

## FAEDE\_Sesión 3

JM

2023-11-09

```
setwd("D:/FAEDE")  
load("D:/FAEDE/master.RData")
```

*## Solución de ejemplos propuestos en las diapositivas de esta sesión*

*#Ejemplo 3.2 Análogamente, para el gen Gdf5, situado en la columna 3396, la hipótesis nula de igualdad de varianzas y la hipótesis nula de igualdad de medias entre los pacientes ALL y AML puede ser testadas.*

Antes del test t de comparación de medias para muestras independientes

```
var.test(leukemia$x.1040~leukemia$Y)  
  
##  
## F test to compare two variances  
##  
## data: leukemia$x.1040 by leukemia$Y  
## F = 0.90082, num df = 46, denom df = 24, p-value = 0.7413  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.4248098 1.7633269  
## sample estimates:  
## ratio of variances  
## 0.9008223
```

```
t.test(leukemia$x.1040~ leukemia$Y,var.equal=T)  
  
##  
## Two Sample t-test  
##  
## data: leukemia$x.1040 by leukemia$Y  
## t = 3.118, df = 70, p-value = 0.002642  
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0  
## 95 percent confidence interval:  
## 0.0880387 0.4005939  
## sample estimates:  
## mean in group 0 mean in group 1  
## 1.891529 1.647213
```

Para un nivel de significación del 5%:  $p\text{-value} = 0.002642 < \alpha = 0.05$ , entonces RECHAZO  $H_0$

*#Ejemplo. Contrastar si existen diferencias de medias entre pulso antes y pulso después en fumadores.*

```
t.test(ElPulso$Pulso1[ElPulso$Fumar=="No Fuma"],ElPulso$Pulso2[ElPulso$Fumar=="No Fuma"], paired=T)

##
## Paired t-test
##
## data: ElPulso$Pulso1[ElPulso$Fumar == "No Fuma"] and ElPulso$Pulso2[ElPulso$Fumar == "No Fuma"]
## t = -4.193, df = 63, p-value = 8.762e-05
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -11.143637 -3.950113
## sample estimates:
## mean difference
## -7.546875
```

Para un nivel de significación del 5%:  $p\text{-value} = 8.762e-05 < \alpha = 0.05$ , entonces RECHAZO  $H_0$



**Imponer más condiciones “fumadoras/es no corredoras/es”:** `ElPulso$Correr=="No"&ElPulso$Fumar=="Fuma"`

```
t.test(ElPulso$Pulso1[ElPulso$Correr=="No"&ElPulso$Fumar=="Fuma"],ElPulso$Pulso2[ElPulso$Correr=="No"&ElPulso$Fumar=="Fuma"],paired=T)

##
## Paired t-test
##
## data: ElPulso$Pulso1[ElPulso$Correr == "No" & ElPulso$Fumar == "Fuma"] and ElPulso$Pulso2[ElPulso$Correr == "No" & ElPulso$Fumar == "Fuma"]
## t = -4.1395, df = 11, p-value = 0.001645
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -21.826755 -6.673245
## sample estimates:
## mean difference
## -14.25
```

Para un nivel de significación del 5%:  $p\text{-value} = 0.001645 < \alpha = 0.05$ , entonces RECHAZO  $H_0$

*#Ejemplo 3.18 Contrastar la hipótesis de que los valores de expresión del gen Gdf5 (x.3395) para pacientes ALL y AML (respectivamente) están normalmente distribuidos.*

```
shapiro.test(leukemia$x.3395[leukemia$Y==0])
```

```
##  
## Shapiro-Wilk normality test  
##
```

```
## data: leukemia$x.3395[leukemia$Y == 0]  
## W = 0.98102, p-value = 0.6348
```

```
shapiro.test(leukemia$x.3395[leukemia$Y==1])
```

```
##  
## Shapiro-Wilk normality test  
##
```

```
## data: leukemia$x.3395[leukemia$Y == 1]  
## W = 0.97114, p-value = 0.674
```

**require(car)** ← *Bandas de confianza en un QQ Plot, para identificar gráficamente la no normalidad de un conjunto de observaciones.*

```
## Loading required package: car
```

```
## Loading required package: carData
```

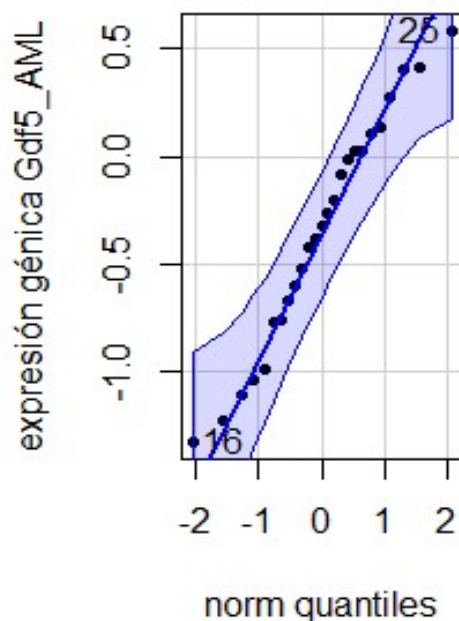
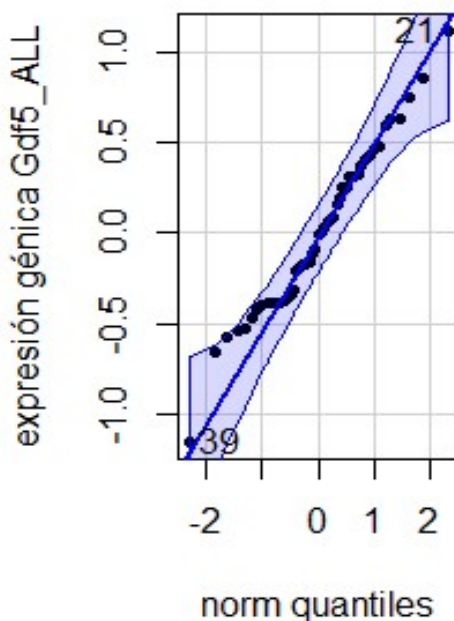
```
o=par(mfrow=c(1, 2))
```

```
qqPlot(leukemia$x.3395[leukemia$factor=="ALL"], pch=20, ylab='expresión génica Gdf5_ALL', m  
ain='QQplot expresión Gdf5 en ALL')
```

```
## [1] 39 21
```

```
qqPlot(leukemia$x.3395[leukemia$factor=="AML"], pch=20, ylab='expresión génica Gdf5_AML',ma  
in='QQplot expresión Gdf5 en AML')
```

## QQplot expresión Gdf5 en ALL      QQplot expresión Gdf5 en AML



```
## [1] 16 25
```

```
par(o)
```

*#Ejemplo 3.2 Análogamente, para el gen Gdf5, situado en la columna 3396, la hipótesis nula de igualdad de varianzas y la hipótesis nula de igualdad de medias entre los pacientes ALL y AML puede ser testadas.*

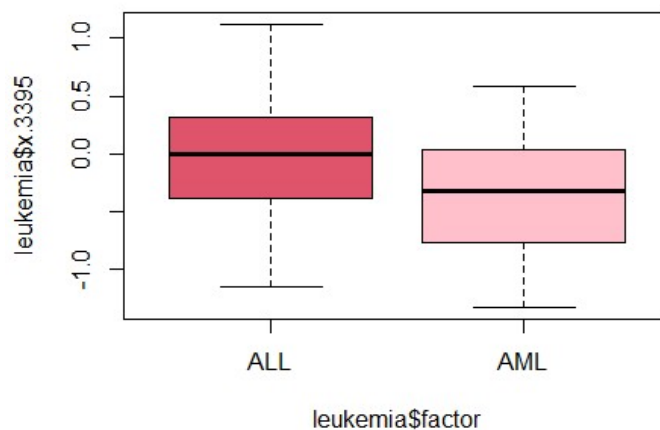
```
var.test(leukemia$x.3395~leukemia$Y)

##
## F test to compare two variances
##
## data: leukemia$x.3395 by leukemia$Y
## F = 0.71067, num df = 46, denom df = 24, p-value = 0.3151
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.3351383 1.3911130
## sample estimates:
## ratio of variances
## 0.7106712
```

```
t.test(leukemia$x.3395~ leukemia$Y,var.equal=T)

##
## Two Sample t-test
##
## data: leukemia$x.3395 by leukemia$Y
## t = 2.8894, df = 70, p-value = 0.005134
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## 0.1083809 0.5914420
## sample estimates:
## mean in group 0 mean in group 1
## 7.579787e-05 -3.498357e-01
```

```
boxplot(leukemia$x.3395~leukemia$factor, col=c(26,"pink"))
```

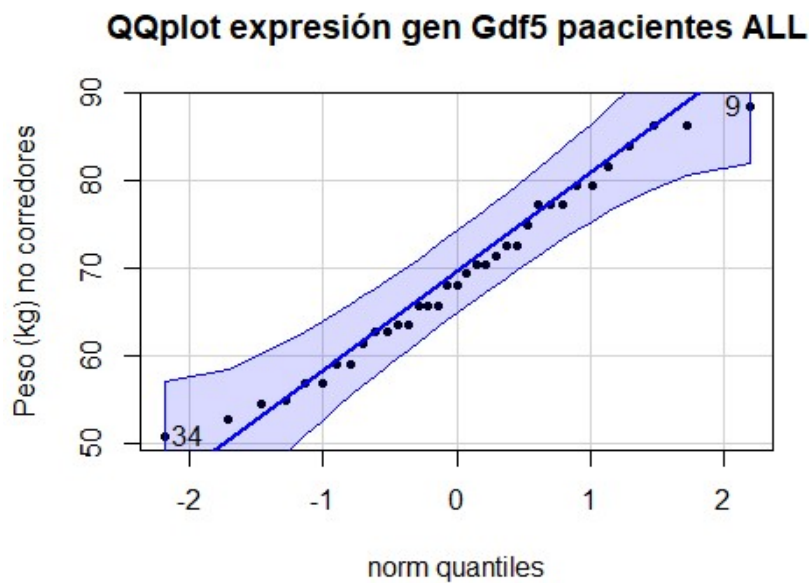


*#Ejemplo 3.20 Contrastar la normalidad de la variable Peso.kg para no corredores, n = 35.*

```
library(nortest)
lillie.test(ElPulso$Peso.kg[ElPulso$Correr=="No"])

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  ElPulso$Peso.kg[ElPulso$Correr == "No"]
## D = 0.076707, p-value = 0.8675
```

```
require(car)
qqPlot(ElPulso$Peso.kg[ElPulso$Correr=="No"], pch=20, ylab='Peso (kg) no corredores', main=
'QQplot expresión gen Gdf5 paacientes ALL')
```



```
## [1] 9 34
```

*#Ejemplo 3.21 Desde el fichero de datos "calorie", 20 observaciones y 4 variables, realizar un contraste de la normalidad de la variable wgt usando el test de Shapiro-Wilk, puesto que n = 20.*

```
shapiro.test(calorie$carb)

##
##  Shapiro-Wilk normality test
##
## data:  calorie$carb
## W = 0.9762, p-value = 0.8762

shapiro.test(calorie$age)

##
##  Shapiro-Wilk normality test
##
## data:  calorie$age
## W = 0.94367, p-value = 0.281
```

```

shapiro.test(calorie$wgt)

##
##  Shapiro-Wilk normality test
##
## data:  calorie$wgt
## W = 0.92421, p-value = 0.1194

shapiro.test(calorie$prot)

##
##  Shapiro-Wilk normality test
##
## data:  calorie$prot
## W = 0.94715, p-value = 0.326

o=par(mfrow=c(1,4))
require(car)
o=par(mfrow=c(2,2))
qqPlot(calorie$carb, pch=20, ylab='carb', main='QQplot')

## [1] 18 12

qqPlot(calorie$age, pch=20, ylab='age', main='QQplot')

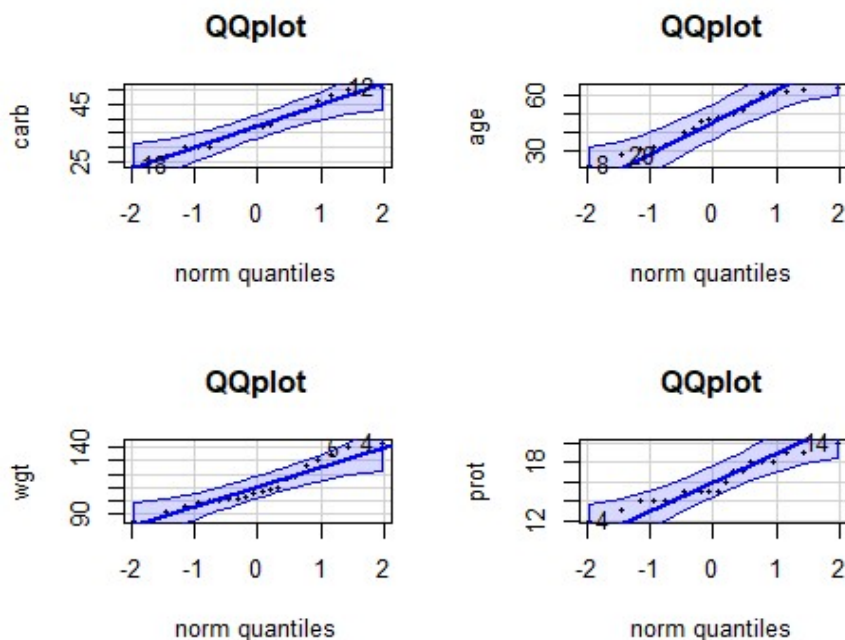
## [1] 8 20

qqPlot(calorie$wgt, pch=20, ylab='wgt', main='QQplot')

## [1] 4 5

qqPlot(calorie$prot, pch=20, ylab='prot', main='QQplot')

```



```

## [1] 14 4

par(o)

```