Ejercicios

Test Binomial, Test X^2, Test K-S y Prueba de rachas

Joaquin Cavieres G.

I. Test Binomial

Ejercicio 1

En una determinada empresa automotora de Valparaíso se registró la venta de 58 vehículos durante el último mes. De los registros obtenidos se observó que 35 vehículos fueron de la marca Ford y el resto de vehículos de la marca Honda. Según estudios previos este tipo de marcas tienen la misma proporción de ventas durante un mes en particular en la región. De lo anterior, ¿Existe evidencia significativa para determinar que la venta de estas marcas no presentan la misma proporción?

Desarrollo

Nosotros estamos interesados en probar la siguiente hipótesis:

$$\mathtt{H}_0 : p = p_0 \qquad \boldsymbol{vs} \qquad \mathtt{H}_1 : p \neq p_0$$

con un nivel de signifcancia $\alpha = 0{,}05$.

a) Plantee el contraste de hipótesis a 'mano'.
b) Plantee el desarrollo análitico mediante R

```r
n = 58          # Venta de vehículos totales
x = n - 35      # Venta de vehiculos Honda
p = 0.5         # Probabilidad a priori asumida en las ventas
pbinom(x, n, prob = p) # Binomial Acumulado (prueba aproximada)
```

```
## [1] 0.0740029
```

Respuesta

El resultado del test Binomial indica que el p-value (0.0740029) $>$ α (0.05), por tanto, No Rechazo H_0. Lo anterior indica que la proporción de venta de automóviles marca Ford y Honda es la misma.

Ejercicio 2

Se probó un nuevo fertilizante en las plantaciones de Palta en una determinada zona. Del total de 1000 Paltos, se utilizó el fertilizante en 500 ejemplares, y luego de un tiempo el 12 % de estos Paltos murieron mediante causas indeterminadas. Mediante estudios previos realizados en diferentes países productores de Paltas, cuando se prueba un nuevo fertilizante el % de aceptación de vida sobre los Paltos debe alcanzar sobre el 85 %, sin embargo en Chile este número se relaja un poco para un porcentaje de aprobación del fertilizante de un 80 % . Se quiere probar si el nuevo fertilizante es efectivo y no presenta problemas en el cultivo de esta fruta.

a) Plantee el contraste de hipótesis a 'mano'.
b) Plantee el desarrollo análitico mediante R
c) Determine los intervalos de confianza

Desarrollo

Nosotros estamos interesados en probar la siguiente hipótesis:

$$\texttt{H}_0 : \text{Se acepta el nuevo fertilizante} \qquad \boldsymbol{vs} \qquad \texttt{H}_1 : \text{No se acepta el nuevo fertilizante}$$

```r
n = 500
x = n - 60
p = 0.80
q = 1 - p
dbinom(x, n, prob = p) # Prueba exacta
```

```
## [1] 6.872639e-07
```

Aproximando mediante z

```r
media = n*p
varianza = n*p*q
if(x > media){
   Z <- ((x - 0.5)-media)/sqrt(varianza)
}else Z <- ((x + 0.5)-media)/sqrt(varianza)
p_value <- pnorm(Z, lower.tail=FALSE)
p_value
```

```
## [1] 5.021762e-06
```

Respuesta

El resultado del test Binomial indica que el p-value (5.021762e-06) $< \alpha$ (0.05), por tanto, se Rechaza H_0. Lo anterior indica que el nuevo fertilizante no se puede aplicar en las plantaciones de Palto ya que no supera los estandares mínimos impuestos.

II. Test X^2

Ejercicio 3

Existen 4 tipos de computadores disponibles a la venta en una tienda comercial. Las cantidades asociadas a estos computaodres tienen la misma proporción inicialmente pero se observó que la tendencia cambio a la siguiente forma: PC1 = 300, PC2 = 350, PC3 = 450, PC4 = 550. Determine si las ventas difieren significativamente de lo presupuestado inicialmente por la tienda comercial con un $\alpha = 0{,}07$

a) Plantee el contraste de hipótesis a 'mano'.
b) Plantee el desarrollo análitico mediante R

Desarrollo

```r
O_i = c(300, 350, 450, 550)  # Frecuencias observadas
E_i = c(1/4, 1/4, 1/4, 1/4)  # Frecuencias esperadas
chisq.test(O_i, correct=FALSE, p=E_i)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  O_i
## X-squared = 89.394, df = 3, p-value < 2.2e-16
```

Respuesta

Dado que el p-value (2.2e-16) $< \alpha$ (0.07) entonces Rechazamos H_0 (con un nivel de significancia del 7 %). Esto significa que los números de ventas de computadores vendidos difieren significativamente a las presupuestadas inicialmente.

Ejercicio 4

En un supermercado los productos liquidos con mayores ventas son los siguientes: cervezas, vinos, bebidas y jugos. Durante los años previos la venta de los jugos y bebidas eran bastante populares y sus ventas eran mayores que las de los otros dos liquidos, sin embargo, durante los últimos años las cervezas y vinos han alcanzado casi la misma cantidad de ventas que bebidas y jugos. Dado lo anterior es que se propone determinar si la demanda de estos productos es uniforme o no. El registro de ventas durante el último mes es el siguiente: Cervezas = 150, Vinos = 120, Bebidas = 115, Jugos = 150.

a) Plantee el contraste de hipótesis a 'mano'.
b) Plantee el desarrollo análitico mediante R

Desarrollo

Nosotros estamos interesados en probar la siguiente hipótesis:

$$\mathtt{H}_0 : \text{Las ventas de los 4 productos es uniforme} \qquad \boldsymbol{vs} \qquad \mathtt{H}_1 : \text{Las ventas de los 4 productos no es uniforme}$$

```r
# Calculo 'a mano'
n = 535
O_i = c(150, 120, 115, 150)      # Frecuencias observadas
E_i = c(0.25, 0.25, 0.25, 0.25)  # Frecuencias esperadas
E = n*E_i

# Estadístico T
T_test = sum((O_i - E)^2/ E)
T_test
```

```
## [1] 7.990654
```

```r
# Calculo manual del p-value con 4 - 1 grados de libertad (k -1)
pchisq(c(T_test), df=3, lower.tail=FALSE)
```

```
## [1] 0.04620525
```

Ahora haremos el cálculo mediante `R`

```r
chisq.test(O_i, correct = FALSE, p = E_i)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  O_i
## X-squared = 7.9907, df = 3, p-value = 0.04621
```

Respuesta

Dado que el p-value (0.04621) $< \alpha$ (0.05) entonces Rechazamos H_0 (con un nivel de significancia del 5 %). Esto significa que los números de ventas los productos no presentan una distribució uniforme entre ellos.

III. Test Kolgomorov-Smirnov (K-S)

Ejercicio 5

Se quiere contrastar la hipotesis nula H_0, la cual considera que $F = F_0$, con una hipótesis alternativa la cual esta definida como $F \neq F_0$. Asuma que se tiene una muestra aleatoria con distribución $\mathcal{N} \sim (0, 1)$ la cual desea ser contrastada.

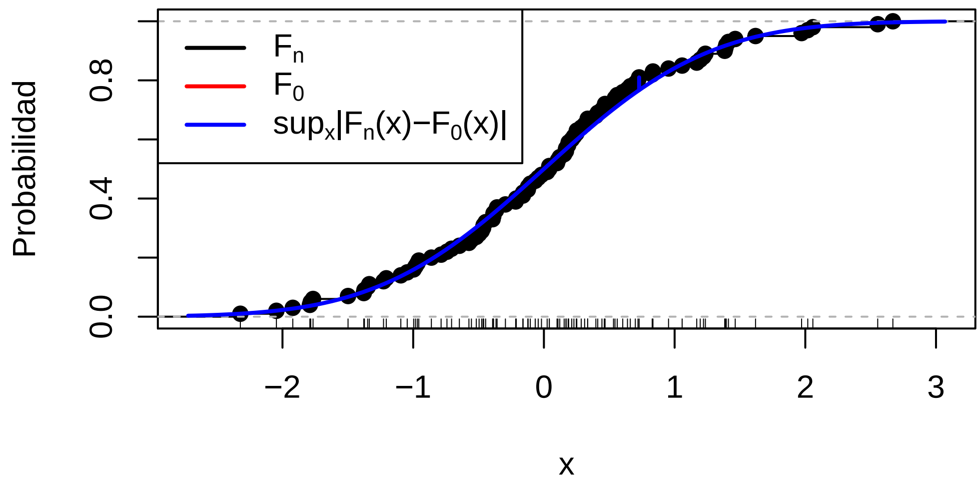
a) Plantee el contraste de hipótesis a 'mano'.
b) Plantee el desarrollo análitico mediante R

Desarrollo

Nosotros estamos interesados en probar la siguiente hipótesis:

$$\mathtt{H}_0 : F = F_0 \qquad \boldsymbol{vs} \qquad \mathtt{H}_1 : F \neq F_0$$

```r
set.seed(1000) # Para que sea reproducible el ejercicio
n = 100
mu = 0
sd = 1
sample = rnorm(n, mu, sd) # muestra con 100 observaciones, mu = 0 y sd = 1
# Calculamos la ecdf y luego la cdf
plot(ecdf(sample), main = "", ylab = "Probabilidad")
curve(pnorm(x, mean = mu, sd = 1), add = TRUE, col = "blue", lwd = 2)
# Calculamos la máxima distancia ("a mano")
sample_ordenada <- sort(sample)
Ui <- pnorm(sample_ordenada, mean = mu, sd = sd)
Dn_sup <- (1:n) / n - Ui
Dn_inf <- Ui - (1:n - 1) / n
i <- which.max(pmax(Dn_sup, Dn_inf))
lines(rep(sample_ordenada[i], 2),
c(i / n, pnorm(sample_ordenada[i], mean = mu, sd = sd)),
col = 4, lwd = 2)
rug(sample)
legend("topleft", lwd = 2, col = c(1:2, 4),
legend = latex2exp::TeX(c("$F_n$", "$F_0$", "sup_x|F_n(x)-F_0(x)|")))
```




```r
D.n = max(Dn_sup,Dn_inf)
D.n
```

```
## [1] 0.04317635
```

Calculamos la disancia máxima mediante la función de `R`

```r
# Tamaños de la muestra
n = length(sample)
# Ordenamos la muestra
x = sort(sample)
mu = 0
sigma = 1
ks.test(x, pnorm, mean = mu, sd = sigma, exact = TRUE)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.043176, p-value = 0.9883
## alternative hypothesis: two-sided
```

Respuesta

Dado que el p-value (0.9883) $> \alpha$ (0.05) entonces No Rechazamos H_0 (con un nivel de significancia del 5 %). Esto significa que la distribución muestral es igual a la funcion de distribución de referencia.

II. Test de rachas (Wald-Wolfowitz)

Ejercicio 6

¿Se puede considerar que el número de goles hechos por cada fecha del campeonato de fútbol en Inglaterra es aleatorio o pueden provenir desde un patrón que no estamos observando? Los resultados del promedio de goles a lo largo de 10 fechas jugadas son los siguientes: 49, 49, 51, 50, 49, 49, 49, 50, 50, 52

Desarrollo

Nosotros estamos interesados en probar la siguiente hipótesis:

$$\mathtt{H}_0 : \text{Los goles presentan aleatoriedad} \qquad \boldsymbol{vs} \qquad \mathtt{H}_1 : \text{Los goles no presentan aleatoriedad}$$

```r
library(tseries)
goles = c(49, 49, 51, 50, 49, 49, 49, 50, 50, 52)
# Calculamos la mediana
median = median(goles)
# Hacemos el test de rachas
library(tseries)
runs.test(as.factor(goles > median))
```

```
##
##  Runs Test
##
## data:  as.factor(goles > median)
## Standard Normal = -1.3416, p-value = 0.1797
## alternative hypothesis: two.sided
```

Respuesta

Dado que el p-value (0.1797) $> \alpha$ (0.05) entonces No Rechazamos H_0 (con un nivel de significancia del 5 %). Esto significa que la distribución de los goles si tienen un comportamiento aleatorio fecha a fecha.

Ejercicio 7

Se quiere evaluar si los números provenientes de una tabla con valores de 1 a 10 son aleatorios. Los resultados muestran la siguiente frecuencia para cada uno de estos valores: $1 = 11$, $2 = 10$, $3 = 12$, $4 = 10$, $5 = 17{,}6 = 26$, $7 = 33$, $8 = 43$, $9 = 27$, $10 = 12$.

Desarrollo

Nosotros estamos interesados en probar la siguiente hipótesis:

$$\mathtt{H}_0 : \text{Los números obtenidos son aleatorios} \qquad \boldsymbol{vs} \qquad \mathtt{H}_1 : \text{Los números obtenidos no son aleatorios}$$

```r
library(tseries)
frecuencia = c(11,  10, 12, 10, 17, 26, 33, 43, 27, 12)
# Calculamos la mediana
median = median(frecuencia)
# Hacemos el test de rachas
runs.test(as.factor(frecuencia > median))
```

```
##
##  Runs Test
##
## data:  as.factor(frecuencia > median)
## Standard Normal = -2.0125, p-value = 0.04417
## alternative hypothesis: two.sided
```

Respuesta

Dado que el p-value $(0.04417) < \alpha$ (0.05) entonces Rechazamos H_0 (con un nivel de significancia del 5 %). Esto significa que la frecuencia de los números obtenidos de una tabla con valores de 1 a 10 no son aleatorios.