Práctica 5 R-Commander

NOMBRE: *Jaime Osés Azcona*

1 - R como calculadora

A lo largo del tiempo en el que las compuertas de un pantano están cerradas, el caudal de un canal de riego (m^3 por segundo) es una variable aleatoria con densidad:

f(x) = $\begin{cases} \frac{1}{6}x^2 e^{-x/2} & x>0 \\ 0 & en\,otro\,caso \end{cases}$

a) En qué porcentaje del tiempo el caudal no llegará a 5 m^3 por segundo.

Como es una distribución continua, para calcular P(X < 5) deberemos realizar la integral de la función distribución f(x) entre ∞ y 5, o lo que es lo mismo, entre 0 y 5. Para ello utilizaremos R como calculadora de la siguiente forma:

```
a=function(x){(1/16)*(x^2)*exp(-x/2)}
integrate (a, lower=0,upper=5)
```

```
> integrate (a, lower=0,upper=5)
0.4561869 with absolute error < 5.1e-15
```

b) ¿Cuál es el caudal medio? ¿Con cuánta desviación?

Para calcular el caudal medio deberemos calcular la esperanza, que es la integral de x*f(x) entre -∞ y ∞, o lo que es lo mismo, entre 0 y ∞.

```
esperanza=function(x){(1/16)*(x^3)*exp(-x/2)}
integrate (esperanza, lower=0,upper=Inf)
```

```
> integrate (a, lower=0,upper=Inf)
6 with absolute error < 0.000034
```

Para calcular la desviación, o lo que es lo mismo, la raíz cuadrada de la varianza , deberemos utilizar la siguiente formula **"Var (x) = E(x^2)-(E(x))^2"** donde E(x^2) será:

```
momento2=function(x){(1/16)*(x^4)*exp(-x/2)}
integrate (momento2, lower=0, upper=Inf)
```

```
> (integrate (momento2, lower=0,upper=Inf))
48 with absolute error < 0.000031
```

Var = 48 – 6^2 = 12

σ = $\sqrt{12}$

c) El 50% de los das el caudal se mantendrá por encima de, ¿cuántos m^3 por segundo?

Para hallar P(x>a)=0.5 realizaremos la integral de f(x) entre a y ∞, poniendo en a cualquier valor hasta encontrar la probabilidad que buscamos.

```
densidad=function(x){(1/16)*(x^2)*exp(-x/2)}
integrate (densidad, lower=a, upper=Inf)
```

En nuestro caso, el valor de a = 5,34

2 - El Menú Distribuciones

1. El tiempo que cuesta rellenar un formulario electrónico se distribuye uniformemente entre 1,5 y 2,2 minutos. Representa sobre la gráfica de densidad y calcula la probabilidad de que cueste menos de dos minutos rellenar el formulario.

Como X sigue una distribución Uniforme diremos que X~U(1.5, 2.2).
Por tanto, para calcular las probabilidades asociadas a este apartado los haremos desde el menú ***"Distribuciones-...continuas-…uniforme"*** e introduciremos los parámetros que necesitemos según lo que nos pida cada apartado.

En este caso, queremos hallar $P(X \leq 2)$. Por tanto, iremos al menú ***"Distribuciones-...continuas-…uniforme-Probabilidades uniformes acumuladas"*** e introduciremos los extremos (1.5, 2.2) y el valor acumulado que queremos obtener, en este caso hasta el 2.

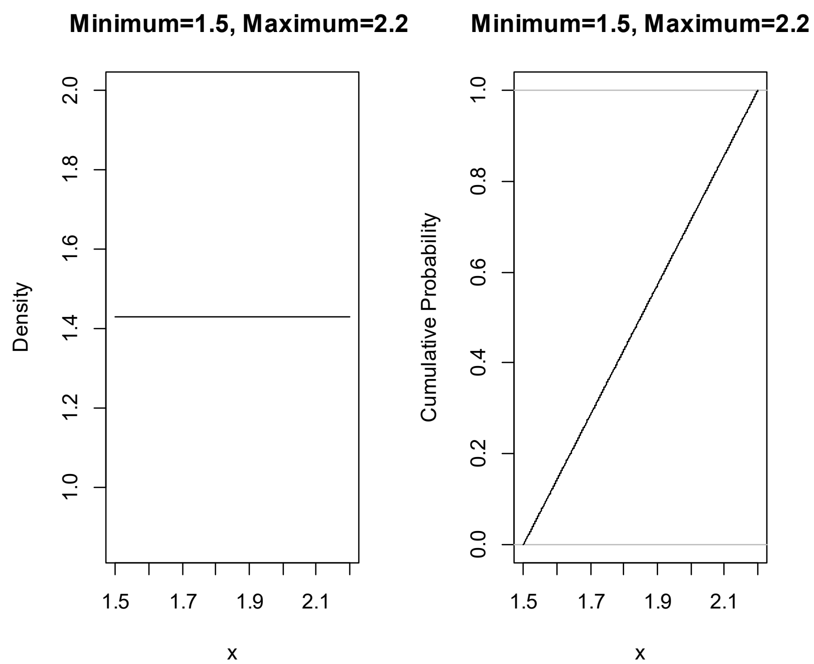
```
> punif(c(2), min=1.5, max=2.2, lower.tail=TRUE)
[1] 0.7142857
```

El 95% de los casos el tiempo mínimo sera de, ¿Cuántos minutos? Dibuja la distribución de probabilidad acumulada y la de densidad en una misma pantalla.
En este caso, queremos hallar lo contrario que en el caso anterior. Por tanto, tendremos que ir al menú **"Distribuciones-*...continuas-…uniforme*-Cuantiles..."** y pondremos el porcentaje que se nos pide, 1-0.95. Lo que hallaremos será lo siguiente :

$$P(X \geq a) = 1 - P(X < a) = 0,05$$

```
> qunif(c(0.05), min=1.5, max=2.2, lower.tail=TRUE)
[1] 1.535
```

Para realizar las gráficas que se nos piden, iremos al menú **"Distribuciones -*...continuas-…uniforme*-Gráfica de ..."** y seleccionaremos la gráfica que queramos, ya que nos da a elegir entre realizar la gráfica de la función de densidad y la gráfica de probabilidades acumuladas.



2. Con un sistema de irrigación automática a altura de las plantas, dos semanas después de la germinación, se distribuye normalmente con media 2,5 cm y desviación de 0,5 cm. ¿Cuál es la probabilidad de que la altura de la planta sea mayor de 2,25 cm?¿Y de que se encuentre entre 2 y 3 cm? ¿Qué altura sobrepasara el 90% de las plantas?

Como X sigue una distribución Normal diremos que X~N(2.5, 0.5).
Por tanto, para calcular las probabilidades asociadas a este apartado los haremos desde el menú ***"Distribuciones-...continuas-…normal"*** e introduciremos los parámetros que necesitemos según lo que nos pida cada apartado.

Queremos calcular P(X>2,25).Para ello, iremos al menú ***"Distribuciones-...continuas-…normal-Probabilidades normales acumuladas"*** e introduciremos la media (2,5) y desviación típica (0,5) y el valor acumulado que queremos obtener, en este caso 2,25. Para hallar la probabilidad deberemos hacer P(X>2,25) = 1 – P(X≤ 2,25).

```
> 1 - pnorm(c(2.25), mean=2.5, sd=0.5, lower.tail=TRUE)
[1] 0.6914625
```

Para hallar P(2<X<3) deberemos hacer P(2<X<3) =P(X<3) -P(X<2), obteniendo dichas probabilidades como en el caso anterior.

```
> pnorm(c(3), mean=2.5, sd=0.5, lower.tail=TRUE) - pnorm(c(2), mean=2.5, sd=0.5, lower.tail=TRUE)
[1] 0.6826895
```

En este caso, queremos hallar lo contrario que en los casos anteriores. Por tanto, tendremos que ir al menú ***"Distribuciones-...continuas-…normal-Cuantiles..."*** y pondremos el porcentaje que se nos pide, 1-0.90. Lo que hallaremos será lo siguiente :

$$P(X \geq a) = 1 - P(X < a) = 0{,}1$$

```
> qnorm(c(0.1), mean=2.5, sd=0.5, lower.tail=TRUE)
[1] 1.859224
```

Si con otro sistema de irrigación el crecimiento se distribuye como una exponencial de media 0,5, ¿con cuál de los dos sistemas hay más probabilidad de que una planta alcance en dos semanas los 3,5 cm?
En este apartado, cuando se dice que la planta alcanza cierta altura, nos estaremos refiriendo a que su altura es igual o mayor que dicha altura. Por tato la probabilidad que queremos hallar es P(X ≥ 3,5)= 1 - P(X < 3,5)

- Distribución Normal

```
> 1- pnorm(c(3.5), mean=2.5, sd=0.5, lower.tail=TRUE)
[1] 0.02275013
```

- Distribución Exponencial

```
> 1 - pexp(c(3.5), rate=2, lower.tail=TRUE)
[1] 0.000911882
```

Por tanto, podemos afirmar que hay más probabilidad de que una planta alcance en 2 semanas los 3,5 cm en el sistema de irrigación que se distribuye como una normal.

3 - Ajuste de distribuciones

1.- Nos dicen que un programa de ordenador genera observaciones de una distribución normal estándar. Como no estamos seguros de ello, obtenemos una muestra aleatoria de 450 observaciones mediante dicho programa y se obtienen los siguientes resultados:

Resultado	< - 2	(-2, -1)	(-1, 0)	(0, 1)	(1, 2)	>2
Frecuencia observada: O	30	80	140	110	60	30
H_0 cierta [frecuencia esperada]: E						

¿Se puede aceptar con un nivel de significación que el programa funciona correctamente?

2.- El Rector de una Universidad opina que el 60% de los estudiantes consideran los cursos que realizan como muy útiles, el 20% como algo útiles y el 20% como nada útiles. Se hace una encuesta a 100 estudiantes y 68 consideran que son muy útiles, 18 poco útiles y 14 nada útiles. ¿Es aceptable la opinión del rector?
Para realizar este problema, haremos un contraste de hipótesis para cada afirmación del rector. Si todas se cumplen, la opinión del rector será aceptable. Pero con que una no se cumpla, la opinión del rector no será aceptable.

Tomaremos para todos ellos un nivel de confianza del 95%.
El primer contraste será el siguiente:
$H_0 : \pi = 0,6$
$H_1 : \pi \neq 0,6$
Para calcular dicho contraste tendremos que hallar la Zobs, $Z_{-\alpha/2}$ y $Z_{\alpha/2}$

$$Z_{obs} = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0,68-0,6}{\sqrt{\frac{0,6(1-0,6)}{100}}} = 1,63$$

```
> qnorm(c(0.025), mean=0, sd=1, lower.tail=TRUE)
[1] -1.959964

> qnorm(c(0.975), mean=0, sd=1, lower.tail=TRUE)
[1] 1.959964
```

Como Z_{obs} está entre $Z_{-\alpha/2}$ y $Z_{\alpha/2}$ no se rechaza la hipótesis nula.

El segundo contraste será el siguiente:
$H_0 : \pi = 0{,}2$
$H_1 : \pi \neq 0{,}2$
Para calcular dicho contraste tendremos que hallar la Zobs, $Z_{-\alpha/2}$ y $Z_{\alpha/2}$

$$Z_{obs} = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0{,}18-0{,}2}{\sqrt{\frac{0{,}2(1-0{,}2)}{100}}} = -0{,}5$$

```
> qnorm(c(0.025), mean=0, sd=1, lower.tail=TRUE)
[1] -1.959964

> qnorm(c(0.975), mean=0, sd=1, lower.tail=TRUE)
[1] 1.959964
```

Como Z_{obs} está entre $Z_{-\alpha/2}$ y $Z_{\alpha/2}$ no se rechaza la hipótesis nula.

El tercer contraste será el siguiente:
$H_0 : \pi = 0{,}2$
$H_1 : \pi \neq 0{,}2$
Para calcular dicho contraste tendremos que hallar la Zobs, $Z_{-\alpha/2}$ y $Z_{\alpha/2}$

$$Z_{obs} = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0{,}14-0{,}2}{\sqrt{\frac{0{,}2(1-0{,}2)}{100}}} = -1$$

```
> qnorm(c(0.025), mean=0, sd=1, lower.tail=TRUE)
[1] -1.959964

> qnorm(c(0.975), mean=0, sd=1, lower.tail=TRUE)
[1] 1.959964
```

Como Z_{obs} está entre $Z_{-\alpha/2}$ y $Z_{\alpha/2}$ no se rechaza la hipótesis nula.

Como se cumplen los 3 contrastes planteados, podremos decir que la opinión del rector es aceptable.

3.- En una tómbola se desarrolla el siguiente sorteo. De un gran bombo opaco se extrae una bola al azar. En el bombo se nos dice que hay 5 bolas con lunares, 45 bolas blancas, 50 azules, 50 rojas y 50 amarillas. Una bola con lunares supone un premio importante, mientras que una blanca es un premio de consolación, el resto no tienen premio. Se observan 600 extracciones y se observa que se reparten 6 premios importantes y 160 de consolación. Se puede afirmar, con una significación del 5% que el bombo realmente contiene la distribución de bolas que se anuncia?

Para realizar este problema, haremos un contraste de hipótesis para ver la proporción de bolas con premio importante y otro para las de premio de consolación. Si todas se cumplen, la distribución de bolas del bombo será la que se anuncia. Pero con que una no se cumpla, la distribución no sera esa.
Hay 200 bolas, 5 de ellas con premio importante, 45 con premio de consolación y 150 sin premio.
El primer contraste será el siguiente:
$H_0 : \pi = 0,025$
$H_1 : \pi \neq 0,025$
Para calcular dicho contraste tendremos que hallar la Zobs, $Z_{-\alpha/2}$ y $Z_{\alpha/2}$

$$Z_{obs} = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0,01-0,025}{\sqrt{\frac{0,025(1-0,025)}{600}}} = -2,35$$

```
> qnorm(c(0.025), mean=0, sd=1, lower.tail=TRUE)
[1] -1.959964

> qnorm(c(0.975), mean=0, sd=1, lower.tail=TRUE)
[1] 1.959964
```

Como Z_{obs} no está entre $Z_{-\alpha/2}$ y $Z_{\alpha/2}$ se rechaza la hipótesis nula.

El segundo contraste será el siguiente:
$H_0 : \pi = 0,225$
$H_1 : \pi \neq 0,225$
Para calcular dicho contraste tendremos que hallar la Zobs, $Z_{-\alpha/2}$ y $Z_{\alpha/2}$

$$Z_{obs} = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0,27-0,225}{\sqrt{\frac{0,225(1-0,225)}{600}}} = 2,63$$

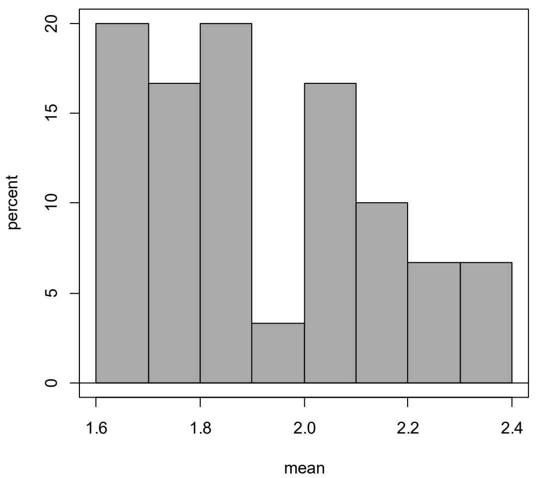
```
> qnorm(c(0.025), mean=0, sd=1, lower.tail=TRUE)
[1] -1.959964

> qnorm(c(0.975), mean=0, sd=1, lower.tail=TRUE)
[1] 1.959964
```

Como Z_{obs} no está entre $Z_{-\alpha/2}$ y $Z_{\alpha/2}$ se rechaza la hipótesis nula.

Ninguno de los contrastes se cumple, por tanto podemos afirmar que la distribución de bolas del bombo no es la que se anuncia.

4 - Teorema central del límite

El tiempo entre llegadas de clientes a un banco se comporta como una exponencial. En el fichero tiempos.rda se tiene el tiempo transcurrido entre la llegada de los 100 primeros clientes en 30 días distintos. Se añade la variable media de estos tiempos.
1. Construye un histograma para los valores de la medias de las 30 muestras y a la luz del gráfico da una estimación del tiempo medio que transcurre entre dos llegadas.
Para construir el histograma iremos al menú **"Gráficas-Histograma..."**.



Para estimar el tiempo medio que transcurre entre dos llegadas haremos un resumen numérico de las medias obtenidas para ser más exactos.

```
> numSummary(tiempos[,"
+    .75,1))
      mean          sd  n
 1.925757 0.2214932 30
```

Por tanto, el parámetro de la exponencial será $\lambda = \frac{1}{\mu}$ = 0,51086

2. El tiempo medio entre llegadas en un 75% de los días es superior a ¿qué valor?
Para hallar lo que se nos pide en el enunciado, $P(X \geq a) = 1 - P(X < a) = 0.25$, iremos al menú **"Estadísticos-Resúmenes-...numéricos..."** y seleccionaremos la opción cuantiles.

```
> numSummary(tiempos[,"mean", drop=FALSE],
+    quantiles=c(.25,.75))
      mean        sd       25%       75%  n
 1.925757 0.2214932 1.744926 2.082095 30
```

3. Resuelve la pregunta anterior utilizando el teorema central del límite y haciendo uso de la estimación del primer apartado.
Sabemos que al distribuirse como una exponencial, la media y la desviación típica serán $\mu = \frac{1}{\lambda}$ y $\sigma = \frac{1}{\lambda^2}$ respectivamente. Tomaremos para ello la estimación del parámetro de la exponencial que hemos hallado en la pregunta **1**. Por tanto, μ = 1,9257 y σ = 3,8318.
Como tenemos una muestra de tamaño 30 podremos aplicar el teorema central del límite.

$X \sim N(1.9257, \frac{3.8318}{\sqrt{30}})$, o lo que es lo mismo, $X \sim N(1.9257, 0.6996)$

En este caso, queremos hallar el valor que deja al 75% a su izquierda. Por tanto, tendremos que ir al menú ***"Distribuciones-...continuas-…normal-Cuantiles..."*** y pondremos el porcentaje que se nos pide, 1-0.75. Lo que hallaremos será lo siguiente :

$$P(X \geq a) = 1 - P(X < a) = 0,25$$

```
> qnorm(c(0.25), mean=1.9257, sd=0.6996, lower.tail=TRUE)
[1] 1.453827
```