

4. PROBABILIDAD

La probabilidad mide la frecuencia con la que ocurre un resultado en un experimento bajo condiciones suficientemente estables. La teoría de la probabilidad se usa en áreas como la estadística, la matemática, la ciencia y la filosofía para sacar conclusiones sobre la posibilidad de sucesos potenciales y la mecánica subyacente de sistemas complejos. La teoría de la probabilidad constituye la base o fundamento de la estadística, ya que las inferencias que se hagan sobre la población o poblaciones en estudio se moverán dentro de unos márgenes de error controlado, que será medido en términos de probabilidad.

Las probabilidades son muy útiles, ya que pueden servir para desarrollar estrategias. Por ejemplo, algunos conductores parecen mostrar una mayor tendencia a aumentar la velocidad si creen que existe un riesgo pequeño de ser multados; los inversionistas estarán más interesados en invertir dinero si las posibilidades de ganar son buenas. El punto central en todos estos casos es la capacidad de cuantificar cuán probable es determinado evento.

Algunas funciones para calcular probabilidades en R requieren cargar paquetes adicionales a los estándar, como `combinat` y `prob`; por tanto, se recomienda cargar estos paquetes antes de utilizar comandos relacionados con probabilidad.

4.1 CONJUNTOS

La palabra conjunto implica la idea de una colección de objetos que se caracterizan por algo común. En matemática tiene el mismo significado, sólo que a estos objetos se les llama elementos o miembros del conjunto. La noción simple de una colección o conjunto de objetos es fundamental en la estructura básica de las matemáticas y fue Georg Cantor, en el año 1870, quien primero llamó la atención de los matemáticos a este respecto. En R, un conjunto se puede pensar como un vector donde cada uno de los elementos que lo componen son los elementos de un conjunto determinado.

En R se pueden trabajar las diferentes operaciones entre conjuntos. Considerando los siguientes conjuntos:

```
>
> A=c(0,1,2,3,4,5,6,7,8,9)
>
>
> B=c(3,6,9,12,15,18,21,24,27,30)
>
```

Imagen 67. Salida R para la determinación de conjuntos

las operaciones unión, intersección, diferencia y diferencia simétrica se llevan a cabo de la siguiente manera:

4.1.1 **Unión.** La unión de los conjuntos A y B es el conjunto formado por todos los elementos que pertenecen a A o a B o a ambos. Se denota por $A \cup B$ y el comando utilizado para calcularla es `union()`.

```
>
> A=c(0,1,2,3,4,5,6,7,8,9)
> B=c(3,6,9,12,15,18,21,24,27,30)
> union(A,B)
[1] 0 1 2 3 4 5 6 7 8 9 12 15 18 21 24 27 30
>
```

Imagen 68. Salida R para la unión de dos conjuntos

4.1.2 **Intersección.** Se define la intersección de dos conjuntos A y B, como el conjunto de elementos que son comunes a A y B, que se lee A intersección B, y el comando utilizado para calcularla es `intersect()`.

```
>
>
> intersect(A,B)
[1] 3 6 9
>
```

Imagen 69. Salida R para la intersección de dos conjuntos

4.1.3 **Diferencia.** Se denomina diferencia de dos conjuntos A y B al conjunto formado por todos los elementos que pertenecen a A, pero que no pertenecen a B. La diferencia se denota por $A - B$ y se lee A diferencia B o A menos B, y en R el comando utilizado para calcularla es `setdiff()`.

```
>
> setdiff(A,B)
[1] 0 1 2 4 5 7 8
>
> setdiff(B,A)
[1] 12 15 18 21 24 27 30
>
```

Imagen 70. Salida R para la diferencia de dos conjuntos

4.1.4 **Diferencia simétrica.** Se denomina diferencia simétrica de dos conjuntos A y B al conjunto formado por todos los elementos que pertenecen a la unión de A y B, pero que no pertenecen a la intersección de A y B. Se lee A diferencia simétrica B. En R se puede calcular la diferencia simétrica de dos conjuntos al realizar la unión entre las diferencias de los conjuntos, como se mostró en la operación anterior.

```

>
> union(setdiff(A,B),setdiff(B,A))
[1] 0 1 2 4 5 7 8 12 15 18 21 24 27 30
>

```

Imagen 71. Salida R para la diferencia simétrica de dos conjuntos

Dos conjuntos son iguales si los elementos que los conforman son los mismos; el comando `setequal()` permite verificar si dos conjuntos son iguales, utilizando los A y B se tiene,

```

> setequal(A,B)
[1] FALSE
>

```

Imagen 72. Salida R para la igualdad de dos conjuntos

Como el resultado es FALSE, estos dos conjuntos no son iguales.

4.2 ESPACIO MUESTRAL

Un espacio muestral es el conjunto de todos los resultados posibles de un evento o muestra. Es el conjunto de todos los resultados posibles de un experimento estadístico, y se representa con la letra S.

4.2.1 Probabilidad de eventos. La probabilidad de un evento A en un espacio muestral S es la suma de todos los puntos muestrales que conforman el evento A dividido en el tamaño de S; por lo tanto $0 < P(A) < 1$, $P() = 0$, y $P(S) = 1$, es decir, la probabilidad de que un evento ocurra está dada mediante un número que va desde de 0 a 1.

4.2.2 Técnicas de conteo. Son aquellas que son usadas para enumerar eventos difíciles de cuantificar; son principios que se usan para contar resultados que no se conocen o que son muy extensos. Se les denomina técnicas de conteo a: combinaciones y permutaciones, cabe destacar que estas proporcionan la información de todas las maneras posibles en que ocurre un evento determinado.

4.2.3 Factorial. El factorial de un número entero positivo se define como el producto de todos los números naturales anteriores o iguales a él. Se escribe $n!$, y se lee “n factorial” (por definición, el factorial de 0 es 1, así: $0! = 1$). Si se quiere calcular $7!$, su desarrollo y comando utilizado es:

$$7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

```

>
> factorial(7)
[1] 5040
>

```

Imagen 73. Salida R para el cálculo de un factorial

4.2.4 Combinaciones. Una combinación es un arreglo de elementos en donde no interesa el lugar o posición que estos ocupan. En una combinación el objetivo está en formar grupos, y el contenido de estos, el número de combinaciones de n objetos tomados de r , se puede calcular mediante el comando `choose(n, r)`.

Ejemplo: Si se cuenta con 5 personas para seleccionar 2, sin importar el orden de selección, el número total de combinaciones posibles entre estas personas es de:

$n C r = \frac{n!}{(n-r)! r!}$	<pre style="margin: 0;">> > choose(5,2) [1] 10 ></pre>
--------------------------------	---

Ecuación 6 e imagen 74. Salida R para el cálculo de una combinación

Con lo que se tiene que existen 10 posibles combinaciones de 2 personas; ahora bien, si se desea establecer el espacio muestral, es decir, observar las 10 combinaciones, se utilizan los comandos dados en la siguiente ilustración; los nombres de las personas son Bibiana, Carlos, Sandra, Carmen y Reinaldo.

```
>
> # Vector con los nombres de las personas a combinar
>
> Per=c("Bibiana","Carlos","Sandra","Carmen","Reinaldo")
>
> # Comando para crear el espacio muestral asociado
>
> urnsamples(Per, size = 2)
      X1      X2
1 Bibiana Carlos
2 Bibiana  Sandra
3 Bibiana  Carmen
4 Bibiana Reinaldo
5 Carlos   Sandra
6 Carlos   Carmen
7 Carlos Reinaldo
8 Sandra   Carmen
9 Sandra Reinaldo
10 Carmen Reinaldo
>
```

Imagen 75. Salida R para determinar las posibles combinaciones

Obsérvese que el comando que permite construir el espacio muestral asociado con la combinatoria es `urnsamples(x, size)`, donde x es el vector que contiene los objetos por combinar y `size` se refiere a la cantidad de objetos tomados en cada grupo.

4.2.5 Permutaciones. Es todo arreglo de elementos en donde interesa el lugar o posición que ocupa cada uno de ellos. El número de permutaciones de n objetos tomados de r en r objetos se puede calcular mediante el comando `nsamp(n,r,ordered=TRUE)`.

Ejemplo: Si se cuenta con 5 personas y se quieren seleccionar 2, importando el orden de selección, el número total de permutaciones posibles es de:

${}_n P_r = \frac{n!}{(n-r)!}$	<pre> > > nsamp(5,2,ordered=TRUE) [1] 20 > > > n=5 > r=2 > Permutaciones=(factorial(n)/factorial(n-r)) > Permutaciones [1] 20 > . </pre>
--------------------------------	--

Ecuación 7 e imágenes 76 y 77. Salida R para el cálculo de una permutación

Con lo que se tiene que existen 20 permutaciones de 2 personas; ahora bien, si se desea establecer el espacio muestral, el procedimiento es (tomando los mismos nombres de las personas del ejemplo de combinatoria):

```

>
> urnsamples(Per,size=2,ordered=TRUE)

      X1      X2
1  Bibiana Carlos
2  Carlos  Bibiana
3  Bibiana  Sandra
4  Sandra  Bibiana
5  Bibiana  Carmen
6  Carmen  Bibiana
7  Bibiana Reinaldo
8  Reinaldo Bibiana
9  Carlos  Sandra
10 Sandra  Carlos
11 Carlos  Carmen
12 Carmen  Carlos
13 Carlos  Reinaldo
14 Reinaldo Carlos
15 Sandra  Carmen
16 Carmen  Sandra
17 Sandra  Reinaldo
18 Reinaldo Sandra
19 Carmen  Reinaldo
20 Reinaldo Carmen
>

```

Imagen 78. Salida R para determinar las posibles permutaciones

El comando que permite construir el espacio muestral asociado con la permutación es el mismo utilizado para la combinación, adicionándole el argumento `ordered=TRUE`. Esta permutación también se puede realizar con el comando `permsn(x, r)`, donde `x` es el vector que contiene los objetos por permutar y `r` se refiere a la cantidad de objetos tomados en cada grupo.

Los paquetes `combinat` y `prob` permiten generar experimentos aleatorios, por tanto, si se desea desarrollar los experimentos aleatorios que se mencionan a continuación, se hace necesario que estos paquetes se carguen previamente. A continuación se especifica cómo realizar estos experimentos aleatorios y la forma de determinar su espacio muestral.

4.2.6 `rolldie(times, nsides=6, makespace=TRUE)` genera el espacio muestral para un experimento aleatorio consistente en el lanzamiento de un dado varias veces. Los argumentos utilizados son: `times`, número de lanzamientos deseados; `nsides`, número de caras del dado, y `makespace=`, que si es `FALSE`, solo genera el espacio muestral, y si es `TRUE`, genera una columna adicional con la probabilidad de cada punto muestral.

Ejemplo: Considere el lanzamiento de un dado que tiene 3 caras, numeradas con 1, 2 y 3, respectivamente, en 2 ocasiones.

```
>
> rolldie(2, nsides = 3, makespace = TRUE)
  x1 x2      probs
1  1  1 0.1111111
2  2  1 0.1111111
3  3  1 0.1111111
4  1  2 0.1111111
5  2  2 0.1111111
6  3  2 0.1111111
7  1  3 0.1111111
8  2  3 0.1111111
9  3  3 0.1111111
>
```

Imagen 79. Salida R para el experimento aleatorio "dados"

4.2.7 `tosscoin(times, makespace=TRUE)` genera el espacio muestral para un experimento aleatorio consistente en el lanzamiento de una moneda varias veces. Los argumentos utilizados son: `times`, número de lanzamientos deseados, y `makespace=`, que si es `FALSE`, solo genera el espacio muestral, y si es `TRUE`, genera una columna adicional con la probabilidad de cada punto muestral.

Ejemplo: Considere el lanzamiento de una moneda en dos ocasiones.

```

>
> tosscoin(2, makespace = FALSE)
  toss1 toss2
1      H      H
2      T      H
3      H      T
4      T      T
> tosscoin(2, makespace = TRUE)
  toss1  toss2 probs
1      H      H 0.25
2      T      H 0.25
3      H      T 0.25
4      T      T 0.25
>

```

Imagen 80. Salida R para el experimento aleatorio "monedas"

4.2.8 `urnsamples(x,size,replace=,ordered=)` crea un espacio muestral asociado con un experimento en el cual se muestrean objetos distinguibles de una urna. Los argumentos utilizados al interior de este comando son: `x`, un vector con el nombre de los objetos a ser muestreados; `size` indica el número de objetos de la muestra; `replace`, si es `TRUE` realiza un muestreo con reemplazamiento, si es `FALSE` realiza el muestreo sin reemplazamiento, y `ordered`, si es `TRUE` importa el orden en la selección de las muestras, si es `FALSE` no importa el orden de selección en las muestras. Los tamaños de cada uno de los escenarios anteriores con un vector `x` de tamaño `n` y tamaño de muestra `size` igual a `r`, son como sigue:

Replace	Ordered	Número de muestras
TRUE	TRUE	n^r
FALSE	TRUE	$\frac{n!}{(n-r)!}$
FALSE	FALSE	$\frac{n!}{(n-r)!} r!$
TRUE	FALSE	$\frac{(n-1+r)!}{(n-r)!} r!$

Imagen 81. Tamaños para los posibles escenarios del comando

4.2.9 **Cálculo de una probabilidad y de probabilidades condicionales de eventos.** En R es posible calcular la probabilidad de un evento, dado que se conoce el espacio muestral; además, es posible calcular la probabilidad de un evento, dado que otro evento ya ha ocurrido. El comando que permite calcular lo mencionado anteriormente es `prob(S,evento,given=evento)`, donde `S` hace referencia al espacio muestral, `evento` define el evento al cual se le calcula la probabilidad y `given=` se usa para determinar una probabilidad condicional, es decir, calcular la probabilidad de un evento, dado que ocurrió otro evento.

Ejemplo: Considere el experimento aleatorio (E.A.) lanzamiento de un dado en dos oportunidades; el espacio muestral está determinado mediante una asignación como se indica:

```
>
> s <- rolldie(times = 2, nsides=6, makespace = TRUE)
>
```

Imagen 82. Construcción experimento aleatorio

El tamaño del espacio muestral para este experimento es de 36, a continuación se muestran los 10 primeros puntos muestrales del espacio:

```
> s
      X1 X2      probs
1     1  1  0.02777778
2     2  1  0.02777778
3     3  1  0.02777778
4     4  1  0.02777778
5     5  1  0.02777778
6     6  1  0.02777778
7     1  2  0.02777778
8     2  2  0.02777778
9     3  2  0.02777778
10    4  2  0.02777778
```

Imagen 83. Salida R para los primeros 10 puntos del E.A.

En el espacio muestral construido, el primer lanzamiento es denotado por X1, y el segundo, por X2. Ahora bien, luego de generar el espacio muestral y de asignarle la letra S para el experimento, se procede a realizar el cálculo de algunas probabilidades de interés.

Probabilidad de que la suma de las caras superiores de los dados sea igual a 4.	<pre>> > probs(s, X1+X2 ==4) [1] 0.08333333 ></pre>
Probabilidad de que la suma de las caras superiores de los dados sea menor a 5.	<pre>> > probs(s, X1+X2 < 5) [1] 0.1666667 ></pre>
Probabilidad de que la suma de las caras superiores de los dados sea mayor a 8.	<pre>> > probs(s, X1+X2 > 8) [1] 0.2777778 ></pre>
Probabilidad de que en el primer lanzamiento caiga 5, dado que la suma de las caras superiores de los dados sea mayor a 9.	<pre>> > probs(s, X1 == 5, given = X1+X2 >9) [1] 0.3333333 ></pre>

Imágenes 84, 85, 86 y 87. Salida R cálculo de probabilidades

4.3 EJERCICIOS

4.3.1 Dados los conjuntos A y B, escríbalos en R y luego realice su unión, intersección, diferencia y diferencia simétrica.

$$A = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25\}$$

$$B = \{8, 9, 10, 11, 12, 13, 14, 15, 16\}$$

4.3.2 En una carrera atlética los cuatro primeros en llegar fueron Andrés, Juan, Felipe y Pablo, sin ser este el orden de llegada, responda:

- De cuántas formas diferentes pudieron llegar los atletas, y
- Construya el espacio muestral.

4.3.3 Una empresa selecciona 3 empleados, entre 8 candidatos, para un curso de ascenso; la selección se realiza al azar y no importa el orden. Empleados (a, b, c, d, e, f, g, h)

- ¿Cuántos grupos diferentes se pueden conformar entre los empleados para tomar el curso de ascenso?
- Construya el espacio muestral.

4.3.4 Considere el lanzamiento en 3 ocasiones de un dado que tiene 6 caras, numeradas del 1 al 6.

- Construya el experimento aleatorio en R
- Calcule la probabilidad de que la suma de las caras superiores sea 14
- Calcule la probabilidad de que la suma de las caras superiores sea mayor a 10.