

Unidad 3

# Funciones de distribución de probabilidad

# Unidad 3.

## Funciones de distribución de probabilidad

### Introducción

Tomar decisiones es un proceso al que nos enfrentamos de manera periódica y constante en cada uno de los ámbitos de nuestra vida. Usualmente, las situaciones en las que debemos elegir entre diversas opciones, implican contar con un mínimo de contexto para elegir qué elementos son relevantes y así analizar las posibles consecuencias de las acciones que emprendamos.

Por consiguiente, para tomar buenas decisiones, es necesario contar con información de calidad, que nos brinde las herramientas y la claridad suficiente sobre el objetivo, estrategia o problema a abordar. Y como ya hemos mencionado en la unidad anterior, los datos son el recurso principal para comprender los procesos y el entorno de las organizaciones, comunidades, ambientes o grupos.

Aunque para algunos, las decisiones estratégicas pueden atribuirse a la intuición, en realidad, no se trata de nada diferente a una habilidad desarrollable para poner en perspectiva los datos, depurarlos y organizarlos con el propósito de resolver problemas en favor de nuestros intereses y objetivos.

La toma de decisiones a partir de los datos proporciona diferentes ventajas para la gestión de una organización. Por tal razón, en esta segunda unidad abordaremos conceptos fundamentales como variables aleatorias y continuas, probabilidad y pruebas de bondad y ajuste; los cuales nos darán una mejor perspectiva para escoger las mejores alternativas que nos permita desarrollar ideas innovadoras o encontrar soluciones eficientes.

## Objetivos de aprendizaje:

- Identificar las funciones de distribución y de densidad de variables aleatorias.
- Utilizar diferentes propiedades de probabilidad característicos de procesos aleatorios continuos o discretos.
- Emplear herramientas de representación gráfica para la distribución probabilística discreta o continua en la resolución de problemas con datos de esta naturaleza.

## Unidad 3. Actividad de aprendizaje 1: Los datos, la predicción y la toma de decisiones

En esta actividad de aprendizaje abordaremos conceptos que nos permitirán predecir valores de población a partir de las muestras. Además, conoceremos qué es una distribución y cómo se aplica a diferentes contextos. Estos temas se presentan como la puerta de entrada a la toma de decisiones basada en datos.

- Los siguientes son los contenidos que hacen parte de la Unidad 3:
- Funciones de distribución de probabilidad
- Funciones de densidad
- Pruebas de bondad de ajuste

Al finalizar el estudio de esta unidad, estarás en capacidad de emplear los datos, y las predicciones que de ellos puedas hacer, para realizar análisis que aporten valor y otorguen una ventaja pensando en la toma de decisiones.

## Unidad 3: Funciones de distribución de probabilidad



### 1. Funciones de distribución de probabilidad y funciones de densidad

Cuando en estadística hablamos de distribución, normalmente nos referimos al proceso de calcular resultados con algún nivel de probabilidad de ocurrencia en el contexto de un determinado experimento. Por tanto, una distribución nos ayuda a describir la probabilidad de ocurrencia de un evento y de todos los posibles resultados, utilizando valores entre 0 y 1. Donde 1 significa que un evento tiene 100% de probabilidad de suceder.

Dicho de un modo un poco más simple, una distribución es una función que nos muestra los valores posibles que puede tomar una variable y la frecuencia con que ocurre.

Debido a que la distribución de probabilidad se genera a partir de una variable determinada, podemos emplear ciertas funciones para asignar probabilidades a los valores de dicha variable y así tener una distribución que se ajuste a los resultados esperados.

## 1.1. Distribuciones variables aleatorias discretas

Una variable aleatoria discreta es aquella que solo incluye valores enteros, por ejemplo, en casos como el clásico lanzamiento de monedas, el número de intentos de ejecución de una acción, la elaboración y revisión de artículos, la fabricación de productos, la identificación de incidentes, entre otros. Dentro del conjunto de distribuciones de variables aleatorias discretas, encontramos las siguientes:

- Binomial
- Binomial negativa
- Poisson
- Geométrica
- Hipergeométrica

## ● Binomial

La distribución de probabilidad de la variable aleatoria binomial mide el número de éxitos probables en  $n$  experimentos independientes, como en el lanzamiento de dados o monedas.

Es un modelo adecuado para procesos donde un determinado ejercicio se repite hasta obtener un número establecido de resultados favorables (por primera vez). El modelo que describe esta distribución de probabilidad es:

$$p(x = k) = \binom{n}{k} p^k \cdot q^{n-k}$$

### En donde:

- $n$ : es el número de pruebas.
- $k$ : es el número de éxitos.
- $p$ : es la probabilidad de éxito.
- $q$ : es la probabilidad de fracaso.



### Problema

¿Cómo podría verse un *script* que permita calcular la probabilidad de elegir al azar uno de dos proveedores, con una simulación de un centenar de eventos de selección?

### Solución

En primer lugar, definimos una variable *proveedor* con un arreglo que contenga dos elementos, uno que identifique a cada proveedor.

```
proveedor <- c("Proveedor 1", "Proveedor 2")
```

Luego, establecemos la variable *elección*, y la inicializamos en valor nulo. En otra línea, inicializamos la variable *n* con el número de simulaciones que queremos correr.

```
eleccion <-NULL
n <- 100 # numero de simulaciones
```

A continuación, utilizamos la función ***sample*** de *R*. Esta función utiliza el arreglo como una caja de la cual se pueden extraer elementos al azar. Los parámetros que se usan pertenecen al arreglo que contiene los elementos para “extraer” y la cantidad de veces que se repetirá la acción. Adicionalmente, empleamos el parámetro ***replace*** para reintroducir el elemento extraído a la caja.

```
eleccion <- sample(proveedor, n, replace = TRUE)
```

Para ver el resultado de cada elección en las ***n*** simulaciones, imprimimos el vector que almacena el muestreo.

```
print(eleccion)
```

Finalmente, utilizamos la función ***table*** para construir la tabla de frecuencias del arreglo que se construyó.

```
table(eleccion)/n
```

En resumen, nuestro *script* debería tener la siguiente apariencia:

```
proveedor <- c("Proveedor 1", "Proveedor 2")
eleccion <-NULL
n <- 20 # numero de simulaciones
eleccion <- sample(proveedor, n, replace = TRUE)
print(eleccion)
table(eleccion)/n
```

Mientras que, en la interfaz de *R*, debería verse así:

```
> proveedor <- c("Proveedor 1", "Proveedor 2")
> eleccion <-NULL
> n <- 20 # numero de simulaciones
> eleccion <- sample(proveedor, n, replace = TRUE)
> print(eleccion)
[1] "Proveedor 2" "Proveedor 2" "Proveedor 2" "Proveedor 2" "Proveedor 2"
[6] "Proveedor 1" "Proveedor 2" "Proveedor 1" "Proveedor 2" "Proveedor 2"
[11] "Proveedor 1" "Proveedor 2" "Proveedor 2" "Proveedor 1" "Proveedor 2"
[16] "Proveedor 2" "Proveedor 2" "Proveedor 1" "Proveedor 1" "Proveedor 1"
> table(eleccion)/n
eleccion
Proveedor 1 Proveedor 2
0.35      0.65
```

*Figura 1.* Script de elección de proveedor en *R*.



Vale la pena señalar que, en *R*, para cada distribución de probabilidad disponible, hay cuatro “comandos” que se ejecutan antecediendo un carácter al nombre de la función:

- ***d***: retorna la función de densidad de probabilidad.
- ***p***: retorna la función de densidad acumulada.
- ***q***: retorna la inversa acumulativa de la función de densidad (cuartiles).
- ***r***: retorna el conjunto de números generados aleatoriamente.



## Problema

En una unidad residencial viven 5 familias. Se asume que cada una de ellas está compuesta por 4 miembros y que al menos uno está contagiado con algún tipo de virus.

¿Si se tomara al azar un miembro de cada familia, cuál es la probabilidad de elegir las cuatro personas que se encuentran contagiadas?

## Solución

En este caso, podemos usar la función de distribución binomial ***binom***. Para ello, tenemos dos formas de hacerlo:

La primera es mediante la sumatoria de las densidades de probabilidad, a la que se le envía como argumento el valor del rango de ***x***, el tamaño ***size***, y el valor de la probabilidad ***prob***.

```
sum(dbinom(x = 0:10, size = 30, prob = 4/20))
```

La segunda forma implica utilizar la densidad acumulada, a la que se le envía el número de ocurrencias máximo, el tamaño y la probabilidad.

```
pbinom(10, size = 30, prob = 4/20)
```

Si se ejecutan en el mismo bloque, ambas instrucciones devuelven el mismo valor. Por otro lado, si se quiere conocer el conjunto de números aleatorios generados, se antecede la letra **r** al nombre de la función.

```
rbinom(10, size = 30, prob = 4/20)
```

El resultado de la aplicación de este *script* sería:

```
> sum(dbinom(x = 0:4, size = 20, prob = 4/20))
[1] 0.6296483
> pbinom(4, size = 20, prob = 4/20)
[1] 0.6296483
> rbinom(4, size = 20, prob = 4/20)
[1] 2 4 4 5
```

*Figura 2.* Script de probabilidad de elección de problemas contagiadas

Se puede concluir, a partir de los resultados, que la probabilidad de elegir las 4 personas contagiadas es del 62.96%.

## ● Binomial negativa

La variable aleatoria  $x$  se puede definir como el número de ensayos de *Bernoulli* necesarios para obtener exactamente  $k$  éxitos.

$$BN(x; k; p) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$$

En donde:

- $x$ : número de ensayos.
- $k$ : número de éxitos.
- $p$ : probabilidad de éxito.
- $1 - p$ : probabilidad de fracaso.



Un ensayo de *Bernoulli* se caracteriza porque solo se puede obtener uno de dos resultados posibles. Esta denominación se debe a su pionero, el matemático suizo Jakob Bernoulli.

## ● Geométrica

Esta distribución nos permite obtener la probabilidad de un número determinado de repeticiones ( $x$ ), antes de que se pueda tener éxito la primera vez. Una condición para esta probabilidad es tener un comportamiento decreciente a medida que aumenta el número de repeticiones.

Se diferencia de la binomial porque el número de repeticiones no se predefine, sino que es el objeto de la variable aleatoria que se calcula. Otra condición importante es que el conjunto de posibles valores de la variable es ilimitado.

$$G(x; p) = p(1-p)^{x-1}$$

En donde:

- $x$ : número de ensayos.
- $p$ : probabilidad de éxito.
- $q = 1 - p$ : probabilidad de fracaso.

## ● Poisson

Algunos autores definen a esta probabilidad como la más importante de las variables aleatorias discretas. Es útil para identificar el número de ocurrencias de un evento en un intervalo de tiempo o espacio, bajo presupuestos de aleatoriedad y con un conjunto específico de restricciones. Además, se especializa en eventos de muy baja probabilidad o “raros”.

$$p(x, \lambda) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

**En donde:**

- $\lambda$ : media (promedio) de éxitos por unidad (área, tiempo, producto).
- $e$ : 2,71828
- $x$ : número de éxitos esperados.



### Problema

En una línea de atención de un centro de contacto, se ha identificado que pueden confluir hasta 15 llamadas por minuto. Por tal razón, se requiere un *script* que permita calcular cuál es la probabilidad de que, en un minuto en particular, ingresen 20 llamadas o menos.

### Solución

Ya que usaremos la distribución de probabilidad de Poisson para resolver el problema, empecemos por almacenar en variables los valores que corresponden al número de llamadas promedio **promLlam**, y al número de llamadas que queremos evaluar **evaLlam**.

```
promLlam = 15
evaLlam = 20
```

Luego, invocamos la función **pois** antecediendo la letra **p**, para obtener la densidad acumulada.

```
ppois(evaLlam-1, lambda=promLlam)
```

```
> promLlam = 15
> evalLlam = 20
> ppois(evalLlam-1, lambda = promLlam)
[1] 0.8752188
```

**Figura 3.** Script de distribución de probabilidad Poisson en R

De acuerdo con el resultado obtenido, podemos concluir que la probabilidad de que en un minuto en particular ingresen 20 llamadas o menos es del 87.52% (0.8752188).

Si modificáramos la pregunta para averiguar la probabilidad de tener un ingreso de 20 llamadas **o más**, debemos recordar entonces que la misma se ubica en la cola derecha de la función de densidad de probabilidad. Esto se puede calcular de la siguiente manera:

```
ppois(evalLlam-1, lambda=promLlam, lower=FALSE)
```

```
> promLlam = 15
> evalLlam = 20
> ppois(evalLlam-1, lambda = promLlam, lower.tail = FALSE)
[1] 0.1247812
```

**Figura 4.** Script de distribución de probabilidad Poisson en R.

Concluimos así que la probabilidad de recibir 20 llamadas o más en un minuto, es de 12.48% (0.1247812), lo que es, por supuesto, una probabilidad menor.

## ● Hipergeométrica

Se caracteriza por estimar la cantidad posible de éxitos de una población a partir de una muestra. En otras palabras, y a manera de ejemplo, se utiliza para determinar el rechazo o aceptación de lotes a través del análisis de muestras más pequeñas sin tener que revisar la totalidad de los productos.

$$g(x; N; n; k) = \frac{\binom{k}{x} * \binom{N - k}{n - x}}{\binom{N}{n}}$$

**En donde:**

- $N$ : tamaño de la población
- $k$ : número de éxitos en la población
- $x$ : número de éxitos en la muestra
- $n$ : número de ensayos

Adicionalmente, el valor de  $x$  debe estar en el intervalo:

$$\max\{0, n - (N - k)\} \leq x \leq \min\{n, k\}.$$



### Problema

Una empresa conforma equipos de soporte con 10 técnicos cada uno. En cada equipo hay 4 funcionarios novatos. Además, se empleará un *software* para elegir aleatoriamente parejas de trabajadores en cada equipo. Por ello, requerimos contar con un *script* que permita calcular la probabilidad de que, en cada pareja elegida, uno de los trabajadores sea novato.

### Solución

Para tener control sobre las variables que forman parte del modelo algorítmico, debemos definirlas y almacenarlas previamente en el *script*.

En este caso, utilizaremos el símbolo numeral o *hashtag* (#) para crear comentarios dentro del código del *script*.

```
N = 10 # Número de opciones (10 técnicos por equipo)
x = 1 # Variable a determinar su probabilidad (1 trabajador novato)
k = 2 # La muestra (El software elige 1 pareja)
m = 4 # Casos exitosos (4 funcionarios novatos)
n = N-m # Casos no Exitosos
```

Luego, utilizamos la función ***hyper()***, antecedida por ***d*** para calcular la función de densidad de probabilidad.

```
dhyper(x = x,m = m, k = k, n = n)
```

```
> N = 10 # Número de opciones (10 técnicos por equipo)
> x = 1 # Variable a determinar su probabilidad (1 trabajador novato)
> k = 2 # La muestra (El software elige 1 pareja)
> m = 4 # Casos exitosos (4 funcionarios novatos)
> n = N-m # Casos no Exitosos
> dhyper(x = x,m = m, k = k, n = n)
[1] 0.5333333
```

**Figura 5.** Script de distribución de probabilidad hipergeométrica en R.

Concluimos así que la probabilidad de que uno de los trabajadores sea novato en cada pareja es del 53.3% (0.5333333).

**Si quisiéramos averiguar, en la elección aleatoria, la probabilidad de que ambos trabajadores sean novatos, el script podría quedar como sigue:**

```
N = 10 # Número de opciones (10 técnicos por equipo)
x = 2 # Variable a determinar su probabilidad (2 trabajadores novatos)
k = 2 # La muestra (El software elige 1 pareja)
m = 4 # Casos exitosos (4 funcionarios novatos)
n = N-m # Casos no Exitosos
dhyper(x = x,m = m, k = k, n = n)
```

```
> N = 10 # Número de opciones (10 técnicos por equipo)
> x = 2 # Variable a determinar su probabilidad (2 trabajadores novatos)
> k = 2 # La muestra (El software elige 1 pareja)
> m = 4 # Casos exitosos (4 funcionarios novatos)
> n = N-m # Casos no Exitosos
> dhyper(x = x,m = m, k = k, n = n)
[1] 0.1333333
```

**Figura 6.** Script de distribución de probabilidad hipergeométrica en R (cola derecha de la distribución)

En este caso, la probabilidad de que una pareja elegida esté compuesta por dos novatos es del 13.33% (0.1333333).

## 1.2. Distribuciones variables aleatorias continuas

La definición de variable aleatoria continua incluye cualquier valor entre 2 predefinidos (teóricamente).

Para ello, se consideran varios elementos:

- Los valores de la variable no se repiten.
- Se permite el uso de unidades no completas, es decir, decimales.
- El número de valores que puede tomar es limitado, dicho de otro modo, la variable no asume valores infinitos. Esto se logra con un número preciso de decimales.
- Por sus características, se asume como un área bajo la curva.

Las funciones más comunes son: uniforme, exponencial y normal. Aunque R incluye funciones para, por lo menos, 11 tipos de distribución.

### ● Uniforme

Es una función constante con respecto al eje vertical. Esta variable le da el mismo valor de probabilidad a todos los valores bajo esa curva, generando un área rectangular.

$$p(x) = \frac{1}{b-a}, x \in [a, b]$$

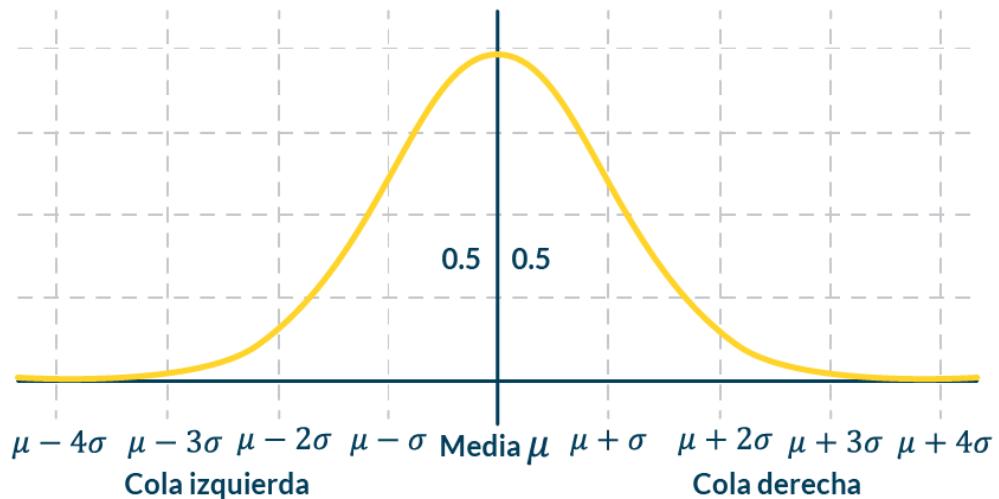
### ● Normal

Esta distribución se representa con una gráfica simétrica antes y después de la media aritmética, tipo campana. Cuando el valor de probabilidad se obtiene a través de la función acumulada, puede ser útil para ayudar a construir esta representación gráfica.

La distribución normal sirve para acercarse a diversas distribuciones de probabilidad discreta, como la distribución binomial y la distribución de *Poisson* (aproximaciones), con el fin de obtener la probabilidad de rechazo o aceptación.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

## Representación gráfica de una distribución normal



**Figura 7.** Representación gráfica de una distribución normal

### 1.3. Funciones de densidad

Una función de densidad de probabilidad describe el comportamiento probable de una población. Dicha función considera la manera en que una variable aleatoria continua ( $X$ ) se distribuye en una población y toma un valor cercano a  $x$ .

Se construye a partir de una expresión que totalice todos los eventos posibles que sumen la unidad. Y para cualquier otro caso, el resultado sería cero (0).

El modelo que describe la función de distribución acumulativa de  $x$  es:

$$F(x) = \int_{-\infty}^x f(\mu) d\mu$$

Y si  $f$  es continua en  $x$

$$f(x) = \frac{d}{dx} F(x)$$



#### Problema

El problema que desarrollaremos a continuación tiene como base un análisis de cinco variedades de aguacates, particularmente, la muestra de la variedad *Hass*. La investigación hace parte del trabajo de [Rojas y Flórez \(2020\)](#), quienes registraron los resultados de los cálculos promedio por árbol de aguacate, en una tabla que se reproduce a continuación, y en la cual solo se mostrarán los datos correspondientes a dichos promedios.

Planta	Promedio frutos por árbol
Planta uno	19.6
Planta dos	20.1
Planta tres	17.3
Planta cuatro	17.6
Planta cinco	23.1
Planta seis	17.8
Planta siete	17.2
Planta ocho	20.6
Planta nueve	19.3
Planta diez	18.3

A partir de esta información, debemos calcular y representar gráficamente la distribución de densidad de probabilidad.

## Solución

En primer lugar, cargaremos los datos en un arreglo para operar con ellos.

```
obs = c(19.6, 20.1, 17.3, 17.6, 23.1, 17.8, 17.2, 20.6, 19.3, 18.3)
```

Después, calculamos y almacenamos en una variable: tamaño, media, cuasidesviación y desviación estándar de la muestra.

```
n = length(obs)
mu = mean(obs)
desv = sd(obs, na.rm = FALSE)
```

A continuación, actualizamos la variable ***n***, otorgándole el valor de la nueva simulación de la muestra que obtenemos con un generador de valores aleatorios llamado ***rnorm***, con una distribución normal a partir de la media y la desviación estándar.

```
n = 5000
pob = rnorm(n = n, mean = mu, sd = desv)
muestra = data.frame(pob = pob)
```

Calculamos el tamaño del ancho de las barras del histograma (***bins***), utilizando la regla *Freedman–Diaconis*, con el siguiente modelo:

$$\text{ancho de la barra} = 2 \frac{IQR_{(x)}}{\sqrt[3]{n}}$$

En donde:

- $IQR_{(x)}$  : equivale al rango intercuartílico de los datos (número de observaciones en la muestra)

```
frec = 2*IQR(pob)*length(pob)**(-1/3)
```

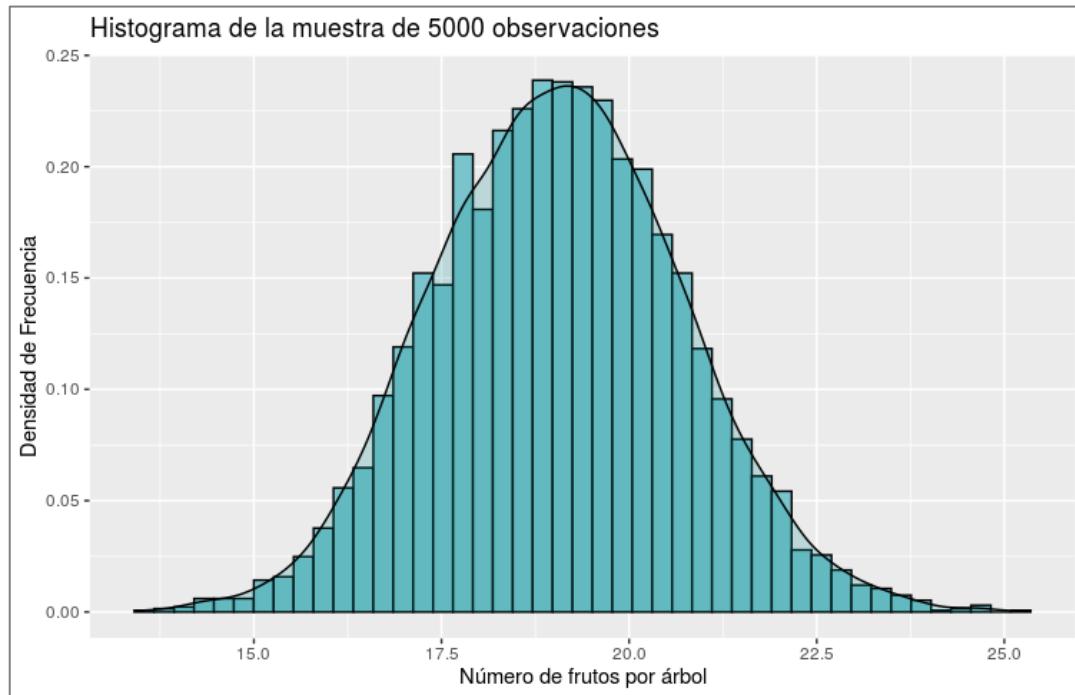
En seguida, graficamos el histograma con ***geom\_histogram***, con algunos parámetros dentro de los cuales destacamos el ***binwidth***, que utilizará el valor calculado en el paso anterior. Finalmente, definimos la curva de función de densidad con ***geom\_density***.

```
ggplot(muestra,aes(x = pob)) +  
  geom_histogram(aes(y=..density..,colour = "Density"),  
    binwidth = freq,  
    colour = "black",  
    fill = "cadetblue3") +  
  xlab("Número de frutos por árbol") + ylab("Densidad de Frecuencia") +  
  geom_density(alpha = .2, fill = "darkcyan") +  
  ggtitle("Histograma de la muestra de 5000 observaciones")
```

Luego de tener los componentes necesarios para la solución, podemos presentar el *script* completo para arrojar la representación que esperamos (si se ejecuta por pasos, funcionará igual).

```
obs = c(19.6, 20.1, 17.3, 17.6, 23.1, 17.8, 17.2, 20.6, 19.3, 18.3)  
n = length(obs)  
mu = mean(obs)  
cuasiDesv = sd(obs, na.rm = FALSE)  
desv = cuasiDesv * ((n-1)/n)  
#Actualizamos el valor de n  
n = 5000  
pob = rnorm(n = n, mean = mu, sd = desv)  
muestra = data.frame(pob = pob)  
freq = 2*IQR(pob)*length(pob)**(-1/3)  
ggplot(muestra,aes(x = pob)) +  
  geom_histogram(aes(y=..density..,colour = "Density"),  
    binwidth = freq,  
    colour = "black",  
    fill = "cadetblue3") +  
  xlab("Número de frutos por árbol") + ylab("Densidad de Frecuencia") +  
  geom_density(alpha = .2, fill = "darkcyan") +  
  ggtitle("Histograma de la muestra de 5000 observaciones")
```

El resultado de la representación se muestra en la figura 8.



**Figura 8.** Histograma y curva de densidad. Problema del aguacate



Para aprovechar la amplia oferta de colores de R en los gráficos, puedes consultar el siguiente enlace:  
<https://stat.columbia.edu/~tzhang/files/Rcolor.pdf>

## 2. Pruebas de bondad de ajuste

Las denominadas pruebas de bondad de ajuste (pruebas de hipótesis) permiten identificar si los datos de las observaciones en una muestra aleatoria se ajustan al modelo estadístico de alguna distribución de probabilidad con cierto nivel de significancia.

En general, se plantea una hipótesis que se debe verificar (denominada hipótesis nula), cuya validez está dada por el cumplimiento del estadístico de la prueba.



Si deseas ampliar el tema de pruebas de hipótesis, te recomendamos consultar el siguiente artículo:

<https://revistachilenadeanestesia.cl/inferencia-estadistica-pruebas-de-hipotesis/>

En cuanto a los generadores de variables aleatorias, es preciso indicar que, para obtener muestras de valores aleatorios, se requiere de números aleatorios como las “semillas”.

### 2.1. Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste permiten comparar los resultados y encontrar la discrepancia en una muestra con aquellos que se esperaría tener cuando la hipótesis nula sea verdadera.

La finalidad de estas pruebas es establecer si el conjunto de datos se ajusta a una determinada distribución, que puede ser especificada (hipótesis simple), o si pertenece a la clase paramétrica (hipótesis compuesta). Entre ellas, se destacan las pruebas de chi-cuadrado y la *de Kolmogorov-Smirnov*.

## ● Prueba de chi-cuadrado

Es una prueba no paramétrica empleada para distribuciones continuas y discretas, la cual permite encontrar la distribución de una serie de datos y comparar las frecuencias esperadas con las halladas. Para ello, utiliza la siguiente fórmula:

$$X^2 = \frac{\sum(x - \bar{x})^2}{\sigma^2}$$

Donde:

- $x^2$  : valor de la variable aleatoria.
- $\Sigma(x - \bar{x})^2$  : cuadrado de la sumatoria de las desviaciones.
- $\sigma^2$ : valor de la varianza.

En el cálculo de una prueba de distribución de este tipo, se asumen dos hipótesis contrarias, la hipótesis nula supone que los datos estimados y los obtenidos no tienen una diferencia significativa. Si se presenta una diferencia significativa, la hipótesis nula se rechaza.

**Hipótesis:**

- **$H_0$ :** la muestra se ajusta a una distribución teórica esperada.
- **$H_1$ :** la muestra no se ajusta a una distribución teórica esperada.

## Pasos para realizar la prueba de chi-cuadrado

Suponiendo que los datos son normales y que ya conocemos la media y desviación, procedemos a desarrollar los siguientes pasos:

1. Determinar el número de intervalos y el tamaño de cada uno de ellos.
2. Establecer la frecuencia observada por cada intervalo.
3. Hallar la frecuencia relativa esperada acumulada, teniendo en cuenta la función de distribución a utilizar, el límite superior, la media y la desviación.
4. Encontrar la frecuencia relativa esperada, restando la frecuencia relativa esperada acumulada del dato anterior de la frecuencia dentro de la columna.
5. Averiguar la frecuencia observada esperada (FOE), multiplicando la frecuencia relativa esperada por la suma de los datos de la frecuencia observada.
6. Calcular el estimador a partir de la fórmula de chi-cuadrado.
7. Sumar los datos calculados en el paso anterior.

8. Determinar los grados de libertad ( $V$ ) restando 1 al número de intervalos. Y, teniendo en cuenta la suma anterior, se buscan en la siguiente tabla:

**Tabla 1.**

### Distribución chi-cuadrado $\chi^2$ (niveles de confianza entre 0.001 y 0.5)

$P$  = probabilidad de encontrar un valor mayor o igual que el Chi cuadrado tabulado;

$V$  = grados de libertad

V/P	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415	2,7055	2,0722	1,6424	1,3233	1,0742	0,8735	0,7083	0,5707	0,4549
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052	3,7942	3,2189	2,7726	2,4079	2,0996	1,8326	1,5970	1,3863
3	16,2660	14,3202	12,8381	11,3449	9,3484	7,8147	6,2514	5,3170	4,6416	4,1083	3,6649	3,2831	2,9462	2,6430	2,3660
4	18,4662	16,4238	14,8602	13,2767	11,1433	9,4877	7,7794	6,7449	5,9886	5,3853	4,8784	4,4377	4,0446	3,6871	3,3567
5	20,5147	18,3854	16,7496	15,0863	12,8325	11,0705	9,2363	8,1152	7,2893	6,6257	6,0644	5,5731	5,1319	4,7278	4,3515
6	22,4575	20,2491	18,5475	16,8119	14,4494	12,5916	10,6446	9,4461	8,5581	7,8408	7,2311	6,6948	6,2108	5,7652	5,3481
7	24,3213	22,0402	20,2777	18,4753	16,0128	14,0671	12,0170	10,7479	9,8032	9,0371	8,3834	7,8061	7,2832	6,8000	6,3458
8	26,1239	23,7742	21,9549	20,0902	17,5345	15,5073	13,3616	12,0271	11,0301	10,2189	9,5245	8,9094	8,3505	7,8325	7,3441
9	27,8767	25,4625	23,5893	21,6660	19,0228	16,9190	14,6837	13,2880	12,2421	11,3887	10,6564	10,0060	9,4136	8,8632	8,3428
10	29,5879	27,1119	25,1881	23,2093	20,4832	18,3070	15,9872	14,5339	13,4420	12,5489	11,7807	11,0971	10,4732	9,8922	9,3418
11	31,2635	28,7291	26,7569	24,7250	21,9200	19,6752	17,2750	15,7671	14,6314	13,7007	12,8987	12,1836	11,5298	10,9199	10,3410
12	32,9092	30,3182	28,2997	26,2170	23,3367	21,0261	18,5493	16,9893	15,8120	14,8454	14,0111	13,2661	12,5838	11,9463	11,3403
13	34,5274	31,8830	29,8193	27,6882	24,7356	22,3620	19,8119	18,2020	16,9848	15,9839	15,1187	14,3451	13,6356	12,9717	12,3398
14	36,1239	33,4262	31,3194	29,1412	26,1189	23,6848	21,0641	19,4062	18,1508	17,1169	16,2221	15,4209	14,6853	13,9961	13,3393
15	37,6978	34,9494	32,8015	30,5780	27,4884	24,9958	22,3071	20,6030	19,3107	18,2451	17,3217	16,4940	15,7332	15,0197	14,3389
16	39,2518	36,4555	34,2671	31,9999	28,8453	26,2962	23,5418	21,7931	20,4651	19,3689	18,4179	17,5646	16,7795	16,0425	15,3385
17	40,7911	37,9462	35,7184	33,4087	30,1910	27,5871	24,7690	22,9770	21,6146	20,4887	19,5110	18,6330	17,8244	17,0646	16,3382
18	42,3119	39,4220	37,1564	34,8052	31,5264	28,8693	25,9894	24,1555	22,7595	21,6049	20,6014	19,6993	18,8679	18,0860	17,3379
19	43,8194	40,8847	38,5821	36,1908	32,8523	30,1435	27,2036	25,3289	23,9004	22,7178	21,6891	20,7638	19,9102	19,1069	18,3376
20	45,3142	42,3358	39,9969	37,5663	34,1696	31,4104	28,4120	26,4976	25,0375	23,8277	22,7745	21,8265	20,9514	20,1272	19,3374
21	46,7963	43,7749	41,4009	38,9322	35,4789	32,6706	29,6151	27,6620	26,1711	24,9348	23,8578	22,8876	21,9915	21,1470	20,3372
22	48,2676	45,2041	42,7957	40,2894	36,7807	33,9245	30,8133	28,8224	27,3015	26,0393	24,9390	23,9473	23,0307	22,1663	21,3370
23	49,7276	46,6231	44,1814	41,6383	38,0756	35,1725	32,0069	29,9792	28,4288	27,1413	26,0184	25,0055	24,0689	23,1852	22,3369
24	51,1790	48,0336	45,5584	42,9798	39,3641	36,4150	33,1962	31,1325	29,5533	28,2412	27,0960	26,0625	25,1064	24,2037	23,3367
25	52,6187	49,4351	46,9280	44,3140	40,6465	37,6525	34,3816	32,2825	30,6752	29,3388	28,1719	27,1183	26,1430	25,2218	24,3366
26	54,0511	50,8291	48,2898	45,6416	41,9231	38,8851	35,5632	33,4295	31,7946	30,4346	29,2463	28,1730	27,1789	26,2395	25,3365
27	55,4751	52,2152	49,6450	46,9628	43,1945	40,1133	36,7412	34,5736	32,9117	31,5284	30,3193	29,2266	28,2141	27,2569	26,3363
28	56,8918	53,5939	50,9936	48,2782	44,4608	41,3372	37,9159	35,7150	34,0266	32,6205	31,3909	30,2791	29,2486	28,2740	27,3362

29	58,3006	54,9662	52,3355	49,5878	45,7223	42,5569	39,0875	36,8538	35,1394	33,7109	32,4612	31,3308	30,2825	29,2908	28,3361
30	59,7022	56,3325	53,6719	50,8922	46,9792	43,7730	40,2560	37,9902	36,2502	34,7997	33,5302	32,3815	31,3159	30,3073	29,3360
31	61,0980	57,6921	55,0025	52,1914	48,2319	44,9853	41,4217	39,1244	37,3591	35,8871	34,5981	33,4314	32,3486	31,3235	30,3359
32	62,4873	59,0461	56,3280	53,4857	49,4804	46,1942	42,5847	40,2563	38,4663	36,9730	35,6649	34,4804	33,3809	32,3394	31,3359
33	63,8694	60,3953	57,6483	54,7754	50,7251	47,3999	43,7452	41,3861	39,5718	38,0575	36,7307	35,5287	34,4126	33,3551	32,3358
34	65,2471	61,7382	58,9637	56,0609	51,9660	48,6024	44,9032	42,5140	40,6756	39,1408	37,7954	36,5763	35,4438	34,3706	33,3357
35	66,6192	63,0760	60,2746	57,3420	53,2033	49,8018	46,0588	43,6399	41,7780	40,2228	38,8591	37,6231	36,4746	35,3858	34,3356
36	67,9850	64,4097	61,5811	58,6192	54,4373	50,9985	47,2122	44,7641	42,8788	41,3036	39,9220	38,6693	37,5049	36,4008	35,3356
37	69,3476	65,7384	62,8832	59,8926	55,6680	52,1923	48,3634	45,8864	43,9782	42,3833	40,9839	39,7148	38,5348	37,4156	36,3355
38	70,7039	67,0628	64,1812	61,1620	56,8955	53,3835	49,5126	47,0072	45,0763	43,4619	42,0450	40,7597	39,5643	38,4302	37,3354
39	72,0550	68,3830	65,4753	62,4281	58,1201	54,5722	50,6598	48,1263	46,1730	44,5395	43,1053	41,8040	40,5935	39,4446	38,3354
40	73,4029	69,6987	66,7660	63,6908	59,3417	55,7585	51,8050	49,2438	47,2685	45,6160	44,1649	42,8477	41,6222	40,4589	39,3353
45	80,0776	76,2229	73,1660	69,9569	65,4101	61,6562	57,5053	54,8105	52,7288	50,9849	49,4517	48,0584	46,7607	45,5274	44,3351
50	86,6603	82,6637	79,4898	76,1538	71,4202	67,5048	63,1671	60,3460	58,1638	56,3336	54,7228	53,2576	51,8916	50,5923	49,3349
55	93,1671	89,0344	85,7491	82,2920	77,3804	73,3115	68,7962	65,8550	63,5772	61,6650	59,9804	58,4469	57,0160	55,6539	54,3348
60	99,6078	95,3443	91,9518	88,3794	83,2977	79,0820	74,3970	71,3411	68,9721	66,9815	65,2265	63,6277	62,1348	60,7128	59,3347
70	112,3167	107,8079	104,2148	100,4251	95,0231	90,5313	85,5270	82,2553	79,7147	77,5766	75,6893	73,9677	72,3583	70,8236	69,3345
80	124,8389	120,1018	116,3209	112,3288	106,6285	101,8795	96,5782	93,1058	90,4053	88,1303	86,1197	84,2840	82,5663	80,9266	79,3343
90	137,2082	132,2554	128,2987	124,1162	118,1359	113,1452	107,5650	103,9040	101,0537	98,6499	96,5238	94,5809	92,7614	91,0234	89,3342
100	149,4488	144,2925	140,1697	135,8069	129,5613	124,3421	118,4980	114,6588	111,6667	109,1412	106,9058	104,8615	102,9459	101,1149	99,3341
120	173,6184	168,0814	163,6485	158,9500	152,2113	146,5673	140,2326	136,0620	132,8063	130,0546	127,6159	125,3833	123,2890	121,2850	119,3340
140	197,4498	191,5653	186,8465	181,8405	174,6478	168,6130	161,8270	157,3517	153,8537	150,8941	148,2686	145,8629	143,6043	141,4413	139,3339
160	221,0197	214,8081	209,8238	204,5300	196,9152	190,5164	183,3106	178,5517	174,8283	171,6752	168,8759	166,3092	163,8977	161,5868	159,3338
180	244,3723	237,8548	232,6198	227,0563	219,0442	212,3039	204,7036	199,6786	195,7434	192,4086	189,4462	186,7282	184,1732	181,7234	179,3338
200	267,5388	260,7350	255,2638	249,4452	241,0578	233,9942	226,0210	220,7441	216,6088	213,1022	209,9854	207,1244	204,4337	201,8526	199,3337
250	324,8306	317,3609	311,3460	304,9393	295,6885	287,8815	279,0504	273,1944	268,5987	264,6970	261,2253	258,0355	255,0327	252,1497	249,3337
300	381,4239	373,3509	366,8439	359,9064	349,8745	341,3951	331,7885	325,4090	320,3971	316,1383	312,3460	308,8589	305,5741	302,4182	299,3336
500	603,4458	593,3580	585,2060	576,4931	563,8514	553,1269	540,9303	532,8028	526,4014	520,9505	516,0874	511,6081	507,3816	503,3147	499,3335
600	712,7726	701,8322	692,9809	683,5155	669,7690	658,0936	644,8004	635,9329	628,8157	622,9876	617,6713	612,7718	608,1468	603,6942	599,3335

**Tabla 2.****Distribución chi-cuadrado  $\chi^2$   
(niveles de confianza entre 0.55 y 0.999)****P** = probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado;**V** = grados de libertad

V/P	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995	0,9975	0,999
<b>1</b>	0,3573	0,2750	0,2059	0,1485	0,1015	0,0642	0,0358	0,0158	0,0039	0,0010	0,0002	0,0000	0,0000	0,0000
<b>2</b>	1,1957	1,0217	0,8616	0,7133	0,5754	0,4463	0,3250	0,2107	0,1026	0,0506	0,0201	0,0100	0,0050	0,0020
<b>3</b>	2,1095	1,8692	1,6416	1,4237	1,2125	1,0052	0,7978	0,5844	0,3518	0,2158	0,1148	0,0717	0,0449	0,0243
<b>4</b>	3,0469	2,7528	2,4701	2,1947	1,9226	1,6488	1,3665	1,0636	0,7107	0,4844	0,2971	0,2070	0,1449	0,0908
<b>5</b>	3,9959	3,6555	3,3251	2,9999	2,6746	2,3425	1,9938	1,6103	1,1455	0,8312	0,5543	0,4118	0,3075	0,2102
<b>6</b>	4,9519	4,5702	4,1973	3,8276	3,4546	3,0701	2,6613	2,2041	1,6354	1,2373	0,8721	0,6757	0,5266	0,3810
<b>7</b>	5,9125	5,4932	5,0816	4,6713	4,2549	3,8223	3,3583	2,8331	2,1673	1,6899	1,2390	0,9893	0,7945	0,5985
<b>8</b>	6,8766	6,4226	5,9753	5,5274	5,0706	4,5936	4,0782	3,4895	2,7326	2,1797	1,6465	1,3444	1,1042	0,8571
<b>9</b>	7,8434	7,3570	6,8763	6,3933	5,8988	5,3801	4,8165	4,1682	3,3251	2,7004	2,0879	1,7349	1,4501	1,1519
<b>10</b>	8,8124	8,2955	7,7832	7,2672	6,7372	6,1791	5,5701	4,8652	3,9403	3,2470	2,5582	2,1558	1,8274	1,4787
<b>11</b>	9,7831	9,2373	8,6952	8,1479	7,5841	6,9887	6,3364	5,5778	4,5748	3,8157	3,0535	2,6032	2,2321	1,8338
<b>12</b>	10,7553	10,1820	9,6115	9,0343	8,4384	7,8073	7,1138	6,3038	5,2260	4,4038	3,5706	3,0738	2,6612	2,2141
<b>13</b>	11,7288	11,1291	10,5315	9,9257	9,2991	8,6339	7,9008	7,0415	5,8919	5,0087	4,1069	3,5650	3,1118	2,6172
<b>14</b>	12,7034	12,0785	11,4548	10,8215	10,1653	9,4673	8,6963	7,7895	6,5706	5,6287	4,6604	4,0747	3,5820	3,0407
<b>15</b>	13,6790	13,0298	12,3809	11,7212	11,0365	10,3070	9,4993	8,5468	7,2609	6,2621	5,2294	4,6009	4,0697	3,4825
<b>16</b>	14,6555	13,9827	13,3096	12,6243	11,9122	11,1521	10,3090	9,3122	7,9616	6,9077	5,8122	5,1422	4,5734	3,9417
<b>17</b>	15,6328	14,9373	14,2406	13,5307	12,7919	12,0023	11,1249	10,0852	8,6718	7,5642	6,4077	5,6973	5,0916	4,4162
<b>18</b>	16,6108	15,8932	15,1738	14,4399	13,6753	12,8570	11,9462	10,8649	9,3904	8,2307	7,0149	6,2648	5,6234	4,9048
<b>19</b>	17,5894	16,8504	16,1089	15,3517	14,5620	13,7158	12,7727	11,6509	10,1170	8,9065	7,6327	6,8439	6,1673	5,4067
<b>20</b>	18,5687	17,8088	17,0458	16,2659	15,4518	14,5784	13,6039	12,4426	10,8508	9,5908	8,2604	7,4338	6,7228	5,9210
<b>21</b>	19,5485	18,7683	17,9843	17,1823	16,3444	15,4446	14,4393	13,2396	11,5913	10,2829	8,8972	8,0336	7,2889	6,4467
<b>22</b>	20,5288	19,7288	18,9243	18,1007	17,2396	16,3140	15,2787	14,0415	12,3380	10,9823	9,5425	8,6427	7,8648	6,9829
<b>23</b>	21,5095	20,6902	19,8657	19,0211	18,1373	17,1865	16,1219	14,8480	13,0905	11,6885	10,1957	9,2604	8,4503	7,5291
<b>24</b>	22,4908	21,6525	20,8084	19,9432	19,0373	18,0618	16,9686	15,6587	13,8484	12,4011	10,8563	9,8862	9,0441	8,0847
<b>25</b>	23,4724	22,6156	21,7524	20,8670	19,9393	18,9397	17,8184	16,4734	14,6114	13,1197	11,5240	10,5196	9,6462	8,6494
<b>26</b>	24,4544	23,5794	22,6975	21,7924	20,8434	19,8202	18,6714	17,2919	15,3792	13,8439	12,1982	11,1602	10,2561	9,2222
<b>27</b>	25,4367	24,5440	23,6437	22,7192	21,7494	20,7030	19,5272	18,1139	16,1514	14,5734	12,8785	11,8077	10,8733	9,8029
<b>28</b>	26,4195	25,5092	24,5909	23,6475	22,6572	21,5880	20,3857	18,9392	16,9279	15,3079	13,5647	12,4613	11,4973	10,3907
<b>29</b>	27,4025	26,4751	25,5391	24,5770	23,5666	22,4751	21,2468	19,7677	17,7084	16,0471	14,2564	13,1211	12,1278	10,9861
<b>30</b>	28,3858	27,4416	26,4881	25,5078	24,4776	23,3641	22,1103	20,5992	18,4927	16,7908	14,9535	13,7867	12,7646	11,5876
<b>31</b>	29,3694	28,4087	27,4381	26,4397	25,3901	24,2551	22,9762	21,4336	19,2806	17,5387	15,6555	14,4577	13,4073	12,1961
<b>32</b>	30,3533	29,3763	28,3889	27,3728	26,3041	25,1478	23,8442	22,2706	20,0719	18,2908	16,3622	15,1340	14,0555	12,8104
<b>33</b>	31,3375	30,3444	29,3405	28,3069	27,2194	26,0422	24,7143	23,1102	20,8665	19,0467	17,0735	15,8152	14,7092	13,4312
<b>34</b>	32,3219	31,3130	30,2928	29,2421	28,1361	26,9383	25,5864	23,9522	21,6643	19,8062	17,7891	16,5013	15,3679	14,0568
<b>35</b>	33,3065	32,2821	31,2458	30,1782	29,0540	27,8359	26,4604	24,7966	22,4650	20,5694	18,5089	17,1917	16,0315	14,6881
<b>36</b>	34,2913	33,2517	32,1995	31,1152	29,9730	28,7350	27,3363	25,6433	23,2686	21,3359	19,2326	17,8868	16,7000	15,3243
<b>37</b>	35,2764	34,2216	33,1539	32,0532	30,8933	29,6355	28,2138	26,4921	24,0749	22,1056	19,9603	18,5859	17,3730	15,9652
<b>38</b>	36,2617	35,1920	34,1089	32,9919	31,8146	30,5373	29,0931	27,3430	24,8839	22,8785	20,6914	19,2888	18,0501	16,6109
<b>39</b>	37,2472	36,1628	35,0645	33,9315	32,7369	31,4405	29,9739	28,1958	25,6954	23,6543	21,4261	19,9958	18,7318	17,2612
<b>40</b>	38,2328	37,1340	36,0207	34,8719	33,6603	32,3449	30,8563	29,0505	26,5093	24,4331	22,1642	20,7066	19,4171	17,9166
<b>45</b>	43,1638	41,9950	40,8095	39,5847	38,2910	36,8844	35,2895	33,3504	30,6123	28,3662	25,9012	24,3110	22,8994	21,2509
<b>50</b>	48,0986	46,8638	45,6100	44,3133	42,9421	41,4492	39,7539	37,6886	34,7642	32,3574	29,7067	27,9908	26,4636	24,6736

55	53,0367	51,7391	50,4204	49,0554	47,6105	46,0356	44,2448	42,0596	38,9581	36,3981	33,5705	31,7349	30,0974	28,1731
60	57,9775	56,6200	55,2394	53,8091	52,2938	50,6406	48,7587	46,4589	43,1880	40,4817	37,4848	35,5344	33,7909	31,7381
70	67,8664	66,3961	64,8990	63,3460	61,6983	59,8978	57,8443	55,3289	51,7393	48,7575	45,4417	43,2753	41,3323	39,0358
80	77,7631	76,1879	74,5825	72,9153	71,1445	69,2070	66,9938	64,2778	60,3915	57,1532	53,5400	51,1719	49,0430	46,5197
90	87,6661	85,9925	84,2854	82,5111	80,6247	78,5584	76,1954	73,2911	69,1260	65,6466	61,7540	59,1963	56,8918	54,1559
100	97,5744	95,8078	94,0046	92,1290	90,1332	87,9453	85,4406	82,3581	77,9294	74,2219	70,0650	67,3275	64,8571	61,9182
120	117,4041	115,4646	113,4825	111,4186	109,2197	106,8056	104,0374	100,6236	95,7046	91,5726	86,9233	83,8517	81,0726	77,7555
140	137,2476	135,1491	133,0028	130,7657	128,3800	125,7580	122,7476	119,0293	113,6594	109,1368	104,0343	100,6547	97,5908	93,9253
160	157,1019	154,8555	152,5564	150,1583	147,5988	144,7834	141,5475	137,5457	131,7560	126,8700	121,3457	117,6791	114,3496	110,3592
180	176,9652	174,5799	172,1373	169,5879	166,8653	163,8682	160,4206	156,1526	149,9687	144,7413	138,8205	134,8843	131,3050	127,0114
200	196,8359	194,3193	191,7409	189,0486	186,1717	183,0028	179,3550	174,8353	168,2785	162,7280	156,4321	152,2408	148,4262	143,8420
250	246,5387	243,7202	240,8297	237,8085	234,5768	231,0128	226,9048	221,8059	214,3915	208,0978	200,9387	196,1604	191,8020	186,5537
300	296,2700	293,1786	290,0062	286,6878	283,1353	279,2143	274,6901	269,0679	260,8781	253,9122	245,9727	240,6631	235,8126	229,9620
500	495,3734	491,3709	487,2569	482,9462	478,3231	473,2099	467,2962	459,9261	449,1467	439,9360	429,3874	422,3034	415,8081	407,9458
600	594,9938	590,6057	586,0930	581,3623	576,2859	570,6681	564,1661	556,0560	544,1801	534,0185	522,3654	514,5285	507,3385	498,6219

9. Si el estimador  $X^2$  es menor o igual al valor correspondiente en la tabla, entonces, se acepta  $H_0$ ; en caso contrario, se rechaza.

## ● Prueba de Kolmogorov-Smirnov

Es una prueba no paramétrica la cual se emplea solo para distribuciones continuas. Esta tiene como objetivo encontrar el tipo de distribución de una serie de datos, se considera más eficiente que la prueba de chi-cuadrado debido a que trabaja con la distribución de probabilidad acumulada: la distribución acumulada de los datos observados y la distribución acumulada teórica correspondiente al modelo elegido.

### Ventajas:

- Es una prueba de fácil aplicación, ya que no requiere que los datos sean agrupados.
- Permite juzgar la discrepancia entre la distribución de frecuencias observada y la distribución de frecuencias esperada, puesto que la distribución de probabilidad ( $D_n$ ) depende del tamaño de muestra  $n$ , pero es independiente de la distribución de frecuencia esperada.

### Pasos para realizar la prueba de Kolmogorov-Smirnov

Suponiendo que los datos son normales y que ya conocemos la media y desviación, procedemos a desarrollar los siguientes pasos:

1. Identificar la muestra de la población a utilizar.
2. Plantear la hipótesis para la muestra:
  - $H_0$ : hipótesis nula.
  - $H_1$ : hipótesis alternativa.
3. Calcular la frecuencia observada de cada uno de los intervalos. Luego, se suman todas las frecuencias observadas.
4. Determinar la frecuencia observada relativa (frecuencia observada de cada intervalo/ sumatoria total de la frecuencia observada).
5. Averiguar la frecuencia observada relativa acumulada (FORA) y la frecuencia esperada relativa acumulada (FERA).
6. Se calcula el *estadístico de prueba* (D) de cada intervalo con la siguiente fórmula:

$$D = \text{abs} (\text{FOR Acum} - \text{FER Acum})$$

**7.** Buscar el valor esperado en la siguiente tabla de acuerdo con el tamaño de la muestra y un alfa ( $\alpha$ ):

- $n < 40$ : se realiza el procedimiento normal.
- $n > 40$ : se aplica la fórmula que se expone en la tabla.

**Tabla 3. Valores prueba de Kolmogorov-Smirnov**

Grados de libertad $n$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,1$	$\alpha = 0,15$	$\alpha = 0,2$
1	0,995	0,975	0,95	0,925	0,9
2	0,929	0,842	0,776	0,726	0,684
3	0,828	0,708	0,642	0,597	0,265
4	0,733	0,624	0,564	0,525	0,494
5	0,669	0,565	0,51	0,474	0,446
6	0,618	0,521	0,47	0,436	0,41
7	0,577	0,486	0,438	0,405	0,381
8	0,543	0,457	0,411	0,381	0,358
9	0,514	0,432	0,388	0,36	0,339
10	0,49	0,401	0,368	0,342	0,322
11	0,468	0,391	0,352	0,326	0,306
12	0,45	0,375	0,338	0,313	0,295
13	0,433	0,361	0,325	0,302	0,284
14	0,418	0,349	0,314	0,292	0,274
15	0,404	0,338	0,304	0,283	0,266
16	0,392	0,328	0,295	0,274	0,258
17	0,382	0,318	0,286	0,266	0,25
18	0,371	0,309	0,278	0,259	0,244
19	0,363	0,301	0,272	0,252	0,237
20	0,356	0,295	0,264	0,265	0,231
25	0,32	0,27	0,24	0,22	0,21
30	0,29	0,24	0,22	0,2	0,19
35	0,27	0,23	0,21	0,19	0,18
40	0,25	0,21	0,19	0,18	0,17
45	0,24	0,2	0,18	0,17	0,16
50	0,23	0,19	0,17	0,16	0,15
Recurrencia para $n$ mayor	$\frac{1.63}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.07}{\sqrt{n}}$

# ¡Felicitaciones!

Has completado con éxito el estudio de la actividad de aprendizaje de la Unidad 3, en la cual abordamos temas relacionados a los datos, la predicción y la toma de decisiones.

Aprovecha los conocimientos que has adquirido para desarrollar habilidades que te permitan interpretar y analizar datos para descifrar contextos, encontrar soluciones o proponer ideas innovadoras en todos los ámbitos de tu vida.

Ahora puedes realizar la segunda evidencia de aprendizaje del curso para poner a prueba lo que has aprendido. Lee muy bien el problema a resolver, identifica sus condiciones y restricciones, y define cuál es la mejor forma de solucionarlo.

## Evidencia de aprendizaje (EA) de la Unidad 3:

### Problema de satisfacción de clientes en las tiendas virtuales reportadas en el portal de datos abiertos del Gobierno.

<b>Nombre de la evidencia de aprendizaje</b>	Problema de producción de repuestos de lámparas direccionales para vehículos.
<b>Objetivo de la evidencia de aprendizaje</b>	A partir de los datos de producción de la planta, calcular y representar gráficamente la distribución de probabilidad mediante procesos aleatorios discretos.
<b>Contenidos</b>	Los datos que se utilizan para el cálculo del modelo de distribución de probabilidad se aprovisionan en el planteamiento del problema.
<b>Descripción de lo que debe hacer el estudiante</b>	El estudiante debe construir la estructura que permita almacenar los datos señalados en el planteamiento del problema. Posteriormente, con ayuda de <i>R Studio</i> , diseñar el <i>script</i> que permita resolver las cuestiones planteadas en el problema en términos de la distribución de probabilidad y la representación gráfica.
<b>Especifique lo que debe entregar el estudiante</b>	El estudiante debe diseñar un <i>script</i> en <i>R</i> , utilizando la plataforma <i>online</i> de <i>R Studio</i> . Después, debe descargar el <i>script</i> y subirlo como evidencia a la plataforma.



## Referencias Bibliográficas

- Banks, J. (2010). Discrete event system simulation. En Bigdoli, H. (Ed.). *Encyclopedia of Information Systems* (pp. 663-671). Elsevier. <https://doi.org/10.1016/b0-12-227240-4/00045-9>
- Conexión Esan. (2016). *Conoce las principales distribuciones de probabilidad*. <https://www.esan.edu.pe/apuntes-empresariales/2016/10/conoce-las-principales-distribuciones-de-probabilidad/>
- El Economist. (2019). *Discrepancia Estadística*. <https://www.eleconomist.com/2013/01/discrepancia-estadistica.html>
- Glosarios especializados. (2019). *Glosario de términos estadísticos*. <https://glosarios.servidor-alicante.com/terminos-estadistica/>
- Histograma. (2021). En *Wikipedia*. <https://es.wikipedia.org/w/index.php?title=Histograma&oldid=135404426>
- Kelmansky, D. (2006). *Histogramas de Áreas - Curvas de densidad*. Universidad de Buenos Aires. [http://www.dm.uba.ar/materias/analisis\\_de\\_datos/2006/1/teoricas/Teor2b.pdf](http://www.dm.uba.ar/materias/analisis_de_datos/2006/1/teoricas/Teor2b.pdf)
- Khan Academy. (2017). *Probabilidad: conceptos básicos*. <https://es.khanacademy.org/math/statistics-probability/probability-library/basic-theoretical-probability/a/probability-the-basics>
- Levin, R. & Rubin, D. (2004). *Estadística para administración y economía*. Pearson Educación.
- López, J. (2019). *Estadística paramétrica*. Economipedia. <https://economipedia.com/definiciones/estadistica-parametrica.html>
- Marco, J. (2017). *Frecuencia relativa*. Economipedia. <https://economipedia.com/definiciones/frecuencia-relativa.html>

- Massey, F. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46, 68-78. doi:10.2307/2280095
- Mendoza, H. (2017). *Probabilidad y Estadística*. Universidad Nacional.  
[http://red.unal.edu.co/cursos/ciencias/2001065/html/un1/cont\\_116\\_16.html](http://red.unal.edu.co/cursos/ciencias/2001065/html/un1/cont_116_16.html)
- Minitab. (2019, 18 de abril). *¿Qué son los Grados de Libertad en Estadística?*  
<https://blog.minitab.com/es/que-son-los-grados-de-libertad-en-estadistica>
- Montoro, D. (2007). *Métodos Estadísticos en la Ingeniería*. Universidad de Jaén  
<http://www4.ujaen.es/~dmontoro/temarioMetodos.htm>
- Proyecto de Cooperación UE – CAN en Materia de Estadísticas. (2007). *Glosario de Términos Estadísticos*. Comunidad Andina.  
[http://bit.ly/term\\_estadisticos\\_comandina](http://bit.ly/term_estadisticos_comandina)
- QuestionPro. (2019). *¿Cuáles son los tipos de datos estadísticos?*  
<https://www.questionpro.com/blog/es/tipos-de-datos-estadisticos/>
- Software DELSOL. (2019). *Hipótesis nula.*  
<https://www.sdelsol.com/glosario/hipotesis-nula/>
- Universitat de Valencia. (2014). *Medida de Asimetría.*  
<https://www.uv.es/ceaces/base/descriptiva/simetri.htm>
- Valdez, I. (2018). *Pruebas de bondad de ajuste*. Universidad Nacional Autónoma de México. <http://dcb.fi-c.unam.mx/profesores/irene/Notas/PruebaBondad.pdf>
- Westreicher, G. (2020). *Predictión (estadística)*. Economipedia.  
<https://economipedia.com/definiciones/predictcion-estadistica.html>
- Wolters Kluwer. (2020). *Función de densidad de probabilidad.*  
[https://guiasjuridicas.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAA\\_AAAAEAMtMSbF1jTAAASNjS3MTtbLUouLM\\_DxblwMDS0NDA1OQQGZapUt-ckhlQaptWmJOcSoAdl8gBjAAAAA=WKE](https://guiasjuridicas.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAA_AAAAEAMtMSbF1jTAAASNjS3MTtbLUouLM_DxblwMDS0NDA1OQQGZapUt-ckhlQaptWmJOcSoAdl8gBjAAAAA=WKE)



## Lecturas y material complementario

- **Diseño Gráfico:** Steven Miranda Cardona. Dirección de Tecnología. IUD
- **Video**

Autor: UOC - Universitat Oberta de Catalunya  
Título: R Studio Cloud en la docencia online en asignaturas de metodología y estadística  
URL: <https://youtu.be/Rw86uqi3o5E?t=900> (Minuto 15:00)
- **Video**

Autor: Juan Correa  
Título: Tutorial para usar R Studio Cloud  
URL: <https://youtu.be/zr63MPKI2kw?t=115> (Minuto 1:55)
- **Video**

Autor: Mate316  
Título: Distribución hipergeométrica con R  
URL: <https://youtu.be/tPnPZdLXy88>
- **Lectura**

Autor: Stat Columbia  
Título: Color in R  
URL: <https://stat.columbia.edu/~tzheng/files/Rcolor.pdf>
- **Lectura**

Autores: Yilson Rojas y Yamid Flórez  
Título: Análisis de cinco variedades de aguacate  
URL:  
<https://repository.unad.edu.co/bitstream/handle/10596/33533/ysrojasp.pdf?sequence=1&isAllowed=y>
- **Lectura**

Autor: Carlos Gil Bellosta  
Título: R para profesionales de los datos: una introducción  
URL: [https://www.datanalytics.com/libro\\_r/index.html](https://www.datanalytics.com/libro_r/index.html)

- **Lectura**

Autores: Javier Gorgas y Nicolás Cardiel

Título: Introducción a R

URL: [https://www.ucm.es/data/cont/docs/339-2016-09-29-Introduccion%20a%20R\\_v1617.pdf](https://www.ucm.es/data/cont/docs/339-2016-09-29-Introduccion%20a%20R_v1617.pdf)

Esta licencia permite a otros distribuir, remezclar, retocar, y crear a partir de esta obra de manera no comercial y, a pesar que sus nuevas obras deben siempre mencionar a la IU Digital y mantenerse sin fines comerciales, no están obligados a licenciar obras derivadas bajo las mismas condiciones.

