Ejercicios 2

Estadística No Paramétrica

Joaquin Cavieres G.

Test Binomial

Este tipo de test está basado en que el estadístico de prueba utilizado para contrastar la hipótesis nula $\mathtt{H_0}$ sigue una distribución Binomial.

Considere a $X_1, ..., X_n$ una variable aleatoria proveniente de algún fenómeno en estudio y que sólo admite dos posibles resultados. Si el resultado es positivo entonces la probabilidad asociada es p y si el resultado es negativo la probabilidad asociada es $1-p$. p se puede definir como el parámetro asociado a la proporción (o probabilidad) con que observamos X_i para cualquier $i = 1, ..., n$.

a) **Prueba de una cola**

Suponga que estamos interesados en probar las siguientes hipótesis:

$$\mathtt{H_0} : p = p_0 \qquad \boldsymbol{vs} \qquad \mathtt{H_1} : p > p_0$$

o queremos probar lo siguiente:

$$\mathtt{H_0} : p \leq p_0 \qquad \boldsymbol{vs} \qquad \mathtt{H_1} : p > p_0$$

Como podemos observar en ambos casos, nosotros queremos evaluar que $p > p_0$ y lo único que esta cambiando es la hipótesis nula, ya que en el primer caso $p = p_0$ y en el segundo caso $p \leq p_0$. Sin embargo, nuestro análisis debe centrarse en proporcionar información empírica para rechazar $\mathtt{H_0}$ dado que $p > p_0$. Por lo tanto queremos probar que el verdadero p efectivamente sea mayor que p_0 propuesto. Establezcamos el estadístico de prueba determinado por:

$$F = \sum_{i=1}^{n} \mathbf{1}_{(X_i \in C_1)} > w_{1-\alpha} \tag{1}$$

donde $w_{1-\alpha}$ es el cuantil $1-\alpha$ de la distribución Binomial(n, p_0).

Ejercicio

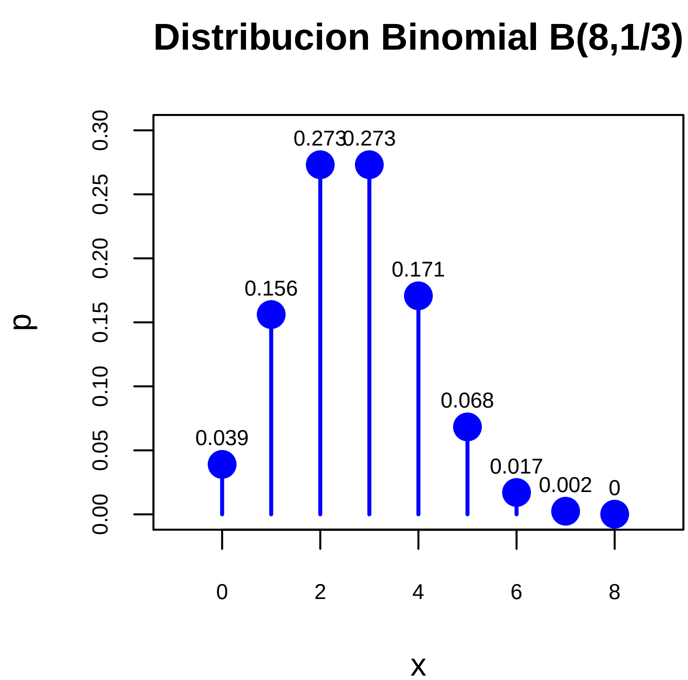
Compruebe la siguiente hipótesis planteada:

$$\begin{array}{ccc} \mathtt{H}_0 : p = \dfrac{1}{3} & \boldsymbol{vs} & \mathtt{H}_1 : p > \dfrac{1}{3} \\ \mathtt{H}_0 : p \leq \dfrac{1}{3} & \boldsymbol{vs} & \mathtt{H}_1 : p > \dfrac{1}{3} \end{array}$$

Desarrollo

En este caso estamos interesados en que el lado derecho acumule α de probabilidad, esto significa que iremos acumulando las probabilidades de derecha a izquierda hasta llegar al valor deseado de α.

```r
n = 8    # número de observaciones
p = 1/3   # Proporción (probabilidad)
x = 0:8
p = dbinom(x, size = n, prob = p)
plot(x,p,type="h",xlim=c(-1,9),ylim=c(0,0.3),lwd=2,col="blue",ylab="p",
main="Distribucion Binomial B(8,1/3)",cex.axis=0.7)
points(x,p,pch=16,cex=2,col="blue")
text(x,p,round(p,3),pos=3,cex=0.7)
```



El gráfico anterior podemos expresarlo en una tabla númerica para una mejor comprensión:

```r
data_prob = as.data.frame(cbind(x,round(p,4)))
colnames(data_prob) = c("F","Pr")
data_prob
```

```
##   F     Pr
## 1 0 0.0390
## 2 1 0.1561
## 3 2 0.2731
## 4 3 0.2731
## 5 4 0.1707
## 6 5 0.0683
## 7 6 0.0171
## 8 7 0.0024
## 9 8 0.0002
```

Ya que necesitamos obtener las probabilidades de derecha a izquierda y, como estamos trabajando con una variable aleatoria discreta, el valor acumulado de α en el cual estamos interesados es de $F = 5$. ¿Por que cree que se considera este valor?

$$\mathbb{P}(F = 8) + \mathbb{P}(F = 7) + \mathbb{P}(F = 6) + \mathbb{P}(F = 5) = 0{,}0898$$

Para este caso particular tenemos que con un $\alpha = 0{,}005$ se rechaza $\mathtt{H}_0$ si $F > 5$.

Ya que hemos sobrepasado el α deseado acumulado hasta $F = 6$ dado por:

$$\mathbb{P}(F = 8) + \mathbb{P}(F = 7) + \mathbb{P}(F = 6) = 0{,}0215$$

la prueba en este caso tendría una significancia del $\alpha = 0{,}0215$ con un cuantil $w_{1-\alpha}$ de 5 y expresado como:

$$F > w_{1-\alpha} = 5$$

Por otra parte, también podemos hacer la prueba para la otra cola, esto sería:

$$\begin{array}{ccc} \mathtt{H}_0 : p = p_0 & \boldsymbol{vs} & \mathtt{H}_1 : p < p_0 \\ \mathtt{H}_0 : p \geq p_0 & \boldsymbol{vs} & \mathtt{H}_1 : p < p_0 \end{array}$$

Debemos repetir el proceso de la misma manera en como lo hicimos previamente pero ahora debemos acumular las probabilidades de la cola izquierda y rechazar $\mathtt{H}_0$ si:

$$F = \sum_{i=1}^{n} \mathbf{1}_{(X_i \in C_1)} \leq w_\alpha$$

b) **Prueba de dos colas**

Para evaluar este tipo de problemas (tambien conocido como “prueba de colas en ambos lados”) primero planteamos la hipótesis:

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

y de igual manera, al igual que en la prueba de una cola, poodemos definir el estadístico de prueba como:

$$F = \sum_{i=1}^{n} \mathbf{1}_{(X_i \in C_1)} = \text{número de observaciones en } C_1 \tag{2}$$

El estadístico de prueba $F = \sum_i^n \mathbf{1}_{X_i \in C_1}$ es la suma de F variables aleatorias Bernoulli, por tanto $F \sim Binomial(n, p_0)$, entonces si $\mathtt{H}_0$ es verdadera, nosotros esperamos que F se concentre en la parte densa de la densidad Binomial. Al contrario, si F toma valores pequeños o grandes, entonces rechazamos $\mathtt{H}_0$ con un nivel de significancia α si:

$$F \leq w_{\alpha_1} \quad o \quad F > w_{1-\alpha_2} \tag{3}$$

donde $\alpha_1 + \alpha_2 = \alpha$. Este es un caso en donde encontrar el valor de α no es simple ya que es una variable aleatoria discreta, por tanto encontrar una significancia exacta a α es casi imposible, es por eso que vamos a encontrar los cuantiles tales que $\alpha_1 + \alpha_2 = \alpha_0 \leq \alpha$. En esta notación α_0 es la probabilidad de cometer el error tipo I mas cercano inferiormente a α. Si la distribución fuese simétrica ($p_0 = 1/2$) sería sencillo encontrar α ya que $\alpha_1 = \alpha_2 = \alpha/2$ pero este no es el caso.

Ejercicio

Se tiene una variable aleatoria $X_1, ..., X_8$ desde un experimento en el cual existen sólo dos posibles resultados tal que $\mathbb{P}(X_i \in C_1) = p$.

Desarrollo

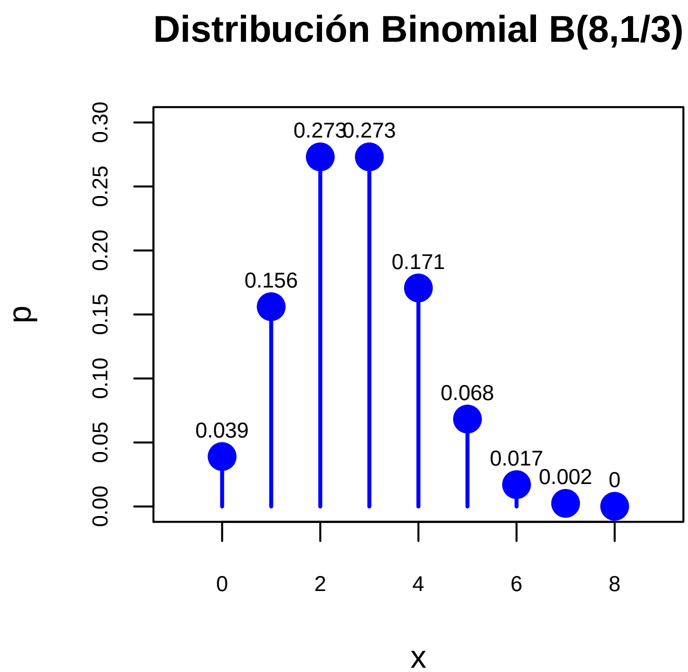
Hipótesis:

$$\mathtt{H}_0 : p = \frac{1}{3} \qquad \boldsymbol{vs} \qquad \mathtt{H}_1 : p \neq \frac{1}{3}$$

Considerando nuestra hipótesis nula $\mathtt{H}_0$ entonces F tiene una distribución Bernoulli con $n = 8$ y $p = 1/3$.

```r
n = 8     # número de observaciones
p = 1/3    # Proporción (probabilidad)
x = 0:8
p = dbinom(x, size = n, prob = p)
```

```r
plot(x, p, type="h", xlim=c(-1,9), ylim=c(0,0.3), lwd=2,
                     col="blue",ylab="p",
main="Distribución Binomial B(8,1/3)",cex.axis=0.7)
points(x,p,pch=16,cex=2,col="blue")
text(x,p,round(p, 3), pos=3,cex=0.7)
```



Si queremos rechazar $\mathtt{H}_0$ con un nivel de signifcancia de α primero debemos encontrar α_1 y α_2.

El gráfico nos muestra que la moda de los datos se encuentra entre $F = 2$ y $F = 3$, con una probabilidad conjunta de 0.546. Si continuamos este proceso hasta encontrar $1 - \alpha$ (considerando en este ejemplo un $\alpha = 0{,}05$), entonces debemos encontrar el acumulado de las probabilidades hasta el valor de 0.95. En ese caso el acumulado deberiamos considerarlo hasta $F \in \{1, ..., 5\}$, por lo tanto, $\mathbb{P}(F \in \{1, \ldots, 5\}) = 0.98$. Por tanto se escoge un $\alpha_1 = \mathbb{P}(F \in \{0\} = 0{,}039)$ y $\alpha_2 = \mathbb{P}(F \in \{6, 7, 8\} = 0{,}019)$. De acuerdo a lo anterior $w_{\alpha_1} = 0$, $w_{1-\alpha_2} = 5$ y rechazamos $\mathtt{H}_0$ si $F < 1 \quad o \quad F > 5$.

Intervalos de confianza

En este tipo de problemas (prueba de dos colas) podemos estimar los intervalos de confianza para p. Si recordamos el test paramétrico para la media de una distribución Normal, generalmente:

$$\mathtt{H}_0 : \mu = \mu_0 \qquad \text{vs} \qquad \mathtt{H}_1 : \mu \neq \mu_0$$

para encontrar una región de rechazo a través de un intervalo de confianza para μ y luego verificar si μ_0 se encuentra dentro de ese intervalo. Dentro de los test no paramétricos esto estaría en función de encontrar los valores de p_0 en los cuales no se rechaza la hipótesis nula.

En este caso debemos ir buscando los distintos valores de p_0 al discretizar el intervalo (0,1) e ir comprobando en cuales de esos valores no se rechaza $\mathtt{H}_0$. Todos los valores de p_0 que cumplan con esta condición son los que formarán un intervalo de confianza. Considere que aquí F es fijo y el valor que va variando es p_0, lo que a su vez provoca la modificación la distribución asociada.

Ejercicio

Considere nuevamente las condiciones previamente vistas pero con un $F = 4$ fijo. Vamos a utilizar la función `binom.test` disponible en `R` la cual nos permite determinar la prueba exacta de una distribución Binomial con su respectivo intervalo de confianza.

Desarrollo

Hipótesis:

$$\mathtt{H}_0 : p = \frac{1}{3} \qquad \boldsymbol{vs} \qquad \mathtt{H}_1 : p \neq \frac{1}{3}$$

```r
alpha = 0.05
binom.test(4,8,1/3,alternative=c("two.sided"),conf.level=1-alpha)
```

```
##
##  Exact binomial test
##
## data:  4 and 8
## number of successes = 4, number of trials = 8, p-value = 0.4537
## alternative hypothesis: true probability of success is not equal to 0.3333333
## 95 percent confidence interval:
##  0.1570128 0.8429872
## sample estimates:
## probability of success
##                    0.5
```

En este caso el intervalo de confianza de $\alpha = 0{,}05$ para la probabilidad p es:

$$(0{,}1570128, 0{,}8429872)$$