

Hypothesis Testing . Quantitative Variables

Curs d'Estadística Bàsica per a la Recerca Biomèdica

UEB – VHIR

Miriam Mota, Santiago Pérez-Hoyos and Alex Sánchez

miriam.mota@vhir.org santi.perezhoyos@vhir.org

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

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	

Questions to solve

- Are samples “comparable” at baseline?
- Has there been a change in BP between month 1 (first measure) and month 12?

Syllabus

1. INTRODUCTION

2. NORMALITY TEST

3. ONE GROUP COMPARISON

**4. TWO GROUPS COMPARISON IN INDEPENDENT
SAMPLES**

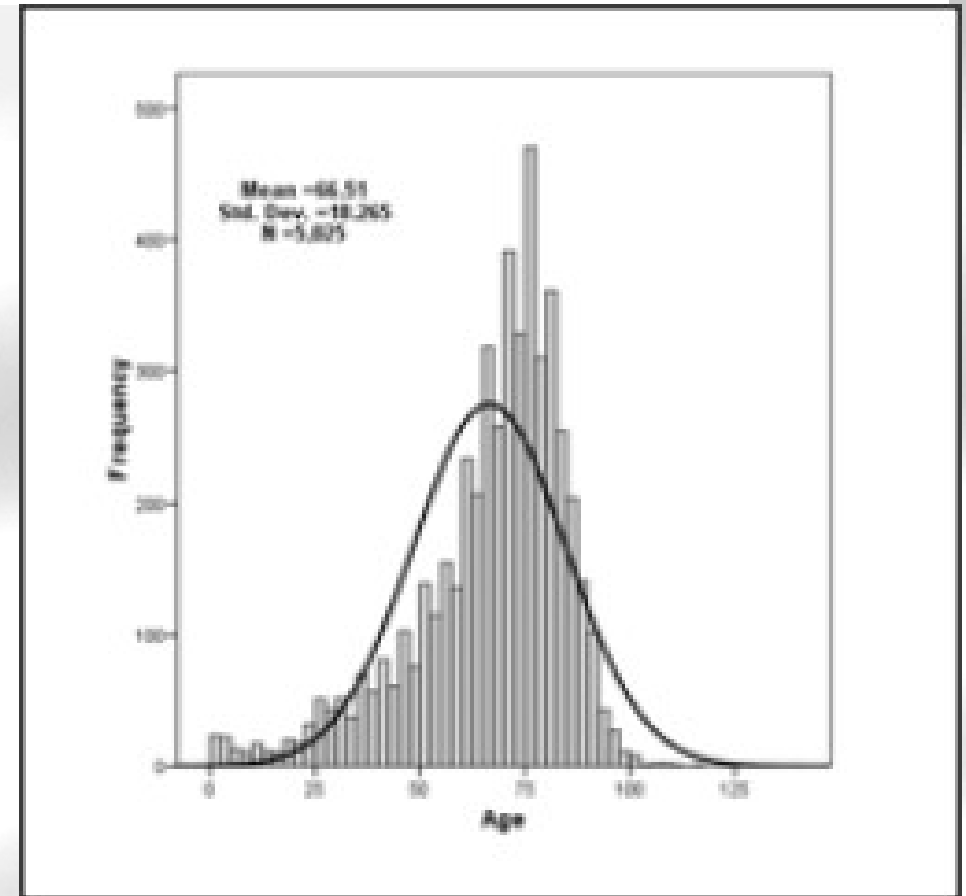
**5. TWO GROUPS COMPARISON IN DEPENDENT
SAMPLES**

**6. MORE THAN TWO GROUPS COMPARISON IN
INDEPENDENT SAMPLES**

7. MULTIPLE COMPARISONS AND MULTIPLE TESTING

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?



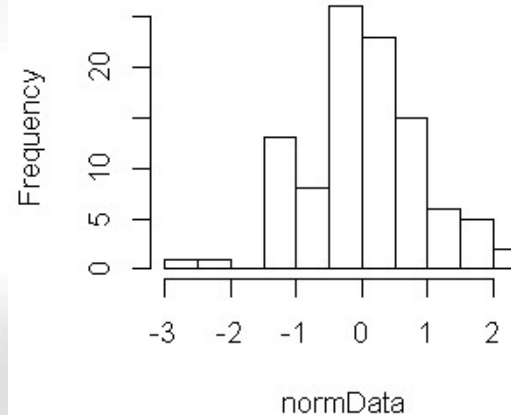
Testing normality

- 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

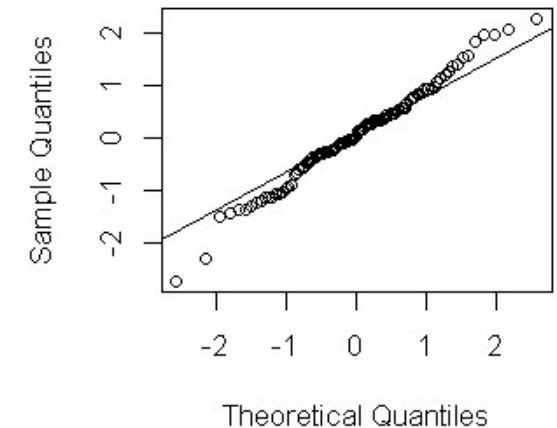
Histograms and QQ-plots

- 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

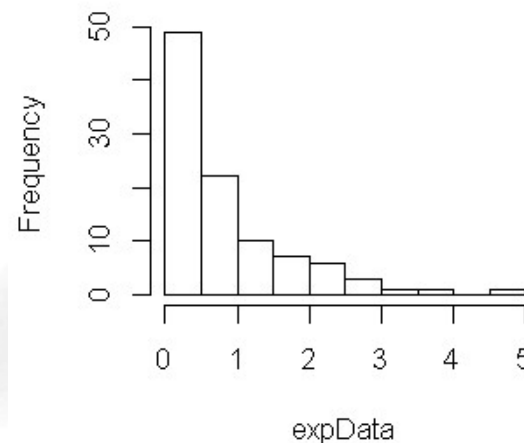
Normal data



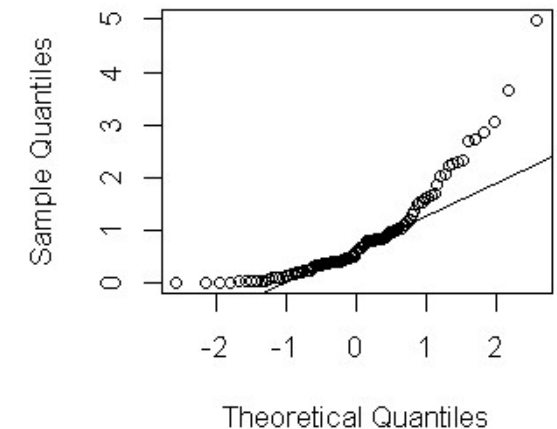
Normal Q-Q Plot



Exponential data



Normal Q-Q Plot

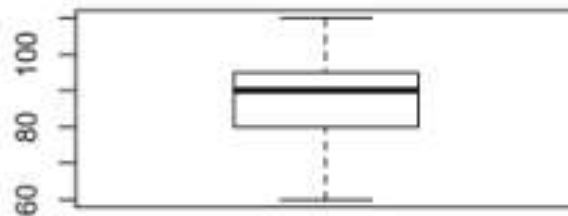
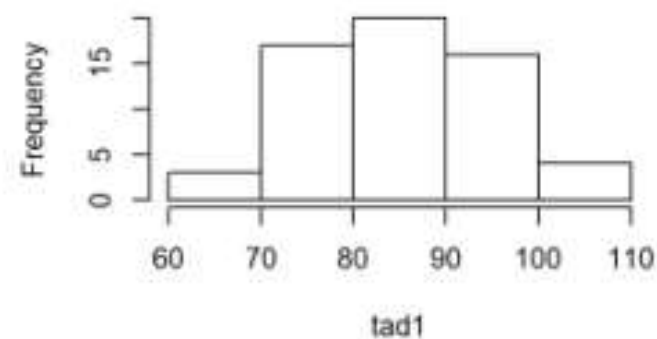
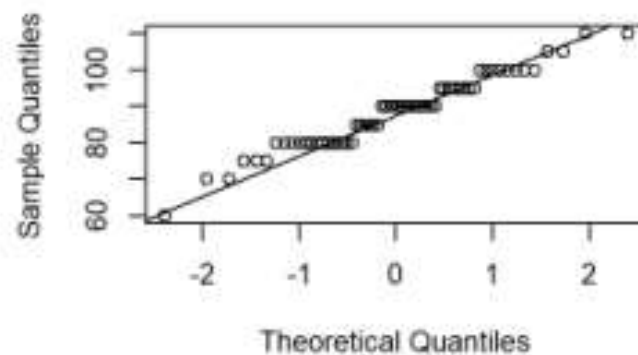



```
oldpar<-par(mfrow=c(1,1)) # Guarda los parámetros para el dibujo
par(mfrow=c(2,2)) # Dibuja cuatro gráficos por grafico
with(hta, boxplot(tad1, main="Box-plot") )

with(hta, hist(tad1) )

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

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

Box-plot**Histogram of tad1****Normal QQplot**

- 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

```
with(hta, shapiro.test(tas1) ) # Shapiro Wilk test
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  tas1  
## W = 0.91869, p-value = 0.0006796
```

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 in Hypertensive 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
```


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

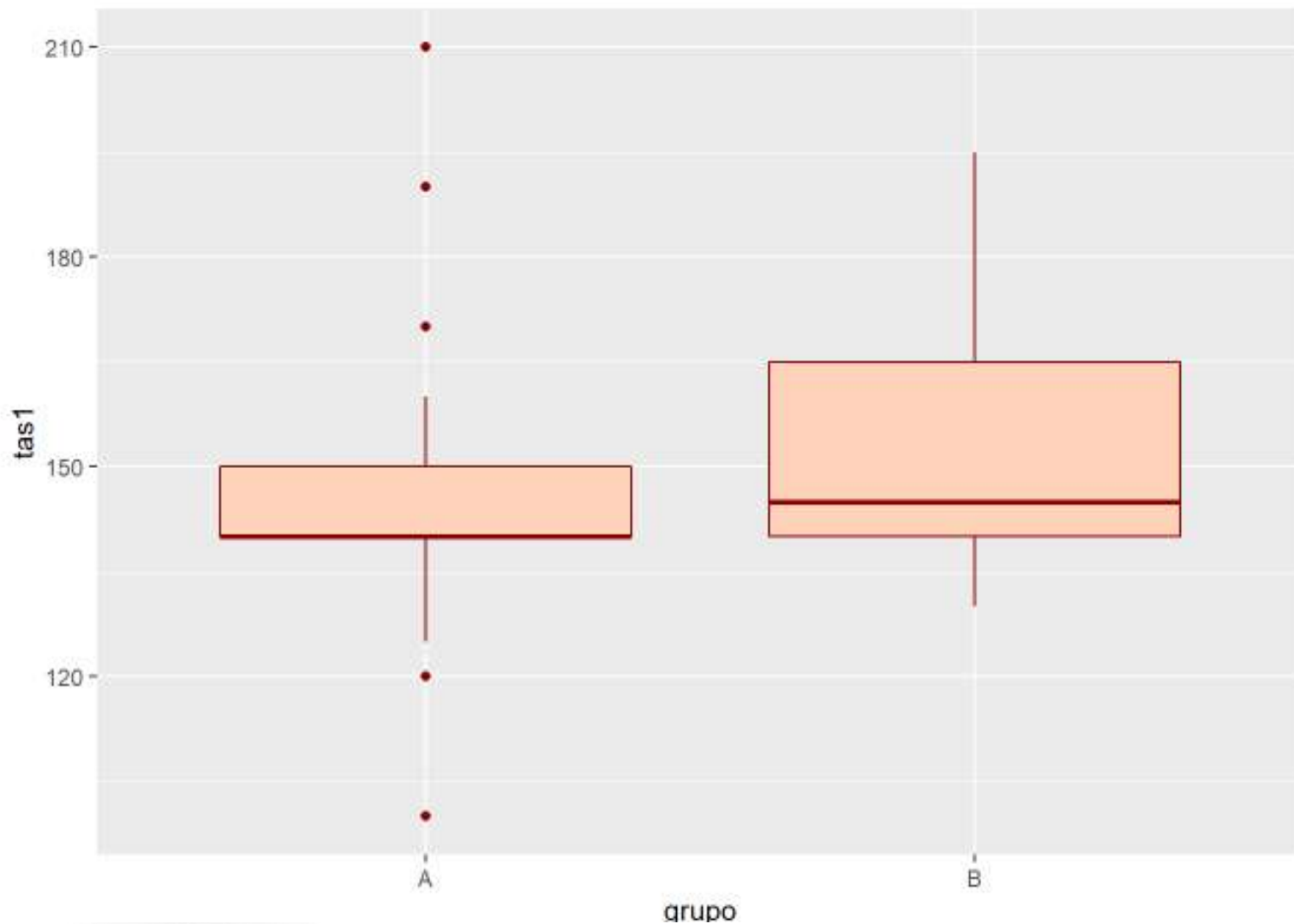
6. MULTIPLE COMPARISONS AND MULTIPLE TESTING

Questions to answer

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

Boxplot TAD, by group

```
#Plot the data
bp <- ggplot(hta, aes(x = grupo, y = tas1)) +
  geom_boxplot(fill = "#ffd3b9", color = "darkred")
bp
```



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



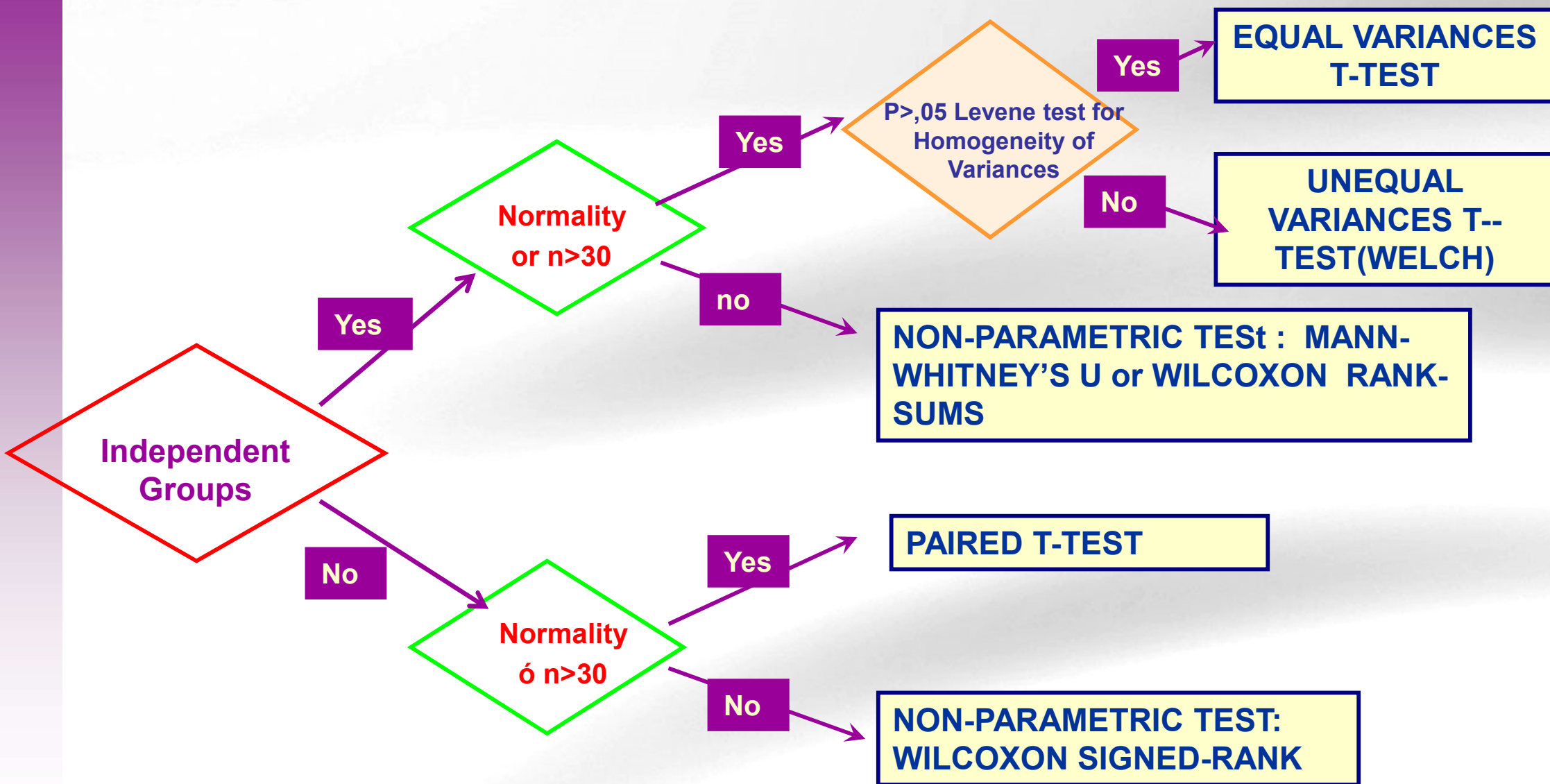
INDEPENDENT

Selected individuals in a group have nothing to do with selected individuals in the other group.

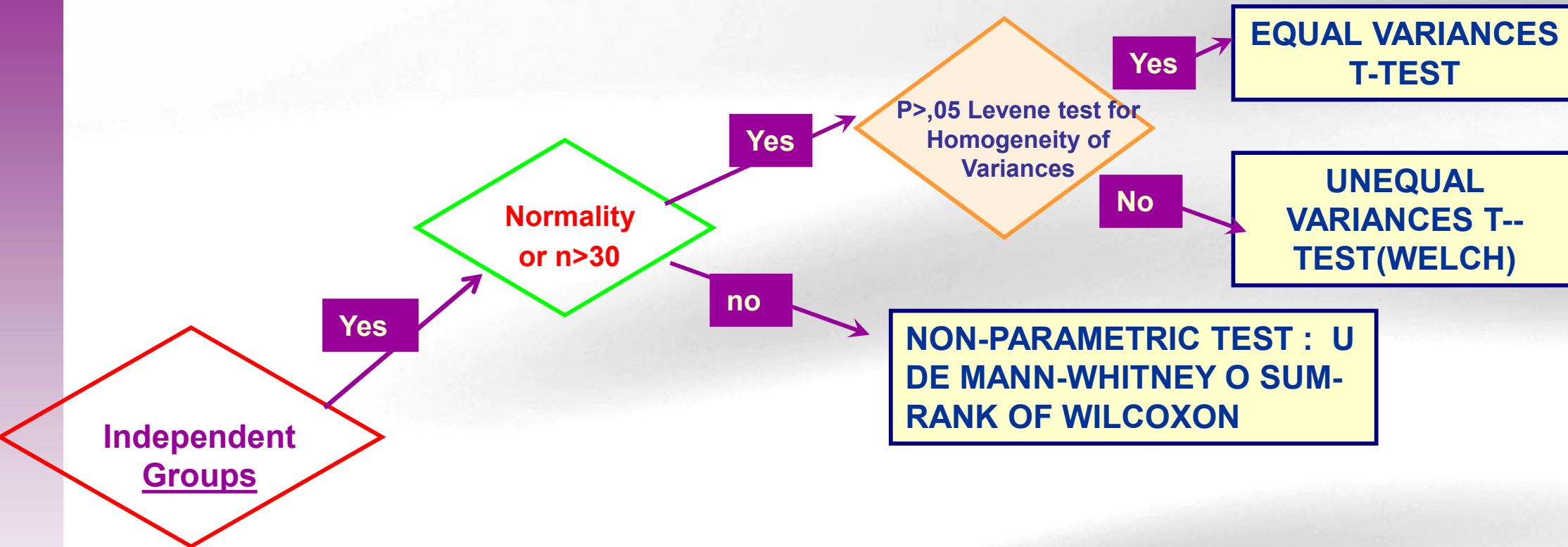
DEPENDENT

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

Homogeneity variance Test

```
library(car)
hta%>%
  group_by(sexo) %>%
  summarize(var = sd(tad1))
```

```
## # A tibble: 2 x 2
##   sexo    var
##   <chr> <dbl>
## 1 MUJER  9.08
## 2 VARON 11.8
```

```
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
```

- p value is over 0.05
- We can assume homogeneity of variances



T test when variances are equal

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

##
##  Two Sample t-test
##
## data:  tad1 by factor(sexo)
## t = 0.35427, df = 58, p-value = 0.7244
## alternative hypothesis: true difference in means is not equal
## 95 percent confidence interval:
##  -4.453505  6.368899
## sample estimates:
## mean in group MUJER mean in group VARON
##           88.78378           87.82609
```

- Type I Error is over than 0.05
- We cannot reject mean equality

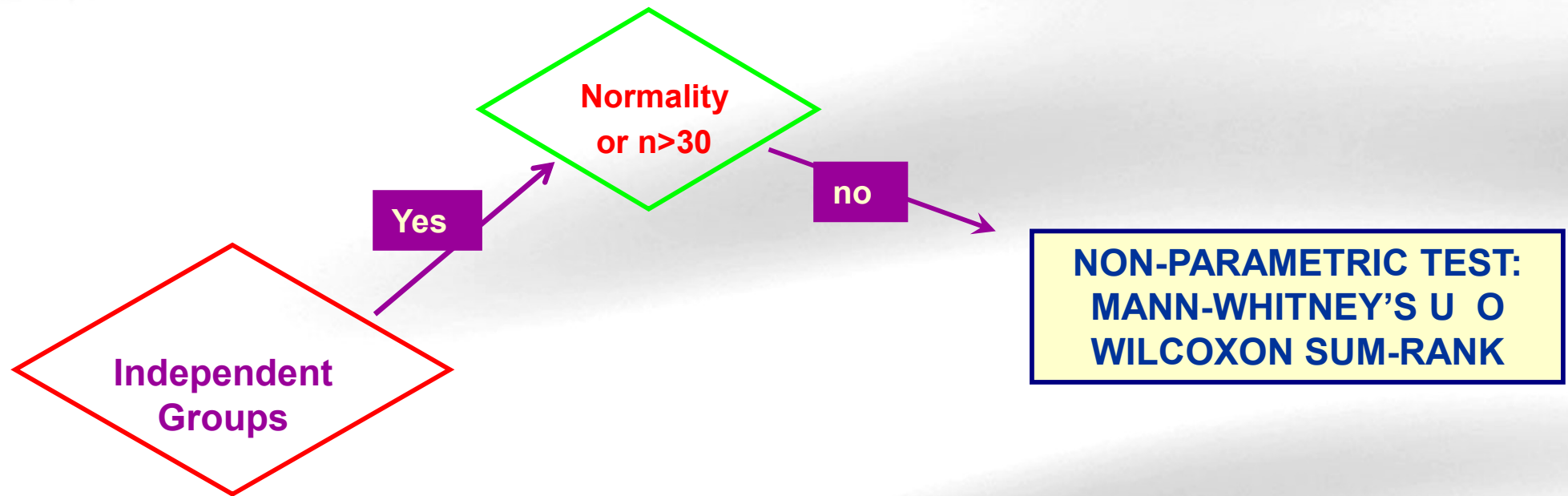
T test when variances are unequal

```
with(hta,t.test(tad1~factor(sexo),var.equal=FALSE ))

##
##  Welch Two Sample t-test
##
## data:  tad1 by factor(sexo)
## t = 0.33362, df = 38.144, p-value = 0.7405
## alternative hypothesis: true difference in means is not equal
## 95 percent confidence interval:
##  -4.852834  6.768228
## sample estimates:
## mean in group MUJER mean in group VARON
##           88.78378           87.82609
```

- Same conclusions as before
- Test is also known as Welch test

Two groups, 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 substituting 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.

U Mann-Whitney or Sum Rank non parametric test

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

```
##
## Wilcoxon rank sum test
##
## data: tad1 by factor(sexo)
## W = 434, p-value = 0.8955
## alternative hypothesis: true location shift is not equal to 0
```

```
hta%>%
  group_by(sexo) %>%
  summarize(median = median(tad1))
```

```
## # A tibble: 2 x 2
##   sexo median
##   <chr>   <dbl>
## 1 MUJER     90
## 2 VARON     90
```

- Null Hypothesis cannot be rejected

Questions to answer

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

Syllabus

1. INTRODUCTION

2. NORMALITY TESTS

3. ONE GROUP COMPARISON

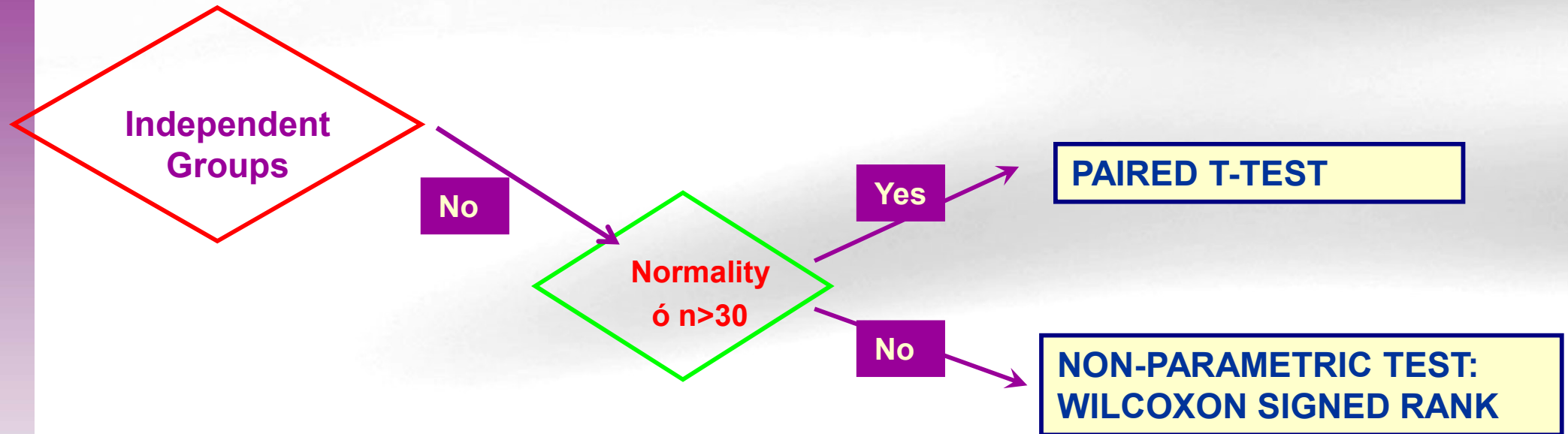
**4. TWO GROUPS COMPARISON IN INDEPENDENT
SAMPLES**

**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



Paired T-test

```
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
```


Syllabus

1. INTRODUCTION

2. NORMALITY TEST

3. ONE GROUP COMPARISON

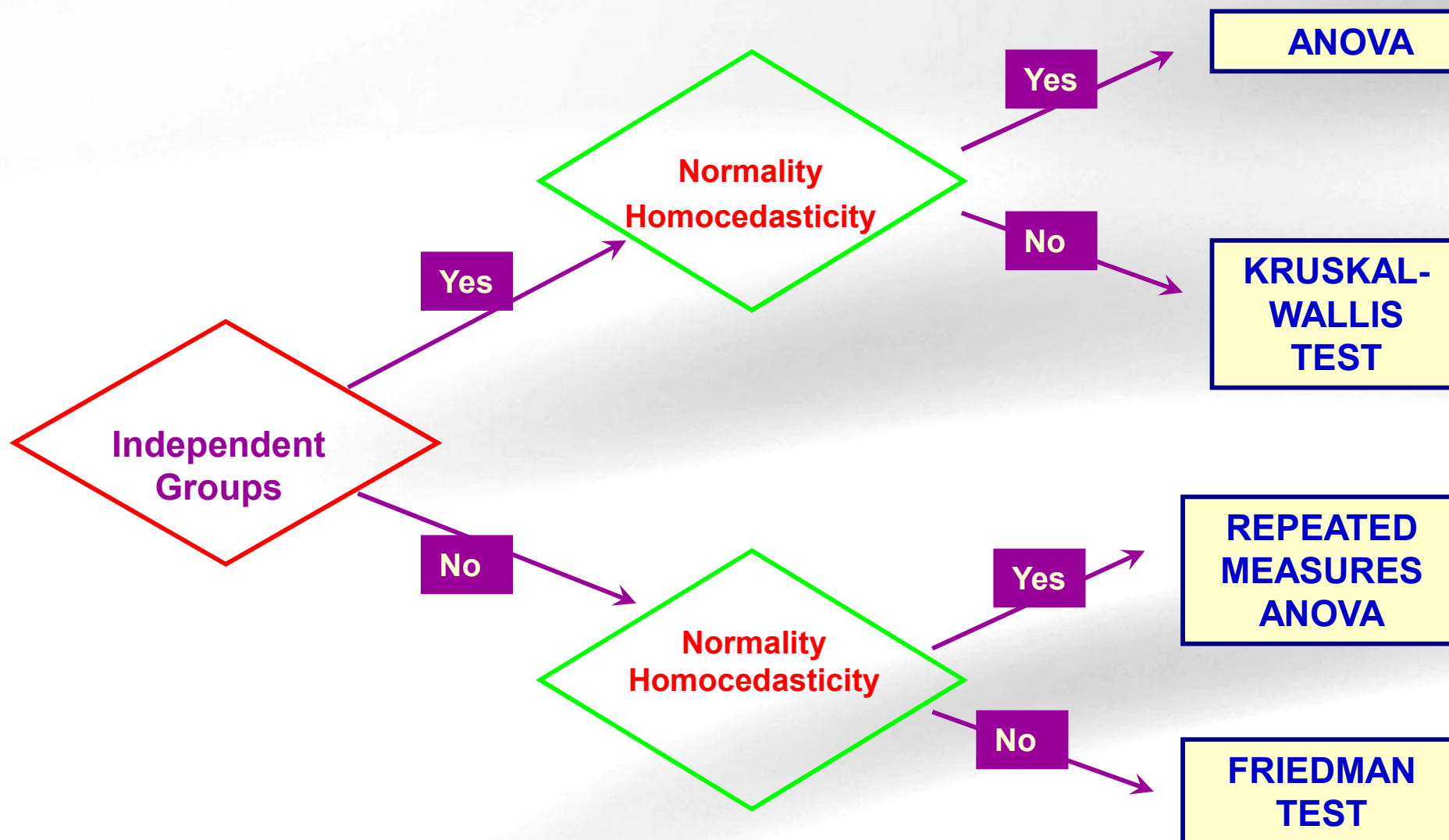
4. TWO GROUPS COMPARISON IN INDEPENDENT
SAMPLES

5. TWO GROUPS COMPARISON IN DEPENDENT
SAMPLES

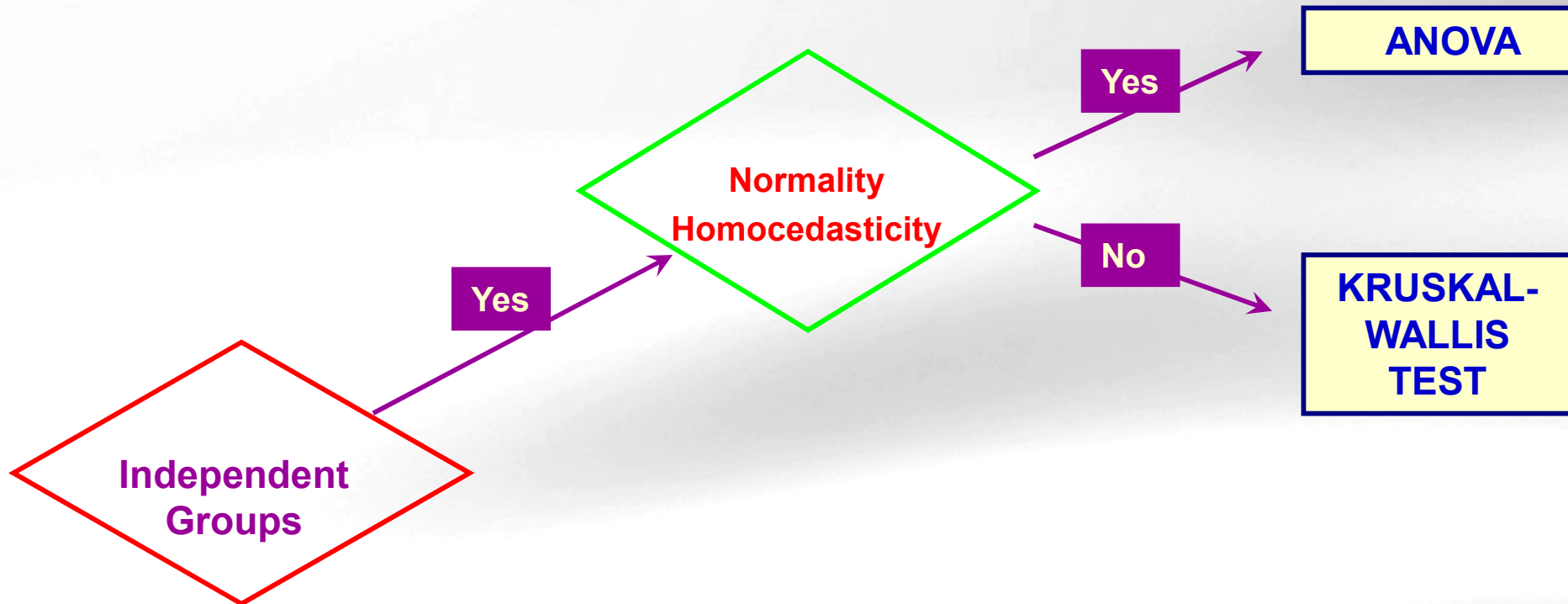
6. MORE THAN TWO GROUPS COMPARISON IN
INDEPENDENT SAMPLES

7. MULTIPLE COMPARISONS AND MULTIPLE TESTING

Three or more groups



Three or more groups



Analysis of the variance

Null Hypothesis

The means of all population are equal

$$H_0 \quad \mu_1 = \mu_2 = \dots = \mu_k$$

Alternative Hypothesis

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

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

Why Analysis of the variance

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

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



drug 1



drug 2









drug 3

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

Why Analysis of the variance

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

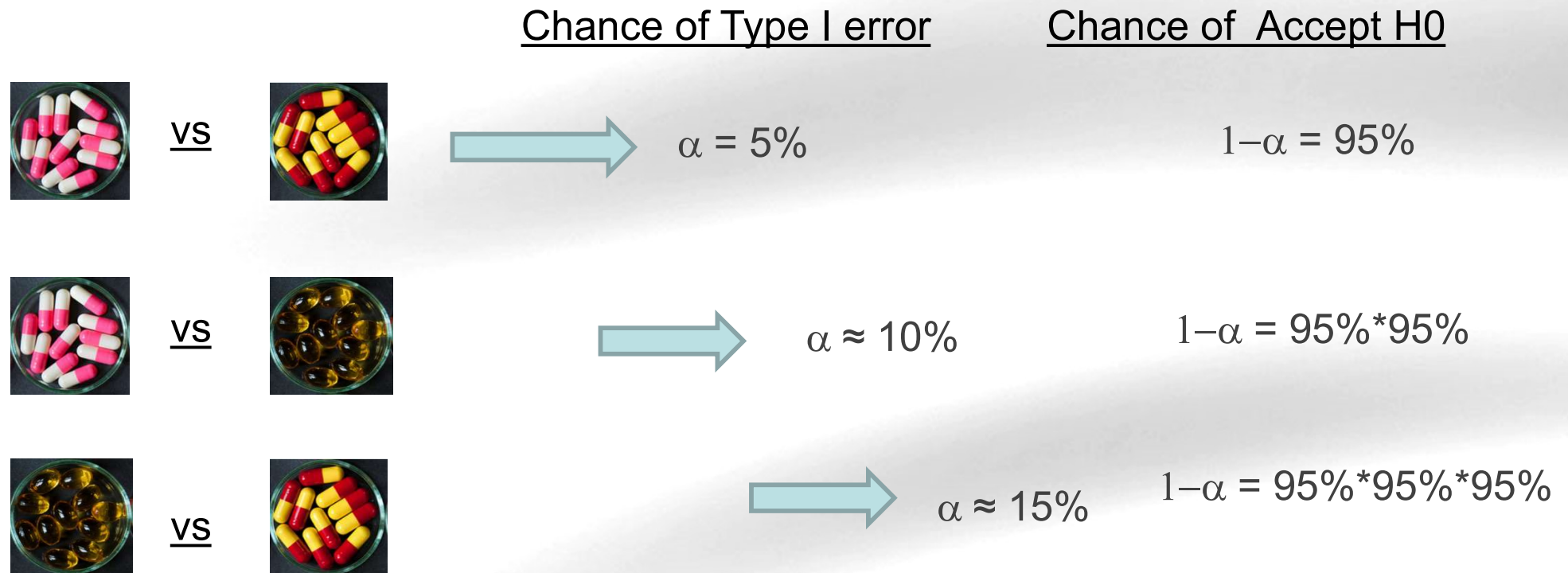
Comparisons with separate t test would be:

		<u>Chance of Type I error</u>	<u>Chance of Accept H0</u>
	<u>vs</u> 	$\alpha = 5\%$	$1-\alpha = 95\%$
	<u>vs</u> 	$\alpha = 5\%$	$1-\alpha = 95\%$
	<u>vs</u> 	$\alpha = 5\%$	$1-\alpha = 95\%$

Why Analysis of the variance

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

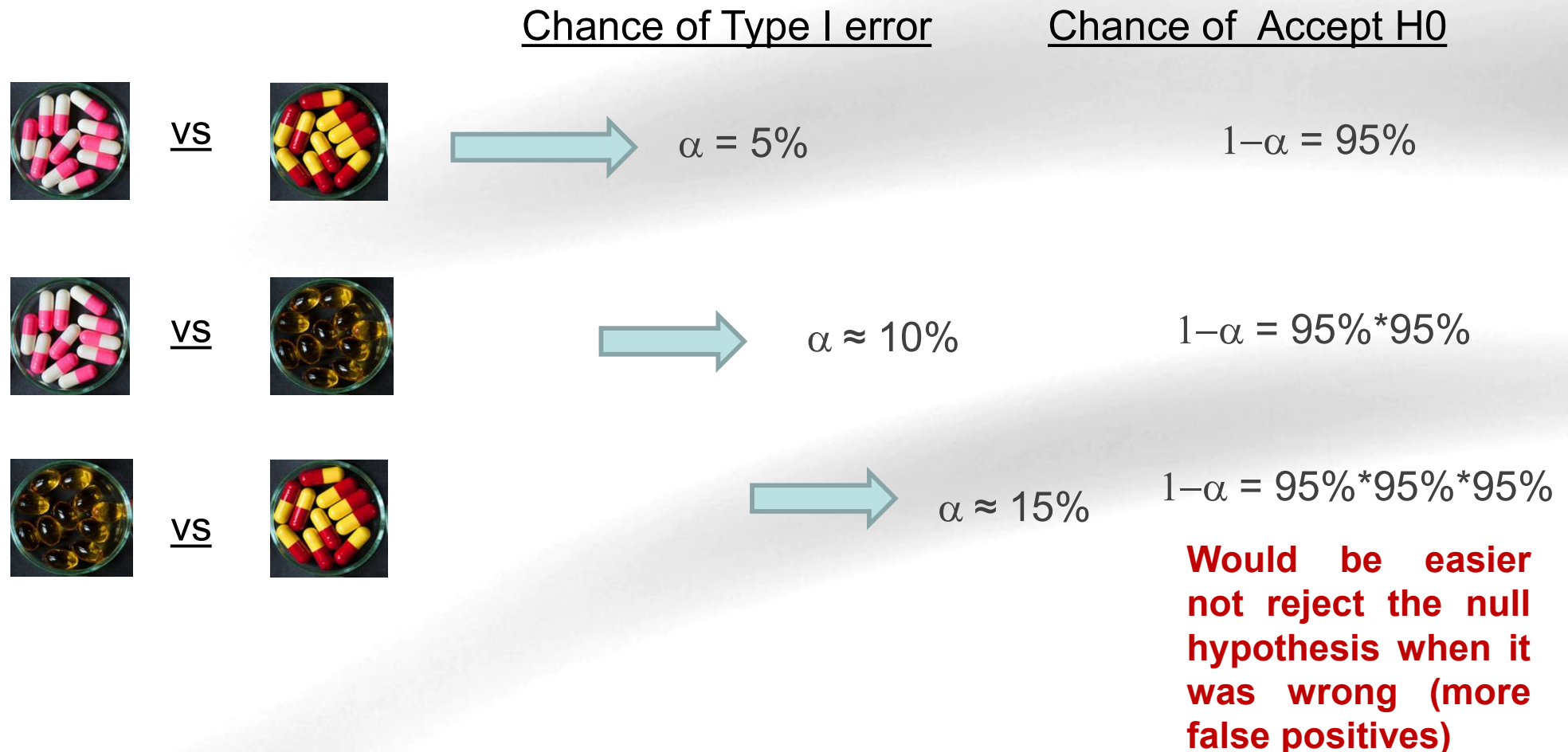
Comparisons with t test would be:



Why Analysis of the variance

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

Comparisons with t test would be:



4. Introduction to ANOVA

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

Comparisons with t test would be:

Chance of Type I error



vs



$$\alpha = 5\%$$



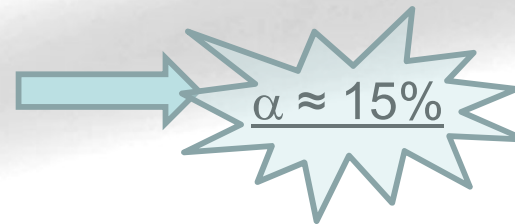
vs



$$\alpha \approx 10\%$$



vs



$$\alpha \approx 15\%$$

Would be easier to
accept the null
hypothesis when it
was wrong (more
false positives)

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_factor %>%
  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<-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.
```

```
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)
```

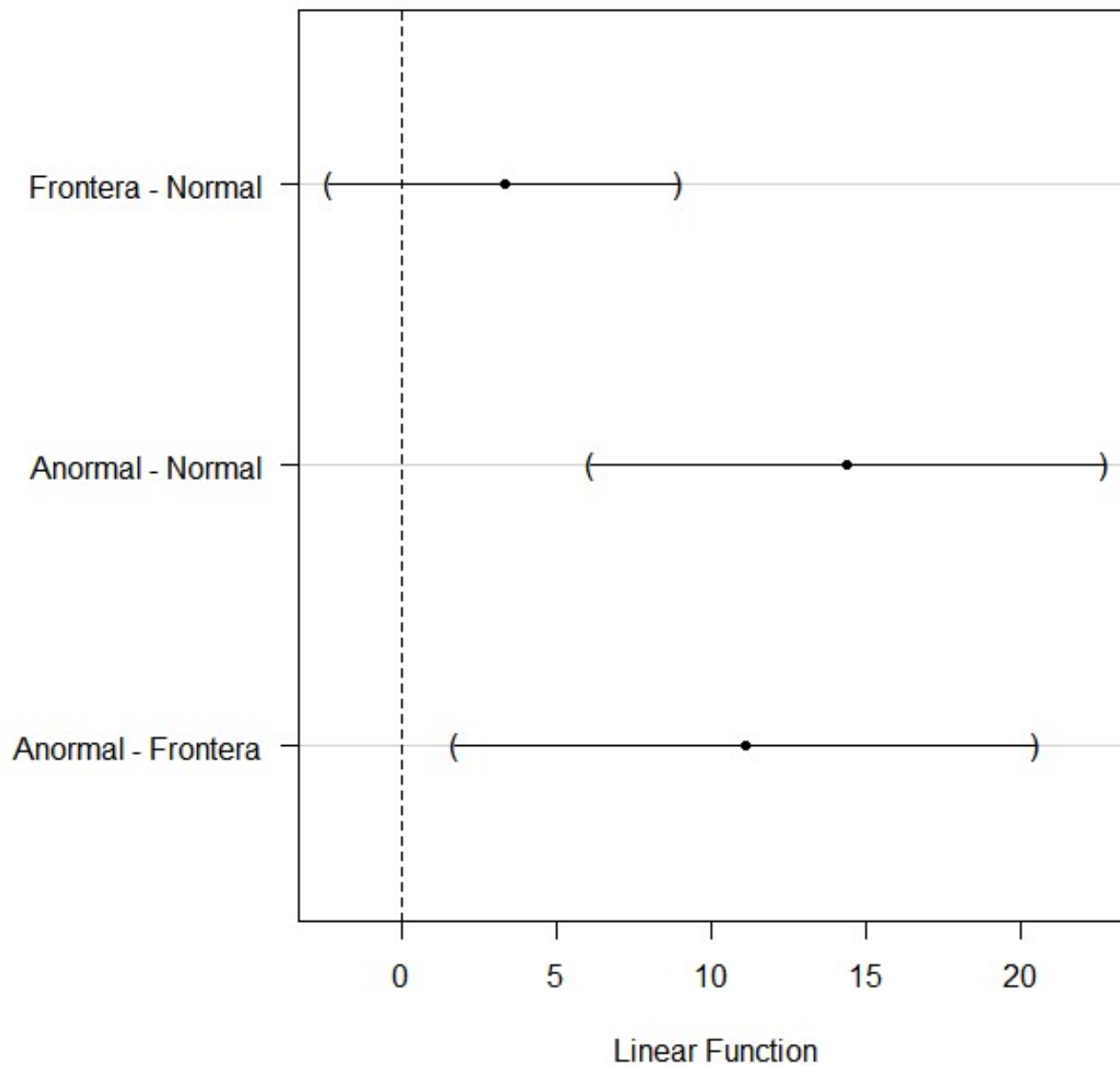



```
print(confint(tuk, level=0.95)) # confidence intervals
```

```
##  
## 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  -3.3103  -8.9534   2.3328
```

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

95% family-wise confidence level



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
##           |
##           |      \
```

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