Universidad de Alcalá

Estadística

Práctica 5

Variables aleatorias y distribuciones discretas

Universidad de Alcalá

Contenido

Universidad de Alcalá

1. Introducción

Con esta práctica se utiliza R y RStudio para calcular probabilidades de sucesos a partir de variables discretas que siguen una distribución binomial o una distribución de Poisson.

2. Variables discretas con distribución binomial

Una variable binomial es una variable discreta (X) que representa el número de éxitos (x) obtenidos en n ensayos independientes (de Bernoulli), conociendo la probabilidad (p) de un éxito.

Se representa como X:B(n,p). Y se dice que la variable tiene una distribución binomial.

Por ejemplo, una variable X:B(10,0.5) que representa: "*Número de caras (éxitos) al lanzar 10 veces una moneda (n=10), sabiendo que la probabilidad de que salga cara (éxito) en cada tirada es del 50% (p=0.5)*".

En R se pueden utilizar las funciones indicadas en la siguiente tabla. En la columna "Ejemplo de uso", se utiliza una variable B(10,0.5).

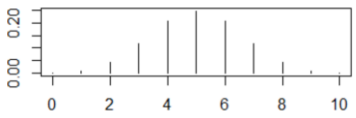
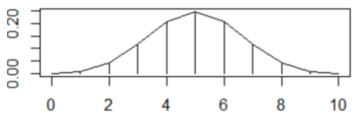
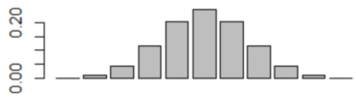
Función	Código R	Resultado	Ejemplo de uso
Masa de probabilidad P{X=x}	`dbinom(x,n,p)`	Calcula la probabilidad de que una variable aleatoria binomial X:B(n,p) tenga el valor x.	`>dbinom(3,10,0.5)` `[1] 0.1171875`
Distribución de probabilidad F(X)=P{X≤x}	`pbinom(x,n,p)`	Calcula la probabilidad de que una variable aleatoria binomial X:B(n,p) tenga un valor igual o menor que x.	`>pbinom(3,10,0.5)` `[1] 0.171875`

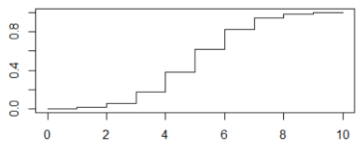
Cuantil	`qbinom(q,n,p)`	Calcula el menor valor de una variable aleatoria binomial X:B(n,p), que cumple que la probabilidad de que la variable sea menor o igual a ese valor, es mayor o igual a q.	`>qbinom(0.7,10,0.5)`
			`[1] 6`
			
			Entonces se cumple:
			P{X≤6} ≥ 0.7
			Pero no se cumple:
			P{X≤5} ≥ 0.7

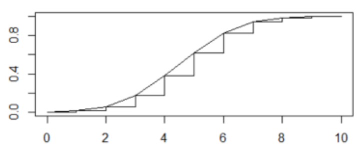
Generación aleatoria de muestras	`rbinom(m,n,p)`	Genera aleatoriamente un vector de m valores aleatorios de una variable binomial X:B(n,p). Cada valor representa el número de éxitos en cada simulación del experimento.	`>rbinom(4,10,0.5)`
			`[1] 6 7 7 8`
			
			`Realiza 4 simulaciones del experimento de lanzar 10 veces la moneda, indicando el número de caras obtenidas en cada simulación.`

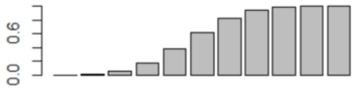
Esperanza E[X]	`n*p`		`>(E=10*0.5)` `[1] 5`
Varianza Var[X]	`n*p*(1-p)`		`>(Var=10*0.5*(1-0.5))` `[1] 2.5`

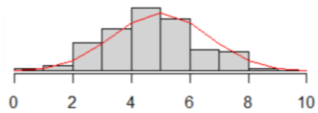
Las funciones anteriores se pueden combinar con otras de generación de gráficos, para obtener los diagramas que se indican en la siguiente tabla.

Diagrama	Código R	Comentarios	Ejemplo de uso
Función de masa de probabilidad	`plot(0:n,dbinom(0:n,n,p),type=tipo)`	`tipo` puede ser: `“l”`: líneas `“h”`: barras `“b”`: líneas y barras	`>plot(0:10,dbinom(0:10,10,0.5),type="h")` 
Función de masa de probabilidad (línea)	`lines(0:n,dbinom(0:n,n,p))`	Los diagramas de línea se dibujan sobre el último diagrama plot o barplot.	`>lines(0:10,dbinom(0:10,10,0.5))` 
Función de masa de probabilidad (barras)	`barplot(dbinom(0:n,n,p))`		`>barplot(dbinom(0:10,10,0.5))` 

Función de distribución de probabilidad	`plot(0:n,pbinom(0:n,n,p),type=”s”)`
	Otra opción:
	`plot(stepfun(0:n-1,pbinom(0:n,n,p))`
		`type` también puede ser:
		`“l”`: líneas
		`“h”`: barras
		`“b”`: líneas y barras
			`>plot(0:10,pbinom(0:10,10,0.5), type="s")`
			

Función de distribución de probabilidad (línea)	`lines(0:n,pbinom(0:n,n,p))`	Los diagramas de línea se dibujan sobre el último diagrama plot o barplot.	`>lines(0:10,pbinom(0:10,10,0.5))` 

Función de distribución de probabilidad (barras)	`barplot(pbinom(0:n,n,p))`		`>barplot(pbinom(0:10,10,0.5))` 

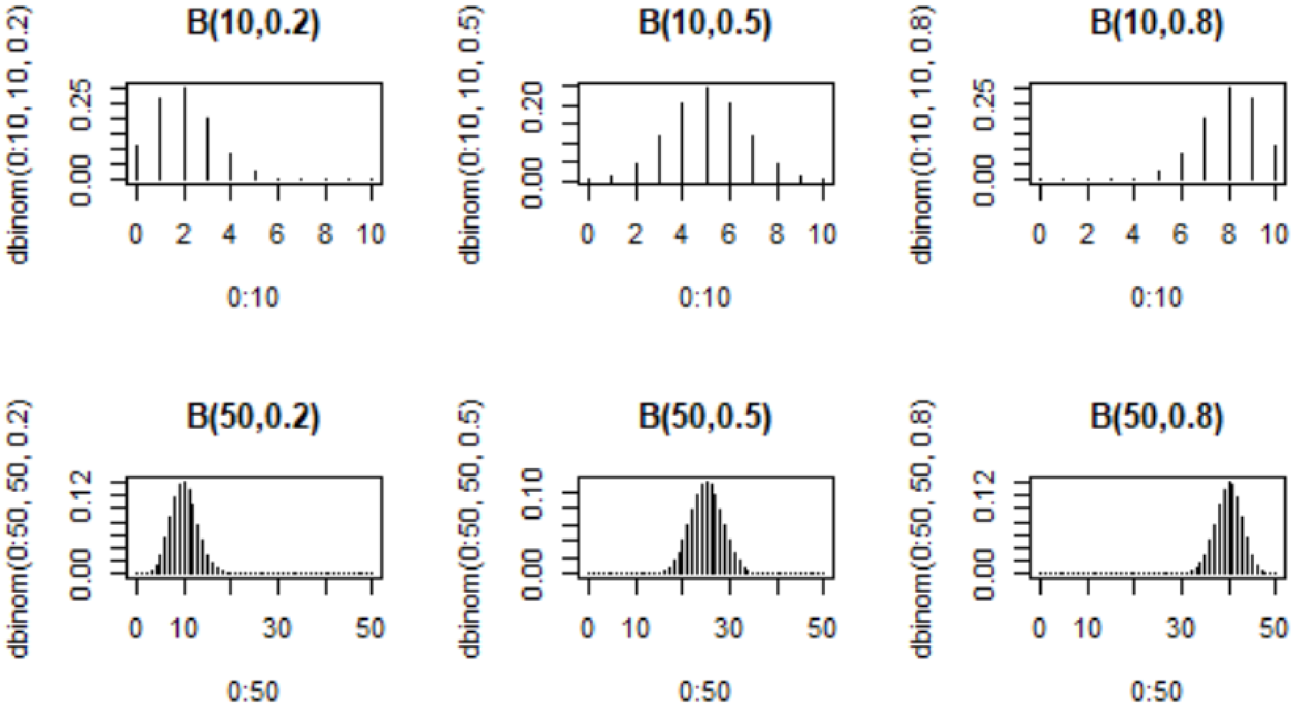
Histograma de una simulación de posibles valores de la variable aleatoria	`hist(rbinom(m,n,p), breaks=0:n, freq=FALSE)`
	
	`lines(0:n,dbinom(0:n,n,p))`
		Se puede representar primero el histograma con los datos de la simulación y después la función de probabilidad binomial teórica, para ver si son similares.
		NOTA: Se podría hacer lo mismo si se dispone de datos reales, para ver si las muestras reales cumplen tienen una distribución binomial. Por ejemplo, lanzando realmente una moneda.
			`>hist(rbinom(100,10,0.5), breaks=0:10, freq=FALSE)`
			
			`>lines(0:10,dbinom(0:10,10,0.5),col="red")`
			

En la función `hist()` con `freq=FALSE` se indica que el eje vertical no represente las frecuencias absolutas, y los valores del eje vertical se ajustan para que el área total del histograma sea 1, y así coincide con el área total bajo la línea de la distribución binomial y los valores del eje son compatibles con la línea dibujada después.

Se puede comparar el efecto de modificar los parámetros n y p, mostrando simultáneamente varios gráficos.

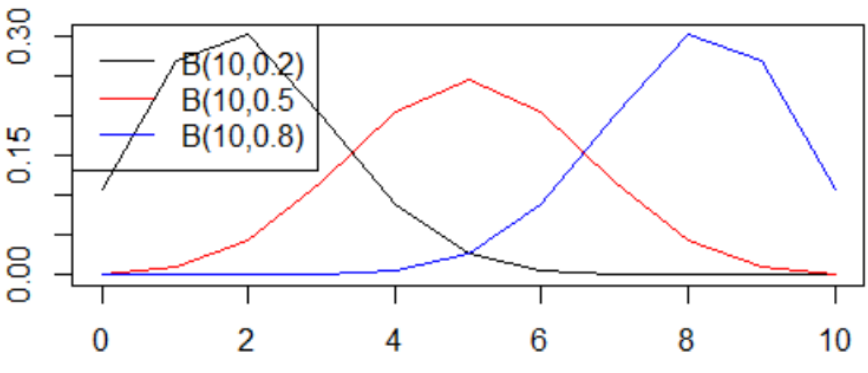
Por ejemplo dividiendo la ventana gráfica en 2 filas y 3 columnas:

```
> par(mfrow=c(2,3))
> plot(0:10,dbinom(0:10,10,0.2), type="h", main = "B(10,0.2)")
> plot(0:10,dbinom(0:10,10,0.5), type="h", main = "B(10,0.5)")
> plot(0:10,dbinom(0:10,10,0.8), type="h", main = "B(10,0.8)")
> plot(0:50,dbinom(0:50,50,0.2), type="h", main = "B(50,0.2)")
> plot(0:50,dbinom(0:50,50,0.5), type="h", main = "B(50,0.5)")
> plot(0:50,dbinom(0:50,50,0.8), type="h", main = "B(50,0.8)")
> par(mfrow=c(1,1))
```



O se pueden dibujar varios diagramas en el mismo gráfico, el primero con plot y el resto con lines:

```
plot(0:10,dbinom(0:10,10,0.2), type="l")
lines(0:10,dbinom(0:10,10,0.5), col="red")
lines(0:10,dbinom(0:10,10,0.8), col="blue")
legend("topleft",legend=c("B(10,0.2)","B(10,0.5","B(10,0.8)"),
       col=c("black", "red", "blue"), bg="transparent", lty=1)
```



Se recomienda investigar diferentes formas de dibujar la leyenda en el gráfico.

Universidad de Alcalá

2.1 Ejemplo: Estudiantes que tienen móviles Android

ENUNCIADO

Mediante una encuesta, se sabe que 50 de 74 estudiantes de un curso de la asignatura Estadística del Grado en Ingeniería en Sistemas de Información de la Universidad de Alcalá tienen teléfonos móviles con sistema operativo Android, y el resto con iOS. Si se pregunta aleatoriamente a 10 de los estudiantes, responder a las siguientes preguntas:

a) **Definir una variable aleatoria binomial para el experimento**
b) **Calcular la esperanza y varianza de la variable aleatoria**
c) **Dibujar las funciones de probabilidad y distribución de la variable aleatoria**
d) **Cuál es la probabilidad de que exactamente 5 de los 10 estudiantes tengan móvil con Android**
e) **Cuál es la probabilidad de que 5 o menos de los 10 estudiantes tengan móvil con Android**
f) **Cuál es la probabilidad de que tengan móvil con Android más de 5 de los 10 estudiantes**
g) **Cuál es la probabilidad de que los 10 estudiantes tengan móvil con Android**
h) **Cuál es la probabilidad de que ninguno de los 10 estudiantes tenga móvil con Android**
i) **Cuál es la probabilidad de que 6 o 7 estudiantes tenga móvil con Android**
j) **Cuál es el valor de los cuartiles de la variable aleatoria**
k) **Si tenemos los datos de la encuesta original a los 74 alumnos, comprobar si tomando las respuestas en grupos de 10 alumnos, el histograma se ajusta realmente a la función de probabilidad binomial.**
l) **Repetir el apartado anterior pero utilizando datos generados aleatoriamente con R, en lugar de los datos reales**

SOLUCIÓN

a) Definir una variable aleatoria binomial para el experimento

X = Número de estudiantes que tienen móvil Android de un conjunto de 10 estudiantes.

La variable sigue una distribución binomial B(n,p), donde n=10 y p=50/74, ya que por las encuestas se sabe que 50 de los 74 alumnos tienen móvil Android, como puede comprobarse leyendo los datos del archivo encuesta.csv y generando una tabla de frecuencias absolutas.

```
> encuesta=read.csv2("encuesta.csv")
> encuesta$SO
 [1] "Android" "iOS"     "Android" "iOS"     "Android" "Android" "Android" "Android" "iOS"     "iOS"
[11] "iOS"     "Android" "Android" "iOS"     "Android" "iOS"     "Android" "iOS"     "Android" "Android"
[21] "Android" "iOS"     "Android" "iOS"     "Android" "iOS"     "Android" "Android" "Android" "iOS"
[31] "Android" "Android" "iOS"     "Android" "Android" "iOS"     "Android" "Android" "Android" "Android"
[41] "iOS"     "Android" "Android" "Android" "Android" "Android" "iOS"     "Android" "Android" "Android"
[51] "iOS"     "iOS"     "iOS"     "Android" "Android" "iOS"     "iOS"     "Android" "iOS"     "Android"
[61] "Android" "iOS"     "Android" "iOS"     "Android" "Android" "Android" "Android" "Android" "Android"
[71] "Android" "Android" "Android" "Android"
```

```
> table(encuesta$SO)

Android     iOS
     50      24
```

Una vez confirmados los datos, podemos dar valores a los parámetros n y p. Por simplicidad vamos a redondear todos los valores con dos cifras decimales.

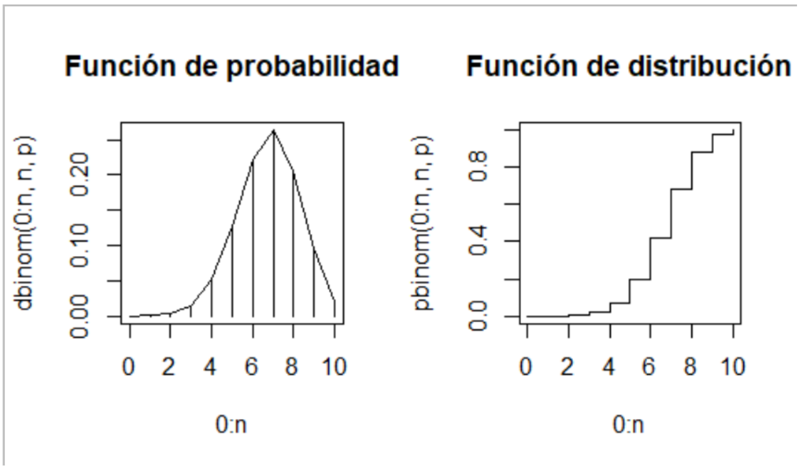
```
> (n=10)
[1] 10
> (p=round(50/74,2))
[1] 0.68
```

b) Calcular la esperanza y varianza de la variable aleatoria

```
> (Esperanza = round(n*p,2))
[1] 6.8
> (Varianza = round(n*p*(1-p),2))
[1] 2.18
```

c) Dibujar las funciones de probabilidad y distribución.

```
> par(mfrow=c(1,2))
> plot(0:n,dbinom(0:n,n,p),type="h",main="Función de probabilidad")
> lines(0:n,dbinom(0:n,n,p))
> plot(0:n,pbinom(0:n,n,p),type="s",main="Función de distribución")
> par(mfrow=c(1,1))
```



d) Cuál es la probabilidad de que exactamente 5 de los 10 estudiantes tengan móvil con Android

Se trata de calcular P{X=5}, usando la función de probabilidad `dbinom()`:

```
> (P.5=round(dbinom(5,n,p),2))
[1] 0.12
```

e) Cuál es la probabilidad de que 5 o menos de los 10 estudiantes tengan móvil con Android

Se trata de calcular P{X≤5}. Se puede usar, por tanto la función de distribución `pbinom()`:

```
> (P.menor.igual.5=round(pbinom(5,n,p),2))
[1] 0.19
```

También se podría calcular como P{X≤5} = P{X=0}+P{X=1}+ P{X=2}+P{X=3}+ P{X=4}+P{X=5}, es decir usando la función de probabilidad en lugar de la de distribución:

```
> (P.menor.igual.5=round(dbinom(0,n,p)+dbinom(1,n,p)+dbinom(2,n,p)+dbinom(3,n,p)
+dbinom(4,n,p)+dbinom(5,n,p),2))
[1] 0.19
```

f) Cuál es la probabilidad de que tengan móvil con Android más de 5 de los 10 estudiantes

Como ya sabemos P{X≤5}, se puede calcular como P{X>5}=1- P{X≤5}:

```
> (P.mayor.5=1-P.menor.o.igual.5)
[1] 0.81
```

También se podría calcular como P{X>5}=P{X=6}+ P{X=7}+P{X=8}+ P{X=9}+P{X=10}:

```
>
(P.mayor.5=round(dbinom(6,n,p)+dbinom(7,n,p)+dbinom(8,n,p)+dbinom(9,n,p)+dbinom(1
0,n,p),2))
[1] 0.81
```

g) Cuál es la probabilidad de que los 10 estudiantes tengan móvil con Android

```
> (P.todos=round(dbinom(10,n,p),2))
[1] 0.02
```

h) Cuál es la probabilidad de que ninguno de los 10 estudiantes tenga móvil con Android

```
> (P.ninguno=dbinom(0,n,p))
[1] 1.287642e-05
```

Para que el resultado no se obtenga con notación científica, puede cambiarse la opción scipen antes de ejecutar el cálculo. Para volver a la notación científica puede ejecutarse `options(scipen=0)`.

```
> options(scipen = 10)
> (P.ninguno=dbinom(0,n,p))
[1] 0.00001287642
```

i) Cuál es la probabilidad de que entre 6 y 8 estudiantes tengan móvil con Android

Se trata de calcular P{X=6}+P{X=7}+P{X=8}:

```
> (P.entre.6.y.8=round(dbinom(6,n,p)+dbinom(7,n,p)+dbinom(8,n,p),2))
[1] 0.69
```

También se puede calcular como P{X≤8}-P{X≤5}, es decir usando la función de distribución en lugar de la de probabilidad:

```
> (P.entre.6.y.8=round(pbinom(8,n,p)-pbinom(5,n,p),2))
[1] 0.69
```

j) Cuál es el valor de los cuartiles de la variable aleatoria (es decir los percentiles 25%, 50%, 75%, 100%, o cuantiles 0.25, 0.50, 0.75, 1)

En lugar de ejecutar 4 veces la función qbinom(), una para cada cuartil, se puede pasar como primer argumento, un vector con los valores de los cuantiles correspondientes a cada cuartil:

```
> qbinom(c(0.25,0.50,0.75,1),n,p)
[1]  6  7  8 10
```

Lo que indican estos resultados es que:

- P{X≤6} ≥ 0.25
- P{X≤7} ≥ 0.50
- P{X≤8} ≥ 0.75
- P{X≤10} = 1

Puede comprobarse que es cierto con la función de distribución, viendo el diagrama de esta función en el apartado c) o calculando lo valores con pbinom(). Por ejemplo, en el caso del cuartil 0.50, no puede ser el valor 6, porque P{X≤6} es 0.40 y por tanto no llega a 0.50.

```
> round(pbinom(6,n,p),2)
[1] 0.4
> round(pbinom(7,n,p),2)
[1] 0.67
```

k) Si tenemos los datos de la encuesta original a los 74 estudiantes, comprobar si tomando las respuestas de la encuesta en grupos de 10 estudiantes, el histograma se ajusta realmente a la función de probabilidad binomial.

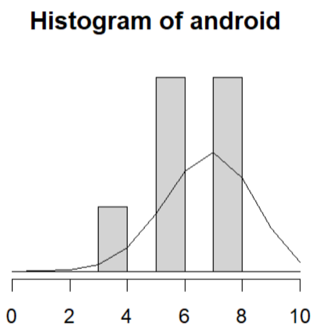
Las respuestas sobre el sistema operativo que utilizaban los estudiantes fueron:

```
> encuesta$SO
[1] "Android" "iOS"     "Android" "iOS"     "Android" "Android" "Android" "Android" "iOS"     "iOS"
[11] "iOS"     "Android" "Android" "iOS"     "Android" "iOS"     "Android" "iOS"     "Android" "Android"
[21] "Android" "iOS"     "Android" "iOS"     "Android" "iOS"     "Android" "Android" "Android" "iOS"
[31] "Android" "Android" "iOS"     "Android" "Android" "iOS"     "Android" "Android" "Android" "Android"
[41] "iOS"     "Android" "Android" "Android" "Android" "Android" "iOS"     "Android" "Android" "Android"
[51] "iOS"     "iOS"     "iOS"     "Android" "Android" "iOS"     "iOS"     "Android" "iOS"     "Android"
[61] "Android" "iOS"     "Android" "iOS"     "Android" "Android" "Android" "Android" "Android" "Android"
[71] "Android" "Android" "Android" "Android"
```

Al aparecer los resultados en bloques de 10 estudiantes, podemos contar las veces que aparece "Android" en cada uno de los 7 bloques de 10 estudiantes, y se obtiene: 6, 6, 6, 8, 8, 4, 8.

Podemos crear un vector con esos valores y dibujar su histograma y superponer las líneas de la función de probabilidad binomial.

```
> (android = c(6,6,6,8,8,4,8))
[1] 6 6 6 8 8 4 8
> hist(android, breaks=0:n, freq=FALSE)
> lines(0:n,dbinom(0:n,n,p))
```

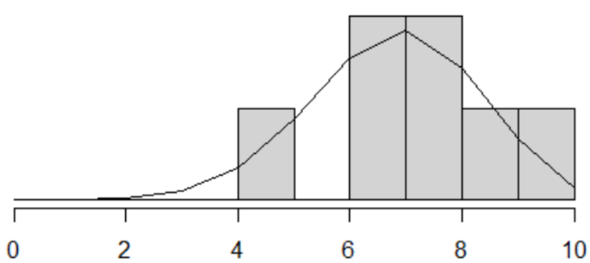


Puede comprobarse visualmente que el histograma con los datos reales se ajusta aproximadamente a la forma de la distribución binomial usada en este ejemplo.

l) Repetir el apartado anterior pero utilizando datos generados aleatoriamente con R, en lugar de los datos reales

Repetimos los comandos del apartado anterior, pero en este caso generamos aleatoriamente una muestra con los resultados de repetir 7 veces el experimento con los mismos valores de n y p que en el caso anterior, pero ahora utilizando la función `rbinom()`.

```
> (muestra=rbinom(7,n,p))
[1]  8  9  7  7 10  5  8
> hist(muestra, breaks = 0:n, freq=FALSE)
> lines(0:n,dbinom(0:n,n,p))
```



Si se ejecuta la función `rbinom()` varias veces, en cada ocasión se obtendrá valores diferentes, pero en general puede comprobarse que se ajustan a la distribución binomial.

3. Variables discretas con distribución de Poisson

Una variable aleatoria que tiene una distribución de Poisson es una variable discreta (X) que representa el número de eventos "raros" que ocurren en un determinado periodo de tiempo conociendo la frecuencia o número medio de esos sucesos que ocurren por unidad de tiempo (λ)[1]. Se representa como X:P(λ).

Por ejemplo, una variable X:P(10) que representa: "*Número de visitas a un sitio web, sabiendo que la media de visitas en una hora es de 10 (λ=10)"*.

En R se pueden utilizar las funciones indicadas en la siguiente tabla. En la columna "Ejemplo de uso", se utiliza una variable P(10).

Función	Código R	Resultado	Ejemplo de uso
Masa de probabilidad P{X=x}	`dpois(x,λ)`	Calcula la probabilidad de que una variable aleatoria de Poisson X:P(λ) tenga el valor x.	`> dpois(15,10)` `[1] 0.03471807`
Distribución de probabilidad F(X)=P{X≤x}	`ppois(x,λ)`	Calcula la probabilidad de que una variable aleatoria de Poisson X:P(λ) tenga un valor igual o menor que x.	`> ppois(15,10)` `[1] 0.9512596`

Cuantil	`qpois(q,λ)`	Calcula el menor valor de una variable aleatoria de Poisson X:P(λ) que cumple que la probabilidad de que la variable sea menor o igual a ese valor es mayor o igual a q.	`> qpois(0.7,10)`
			`[1] 12`
			
			Entonces se cumple:
			P{X≤12} ≥ 0.7
			Pero no se cumple:
			P{X≤11} ≥ 0.7

Generación aleatoria de muestras	`rpois(m,λ)`	Genera aleatoriamente un vector de m valores aleatorios de una variable de Poisson X:P(λ).	`> rpois(4,10)` `[1]  7  9  8 16`
Esperanza E[X]	`λ`		`>(E=10)` `[1] 10`
Varianza Var[X]	`λ`		`>(Var=10)` `[1] 10`

[1] El símbolo λ (letra griega lambda) se escribe pulsando Alt + 955.

Las funciones anteriores se pueden combinar con otras de generación de gráficos, para obtener los diagramas que se indican en la siguiente tabla.

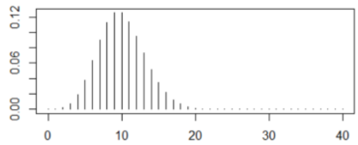
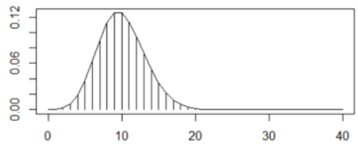
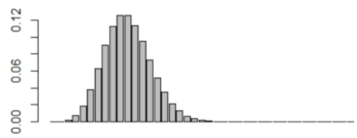
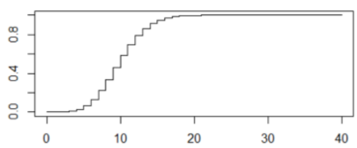
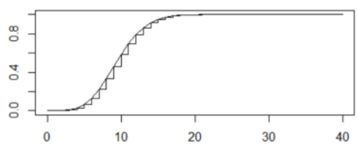
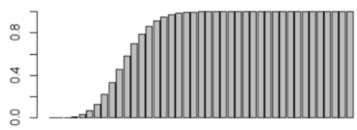
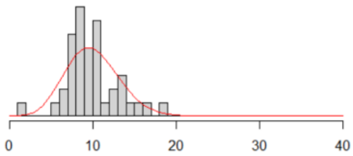
Diagrama	Código R	Comentarios	Ejemplo de uso
Función de masa de probabilidad	`plot(0:n,dpois(0:n,λ),type=tipo)`	`tipo` puede ser: `"l"`: líneas `"h"`: barras `"b"`: líneas y barras	`>plot(0:40,dpois(0:40,10),type="h")` 
Función de masa de probabilidad (línea)	`lines(0:n,dpois(0:n,λ))`	Los diagramas de línea se dibujan sobre el último diagrama plot o barplot.	`>lines(0:40,dpois(0:40,10))` 
Función de masa de probabilidad (barras)	`barplot(dpois(0:n,λ))`		`>barplot(dpois(0:40,10))` 
Función de distribución de probabilidad	`plot(0:n,ppois(0:n,λ),type="s")`	`type` también puede ser: `"l"`: líneas `"h"`: barras `"b"`: líneas y barras	`>plot(0:40,ppois(0:40,10), type="s")` 
Función de distribución de probabilidad (línea)	`lines(0:n,ppois(0:n,λ))`	Los diagramas de línea se dibujan sobre el último diagrama plot o barplot.	`>lines(0:40,ppois(0:40,10))` 
Función de distribución de probabilidad (barras)	`barplot(ppois(0:n,λ))`		`>barplot(ppois(0:40,10))` 

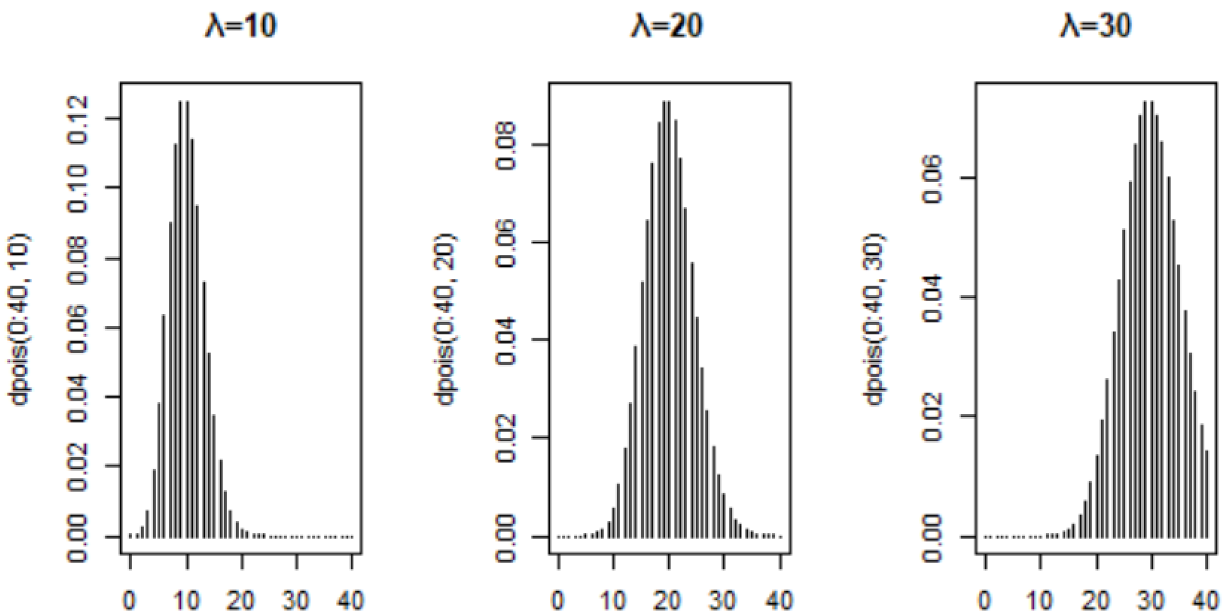
Diagrama	Código R	Comentarios	Ejemplo de uso

Histograma de una simulación de posibles valores de la variable aleatoria	`hist(rpois(m,λ), breaks=0:m, freq=FALSE)`
	
	`lines(0:m,dpois(0:m, λ))`
		Se puede representar primero el histograma con los datos de la simulación y después la función de probabilidad de Poisson teórica, para ver si son similares.
		NOTA: Se podría hacer lo mismo si se dispone de datos reales, para ver si las muestras reales cumplen tienen una distribución de Poisson.
			`>hist(rpois(40,10), breaks=0:40, freq=FALSE)`
			
			`>lines(0:40,dpois(0:40,10),col="red")`
			

Se puede comparar el efecto de modificar el parámetro λ, mostrando simultáneamente varios gráficos.

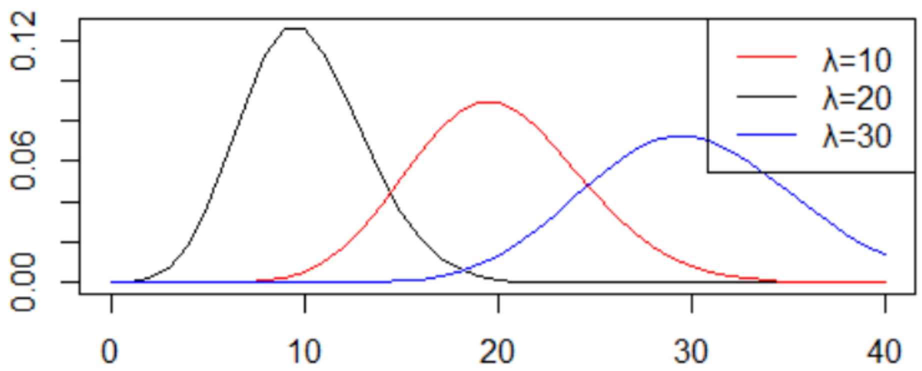
Por ejemplo dividiendo la ventana gráfica en 3 columnas:

```
> par(mfrow=c(1,3))
> plot(0:40,dpois(0:40,10), type="h", main = "λ=10")
> plot(0:40,dpois(0:40,20), type="h", main = "λ=20")
> plot(0:40,dpois(0:40,30), type="h", main = "λ=30")
> par(mfrow=c(1,1))
```



O se pueden dibujar varios diagramas en el mismo gráfico, el primero con plot y el resto con lines:

```
> plot(0:40,dpois(0:40,10), type="l")
> lines(0:40,dpois(0:40,20), col="red")
> lines(0:40,dpois(0:40,30), col="blue")
> legend("topright",legend=c("λ=10","λ=20","λ=30"),
       col=c("red", "black", "blue"), bg="transparent", lty=1)
```



Se recomienda investigar diferentes formas de dibujar la leyenda en el gráfico.

3.1 Ejemplo: Correo electrónico no deseado (spam)

ENUNCIADO

Una persona ha comprobado que recibe una media de 5 mensajes al día de correo electrónico no deseado (spam).

- **a) Definir una variable aleatoria de Poisson para el número de mensajes no deseados recibidos en una semana (7 días)**
- **b) Calcular la esperanza y varianza de la variable aleatoria**
- **c) Dibujar las funciones de probabilidad y distribución de la variable aleatoria**
- **d) Cuál es la probabilidad de que se reciban 20 mensajes no deseados en una semana**
- **e) Cuál es la probabilidad de que se reciban 20 o menos mensajes no deseados en una semana**
- **f) Cuál es la probabilidad de que se reciban más de 20 mensajes no deseados en una semana**
- **g) Cuál es la probabilidad de que no se reciba ningún mensaje en una semana**
- **h) Cuál es la probabilidad de que se reciban entre 20 y 30 mensajes no deseados en una semana**
- **i) Cuál es el valor de los cuartiles de la variable aleatoria**
- **j) Generar aleatoriamente 100 valores para la variable y comprobar si el histograma se ajusta realmente a la función de probabilidad de Poisson.**

k) **Definir una nueva variable aleatoria para calcular la probabilidad de que en un día se reciban 3 mensajes.**
l) **Generar aleatoriamente 100 valores para la nueva variable y comprobar si el histograma se ajusta realmente a la función de probabilidad de Poisson.**

SOLUCIÓN

a) Definir una variable aleatoria de Poisson para el número de mensajes no deseados recibidos en una semana (7 días)

X = Mensajes no deseados recibidos en una semana, sabiendo que la media de mensajes no deseados recibidos en un día es de 5 mensajes.

La variable sigue una distribución de Poisson P(λ), donde λ=5*7=35 mensajes/semana. Por simplicidad se utiliza "lambda" en lugar de "λ" en el código R.

```
> (lambda=35)
[1] 35
```

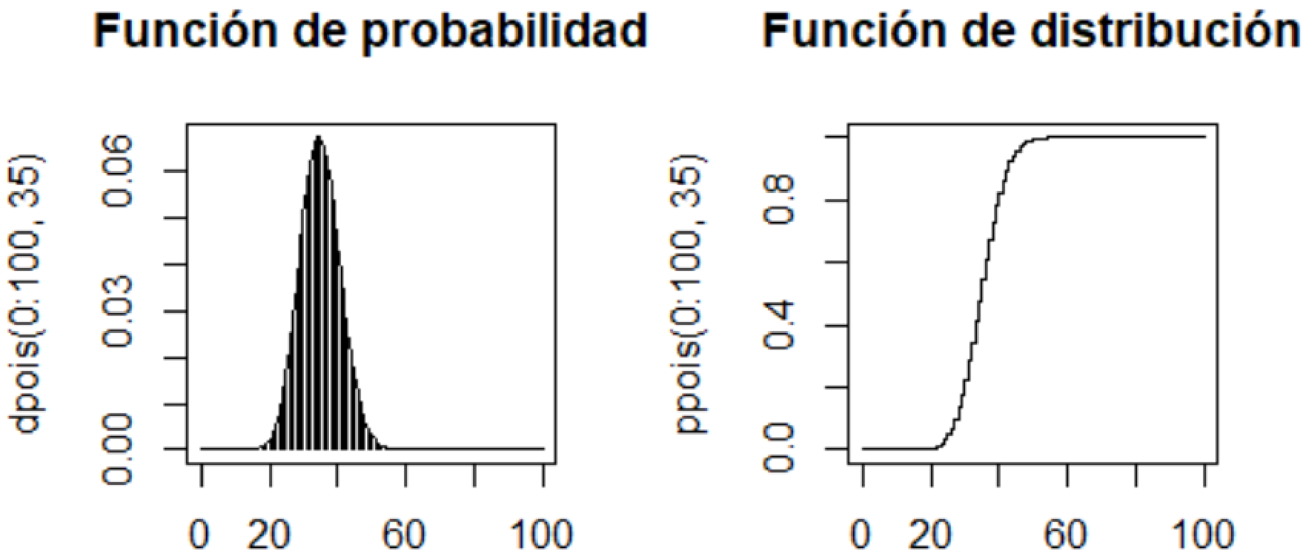
b) Calcular la esperanza y varianza de la variable aleatoria

La esperanza y varianza coinciden con λ, por lo que su valor es 35.

```
> (Esperanza = lambda)
[1] 35
> (Varianza = lambda)
[1] 35
```

c) Dibujar las funciones de probabilidad y distribución de la variable aleatoria

```
> par(mfrow=c(1,2))
> plot(0:100,dpois(0:100,lambda),type="h",main="Función de probabilidad")
> lines(0:100,dpois(0:100,lambda))
> plot(0:100,ppois(0:100,lambda),type="s",main="Función de distribución")
> par(mfrow=c(1,1))
```



d) Cuál es la probabilidad de que se reciban 20 mensajes no deseados en una semana

Se trata de calcular P{X=20}, usando la función de probabilidad `dpois()`:

```
> dpois(20,lambda)
[1] 0.001972102
```

e) Cuál es la probabilidad de que se reciban 20 o menos mensajes no deseados en una semana

Se trata de calcular P{X≤20}. Se puede usar, por tanto la función de distribución `ppois()`:

```
> ppois(20,lambda)
[1] 0.004296609
```

f) Cuál es la probabilidad de que se reciban más de 20 mensajes no deseados en una semana

Como ya sabemos P{X≤20}, se puede calcular como P{X>20}=1- P{X≤20}:

```
> 1-ppois(20,lambda)
[1] 0.9957034
```

g) Cuál es la probabilidad de que no se reciba ningún mensaje en una semana

```
> dpois(0,lambda)
[1] 6.305117e-16
```

Para que el resultado no se obtenga con notación científica, puede cambiarse la opción scipen antes de ejecutar el cálculo. Para volver a la notación científica puede ejecutarse `options(scipen=0)`.

```
> options(scipen = 100)
> dpois(0,lambda)
[1] 0.0000000000000006305117
```

h) Cuál es la probabilidad de que se reciban entre 20 y 30 mensajes no deseados en una semana

Se trata de calcular P{X=20}+P{X=21}+…+P{X=30}, pero es más fácil calcularlo como P{X≤30}-P{X≤20}, usando la función de distribución en lugar de la de probabilidad:

```
> ppois(30,lambda)-ppois(19,lambda)
[1] 0.2246179
```

i) Cuál es el valor de los cuartiles de la variable aleatoria (es decir los percentiles 25%, 50%, 75%, 100%, o cuantiles 0.25, 0.50, 0.75, 1)

En lugar de ejecutar 4 veces la función qpois(), una para cada cuartil, se puede pasar como primer argumento, un vector con los valores de los cuantiles correspondientes a cada cuartil:

```
> qpois(c(0.25,0.50,0.75,1),lambda)
[1]  31  35  39 Inf
```

Lo que indican estos resultados es que:

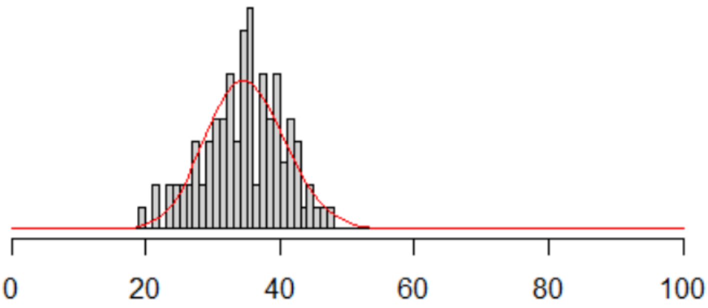
- P{X≤31} ≥ 0.25
- P{X≤35} ≥ 0.50
- P{X≤39} ≥ 0.75
- P{X≤Infinito} = 1

Puede comprobarse que es cierto con la función de distribución, viendo el diagrama de esta función en el apartado c) o calculando lo valores con ppois(). Por ejemplo en el caso del cuantil 0.25, no puede ser el valor 30, porque P{X≤30} es 0.23 y por tanto no llega a 0.25.

```
> round(ppois(30,lambda),2)
[1] 0.23
> round(ppois(31,lambda),2)
[1] 0.28
```

j) Generar aleatoriamente 100 valores para la variable y comprobar si el histograma se ajusta realmente a la función de probabilidad de Poisson.

```
> (muestra=rpois(100,lambda))
  [1] 39 39 35 32 40 27 34 39 22 30 25 31 24 28 43 30 36 32 38 36 45 35
 [23] 44 28 41 26 38 32 38 35 35 42 40 29 38 34 42 31 33 37 43 40 34 46
 [45] 28 40 37 33 24 36 36 40 35 31 43 33 38 36 48 41 32 36 30 31 33 27
 [67] 36 25 35 35 41 47 26 20 28 40 36 40 31 42 22 38 39 36 33 30 35 33
 [89] 35 42 36 29 42 32 45 33 43 34 39 38
> hist(muestra, breaks = 0:100, freq=FALSE)
> lines(0:100,dpois(0:100,lambda))
```



Si se ejecuta la función `rpois()` varias veces, en cada ocasión se obtendrá valores diferentes, pero en general puede comprobarse que se ajustan a la distribución de Poisson.

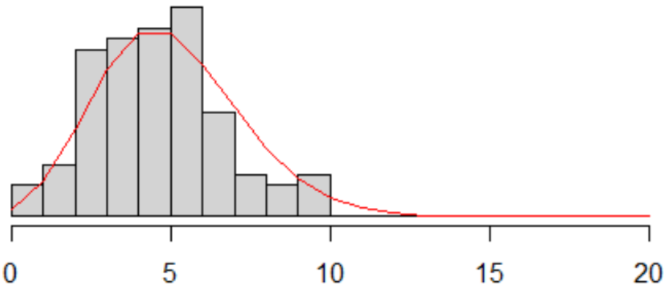
k) Definir una nueva variable aleatoria para calcular la probabilidad de que en un día se reciban 3 mensajes.

En este caso, la nueva variable sería “Mensajes no deseados recibidos en un día, sabiendo que la media de mensajes no deseados recibidos en un día es de 5 mensajes”, y el parámetro λ de la variable sería 5.

```
> dpois(3,5)
[1] 0.1403739
```

l) Generar aleatoriamente 100 valores para la nueva variable y comprobar si el histograma se ajusta realmente a la función de probabilidad de Poisson.

```
> (muestra=rpois(100,5))
  [1]  6  4  9  5  2  5  3  7  3  4  3  8  3  3  3  4  5  6  7  4  3  3
 [23]  5  5  6  5  4  4  6  6  4  6  6  4  6  6  5  3  2  1  1  4  5  4
 [45]  8 10  5  5  6  7  7  2  6  6  5  8  6  8  6  6  7  9  5  4  5  3
 [67]  9  7  6  6  2  7  5  4  4  3  6  3  4 10  1  7  7  6  3  5  5  7
 [89] 10  4 10  4  5  3  4  6  3  3  2  5
> hist(muestra, breaks = 0:20, freq=FALSE)
> lines(0:20,dpois(0:20,5),col="red")
```



3.2 Aproximar una variable binomial a una distribución de Poisson

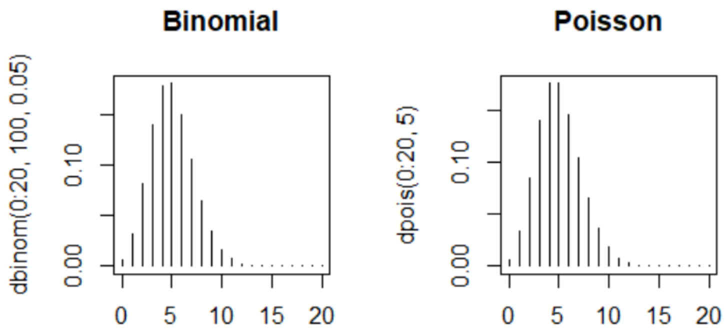
Si en una variable binomial B(n,p) se cumple que n es grande y p pequeño[2], entonces se puede aproximar a una distribución de Poisson P(λ) con λ=n*p.

Se puede comprobar por ejemplo con B(100,0.05), aproximándola a una de Poisson con λ=100*0.05=5.

```
> dbinom(2,100,0.05)
[1] 0.08118177
> dpois(2,5)
[1] 0.08422434
```

Podemos comparar gráficamente sus funciones de probabilidad:

```
> par(mfrow=c(1,2))
> plot(0:20,dbinom(0:20,100,0.05),type="h",main="Binomial")
> plot(0:20,dpois(0:20,5),type="h",main="Poisson")
> par(mfrow=c(1,1))
```



[2] No hay unanimidad respecto a lo que se entiende por “grande” y “pequeño”. Algunos autores afirman que la aproximación es “buena” si n≥20 y p≤0,05 y “muy buena” si n≥100 y p ≤0,01, otros que se cumpla n≥30 y np<10 o n(1-p)<10, otros que n>30 y p<0.1, otros que es suficiente que np<5, otros que np>1 y p<0.1, etc.

4. Otras distribuciones discretas

En R se pueden utilizar otras distribuciones de probabilidad para variables aleatorias discretas, y para todas ellas existen las funciones de probabilidad o masa, de distribución, de cálculo de cuantiles y de generación de valores aleatorios. Se trata de las siguientes:

Distribución	Función de probabilidad (masa)	Función de distribución	Cuantiles	Generación valores aleatorios
Binomial negativa	`dnbinom()`	`pnbinom()`	`qnbinom()`	`rnbinom()`
Geométrica	`dgeom()`	`pgeom()`	`qgeom()`	`rgeom()`
Hipergeométrica	`dhyper()`	`phyper()`	`qhyper()`	`rhypr()`
Multinomial	`dmultinom()`	`pmultinom()`	`qmultinom()`	`rmultinom()`

Para ver la ayuda sobre todas las distribuciones en R se debe ejecutar el comando `help("Distributions")`.

Se pueden encontrar ejemplos en R en https://fhernanb.github.io/Manual-de-R/discretas.html. Y los fundamentos teóricos de éstas y otras distribuciones de variables discretas en https://es.m.wikipedia.org/wiki/Distribución_de_probabilidad.

5. Ejercicios propuestos

1) Si se lanza 50 veces un dado de seis lados con valores 1, 2, 3, 4, 5 y 6 en cada lado:
   a. Definir una variable aleatoria binomial para el experimento
   b. Calcular la esperanza y varianza de la variable aleatoria
   c. Dibujar las funciones de probabilidad y distribución de la variable aleatoria
   d. Calcular la probabilidad de que nunca aparezca el lado con el valor 6.
   e. Calcular la probabilidad de que siempre aparezca el lado con el valor 6.
   f. Calcular la probabilidad de que el lado con el valor 6 aparezca menos de 10 veces.
   g. Calcular la probabilidad de que el lado con el valor 6 aparezca más de 10 veces.
   h. Calcular la probabilidad de que el lado con el valor 6 aparezca entre 20 y 40 veces.
   i. Calcular los cuantiles 50 y 90.
   j. Generar aleatoriamente 100 valores para la variable y comprobar si el histograma se ajusta realmente a una función de probabilidad Binomial.

2) Mediante una encuesta cuyas respuestas están disponibles en el archivo encuesta.csv, y después de eliminar las respuestas de 5 personas que no contestaron alguna pregunta, se sabe que 26 de 74 estudiantes de un curso de la asignatura Estadística del Grado en Ingeniería en Sistemas de Información de la Universidad de Alcalá tienen una línea telefónica contratada con las compañías Jazzte/Orange/Simyo. Si se pregunta aleatoriamente a 6 de los estudiantes, responder a las siguientes preguntas:
   a. Definir una variable aleatoria binomial para el experimento
   b. Calcular la esperanza y varianza de la variable aleatoria
   c. Dibujar las funciones de probabilidad y distribución de la variable aleatoria
   d. Cuál es la probabilidad de que exactamente 3 de los 6 estudiantes tengan contrato con Jazzte/Orange/Simyo
   e. Cuál es la probabilidad de que 3 o menos de los 6 estudiantes tengan contrato con Jazzte/Orange/Simyo
   f. Cuál es la probabilidad de que tengan contrato con Jazzte/Orange/Simyo más de 3 de los 6 estudiantes
   g. Cuál es la probabilidad de que los 6 estudiantes tengan contrato con Jazzte/Orange/Simyo
   h. Cuál es la probabilidad de que ninguno de los 6 estudiantes tenga contrato con Jazzte/Orange/Simyo
   i. Cuál es la probabilidad de que entre 2 y 4 estudiantes tengan contrato con Jazzte/Orange/Simyo
   j. Cuál es el valor de los cuartiles de la variable aleatoria
   k. Si tenemos los datos de la encuesta original a los 74 alumnos, comprobar si tomando las respuestas en grupos de 6 alumnos, el histograma se ajusta realmente a la función de probabilidad binomial.
   l. Repetir el apartado anterior pero utilizando datos generados aleatoriamente con R, en lugar de los datos reales

3) Se ha comprobado que un sitio web tiene una media de 120 visitas al día.
   a. Definir una variable aleatoria de Poisson para el número de visitas en una hora
   b. Calcular la esperanza y varianza de la variable aleatoria
   c. Dibujar las funciones de probabilidad y distribución de la variable aleatoria
   d. Cuál es la probabilidad de que haya 4 visitas en una hora
   e. Cuál es la probabilidad de que haya 4 o menos visitas en una hora
   f. Cuál es la probabilidad de que haya más de 4 visitas en una hora
   g. Cuál es la probabilidad de que no haya ninguna visita en una hora
   h. Cuál es la probabilidad de que haya entre 4 y 6 visitas en una hora
   i. Cuál es el valor de los cuartiles de la variable aleatoria
   j. Generar aleatoriamente 100 valores para la variable y comprobar si el histograma se ajusta realmente a la función de probabilidad de Poisson.
   k. Definir una nueva variable aleatoria para calcular la probabilidad de que en un día haya 100 visitas.
   l. Generar aleatoriamente 100 valores para la nueva variable y comprobar si el histograma se ajusta realmente a la función de probabilidad de Poisson.