Vall d'Hebron Institut de Recerca

Hypothesis Testing. Quantitative Variables

Basic Statistics with R

UEB-VHIR

Mireia Ferrer, Miriam Mota, Alex Sanchez and Santiago Perez-Hoyos

10/05/2021

Vall d'Hebron Hospital · Generalitat de Catalunya Departament de Salut · Institut Català de la Salut · UAB Universitat Autònoma de Barcelona · Generalitat de Catalunya Departament d'Economia i Coneixement · Institució CERCA Centres de Recerca de Catalunya · Instituto de Salud Carlos III

Vall d'Hebron Institut de Recerca

UNITAT D'ESTADÍSTICA I BIOINFORMÀTICA

Syllabus

1. INTRODUCTION
2. TYPE OF TEST
3. NORMALITY TEST
4. **ONE GROUP** COMPARISON
5. **TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
6. **TWO GROUPS** COMPARISON IN DEPENDENT SAMPLES
7. **MORE THAN TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
8. MULTIPLE COMPARISONS AND MULTIPLE TESTING

Introduction

- Once the concept of hypothesis testing is established,
- Researchers face the problem of _which test should be applied at every possible situation_.
- Best (ideal) solution:
  - understand the problem and the questions addressed
  - know available tests for each problem
  - be aware of applicability assumptions of each test and how to check them.
- Easier to say than to do.
  - Sometimes cheatsheets may be helpful,
  - but be warned against a blind use,
  - **that is understand and be critic with the steps**.

How to choose the right statistical test

Example Data/Problem

- A study was designed to compare two distinct hypertension control programs.
- 60 individuals with HTA were randomly assigned to either one or the other group (30 per group)
- Blood pressure was measured each month during a year

A	B	C	D	E	F	G	H	I
numero	sexo	grupo	tas1	tad1	tas2	tad2	tas3	tad3
1	VARON	B	150	100	150	90	170	
2	MUJER	B	160	90	170	90	160	
3	MUJER	B	150	90	110	90	115	
4	VARON	A	120	80	140	90	140	
5	MUJER	A	150	85	145	85	160	
6	MUJER	B	140	75	160	70	135	
7	MUJER	A	150	100	140	90	130	
8	VARON	A	160	90	170	90	170	
9	MUJER	A	145	105	170	95	140	
10	MUJER	A	210	110				
11	MUJER	A	170	100	170	90	170	
12	MUJER	B	140	90	140	90	100	

Questions to answer

- Is diastolic (min) tension above 90, “on average”, at the beginning of the study.
- Are samples “comparable” at baseline?
  - In Age? Sex%? Sist? Diast?
- Has there been a change in BP between month 1 (first measure) and month 12?
- Has this change been different between tretament groups?

Let’s think about it

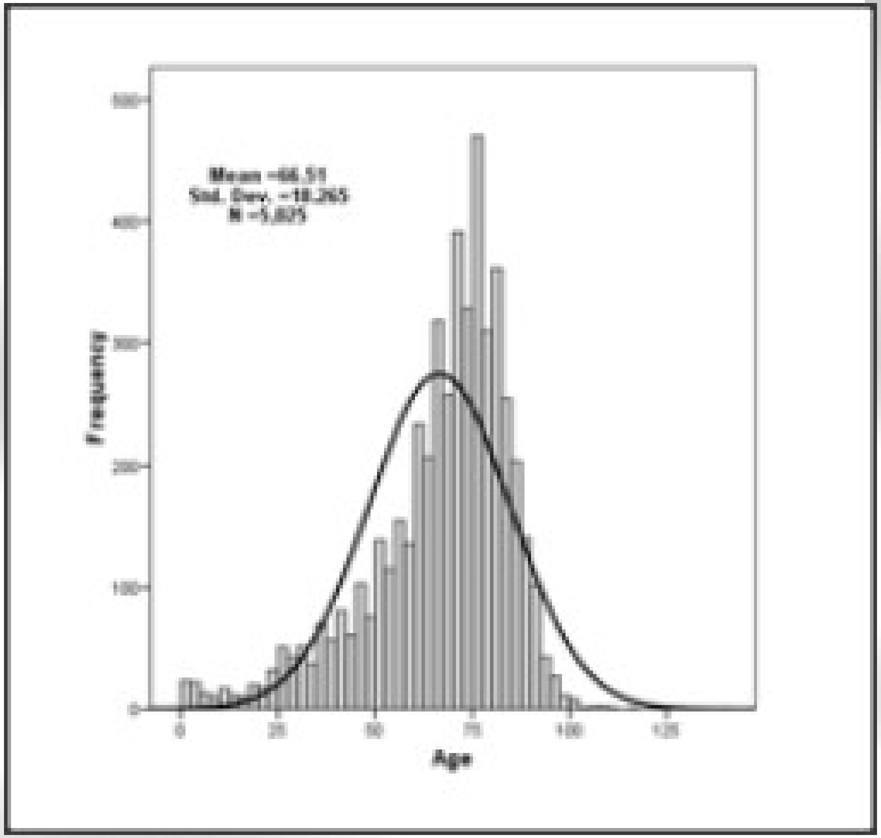
- Is diastolic (min) tension above 90, “on average”, at the beginning of the study.
  - *One variable. Test about the mean*
- Are samples “comparable” at baseline?
  - *Comparison between distinct individuals in two groups (A/B, Male/Fem)*
- Has there been a change in BP between month 1 (first measure) and month 12?
  - *Comparison between same individuals*

Our goal

- Lear how to do tests about characteristics of one variable.
- Distinct scenarios
  - One group, Two groups, Paired data
  - Normally or non normally distributed data
- We start checking normality
  - Depending on what we obtain →
  - Use parametric or non-parametric test
- Be flexible and comprehensive!

Normality test

- Some parametrical test assume data come from a normal population
- How can we check this assumption?
- What can we do if assumption is false?

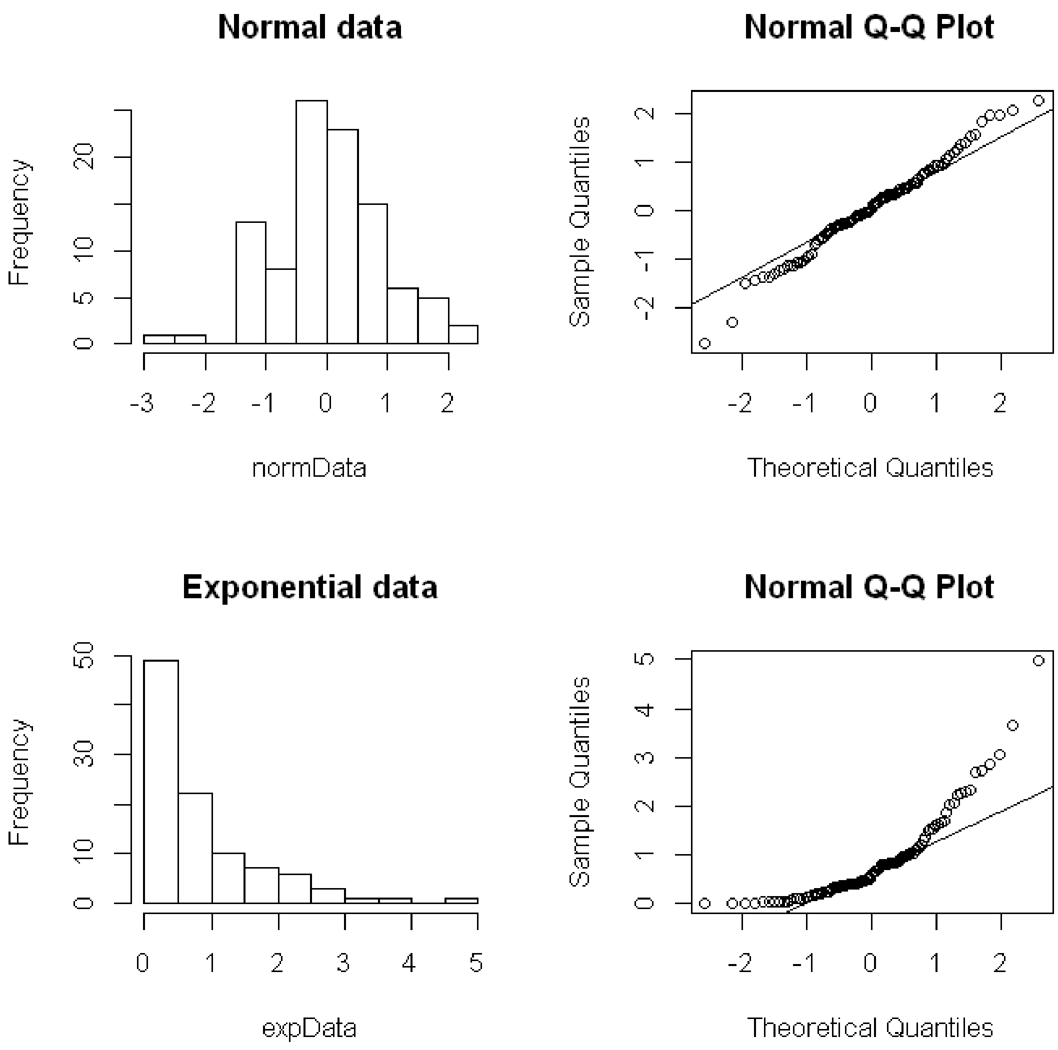


Testing normality

- We can use graphical methods or hypothesis tests
- Graphs
  - Check if it is a symmetric distribution
  - Probability graphs (QQ-plots)
- Hypothesis test (Normality)
  - Kolmogorov-Smirnov test
  - Kolmogorov-Liliefors test
  - Shapiro-Wilks test

Histograms and QQ-plots

- Histogram
  - It should be symmetric with gaussian shape.
- QQ-plot
  - Dots should be over the diagonal line
- Non normal data deviate from normal patterns.
- Difficult to quantify if there are few data

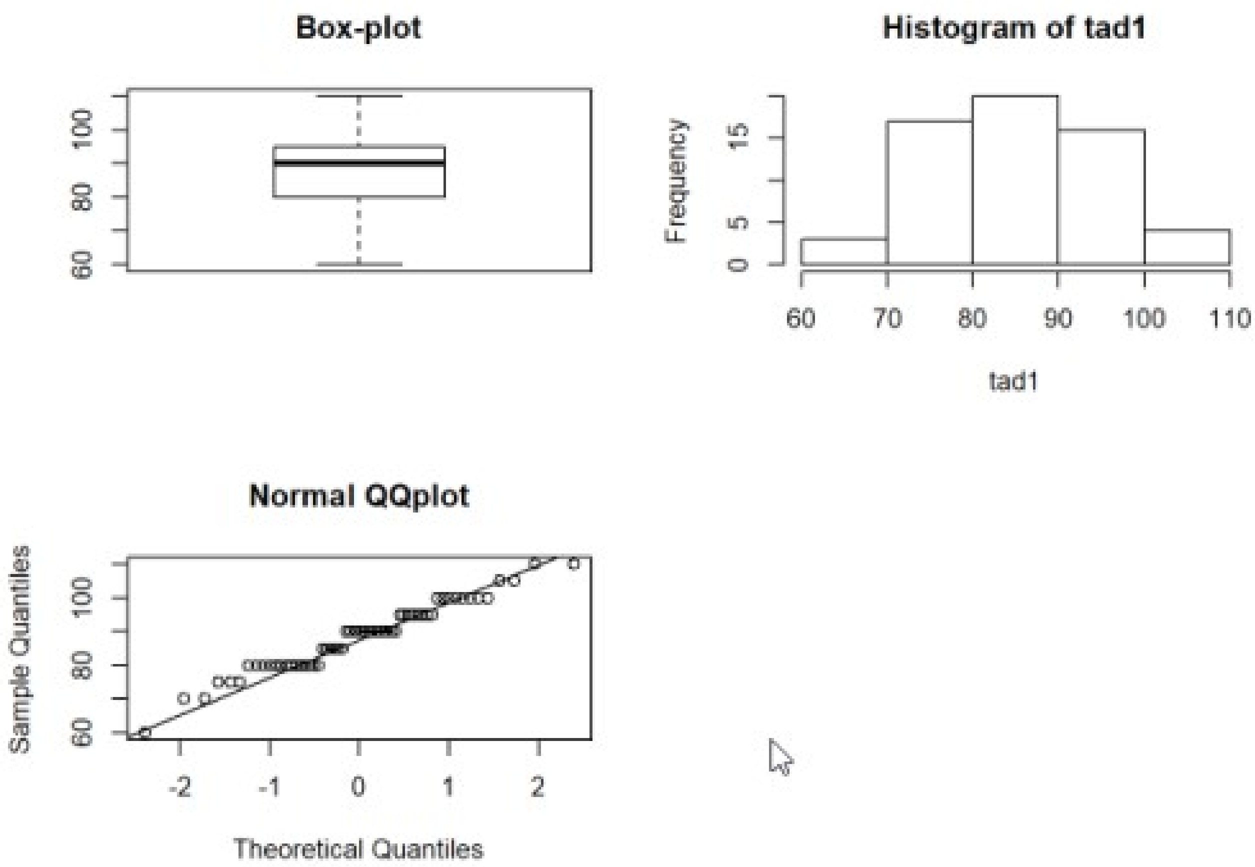


```r
oldpar<-par(mfrow=c(1,1)) # Guarda los parámetros para el dibgujo
par(mfrow=c(2,2)) # Dibuja cuatro gráficos por grafico
with(hta, boxplot(tad1, main="Box-plot") )

with(hta, hist(tad1) )

with(hta, qqnorm(tad1, main="Normal QQplot") )
;with(hta, qqline(tad1) )

par(oldpar) # Vuelve a los parámetros de dibujo originales
```



Normality tests

- Statistical normality test are more precise than graphs. It is possible to calculate a p-value.

- The most used tests are Kolmogorov-Smirnov and Shapiro-Wilks test.

- The hypothesis to test are:
  - H_0: Data follow a normal distribution
  - H_1 : Data do not follow a normal distribution

Normality Test

```
```{r normtest , mysize=TRUE, size='\\small'}
with(hta,shapiro.test(tad1) ) # Shapiro Wilk test

```
```

```
	Shapiro-Wilk normality test

data:  tad1
W = 0.96622, p-value = 0.09512
```

Syllabus

1. INTRODUCTION
2. NORMALITY TEST
3. **ONE GROUP** COMPARISON
4. **TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
5. **TWO GROUPS** COMPARISON IN DEPENDENT SAMPLES
6. **MORE THAN TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
7. MULTIPLE COMPARISONS AND MULTIPLE TESTING

One sample t-test

- We do not use it very often.
- Very similar to estimation questions. It can be solved calculating a confidence interval
- Idea: We want to verify from a sample a previous hypothesis about the mean in a population
- *Can it be accepted that the initial TAD is 90* ***or greater*** *in Hipertensive patients?*

One sample Test

```
with(hta,t.test(tad1,mu=90) ) # One sample T.test
```

```
##
##  One Sample t-test
##
## data:  tad1
## t = -1.2137, df = 59, p-value = 0.2297
## alternative hypothesis: true mean is not equal to 90
## 95 percent confidence interval:
##  85.80626 91.02707
## sample estimates:
## mean of x
##  88.41667
```

Exercise 1

a) Check the normality of *tas1* variable in **hta** dataset

b) Can it be accepted that the initial TAS is 120 in Hipertensive patients?

c) Find the 95% confidence interval for the mean of tas1 variable

d) Extra: Can it be accepted that the initial TAS is *higher than* 120 in Hipertensive *women*?

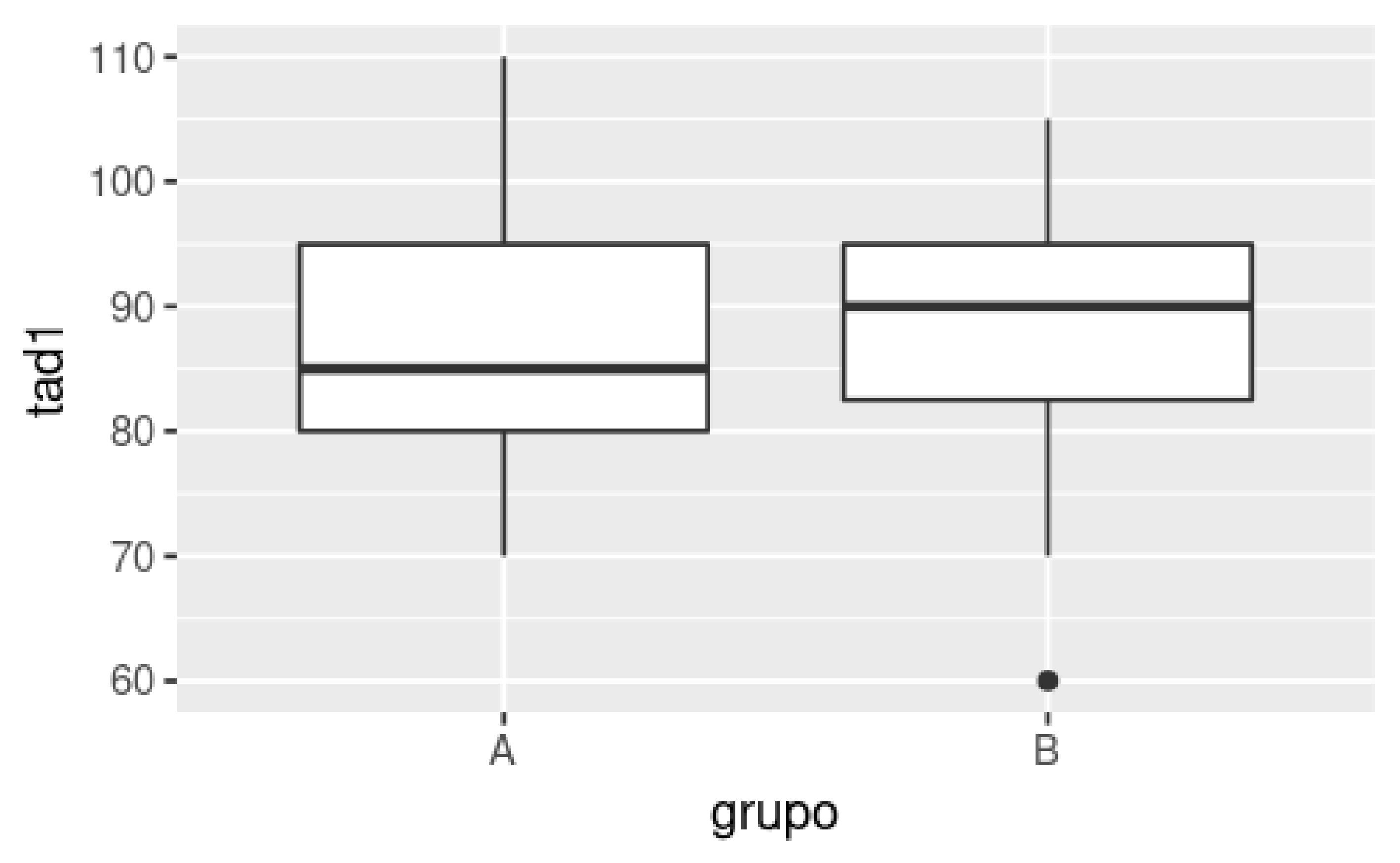
Syllabus

1. INTRODUCTION
2. NORMALITY TEST
3. **TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
4. **TWO GROUPS** COMPARISON IN DEPENDENT SAMPLES
5. **MORE THAN TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
6. MULTIPLE COMPARISONS AND MULTIPLE TESTING

Questions to answer

- Are samples comparable at baseline time
- Is blood pressure comparable between first and 12th measures

Boxplot *tad1*, by *group*



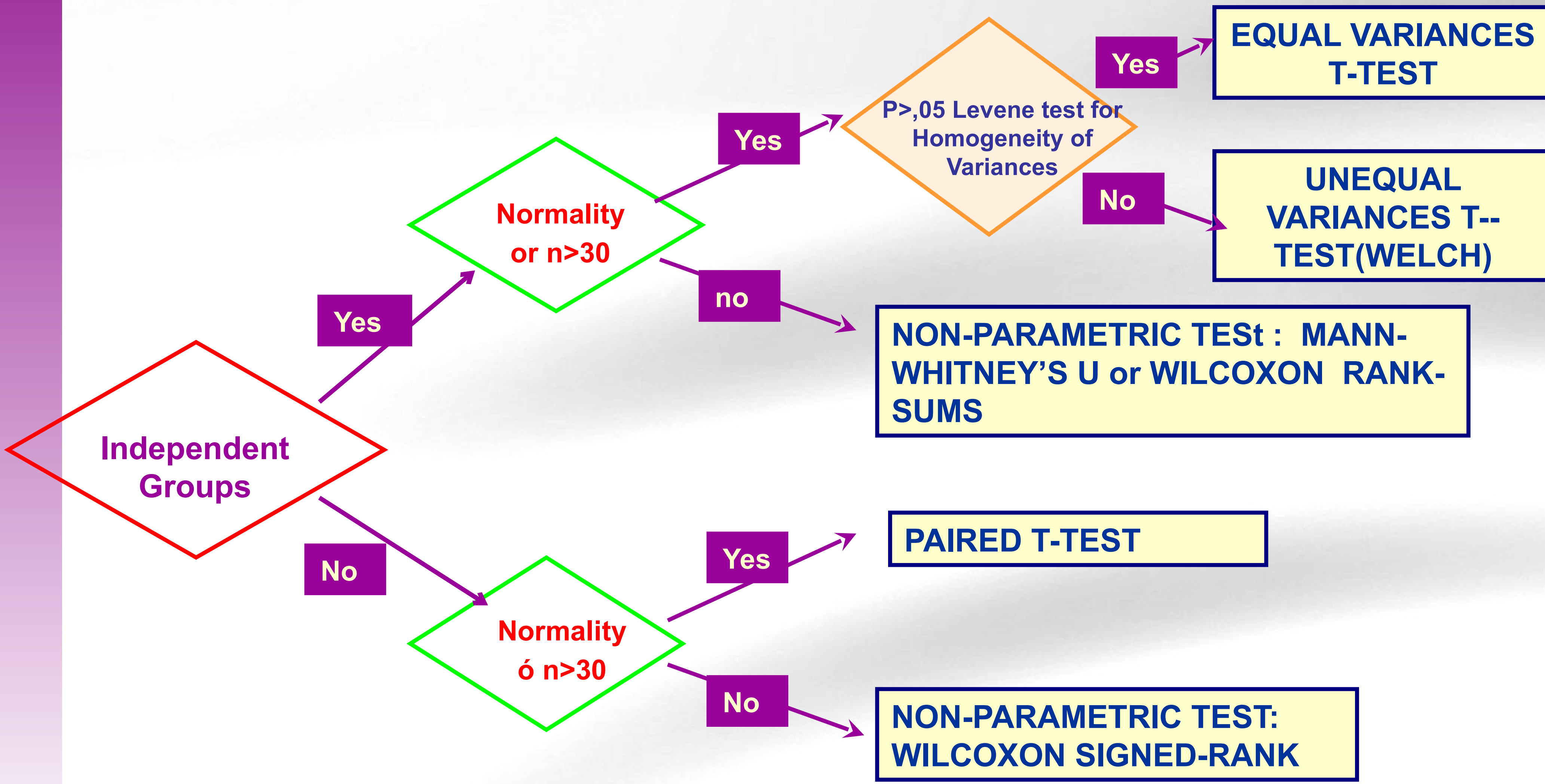
Compare a Quantitative variable in two groups

Null Hypothesis: There is not difference of the variable in two population or groups

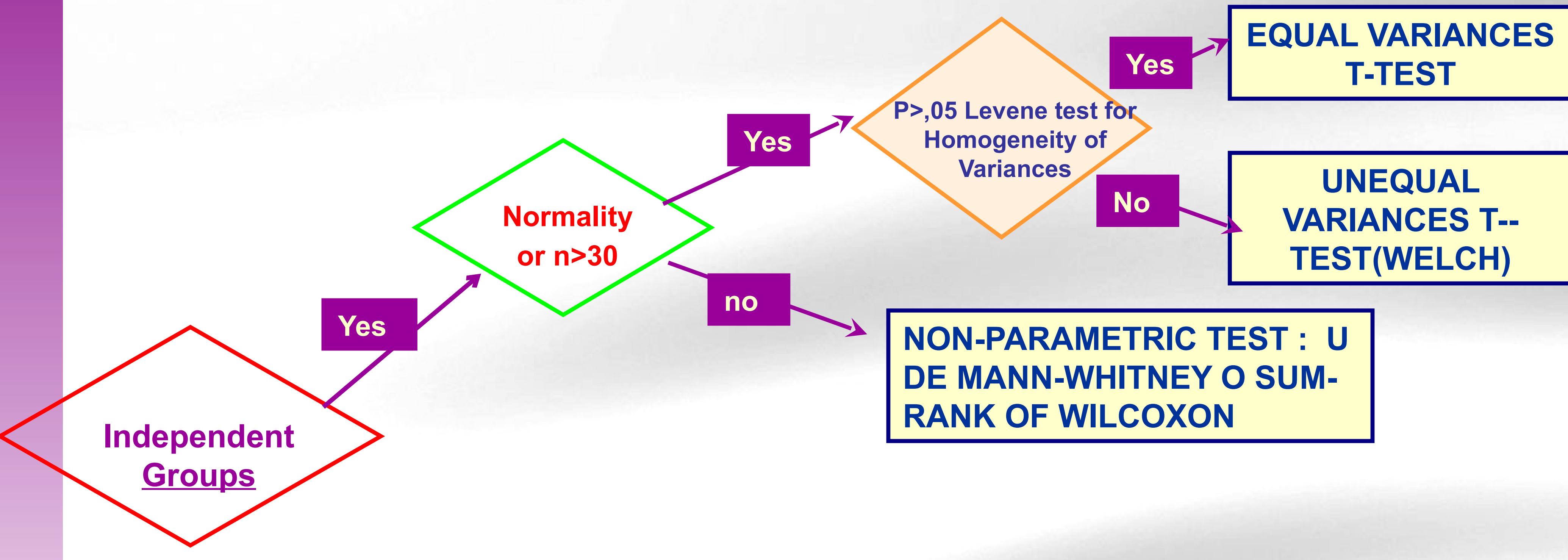
Samples have been generated

INDEPENDENT	DEPENDENT
Selected individuals in a group have nothing to do with selected individuals in the other group.	Each individual in a group has a correspondent in other group. These are ***paired data***.

Two sample tests



Two sample tests (1)



1. Data is normal (normality test) or sample size > 30.
2. Mean is a good summary statistic for this problem.
3. Test homogeneity of variances

Test of homogeneity of variances

```
> library(car)
> hta%>%
+   group_by(grupo) %>%
+   summarise(var = sd(tad1))
# A tibble: 2 × 2
  grupo   var
  <chr> <dbl>
1 A      10.6
2 B       9.79
> with(hta,leveneTest(tad1~factor(sexo),center="median"))
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  1.3506 0.2499
      58
```

We cannot reject H0: Variances cannot be considered to be distinct

t-Test if variances are equal

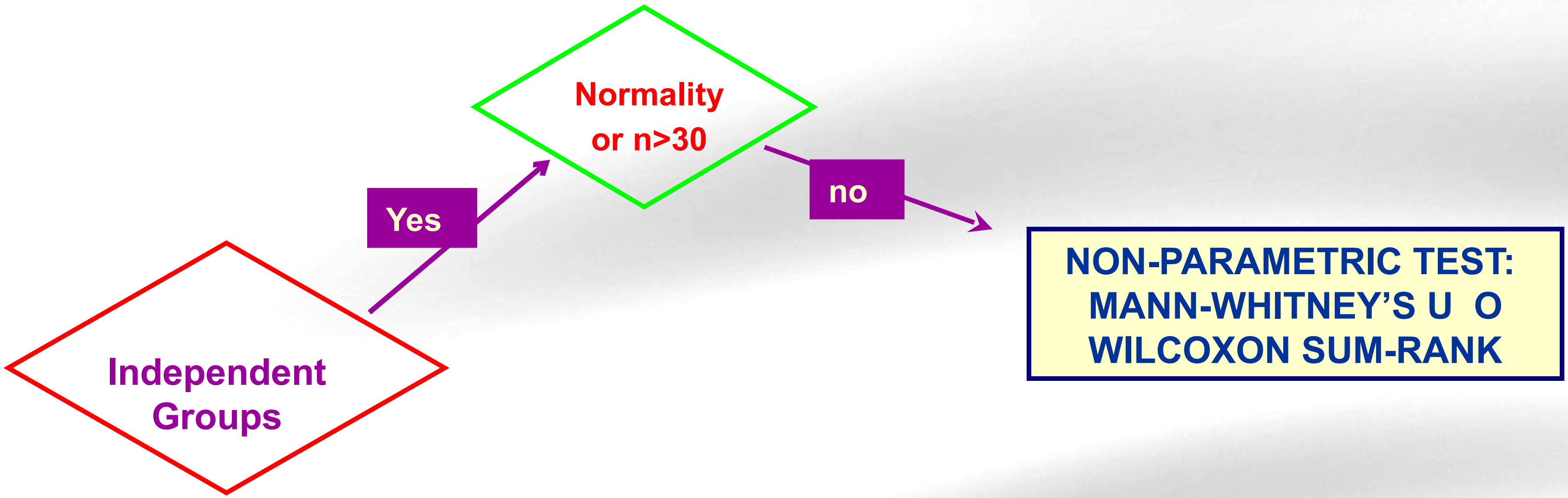
```
> with(hta,t.test(tad1~factor(grupo),var.equal=TRUE ))

Two Sample t-test

data:  tad1 by factor(grupo)
t = 0.15001, df = 58, p-value = 0.8813
alternative hypothesis: true difference in means between group A and group
B is not equal to 0
95 percent confidence interval:
 -4.874549  5.664316
sample estimates:
mean in group A mean in group B
       88.62069        88.22581
```

We cannot reject H0: Mean values are the same in group A and B (cannot be considered to be distinct)

Comparison between two <u>groups</u>, data non normal



Non parametric tests

- If data distribution is unknown or mean is not the best way to summarize data …
  - Non parametric test are not based on the usual parameters from a distribution, such as μ or σ^2.
  - Instead they may be based …
    - On order statistics, such as median or percentiles
  - They take into account the whole distribution.

Test based on ranks (Wilcoxon)

- Based on susbtituting original values by “ranks” in a joint sample
  - 12, 5, 14, 16, 3 ☾ ranks are: 3, 2, 4, 5, 1
- Ranks only depend on the position of each value in the ordered sample.
  - 120, 95, 121, 130, 3 have the same ranks as values in the first sample

- NP test are more robust than parametrics ones
- In the ideal situation where parametric tests are valid they are considered to be preferable.

Wilcoxon rank sums test

```
> hta%>%
+   group_by(sexo) %>%
+   summarise(median = median(tad1))
# A tibble: 2 × 2
  sexo   median
  <chr>   <dbl>
1 MUJER      90
2 VARON      90
```

```
> with(hta,wilcox.test(tad1~factor(grupo),
+ alternative='two.sided',
+                       exact=TRUE, correct=FALSE))

Wilcoxon rank sum test

data:  tad1 by factor(grupo)
W = 432, p-value = 0.7926
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(x = c(80, 85, 100, 90, 105, 110, 100, 100,  :
  cannot compute exact p-value with ties
```

Exercise 2

- Is TAD comparable at baseline time between Men and women?
  a) What is the Hypothesis that we want to test? Describe the null hypothesis and the alternative hypothesis.
  b) What test would be appropiate to answer the question?
  c) Compute and decide
  d) Apply a non-parametric test and compare the results

Questions to answer

- Are samples comparable at baseline time?
- Is blood pressure comparable between first and 12th measures?

Example Data

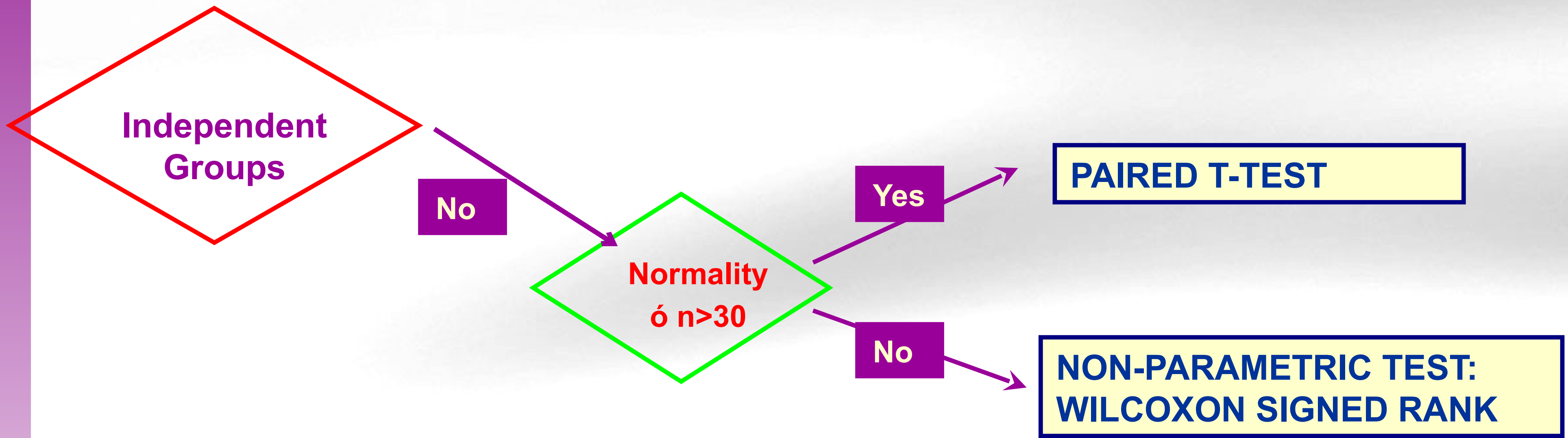
- A study was designed to compare two distinct hypertension control programs.
- 60 individuals with HTA were randomly assigned to either one or the other group (30 per group)
- Blood pressure was measured each month during a year

A	B	C	D	E	F	G	H	I
numero	sexo	grupo	tas1	tad1	tas2	tad2	tas3	tad3
1	VARON	B	150	100	150	90	170	
2	MUJER	B	160	90	170	90	160	
3	MUJER	B	150	90	110	90	115	
4	VARON	A	120	80	140	90	140	
5	MUJER	A	150	85	145	85	160	
6	MUJER	B	140	75	160	70	135	
7	MUJER	A	150	100	140	90	130	
8	VARON	A	160	90	170	90	170	
9	MUJER	A	145	105	170	95	140	
10	MUJER	A	210	110				
11	MUJER	A	170	100	170	90	170	
12	MUJER	B	140	90	140	90	100	

Syllabus

1. INTRODUCTION
2. NORMALITY TESTS
3. **ONE GROUP** COMPARISON
4. **TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
5. **TWO GROUPS** COMPARISON IN DEPENDENT SAMPLES
6. **MORE THAN TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
7. MULTIPLE COMPARISONS AND MULTIPLE TESTING

Two dependent groups



Questions to answer

- Are samples comparable at baseline time?
- Is blood pressure comparable between first and 12th measures?

Paired T-test

```r
with(hta,t.test(tad1,tad12,paired=TRUE))
```

```
##
##  Paired t-test
##
## data:  tad1 and tad12
## t = 1.8507, df = 51, p-value = 0.07001
## alternative hypothesis: true difference in means is not equal
## 95 percent confidence interval:
##  -0.2364274  5.8133505
## sample estimates:
## mean of the differences
##                2.788462
```

- P value is over 0.05

Paired Sign-Rank Wilcoxon Test

```
with(hta,wilcox.test(tad1,tad12,
     exact=TRUE, paired=TRUE))
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  tad1 and tad12
## V = 478.5, p-value = 0.05333
## alternative hypothesis: true location shift is not equal to 0
```

Exercise 3

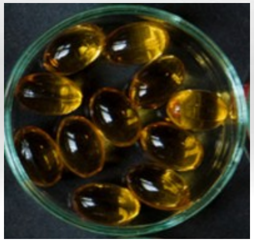
- Is sistolic blood pressure (TAS) comparable between first and 12th measures?

Syllabus

1. INTRODUCTION
2. NORMALITY TEST
3. **ONE GROUP** COMPARISON
4. **TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
5. **TWO GROUPS** COMPARISON IN DEPENDENT SAMPLES
6. **MORE THAN TWO GROUPS** COMPARISON IN INDEPENDENT SAMPLES
7. MULTIPLE COMPARISONS AND MULTIPLE TESTING

Question to answer

A pharmaceutical laboratory wants to test which of three drugs are better:



drug 1

drug 3

drug 2

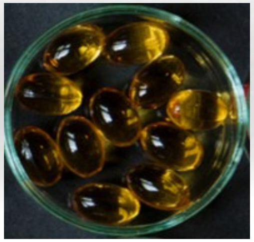
Question to answer

Could we use Student's t test? We will see with and example:

A pharmaceutical laboratory wants to test which of three drugs are better:

drug 1

drug 3



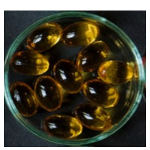
drug 2

To know which of the drugs is the best one, one could think to perform the following comparison using a t test:

Question to answer

Could we use Student's t test? We will see with and example:

Comparisons with separate t test would be:

			Chance of Type I error	Chance of Accept H0
	vs		α = 5%	$1-\alpha$ =95%
	vs		α = 5%	$1-\alpha$ =95%
	vs		α = 5%	$1-\alpha$ =95%

Question to answer

Could we use Student's t test? We will see with and example:

Comparisons with t test would be:

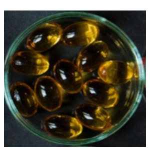
<u>Chance of Type I error</u> | <u>Chance of Accept H0</u>

 vs 

$\alpha = 5\%$ — $1-\alpha = 95\%$

 vs 

$\alpha \approx 10\%$ — $1-\alpha = 95\%*95\%$

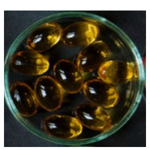
 vs 

$\alpha \approx 15\%$ — $1-\alpha = 95\%*95\%*95\%$

Question to answer

Could we use Student's t test? We will see with and example:

Comparisons with t test would be:

			Chance of Type I error	Chance of Accept H0
	vs		$\alpha = 5\%$	$1-\alpha = 95\%$
	vs		$\alpha \approx 10\%$	$1-\alpha = 95\%*95\%$
	vs		$\alpha \approx 15\%$	$1-\alpha = 95\%*95\%*95\%$

Would be easier not reject the null hypothesis when it was wrong (more false positives)

Analysis of the variance

Null Hypothesis

The means of all population are equal

$$H_0 \quad \mu_1 = \mu_2 = \ldots. \quad = \mu_k$$

Alternative Hypothesis

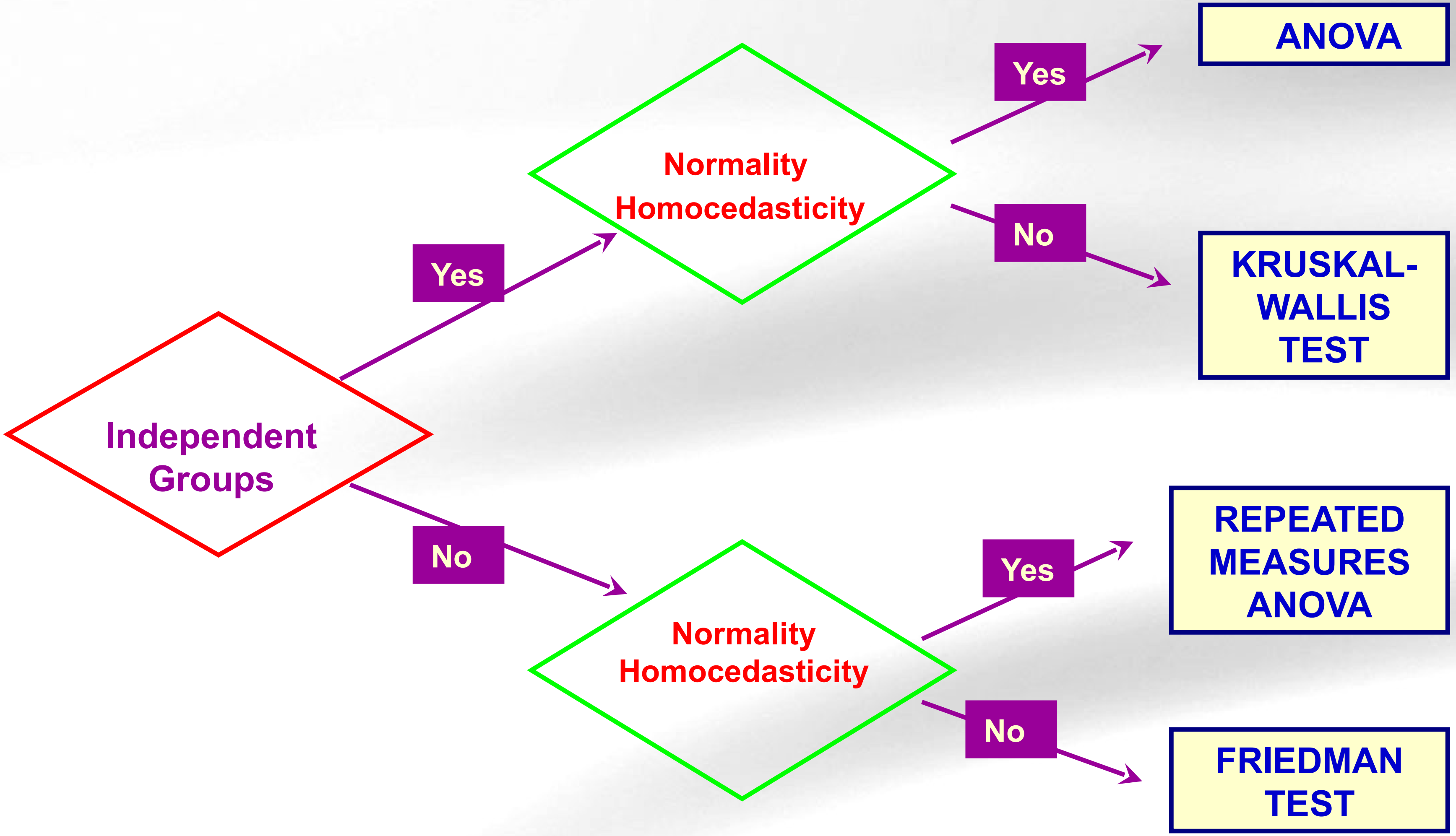
Not all the means are equal. At least there are two different means

$$H_a \, \exists \, i,j \quad \mu_i \neq \mu_j$$

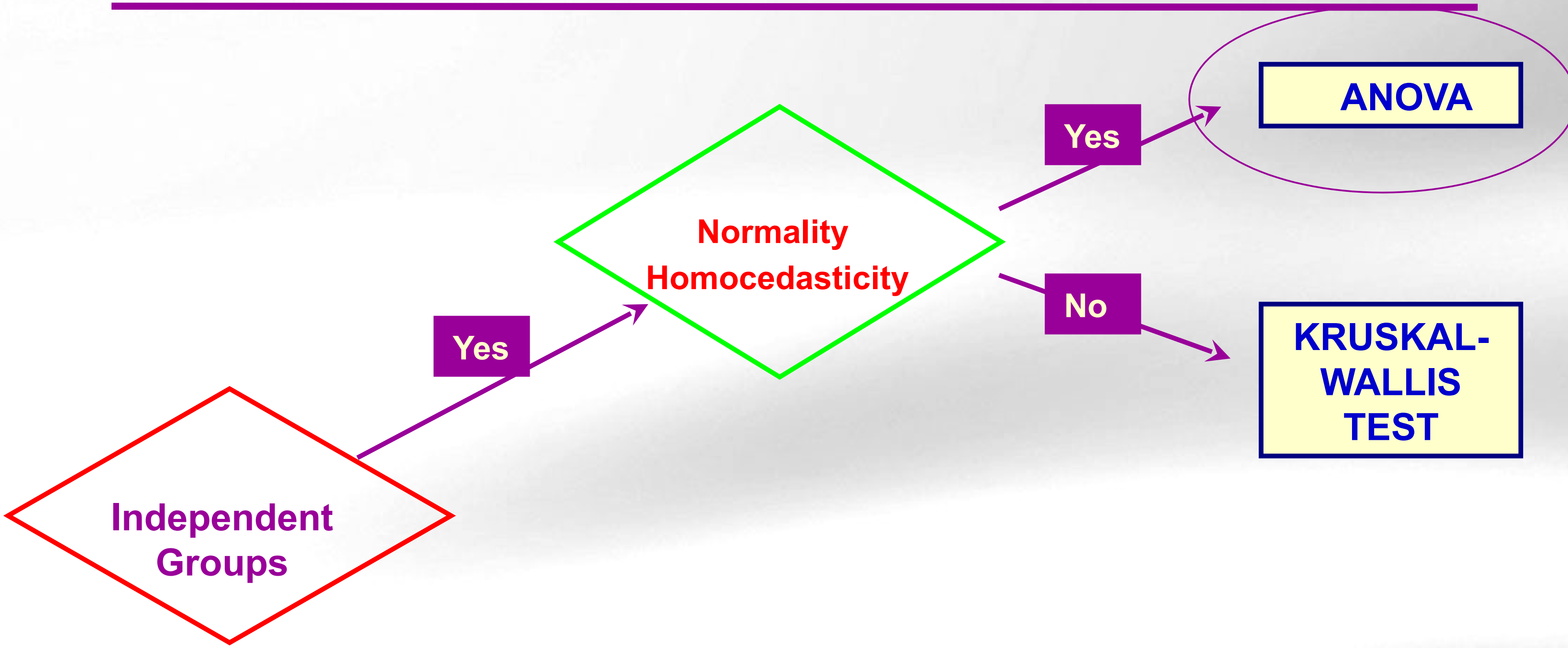
Post-hoc ANOVA tests

- If we wish to compare all means against all means he number of tests increases quickly (to compare all pairs of means if there are k groups (k*k-1)/2 tests are required).
- This is usually called **multiple comparisons** and common methods of adjustment are Tukey, Fisher HSD or Bonferroni.

Three or more groups



Three or more groups



Read diabetes data

```
library(readxl)
library(dplyr)
library(magrittr)
diabetes <- read_excel("datasets/diabetes.xls")
sapply(diabetes, class)
```

```
##    numpacie        mort    tempsviu        edat         bmi    edatdiag
##   "numeric" "character"   "numeric"   "numeric"   "numeric"   "numeric"
##       tabac         sbp         dbp         ecg         chd
## "character"   "numeric"   "numeric" "character" "character"
```

```
diabetes_factor <- diabetes %>%
  mutate_if(sapply(diabetes, is.character), as.factor) %>%
  select (-numpacie)

diabetes%>%
  group_by(ecg) %>%
  summarise( n=n(),
    mean = mean(edat),
            sd=sd(edat))
```

```
## # A tibble: 3 x 4
##   ecg          n  mean    sd
##   <chr>    <int> <dbl> <dbl>
## 1 Anormal     11  64.9  6.76
## 2 Frontera    27  53.8 11.4
## 3 Normal     111  50.5 11.5
```

ANOVA

```
anova<-aov(edat~ecg,data=diabetes)
summary(anova)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## ecg           2   2166  1083.0   8.619 0.00029 ***
## Residuals   146  18347   125.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.
```

Multicomparison

```
library(multcomp)
tuk <- glht(anova, linfct = mcp(ecg = "Tukey"))

  print(summary(tuk)) # pairwise tests
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = edat ~ ecg, data = diabetes_factor)
##
## Linear Hypotheses:
##                           Estimate Std. Error t value Pr(>|t|)
## Frontera - Anormal == 0  -11.094      4.010  -2.767 0.016496 *
## Normal - Anormal == 0    -14.405      3.543  -4.065 0.000217 ***
## Normal - Frontera == 0    -3.310      2.405  -1.376 0.345732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

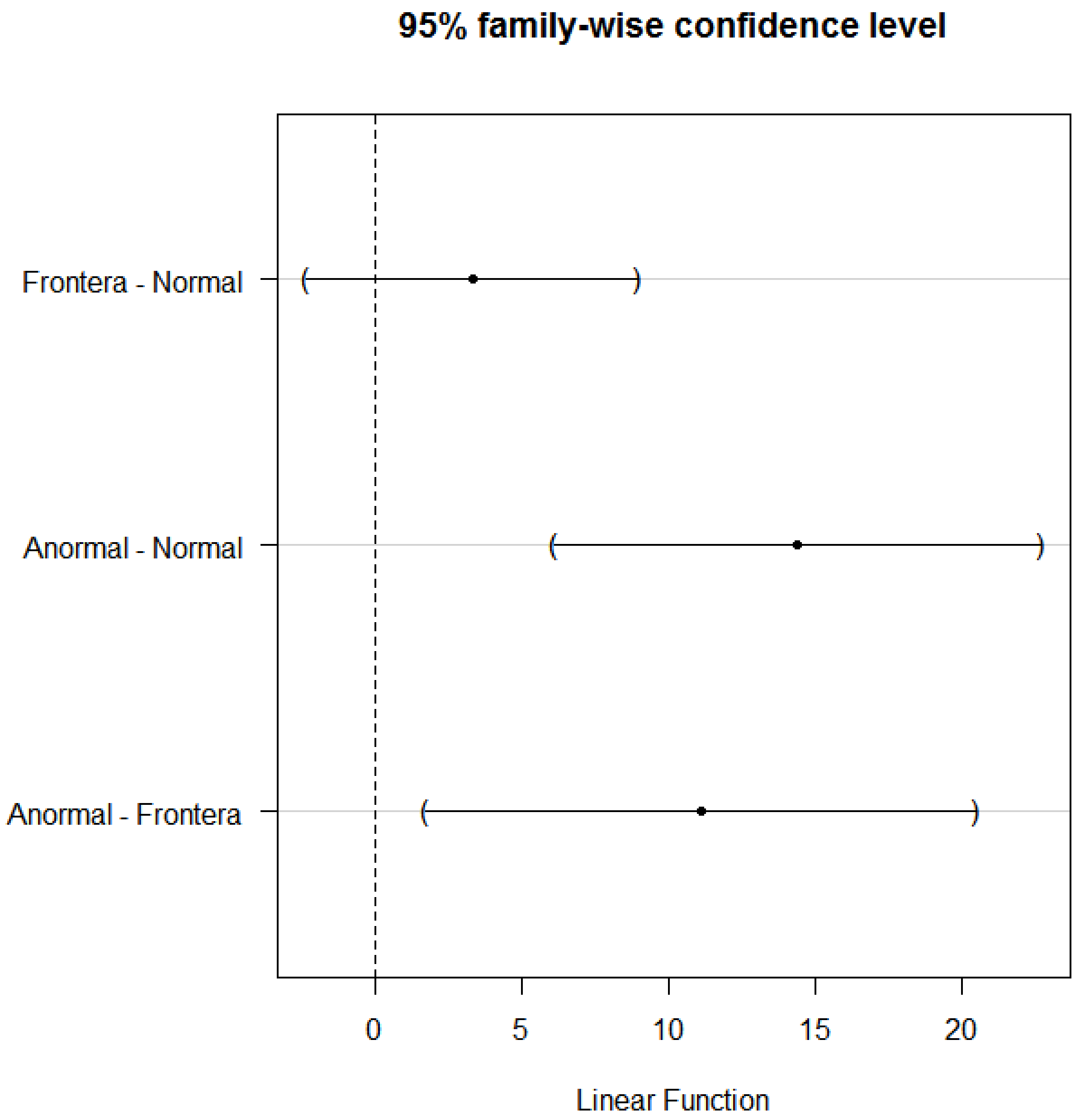
Multicomparison

```r
print(confint(tuk, level=0.95)) # confidence intervals
```

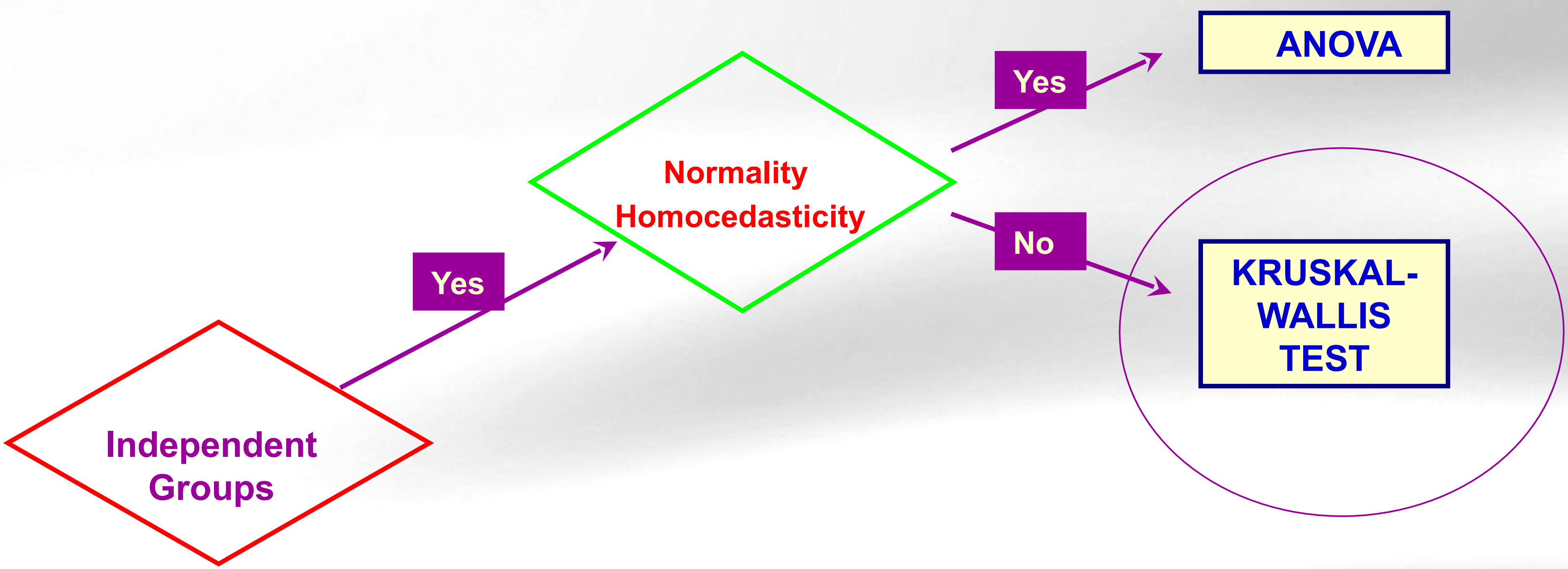
```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = edat ~ ecg, data = diabetes_factor)
##
## Quantile = 2.3459
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                            Estimate lwr      upr
## Frontera - Anormal == 0 -11.0943 -20.5009  -1.6876
## Normal - Anormal == 0   -14.4046 -22.7173  -6.0919
## Normal - Frontera == 0   -[illegible].3103  -8.9534   2.3328
```

Multicomparison

```
plot(confint(tuk))
```



Three or more groups



Kruskal-Wallis Test

```
diabetes_factor%>%
  group_by(ecg) %>%
  summarise(median = median(edat))
```

```
## # A tibble: 3 x 2
##   ecg       median
##   <fct>      <dbl>
## 1 Anormal       64
## 2 Frontera      53
## 3 Normal        49
```

```
kruskal.test(edat~ecg,data=diabetes_factor)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  edat by ecg
## Kruskal-Wallis chi-squared = 17.483, df = 2, p-value = 0.0001
```

Dunn Test for multiple comparison

```
library(dunn.test)
with(diabetes_factor,dunn.test(edat,ecg,method="bonferroni"))
```

```
##   Kruskal-Wallis rank sum test
##
## data: edat and ecg
## Kruskal-Wallis chi-squared = 17.4826, df = 2, p-value = 0
##
##
##                          Comparison of edat by ecg
##                                (Bonferroni)
## Col Mean-|
## Row Mean |    Anormal   Frontera
## ---------+----------------------
## Frontera |   2.721182
##          |    0.0098*
##          |
##   Normal |   4.075469   1.467464
##          |    0.0001*     0.2134
##
```

Exercise 5

- Are there differences between systolic pressure (*sbp*) and *ECG* in diabetic patients?
  a) For each variable, perform some descriptives and check normality assumptions
  b) Set the hypothesis contrast and perform the test
  c) If needed, apply a post-hoc test for multiple comparisons to test which groups are different.
  d) Extra: Analyze the relation between the follow-up time (*tempsviu*) and the *ECG*.

Exercise 6

Indica que tipo de análisis o que pruebas estadísticas utilizarías en cada uno de los casos e indica también si es necesario hacer algún tipo de prueba adicional para llevar a cabo el análisis. Formula la hipótesis de trabajo de acuerdo con las hipótesis seleccionadas.

e) Se estudia un grupo de 33 pacientes afectados de Carcinoma hepatocelular, un grupo de 22 afectados únicamente de cirrosis y un grupo control de 31 donantes de sangre. Se determina la actividad celular NK y el número de células CK en los tres grupos. (Nota: la media de actividad celular es de 39 unidades líticas /10ˆ7 de linfocitos y la mediana es de 28 y la media del número de células es de 178 y la mediana de 163)

Vall d'Hebron
Institut de Recerca

UNITAT
D'ESTADÍSTICA I
BIOINFORMÀTICA

Questions to answer

- Is diastolic (min) tension above 90, “on average”, at the beginning of the study.
- Are samples “comparable” at baseline?
  - In Age? Sex%? Sist? Diast?
- Has there been a change in BP between month 1 (first measure) and month 12?