

Probabilidad con R

Para ser un científico de datos se requiere tener conocimientos de probabilidad y estadística, por lo que en este laboratorio repasaremos las principales funciones de probabilidad de R, las cuales son de mucha utilidad en estadística para el análisis de variables.

Distribución discreta binomial

Comenzarás con la distribución discreta de probabilidad llamada distribución binomial, la cual sirve para tipificar el número de éxitos que se tiene como resultado de n experimentos o ensayos.

Por ejemplo, si en la Universidad Tecmilenio se te aplica un examen rápido que consiste en 10 preguntas, donde cada una de ellas tiene cuatro posibles respuestas y solo una es correcta. Calcula la probabilidad de que pases el examen con 7 respuestas correctas de 10 (la probabilidad de contestar una respuesta correcta es una de cada cuatro, que es 0.25). Para lo anterior, utilizarías la función `dbinom()` de la siguiente forma:

```
> dbinom(7, size = 10, prob = 0.25)
[1] 0.003089905
```

La probabilidad de pasar el examen de forma aleatoria es del 0.3%. Ahora, si lo que deseas es saber la probabilidad de sacar aleatoriamente 10 de 10:

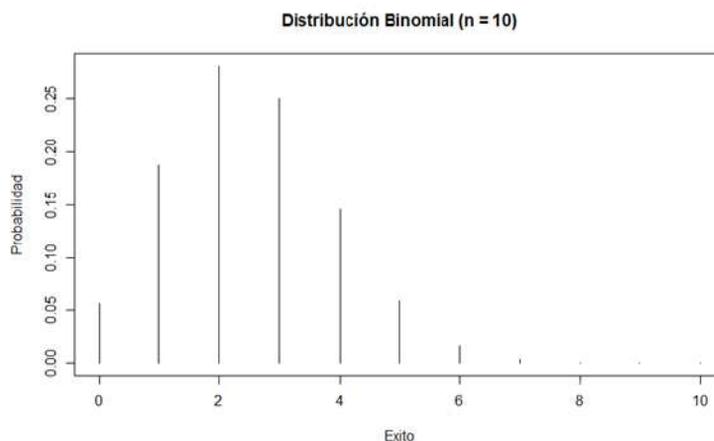
```
> dbinom(10, size = 10, prob = 0.25)
[1] 9.536743e-07
```

Por otro lado, podrías saber las probabilidades binomiales acumuladas con la función `pbinom()`. Si regresamos al ejemplo original, ¿cuál sería la probabilidad de obtener 7 o menos?

```
> pbinom(7, size = 10, prob = 0.25)
[1] 0.9995842
```

Asimismo, se puede graficar la distribución con la función `plot`. A continuación, grafica la función de distribución sobre la probabilidad de obtener el pase del examen con 7 aleatoriamente:

```
> plot(x, dbinom(x, size=10, prob=0.25), xlab="Exito", ylab="Probabilidad", main="Distribución Binomial (n = 10)", type="h")
```



Ahora realizaremos el siguiente ejemplo, ¿cuál es la probabilidad de que al tirar una moneda obtenga 7 caras en 20 tiros si la probabilidad de obtener caras es de 0.45?

Para eso, ejecutaremos la siguiente instrucción de código:

```
> dbinom(7, size = 20, prob = 0.45)
[1] 0.1220721
```

Por otro lado, se ejecutará la siguiente instrucción si lo que deseamos es obtener la probabilidad de tener más de 12 caras o menos bajo las mismas condiciones establecidas en el problema original:

```
> pbinom(12, size = 20, prob = 0.45)
[1] 0.9419659
```

Ahora es tu turno de practicar con el siguiente ejercicio:

Ejercicio:

En un concurso de encestar balones de basquetbol en tiros de 3 puntos se desea conocer la probabilidad de que:

- En un concurso de 10 tiros de 3 puntos, cuál la probabilidad de encestar 3 veces si la probabilidad de encestar es de 0.3.
- ¿Cuál es la probabilidad de encestar 5 o más tiros de 10?
- Realiza las gráficas de probabilidad de éxito para ambos casos.

Distribución discreta de Poisson

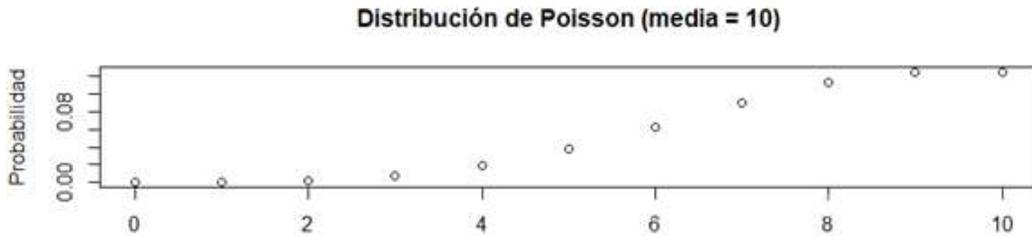
Esta distribución de probabilidad se utiliza para tipificar ocurrencias o eventos independientes, por ejemplo, cuando se desee conocer la probabilidad de que cierto producto sea vendido en línea 10 veces en un minuto, para lo cual necesitarías conocer el promedio de ventas por minuto de ese producto (en este caso es 5).

Si usas la función `dpois()` se tiene la siguiente instrucción:

```
> dpois(10, 5)
[1] 0.01813279
```

La probabilidad es del 1.81% para poder graficar las probabilidades de:

```
> x <- 0:10
> plot(x, dpois(x, lambda = 10), xlab = "x", ylab = "Probabilidad", main = "Distribución de Poisson (media = 10)")
```



Distribución continua exponencial

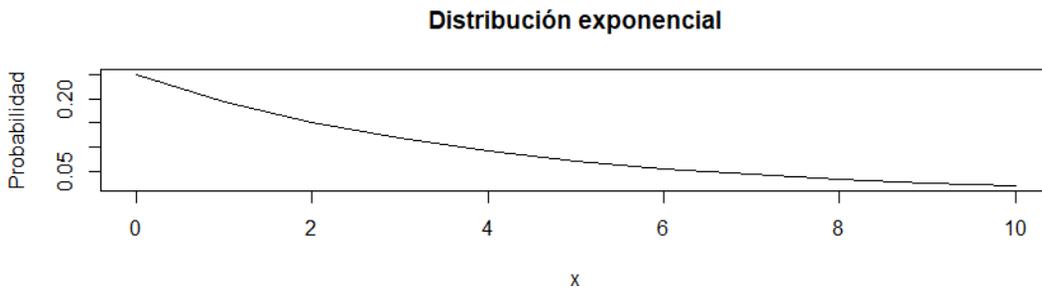
Es una distribución que tiene su aplicación en procesos donde queremos saber la probabilidad en tiempo de que un evento suceda, independientemente de los tiempos pasados. Para usarla en lenguaje R existe la función *pexp()*, por ejemplo, en una panadería se atienden en promedio 15 clientes por hora, ¿cuál es la probabilidad de que llegue un cliente en el próximo minuto?

Se puede graficar la distribución con la función *plot*. A continuación, grafica la función de distribución de probabilidad de obtener el pase del examen con 7 de forma aleatoria:

```
> pexp(1, rate = 1/4)
[1] 0.2211992
```

Para graficar vas a considerar la probabilidad desde 0 hasta 10 clientes por minuto:

```
> x <- 0:10
> plot(x, dexp(x, rate = 1/4), xlab = "x", ylab = "Probabilidad", main = "Distribución exponencial", type="l")
```

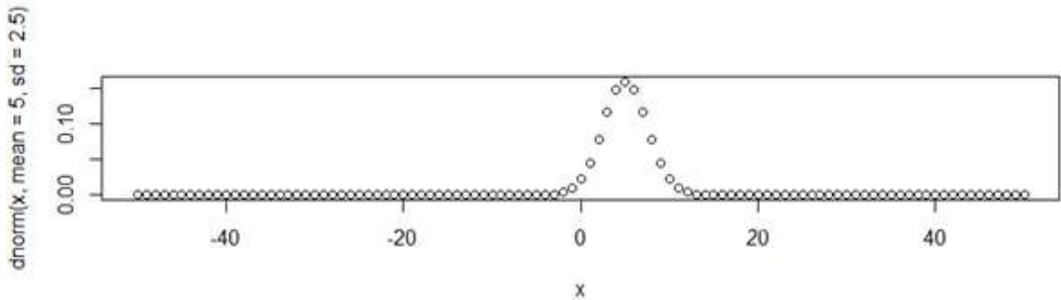


Distribución continua normal

También conocida como de Gauss o campana Gaussiana (por su forma de campana simétrica respecto a la media), es la más utilizada debido a la cantidad de aplicaciones que tiene.

A manera de ejemplo, genera una secuencia de números entre -50 y +50, después grafica la distribución normal asumiendo una media de 5 y una desviación estándar de 2.5:

```
> x <- seq(-50, 50)
> plot(x, dnorm(x, mean = 5, sd = 2.5))
```



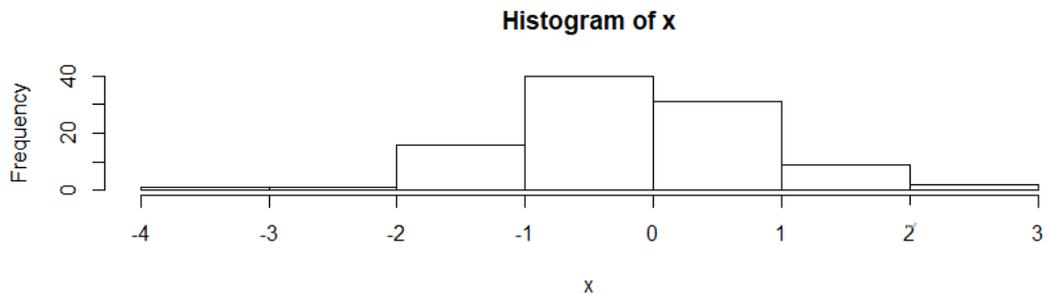
Si la distribución etaria (por edades) de empleados de una empresa tiene una distribución normal, donde la media es 40 y la desviación estándar es 13, ¿qué porcentaje de empleados tienen más de 50 años?

```
> pnorm(50, mean=40, sd=13, lower.tail=FALSE)
[1] 0.2208782
```

La respuesta es 22.08%.

Cuando deseas generar números aleatorios que sigan una distribución normal puedes utilizar la función `rnorm()`, como lo puedes observar en el siguiente ejemplo:

```
> x <- rnorm(100)
> hist(x)
```



Distribución t student

Esta se utiliza cuando se necesita estimar la media de una población normal, pero que cuenta con una muestra pequeña o cuando se desconoce la dispersión que tiene una variable.

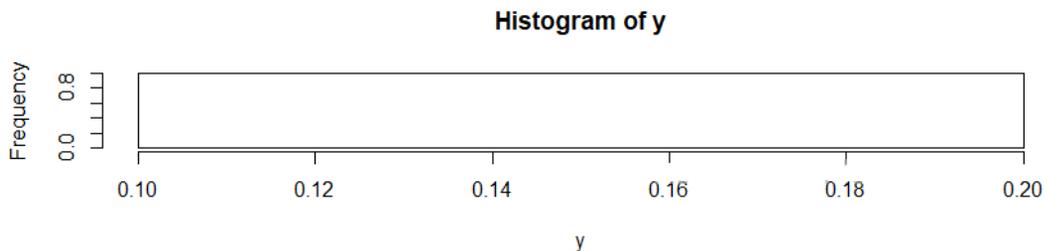
Supongamos que deseas calcular la probabilidad de que la calificación media de 100 estudiantes sea 60 o menos y tienes el conocimiento de que la media poblacional es de 70 y la desviación estándar es de 5.

```
> pnorm(50, mean=40, sd=13, lower.tail=FALSE)
[1] 0.2208782
```

A continuación, realizarás un ejemplo para comprender el teorema del límite central, el cual expresa que cuando se tiene la suma de un número considerable de variables la resultante tendrá una distribución normal, es decir, entre más grande sea la muestra la media muestral será la media poblacional, independientemente de que las muestras no sean normales o que tengan otro tipo de distribución.

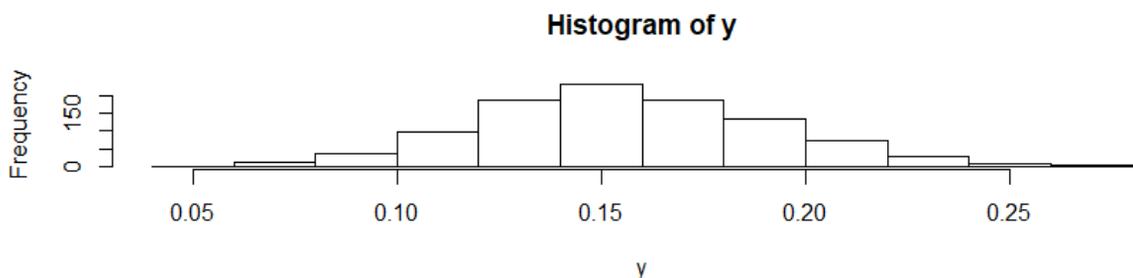
A manera de ejemplo realizarás un experimento en el cual se tomaron 100 muestras durante 1 día; si la probabilidad es de 1/6 tenemos que:

```
> x <- rbinom(1,100,0.16)
> y <- x/100
> hist(y)
```



Al realizarlo 1000 veces:

```
> x <- rbinom(1000,100,0.16)
> y <- x/100
> hist(y)
```



Puedes observar que entre más grande es la muestra, a pesar de provenir de muestras que no son normales, se tiende a una distribución normal.

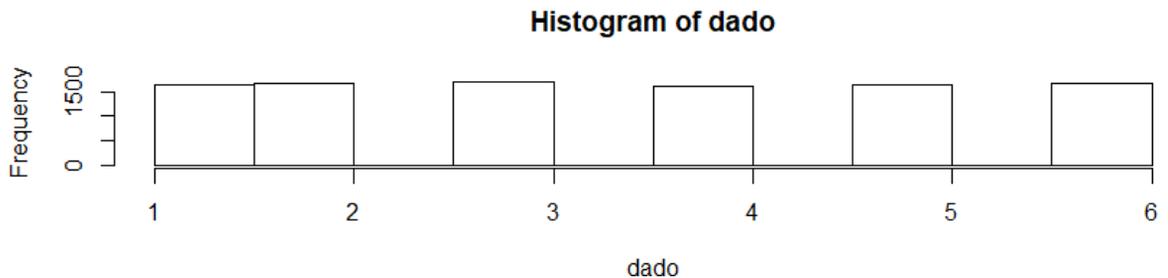
Generación de variables discretas

Para generar variables discretas es posible exponer el ejemplo del lanzamiento de un dado con la función `sample()`. En este caso, imagina que lanzas un dado 15 veces, por tanto, el código que debes escribir es el siguiente:

```
> dado=sample(1:6, 15, replace = TRUE);
> view(dado)
```

	V1
1	1
2	6
3	2
4	2
5	3
6	6
7	1
8	6
9	3
10	1
11	2
12	1
13	1
14	3
15	2

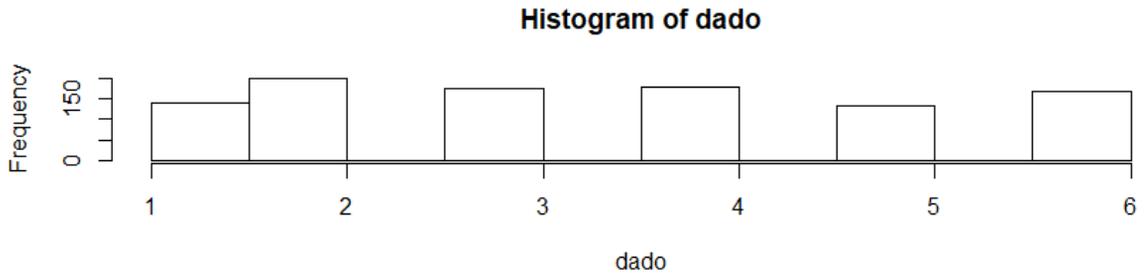
Lo que se obtiene es el siguiente histograma:



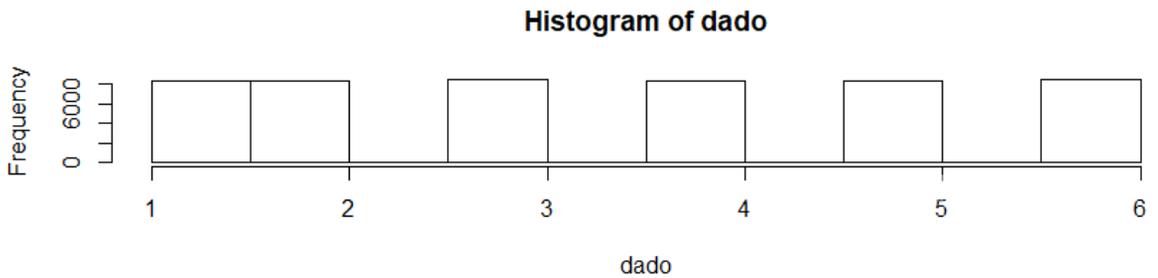
En cambio, si realizaras 1000 tiradas:

```

> dado=sample(1:6, 1000, replace = TRUE);
> hist(dado)
    
```



En 50,000 tiradas:



La obra presentada es propiedad de ENSEÑANZA E INVESTIGACIÓN SUPERIOR A.C. (UNIVERSIDAD TECMILENIO), protegida por la Ley Federal de Derecho de Autor; la alteración o deformación de una obra, así como su reproducción, exhibición o ejecución pública sin el consentimiento de su autor y titular de los derechos correspondientes es constitutivo de un delito tipificado en la Ley Federal de Derechos de Autor, así como en las Leyes Internacionales de Derecho de Autor.

El uso de imágenes, fragmentos de videos, fragmentos de eventos culturales, programas y demás material que sea objeto de protección de los derechos de autor, es exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por UNIVERSIDAD TECMILENIO.

Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de esta obra sin la autorización previa por escrito de UNIVERSIDAD TECMILENIO. Sin embargo, usted podrá bajar material a su computadora personal para uso exclusivamente personal o educacional y no comercial limitado a una copia por página. No se podrá remover o alterar de la copia ninguna leyenda de Derechos de Autor o la que manifieste la autoría del material.