Hoja 4 (b): Inferencia sobre 2 muestras con R

Estadística Computacional I. Grado en Estadística

Departamento de Estadística e Investigación Operativa. Universidad de Sevilla

Ejercicio 1

Comparación de medias con varianzas iguales.

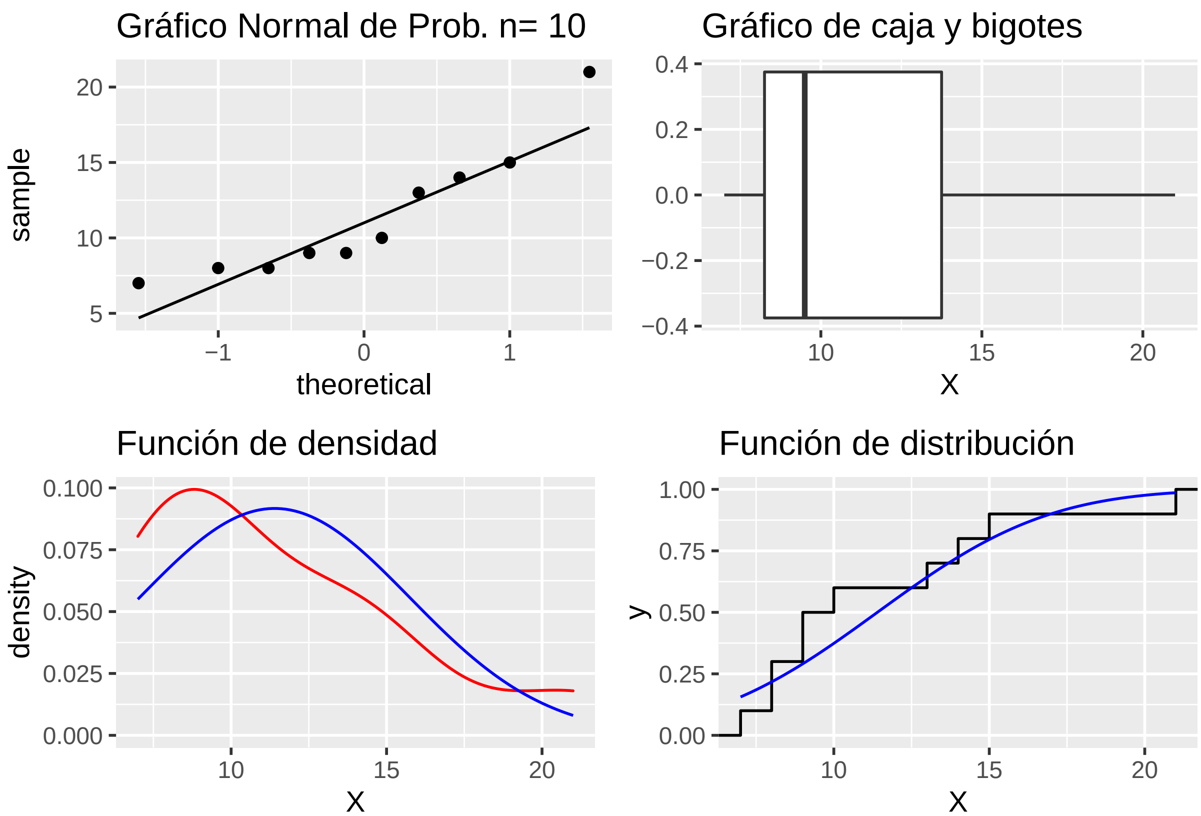
```
x <- c(15, 10, 13, 7, 9, 8, 21, 9, 14, 8)
y <- c(15, 14, 12, 8, 14, 7, 16, 10, 15, 12)
```

Tiempos de recuperación con cierta medicina (x) y grupo placebo (y).

Solución

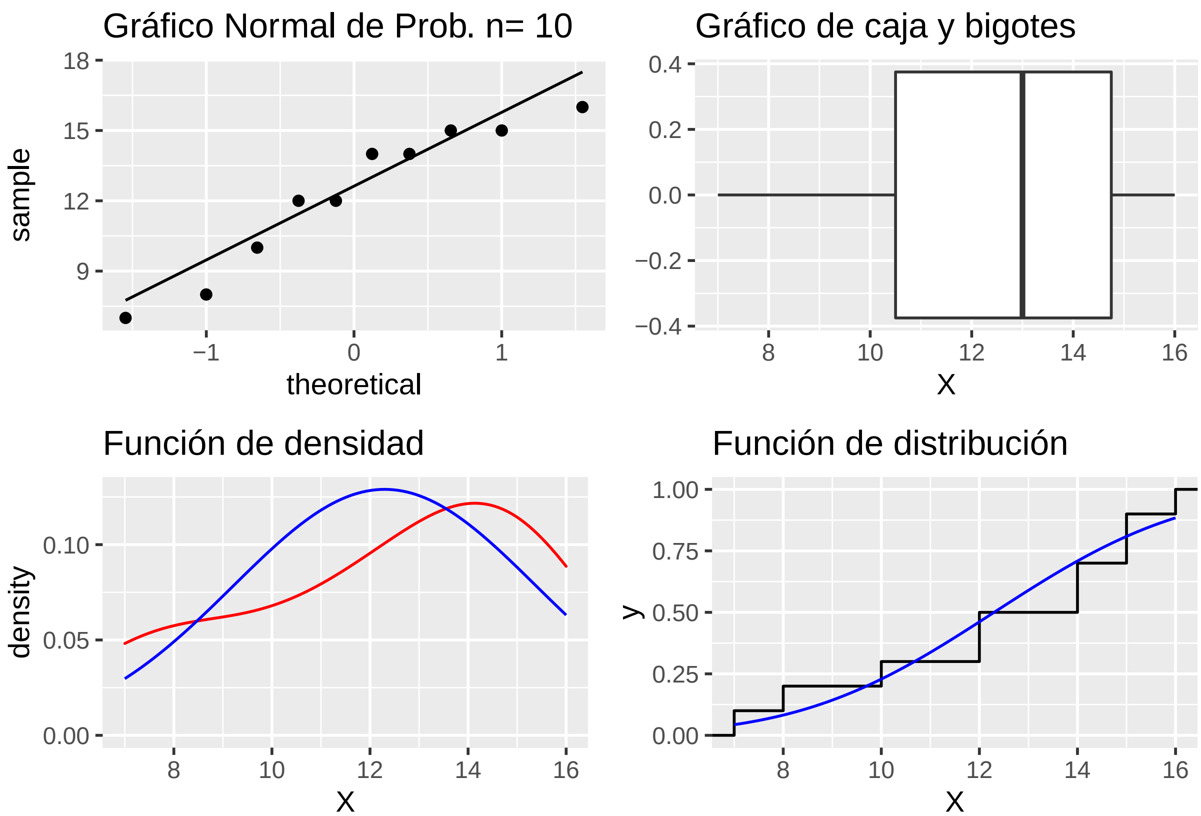
```
x <- c(15, 10, 13, 7, 9, 8, 21, 9, 14, 8)
y <- c(15, 14, 12, 8, 14, 7, 16, 10, 15, 12)
ananor_tidy(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.86663, p-value = 0.09131
```

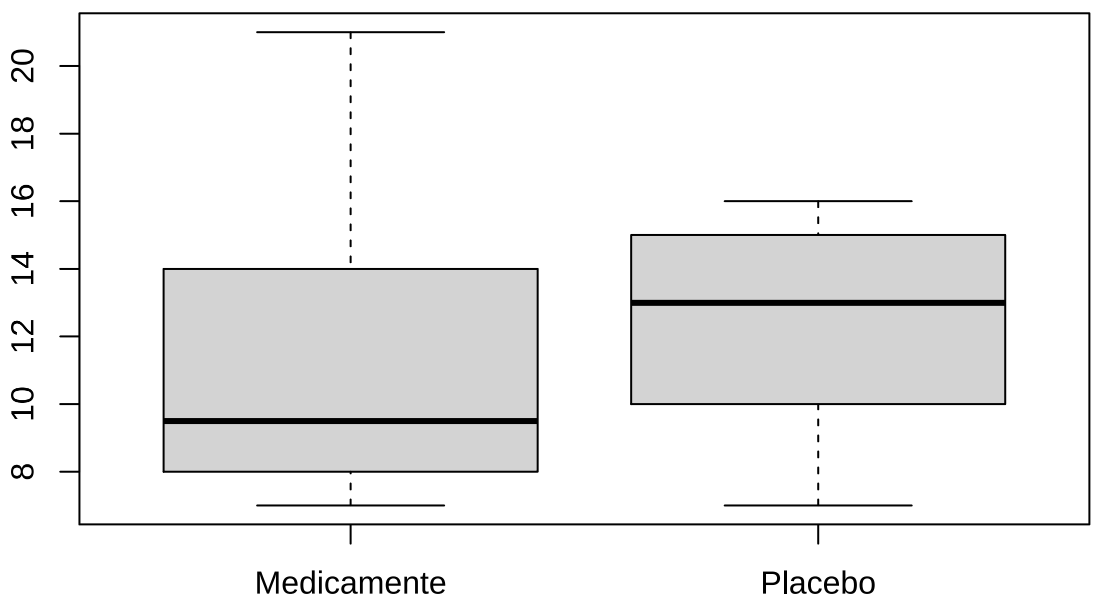



```
ananor_tidy(y)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.91249, p-value = 0.2986
```




```
boxplot(x,y,names = c("Medicamente","Placebo"))
```



Estudiar la igualdad de varianzas:

```
var.test(x,y)
```

```
##
##  F test to compare two variances
##
## data:  x and y
```

```
## F = 1.9791, num df = 9, denom df = 9, p-value = 0.3237
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.491579 7.967821
## sample estimates:
## ratio of variances
##           1.979094
```

Se pueden considerar varianzas iguales.

El estadístico que se está calculando aquí:

```
var(x)/var(y)
```

```
## [1] 1.979094
```

El test de Levene (para igualdad de varianzas):

```
library(car)
```

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

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
car::leveneTest(c(x,y), factor(c(rep("Medicamento",length(x)),
                                 rep("Placebo",length(y)))),
                center = "mean") # "median" es el valor por defecto
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##       Df F value Pr(>F)
## group  1  1.1857 0.2906
##       18
```

Los test paramétricos para comparar 2 muestras independientes

```
t.test(x, y, var.equal = TRUE, alternative = "two.sided")
```

```
##
##  Two Sample t-test
##
## data:  x and y
## t = -0.53311, df = 18, p-value = 0.6005
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.446765  2.646765
## sample estimates:
## mean of x mean of y
##      11.4      12.3
```

No hemos encontrados diferencias significativas para rechazar la igualdad de medias.

Ejercicio 2

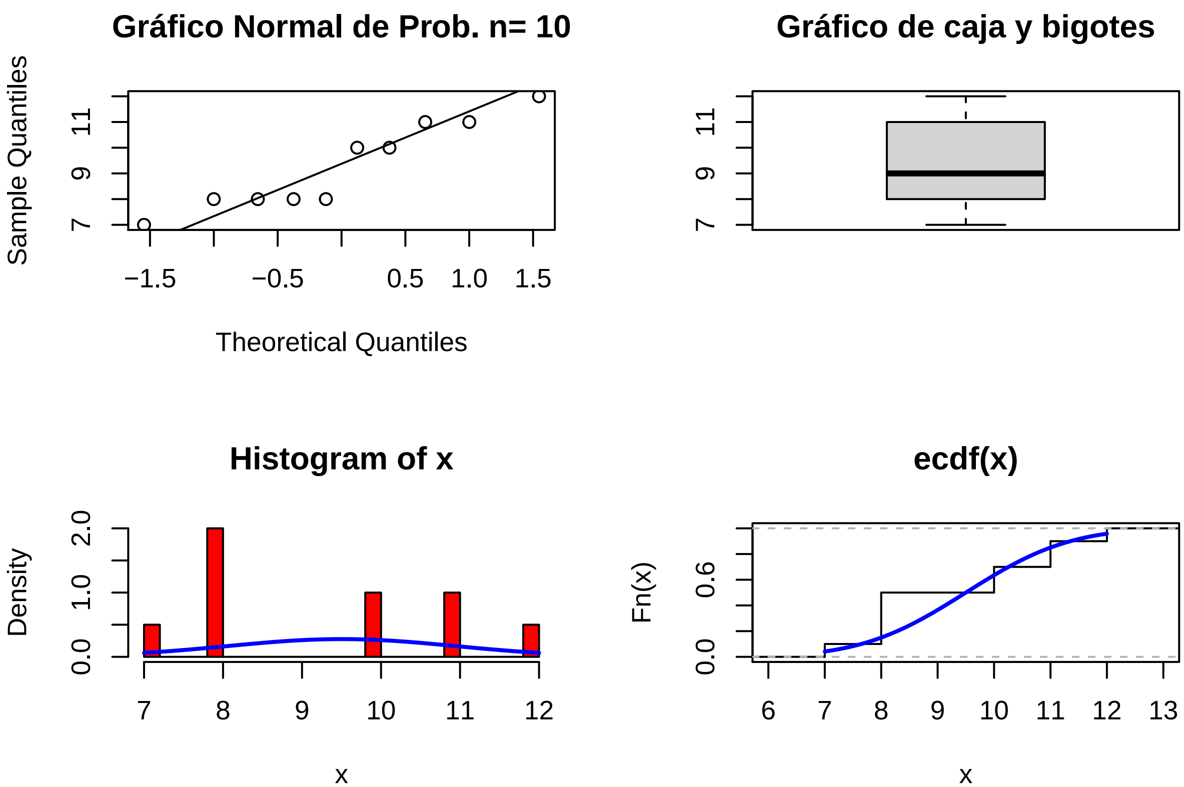
Comparación de medias con varianzas distintas

```
x <- c(11, 10, 8, 8, 10, 7, 12, 8, 11, 8)
y<- c(15, 10, 13, 7, 9, 8, 21, 9, 14, 8)
```

Analizar la normalidad, caja y bigotes, test de varianzas y t.test.

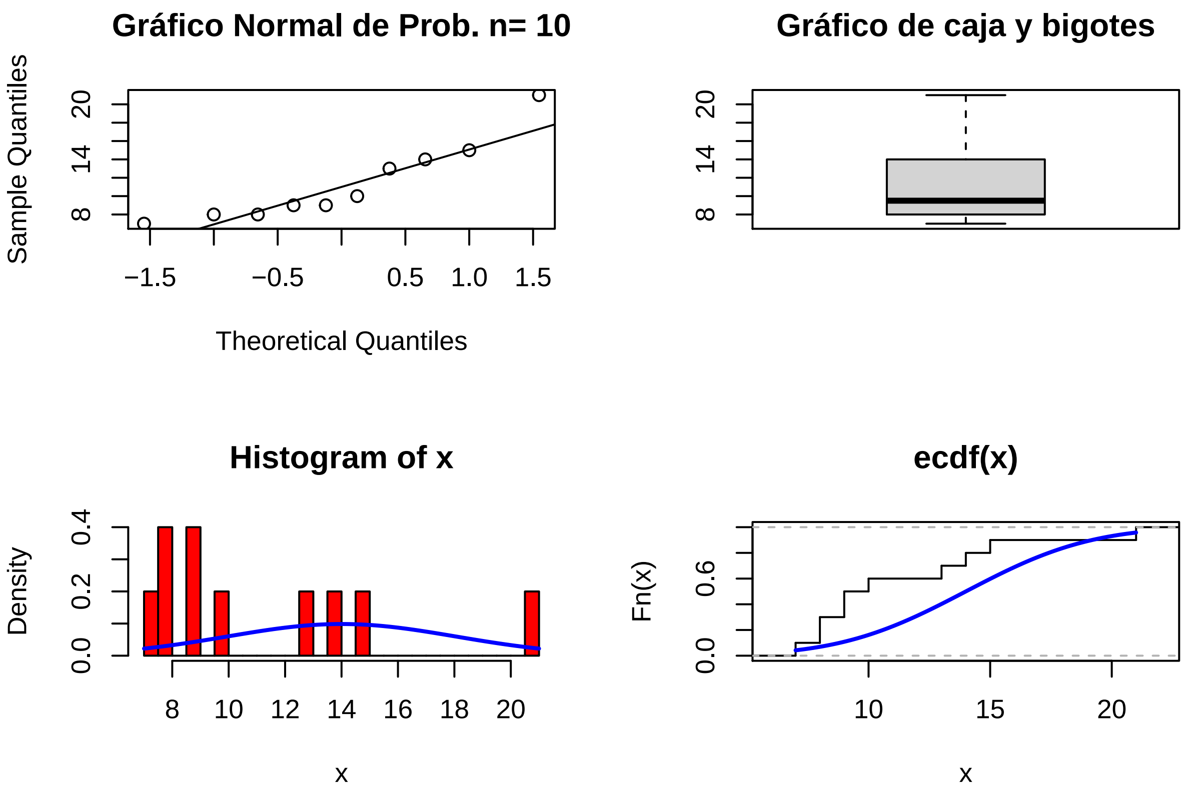
Solución

```
x <- c(11, 10, 8, 8, 10, 7, 12, 8, 11, 8)
y<- c(15, 10, 13, 7, 9, 8, 21, 9, 14, 8)
ananor(x)
```




```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.89172, p-value = 0.1773
```

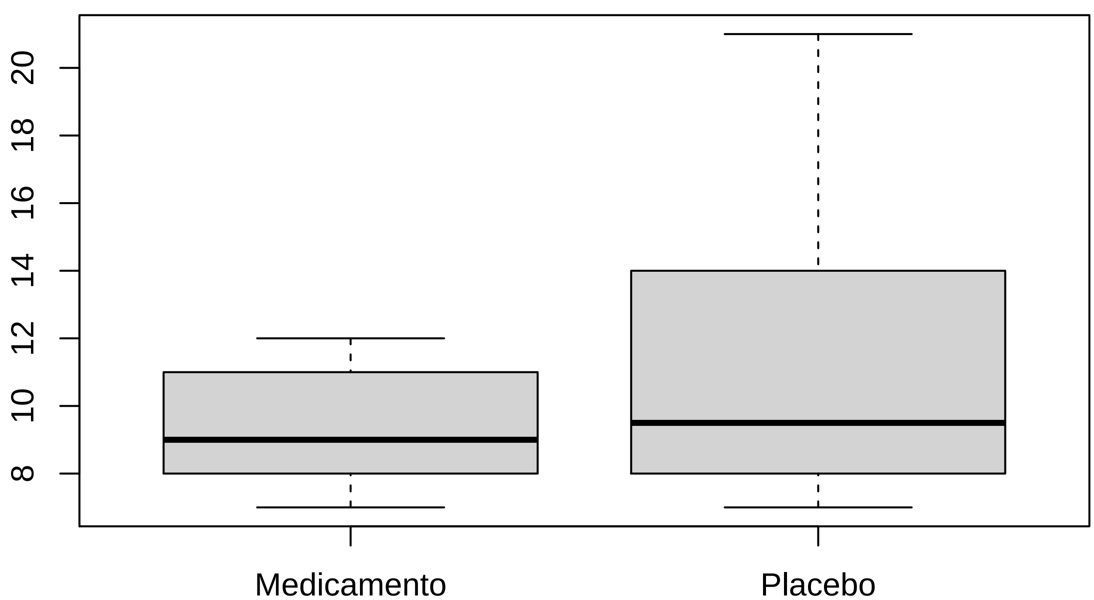
```
ananor(y)
```




```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.86663, p-value = 0.09131
```

"Aceptamos la Normalidad de las muestras".

```
boxplot(x,y,names = c("Medicamento","Placebo"))
```



```
var.test(x,y)
```

```
##
##  F test to compare two variances
##
## data:  x and y
## F = 0.15317, num df = 9, denom df = 9, p-value = 0.01007
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.03804502 0.61665756
## sample estimates:
## ratio of variances
##           0.153169
```

El p-valor $0.01007 < 0.05$, por tanto rechazamos la igualdad de varianzas.

Usamos también el test de Levene

```
leveneTest(c(x,y),
           factor(c(rep("Medicamento",length(x)),
                    rep("Placebo",length(y)))),
           center = "mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##       Df F value  Pr(>F)
## group  1  6.6704 0.01877 *
##       18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparar las medias de las dos muestras

```
t.test(x,y,var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = -1.4212, df = 11.694, p-value = 0.1814
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.32881  1.12881
## sample estimates:
## mean of x mean of y
##       9.3      11.4
```

"Aceptamos la igualdad de medias".

Ejercicio 3

Dos muestras relacionadas

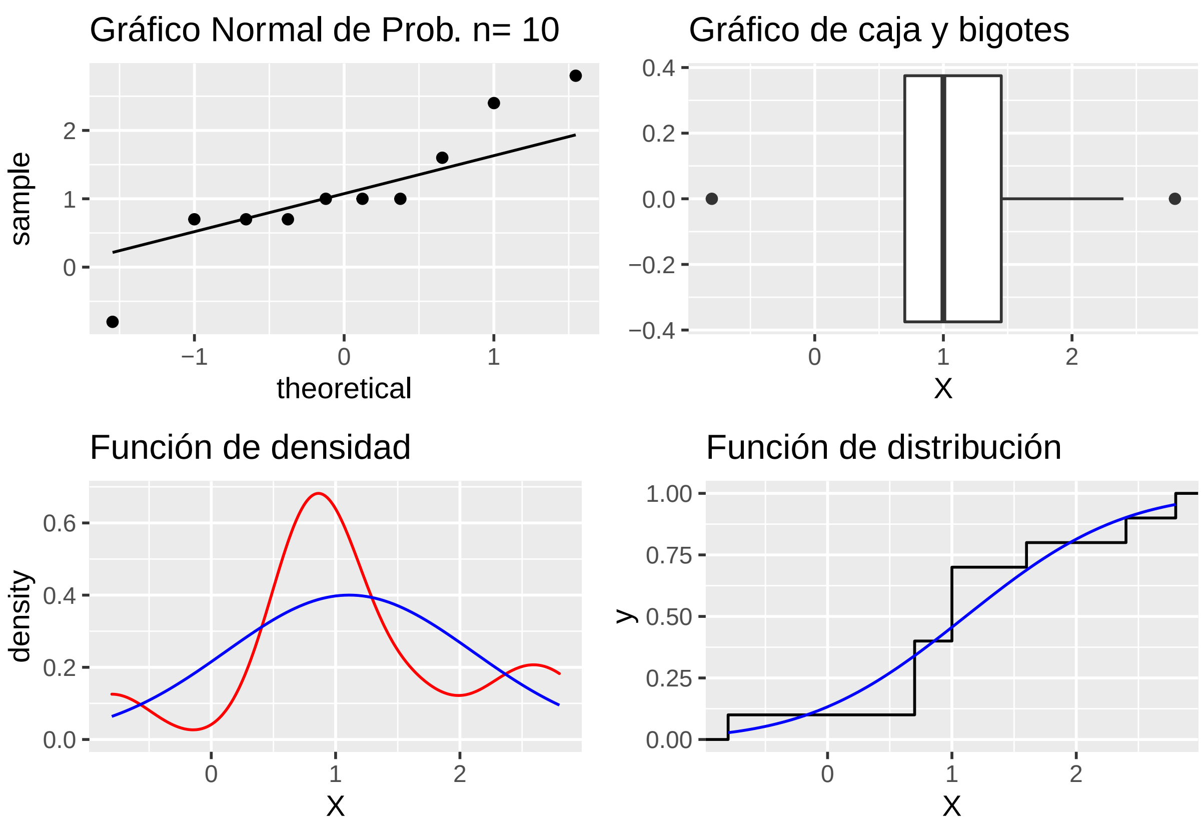
10 vinos son puntuados por dos jurados. Se quiere contrastar que el jurado 1 puntúa más alto que el jurado 2.

```
x <- c(3.1, 0.2, 5.1, 1.9, 4.8,
       4.9, 5.2, 4.5, 4.3, 4.8)
y <-c(2.1, 1, 4.1, 1.2, 4.1,
      3.3, 2.8, 1.7, 3.3, 4.1)
```

Solución

```r
x <- c(3.1, 0.2, 5.1, 1.9, 4.8,
       4.9, 5.2, 4.5, 4.3, 4.8)
y <-c(2.1, 1, 4.1, 1.2, 4.1,
      3.3, 2.8, 1.7, 3.3, 4.1)
Dif = x - y
ananor_tidy(Dif)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.91043, p-value = 0.284
```



Se acepta la Normalidad de la muestra diferencia.

Se quiere contrastar que el jurado 1 puntúa más alto que el jurado 2 es equivalente a que "Media de la Dif = X-Y > 0"

```r
t.test(Dif, alternative = "greater" )
```

```
##
##  One Sample t-test
##
## data:  Dif
## t = 3.5201, df = 9, p-value = 0.003257
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  0.5319636       Inf
```

```
## sample estimates:
## mean of x
##      1.11
```

```r
t.test(x,y,paired = TRUE,alternative = "greater")
```

```
##
##  Paired t-test
##
## data:  x and y
## t = 3.5201, df = 9, p-value = 0.003257
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.5319636       Inf
## sample estimates:
## mean of the differences
##                    1.11
```

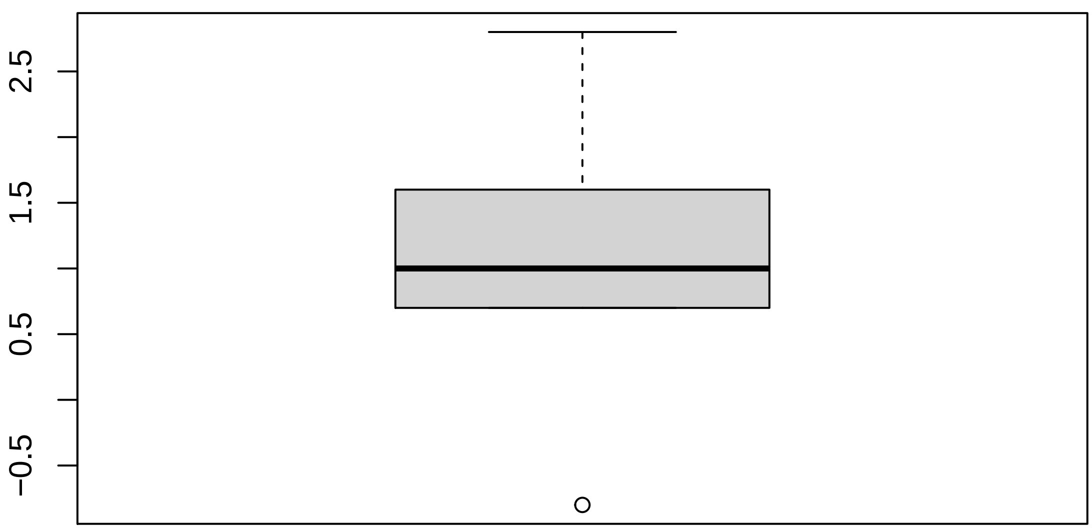
El p-valor es 0.003257 menor que 0.05, por tanto rechazamos H0, en favor de la hipótesis alternativa: Primer jurado puntúa más alto que el segundo.

Ejercicio 4

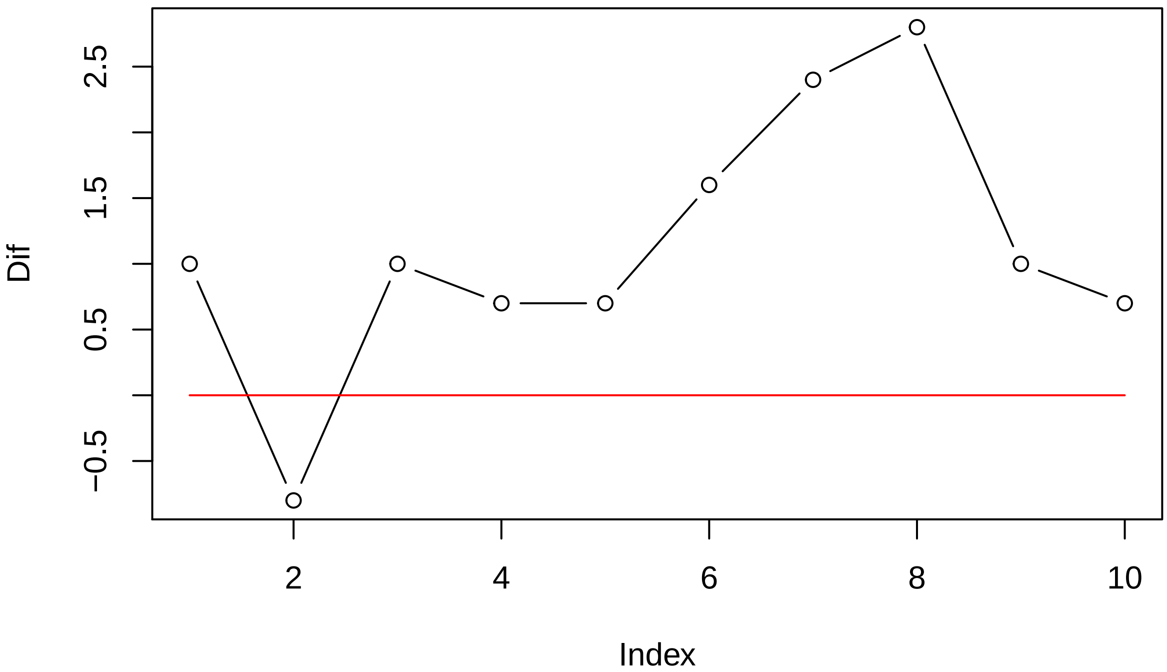
Dibujar ambas muestras (anteriores) con plot caja y bigote de x-y y realizar el contraste.

Solución

```r
boxplot(Dif)
```




```r
plot(Dif,type = "b")
lines(x = 1:10, y = rep(0,10), col = "red")
```



Si hubiera fallado la normalidad

Se aplicaría el test de Wilcoxon para la muestra diferencia

```
wilcox.test(Dif, alternative = "greater")
```

```
## Warning in wilcox.test.default(Dif, alternative = "greater"): cannot compute
## exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  Dif
## V = 51, p-value = 0.009336
## alternative hypothesis: true location is greater than 0
```

Ejercicio 5

Dos muestras independientes

```
x=c(0.11, 0.62, 0.32, 2.41, 3.48,
    0.29, 0.81, 0.43, 1.71, 0.46,0.92)
y=c(0.01, 0.14, 0.23, 0.18, 1.32,
    0.86, 0.97, 0.34, 0.25, 0.72)
```

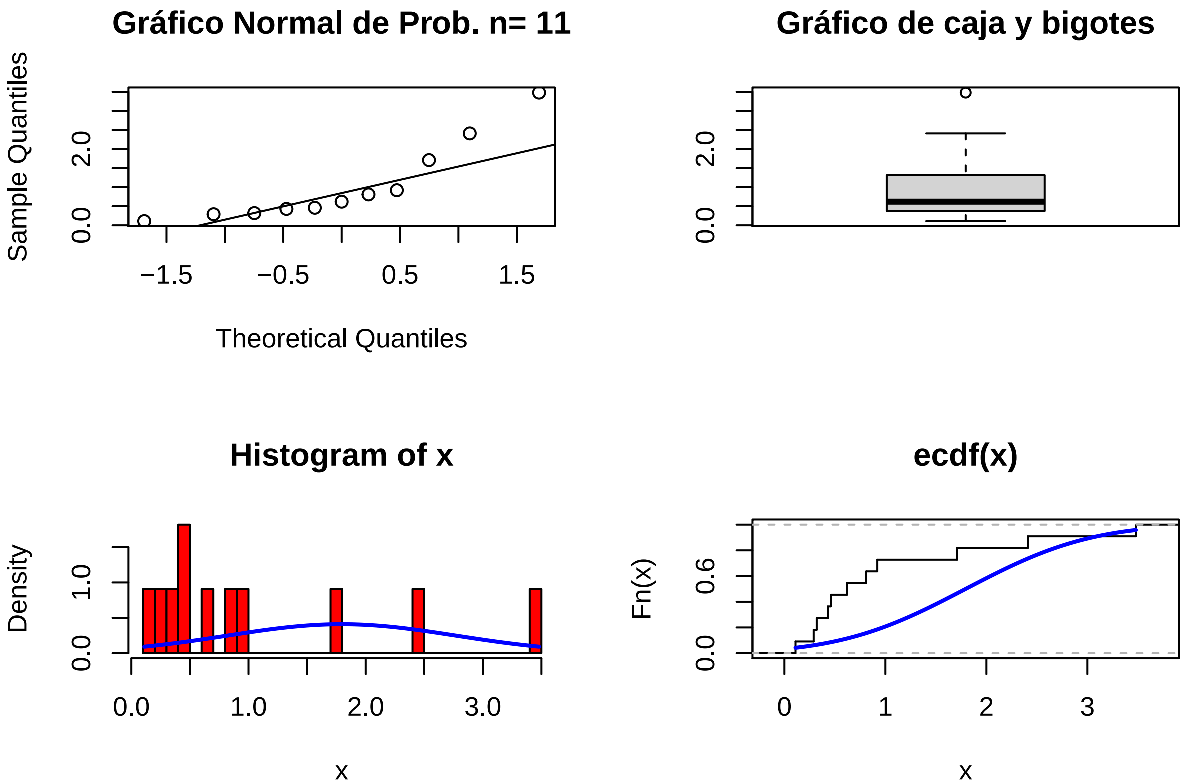
Con test no paramétricos.

Solución

Estudiamos la Normalidad

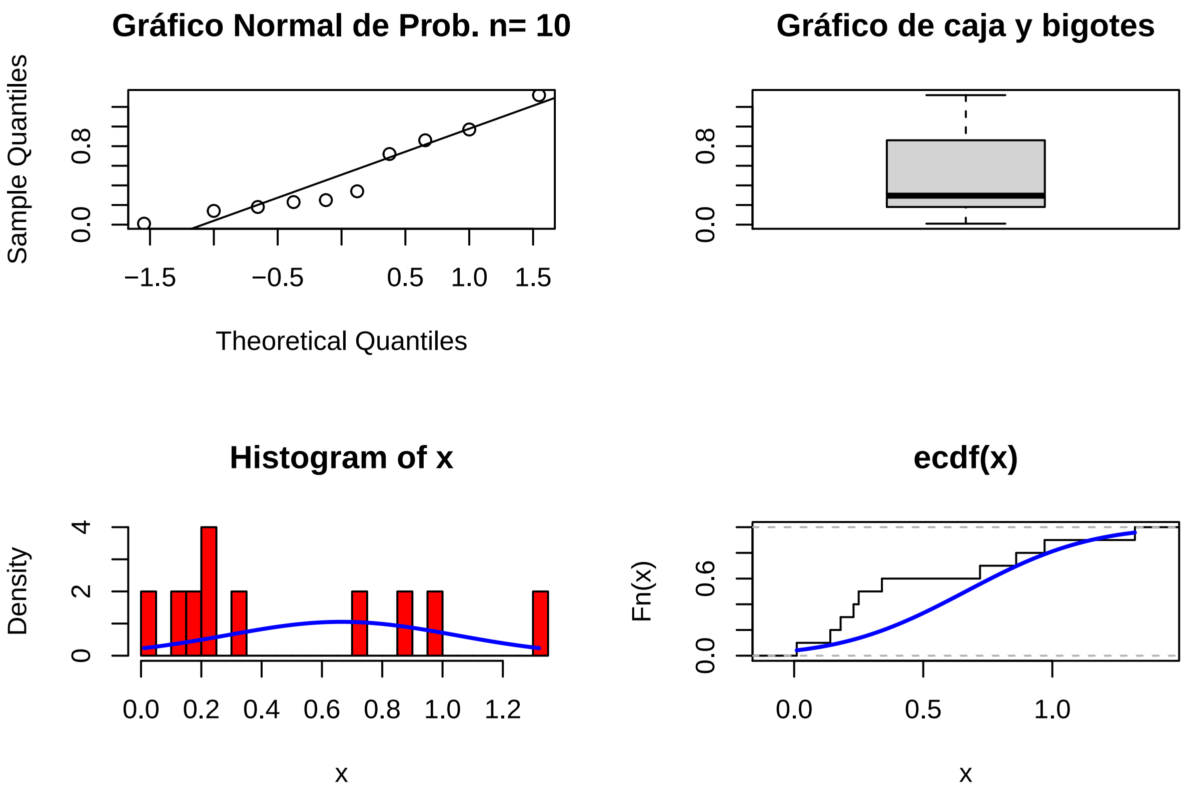
```
x=c(0.11, 0.62, 0.32, 2.41, 3.48,
    0.29, 0.81, 0.43, 1.71, 0.46,0.92)
y=c(0.01, 0.14, 0.23, 0.18, 1.32,
```

```
    0.86, 0.97, 0.34, 0.25, 0.72)
ananor(x)
```




```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.80467, p-value = 0.01084
```

```
ananor(y)
```




```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.89366, p-value = 0.1864
```

Test de U Mann-Whitney o Mann-Whitney-Wilcoxon

```
wilcox.test(x,y,exact = TRUE)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  x and y
## W = 74, p-value = 0.1971
## alternative hypothesis: true location shift is not equal to 0
```

Se acepta la igualdad de "medias" (medianas).

```
wilcox.test(x,y,exact = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 74, p-value = 0.1927
## alternative hypothesis: true location shift is not equal to 0
```