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

Estadística Computacional I. Grado en Estadística

Marta Venegas Pardo

Índice

1. Ejercicio 1: Comparación de medias con varianzas iguales. 1
- 1.1. Representación . . . 3
- 1.2. Estudiar la igualdad de varianzas . . . 4
- 1.3. Test paramétricos para la comparación de muestras independientes (t-test) . . . 4

2. Ejercicio 2: Comparación de medias con varianzas distintas 5
- 2.1. Análisis de normalidad, caja y bigotes, test de varianzas y t.test . . . 5
- 2.2. Comparación de medias en dos muestras, con distintas varianzas. . . . 7

3. Ejercicio 3: Muestras relacionadas 8

4. Ejercicio 4 9
- 4.1. Comparación de medias en muestras relacionadas NO NORMALES- TEST DE WILXCONSON 10

5. Ejercicio 5: Muestras independientes 10
- 5.1. Estudiamos la normalidad . . . 11
- 5.2. Test de Mann-Whitney o Mann-Whitney-Wilconxon (igualdad de medias) . . . 12

Instalamos las librerías necesarias.

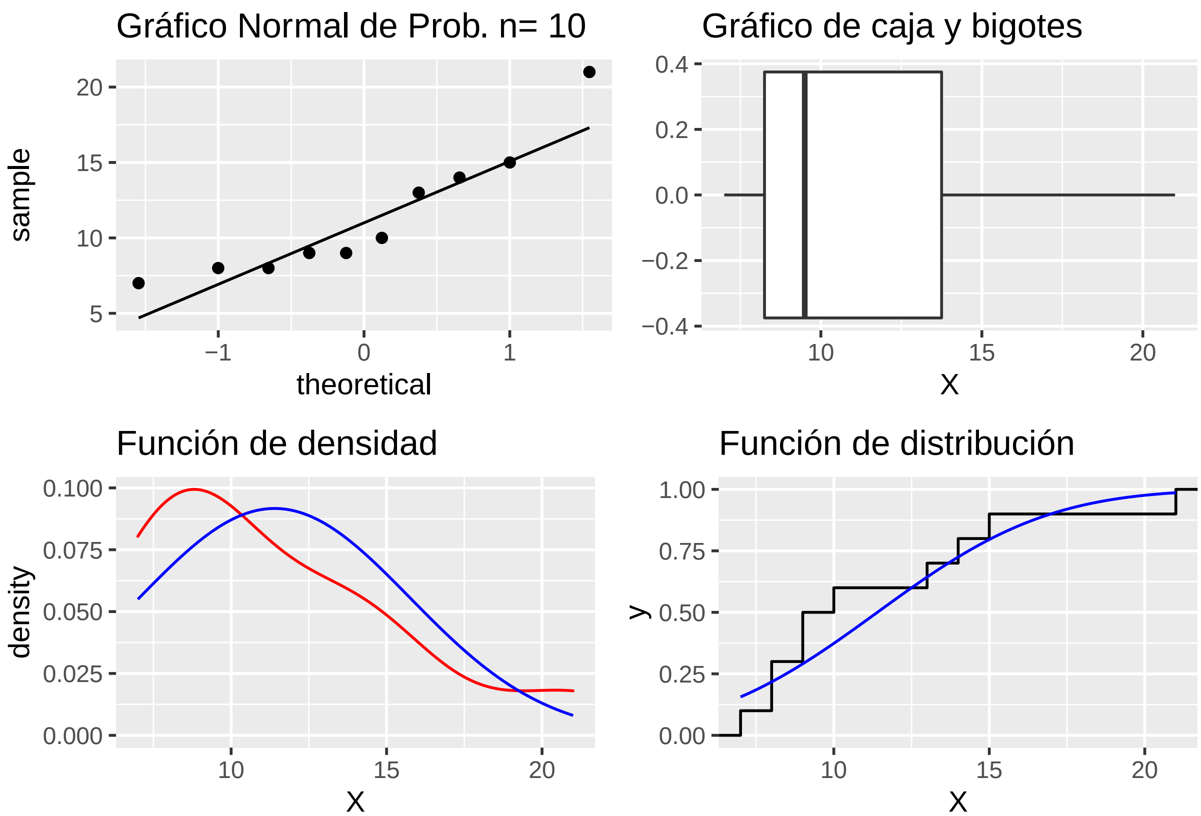
```
library(ggplot2)
source("ananor.R")
library(patchwork)
```

1. 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)
```

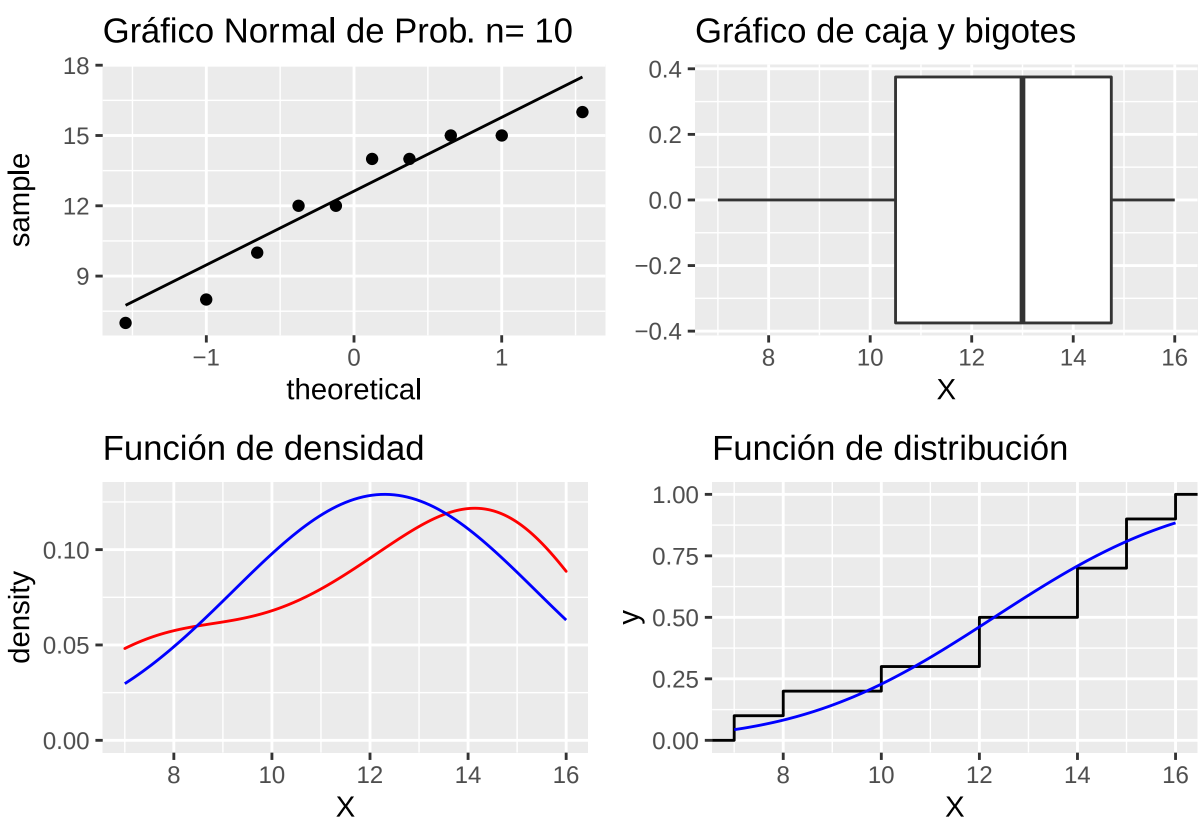
```
ananor_tidy(x)
```

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



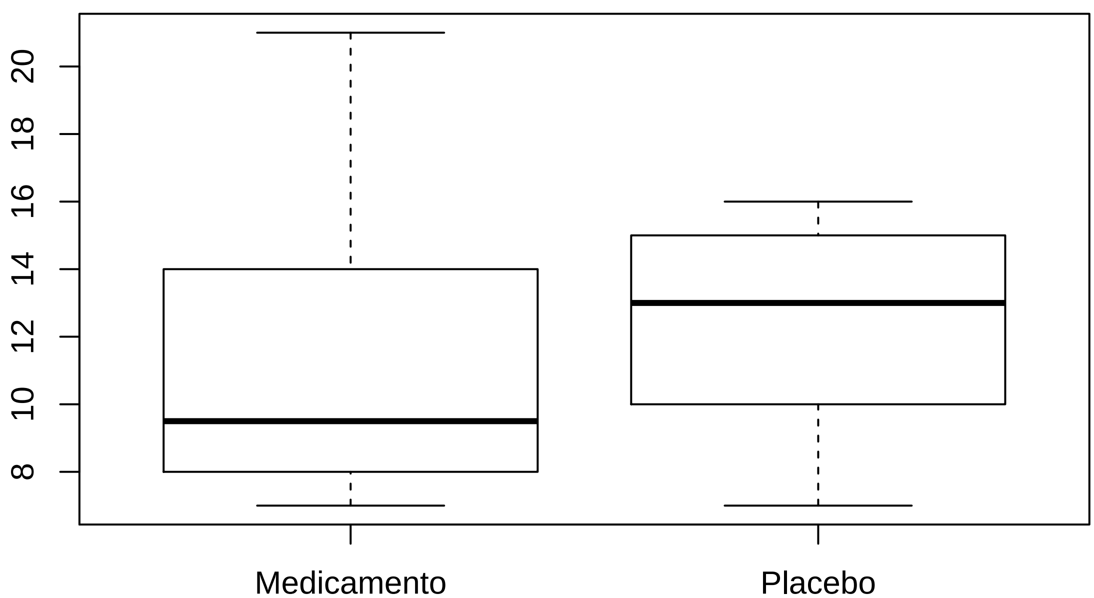

```
ananor_tidy(y)
```

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



1.1. Representación

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



Tiempos de recuperación con cierta medicina (x) y grupo placebo (y). La mediana con placebo es bastante más elevada que con medicamentos.

1.2. Estudiar la igualdad de varianzas

$$H_0 : \sigma_x^2 = \sigma_y^2$$
$$H_1 : \sigma_x^2 \neq \sigma_y^2$$

- 1.Con var.test

```
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, acepto H_0, pvalor>0.05

El estadístico utilizado, es el codiente de las cuasivarianzas, porque n es el mismo para ambas muestras.

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

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

- 2. Otro test: Test de Levenge para igualdad de varianzas

Está en la librería car

```
library(car)
```

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

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

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

```
# median es el valor por defecto
```

Aceptamos la igualdad de varianzas.

1.3. Test paramétricos para la comparación de muestras independientes (t-test)

```
t.test(x,
       y,
       var.equal = TRUE, # Porque hemos aceptado homocedasteceidad.
       alternative = "two.sided") # También los mayor y menor que
```

```
##
##  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 tenemos evidencias para rechazar $H_0 : \mu_1 = \mu_2$, a pesar de ver una gran diferencia en el boxplot.

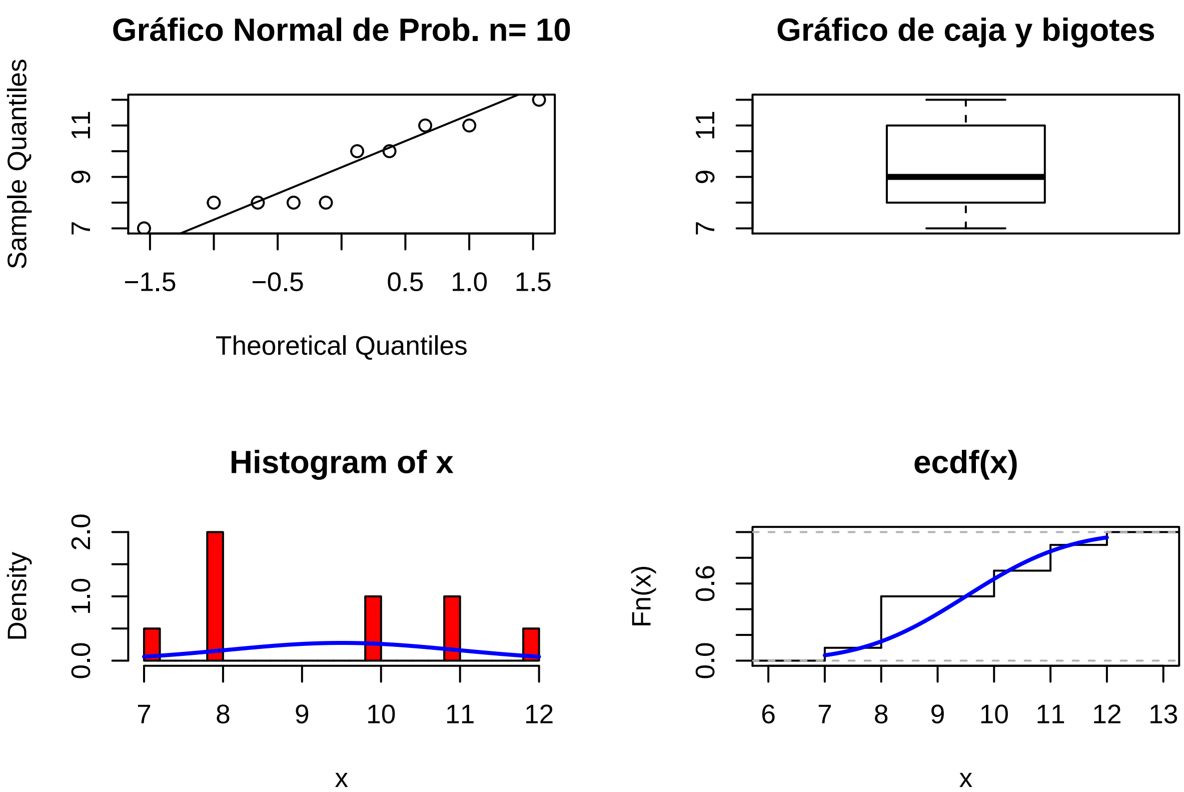
Vemos que en diferencia de medias, el 0 está contenido en el intervalo. Acepto H_0.

2. 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)
```

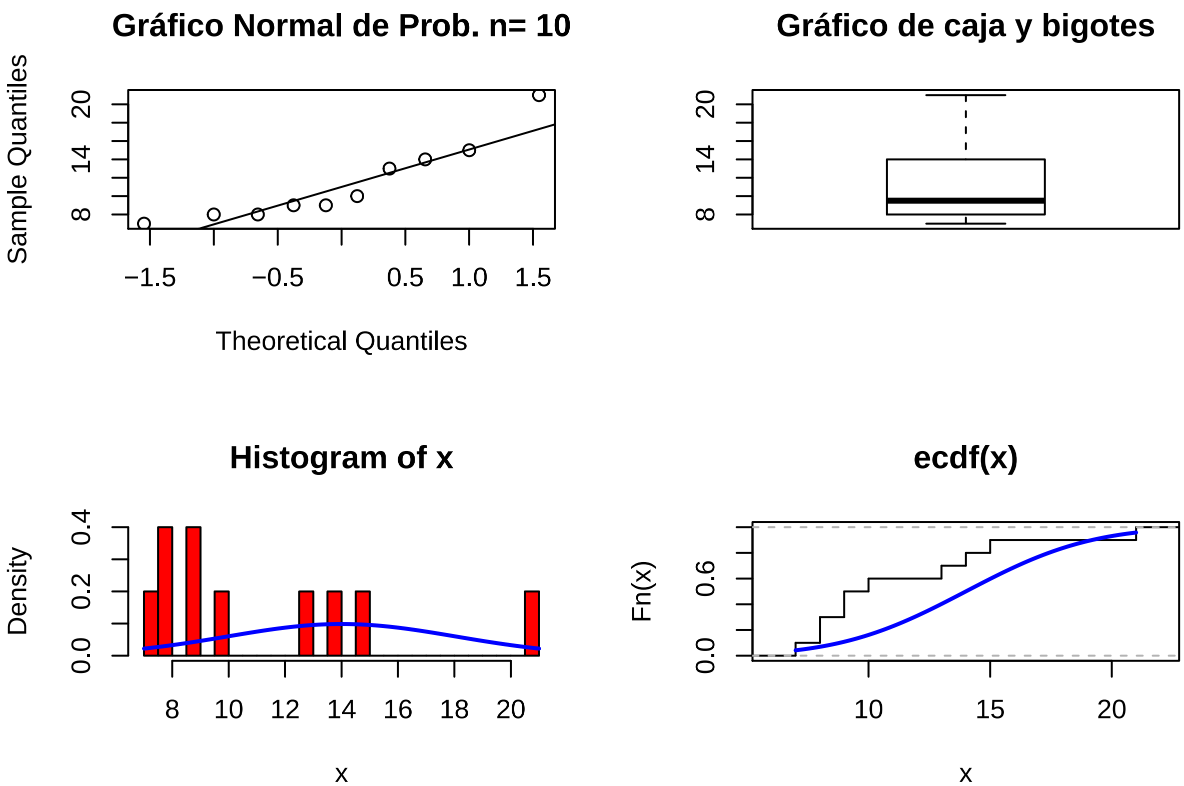
2.1. Análisis de normalidad, caja y bigotes, test de varianzas y t.test

```
ananor(x)
```




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

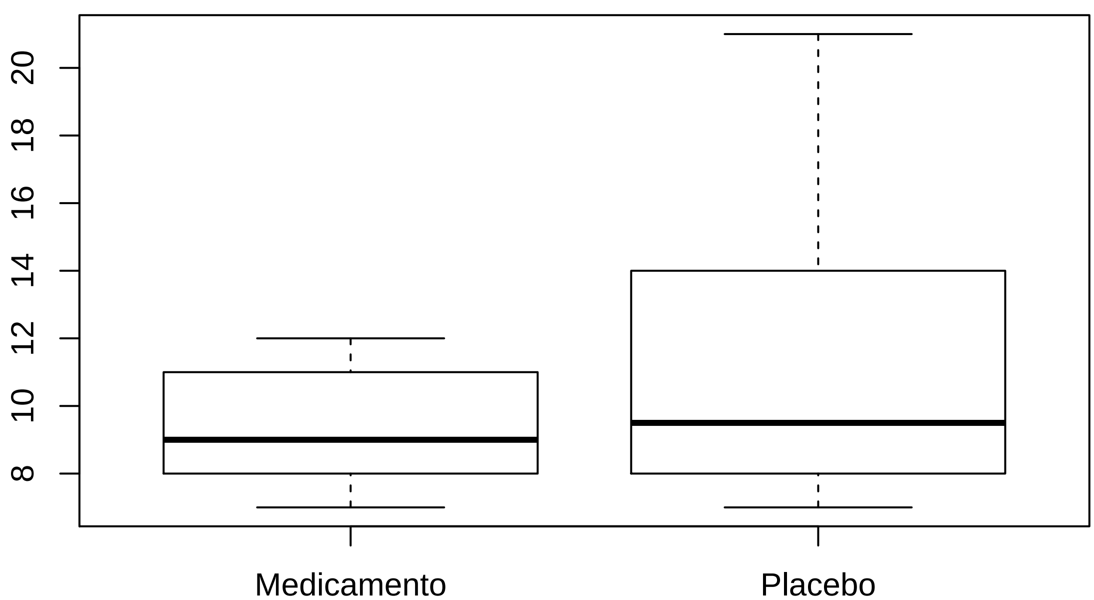
```
ananor(y)
```




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

"Aceptamos la normalidad de las muestras", a vista de los resultados del test de Shapiro-Wilk

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



La mediana para ambas variables es muy sumilar, aunque vemos que con medicamento el tiempo mediano de recuperación es algo inferior.

Realizamos el var.test de igualdad de varianzas.

```
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 $0.01007 < 0.05$, rechazo H_0.

Realizamos también el Test de Levenge:

```
leveneTest(c(x,y),
           factor(c(rep("Medicamento",length(x)),
                    c(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
```

Valor muy poco significativo, vuelvo a tener que rechazar H_0, no existe igualdad de varianzas.

2.2. Comparación de medias en dos muestras, con distintas varianzas.

$$H_0 : \mu_x = \mu_y$$
$$H_1 : \mu_x \neq \mu_y$$

```
t.test(x,
       y,
       var.equal = F # Porque hemos aceptado homocedasteceidad.
)
```

```
##
##  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. Misma media pero las muestras no proceden de la misma distribución normal.

3. Ejercicio 3: Muestras relacionadas

10 vinos son puntuados por dos jurados. Se quiere contrastar que el jurado 1 puntúa más alto que el jurado 2. Esto es equivalente a que la diferencia, en término medio, sea mayor que o $dif = X - Y > 0$

$$H_0 : \mu_x > \mu_y$$
$$H_1 : \mu_x \leq \mu_y$$

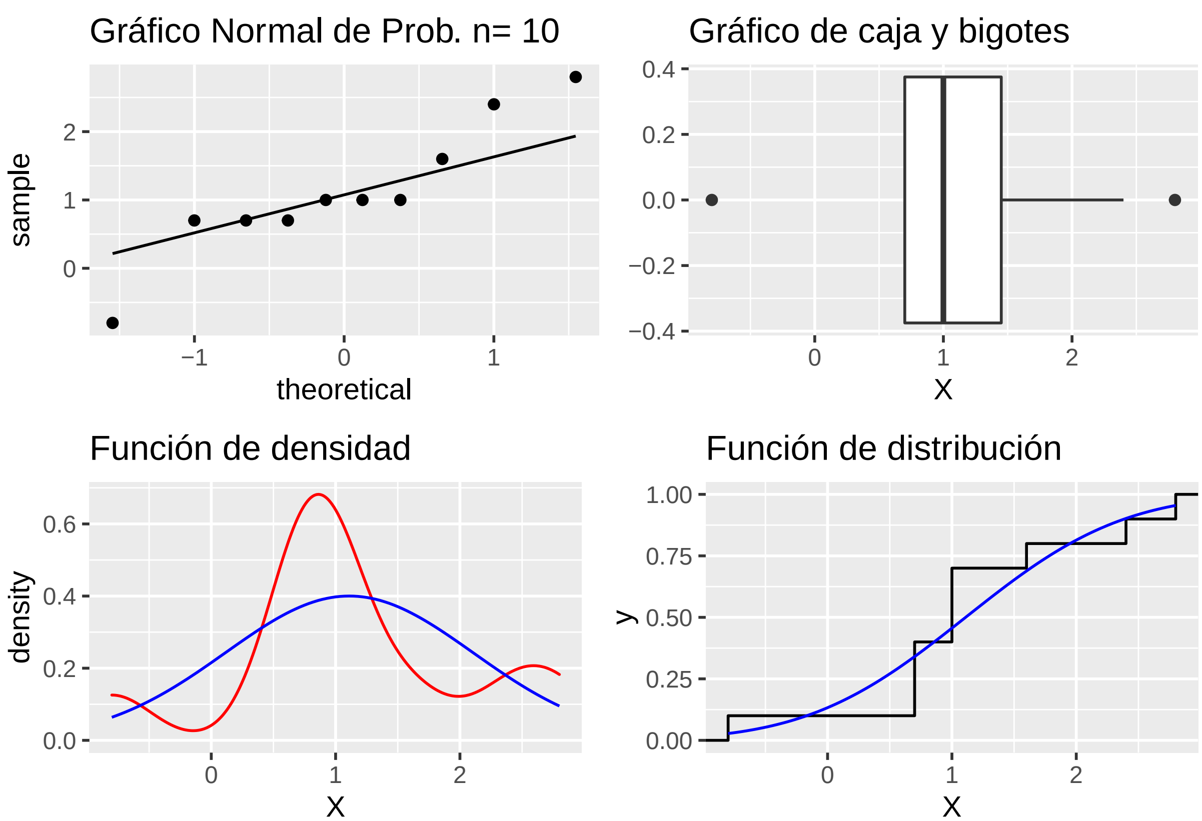
O lo que es equivalente:

$$H_0 : \mu_x - \mu_y > 0$$
$$H_1 : \mu_x - \mu_y \leq 0$$

```
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
```



No hay observaciones outliers y hau algunas salidas que hacen que el p-valor no sea lo suficientemnete grande pero no las suficientes para rechazar la normalidad de la variable diferencia.

```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
```

Rechazo H_0, el segundo jurado puntúa más alto que el 1.

Es lo mismo que:

```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
```

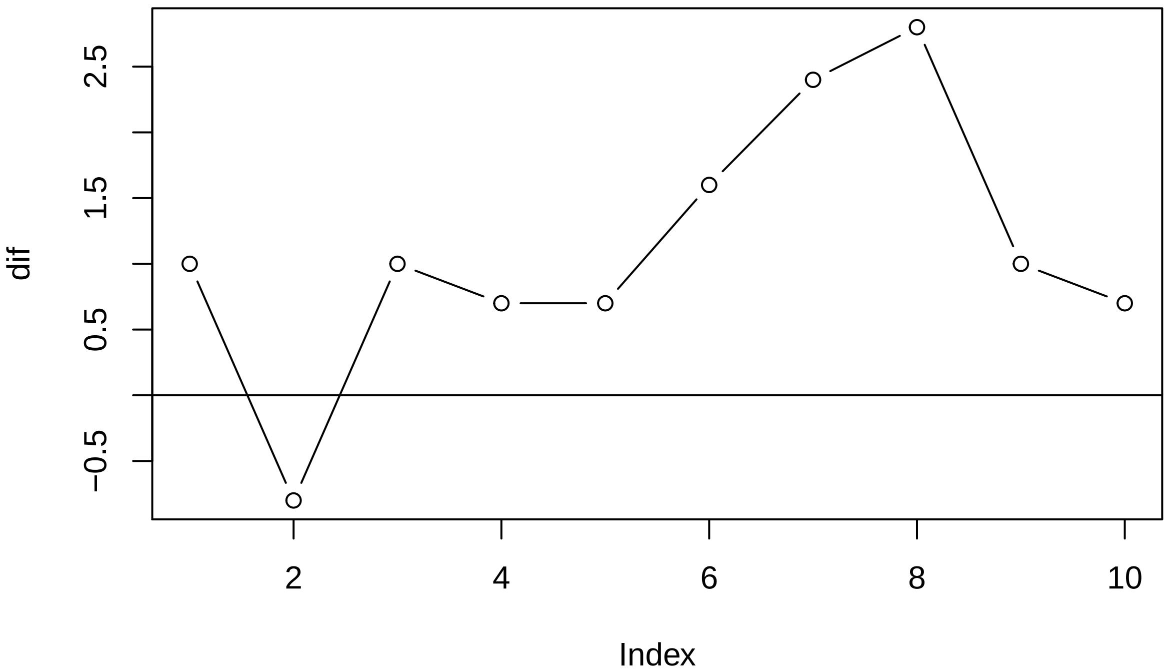
El p-valor < 0.05, por lo que rechazo la hipótesis nula. Por lo que la media de la variable x es mayor. El primer jurado puntúa más alto qu el 2.

4. Ejercicio 4

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

Dibujamos la variable dif

```r
plot(dif, type="b")
abline(h=0, col="black")
```



Tendríamos la línea en el 0, que es donde estarían todos los puntos si la puntuación hubiera sido la misma.

4.1. Comparación de medias en muestras relacionadas NO NORMALES- TEST DE WILXCONSON

Si hubiera fallado la normalidad, aplicaríamos el test de Wilxconson 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
```

Emplea la correción de continuidad, ya estamaos en la normal. De nuevo rechazamos la hipótesis nula y concluímos con que el jurado 2 está puntuando con más nota que el 1.

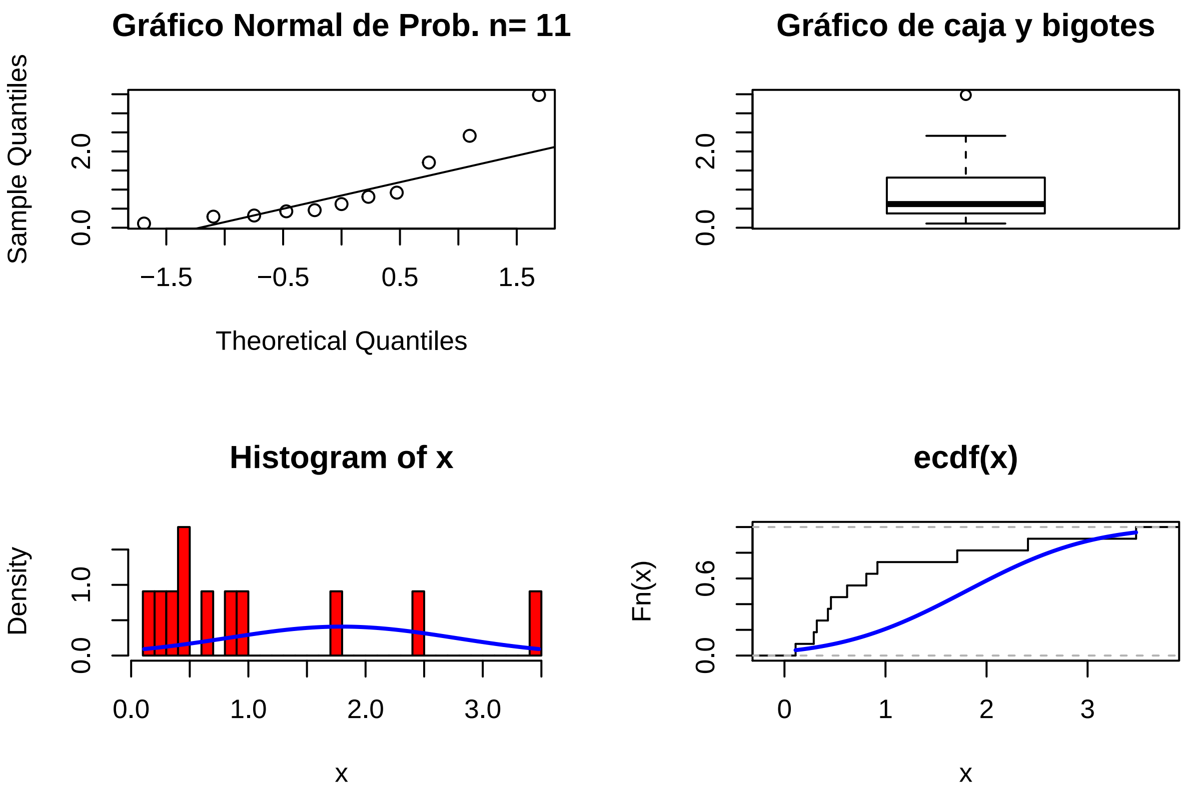
5. Ejercicio 5: Muestras independientes

```
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.

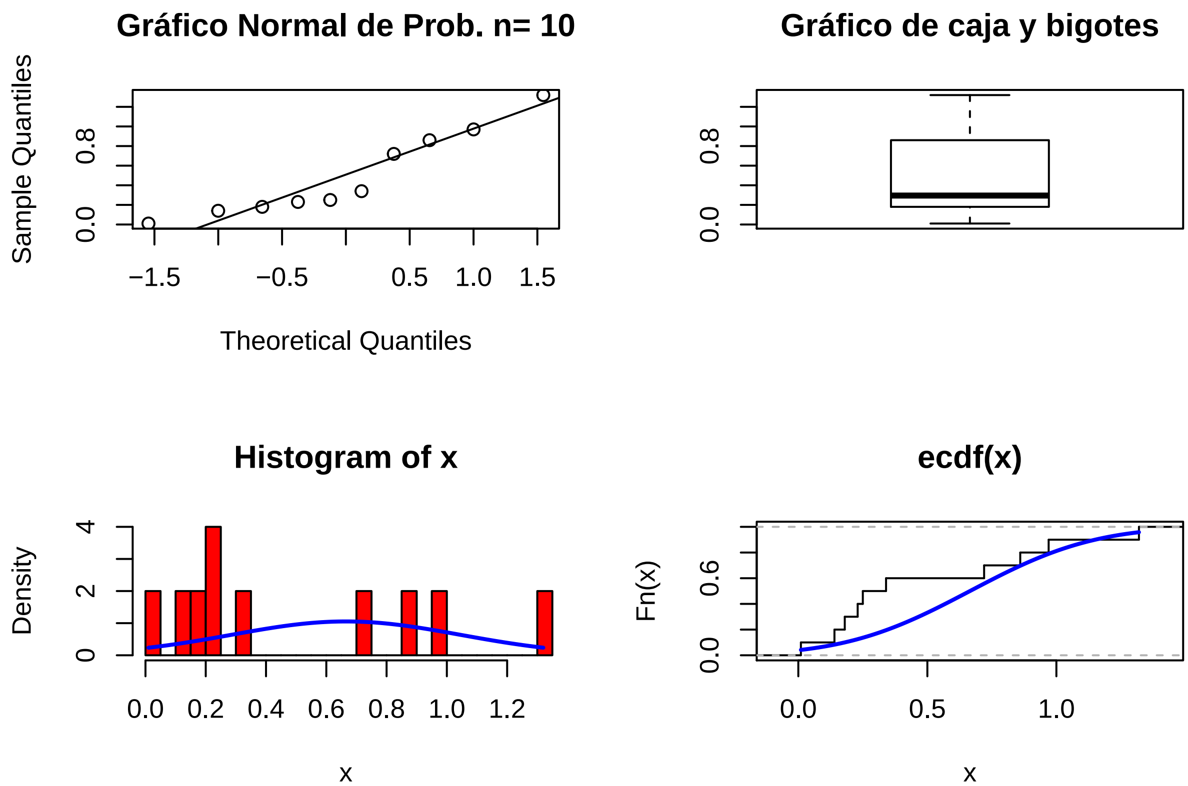
5.1. Estudiamos la normalidad

```
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
```

La variable X no sigue una normal mientras que Y si podemos aceptar que se distribuye según una normal.

5.2. Test de Mann-Whitney o Mann-Whitney-Wilconxon (igualdad de medias)

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

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

```
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
```

Acepto la igualdad de medias (medianas)