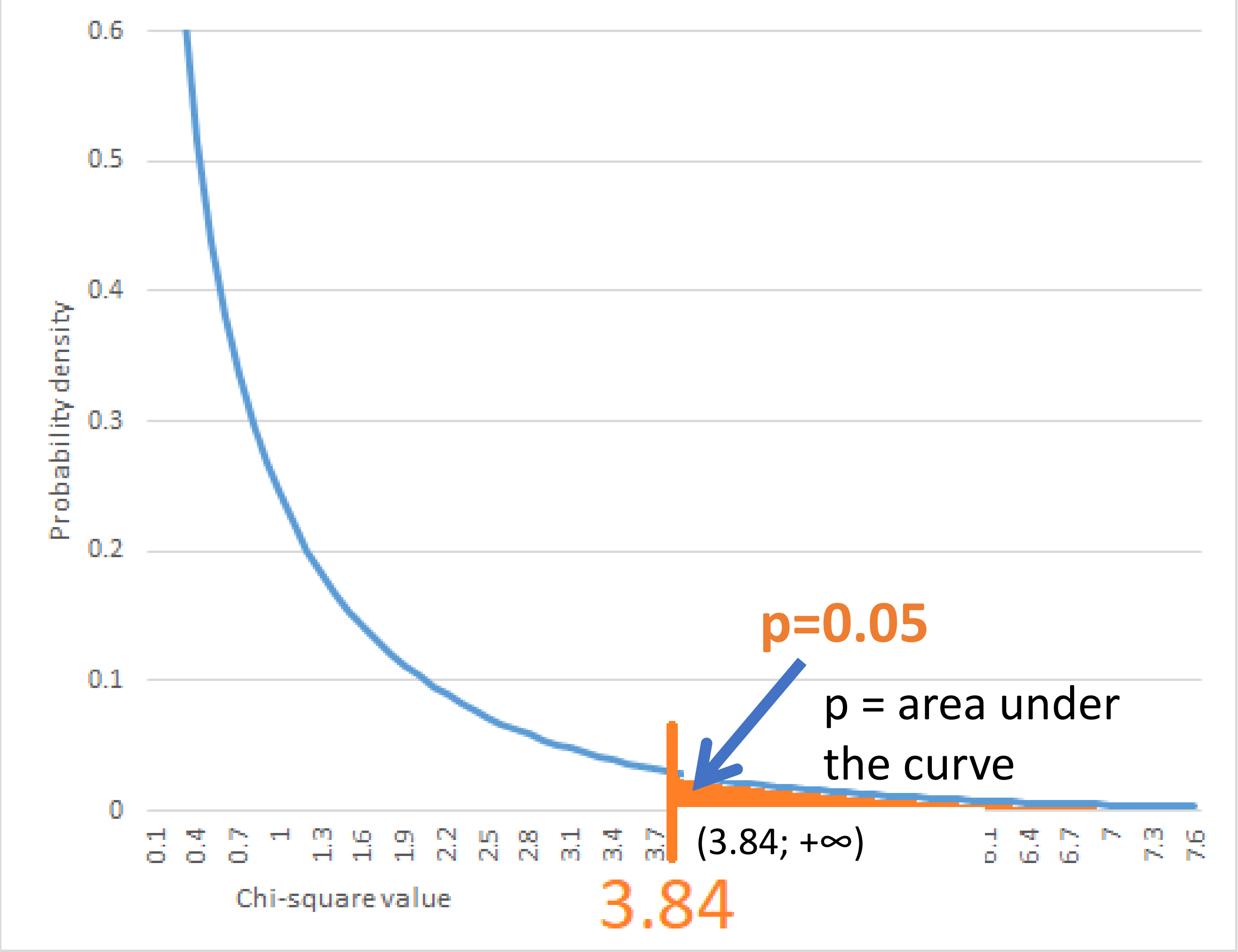
2 line-1) x (2 columns-1) = 1



Choosing the significance level and establishing the critical region

- Statistical significance **α=5%**
- Rejection region (3.84; +∞)

- Acceptance area (0; 3.84]



test parameter follow the Chi-square distribution

- Observed contingency table

	graft rejection +	graft rejection -	Total
Obese+	50	1450	1500
Obese-	62	1438	1500
Total	112	2888	3000

$\neq$?

- Theoretic contingency table

	graft rejection +	graft rejection -	Total
Obese+	56	1444	1500
Obese-	56	1444	1500
Total	112	2888	3000

parameter of the test

$$\chi^2 = \sum_{i=1}^{4} \frac{(f_i^o - f_i^t)^2}{f_i^t}$$

f_i^o = observed frequency, f_i^t = theoretical frequency

- Observed contingency table

	graft rejection +	graft rejection -	
Obese+	50	1450	
Obese-	62	1438	

- Theoretic contingency table

	graft rejection +	graft rejection -	
Obese+	56	1444	
Obese-	56	1444	

$$\chi^2 = \sum_{i=1}^{4} \frac{(f_i^o - f_i^t)^2}{f_i^t} = \frac{(f_1^0 - f_1^t)^2}{f_1^t} + \frac{(f_2^0 - f_2^t)^2}{f_2^t} + \frac{(f_3^0 - f_3^t)^2}{f_3^t} + \frac{(f_4^0 - f_4^t)^2}{f_4^t}$$

difference

$$= \frac{(50 - 56)^2}{56} + \frac{(1450 - 1444)^2}{1444} + \frac{(62 - 56)^2}{56} + \frac{(1438 - 1444)^2}{1444} =$$

=0.64+0.002+0.64+0.02 = 1.33

p=0.248

A. Test decision according to the rejection region:

If χ2 belongs to (3.84; +∞)

we have enough evidence to reject H0, so we accept H1

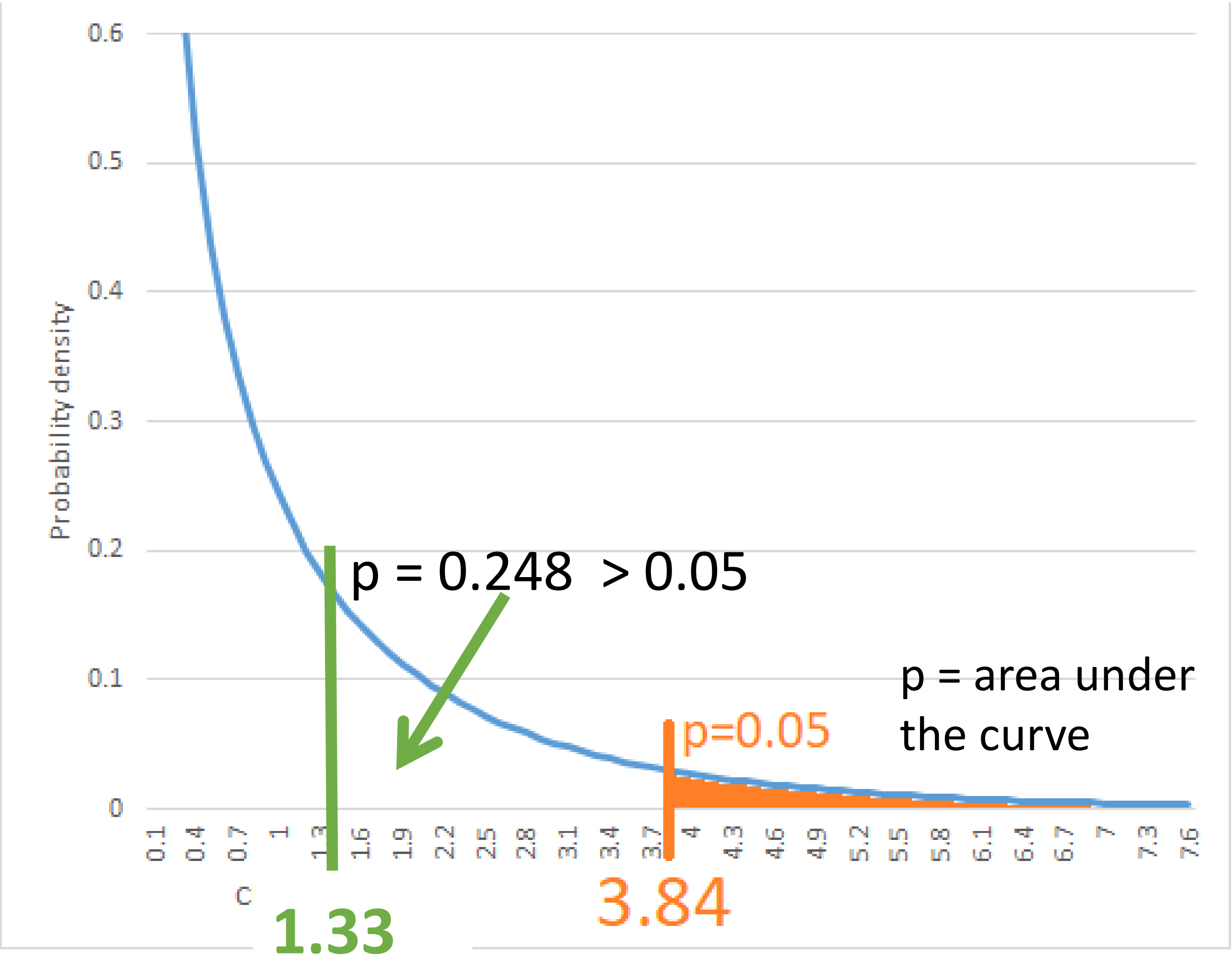
If χ2 DOES NOT belong to (3.84; +∞)

we do NOT have enough evidence to reject H0, so we are in favor of H0

In our case: χ2 = 1.33 does not belong to (3.84, + ∞),

we do NOT have enough evidence to reject H0, so we are in favor of H0

Conclusion: There is no statistically significant association between obesity and graft rejection in the first month after kidney transplantation



B. Test decision based on p probability – the probability of finding a difference equal to or smaller than the one found if we repeat the study:

If $p < 0.05$

we have enough evidence to reject H0, so we accept H1

If $p \geq 0.05$

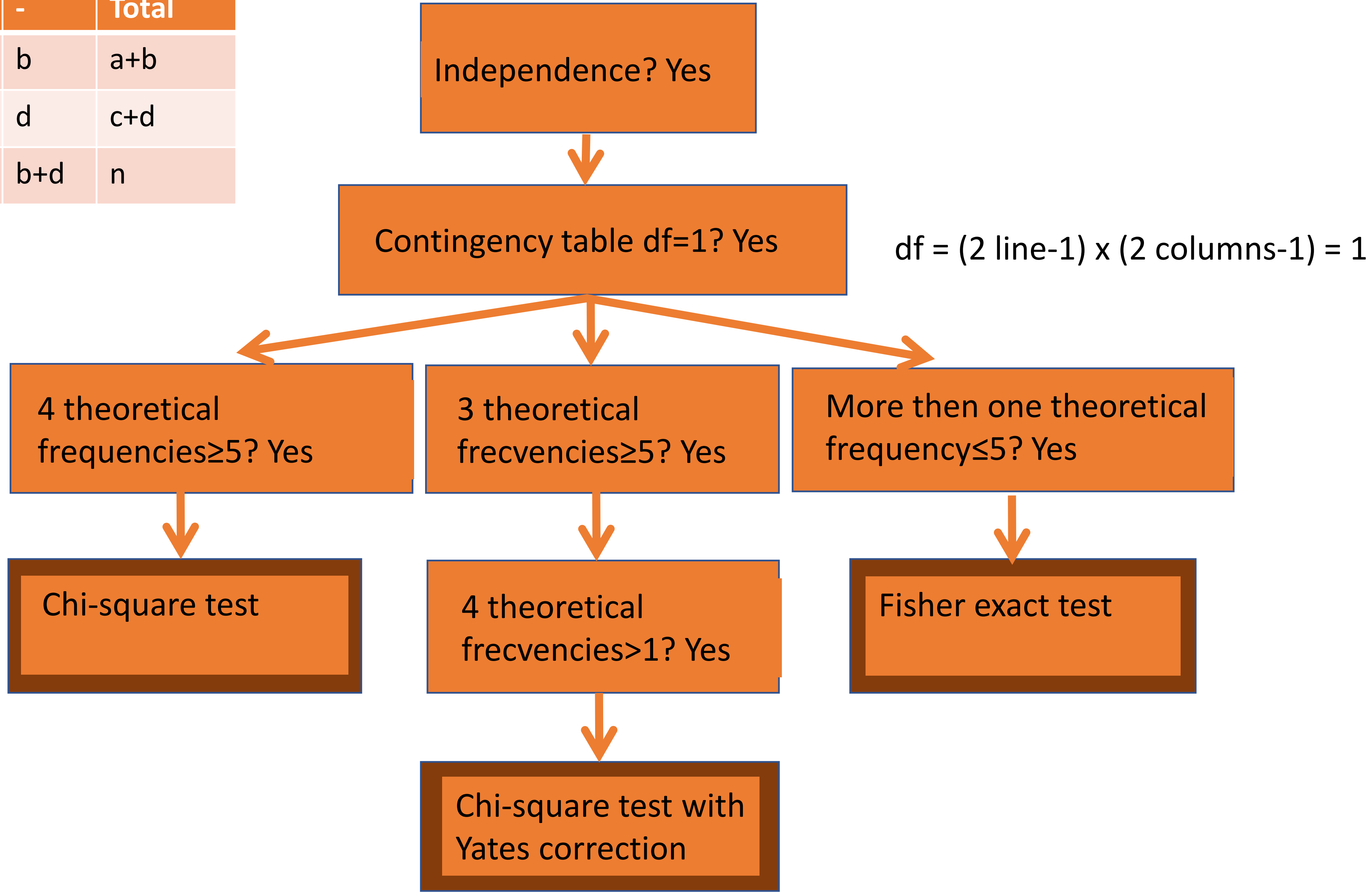
we do NOT have enough evidence to reject H0, so we are in favor of H0

In our case: $p = 0.248 > 0.05$

we do NOT have enough evidence to reject H0, so we are in favor of H0

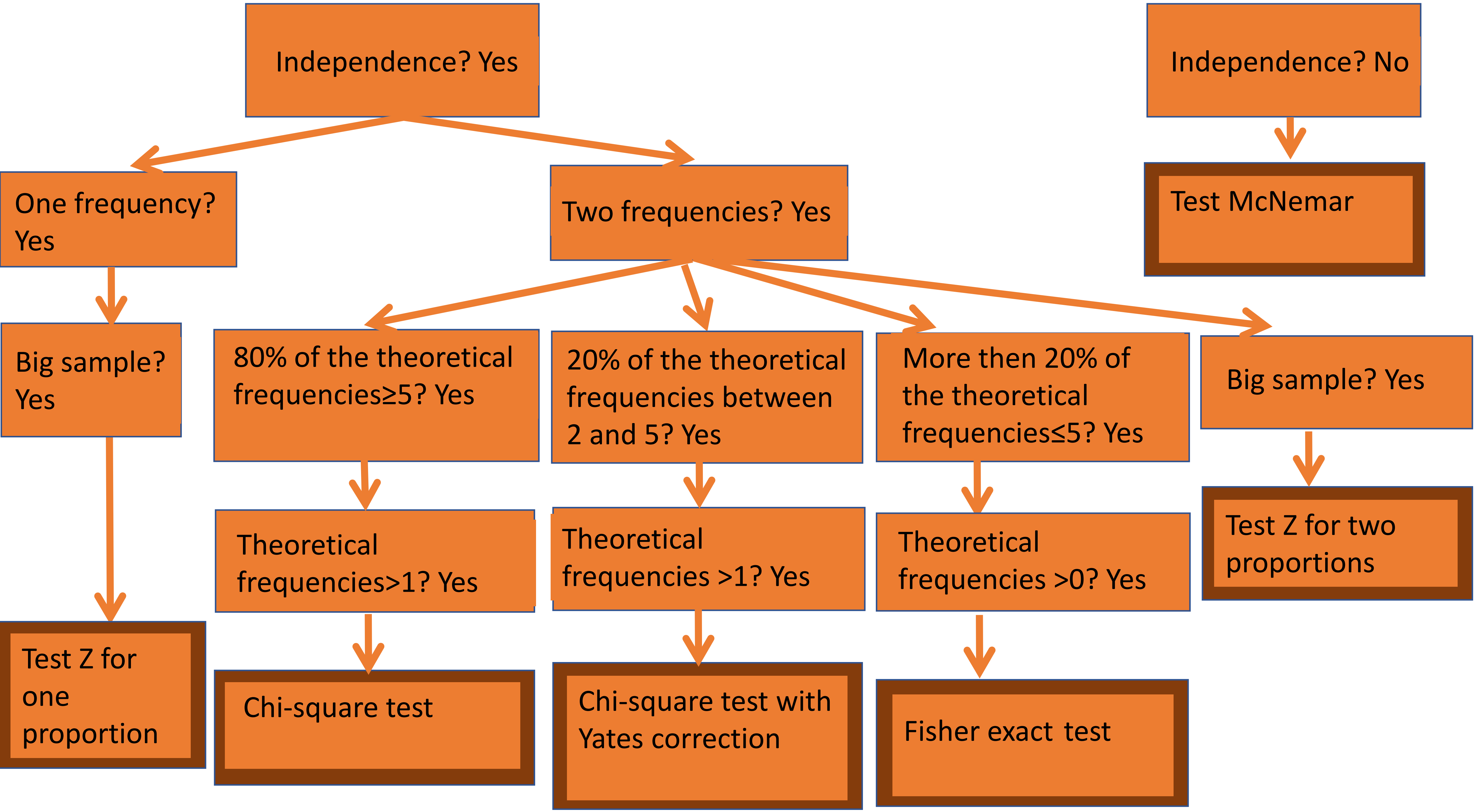
Conclusion: There is no statistically significant association between obesity and graft rejection after (up to one month) kidney transplantation, i.e. the probability of finding a difference equal to or smaller than the found one if we repeat the study is 1.33%

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



df>1

	C1	C2	C3	…	Total
L1	a	b	c	…	a+b+c+…
L2	d	e	f	…	d+e+f+…
L3	g	h	i	…	g+h+i+…
…	…	…	…	…	…
Total	a+d+g+…	b+e+h+…	c+f+i+…	…	n


Independence? Yes
Independence? No
Test McNemar
One frequency? Yes
Two frequencies? Yes
Big sample? Yes
80% of the theoretical frequencies≥5? Yes
20% of the theoretical frequencies between 2 and 5? Yes
More then 20% of the theoretical frequencies≤5? Yes
Big sample? Yes
Theoretical frequencies>1? Yes
Theoretical frequencies >1? Yes
Theoretical frequencies >0? Yes
Test Z for two proportions
Test Z for one proportion
Chi-square test
Chi-square test with Yates correction
Fisher exact test

*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

* Drinking alcohol in large amount is suspected to be associated with the onset of ageusia (lack of taste). A sample of 500 people was studied: 60 presented ageusia from which 30 reported to be heavy drinkers. A number of 410 patients without ageusia and no alcohol drinkers were identified. Which statistical test should be used to test the null hypothesis "Drinking alcohol in large amount and ageusia are independent":

A. Chi-square test

B. This is not a null hypothesis

C. Student t test

D. Levene test

E. Anova test

* Researcher want to see if eating meat is a risk factor for dental caries. A sample of 300 persons was studied: 240 presented dental caries and from which 150 reported to eat meat in large amount. A number of 40 subjects without dental caries and vegetarian were identified. 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 alternative hypothesis H_a

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

D. Reject null hypothesis H_0, accept null hypothesis H_0

E. Fail to reject alternative hypothesis H_a

- Obs. to respond you need to apply the chi-square test. Rejection area $[\chi_\alpha, \infty)$ = $[3.84, \infty)$

- 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, F test p<0.001. Select all correct answers for α = 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, F test p<0.001. Select all correct answers for α = 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!