

Probabilidad-Parte I

Probabilidad Aplicada

Introducción a la probabilidad

Experimento: palabra para describir cualquier proceso que genere un conjunto de datos. Por ejemplo, lanzar una moneda.

Espacio muestral: Es el conjunto de todos los resultados posibles de un experimento estadístico.

Probabilidad: es una medida de la certidumbre asociada al resultado de un experimento o de un evento futuro.

Evento (o suceso): es un subconjunto de un espacio muestral.

Complemento de un evento: es el conjunto de todos los elementos del espacio muestral que no están en el evento.

Definición clásica de probabilidad

La probabilidad de ocurra el suceso o evento A se calcula como:

$$P(A) = \frac{\text{Casos favorables}}{\text{Casos totales}}$$

Donde *Casos favorables* hace referencia al número de casos en los que se obtiene como resultado A y *Casos totales* hace referencia a la cantidad total de resultados posibles, es decir, a la cantidad de elementos del conjunto muestral.

En el ejemplo del lanzamiento de una moneda, el espacio muestral S tiene dos resultados posibles. Cara o Cruz. Entonces $S = \{ "Cara", "Cruz" \}$.

Para conocer la probabilidad de obtener una Cara al lanzar una moneda tenemos que contar todos los resultados posibles de lanzar una moneda. Para ello, contamos la cantidad de elementos del espacio muestra y tenemos que **Casos totales**= 2 (podemos obtener una Cara o una Cruz). Ahora contamos la cantidad de resultados en donde obtenemos como resultado Cara. Tenemos **Casos totales**= 1 (en una moneda solo hay una Cara). Aplicando la definición clásica tenemos:

$$A: \text{ Suceso sacar "Cara" entonces, } P(A) = \frac{1}{2}$$

Veamos el ejemplo de construcción de un experimento.

Utilizamos la base *Insurance*. Para ello, extraemos la columna de la variable *género*.

```
genero <- insurance %>%
  select(genero)
genero
```

```
## # A tibble: 1,338 x 1
##   genero
##   <chr>
## 1 Femenino
## 2 Masculino
## 3 Masculino
## 4 Masculino
## 5 Masculino
## 6 Femenino
## 7 Femenino
## 8 Femenino
## 9 Masculino
## 10 Femenino
## # i 1,328 more rows
```

Ahora, vamos contar cuántas mujeres y hombres hay en nuestra muestra.

```
can_mujeres_hombres <- genero %>%
  group_by(genero) %>%
  summarise(cantidad = n())
can_mujeres_hombres
```

```
## # A tibble: 2 x 2
##   genero   cantidad
##   <chr>     <int>
## 1 Femenino     662
## 2 Masculino    676
```

Tenemos 662 Mujeres y 676 hombres.

¿Cuál es la probabilidad de seleccionar una mujer al azar?

- Casos favorables: Cantidad de mujeres que pueden ser seleccionadas (662)
- Casos totales: Cantidad de selecciones posibles (1338)

```
P <- can_mujeres_hombres %>%
  mutate(Probabilidad = cantidad/nrow(insurance))
P
```

```
## # A tibble: 2 x 3
##   genero   cantidad Probabilidad
##   <chr>     <int>     <dbl>
## 1 Femenino     662         0.495
## 2 Masculino    676         0.505
```

Con esta salida, tenemos que la probabilidad de seleccionar una mujer al azar en esta muestra, es de 0.4947. Pero también podemos conocer la probabilidad de seleccionar un hombre observado nuestra tabla de probabilidades.

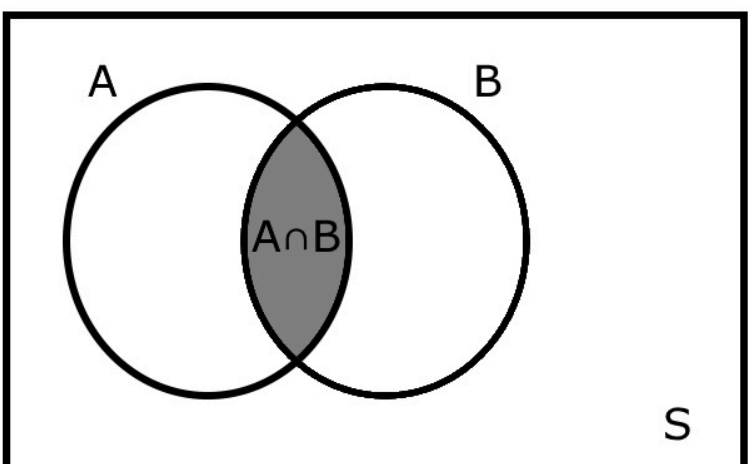
Propiedades de la probabilidad

Sea un evento A del espacio muestral S . Entonces:

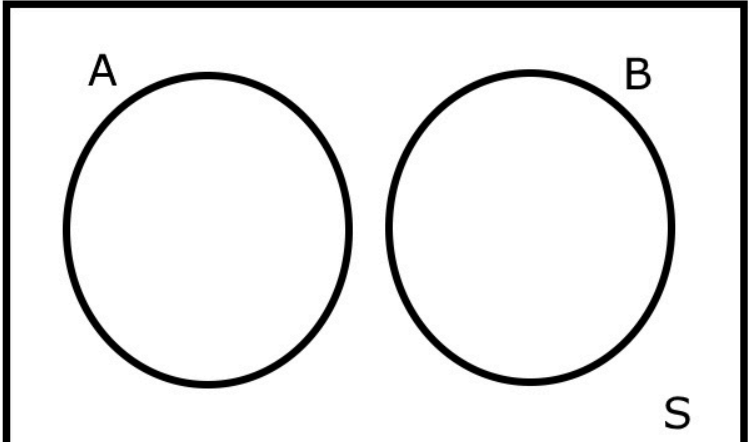
- $0 \leq P(A) \leq 1$
- $P(\emptyset) = 0$
- $P(S) = 1$

Más definiciones

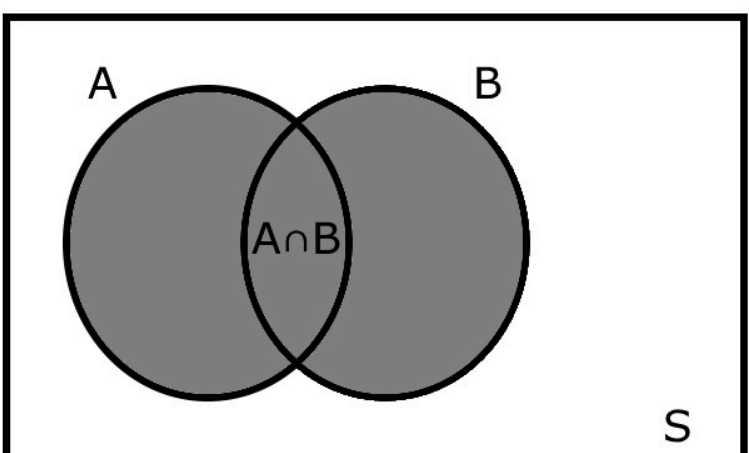
Intersección de dos eventos A y B: Se simboliza $A \cap B$ y es el evento que contiene a los elementos comunes de A y B .



Eventos mutuamente excluyentes o disjuntos: Dados dos eventos A y B , se dicen mutuamente excluyentes o disjuntos si $A \cap B = \emptyset$



Unión de dos eventos A y B: Se simboliza $A \cup B$ y es el evento que contiene a los elementos de A o B , o ambos.



Más propiedades

- Si A y B son dos eventos, entonces: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Si A y B son dos eventos mutuamente excluyentes, entonces: $P(A \cup B) = P(A) + P(B)$
- Si A y A' son eventos complementarios, entonces $P(A) + P(A') = 1$

Probabilidad condicional

La probabilidad de que ocurra un evento B cuando se sabe que ya ocurrió algún evento A se llama **Probabilidad condicional** y se simboliza $P(B|A)$. Esto se lee como "la probabilidad de que ocurra B dado que ocurrió A "

Se calcula como:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \text{ siempre que } P(A) > 0.$$

Veamos un ejemplo. La siguiente tabla resume el espacio muestral S constituido por la población de adultos de una pequeña ciudad que cumplen con los requisitos para obtener un título universitario. Están clasificados de acuerdo con su género y situación labora.

	Empleado	Desempleado	Total
Hombre	460	40	500
Mujer	140	260	400
Total	600	300	900

Clasificación de los adultos de una ciudad pequeña

Se seleccionará la azar a uno de estos individuos. Nos interesamos en los siguientes eventos:

B : Se elige a un hombre

A : el elegido tiene empleo

¿Cuál es la probabilidad de seleccionar un hombre dado que el elegido tiene empleo? es decir, $P(B|A) = ?$

Para poder calcular $P(B|A)$ necesitamos determinar:

$P(A \cap B)$: la probabilidad de seleccionar un hombre y que tenga empleo y $P(A)$: La probabilidad de seleccionar a alguien con empleo.

$$P(A \cap B) = \frac{\text{Cantidad de personas hombres y con empleo de la ciudad}}{\text{Cantidad de personas de la ciudad}} = \frac{460}{900}$$

$$P(A) = \frac{\text{Cantidad de personas con empleo de la ciudad}}{\text{Cantidad de personas de la ciudad}} = \frac{600}{900}$$

Entonces, reemplazando en la fórmula:

$$P(B|A) = \frac{\frac{460}{900}}{\frac{600}{900}} = \frac{23}{30} \approx 0.76$$

Tablas de contingencia

Una **tabla de contingencia** es una forma de caracterizar los datos, para describir o analizar la asociación entre dos variables. Facilita el calculo de probabilidades.

Un ejemplo de tabla de contingencia es el de la tabla de la clasificación de los adultos de una ciudad pequeña, cuando se estudió la probabilidad condicional:

	Empleado	Desempleado	Total
Hombre	460	40	500
Mujer	140	260	400
Total	600	300	900

Clasificación de los adultos de una ciudad pequeña

Tabla de contingencia en R: Package “expss”

Primero instalamos y cargamos la librería **expss**

```
install.packages("expss")
library(expss)
```

Supongamos que deseamos construir una tabla de contingencia por género y por región de la base *insurance*.

Entonces, en primer lugar creamos nuestra tabla utilizando la función **cross_case()** de la librería **expss**. La llamamos **TC01**

```
TC01 <- cross_cases(insurance, genero, region)
```

	region			
	noreste	noroeste	sureste	suroeste
genero				
Femenino	161	164	175	162
Masculino	163	161	189	163
#Total cases	324	325	364	325

Esta primera tabla de contingencia, nos cuenta los casos en cada combinación de las variables género y región. Además, la última fila nos provee los totales marginales por columnas. Lo único que nos falta, es incorporar el total marginal por filas. Veamos cómo...

```
TC02 <- TC01 %>%
  mutate(Total = rowSums(TC01[, 2:5]))
```

Con la función **mutate()** agregamos la columna **Total**. Esta columna debe sumar las celdas por filas (con la función **rowSum()**, de las columnas de TC01 que van de la 2 a la 5. Obtenemos:

	region				Total
	noreste	noroeste	sureste	suroeste	
genero					
Femenino	161	164	175	162	662
Masculino	163	161	189	163	676
#Total cases	324	325	364	325	1338