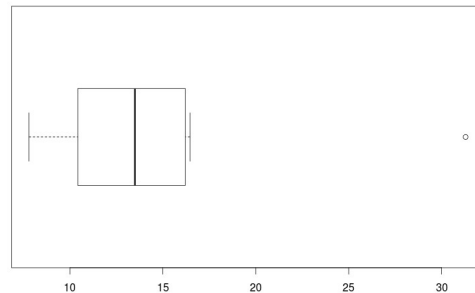


## SESIÓN 1: ESTADÍSTICA BÁSICA

**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

```
> x <- c(7.80, 12.60, 15.96, 13.50, 8.25, 31.29, 16.46)
> summary(x)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  7.80   10.42   13.50   15.12   16.46   31.29
> sd(x)
[1] 7.892234
> boxplot(x, horizontal=TRUE)
```



**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.

Las muestras de expertos (no probabilísticas) no pueden generalizarse a la población, pues son muestras dirigidas. Sin embargo el muestreo no probabilístico tiene su utilidad para determinados tipos de investigación.

**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:	- método seleccionado para medida de tiempo - precisión del instrumento de medida de tiempo
Factores aleatorios:	- tiempo (sol, lluvia, tormenta) - semáforos - avería del vehículo (si voy en coche) - incidente en las vías (si voy en RENFE)

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

Cualquier mecanismo generador de una variable aleatoria se aproximará a una distribución gaussiana si se realizan el número suficiente de experimentos, tomando muestras independientes.

**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).

A modo de ejemplo, tomemos los precios históricos del archivo 'preus\_3hab\_Leixample.r' (alternativamente deberíamos construir una distribución de referencia) y calculamos su p.value:

```
> length(preu_3hab)
[1] 221
> x = preu_3hab[preu_3hab<8]
> x
[1] 7.80 6.21 7.52 7.14 6.76
> length(x)
[1] 5
> length(x)/length(preu_3hab)
[1] 0.02262443
```

El p.value nos indica que solamente encontraremos un 2.26% de pisos a precios más bajos con lo que podemos decir que es una gran oferta (dejando a un lado todos los otros factores intervinientes a parte del precio).

**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?

....

**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?

```
> eixample2 = c(13.57, 14.80, 22.36, 29.29, 22.70)
> summary(eixample2)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
13.57  14.80   22.36   20.54  22.70   29.29
> sd(eixample2)
[1] 6.443387
```

No. Ha habido un cambio significativo en la media (+22%) por ejemplo.

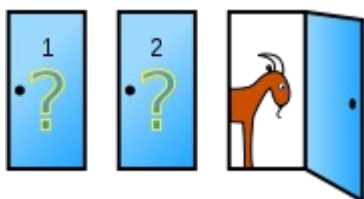
**8.** Explicar cómo se calcula el p\_valor en una prueba de permutaciones del problema anterior.

...

**9.** 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?

Si, pues la probabilidad estima que se producirían solo 46, obteniend en realidad más del doble.

**10.** Hace mucho tiempo, cuando la televisión era en blanco y negro, empezó en TVE un programa concurso de gran éxito, se llamaba “Un, dos, tres, responde 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?

El concursante debe cambiar siempre de puerta. Si no se realiza el cambio simplemente tenemos la probabilidad inicial de 0.33 de haber escogido el premio, en cambio la probabilidad de 0.66 de haber escogido una puerta sin premio que si luego la cambiamos nos llevará al premio.

Supongamos que el premio está en la puerta del medio (2ª):

1. El concursante elige la 1ª puerta. El presentador le abre la 3ª.  
Si decidiese no cambiar de puerta – **no se lleva nada**  
Si decidiese cambiar de puerta – **se lleva el premio**
2. El concursante elige la 2ª puerta. El presentador le abre la 1ª o la 3ª.  
Si decidiese no cambiar de puerta – **se lleva el premio**  
Si decidiese cambiar de puerta – **no se lleva nada**
3. El concursante elige la 3ª puerta. El presentador le abre la 1ª.  
Si decidiese no cambiar de puerta – **no se lleva nada**  
Si decidiese cambiar de puerta – **se lleva el premio**

Queda demostrado que el cambio aumenta las probabilidades de ganar.

## 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, ...).

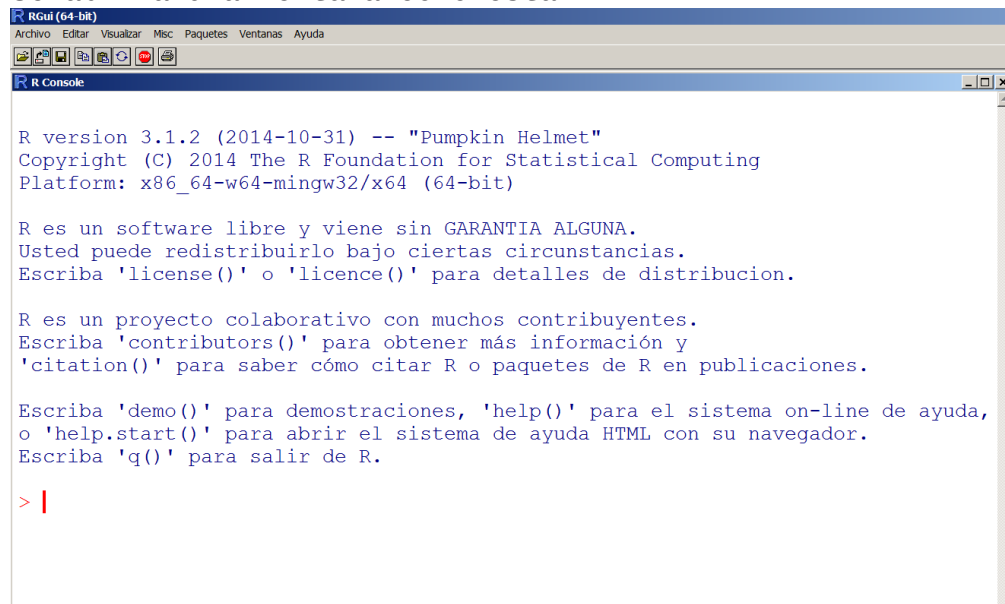
## Instalación de R

<http://cran.r-project.org/>  
Download R for ...

Ejecutamos el fixero **exe** descargado. Veremos aparecer el icono de **R** 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")
```

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

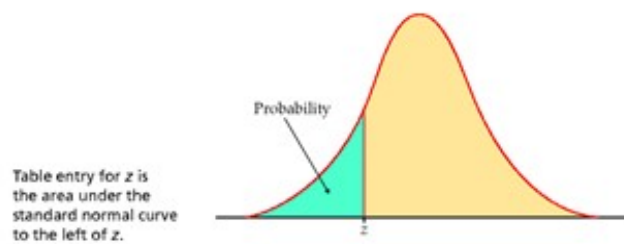
## Ejecutar un script en R

```
> source("fichero.r")
```

0 En Archivo - Abrir script ... ESPECIFICAR EL ARCHIVO  
CONTENIENDO EL PROGRAMA EN R

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

## Trabajando con la $N(0,1)$

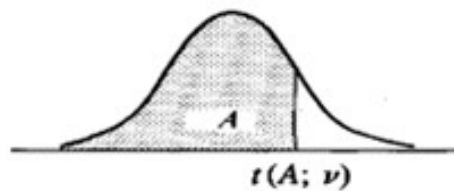


In 

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

## Trabajando con la Student's $t$ con $\nu = n-1$ grados de libertad (df)

Entry is  $t(A; \nu)$  where  $P\{t(\nu) \leq t(A; \nu)\} = A$



In 

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