Exercise Inference 27-11

Lais Silva Almeida Zanchetta, Daniel Ferreira Zanchetta

11/27/2019

### 1) Enumere cuales son los elementos clave de una prueba de hipótesis y su significado.

Resp.: En primer lugar, siguiendo el principio de Karl Popper, una hipotesis solo puede ser entendida como cientifica si puede ser falsificable. Teniendo esto en cuenta, los elementos claves de una prueba de hipótesis son: - Null hypothesis: la hipótesis nula és la hipótesis contraria a que nos gustaría probar (que es la llamada alternative hypothesis). - Alternative hypothesis: es la hipotesis que se quiere probar. Es decir, és una variación de la hipotésis nula que se desea validar. - Test Statistic: es un valor estadístico adecuado para comparar la hipótesis nula con la hipótesis alternativa (variable aleatória) - Reference distribution: prueba estadística de la hipótesis nula, si la hipótesis nula és verdadera - p\_value: Probabilidad estadística de la prueba observada. Si el p\_value es muy bajo la hipótesis nula és rechazada (hasta los dias de hoy es considerada rechazada si el p\_value < 0,05)

### 2) Un elemento clave en toda prueba de hipótesis es la distribución de referencia del estadístico de la prueba. Diga de que maneras podemos conocer (o aproximar) la distribución de referencia para la prueba de hipótesis de querer ver si una media de una muestra es igual o no a un cierto valor nominal. Especifique como obtendría o cual sería esta distribución de referencia.

Resp.: Podemos conocer la distribución de referencia a través de (1) datos historicos, y tambien a través de (2) sin datos historicos. Para (1) obtendría una amuestra y calcularía su media. En seguida, calcularía cuantas amuestras son mas pequeñas o iguales al valor clave de la hipotesis. Para (2), se hace necesario crear una distribución de referencia haciendo suposición y utilizando la teoria. Tambien es calculada la media de la amuestra. En seguida debemos verificar la media de la amuestra en la distribución de referencia, y cuanto sería su p-value.

### 3) Un amigo me ofrece un piso por valor de 8MPts. Dejando de lado todos los otros factores intervinientes en la decisión de compra de un piso y solo teniendo en cuenta su precio, puedo considerar que se trata de ocasión para comprar?. Que suposiciones necesito hacer para resolver el problema. (Para resolver esta pregunta utilice los datos disponibles que crea conveniente).

Resp.: No se puede afirmar si se trata de ocasión para comprar simplemente con el nivel de detalle ofrecido. Es necesario considerar, al menos, tener una ditribución de referencia de los precios promedio de los pisos de donde el piso esta localizado, y tambien considerar con qué quieres hacer la comparación (por ejemplo, si quieres considerar otros pisos de la misma zona y entre zonas/comunidades distintas). Además de esto, tambien pueden ser necesarios mas detalles sobre el piso en si, como por ejemplo la cantidad de habitaciones, planta, garaje, etc… Teniendo esto en cuenta, si tomamos, por ejemplo, el dataset bcn\_pisos facilitado en la clase anterior (para hacer distribución de referencia), llegamos a un valor de p-value de 5,9% para los pisos que tienen Valor más pequeños que 8MPts. Para una hipotesis, como la siguiente: 1) H0: el piso no puede ser considerado que se trata de ocasión de compra 2) H1: el piso si que puede ser considerado que se trata de ocasión de compra

Teoricamente, con el p-value de 5,9% se puede decir que la hipotesis nula sea rechazada, pues entaria entre los pisos con mejoro valor de la distribución.

bcnpisos<-read.csv("C:/Users/Daniel/Documents/Certificados & Faculdade/UPC Master Big Data/Data Analytics/Aula 2 - 13-11/exer\_Descr/bcn\_pisos.txt", header=TRUE,sep="\t")  
bcnpisosnotdupl<-bcnpisos[!duplicated(bcnpisos),]  
nrow(bcnpisosnotdupl)

## [1] 2254

nrow(bcnpisos)

## [1] 2329

bcnpisoslowerprice<-nrow(bcnpisosnotdupl[bcnpisosnotdupl$Valor<8000000,])  
pvalue<-bcnpisoslowerprice/nrow(bcnpisosnotdupl)  
pvalue

## [1] 0.05989352

### 4. Sabemos que la media de los precios de los pisos (de 3 habitaciones) en l’Eixample es de 16.81 y su desviación tipo es de 5.91. Suponiendo que la muestra obtenida en el ejercicio 1 (7.80, 12.60, 15.96, 13.50, 8.25, 31.29, 16.46) es aleatoria. ¿Podemos asegurar de que se trata de una muestra de pisos de l’Eixample?

Resp.: Con el resultado obtenido de 70,31% se acepta la hipotesis nula, entonces en este caso no se puede afirmar que se trata de una muestra de pisos de l’Eixample.

x1 <- c(7.80, 12.60, 15.96, 13.50, 8.25, 31.29, 16.46)  
  
mu0 <- 16.81  
sigma <- 5.916  
  
m<-mean(x1)  
n<-length(x1)  
  
std<-sd(x1)/sqrt(n)  
  
t<-(mu0 - m)/std  
pt(t,df=n-1)

## [1] 0.7039106

### 5. El siguiente año, los precios de una muestra aleatoria de pisos de 3hab. en l’Eixample han sido 13.57 14.80 22.36 29.29 22.70. Puedo afirmar de que no ha habido cambio de precio entre los dos años?

Considerando: H0 –> no hay cambio de precio entre los dos años H1 –> hay cambio de precio entre los dos años

Con el resultado de 67,32% no es posible rechazar la hipotesis nula, por lo tanto, no ha tenido cambio de precio entre los dos años (H0 no rechazada).

x1 <- c(7.80, 12.60, 15.96, 13.50, 8.25, 31.29, 16.46)  
x2 <- c(13.57,14.80,22.36,29.29,22.70)  
meanx1<-mean(x1)  
meanx2<-mean(x2)  
nx1 <- length(x1)  
nx2 <- length(x2)  
s\_pool<-sqrt(((nx1-1)\*var(x1))+((nx2-1)\*var(x2))/nx1+nx2-2)  
s\_pool

## [1] 20.0112

t<-(meanx1-meanx2)/(s\_pool\*sqrt((1/nx1)+(1/nx2)))  
pt(t,df=nx1+nx2-2,lower.tail = F)

## [1] 0.6732438

### 6. Calcula el p\_valor en para la misma prueba del problema anterior, usando el método de permutaciones.

x1 <- c(7.80, 12.60, 15.96, 13.50, 8.25, 31.29, 16.46)  
x2 <- c(13.57,14.80,22.36,29.29,22.70)  
x3<-c(x1,x2)  
x3

## [1] 7.80 12.60 15.96 13.50 8.25 31.29 16.46 13.57 14.80 22.36 29.29  
## [12] 22.70

dif\_per<-NULL  
for (i in 1:1000){  
 random<-sample(1:12,6)  
 dif\_per[i]<-mean(x3[random])-mean(x3[-random])  
}  
difmean<-mean(x2)-mean(x1)  
sum(dif\_per>=difmean)/length(dif\_per)

## [1] 0.123

p-value de uno de los intentos habia dado 11,9%.

### 7. Sabemos que la probabilidad de compra de un producto en el canal internet es de 0.02. En un mes se han conectado 2300 visitantes, de los cuales 94 han comprado nuestro producto, puedo pensar que ha habido un incremento en la probabilidad de compra por internet?

Considerando… H0 –> no ha habido un incremento en la probabilidad de compra H1 –> ha habido un incremento en la probabilidad de compra

p-value es muy pequeño, por lo tanto, rechazamos la H0. Entonces, ha habido un incremento en la probabilidad de compra.

nevents <- 2300  
compras <- 94  
prob\_true<-0.02  
phat <- compras/nevents  
phat

## [1] 0.04086957

n<-sqrt(prob\_true\*(1-prob\_true)/nevents)  
pnorm(phat,mean = prob\_true,sd=n, lower.tail=F)

## [1] 4.368528e-13

### 8. Por otro lado, en el mismo mes de la pregunta anterior se ha lanzado una campaña de márqueting directo con un target preseleccionado de 1000 clientes potenciales, obteniendo una respuesta positiva, esto es la compra del producto, en 56 casos. ¿Podemos afirmar que la tasa de respuesta obtenida en el target preseleccionado es mejor que la obtenida por internet?.

Considerando que el p-value ha sido un valor elevado (96,48%), entonces no rechazamos la hipotesis nula de que la tasa de respuesta obtenida NO es mejor. En este caso no ha habido mejora.

targetmkt <- 1000  
compramkt <- 56  
phatmkt <- compramkt/targetmkt  
  
nevents <- 2300  
compras <- 94  
phat <- compras/nevents  
  
z <- (phatmkt - phat)/sqrt((phatmkt\*(1-phatmkt)/targetmkt)+(phat\*(1-phat))/nevents)  
pnorm(z)

## [1] 0.9648237

### 9. Un día me encuentro con un amigo al que hace tiempo que no veía, va acompañado por su hijo. Me dice pero que tiene dos hijos. Cuál es la probabilidad de que su otro hijo también sea varón.

Resp.: Para descobrir la probabilidad necesitamos de un desenvolvimento binomial, donde el ‘X’ utilizaremos para representar el niño y ‘y’ la niña.

(x + y)2 → (x + y) \* (x + y) → x² + xy + xy + y² → x² + 2xy + y²

Probabilidad de tener dos hijos hombres → x² → (1/2)² → 1/4 → 25%

Como resultado de la función llegamos a conclusión que la probabilidad del segundo hijo ser un varón es seria del 25%.

### 10. Hace mucho tiempo, cuanto la televisión era en blanco y negro, empezó en TVE un programa concurso de gran éxito, se llamaba “Un, dos, tres, responda otra vez”. Su primer presentador fue el gran Kiko Ledgard. Una situación típica en dicho programa era cuando al concursante se le ofrecían tres puertas, detrás de una sola de las cuales había el premio. El concursante escogía una de las puertas, y entonces Kiko Ledgard abría una de las dos puertas no escogidas en donde NO había el premio y preguntaba al concursante si quería cambiar de opción (problema de Monty Hall en honor de su creador). ¿Cuál es la mejor opción para el concursante, mantenerse en su primera opción o cambiar de puerta?

Resp.: Cambiar de puerta, porque tendriás el doble de probabilidad de ganar el premio. La explicación con la cual es el doble de probabilidad si la puerta es cambiada, es debido al siguiente racional: -La chance de la primera vez pillar una puerta donde esta el premio es de 1/3 (33%); -La chance de la primera vez pillar una puerta donde NO esta el premio es de 2/3 (66%). Entonces, si NO es cambiada, hay el 66% de chance de no pillar la puerta con el premio.

Sin embargo, al cambiar de puerta, la situación se invierte. Porque como en la primera vez, las chances de NO conseguir pillar el premio todavia es de 66%, entonces cuando la puerta es cambiada, los 66% se invierten en la probabilidad de conseguir el premio. Todavia considerando este escenario, hay solo el 33% de chance de haber pillado por primera vez el premio, y ahí si cambiada la puerta, seria la que no tiene el premio. Pero esto representa el 33% de las veces.