Inferencia Estadística

Profesor(es): Jarnishs Beltran
Ayudante: Pablo Rivera

Pauta ayudantía Nº7
Otoño 2020

Ejercicio en R

i) La prueba Test-Student es una prueba de contraste parametrica Se utiliza para comprobar la igualdad de las medias de dos muestras o una muestra.
También para comprobar si la media de una muestra es igual a una media teórica determinada. Los datos tienen que tener distribución normal (véase la prueba de Shapiro-Wilk).

a) Compare dos muestras que Ud. crea, 100 variables aleatorias con media 10 y otras 100 con media 10,5. Interprete con un gráfico de cajas.

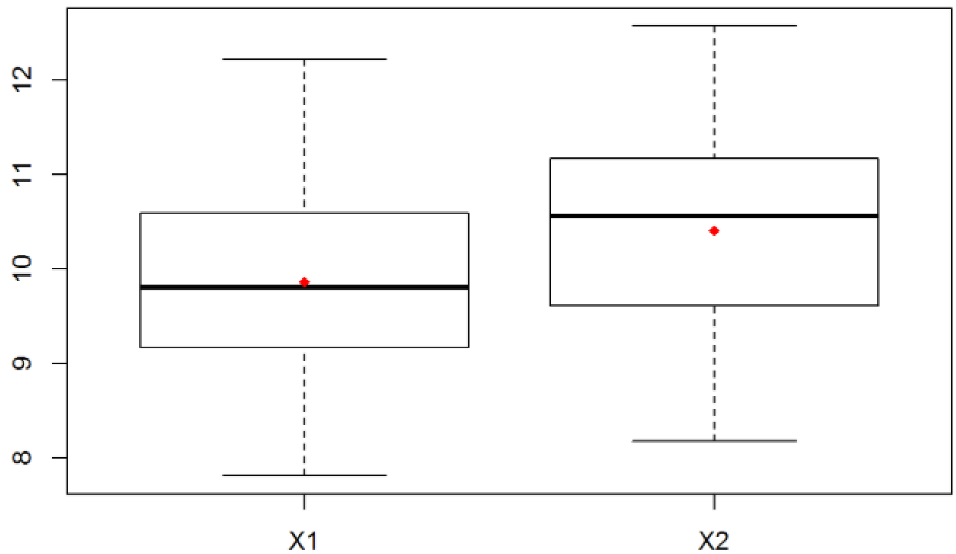
```
set.seed(10)
x1 <- rnorm(100,10) # Variable aleatoria de media 10
x2 <- rnorm(100,10.5) # Variable aleatoria de media 10.5

test <- t.test(x1,x2) # Prueba t de Student

print(test)
```

```
##
##  Welch Two Sample t-test
##
## data:  x1 and x2
## t = -4.0081, df = 197.83, p-value = 8.665e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8080508 -0.2751220
## sample estimates:
## mean of x mean of y
##  9.863451 10.405037
```

```
boxplot(x1,x2,names=c("X1","X2"))#Muestra las diagramas
medias <- c(mean(x1),mean(x2))#Muestra la Media mediante un punto
points(medias,pch=18,col="red")#Resalta la media de un color
```



b) Ahora realice una comparación de la media muestral (de 10) con la media misma.

```r
set.seed(10)
x <- rnorm(100,10) # Creación de una variable aleatoria de media 10
Media <- 10

test <- t.test(x, mu=Media) # Comparación de la media muestral con la media

print(test)
```

```
##
##  One Sample t-test
##
## data:  x
## t = -1.4507, df = 99, p-value = 0.15
## alternative hypothesis: true mean is not equal to 10
## 95 percent confidence interval:
##   9.676689 10.050213
## sample estimates:
## mean of x
##  9.863451
```

Como p-value > 0.05 no podemos rechazar la hipótesis de que la muestra tiene media 10.

ii) Queremos comparar una variable dependiente entre dos grupos diferentes. Estos últimos pueden ser independientes entre si (no pareados) o estar relacionados (pareados). En este último caso podría ser un grupo de personas que se someten a una dieta para engordar (peso antes y después de la dieta).

a) Crea dos grupos, muestre el aspecto de éstos mediante un gráfico de cajas.

```
gr1 = c(55,65,76,89,86,78,90,76,49,89,99,78,67,78,90,99,65,66)
gr2 = c(59,69,76,89,90,87,97,80,58,98,92,87,67,79,91,98,69,76)

str(gr1)
```

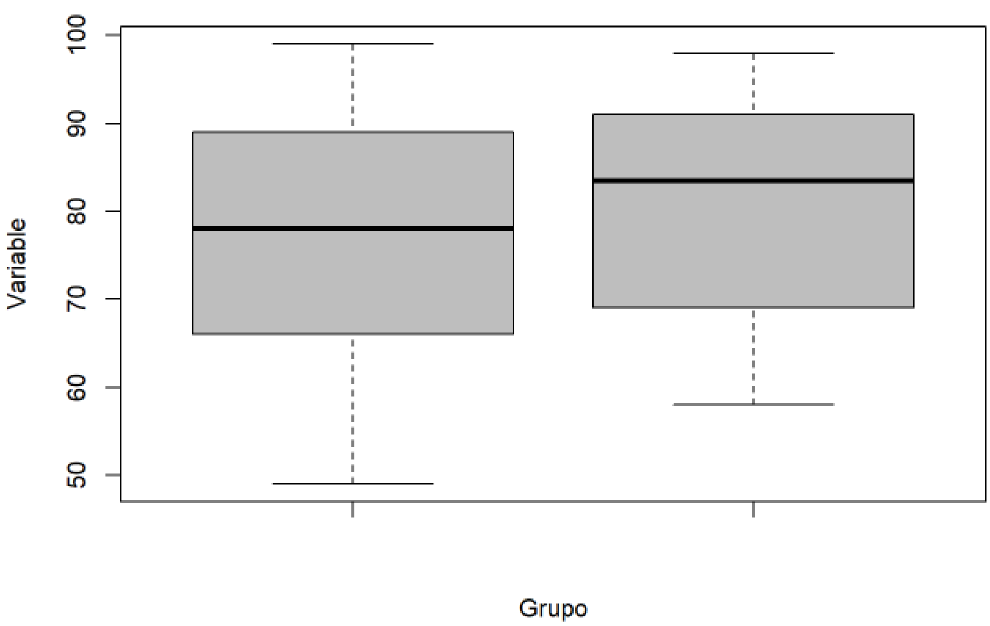
```
##  num [1:18] 55 65 76 89 86 78 90 76 49 89 ...
```

```
str(gr2)
```

```
##  num [1:18] 59 69 76 89 90 87 97 80 58 98 ...
```

¿Qué aspecto tienen nuestros datos? Los vamos a representar gráficamente por medio de un `boxplot`:

```
boxplot (gr1,gr2, xlab="Grupo", ylab="Variable", col="grey")
```



b) Comprobar si nuestros datos son normales, además, realice análisis visual de la normalidad

```
shapiro.test(gr1)
```

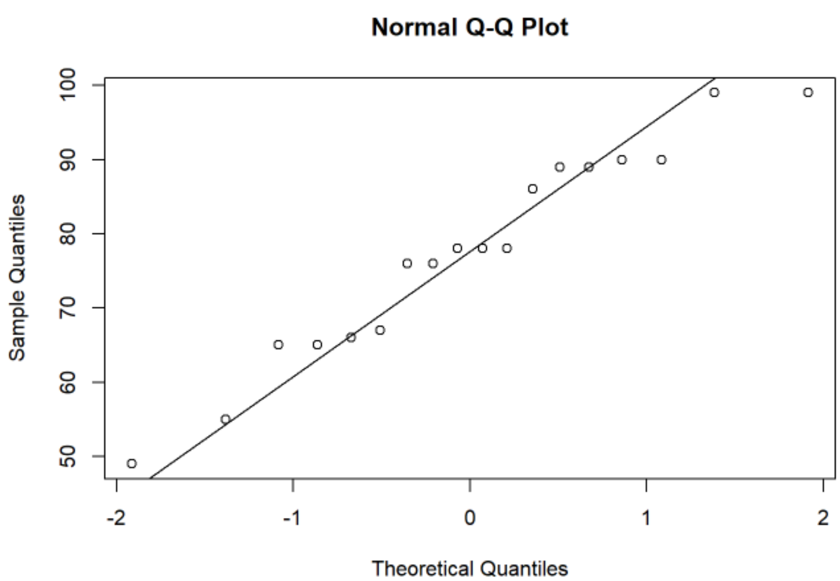
```
##
##  Shapiro-Wilk normality test
##
## data:  gr1
## W = 0.95493, p-value = 0.5075
```

```
shapiro.test(gr2)
```

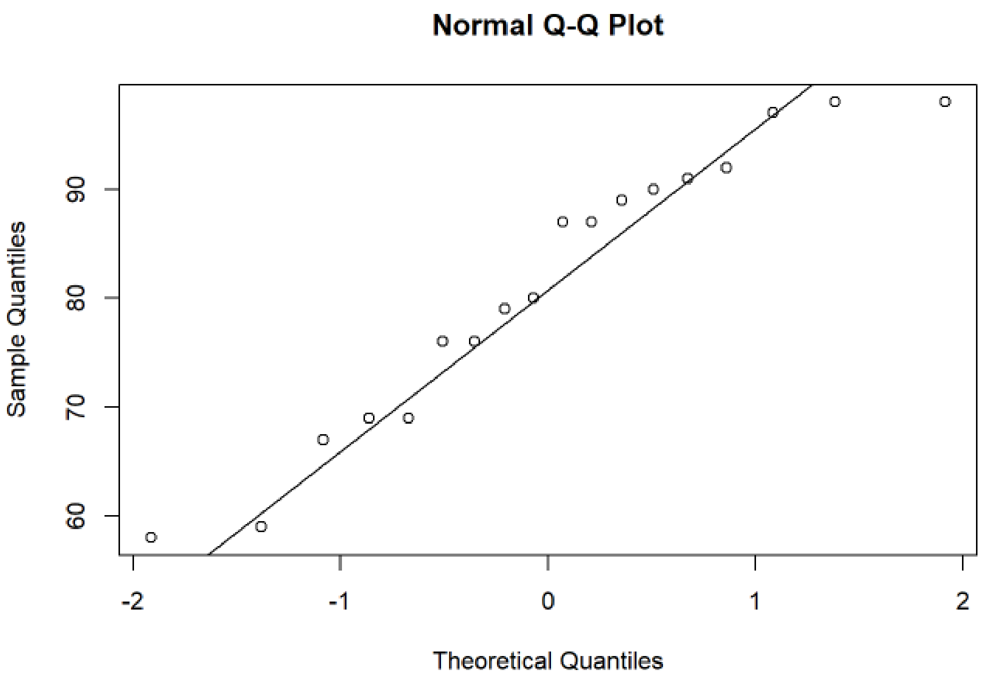
```
##
##  Shapiro-Wilk normality test
##
## data:  gr2
## W = 0.93285, p-value = 0.2179
```

Podemos también hacer un análisis visual de la normalidad (los puntos de la figura cuanto más cerca de la línea mucho mejor):

```
qqnorm(gr1)
qqline(gr1)
```



```
qqnorm(gr2)
qqline(gr2)
```



c) Realice un test t no pareados, pareados y diferencias de cero.

```
t.test(gr1,gr2, var.equal=TRUE, paired=FALSE)
```

```
##
##  Two Sample t-test
##
## data:  gr1 and gr2
## t = -0.82193, df = 34, p-value = 0.4168
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12.925515   5.481071
## sample estimates:
## mean of x mean of y
##  77.50000  81.22222
```

```
t.test(gr1,gr2, var.equal=TRUE, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  gr1 and gr2
## t = -3.4528, df = 17, p-value = 0.003039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.996646 -1.447798
## sample estimates:
## mean of the differences
##               -3.722222
```

```
t.test(gr1)
```

```
##
##  One Sample t-test
##
## data:  gr1
## t = 23.073, df = 17, p-value = 2.861e-14
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  70.41319 84.58681
## sample estimates:
## mean of x
##      77.5
```

```
t.test(gr2)
```

```
##
##  One Sample t-test
##
## data:  gr2
## t = 26.741, df = 17, p-value = 2.483e-15
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  74.81385 87.63060
## sample estimates:
## mean of x
##  81.22222
```

En ambos casos el conjunto de datos es diferente de cero.

d) Crea dos grupos esta vez con datos no normales, ¿qué se podría hacer en este caso?

Para datos no normales: U de Mann-Whitney

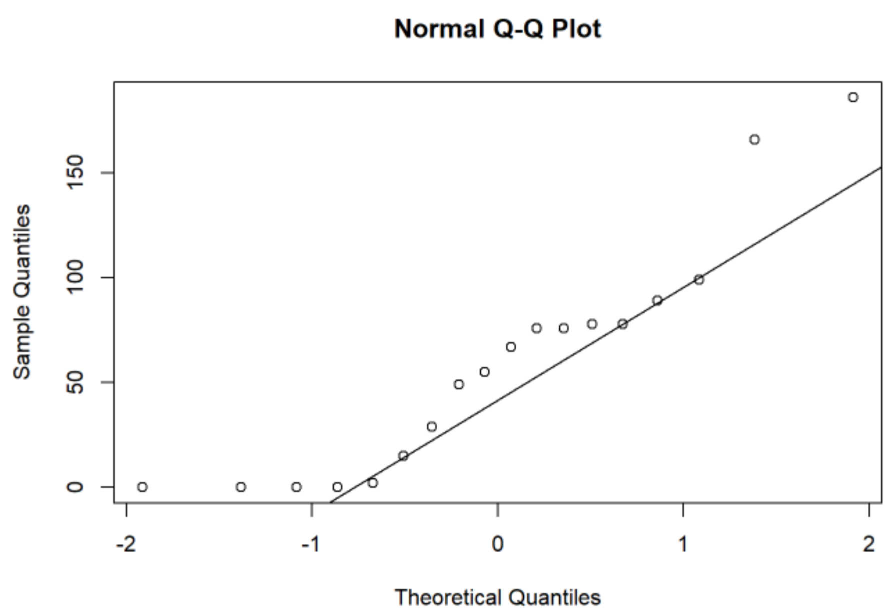
```
nn1 = c(55,0,76,89,186,0,0,76,49,2,29,78,67,78,0,99,15,166)
nn2 = c(0,2,76,89,0,87,97,0,58,98,92,87,67,179,1,98,69,6)
str(nn1)
```

```
##  num [1:18] 55 0 76 89 186 0 0 76 49 2 ...
```

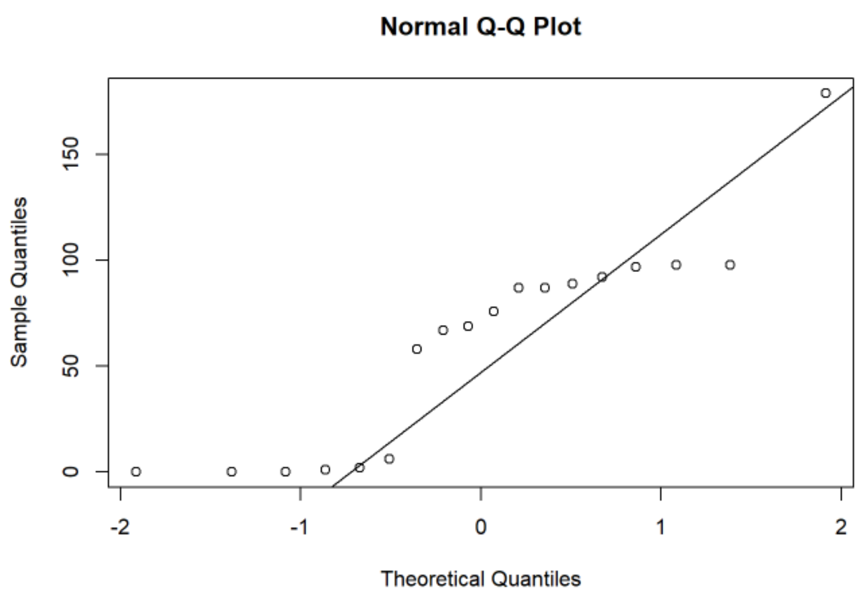
```
str(nn2)
```

```
##  num [1:18] 0 2 76 89 0 87 97 0 58 98 ...
```

```
qqnorm(nn1)
qqline(nn1)
```



```
qqnorm(nn2)
qqline(nn2)
```



Realizamos el test de Mann-Whitney o Wilcox test:

```
wilcox.test(nn1,nn2, correct=FALSE, exact=FALSE)
```

```
##
##  Wilcoxon rank sum test
##
## data:  nn1 and nn2
## W = 145.5, p-value = 0.6001
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(nn1,nn2, correct=FALSE, exact=FALSE, paired=TRUE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  nn1 and nn2
## V = 46.5, p-value = 0.443
## alternative hypothesis: true location shift is not equal to 0
```

Si queremos comparar cada grupo con respecto a cero entonces (con `mu` podemos testar otros valores):

```
wilcox.test(nn1, mu=0, alternative="two.sided")
```

```
## Warning in wilcox.test.default(nn1, mu = 0, alternative = "two.sided"):
## cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(nn1, mu = 0, alternative = "two.sided"):
## cannot compute exact p-value with zeroes
```