

Estadística

Pablo de la Cuesta García

2024-01-08

Contents

1	Introducción a la Probabilidad	2
1.1	Operaciones con sucesos	2
1.2	Leyes de Conjuntos en Probabilidad	2
1.3	Tipos de datos	3
1.4	Tablas de frecuencias	4
2	Estadística Descriptiva. Visualización de datos	5
2.1	Introducción	5
2.2	Resumen numérico y gráfico	6
2.3	Ejemplos	6
3	Probabilidad	6
3.1	Definición de probabilidad	6
3.2	Probabilidad condicionada	6
3.3	Dependencia e independencia de sucesos	6
3.4	Teorema de Bayes	6
4	Variables aleatorias	6
4.1	Variables aleatorias discretas y continuas	6
4.2	Función de probabilidad puntual	6
4.3	Función de densidad de probabilidad	6
4.4	Función de distribución de probabilidad	6
5	Principales distribuciones de probabilidad	6
5.1	Distribuciones discretas	6
5.2	Distribución uniforme	6
5.3	Distribución binomial	6

6	Distribuciones de variables discretas. Distribución de Poisson.	6
6.1	Función de densidad de probabilidad	6
6.2	Función de distribución de probabilidad	6
6.3	Condiciones	6

1 Introducción a la Probabilidad

La probabilidad es una medida que cuantifica la incertidumbre asociada a un evento o fenómeno. Se expresa como un número entre 0 y 1, donde 0 indica imposibilidad y 1 indica certeza absoluta.

Para poder medir la probabilidad de un evento, es necesario definir varios conceptos básicos:

- **Experimento aleatorio.** Es un experimento que se puede repetir en las mismas condiciones y que puede tener varios resultados posibles. Por ejemplo, lanzar un dado.
- **Espacio muestral.** Es el conjunto de todos los posibles resultados de un experimento aleatorio. Se denota por Ω .
- **Suceso.** Es un subconjunto del espacio muestral. Se denota por A .
- **Suceso elemental.** Es un elemento del espacio muestral. Se denota por ω .
- **Suceso seguro.** Es el suceso que contiene todos los elementos del espacio muestral. Se denota por Ω .
- **Suceso imposible.** Es el suceso que no contiene ningún elemento del espacio muestral. Se denota por \emptyset .

1.1 Operaciones con sucesos

- **Unión.** Se denota por $A \cup B$ y es el suceso que contiene todos los elementos de A y B .
- **Intersección.** Se denota por $A \cap B$ y es el suceso que contiene los elementos comunes de A y B .
- **Complementario.** Se denota por A^c y es el suceso que contiene todos los elementos del espacio muestral que no están en A .
- **Diferencia.** Se denota por $A \setminus B$ y es el suceso que contiene los elementos de A que no están en B .

1.2 Leyes de Conjuntos en Probabilidad

- **Unión.** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Intersección.** $P(A \cap B) = P(A) + P(B) - P(A \cup B)$
- **Complementario.** $P(A^c) = 1 - P(A)$
- **Probabilidad total.** $P(A) = P(A \cap B) + P(A \cap B^c)$

1.2.1 Ejemplo

Imagina que lanzamos dos monedas. El espacio muestral Ω es $\{CC, CS, SC, SS\}$, donde C representa cara y S sello.

- Evento A: al menos obtenemos una cara $A = CC, CS, SC$.
- Evento B: al menos obtenemos un sello $B = CS, SC, SS$.

Por tanto:

- $P(A) = 3/4$

- $P(B) = 3/4$
- $P(A \cap B) = 1/2$
- $P(A \cup B) = 1$

Usando la ley de la suma: $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 3/4 + 3/4 - 1/2 = 1$.

1.3 Tipos de datos

1.3.1 Datos Cualitativos

Los datos cualitativos o categóricos describen cualidades o características. No se expresan en números y por lo general se clasifica en categorías. Por ejemplo, el color de los ojos, el género, el estado civil, etc.

Dentro de los datos cualitativos, podemos distinguir entre:

- **Nominales.** No existe un orden entre las categorías. Por ejemplo, el color de los ojos.
- **Ordinales.** Existe un orden entre las categorías. Por ejemplo, el nivel de satisfacción de un cliente.

Este código crea un vector con colores de camisetas, un ejemplo típico de datos nominales donde cada color es una categoría sin un orden inherente.

```
# Crear un vector con datos nominales
colores_camisetas <- c("rojo", "azul", "verde", "azul", "rojo")

# Tabla de frecuencias
table(colores_camisetas)

## colores_camisetas
## azul rojo verde
##      2      2      1
```

Aquí, `niveles_educacion` es un factor con orden. Los niveles de educación tienen un orden inherente: primaria, secundaria y universitaria.

```
# Crear un factor con datos ordinales
niveles_educacion <- factor(c("primaria", "universitaria", "secundaria", "secundaria", "universitaria"),
                           levels = c("primaria", "secundaria", "universitaria"),
                           ordered = TRUE)

# Tabla de frecuencia
table(niveles_educacion)

## niveles_educacion
##      primaria      secundaria universitaria
##              1              2              2
```

1.3.2 Datos Cuantitativos

Los datos cuantitativos describen cantidades. Se expresan en números y se pueden realizar operaciones aritméticas con ellos. Por ejemplo, la edad, el peso, la altura, etc.

Dentro de los datos cuantitativos, podemos distinguir entre:

- **Discretos.** Los valores posibles son numerables. Por ejemplo, el número de hijos.
- **Continuos.** Los valores posibles son infinitos. Por ejemplo, la altura.

Este ejemplo muestra un vector con el número de estudiantes en diferentes clases, un claro ejemplo de datos discretos que son contables.

```
# Crear un vector con datos discretos
numero_estudiantes <- c(25, 30, 22, 28, 31)

# Calcular estadísticas básicas
summary(numero_estudiantes)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      22.0   25.0   28.0   27.2   30.0   31.0
```

En este caso, altura_estudiantes representa un conjunto de datos continuos, ya que la altura puede tomar cualquier valor en un rango y puede ser tan precisa como se desee.

```
# Crear un vector con datos continuos
altura_estudiantes <- c(1.70, 1.65, 1.80, 1.75, 1.60)

# Calcular estadísticas básicas
summary(altura_estudiantes)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.60   1.65   1.70   1.70   1.75   1.80
```

1.4 Tablas de frecuencias

Las tablas de frecuencias son una herramienta estadística fundamental para organizar y resumir datos. Son especialmente útiles para entender la distribución de los datos en un conjunto y para la visualización de datos categóricos o numéricos discretos.

Por tanto, una tabla de frecuencia es una representación organizada de los datos que muestra la frecuencia que cada valor de un conjunto de datos ocurre. Básicamente, clasifica los valores de un conjunto de datos en categorías y proporciona el número de observaciones que pertenecen a cada categoría.

1.4.1 Componentes de una tabla de frecuencias

- **Categorías.** Son los valores únicos de un conjunto de datos. Por ejemplo, los colores de las camisetas.
- **Frecuencia.** Es el número de observaciones que pertenecen a cada categoría. Por ejemplo, el número de camisetas de cada color.
 - **Frecuencia absoluta.** Es el número de observaciones que pertenecen a cada categoría.
 - **Frecuencia relativa.** Es la proporción de observaciones que pertenecen a cada categoría.
 - **Frecuencia acumulada.** Es el número de observaciones que pertenecen a cada categoría y a todas las categorías anteriores.

1.4.2 Ejemplo

Utilizaremos un ejemplo de calificaciones de estudiantes en R para ilustrar cómo crear una tabla de frecuencias. El conjunto de datos `calificaciones` contiene las calificaciones de 10 estudiantes en un examen de matemáticas.

```
# Crear el vector con las calificaciones
calificaciones <- c(5, 7, 8, 5, 10, 7, 6, 5, 9, 8)

# Calcular la frecuencia absoluta
frecuencia_absoluta <- table(calificaciones)

# Calcular la frecuencia relativa
frecuencia_relativa <- prop.table(frecuencia_absoluta) * 100

# Frecuencias acumuladas
frecuencia_acumulada <- cumsum(table(calificaciones))

# Representación de las frecuencias
tabla_frecuencias <- data.frame(
  "Calificación" = names(frecuencia_absoluta),
  "Frecuencia Absoluta" = as.integer(frecuencia_absoluta),
  "Frecuencia Relativa" = frecuencia_relativa,
  "Frecuencia Acumulada" = frecuencia_acumulada
)
```

```
## Warning in data.frame(Calificación = names(frecuencia_absoluta), 'Frecuencia
## Absoluta' = as.integer(frecuencia_absoluta), : row names were found from a
## short variable and have been discarded
```

```
# Cambiar los nombres de las columnas con colnames()
colnames(tabla_frecuencias) <- c("Calificación", "Frecuencia Absoluta", "Frecuencia Relativa", "Frecuencia Acumulada", "NA")

tabla_frecuencias
```

```
##   Calificación Frecuencia Absoluta Frecuencia Relativa Frecuencia Acumulada NA
## 1           5                 3             5          30 2
## 2           6                 1             6          10 4
## 3           7                 2             7          20 5
## 4           8                 2             8          20 2
## 5           9                 1             9          10 4
## 6          10                 1            10          10 5
```

2 Estadística Descriptiva. Visualización de datos

2.1 Introducción

- Una variable
 - Variables cualitativas y cuantitativas
 - Tablas de frecuencias
 - Representaciones gráficas

- Dos variables
 - Tablas de contingencia
 - Representaciones gráficas
 - Medidas de asociación

2.2 Resumen numérico y gráfico

2.3 Ejemplos

3 Probabilidad

3.1 Definición de probabilidad

3.2 Probabilidad condicionada

3.3 Dependencia e independencia de sucesos

3.4 Teorema de Bayes

4 Variables aleatorias

4.1 Variables aleatorias discretas y continuas

4.2 Función de probabilidad puntual

4.3 Función de densidad de probabilidad

4.4 Función de distribución de probabilidad

5 Principales distribuciones de probabilidad

5.1 Distribuciones discretas

5.2 Distribución uniforme

5.3 Distribución binomial

6 Distribuciones de variables discretas. Distribución de Poisson.

6.1 Función de densidad de probabilidad

6.2 Función de distribución de probabilidad

6.3 Condiciones