

## SESIÓN PROBABILIDAD Y ESTADÍSTICA

### CONTENIDOS:

- La teoría de probabilidades.
- Experimento aleatorio y determinístico.
- Espacio muestral.
- Eventos aleatorios.
- Árbol de probabilidades.
- Unión e intersección de eventos.
- Muestra y muestreo.
- Muestra aleatoria y sesgo de muestra.
- Buenas prácticas para obtener muestras representativas:
  - Muestreo aleatorio simple.
  - Muestreo aleatorio estratificado.
  - Muestreo por conglomerados.
  - Muestreo de varias etapas.

### LA TEORÍA DE PROBABILIDADES

La teoría de probabilidades es una rama de las matemáticas que estudia los fenómenos aleatorios. Se utiliza para cuantificar la incertidumbre y predecir resultados en situaciones donde el azar juega un papel importante. La probabilidad es una medida numérica que varía entre 0 y 1, donde 0 indica imposibilidad y 1 certeza.

- Aplicaciones: Predicción del clima, análisis de riesgos financieros, Machine Learning, entre otros.
- Ejemplo: La probabilidad de que salga "cara" al lanzar una moneda es 0.5.

En el contexto de los datos, la teoría de probabilidades se utiliza para modelar y analizar situaciones que involucran incertidumbre, como la predicción de comportamientos futuros basados en datos pasados. Los conceptos fundamentales incluyen eventos, espacios muestrales, probabilidad condicionada, variables aleatorias y distribuciones de probabilidad.

## EXPERIMENTO ALEATORIO Y DETERMINÍSTICO

En la teoría de probabilidades y estadística, los experimentos pueden clasificarse en dos categorías fundamentales: aleatorios y determinísticos. Comprender la diferencia entre ambos tipos de experimentos es esencial para el análisis de datos y la toma de decisiones informadas, especialmente en campos como la ciencia de datos, la estadística y la probabilidad.

### Experimento Aleatorio

Un experimento aleatorio es aquel en el que los resultados posibles son conocidos, pero el resultado específico no se puede predecir con certeza antes de realizarlo, debido a la influencia del azar. Estos experimentos se caracterizan por tener múltiples resultados posibles, y la probabilidad es una herramienta clave para describir la incertidumbre asociada con los resultados. La probabilidad se utiliza para asignar una medida de certeza a cada posible resultado, y se puede aplicar en contextos que requieren modelar situaciones impredecibles.

Ejemplo:

Lanzar un dado. Los resultados posibles son {1, 2, 3, 4, 5, 6}, pero el número que aparecerá no se puede predecir antes de lanzar el dado, ya que está sujeto a la aleatoriedad del proceso.

### Experimento Determinístico

Un experimento determinístico es aquel cuyo resultado es completamente predecible y no está influenciado por el azar. En estos experimentos, si se conocen todas las condiciones iniciales y las reglas que rigen el proceso, se puede calcular con precisión el resultado. Los experimentos determinísticos son fundamentales en áreas como las matemáticas y las ciencias físicas, donde las leyes de la naturaleza siguen patrones fijos y sin variabilidad aleatoria.

Ejemplo:

Calcular el área de un círculo dado su radio. Si se conoce el radio ( $r$ ), se puede usar la fórmula  $A = \pi r^2$  para obtener el área, y el resultado será siempre el mismo para ese valor de  $r$ .

## ESPACIO MUESTRAL

El espacio muestral (denotado como  $S$ ) es el conjunto de todos los resultados posibles que pueden surgir de un experimento aleatorio. Es uno de los conceptos fundamentales en la teoría de probabilidades, ya que sirve como la base para el análisis de los posibles resultados de un experimento y la asignación de probabilidades a dichos resultados. En otras palabras, el espacio muestral define todas las opciones que pueden ocurrir durante un experimento.

Ejemplo 1:

Al lanzar un dado, los resultados posibles son los números del 1 al 6. Por lo tanto, el espacio muestral en este caso sería:

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Cada uno de los números en este conjunto representa un posible resultado del experimento.

Ejemplo 2:

Si se lanza una moneda, el espacio muestral es  $S = \{\text{cara}, \text{cruz}\}$ , representando las dos posibilidades del experimento.

## Representación Visual

El espacio muestral puede representarse visualmente de diferentes formas, dependiendo del tipo de experimento y la complejidad del espacio muestral:

1. **Diagrama de Venn:** Se puede usar un diagrama de Venn para representar conjuntos dentro del espacio muestral. En este tipo de representación, el espacio muestral es el conjunto universal, y se pueden trazar subconjuntos dentro de este espacio, como los resultados favorables o los que cumplen con una condición específica.

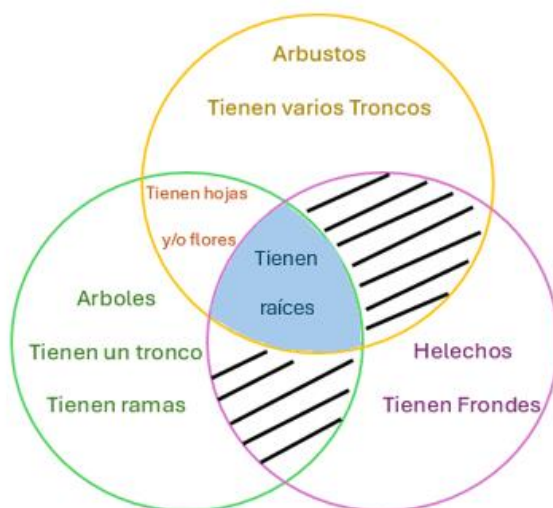


Ilustración 1 Diagrama de Venn

2. **Tabla de resultados:** En algunos casos, especialmente cuando los resultados posibles son pocos y discretos, se puede usar una tabla para mostrar todos los posibles resultados del experimento. Esto ayuda a organizar y visualizar fácilmente el espacio muestral, así como los posibles eventos de interés.

## EVENTOS ALEATORIOS

Un evento es cualquier conjunto de resultados posibles dentro de un espacio muestral. Los eventos se dividen en dos tipos principales: simples y compuestos. Un evento simple corresponde a un solo resultado del experimento, mientras que un evento compuesto incluye varios resultados dentro del espacio muestral.

### Tipos de Eventos:

- **Evento Simple:** Es un evento que consta de un solo resultado dentro del espacio muestral. Ejemplo: Al lanzar un dado, el evento "obtener un número impar" es  $E = \{1, 3, 5\}$ .
- **Evento Compuesto:** Es un evento que involucra varios resultados del espacio muestral. Ejemplo: En el lanzamiento de un dado, el evento "obtener un número par" es  $E = \{2, 4, 6\}$ . Este es un evento compuesto porque abarca más de un resultado posible.

### Probabilidad de un Evento

La probabilidad de un evento es la medida de la posibilidad de que dicho evento ocurra. Se calcula utilizando la fórmula:

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

*Ilustración 2 Fórmula probabilidad de un evento*

Donde:

- **Número de resultados favorables** es el número de elementos del espacio muestral que pertenecen al evento.
- **Número total de resultados posibles** es el número de elementos en el espacio muestral total.

## ÁRBOL DE PROBABILIDADES

Un árbol de probabilidades es una herramienta visual que permite representar de manera ordenada los diferentes resultados de un experimento aleatorio y sus probabilidades asociadas. Cada nivel del árbol representa un paso del experimento, y cada rama muestra una posible elección o resultado, junto con su probabilidad.

Esta técnica es particularmente útil cuando se analizan eventos compuestos, ya que facilita el cálculo de probabilidades conjuntas y condicionales.

### Ejemplo: Lanzar una moneda dos veces

Supongamos que lanzamos una moneda dos veces y queremos visualizar todos los posibles resultados:

1. Primer lanzamiento:
  - Cara (C) con una probabilidad de 0.5.
  - Cruz (X) con una probabilidad de 0.5.
2. Segundo lanzamiento:
  - Si el primer lanzamiento fue Cara, en el segundo también puede ser Cara (0.5) o Cruz (0.5).
  - Si el primer lanzamiento fue Cruz, en el segundo también puede ser Cara (0.5) o Cruz (0.5).

El árbol de probabilidades se representaría de la siguiente manera:

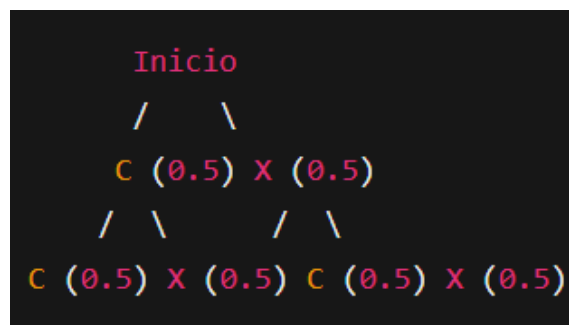



Ilustración 3 Árbol de probabilidades



Los posibles resultados del experimento son:

- (C, C) con una probabilidad de  $0.5 \times 0.5 = 0.25$
- (C, X) con una probabilidad de  $0.5 \times 0.5 = 0.25$
- (X, C) con una probabilidad de  $0.5 \times 0.5 = 0.25$
- (X, X) con una probabilidad de  $0.5 \times 0.5 = 0.25$

El árbol de probabilidades permite ver claramente que la suma de todas las probabilidades es 1, asegurando que todos los resultados posibles están contemplados.

## UNIÓN E INTERSECCIÓN DE EVENTOS

En probabilidad, dos eventos pueden combinarse de distintas maneras. Las dos operaciones más importantes son la unión y la intersección.

### Unión de eventos ( $A \cup B$ )

La unión de dos eventos A y B se refiere al suceso en el que ocurre al menos uno de los dos eventos. Es decir,  $A \cup B$  ocurre si sucede A, sucede B o ambos suceden simultáneamente.

Ejemplo:

Supongamos que extraemos una carta de un mazo de 52 cartas:

A = "Sacar un as"

B = "Sacar una carta de corazones"

El evento  $A \cup B$  significa que la carta extraída es un as o una carta de corazones (o ambas).

La probabilidad de la unión se calcula con la fórmula:

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

*Ilustración 4 Fórmula de probabilidad de unión*

Donde  $P(A \cap B)$  es la probabilidad de que ambos eventos ocurran simultáneamente

### Intersección de eventos ( $A \cap B$ )

La intersección de dos eventos A y B representa la situación en la que ambos eventos ocurren al mismo tiempo. Es decir,  $A \cap B$  ocurre si se cumplen simultáneamente A y B.

Ejemplo:

Siguiendo el ejemplo anterior:

A = "Sacar un as"

B = "Sacar una carta de corazones"

El evento  $A \cap B$  significa que la carta extraída es el **as de corazones**.

En términos de probabilidad, si el mazo tiene 52 cartas:

- Hay 4 ases, por lo que  $P(A) = \frac{4}{52}$
- Hay 13 corazones, por lo que  $P(B) = \frac{13}{52}$
- Solo hay 1 as de corazones, por lo que  $P(A \cap B) = \frac{1}{52}$

Estos conceptos son fundamentales para el cálculo de probabilidades en situaciones más complejas, como en estadística, Machine Learning y toma de decisiones basada en datos.





## MUESTRA Y MUESTREO

En estadística y análisis de datos, la muestra y el muestreo son conceptos clave para realizar estudios eficientes y precisos sin necesidad de examinar a toda la población.

### Muestra

Una muestra es un subconjunto de una población que se selecciona para su análisis con el fin de obtener información y hacer inferencias sobre toda la población. El objetivo es que la muestra sea representativa, es decir, que sus características reflejen fielmente las de la población completa. Esto es crucial para que las conclusiones obtenidas sean precisas y útiles.

Ejemplo:

Si deseamos estudiar las preferencias de los estudiantes de una universidad con 10,000 estudiantes, podríamos tomar una muestra de 500 estudiantes para representar la población general.

### Muestreo

El muestreo es el proceso mediante el cual se selecciona una muestra de una población. Existen diferentes métodos de muestreo, dependiendo de los objetivos del estudio y las características de la población. Entre los tipos de muestreo más comunes están:

- Muestreo aleatorio simple: Cada miembro de la población tiene la misma probabilidad de ser seleccionado.
- Muestreo estratificado: La población se divide en subgrupos homogéneos (estratos) y se selecciona una muestra de cada estrato.
- Muestreo sistemático: Se selecciona un punto de inicio aleatorio y luego se elige cada  $n$ -ésimo miembro de la población.
- Muestreo por conglomerados: Se divide la población en grupos o conglomerados, y luego se selecciona aleatoriamente algunos de estos grupos para ser estudiados.



## **Importancia del muestreo**

El muestreo es esencial porque permite realizar estudios sobre una población sin la necesidad de analizarla en su totalidad, lo cual podría ser costoso y consumir mucho tiempo. Al seleccionar una muestra representativa y aplicar métodos estadísticos apropiados, es posible inferir características de toda la población a partir de los datos obtenidos de la muestra.

Ejemplo:

Si quisiéramos medir la satisfacción de los clientes de una tienda en línea con millones de usuarios, sería inviable encuestar a todos. En su lugar, seleccionaríamos una muestra representativa de clientes y, con base en los resultados, inferiríamos la satisfacción general de todos los clientes.

## **Representatividad de la muestra**

Para que las conclusiones sean válidas, la muestra debe ser representativa de la población. Esto significa que todas las características relevantes de la población deben estar presentes en la muestra en las proporciones adecuadas. Si la muestra es sesgada o no representativa, los resultados pueden ser erróneos o poco precisos.

Ejemplo de problema de representatividad:

Si en una encuesta sobre la preferencia de un producto solo se encuestan a personas jóvenes y se excluye a personas mayores, los resultados estarán sesgados y no representarán la preferencia de toda la población.

Por ello, el diseño del muestreo es una etapa crítica en cualquier investigación, ya que impacta directamente en la validez y fiabilidad de los resultados obtenidos.



## **MUESTRA ALEATORIA Y SESGO DE MUESTRA**

En estadística, la selección de una muestra adecuada es fundamental para obtener resultados confiables. Dos conceptos clave en este proceso son la muestra aleatoria y el sesgo de muestra, los cuales influyen directamente en la validez de los análisis e inferencias realizadas.

### **Muestra Aleatoria**

Una muestra aleatoria es aquella en la que cada individuo de la población tiene la misma probabilidad de ser seleccionado. Este tipo de muestra es fundamental en la estadística, ya que garantiza que los resultados obtenidos sean representativos y no estén influenciados por factores externos.

Características de una muestra aleatoria:

- Todos los elementos de la población tienen la misma posibilidad de ser elegidos.
- Se evita la influencia de factores subjetivos en la selección.
- Reduce la probabilidad de sesgo en los resultados.

Ejemplo:

Si una empresa quiere conocer la satisfacción de sus empleados, en lugar de seleccionar solo a los trabajadores de un departamento específico, se puede hacer un sorteo entre todos los empleados para garantizar una selección imparcial y representativa.

### **Sesgo de Muestra**

El sesgo de muestra ocurre cuando la muestra seleccionada no es representativa de la población, lo que puede llevar a resultados erróneos o distorsionados. Este problema puede surgir de diversas maneras, como la selección inadecuada de la muestra o la falta de diversidad en los participantes.

Causas del sesgo de muestra:

- Muestreo no aleatorio: Cuando la selección de los individuos no es equitativa.
- Falta de representatividad: Cuando ciertos grupos de la población están subrepresentados o sobre-representados en la muestra.

- **Sesgo de respuesta:** Cuando ciertos participantes tienen más probabilidad de responder que otros, por ejemplo, en encuestas voluntarias.

Ejemplo:

Si un estudio sobre hábitos de alimentación solo encuestara a personas en gimnasios, los resultados no serían representativos de toda la población, ya que es probable que las personas que asisten al gimnasio tengan hábitos diferentes a aquellos que no lo hacen.

### **Importancia de evitar el sesgo en el muestreo**

El sesgo en la selección de la muestra puede generar conclusiones incorrectas, lo que afecta la validez de los estudios y decisiones basadas en los datos. Para evitarlo, es crucial aplicar técnicas de muestreo adecuadas, asegurarse de que la muestra sea representativa y verificar que los datos recolectados reflejen la diversidad de la población estudiada.

### **Métodos para reducir el sesgo de muestra:**

- Usar muestreo aleatorio siempre que sea posible.
- Asegurar que la muestra incluya individuos de diferentes características dentro de la población.
- Aumentar el tamaño de la muestra para mejorar su representatividad.
- Aplicar muestreo estratificado, dividiendo la población en subgrupos homogéneos y tomando muestras de cada uno.

Si se minimiza el sesgo y se selecciona correctamente la muestra, los estudios y análisis serán más precisos, permitiendo tomar mejores decisiones basadas en datos confiables.



## **BUENAS PRÁCTICAS PARA OBTENER MUESTRAS REPRESENTATIVAS**

Para garantizar que un estudio estadístico proporcione resultados precisos y generalizables, es crucial utilizar métodos de muestreo adecuados. Una muestra representativa permite inferir características de una población sin necesidad de analizarla en su totalidad. A continuación, se presentan diversas técnicas de muestreo que ayudan a mejorar la calidad de los datos recopilados.

### **Muestreo Aleatorio Simple**

Cada individuo de la población tiene la misma probabilidad de ser seleccionado, lo que reduce el sesgo y asegura una muestra equitativa.

Ejemplo:

Seleccionar 100 nombres al azar de una lista de 1000 personas garantiza que cada individuo tenga la misma oportunidad de ser elegido, haciendo que la muestra sea representativa de la población.

### **Muestreo Aleatorio Estratificado**

Se divide la población en estratos (grupos homogéneos) según características relevantes, y luego se seleccionan muestras proporcionales de cada estrato. Este método es útil cuando se sabe que ciertos subgrupos tienen diferencias significativas.

Ejemplo:

En una universidad, los estudiantes pueden dividirse en facultades (ingeniería, medicina, derecho, etc.). Luego, se selecciona un número proporcional de estudiantes de cada facultad para garantizar que todos los grupos estén representados en el estudio.



### **Muestreo por Conglomerados**

La población se divide en conglomerados (grupos heterogéneos) y se seleccionan algunos conglomerados al azar en lugar de individuos individuales. Este método es útil cuando la población está distribuida en grupos naturales y facilita la recolección de datos.

Ejemplo:

Si se quiere analizar los hábitos de consumo en una ciudad, en lugar de encuestar a todos los habitantes, se pueden seleccionar algunos barrios al azar y encuestar a todas las personas dentro de esos barrios.

### **Muestreo de Varias Etapas**

Este método combina diferentes técnicas de muestreo para optimizar la representatividad de la muestra y facilitar el proceso de selección.


Ejemplo:

En un estudio educativo, se pueden seleccionar escuelas al azar (conglomerados) y luego, dentro de cada escuela seleccionada, elegir a ciertos estudiantes aleatoriamente. Esto permite una selección más eficiente sin necesidad de encuestar a todos los estudiantes de todas las escuelas.

### **Importancia de un Muestreo Adecuado**

La correcta selección de una muestra es fundamental para la validez de un estudio estadístico. Un muestreo inadecuado puede llevar a conclusiones erróneas y decisiones basadas en datos poco representativos. Para evitar estos problemas, es recomendable:

- Usar métodos de muestreo adecuados según la estructura de la población.
- Asegurar que todos los subgrupos relevantes estén representados.
- Minimizar el sesgo mediante técnicas de aleatorización.
- Evaluar el tamaño de la muestra para garantizar resultados significativos.



La probabilidad y la estadística son esenciales en el análisis de datos, y una buena práctica en el muestreo garantiza que las conclusiones extraídas sean válidas y representativas. Aplicar correctamente estos conceptos en la recolección de datos permite obtener información confiable, reducir el margen de error y mejorar la toma de decisiones en diversos campos, como la investigación científica, el marketing y la planificación social.