Universidad Manuela Beltrán

Diplomado en Bioestadística
Modulo III: Introducción a R

Elaborado por

Luz Angélica Walteros Rodríguez

Magíster en Ciencias (Física)

Bogotá, Colombia
Febrero 2024.

Índice general

Capítulo 1

Probabilidad

La **probabilidad** es una rama de las matemáticas que se ocupa del estudio de la incertidumbre. Es la medida cuantitativa de la posibilidad de que ocurra un evento dentro de un espacio de posibles resultados. En términos sencillos, la probabilidad nos ayuda a predecir cuán probable es que ocurra un determinado suceso basado en la información disponible.

1.1. Conceptos básicos

- **Espacio muestral**: Es el conjunto de todos los posibles resultados de un experimento aleatorio. Por ejemplo, al lanzar un dado, el espacio muestral S es $S = \{1, 2, 3, 4, 5, 6\}$, ya que estos son los posibles resultados.

- **Evento**: Es cualquier subconjunto del espacio muestral. Un evento puede ser simple o compuesto:

  - **Evento simple**: Un solo resultado, como $A = \{1\}$, que representa la ocurrencia del número 1 al lanzar el dado.
  - **Evento compuesto**: Un conjunto de resultados, como $B = \{1, 3, 5\}$, que representa obtener un número impar al lanzar el dado.

1.2. Aplicaciones de la Probabilidad

La probabilidad se aplica en numerosos campos, entre los que se incluyen:

- **Estadística**: Para hacer inferencias sobre poblaciones a partir de muestras.

- **Ciencias**: En física, biología, economía y otros campos para modelar fenómenos aleatorios y realizar predicciones.

- **Juegos de azar**: Para calcular las probabilidades de ganar en juegos como el póker, ruleta, dados, etc.
- **Ingeniería**: En el análisis de fiabilidad de sistemas y en el diseño de sistemas con seguridad probabilística.
- **Machine Learning**: Para construir modelos probabilísticos y realizar inferencias a partir de datos observados.

1.3. Propiedades de la Probabilidad

La probabilidad sigue varias reglas fundamentales:

1.3.1. Regla de Adición

Si los eventos A y B son mutuamente excluyentes (es decir, no pueden ocurrir al mismo tiempo), la probabilidad de que ocurra A o B es:

$$P(A \cup B) = P(A) + P(B)$$

Si los eventos no son mutuamente excluyentes, entonces:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Donde $A \cup B$ es la unión de A y B (es decir, el evento de que ocurra A o B), y $A \cap B$ es la intersección de A y B (es decir, el evento de que ocurra tanto A como B).

1.3.2. Regla de Multiplicación

La probabilidad de que ocurran dos eventos A y B simultáneamente (es decir, que ambos eventos ocurran) es:

$$P(A \cap B) = P(A) \cdot P(B|A)$$

Donde $P(B|A)$ es la probabilidad de que ocurra B dado que ya ha ocurrido A (probabilidad condicional).

1.3.3. Complemento

La probabilidad de que no ocurra un evento A es:

$$P(A^c) = 1 - P(A)$$

Donde A^c representa el evento complementario de A, es decir, el evento de que no ocurra A.

1.4. Tipos de Probabilidad

Existen varios enfoques y métodos para definir y calcular la probabilidad, dependiendo de la naturaleza del experimento o el sistema que se está analizando. Los tipos más comunes son:

1.4.1. Probabilidad Clásica

Se basa en el cálculo de probabilidades en experimentos que tienen resultados igualmente probables. Es útil cuando se conocen todos los posibles resultados del espacio muestral. Ejemplo: lanzar un dado.

$$P(A) = \frac{\text{Número de resultados favorables}}{\text{Número total de resultados posibles}}$$

1.4.2. Probabilidad Frecuentista

Se basa en la observación de experimentos repetidos. La probabilidad de un evento se define como el límite de su frecuencia relativa a medida que el número de experimentos tiende al infinito. Ejemplo: realizar un experimento de lanzar una moneda muchas veces y observar la frecuencia de caras.

1.4.3. Probabilidad Subjetiva

Es la estimación de la probabilidad de un evento basado en el juicio personal, la experiencia o el conocimiento de una persona. Es común en situaciones donde no hay suficiente información para aplicar enfoques clásicos o frecuentistas.

1.5. Ejemplo Práctico: Lanzamiento de un Dado

Supongamos que tenemos un dado con 6 caras numeradas del 1 al 6. El espacio muestral es $S = \{1, 2, 3, 4, 5, 6\}$, y cada cara tiene la misma probabilidad de salir, es decir, $P(\text{cara}) = \frac{1}{6}$.

Si queremos calcular la probabilidad de que salga un número par, el evento favorable es $A = \{2, 4, 6\}$, así que:

$$P(A) = \frac{\text{Número de resultados favorables}}{\text{Número total de resultados}} = \frac{3}{6} = 0{,}5$$

Es decir, la probabilidad de obtener un número par es $P(A) = 0{,}5$.

1.6. Probabilidad condicional

La **probabilidad condicional** es un concepto fundamental en la teoría de la probabilidad que describe la probabilidad de que ocurra un evento dado que otro evento ya ha ocurrido. Se denota como $P(A \mid B)$, que se lee como "la probabilidad de A dado B". Esto significa que estamos interesados en la probabilidad de que ocurra el evento A, bajo la condición de que sabemos que B ha sucedido.

1.6.1. Fórmula de la Probabilidad Condicional

La probabilidad condicional se define formalmente de la siguiente manera:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{si } P(B) > 0$$

donde:

- $P(A \mid B)$ es la probabilidad de que ocurra A, dado que B ha ocurrido.
- $P(A \cap B)$ es la probabilidad de que ocurran ambos eventos A y B simultáneamente (la intersección de A y B).
- $P(B)$ es la probabilidad de que ocurra el evento B.

1.6.2. Propiedades de la Probabilidad Condicional

Multiplicación de Probabilidades

Usando la definición de probabilidad condicional, podemos reescribir la probabilidad de la intersección de dos eventos A y B como:

$$P(A \cap B) = P(A \mid B)P(B)$$

1.6.3. Ejemplo

Supongamos que tenemos una baraja de cartas. Si sacamos una carta al azar, la probabilidad de que sea un *as* es:

$$P(A) = \frac{4}{52} = \frac{1}{13}$$

Ahora, si sabemos que la carta es un *corazón* (B), entonces la probabilidad condicional de que sea un *as* dado que es un *corazón* es:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{52}}{\frac{13}{52}} = \frac{1}{13}$$

Esto significa que, dado que el evento B ha ocurrido (la carta es un corazón), la probabilidad de que sea el as de corazones es $\frac{1}{13}$.

1.6.4. Probabilidad condicional en RStudio

A continuación se muestran ejemplos simples utilizando RStudio y conjuntos de datos simulados.

Ejemplo 1: Probabilidad Condicional con un Dado

Supongamos que tenemos un dado y estamos interesados en los siguientes eventos:

- A: El evento de obtener un número mayor o igual a 4.
- B: El evento de obtener un número impar.

Queremos calcular la probabilidad condicional $P(A \mid B)$, es decir, la probabilidad de obtener un número mayor o igual a 4 dado que el número obtenido es impar.

Código en R

El siguiente código en R calcula las probabilidades necesarias para encontrar $P(A \mid B)$:

```r
# Definir el espacio muestral (resultados posibles de tirar
#un dado)
dado <- 1:6

# Evento A: Obtenemos un número mayor o igual a 4
evento_A <- dado[dado >= 4]

# Evento B: Obtenemos un número impar
evento_B <- dado[dado %% 2 != 0]

# Intersección de A y B: Números que son tanto mayores o
#iguales a 4 como impares
interseccion_AB <- intersect(evento_A, evento_B)

# Probabilidad de A (P(A)) - Probabilidad de que salga un
#número mayor o igual a 4
P_A <- length(evento_A) / length(dado)

# Probabilidad de B (P(B)) - Probabilidad de que salga un
#número impar
```

```r
P_B <- length(evento_B) / length(dado)

# Probabilidad de la intersección A y B- Probabilidad de
#que salga un número mayor o igual a 4 e
#impar
P_A_inter_B <- length(interseccion_AB) / length(dado)

# Probabilidad condicional
P_A_given_B <- P_A_inter_B / P_B

# Imprimir resultados
cat("Probabilidad de A (P(A)): ", P_A, "\n")
cat("Probabilidad de B (P(B)): ", P_B, "\n")
cat("Probabilidad de la intersección A y B: ",
P_A_inter_B, "\n")
cat("Probabilidad condicional P(A | B): ", P_A_given_B,
"\n")
```

Ejemplo 2: Probabilidad Condicional con Datos Reales

En este ejemplo, trabajamos con un conjunto de datos simulado sobre la presencia de una enfermedad y si una persona fuma o no. Calcularemos la probabilidad condicional de que una persona tenga la enfermedad dado que fuma.

Código en R

El siguiente código genera un conjunto de datos simulado y calcula la probabilidad condicional $P(\text{Enfermedad} \mid \text{Fuma})$:

```r
#Datos de ejemplo: Un conjunto de datos sobre la presencia
#de enfermedad y fumar
set.seed(123)

# Generamos un conjunto de datos con 1000 personas
n <- 1000
enfermedad <- sample(c(0, 1), n, replace = TRUE)  # 0 = No
#tiene enfermedad, 1 = Tiene enfermedad
fuma <- sample(c(0, 1), n, replace = TRUE)  # 0 = No fuma,
#1 = Fuma

# Evento A: Personas que tienen la enfermedad
evento_A <- enfermedad == 1
```

```r

# Evento B: Personas que fuman
evento_B <- fuma == 1

#Probabilidad de A(P(A)):Personas que tienen la enfermedad
P_A <- mean(evento_A)

# Probabilidad de B (P(B)): Personas que fuman
P_B <- mean(evento_B)

#Intersección de A y B : Personas que tienen la enfermedad
#y fuman
P_A_inter_B <- mean(evento_A & evento_B)

# Probabilidad condicional P(A | B): Personas que tienen
#la enfermedad dado que fuman
P_A_given_B <- P_A_inter_B / P_B

# Imprimir resultados
cat("Probabilidad de A (P(A)): ", P_A, "\n")
cat("Probabilidad de B (P(B)): ", P_B, "\n")
cat("Probabilidad de la intersección A y B:", P_A_inter_B,
"\n")
cat("Probabilidad condicional P(A | B): ", P_A_given_B,
"\n")
```

Resultados Esperados

Al ejecutar los ejemplos en RStudio, el resultado esperado para el primer ejemplo podría ser algo como:

$$P(A \mid B) \approx 0{,}3333$$

En el segundo ejemplo, los resultados podrían ser similares a:

$$P(A \mid B) \approx 0{,}492$$

Esto significa que, dado que sabemos que una persona fuma, la probabilidad de que también tenga la enfermedad es de aproximadamente 49,2 %.

1.7. Variables aleatorias

Una *variable aleatoria* es una función que asigna un valor numérico a cada resultado de un experimento aleatorio. Su valor es incierto hasta que se realice el experimento, y depende de los resultados aleatorios del mismo.

Formalmente, una variable aleatoria X es una función $X : \Omega \to \mathbb{R}$, donde Ω es el espacio muestral (el conjunto de todos los resultados posibles) y $\mathbb{R}$ es el conjunto de los números reales.

1.7.1. Tipos de Variables Aleatorias

Las variables aleatorias se dividen en dos grandes categorías:

1. Variables Aleatorias Discretas

Una variable aleatoria es *discreta* si puede tomar un número finito o numerablemente infinito de valores posibles.

- Ejemplos típicos:
  - El número de caras que aparece al lanzar un dado.
  - El número de llamadas que recibe una central telefónica en una hora.
- Su función de probabilidad se describe mediante la *función de masa de probabilidad (PMF)*, que asigna una probabilidad a cada valor específico que puede tomar la variable.

2. Variables Aleatorias Continuas

Una variable aleatoria es *continua* si puede tomar cualquier valor en un intervalo o conjunto de intervalos de números reales.

- Ejemplos típicos:
  - La altura de una persona.
  - El tiempo que tarda un coche en recorrer una distancia determinada.
- Su función de distribución se describe mediante la *función de densidad de probabilidad (PDF)*, la cual no da directamente la probabilidad de que la variable tome un valor específico, sino la probabilidad de que la variable caiga dentro de un rango de valores.

1.7.2. Funciones Asociadas a las Variables Aleatorias

Para cada tipo de variable aleatoria, existen ciertas funciones fundamentales:

1. Función de Masa de Probabilidad (PMF)

Se aplica a variables aleatorias discretas. La PMF, $P(X = x)$, da la probabilidad de que la variable aleatoria X tome el valor x.

Ejemplo: Si lanzamos un dado, la PMF de una variable aleatoria X que represente el número obtenido es:

$$P(X = x) = \frac{1}{6}, \quad x \in \{1, 2, 3, 4, 5, 6\}$$

2. Función de Densidad de Probabilidad (PDF)

Se aplica a variables aleatorias continuas. La PDF, $f(x)$, describe la densidad de probabilidad de que la variable aleatoria tome un valor dentro de un intervalo determinado. La probabilidad de que la variable tome un valor en el intervalo $[a, b]$ se calcula mediante la integral de la PDF en ese intervalo:

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

3. Función de Distribución Acumulada (CDF)

Tanto para variables aleatorias discretas como continuas, la función de distribución acumulada, $F(x)$, es la probabilidad de que la variable aleatoria X tome un valor menor o igual que x:

$$F(x) = P(X \leq x)$$

1.7.3. Propiedades de las Variables Aleatorias

Las variables aleatorias tienen varias propiedades importantes que se utilizan para describir sus distribuciones y comportamientos.

1. Esperanza (Valor Esperado o Media)

Es el valor promedio ponderado de todas las posibles salidas de una variable aleatoria, tomando en cuenta sus probabilidades. Se denota como $\mathbb{E}[X]$.

- Para una variable aleatoria discreta:

$$\mathbb{E}[X] = \sum_x x \cdot P(X = x)$$

- Para una variable aleatoria continua:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

2. Varianza

La varianza mide la dispersión de los valores de la variable aleatoria respecto a su media. Se denota como $\text{Var}(X)$.

- Para una variable aleatoria discreta:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

- Para una variable aleatoria continua:

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x)\, dx$$

3. Desviación Estándar

La desviación estándar es la raíz cuadrada de la varianza y mide la dispersión de los valores de la variable aleatoria en las mismas unidades que la variable:

$$\sigma_X = \sqrt{\text{Var}(X)}$$

1.7.4. Ejemplo de Variables Aleatorias

Variable Aleatoria Discreta

Supongamos que lanzamos un dado. La variable aleatoria X que representa el número obtenido tiene la siguiente distribución:

$$P(X = x) = \begin{cases} \frac{1}{6} & \text{si } x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{de lo contrario} \end{cases}$$

El valor esperado (media) de X es:

$$\mathbb{E}[X] = \sum_{x=1}^{6} x \cdot \frac{1}{6} = 3{,}5$$

Variable Aleatoria Continua

Supongamos que una variable aleatoria X sigue una distribución uniforme en el intervalo $[0, 1]$. La PDF de X es:

$$f(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{de lo contrario} \end{cases}$$

La probabilidad de que X esté en el intervalo $[a, b]$ es:

$$P(a \leq X \leq b) = \int_a^b 1\, dx = b - a$$

El valor esperado de X es:

$$\mathbb{E}[X] = \int_0^1 x \cdot 1 \, dx = \frac{1}{2}$$

Capítulo 2

Distribucciones de probabilidad

Las distribuciones de probabilidad son funciones matemáticas que describen cómo se distribuyen los posibles resultados de un experimento aleatorio. En otras palabras, nos indican la probabilidad de que ocurran ciertos eventos o resultados dentro de un conjunto dado. Las distribuciones de probabilidad tienen características clave, como:

- **Función de densidad de probabilidad** (en distribuciones continuas) o **función de masa de probabilidad** (en distribuciones discretas).
- **Esperanza matemática (valor esperado)**: Es el promedio ponderado de todos los posibles resultados, según sus probabilidades.
- **Varianza y desviación estándar**: Miden la dispersión de los resultados alrededor del valor esperado.

Estas distribuciones son fundamentales en estadística, ya que nos permiten modelar la incertidumbre y tomar decisiones informadas en base a la probabilidad de distintos eventos.

2.1. Distribuciones de probabilidad Discretas

Una **distribución de probabilidad discreta** es un tipo de distribución en la que la variable aleatoria puede tomar solo un número finito o numerable de valores. En otras palabras, se refiere a aquellos casos en los que los resultados de un experimento aleatorio pueden contarse de forma clara y no son continuos, sino que tienen valores específicos.

2.1.1. Características principales

- **Valores discretos:** Los resultados posibles son finitos o numerables. Por ejemplo, el número de caras que puede salir al lanzar un dado es un conjunto discreto de valores: 1, 2, 3, 4, 5 y 6.

- **Función de masa de probabilidad (FMP):** Para las distribuciones discretas, se usa la **función de masa de probabilidad** (PMF, por sus siglas en inglés) para asignar una probabilidad específica a cada valor posible de la variable aleatoria. La PMF se denota generalmente como $P(X = x)$, donde X es la variable aleatoria y x es un valor específico de esa variable.

- **Suma de probabilidades:** La suma de todas las probabilidades debe ser igual a 1, es decir:
$$\sum_{x \in S} P(X = x) = 1$$
donde S es el conjunto de todos los valores posibles de la variable aleatoria.

- **Probabilidad no negativa:** La probabilidad de que ocurra cualquier resultado debe ser un número entre 0 y 1 (inclusive), es decir:
$$0 \leq P(X = x) \leq 1$$

- **Valor esperado (esperanza matemática):** Es el promedio ponderado de los posibles valores de la variable aleatoria, según sus probabilidades. Se calcula como:
$$E(X) = \sum_{x \in S} x \cdot P(X = x)$$

- **Varianza y desviación estándar:** Miden la dispersión de los resultados alrededor del valor esperado. La varianza $\mathrm{Var}(X)$ se calcula como:
$$\mathrm{Var}(X) = \sum_{x \in S} (x - E(X))^2 \cdot P(X = x)$$
Y la desviación estándar es la raíz cuadrada de la varianza:
$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

2.2. Distribución Binomial

La **distribución binomial** es una de las distribuciones de probabilidad más utilizadas en estadística, especialmente en experimentos que involucran resultados binarios, es decir, aquellos que solo tienen dos posibles resultados (como éxito o fracaso, cara o cruz, aprobado o desaprobado).

2.2.1. Definición Formal

La distribución binomial describe el número de **éxitos** en una secuencia de **n** ensayos independientes, donde en cada ensayo la probabilidad de éxito es constante. Es un caso particular de la distribución de probabilidad discreta, y está definida por dos parámetros:

- **n**: El número de ensayos o pruebas.
- **p**: La probabilidad de éxito en un solo ensayo.

Si X es una variable aleatoria que representa el número de éxitos en n ensayos, se dice que X sigue una distribución binomial, y se denota como:

$$X \sim \text{Binomial}(n, p)$$

donde:

- n es el número de ensayos.
- p es la probabilidad de éxito en cada ensayo.

2.2.2. Función de Masa de Probabilidad (PMF)

La **función de masa de probabilidad** (PMF) para una distribución binomial es la fórmula que nos da la probabilidad de obtener exactamente k éxitos (donde k es un valor específico entre 0 y n) en n ensayos. La fórmula es:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

donde:

- $\binom{n}{k}$ es el coeficiente binomial, que calcula el número de formas en que se pueden seleccionar k éxitos de n ensayos:

$$\binom{n}{k} = \frac{n!}{k!(n - k)!}$$

- p^k es la probabilidad de tener exactamente k éxitos.
- $(1 - p)^{n-k}$ es la probabilidad de tener $n - k$ fracasos.

2.2.3. Propiedades de la Distribución Binomial

- **Valor esperado (media):** El valor esperado $E(X)$ de una distribución binomial es el número promedio de éxitos que se espera en n ensayos. Se calcula como:

$$E(X) = n \cdot p$$

- **Varianza:** La varianza $\text{Var}(X)$ mide la dispersión de la distribución binomial, es decir, cuán dispersos están los posibles valores de X alrededor de la media. Se calcula como:

$$\text{Var}(X) = n \cdot p \cdot (1 - p)$$

- **Desviación estándar:** La desviación estándar σ_X es la raíz cuadrada de la varianza, y mide la dispersión en las mismas unidades que la variable X. Se calcula como:

$$\sigma_X = \sqrt{n \cdot p \cdot (1 - p)}$$

- **Distribución simétrica:** La distribución binomial es simétrica si $p = 0{,}5$ (cuando la probabilidad de éxito y fracaso es igual). Para valores de p cercanos a 0 o a 1, la distribución se vuelve sesgada hacia los valores extremos (es decir, hacia 0 o hacia n).

2.2.4. Condiciones para Usar la Distribución Binomial

Para que un experimento siga una distribución binomial, debe cumplir con las siguientes condiciones:

1. **Número fijo de ensayos (n):** El número de ensayos debe ser fijo y determinado de antemano.

2. **Dos resultados posibles en cada ensayo:** En cada ensayo debe haber solo dos resultados posibles: éxito (probabilidad p) o fracaso (probabilidad $1 - p$).

3. **Independencia de los ensayos:** Los ensayos deben ser independientes entre sí, es decir, el resultado de un ensayo no debe afectar a los resultados de los demás.

4. **Probabilidad constante de éxito:** La probabilidad de éxito debe ser constante en todos los ensayos.

2.2.5. Ejemplo Práctico

Supongamos que tenemos una moneda sesgada que tiene una probabilidad de 0.6 de caer en cara (éxito) y 0.4 de caer en cruz (fracaso). Si lanzamos la moneda 10 veces, la distribución de los resultados sigue una distribución binomial con $n = 10$ y $p = 0{,}6$.

La probabilidad de obtener exactamente 7 caras en 10 lanzamientos de la moneda sería:

$$P(X = 7) = \binom{10}{7}(0{,}6)^7(0{,}4)^{10-7}$$

Calculando este valor:

$$P(X = 7) = \frac{10!}{7!3!}(0{,}6)^7(0{,}4)^3 \approx 0{,}21499$$

Por lo tanto, la probabilidad de obtener exactamente 7 caras es aproximadamente 21.5 %.

2.2.6. Funciones principales para la distribución binomial en R

En RStudio (o R en general), trabajar con la distribución binomial es sencillo gracias a varias funciones específicas que R proporciona. A continuación se describen las principales funciones para trabajar con la distribución binomial:

- **dbinom():** Para calcular la función de masa de probabilidad (PMF), es decir, la probabilidad de obtener un número exacto de éxitos.
- **pbinom():** Para calcular la función de distribución acumulada (CDF), es decir, la probabilidad de obtener un número de éxitos menor o igual a un valor dado.
- **qbinom():** Para obtener el cuantil de la distribución binomial, es decir, el valor de k para una probabilidad acumulada dada.
- **rbinom():** Para generar números aleatorios de una distribución binomial.

Sintaxis básica

La sintaxis para trabajar con la distribución binomial en R es la siguiente:

- **dbinom(k, size, prob)**: Calcula la probabilidad de obtener exactamente k éxitos en una distribución binomial con n ensayos (`size`) y una probabilidad de éxito (`prob`).

- **pbinom(k, size, prob)**: Calcula la probabilidad acumulada de obtener k o menos éxitos.
- **qbinom(p, size, prob)**: Devuelve el valor k tal que la probabilidad acumulada es p.
- **rbinom(n, size, prob)**: Genera n números aleatorios de una distribución binomial con los parámetros *size* y *prob*.

Ejemplos prácticos

Supongamos que estamos realizando un experimento con los siguientes parámetros:

- Número de ensayos $n = 10$.
- Probabilidad de éxito $p = 0{,}6$.

1. Calcular la probabilidad de obtener exactamente 7 éxitos

Usamos la función `dbinom()`:

```r
# Parámetros
n <- 10  # número de ensayos
p <- 0.6  # probabilidad de éxito

# Calcular la probabilidad de obtener exactamente 7 éxitos
prob_7 <- dbinom(7, size = n, prob = p)
prob_7
```

2. Calcular la probabilidad de obtener 7 o menos éxitos

Usamos la función `pbinom()` para calcular la probabilidad acumulada:

```r
# Calcular la probabilidad de obtener 7 o menos éxitos
prob_7_o_menos <- pbinom(7, size = n, prob = p)
prob_7_o_menos
```

3. Calcular la probabilidad de obtener más de 7 éxitos

Para obtener la probabilidad de obtener más de 7 éxitos, podemos restar la probabilidad acumulada de 7 de 1:

```r
# Calcular la probabilidad de obtener más de 7 éxitos
prob_mas_7 <- 1 - pbinom(7, size = n, prob = p)
prob_mas_7
```

4. Generar números aleatorios de una distribución binomial

Si deseas generar n muestras aleatorias de la distribución binomial, puedes usar `rbinom()`:

```
# Generar 1000 muestras aleatorias de la distribución
#binomial
muestras <- rbinom(1000, size = n, prob = p)

# Ver las primeras 10 muestras generadas
head(muestras, 10)
```

5. Calcular el cuantil para una probabilidad acumulada dada

Si deseas saber cuántos éxitos se necesitan para que la probabilidad acumulada sea $p = 0{,}95$, puedes usar `qbinom()`:

```
# Calcular el cuantil para una probabilidad acumulada de
#0.95
cuantil_95 <- qbinom(0.95, size = n, prob = p)
cuantil_95
```

2.2.7. Visualización: Histograma de la distribución binomial

Puedes usar la función `hist()` para graficar la distribución binomial generando primero un conjunto de datos aleatorios y luego visualizándolos en un histograma:

```
# Generar 1000 muestras aleatorias de la distribución
#binomial
muestras <- rbinom(1000, size = n, prob = p)

# Crear el histograma
hist(muestras, main = "Histograma de la distribución
#binomial",
     xlab = "Número de éxitos", ylab = "Frecuencia",
     col = "lightblue", border = "black")
```

2.3. Distribución Poisson

La **distribución de Poisson** es una distribución de probabilidad discreta que describe el número de eventos que ocurren en un intervalo de tiempo o espacio determinado, bajo la condición de que estos eventos ocurren de manera independiente entre sí y a una tasa constante. Es especialmente útil para modelar fenómenos como el número de llamadas que recibe una central telefónica en una hora, el número de autos que pasan por un peaje en un día, o el número de fallos en una máquina en un determinado período de tiempo.

2.3.1. Función de Masa de Probabilidad (PMF)

El parámetro principal de la distribución es λ, que representa la tasa promedio de ocurrencia de eventos en el intervalo especificado (puede ser el número de eventos por unidad de tiempo, espacio, etc.). Si X sigue una distribución de Poisson con parámetro λ, la función de probabilidad de X se expresa como:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, 3, \ldots$$

donde:

- $P(X = k)$ es la probabilidad de que ocurran exactamente k eventos en el intervalo.
- λ es la tasa media de ocurrencia de eventos en el intervalo de interés.
- e es la base del logaritmo natural (aproximadamente 2.71828).
- $k!$ es el factorial de k.

2.3.2. Condiciones para la distribución de Poisson

Para que un fenómeno siga una distribución de Poisson, deben cumplirse las siguientes condiciones:

- Los eventos deben ser **independientes** entre sí.
- La tasa de ocurrencia λ debe ser **constante** en el intervalo de observación.
- El número de eventos no puede ser negativo, por lo tanto, k toma valores enteros no negativos: $k = 0, 1, 2, \ldots$.

2.3.3. Propiedades de la Distribución de Poisson

- **La media** (o esperanza matemática) de la distribución de Poisson es igual a λ:

$$E[X] = \lambda$$

- **La varianza** de la distribución de Poisson también es igual a λ:

$$\mathrm{Var}(X) = \lambda$$

  Esto implica que en una distribución de Poisson, la varianza y la media son iguales.

2.3.4. Ejemplo de la distribución de Poisson

Imaginemos que en una tienda se sabe que, en promedio, 5 clientes llegan por hora. Si $\lambda = 5$, podemos usar la fórmula de Poisson para calcular la probabilidad de que lleguen, por ejemplo, exactamente 3 clientes en una hora:

$$P(X = 3) = \frac{5^3 e^{-5}}{3!} = \frac{125 e^{-5}}{6} \approx 0{,}1404$$

Por lo tanto, la probabilidad de que lleguen exactamente 3 clientes en una hora es aproximadamente 0.1404, o un 14.04 %.

2.3.5. Usos de la distribución de Poisson

La distribución de Poisson es útil para modelar eventos que ocurren a una tasa constante e independiente en un intervalo fijo de tiempo o espacio. Algunos ejemplos incluyen:

- Modelar la frecuencia de ocurrencia de eventos raros, como el número de accidentes en una intersección o el número de defectos en un proceso de fabricación.
- Análisis de la tasa de llegada de clientes a una tienda, el número de llamadas entrantes a una central telefónica, o el número de autos que pasan por un peaje en un día.
- Evaluación de fenómenos en telecomunicaciones, como el número de paquetes de datos que llegan a un servidor por unidad de tiempo.

2.3.6. Funciones principales de Poisson en R

En `R`, existen varias funciones predefinidas para trabajar con la distribución de Poisson:

- `dpois(x, lambda)`: Calcula la función de densidad de probabilidad (PDF) o la probabilidad de obtener exactamente x eventos, dado un valor de λ (tasa promedio de ocurrencia de eventos).
- `ppois(q, lambda)`: Calcula la función de distribución acumulada (CDF), es decir, la probabilidad de que el número de eventos sea menor o igual a q.
- `qpois(p, lambda)`: Devuelve el valor de x tal que la probabilidad acumulada de obtener hasta x eventos es igual a p.
- `rpois(n, lambda)`: Genera n valores aleatorios que siguen una distribución de Poisson con parámetro λ.

Ejemplo: Calcular probabilidades con Poisson

Supongamos que tenemos un parámetro $\lambda = 5$, lo que significa que, en promedio, esperamos que ocurran 5 eventos por unidad de tiempo.

1. Calcular la probabilidad de obtener exactamente 3 eventos

Usamos la función `dpois()` para calcular la probabilidad de que ocurran exactamente 3 eventos:

```
# Parámetro lambda
lambda <- 5

# Número de eventos (x)
x <- 3

# Probabilidad de exactamente 3 eventos
prob_exacta <- dpois(x, lambda)
prob_exacta
```

2. Calcular la probabilidad de obtener hasta 3 eventos (probabilidad acumulada)

Usamos la función `ppois()` para calcular la probabilidad de que ocurran hasta 3 eventos:

```
# Probabilidad acumulada de 3 eventos o menos
prob_acumulada <- ppois(3, lambda)
prob_acumulada
```

3. Calcular la probabilidad de obtener más de 3 eventos

Podemos calcular la probabilidad de obtener más de 3 eventos restando la probabilidad acumulada hasta 3 de 1:

```
# Probabilidad de más de 3 eventos
prob_mas_3 <- 1 - ppois(3, lambda)
prob_mas_3
```

4. Generar valores aleatorios siguiendo una distribución de Poisson

Si queremos generar, por ejemplo, 100 valores aleatorios que sigan una distribución de Poisson con $\lambda = 5$, usamos `rpois()`:

```
# Generar 100 valores aleatorios de Poisson
n <- 100
valores_aleatorios <- rpois(n, lambda)
head(valores_aleatorios)  # Mostrar los primeros 6 valores generados
```

2.3.7. Visualización de la distribución de Poisson

Podemos visualizar la distribución de Poisson con diferentes valores de λ utilizando un gráfico de barras. Por ejemplo, para $\lambda = 5$, generamos un histograma de 1000 muestras aleatorias:

```
# Generar 1000 muestras de Poisson con lambda = 5
valores_aleatorios <- rpois(1000, lambda)

# Crear un histograma
hist(valores_aleatorios, breaks = 30, main = "Distribución de Poisson (la
     xlab = "Número de eventos", ylab = "Frecuencia", col = "lightblue", 
```

2.3.8. Ajuste de un modelo de Poisson

Si deseas ajustar un modelo de regresión de Poisson a datos de conteo, puedes usar la función `glm()` para ajustar un modelo de regresión generalizada. Un ejemplo básico de ajuste a un modelo de Poisson es el siguiente:

```
# Datos de ejemplo: número de eventos observados en diferentes intervalos
eventos <- c(3, 5, 4, 8, 2, 7, 5, 6, 10, 4)
tiempo <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)  # Intervalos de tiempo (todos

# Ajustar un modelo de Poisson
modelo_poisson <- glm(eventos ~ tiempo, family = poisson())

# Ver los resultados del modelo
summary(modelo_poisson)
```

Ejemplo completo en RStudio

Aquí tienes un ejemplo completo de cómo trabajar con la distribución de Poisson en RStudio:

```
# Parámetro lambda
lambda <- 5

# a) Calcular la probabilidad de exactamente 3 eventos
```

```r
prob_exacta <- dpois(3, lambda)
cat("Probabilidad de exactamente 3 eventos:", prob_exacta, "\n")

# b) Calcular la probabilidad acumulada hasta 3 eventos
prob_acumulada <- ppois(3, lambda)
cat("Probabilidad de hasta 3 eventos:", prob_acumulada, "\n")

# c) Calcular la probabilidad de más de 3 eventos
prob_mas_3 <- 1 - ppois(3, lambda)
cat("Probabilidad de más de 3 eventos:", prob_mas_3, "\n")

# d) Generar 100 valores aleatorios de Poisson
valores_aleatorios <- rpois(100, lambda)
cat("Primeros 6 valores aleatorios:", head(valores_aleatorios), "\n")

# e) Crear un histograma de la distribución de Poisson
hist(valores_aleatorios, breaks = 30, main = "Distribución de Poisso
     xlab = "Número de eventos", ylab = "Frecuencia", col = "lightbl
```

2.4. Distribuciones de probabilidad continuas

Las **distribuciones continuas de probabilidad** son un tipo de distribución de probabilidad que describe variables aleatorias continuas, es decir, variables que pueden tomar un número infinito de valores dentro de un intervalo dado. A diferencia de las distribuciones discretas, donde las probabilidades se asignan a valores específicos, en las distribuciones continuas la probabilidad de que una variable aleatoria tome un valor exacto es cero. En cambio, las probabilidades se asignan a intervalos de valores, y la probabilidad de que la variable tome un valor dentro de un intervalo se calcula mediante una **función de densidad de probabilidad** (FDP).

Características de las Distribuciones Continuas

- **Función de Densidad de Probabilidad (FDP)**: La distribución de probabilidad de una variable continua se describe mediante una función llamada *función de densidad de probabilidad* $f(x)$, que tiene las siguientes propiedades:

  $$f(x) \geq 0 \quad \text{para todos los valores de } x.$$

  $$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

  Esto asegura que la probabilidad total sea 1.

- **Probabilidad en Intervalos**: La probabilidad de que una variable continua tome un valor dentro de un intervalo $[a, b]$ se calcula integrando la FDP en ese intervalo:
$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$
- **Probabilidad de un Valor Exacto**: La probabilidad de que una variable continua tome un valor exacto $x = a$ es siempre cero:
$$P(X = a) = 0$$
- **Función de Distribución Acumulada (FDA)**: La *función de distribución acumulada* (FDA) de una variable continua X, denotada por $F(x)$, es la probabilidad de que X tome un valor menor o igual a x:
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$
La FDA es siempre una función no decreciente que varía entre 0 y 1.

2.5. Distribución Normal

La **distribución normal**, también conocida como **distribución gaussiana**, es una de las distribuciones de probabilidad más importantes en estadística. Se utiliza para modelar muchos fenómenos naturales y sociales, ya que describe cómo se distribuyen las variables aleatorias en muchos procesos del mundo real. Esta distribución tiene una forma simétrica, en forma de campana, que es característica de los fenómenos distribuidos de manera normal.

2.5.1. Definición y Características Generales

La distribución normal describe cómo se distribuyen los valores de una variable aleatoria continua. Sus principales características son:

- **Simetría**: La distribución normal es perfectamente simétrica alrededor de su media. Esto significa que los valores a la izquierda y a la derecha de la media tienen la misma probabilidad de ocurrir.
- **Forma de campana**: La gráfica de la distribución tiene la forma de una campana, lo que significa que la mayoría de los valores se concentran cerca de la media, y la probabilidad disminuye a medida que nos alejamos de ella.
- **Media, Mediana y Moda**: En una distribución normal, la media, la mediana y la moda coinciden en el mismo valor. La curva tiene un único pico en el centro.

- **Regla Empírica (68-95-99.7)**:
    - Aproximadamente el **68 %** de los valores se encuentran dentro de un intervalo de una desviación estándar alrededor de la media $\mu \pm \sigma$.
    - Aproximadamente el **95 %** de los valores se encuentran dentro de dos desviaciones estándar $\mu \pm 2\sigma$.
    - Aproximadamente el **99.7 %** de los valores se encuentran dentro de tres desviaciones estándar $\mu \pm 3\sigma$.
- **Área bajo la curva**: El área total bajo la curva de la distribución normal es igual a 1. Esto es una propiedad de todas las funciones de densidad de probabilidad, ya que representan probabilidades.
- **Asintoticidad**: Los "brazos"de la curva se acercan al eje horizontal (el eje x) pero nunca lo tocan. La distribución nunca termina; se extiende infinitamente en ambas direcciones.

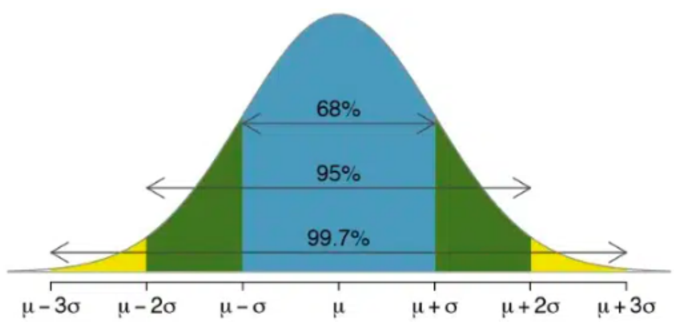

Figura 2.1: Distribución Normal (Recuperada de Miraltabank)

2.5.2. Función de Densidad de Probabilidad (FDP)

La función de densidad de probabilidad (FDP) de una distribución normal está dada por la siguiente fórmula:

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

Donde:

- μ es la *media* de la distribución, que indica el valor central o el promedio de los datos.
- σ es la *desviación estándar*, que mide la dispersión de los datos respecto a la media. Una desviación estándar mayor indica una distribución más extendida.

- e es la base de los logaritmos naturales (aproximadamente 2.71828).
- π es la constante matemática (aproximadamente 3.14159).

La distribución normal está completamente determinada por sus parámetros μ y σ. Si $\mu = 0$ y $\sigma = 1$, entonces la distribución es conocida como **distribución normal estándar**.

2.5.3. Parámetros de la Distribución Normal

La distribución normal se caracteriza por dos parámetros:

- **Media** μ: Es el valor esperado o promedio de la distribución. Determina la ubicación del centro de la curva.
- **Desviación estándar** σ: Mide la dispersión de los datos respecto a la media. Cuanto mayor sea la desviación estándar, más ancha será la curva.

Una variante muy común es la **distribución normal estándar**, que tiene media $\mu = 0$ y desviación estándar $\sigma = 1$.

2.5.4. Distribución Normal Estándar

La **distribución normal estándar** es un caso particular de la distribución normal donde:

- La media μ es igual a 0.
- La desviación estándar σ es igual a 1.

La función de densidad de la distribución normal estándar se representa como:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Los valores de la distribución normal estándar se denotan con la letra Z y se utilizan en muchas áreas de la estadística para hacer cálculos de probabilidades, usando tablas estándar o el concepto de puntuaciones Z.

2.5.5. Cálculo de Probabilidades en la Distribución Normal

En general, la probabilidad de que una variable aleatoria X siga una distribución normal tome un valor dentro de un intervalo específico $[a, b]$ se calcula mediante la integral de la función de densidad de probabilidad entre esos límites:

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

Para la distribución normal estándar, se usa la tabla de la normal estándar o el valor Z correspondiente, que se calcula como:

$$Z = \frac{X - \mu}{\sigma}$$

Una vez que tenemos el valor Z, podemos encontrar la probabilidad asociada usando tablas de la normal estándar o software estadístico.

2.5.6. Aplicaciones de la Distribución Normal

La distribución normal tiene muchas aplicaciones prácticas, algunas de las más comunes incluyen:

- **Control de calidad**: Para medir la variabilidad en los procesos de producción y asegurarse de que los productos cumplen con las especificaciones.

- **Evaluaciones psicológicas y educativas**: Muchos test, como los tests de coeficiente intelectual (IQ), están diseñados para seguir una distribución normal.

- **Errores de medición**: La distribución normal se usa para modelar los errores inherentes a cualquier tipo de medición o proceso experimental.

- **Finanzas**: En la teoría de carteras y la gestión de riesgos, la distribución normal se usa para modelar los retornos de las inversiones bajo ciertas condiciones.

2.5.7. Ejemplo Práctico

Imaginemos que el peso de los adultos en una población sigue una distribución normal con media $\mu = 70$ kg y desviación estándar $\sigma = 10$ kg. Si queremos calcular la probabilidad de que una persona pese entre 60 y 80 kg, utilizamos la distribución normal para calcular el área bajo la curva entre esos dos valores. Primero, transformamos los valores a puntuaciones Z, y luego usamos tablas o un software estadístico para obtener la probabilidad.

2.5.8. Funciones Principales para la Distribución Normal en R

- **`dnorm(x, mean = 0, sd = 1)`**: Esta es la función de *densidad de probabilidad* (FDP) de la distribución normal. Devuelve la densidad de probabilidad de obtener un valor x para una distribución normal con media μ y desviación estándar σ.

- **pnorm(q, mean = 0, sd = 1)**: Esta función calcula la *distribución acumulada* (CDF) de la distribución normal. Devuelve la probabilidad acumulada de obtener un valor menor o igual que q.

- **qnorm(p, mean = 0, sd = 1)**: Esta es la función *cuantil* de la distribución normal. Dado un valor de probabilidad p, devuelve el valor x tal que la probabilidad acumulada hasta x sea p.

- **rnorm(n, mean = 0, sd = 1)**: Esta función genera n números aleatorios de una distribución normal con media μ y desviación estándar σ.

Generar Números Aleatorios con una Distribución Normal

Para generar valores aleatorios que siguen una distribución normal con media μ y desviación estándar σ, se utiliza la función `rnorm()`. Aquí mostramos cómo generar 1000 números aleatorios con una media de 10 y una desviación estándar de 2:

```
# Generar 1000 números aleatorios con media = 10 y
#desviación estándar = 2
n <- 1000
media <- 10
desviacion_estandar <- 2
valores <- rnorm(n, mean = media, sd = desviacion_estandar)

# Mostrar los primeros 10 valores generados
head(valores)
```

Calcular la Probabilidad Acumulada

La función `pnorm()` nos permite calcular la probabilidad acumulada de que una variable aleatoria siga una distribución normal. A continuación, calculamos la probabilidad de que un valor sea menor o igual a 12 para una distribución con media 10 y desviación estándar 2:

```
# Probabilidad acumulada de que X <= 12
probabilidad <- pnorm(12, mean = media, sd = desv_estandar)
print(probabilidad)
```

Encontrar el Cuantil Correspondiente a una Probabilidad

Si queremos encontrar el valor x tal que la probabilidad acumulada es 0.95, usamos la función `qnorm()`:

```r
# Encontrar el valor que corresponde a la probabilidad
#acumulada de 0.95
cuantil <- qnorm(0.95, mean = media, sd = desv_estandar)
print(cuantil)
```

Ejemplo Completo: Analizando una Muestra de Datos

A continuación, mostramos un ejemplo completo en el que generamos una muestra de datos, calculamos la media y desviación estándar de la muestra, y graficamos la distribución junto con la probabilidad acumulada y el cuantil.

```r
# Generar 1000 datos con media 10 y desviación estándar 2
n <- 1000
media <- 10
desviacion_estandar <- 2
valores <- rnorm(n, mean = media, sd = desviacion_estandar)

# Calcular la media y desviación estándar de la muestra
media_muestra <- mean(valores)
desviacion_estandar_muestra <- sd(valores)

# Imprimir los resultados
cat("Media de la muestra:", media_muestra, "\n")
cat("Desviación estándar de la muestra:", desv_estandar_m,
"\n")

# Graficar el histograma con la densidad teórica
hist(valores, probability = TRUE, col = "lightblue",
main = "Histograma de la distribución normal",
     xlab = "Valores", ylab = "Densidad", border = "white")
curve(dnorm(x, mean = media, sd = desviacion_estandar),
col = "red", lwd = 2, add = TRUE)

# Calcular la probabilidad acumulada para X = 12
prob_12 <- pnorm(12, mean = media, sd = desv_estandar)
cat("Probabilidad acumulada para X <= 12:", prob_12, "\n")

# Encontrar el cuantil correspondiente a la probabilidad
#0.95
cuantil_95 <- qnorm(0.95, mean = media, sd = desv_estandar)
cat("Valor correspondiente a la p acumulada 0.95:",
cuantil_95, "\n")
```

2.6. Distribución t-student

La distribución t-Student es una distribución de probabilidad continua que es especialmente útil en situaciones en las que el tamaño de la muestra es pequeño y la varianza de la población es desconocida. Se utiliza comúnmente en pruebas de hipótesis, especialmente en la prueba t para una media de una sola muestra, la prueba t para dos muestras independientes, y en análisis de regresión.

2.6.1. Función de densidad de Probabilidad (PDF)

Dada una variable aleatoria T que sigue una distribución t-Student con ν grados de libertad (también denotados como df), su función de densidad de probabilidad está dada por:

$$f(t;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

donde:

- $f(t;\nu)$ es la densidad de probabilidad para el valor t.
- ν es el número de **grados de libertad**.
- $\Gamma(\cdot)$ es la **función gamma**, que generaliza el factorial para números reales (por ejemplo, $\Gamma(n) = (n-1)!$).
- π es la constante matemática pi ($\approx 3{,}14159$).

2.6.2. Propiedades de la Distribución t-Student

1. **Simetría**:
   - La distribución t-Student es simétrica respecto al cero, similar a la distribución normal.
2. **Colas más gruesas**:
   - Tiene colas más gruesas en comparación con la distribución normal, lo que indica una mayor probabilidad de obtener valores alejados de la media, especialmente cuando los grados de libertad son bajos.
3. **Grados de libertad** (ν):
   - A medida que el número de grados de libertad aumenta ($\nu \to \infty$), la distribución t-Student se aproxima a una **distribución normal estándar** ($\mu = 0, \sigma^2 = 1$).

2.6.3. Condiciones para Usar la Distribución t-Student

La **distribución t-Student** se utiliza en análisis estadísticos cuando ciertas condiciones están presentes. Esta distribución es especialmente útil cuando se trabaja con muestras pequeñas o cuando la varianza de la población es desconocida. A continuación, se detallan las principales condiciones para utilizar la distribución t-Student:

1. **Muestra pequeña**:
   - La distribución t-Student es apropiada cuando el tamaño de la muestra es relativamente pequeño, típicamente $n < 30$.
2. **Varianza poblacional desconocida**:
   - Se utiliza la distribución t-Student cuando la varianza (σ^2) o la desviación estándar (σ) de la población no son conocidas. En este caso, se emplea la **desviación estándar de la muestra** (s) como un estimador de σ.
3. **Población con distribución aproximadamente normal**:
   - La población de la cual se extrae la muestra debe seguir una **distribución aproximadamente normal**. Si bien la distribución t-Student es robusta ante desviaciones moderadas de la normalidad, este requisito es crucial cuando el tamaño de la muestra es muy pequeño ($n < 15$).
   - Para tamaños de muestra más grandes ($n \geq 30$), la t-Student se vuelve menos sensible a la normalidad debido al **Teorema Central del Límite**.
4. **Muestras independientes**:
   - Las observaciones en la muestra deben ser **independientes** entre sí. Esto significa que el valor de una observación no debe influir en los valores de otras observaciones.
5. **Datos numéricos continuos**:
   - La variable de interés debe ser **continua** y medida en una escala numérica. Por ejemplo, alturas, pesos, tiempos, etc.

2.6.4. Ejemplo Práctico

Un investigador quiere evaluar si el tiempo promedio de respuesta de un nuevo software es menor que 5 segundos. Para ello, selecciona una muestra aleatoria de $n = 10$ mediciones de tiempo (en segundos) y obtiene los siguientes datos:

$$4{,}8,\ 5{,}2,\ 4{,}9,\ 5{,}1,\ 4{,}7,\ 4{,}9,\ 5{,}0,\ 4{,}8,\ 5{,}3,\ 4{,}6$$

El investigador asume que el tiempo de respuesta sigue una distribución aproximadamente normal y que la varianza poblacional es desconocida. Quiere probar la hipótesis con un nivel de significancia de $\alpha = 0{,}05$.

Hipótesis

Las hipótesis a probar son:

$$\begin{cases} H_0 : \mu = 5 \text{ segundos (la media poblacional es 5 segundos)} \\ H_a : \mu < 5 \text{ segundos (la media poblacional es menor a 5 segundos)} \end{cases}$$

Este es un caso de **prueba t para una muestra** con una **prueba unilateral** a la izquierda.

Cálculo del Estadístico de Prueba

Primero, calculamos la **media muestral** ($\bar{x}$) y la **desviación estándar muestral** (s):

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Supongamos que estos cálculos dan como resultado:

$$\bar{x} = 4{,}93, \quad s = 0{,}233$$

El **estadístico de prueba t** se calcula como:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Sustituyendo los valores:

$$t = \frac{4{,}93 - 5}{\frac{0{,}233}{\sqrt{10}}} = \frac{-0{,}07}{0{,}0736} \approx -0{,}951$$

Cálculo del Valor Crítico usando la función inversa de t

El valor crítico para una prueba t-Student con $\alpha = 0{,}05$ y $n - 1 = 9$ grados de libertad se puede obtener usando la función de distribución acumulativa inversa:

$$t_{\text{crítico}} = t_{\alpha,\nu} = t_{0{,}05,9} = \text{invT}(0{,}05, 9) \approx -1{,}833$$

Decisión

Dado que el estadístico calculado $t = -0{,}951$ es mayor que el valor crítico $t_{\text{crítico}} = -1{,}833$, no podemos rechazar la hipótesis nula:

$$t > t_{\text{crítico}} \implies \text{No se rechaza } H_0$$

Esto significa que, con un nivel de significancia del 5 %, no hay suficiente evidencia para concluir que el tiempo promedio de respuesta del software es menor a 5 segundos.

Conclusión

La prueba t-Student para una muestra indica que no hay evidencia estadísticamente significativa para afirmar que el nuevo software tiene un tiempo de respuesta menor a 5 segundos. Por lo tanto, el investigador no puede rechazar la hipótesis nula.

2.6.5. Funciones principales para la distribución t-student en R

En RStudio (o R en general), trabajar con la distribución t-student es sencillo gracias a varias funciones específicas que R proporciona. A continuación se describen las principales funciones para trabajar con la distribución:

- **dt()**: Calcula la densidad (función de masa de probabilidad) en un punto x para una distribución t-Student con *df* grados de libertad.
- **pt()**: Calcula la probabilidad acumulada hasta un valor q para una distribución t-Student con *df* grados de libertad.
- **qt()**: Calcula el cuantil correspondiente a una probabilidad acumulada p para una distribución t-Student con *df* grados de libertad.
- **rt()**: Genera n números aleatorios de una distribución t-Student con *df* grados de libertad.

Sintaxis básica

Las funciones mencionadas tienen la siguiente sintaxis en R:

- **dt(x, df)**:Calcula la densidad de probabilidad en el valor especificado x para una distribución t-Student con los grados de libertad proporcionados.
- **pt(q, df)**: Calcula la probabilidad acumulada hasta el valor q para df grados de libertad.
- **qt(p, df)**: Calcula el cuantil correspondiente a la probabilidad acumulada p para df grados de libertad.
- **rt(n, df)**: Genera n números aleatorios de una distribución t-Student con df grados de libertad.

Ejemplos prácticos

Supongamos que estamos realizando un estudio sobre el rendimiento académico de estudiantes, donde tenemos una muestra pequeña (n = 12) y no conocemos la varianza poblacional. Queremos usar la distribución t-Student para hacer algunas estimaciones.ueremos calcular la densidad de probabilidad de obtener un valor $t = 1{,}8$ en una distribución t-Student con $df = 11$ grados de libertad.

1. Calcular la Densidad en un Punto con `dt()`

Queremos calcular la densidad de probabilidad de obtener un valor $t = 1{,}8$ en una distribución t-Student con $df = 11$ grados de libertad. Esto se puede hacer con la función `dt()` en R de la siguiente manera:

```r
# Parámetros
t <- 1.8          # valor de t
df <- 11          # grados de libertad

# Calcular la densidad en t = 1.8
densidad <- dt(t, df = df)
densidad
```

Este cálculo nos dará la densidad de probabilidad de obtener un valor t igual a 1.8 para 11 grados de libertad.

2. Calcular la Probabilidad Acumulada con R

Código:

```r
# Parámetros
q <- 1.5          # probabilidad acumulada
df <- 15          # grados de libertad

# Calcular la probabilidad acumulada
probabilidad_acumulada <- pt(q, df = df)
probabilidad_acumulada
```

Este resultado nos indica que la probabilidad acumulada hasta $t = 1{,}5$ para $df = 15$ es aproximadamente 0.9331.

3. 3. Calcular el Cuantil con `qt()`

```r
# Parámetros
p <- 0.975        # probabilidad acumulada cuantil
df <- 12          # grados de libertad

# Calcular el cuantil
cuantil <- qt(p, df = df)
cuantil
```

Este resultado nos indica que el valor t correspondiente al cuantil 97.5 % para $df = 12$ es aproximadamente 2.201.

3. 3. Calcular el Cuantil con `qt()`

Código:

```r
# Parámetros
n <- 5            # número de valores aleatorios
df <- 10          # grados de libertad

# Generar números aleatorios
numeros_aleatorios <- rt(n, df = df)
numeros_aleatorios
```

Este resultado nos muestra una muestra aleatoria de 5 valores generados a partir de una distribución t-Student con $df = 10$.

Bibliografía

[1] https://es.overleaf.com/learn.

[2] Shahbaba. B. (2011). Biostatistics with R: An introduction to Statistics Through Biological Data. Springer: Baltimore.