Inferencia para una muestra

Marta Venegas Pardo

Contents

Introducción

Sea el siguiente contraste de hipótesis, para un nivel de confianza α:

$$H_0 : X \sim N(\mu, \sigma^2) H_1 : X \neq N(\mu, \sigma^2)$$

Test Shapiro-Wilk:

En el paquete base con la instrucción **shapiro.test**

Creamos una variable aleatoria:

```
set.seed(12345)
x=rnorm(15,0,1)
x
```

```
##  [1]  0.5855288  0.7094660 -0.1093033 -0.4534972  0.6058875 -1.8179560
##  [7]  0.6300986 -0.2761841 -0.2841597 -0.9193220 -0.1162478  1.8173120
## [13]  0.3706279  0.5202165 -0.7505320
```

Test de Shapiro Wilk para ver si mi variable aleatoria sigue una normal.

```
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.96152, p-value = 0.7189
```

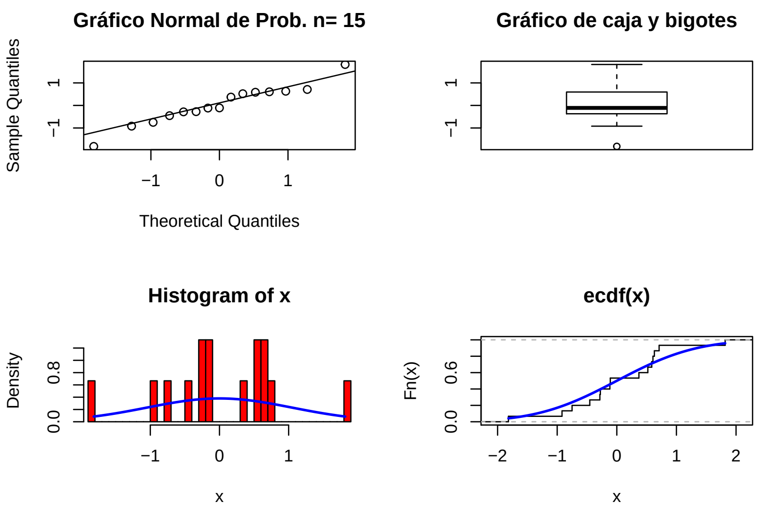
Tenemos que el p-valor=0.7189 por lo que acepto la hipótesis nula, es decir, los datos provienen de una variable aleatoria con distribución normal.

Acepto H_0

Estudio de normalidad gráficamente

Análisis de la normalidad una muestra

```
ananor(x)
```

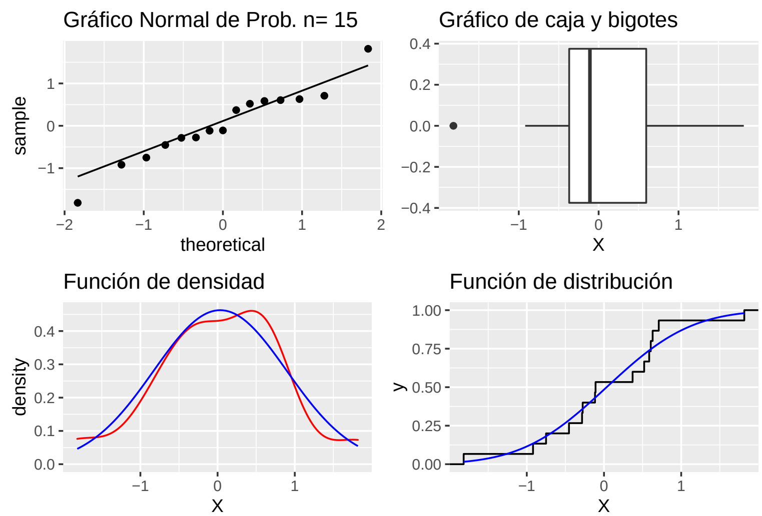



```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.96152, p-value = 0.7189
```

Ahora con el paquete tidyverse

```
ananor_tidy(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.96152, p-value = 0.7189
```



Test de normalidad en el paquete fBasics

Instalamos la librería.

```
library(fBasics)
```

Test de normalidad:

```
ksnormTest(x)
```

```
##
## Title:
##  One-sample Kolmogorov-Smirnov test
##
## Test Results:
##   STATISTIC:
##     D: 0.1724
##   P VALUE:
##     Alternative Two-Sided: 0.7025
##     Alternative      Less: 0.5779
##     Alternative   Greater: 0.3683
##
## Description:
##  Fri Apr 30 11:04:26 2021 by user:
```

Aceptamos H_0 en todos los casos, la distribución de la variable X se aproxima a una Normal Univariante.

Test de Shapiro-Wilk

```
shapiroTest(x)
```

```
##
## Title:
##  Shapiro - Wilk Normality Test
##
## Test Results:
##   STATISTIC:
##     W: 0.9615
```

```
##   P VALUE:
##     0.7189
##
## Description:
##  Fri Apr 30 11:04:26 2021 by user:
```

Test de Jaquerbera

```r
jarqueberaTest(x)
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 0.1005
##   P VALUE:
##     Asymptotic p Value: 0.951
##
## Description:
##  Fri Apr 30 11:04:26 2021 by user:
```

En el paquete *nortest* vienen recogidos otros test.

Test de normalidad en el paquete nortest

```r
library(nortest)
ad.test(x) # Anderson Darling
```

```
##
##  Anderson-Darling normality test
##
## data:  x
## A = 0.33942, p-value = 0.4491
```

```r
cvm.test(x)
```

```
##
##  Cramer-von Mises normality test
##
## data:  x
## W = 0.053087, p-value = 0.4423
```

```r
lillie.test(x)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  x
## D = 0.15004, p-value = 0.4832
```

Intervalo de confianza y Esperanza = 15

Para hacer contrastes paramétricos deberíamos hacer antes contrastes de normalidad previos.

Estudiar gráficamente la normalidad.

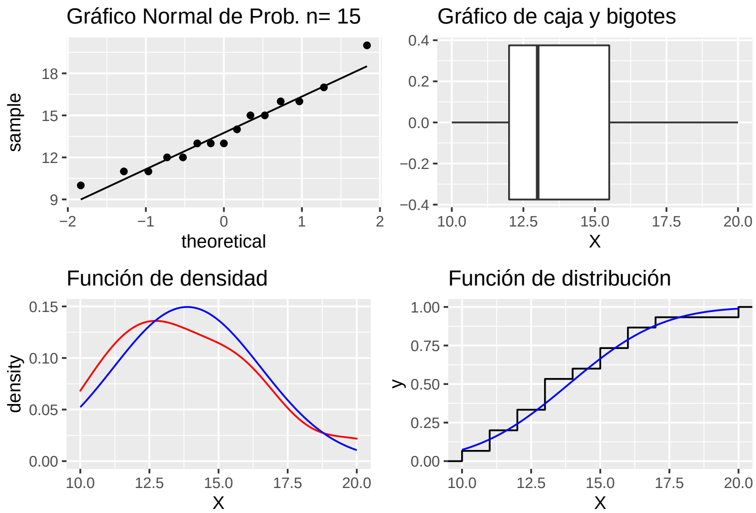
Tenemos 15 empresas y medimos a cada una de ellas la variable X=Gasto en publicidad en miles de euros. Construir un IC.

```r
x = c(17,12,15,16,15,11,12,13,20,16,14,13,11,10,13)
#length(x)
```

Compruebo la hipótesis de normalidad:

```r
ananor_tidy(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.95469, p-value = 0.6011
```



Ligera asimetría a la derecha pero no la suficiente para rechazar la normalidad.

Caja y bigotes no revela que haya observaciones outliers.

```r
summary(x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   12.00   13.00   13.87   15.50   20.00
```

$$H_0 : \mu = 15$$
$$H_a : \mu \neq 15$$

Se pueden utilizar test paramétricos:

```r
t.test(x)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 20.122, df = 14, p-value = 9.905e-12
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  12.38860 15.34474
```

```
## sample estimates:
## mean of x
##  13.86667
```

Si no ponemos nada, estudia si la media es 0 o no. Por ello aparece un p-valor muy pequeño. Debemos fijar $H_0 : \mu = 15$

Luego,

```
t.test(x, mu=15)
```

```
##
##  One Sample t-test
##
## data:  x
## t = -1.6446, df = 14, p-value = 0.1223
## alternative hypothesis: true mean is not equal to 15
## 95 percent confidence interval:
##  12.38860 15.34474
## sample estimates:
## mean of x
##  13.86667
```

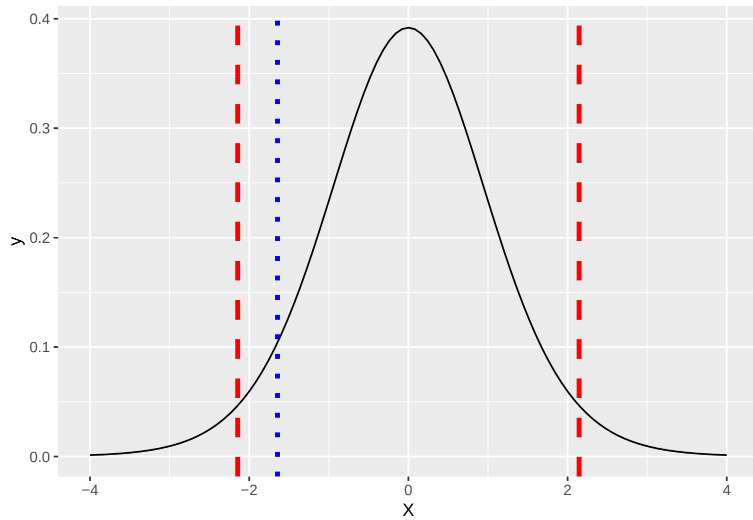
```
# alternative="less"
# alternative="greater"
```

Ejercicio 1

Dibujar la densidad de la t-Student bajo H_0, los cuantiles que definen los puntos críticos y el valor del estadístico t.

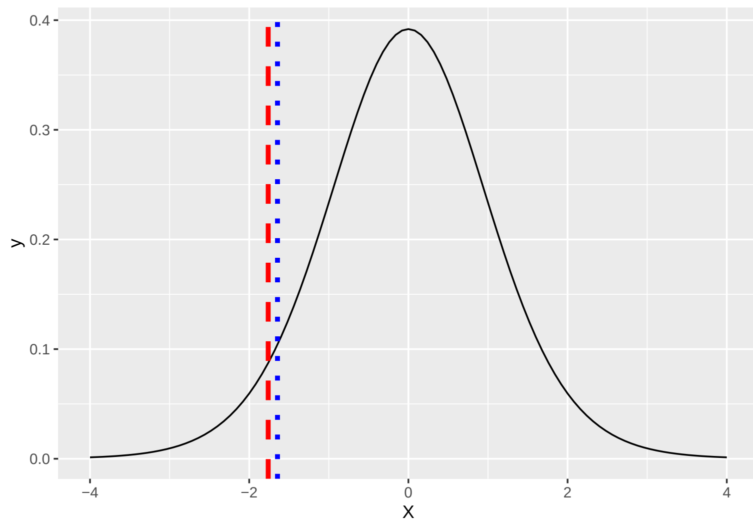
```
library(ggplot2)
funcion_g_dt_test=function(x,mu_=15){
#mu_=15
datos=data.frame(X=x)
resul=t.test(datos$X,mu=mu_)
  ggplot(datos,aes(x=X)) +
  stat_function(aes(x=seq(-4,4,length=length(X))),
              fun=dt,args = list(df=length(datos$X)-1),color="black")+
  geom_vline(aes(xintercept=resul$statistic),color="blue",
             linetype="dotted",size=1.4)+
  geom_vline(aes(xintercept=qt(0.025,df=resul$parameter)),color="red",
             linetype="dashed",size=1.4)+
  geom_vline(aes(xintercept=qt(0.975,df=resul$parameter)),color="red",
             linetype="dashed",size=1.4)
}
```

```
funcion_g_dt_test(x,mu_ = 15)
```



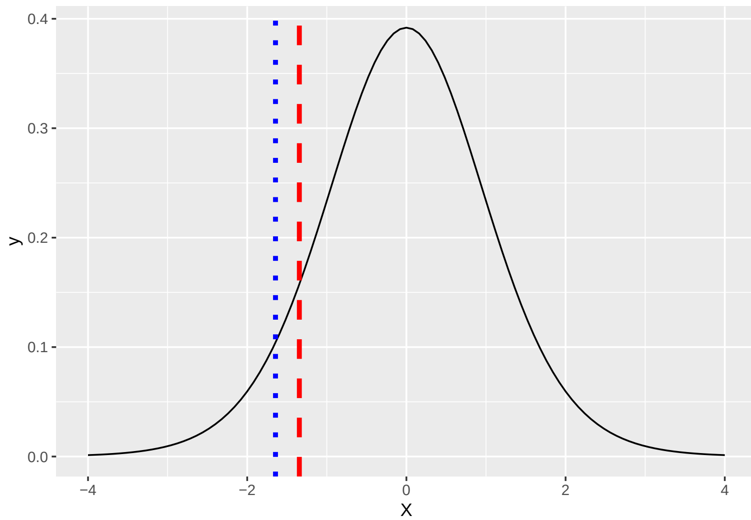
Ejercicio 2

```
func_g_dt_test_uni(x,mu_ = 15)
```




```
func_g_dt_test_uni(x,mu_ = 15, alternativa = "two.sided")
```

```
func_g_dt_test_uni(x,mu_ = 15, alternativa = "less",alpha = 0.10)
```



Ejercicio 3

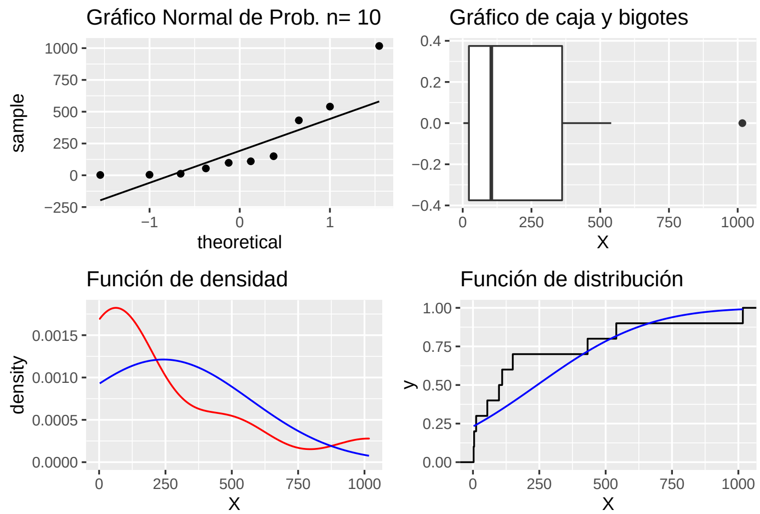
$$110, 12, 2.5, 98, 1017, 540, 54, 4.3, 150, 432$$

Se desea contrastar H_0 precio medio igual a 500

$$H_0 : \mu = 500$$
$$H_a : \mu \neq 500$$

```
x=c(110,12,2.5,98,1017,540,54,4.3,150,432)
ananor_tidy(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.76013, p-value = 0.004741
```



No podemos usar la estadística paramétrica para estudiar esta muestra, habrá que emplear técnicas no paramétricas, como el test de rangos signos de Wilcoxson.

```
wilcox.test(x,conf.int = TRUE,mu=500)
```

```
##
##  Wilcoxon signed rank test
##
## data:  x
## V = 11, p-value = 0.1055
## alternative hypothesis: true location is not equal to 500
## 95 percent confidence interval:
##   33.0 514.5
## sample estimates:
## (pseudo)median
##            150
```

El intervalo incluye al valor 500, el pvalor= 0.1055, acepto H_0

Apartado a

Calcular directamente W+ (test de rango-signo de Wilcoxon)

```
#W+=suma(rangos(|Xi|),xi>0)
# H0 = mu=0
mu_=500
rangos=rank(abs(x-mu_))
rangos[(x-mu_) >0]
```

```
## [1] 10  1
```

```
(est.w=sum(rangos[(x-mu_) >0]))
```

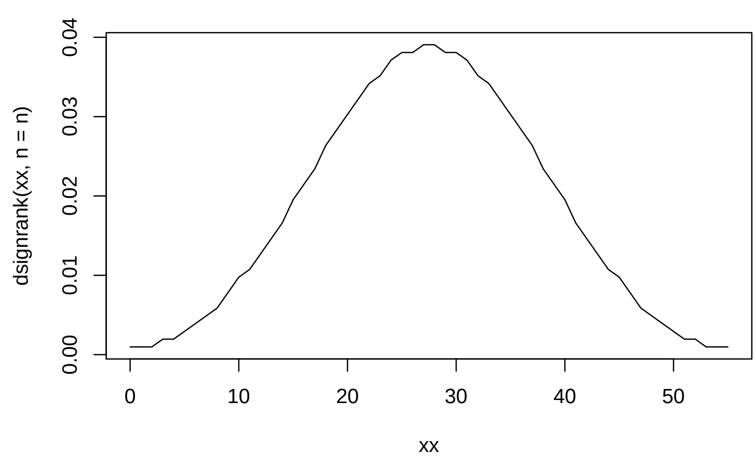
```
## [1] 11
```

Obtenemos el valor V del test anterior.

Apartado b

Dibujar la función de probabilidad (densidad) de W+ para esta n, usando *dsignrank*

```
#dsignrank()
n=length(x)
xx=seq(0,n*(n+1)/2,1)
plot(xx,dsignrank(xx,n=n),type="l")
```



Tengo en el eje X las x y en el Y la f(x).

Esperanza

```
# E[W+]
# Un estiamdor sería:
sum(xx*dsignrank(xx,n=n))
```

```
## [1] 27.5
```

```
n*(n+1)/4
```

```
## [1] 27.5
```

Varianza

```
sum(xx^2*dsignrank(xx,n=n))-(n*(n+1)/4)^2
```

```
## [1] 96.25
```

```
n*(n+1)*(2*n+1)/24
```

```
## [1] 96.25
```

Ejercicio 4

En este ejemplo se considera la hipótesis nula de que la progenie de un cruce de plantas produce como resultado plantas de tipo A o B con probabilidades respectivas 1/4 y 3/4.

En un experimento se obtienen 243 de tipo A y 682 de tipo B.

$$H_0 : p = \frac{3}{4}$$
$$H_1 : p \neq \frac{3}{4}$$

Con el test binomial, realizamos el contraste bilateral

```
binom.test(c(682,243),p=3/4)
```

```
##
##  Exact binomial test
##
## data:  c(682, 243)
## number of successes = 682, number of trials = 925, p-value = 0.3825
## alternative hypothesis: true probability of success is not equal to 0.75
## 95 percent confidence interval:
##  0.7076683 0.7654066
## sample estimates:
## probability of success
##              0.7372973
```

Acepto H_0, podemos afirmar que p=3/4.

Apartado a

Calcular el estadístico chi-cuadrado y comprobar que no coincide con Z.

```
n=682+243
pg=682/n
```

```r
Z=(pg-(3/4))/sqrt(0.75*(1-0.75)/n)
Z^2
```

```
## [1] 0.796036
```

```r
E0 = n*(1/4)
E1 = n*(3/4)
Ob0 = 243
Ob1 = 682
((E0-Ob0)^2)/E0 + ((E1-Ob1)^2)/E1
```

```
## [1] 0.796036
```

Apartado b

Calcular el estadístico chi-cuadrado con la corrección de Yates y comprobar que coincide con el estadístico que da prop.test.

```r
prop.test(x=682,n=682+243,p=3/4)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  682 out of 682 + 243, null probability 3/4
## X-squared = 0.72973, df = 1, p-value = 0.393
## alternative hypothesis: true p is not equal to 0.75
## 95 percent confidence interval:
##  0.7074391 0.7651554
## sample estimates:
##         p
## 0.7372973
```

Este contraste usa la aproximación normal. Aceptamos H_0

Con la corrección de Yates:

```r
((abs(E0-Ob0)-0.5)^2)/E0 + ((abs(E1-Ob1)-0.5)^2)/E1
```

```
## [1] 0.7297297
```

Apartado c

FALTA Escribir el test

```r
prop.test(35,80,p=0.8,alternative = "less",conf.level = 0.95)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  35 out of 80, null probability 0.8
## X-squared = 63.457, df = 1, p-value = 8.195e-16
## alternative hypothesis: true p is less than 0.8
## 95 percent confidence interval:
##  0.0000000 0.5354685
## sample estimates:
##      p
## 0.4375
```