**SESIÓN 1: ESTADÍSTICA BÁSICA**

**1.** Calcular la media, la mediana, la desviación tipo, el primer cuartil, el tercer cuartil, el rango intercuartílico y dibujar el correspondiente boxplot, de la siguiente muestra de precios de pisos (de 3 habitaciones): 7.80, 12.60, 15.96, 13.50, 8.25, 31.29, 16.46

Con el comando *summary* obtenemos gran parte de la información:

*> summary(x)  
 Min. 1st Qu. Median Mean 3rd Qu. Max.  
 7.800 9.338 13.050 14.980 15.720 31.290*

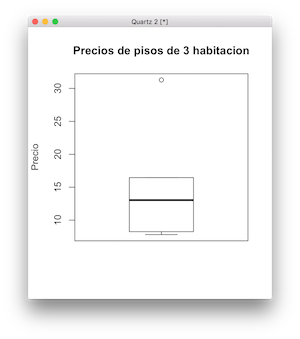
Media: 14.980  
Mediana: 13.050  
Primer cuartil: 9.338  
Tercer cuartil: 15.720

Para la desviación:

*> sd(x)  
 [1] 8.636047*

Y para el Boxplot:

*> boxplot(x, main="Precios de pisos de 3 habitaciones", ylab="Precio")*



**2.** Diga que razones le parecen convincentes para justificar que la media de una muestra escogida al azar es un estimador más fiable que la media de una muestra de expertos.

La media de expertos no será imparcial, estará sesgada por variables desconocidas y que falsearán los resultados de los análisis que se hagan. Mientras que la muestra escogida al azar, al ser por definición, aleatoria, será independiente.

**3.** Suponga la variable aleatoria “tiempo que tardo en ir de mi casa al trabajo” y diferencie factores sistemáticos y factores aleatorios intervinientes en su medición.

Factores sistemáticos serían las variables fijas que no cambian en el tiempo, como distancia de la oficina a casa, número de cruces con semáforos, líneas de metro que unen ambos puntos…

Los factores aleatorios serían por ejemplo el tiempo meteorológico (que influye en el tránsito)

**4.** Cuando podemos asegurar que el mecanismo generador de una variable aleatoria se aproxima a una distribución gaussiana.

Cuando la representación gráfica de las variables, en abscisas y ordenadas, se aproximan a una curva normal. Es decir, con un solo pico en el centro de la distribución donde coincide la media y la mediana. A ambos lados de este pico se distribuyen el resto de valores, la mitad a un lado y la mitad al otro y las colas se alargan de forma indefinida.

**5.** 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).

La hipótesis nula es que ese piso es una buena ocasión de compra.

Sin datos históricos, y sin referencia del total del conjunto, utilizamos la t de student para calcular la p

*> t <- (8-mean(x))/sd(x)*

*> pt(t, df=5) = 0.227*

Con esos datos tenemos que rechazar la hipótesis y no sabemos si es o no es una buena opción de compra.

**6.** 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 apartado 1 es aleatoria. ¿Podemos asegurar de que se trata de una muestra de pisos de l’Eixample?

mu <- 16.81

sigma <- 5.91

x <- (7.80, … 16.46)

n = length

curve(dnorm(x, mean=mu, sd=sigma… ))

mean(x)

z = (mean(x) – mu) / (sigma / sqrt(n))

pnorm(z) = 0.225 así que sí se considera que es una muestra de pisos de l’Eixample

**7.** 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?

x1 = 7.80…

x2 = 13.57…

n1 = length(x1)

n2 = length(x2)

mitj\_1 = mean(x1)

mitj\_2 = mean(x2)

z = (mitj\_2 – mitj\_1) / (sigma\*sqrt(1/n1+1/n2))

pnorm(x, lower.tail=F) = 0.058

Como 0.058 es mayor que 0.05 se podría decir que no ha habido cambio de precios entre los dos años, pero en realidad el valor está muy cerca

**8.** 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?

n = 2300

x = 94

p = 0.02

p\_hat = x/n

z = (p\_hat – p) / sqrt(p\*(1-p)/n)

pnorm(z, lower.tail=F) = 4.3e-13

Se rechaza la hipótesis nula y se considera que algo ha pasado

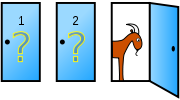
**9.** Un día me encuentro con un amigo al que hace tiempo que no veía, va acompañado por su hijo. Me dice pero que tiene dos hijos. Cuál es la probabilidad de que su otro hijo también sea varón.

Presuponiendo que el amigo vive en España, y según los datos del INE para 2015:

🡪 hombres: 22.809.420, mujeres: 23.636.408, total: 46.445.828

MM - MF - FM - FF

1/3 es la probabilidad

**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?

La mejor opción es cambiar de puerta, ya que al principio cuando le pidieron elegir, tenía un 33% de probabilidad de que le tocase el premio y un 66% de que no eligiese la puerta correcta. Al enseñar una puerta sin premio, si cambia de puerta la probabilidad aumenta a un 50% frente al 33 % de si se queda con la puerta.

*Utilice R para la solución de los ejercicios*

El software que utilizaremos es R. R es un entorno general de cálculo estadístico y de programación, open-source, gratuito, multi-plataforma y en continua evolución con la incorporación de nuevos paquetes específicos a distintas áreas de investigación (genómica, big data, …).

*Instalación de R*

http://cran.r-project.org/

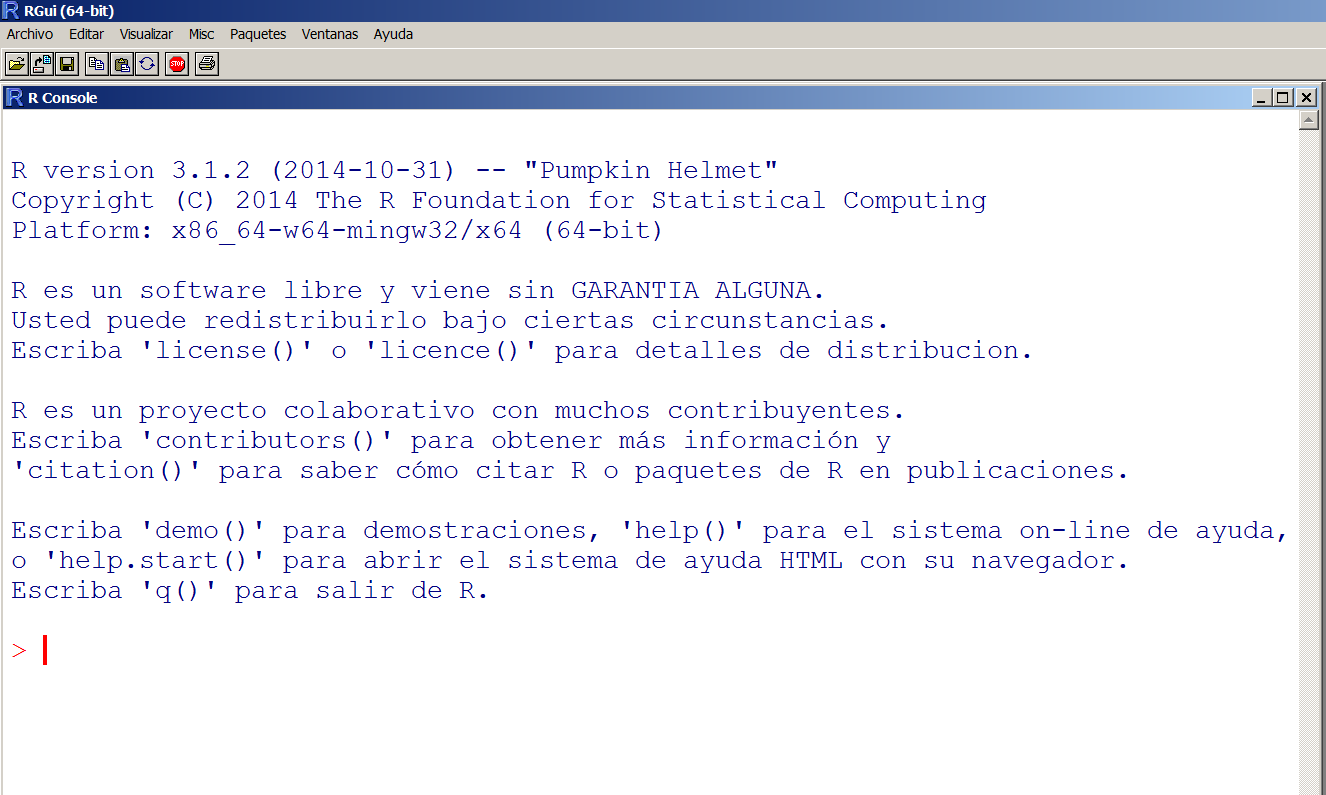
Download R for ...

Ejecutamos el fixero **exe** descargado. Veremos aparecer el icono de http://i657.photobucket.com/albums/uu300/BlogAqueronte/Logos/th_Icono_R.gif para trabajar con él.

*Iniciar una sesión de trabajo en R*

CLICKAR SOBRE EL ICONO DE R

Se abrirá una ventana como esta:



*Especificar el directorio de trabajo en R*

> setwd("C:/path/to/directory")

O En Archivo - Cambiar dir ... ESPECIFICAR EL ARCHIVO DE TRABAJO

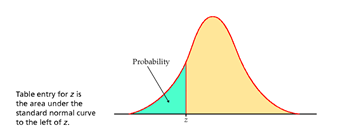
*Ejecutar un script en R*

> source("fichero.r")

O En Archivo – Abrir script ... ESPECIFICAR EL ARCHIVO CONTENIENDO EL PROGRAMA EN R

O COPIANDO fichero.r EN LA ZONA DE COMANDOS DE R

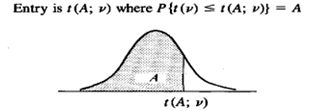
***Trabajando con la N(0,1)***



In 

pnorm(z, mean=0, sd=1, lower.tail=TRUE)

***Trabajando con la Student’s t con ν = n-1 grados de libertad (df)***



In 

pt(t, df, lower.tail=TRUE)