# Guía de estudio - Estadística descriptiva y probabilidades II

¡Hola! Te damos la bienvenida a esta nueva guía de estudio.

## ¿En qué consiste esta guía?

La siguiente guía de estudio tiene como objetivo practicar y ejercitar los contenidos que hemos visto en clase, además de profundizar temas adicionales que complementan aquellos vistos en la clase.

¡Bienvenidos al mundo de la probabilidad!

**¡Vamos con todo!**

# Tabla de contenidos



## ¡Comencemos!

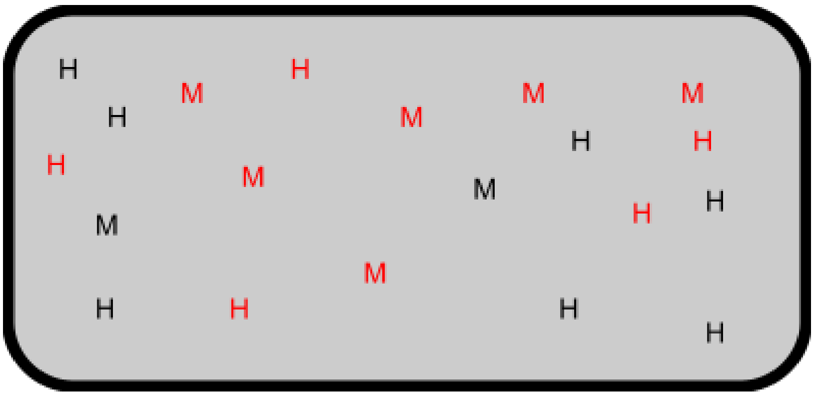
# Probabilidad - definiciones básicas

La probabilidad es una medida matemática que nos permite cuantificar la incertidumbre asociada a un evento. Para modelarla, utilizaremos las siguientes definiciones básicas:

- **Experimento aleatorio**: Es un experimento cuyo resultado no puede anticiparse con total seguridad antes de efectuarlo, es decir, está determinado por el azar.
- **Espacio muestral:** Es el conjunto de todos los posibles resultados de un experimento. Suele anotarse con la letra $\Omega$. Así, en el caso de lanzar una moneda, el espacio muestral sería $\Omega$ = {cara, sello}
- **Evento o suceso:** Es un subconjunto del espacio muestral, es decir, un conjunto de resultados posibles. Se denota comúnmente con letras mayúsculas como "A", "B", etc. Por ejemplo, si consideramos el evento "obtener cara" al lanzar una moneda, sería el conjunto {cara}.
- **Probabilidad:** Es una medida numérica que asigna un valor entre 0 y 1 a un evento. Representa la posibilidad de que un evento ocurra. Un evento con probabilidad 0 significa que es imposible, mientras que un evento con probabilidad 1 significa que es seguro que ocurra.
- **Probabilidad clásica:** Es un enfoque de la probabilidad que se basa en la suposición de que todos los resultados en el espacio muestral son igualmente probables. Por ejemplo, al lanzar un dado justo de seis caras, cada número tiene una probabilidad de 1/6.
- **Probabilidad frecuencial**: Es un enfoque de la probabilidad que se basa en observaciones empíricas o datos recopilados. Se calcula dividiendo el número de veces que ocurre un evento entre el número total de repeticiones del experimento.

## Probabilidades - Cálculo y operaciones

A partir de las definiciones anteriores, podemos extender algunas y realizar deducciones. Para ello, considera una sala en la que hay hombres y mujeres (que anotaremos con las letras H y M). Las personas de esta sala que son bilingües las señalaremos con color rojo, mientras que las que no lo son las señalaremos con color negro.




Fuente: Desafío Latam

Si consideramos el experimento de extraer una persona al azar de esta sala, podemos definir algunos sucesos y su probabilidad. Por ejemplo:

$$B: \text{ser bilingüe} \rightarrow P(B) = 11/20$$

Podemos definir el **complemento** de B como el suceso "no ser bilingüe". Observamos que:

$$P(B^C) = P(\overline{B}) = 1 - P(B)$$

Definimos la **intersección** de dos sucesos a la ocurrencia de uno **y** el otro. Por ejemplo, si consideramos

$$A: \text{ser hombre}$$

tenemos que

$$A \text{ y } B = A \cap B : \text{ser hombre y bilingüe}$$

En nuestro caso, podemos identificar los casos favorables contando la cantidad de letras H **en color rojo**. Tenemos así que

$$P(A \cap B) = 4/20 = 1/5$$

Definimos la **unión** de dos sucesos a la ocurrencia de uno **o** el otro. En nuestro ejemplo anterior:

$$A \text{ o } B = A \cup B : \text{ser hombre o bilingüe}$$

Podemos identificar los casos favorables contando la cantidad de letras que son H, **o bien que son de color rojo**. Tenemos así que

$$P(A \cup B) = 18/20 = 9/10$$

Observa que:

- los casos favorables al suceso "ser hombre" son 12
- los casos favorables al suceso "ser bilingüe" son 11

Podríamos tener la tentación de sumar directamente las probabilidades, y tendríamos un total de 11 + 12 = 23 casos favorables, por lo que la probabilidad buscada sería 23 / 20 > 1 ¿Qué error hemos cometido?

Evidentemente, que hemos contado **dos veces** a los **hombres bilingües**, es decir, la intersección de estos dos sucesos. Podemos entonces deducir que, si sumamos las probabilidades, debemos **restar** la probabilidad de la intersección para no contarla dos veces.

Tenemos así un importante resultado:

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

Si los sucesos no tienen casos en común, decimos que son **disjuntos**, por lo que la probabilidad de la intersección es igual a cero… ¡y simplemente, sumamos las probabilidades!

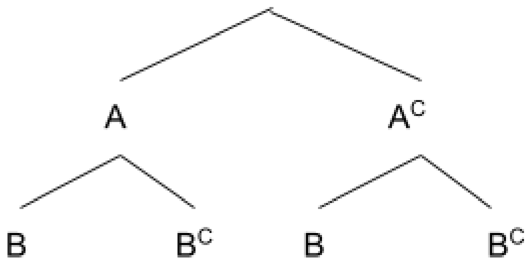
## Probabilidad condicional

Si al levantarnos vemos por la ventana que en el cielo hay nubes negras, podemos pensar que es probable que llueva… ¡al menos, más probable que si lo que vemos es un sol esplendoroso! Es posible que luego la situación cambie: las nubes se disipen o, por el contrario, el sol desaparezca, vengan nubes y la lluvia.

Si consideramos un día al azar, podemos definir los siguientes sucesos:

A: amanece nublado
B: llueve durante la mañana

Se llama **probabilidad de B, dado A**, a la probabilidad de que el suceso B ocurra **si se sabe que ha ocurrido el suceso A**. Lo anotamos $P(B|A)$. Observa que, para un día cualquiera, puede amanecer nublado o no, y podemos considerar a partir de esto que llueve por la mañana o no. Representaremos esto en el siguiente diagrama de árbol.




Vamos a suponer que conocemos la probabilidad de que un día escogido al azar amanezca nublado o no, Además, que conocemos la probabilidad de que llueva si amanece nublado, y de que llueva si no amanece nublado.

$$P(A) = 0{,}7 \rightarrow P(A^C) = 0{,}3$$
$$P(B/A) = 0{,}8$$
$$P(B/A^C) = 0{,}1$$

A partir de estos valores, podemos deducir que:

$$P(B^C / A) = 1 - 0{,}8 = 0{,}2$$
$$P(B^C / A^C) = 1 - 0{,}1 = 0{,}9$$

## Ley de probabilidad total

Si escogemos un día del año, al azar, ¿cuál es la probabilidad de que llueva? Para determinarlo debemos considerar, por un lado, la probabilidad de que amanezca nublado **y** llueva, **o bien** que no amanezca nublado y llueva. La **ley de probabilidad total** nos indica que:

$$P(B) = P(A) \cdot P(B/A) + P(A^C) \cdot P(B/A^C)$$

En nuestro caso, tenemos que:

$$P(B) = 0,7 \cdot 0,8 + 0,3 \cdot 0,1$$

= 0,56 + 0,03

= 0,59

**Importante**: en general, para calcular la probabilidad de que suceda algo, debemos considerar todas las posibles combinaciones de condiciones.

## Dependencia e Independencia

Hemos calculado la probabilidad de que un día amanezca nublado **y** llueva multiplicando la probabilidad de que amanezca nublado (P(A)) por la probabilidad de que llueva dado que amaneció nublado (P(B / A)). Podríamos decir entonces que:

$$P(A\, y\, B) = P(A \cap B) = P(A) \cdot P(B/A)$$

Hemos definido el suceso A como "amanece nublado". ¿Qué ocurriría si lo hubiéramos definido como "es jueves"? Intuitivamente, que sea jueves o no, no debería influir en la probabilidad de que llueva o no.

Decimos que dos sucesos son **independientes** si **la ocurrencia de uno no afecta la probabilidad de ocurrencia del otro** (en caso contrario, son **dependientes**). Observa que si los sucesos son independientes, evidentemente tendremos que

P(B) = P(B/A)

Y con ello:

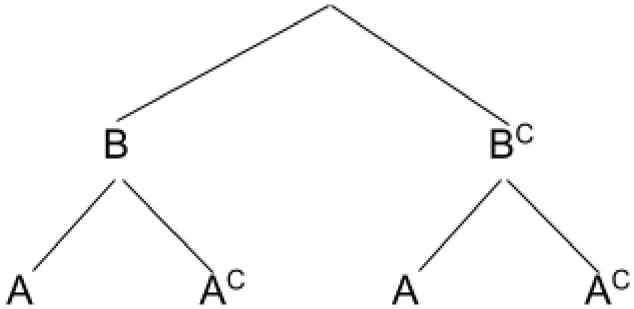
$$P(A\, y\, B) = P(A \cap B) = P(A) \cdot P(B/A)$$

$$P(A\, y\, B) = P(A \cap B) = P(A) \cdot P(B)$$

## Teorema de Bayes

Podemos plantearnos también la pregunta inversa: si un día escogido al azar llueve, ¿cuál es la probabilidad de que haya amanecido nublado?

Expresándolo de la forma en que hemos venido haciéndolo, equivale a la **probabilidad de A dado B**. Podemos, entonces, invertir el planteamiento de este problema y representarlo en el siguiente diagrama de árbol




Podemos ver que la probabilidad de que llueva y esté nublado corresponderá, entonces, a

$$P(AyB) = P(B) \cdot P(A/B)$$
$$\Rightarrow P(A/B) = \frac{P(AyB)}{P(B)}$$

Lo que conocemos como **Teorema de Bayes**

De nuestros cálculos anteriores sabemos que

$$P(AyB) = P(A) \cdot P(B/A)$$
$$= 0,7 \cdot 0,8$$
$$= 0,56$$

$$P(B) = 0,59$$

Por lo tanto:

$$P(A/B) = \frac{P(AyB)}{P(B)}$$
$$P(A/B) = \frac{0,56}{0,59}$$
$$P(A/B) = 0,949152542$$

Hemos calculado así la probabilidad de que, **si un día llueve, ese día haya amanecido nublado**. El teorema de Bayes tiene aplicaciones en una amplia gama de campos, incluyendo la estadística, la inteligencia artificial, la medicina, la economía y muchas otras disciplinas donde se necesita analizar y evaluar la probabilidad de eventos en función de la información disponible.

## Actividad guiada: Cabellos y probabilidades

Supongamos que en una clase de estudiantes, se sabe que el 60% de los estudiantes son mujeres y el 40% son hombres. También se sabe que el 70% de las mujeres tienen cabello largo, mientras que el 30% de los hombres tienen cabello largo. Si seleccionamos al azar a un estudiante de la clase y se sabe que tiene cabello largo, ¿cuál es la probabilidad de que sea mujer?

Para resolver este problema, vamos a utilizar la probabilidad condicional. Definimos los eventos de la siguiente manera:

A: El estudiante seleccionado es una mujer.
B: El estudiante seleccionado tiene cabello largo.

Se nos pide calcular P(A|B), es decir, la probabilidad de que el estudiante sea mujer dado que tiene cabello largo.

La probabilidad de que una persona sea mujer (A) es del 60% (0,6) y la probabilidad de que una persona tenga cabello largo (B) es la suma de las probabilidades de que una mujer tenga cabello largo (0,6 * 0,7) y de que un hombre tenga cabello largo (0,4 * 0,3), que es igual a 0,42 + 0,12 = 0,54.

Por lo tanto, utilizando la fórmula de la probabilidad condicional, tenemos:

$$P(A|B) = P(A \cap B) / P(B)$$
$$P(A|B) = (0{,}6 * 0.7) / 0{,}54$$
$$P(A|B) \approx 0{,}7778$$

En consecuencia, la probabilidad de que el estudiante seleccionado sea una mujer dado que tiene cabello largo es aproximadamente 0,7778 o 77,78%.

## ¡Manos a la obra! - Probabilidad y teorema de Bayes

Resuelve las siguientes situaciones:

1. Se estima que el 13% de la población adulta de un país padece de una enfermedad, pero que el 75% de todos los adultos creen no tener este problema. Se estima también que el 4% de la población tiene la enfermedad, aunque no lo sabe. Si una persona adulta al azar piensa que está sana, ¿cuál es la probabilidad de que realmente esté enferma?
2. Se realiza un test para detectar la presencia de un tipo de bacterias en el agua. Si las bacterias están presentes en el agua, el test tiene una probabilidad de 0,9 de

detectarlas; si las bacterias no están presentes, el test tiene una probabilidad de 0.8 de indicar su ausencia. Se sabe además que la probabilidad de que una muestra de agua contenga bacterias de ese tipo T es de 0,2.

a. ¿Cuál es la probabilidad de que realmente haya presencia de bacterias cuando el test ha dado resultado positivo?
b. ¿Cuál es la probabilidad de que realmente haya presencia de bacterias cuando el test ha dado resultado negativo?
c. ¿Cuál es la probabilidad de que haya bacterias y además el test dé positivo?

# Probabilidades con Python

Para calcular probabilidades con Python, más que hacerlo directamente utilizaremos funciones de conteo de casos, considerando los filtros que podemos utilizar para las DataFrames.

## Conteo de casos en una columna

Podemos contar la cantidad de veces que aparece **'Valor1"** en la columna **"columna1"** de nuestro DataFrame **df** utilizando el siguiente comando:

```
conteo = df['columna1'].value_counts()['Valor1']
```

Si escribimos el siguiente comando

```
count_group = df['columna1'].value_counts()
```

obtendremos de igual manera un conteo de cada registro de la columna

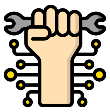
## Conteo con condiciones

Podemos definir condiciones considerando valores y columnas. Por ejemplo, nuestra **condicion1** corresponderá a que en la **columna1** aparezca **valor1**, y nuestra **condicion2** corresponderá a que en la **columna2** aparezca **valor2**

```
condicion1 = df['columna1'] == 'valor1'
condicion2 = df['columna2'] =='valor2'
```

Lo anterior nos permitirá contar los casos en que se cumplen las condiciones simultáneamente (intersección), o bien una o la otra (unión)

```
conteo_interseccion = (condicion1 & condicion2).sum()
conteo_union = (condicion2 | condicion3).sum()
```



## ¡Manos a la obra! - Calculando con Python

Utilizando el dataset **exams.csv**, calcula las siguientes probabilidades al seleccionar a un estudiante al azar

1. Que su grupo étnico sea A o B
2. Que su preparación para el examen sea completa
3. Que tenga un puntaje en matemáticas superior o igual a 50
4. Que el nivel educacional de sus padres sea ‘high school”
5. Que el nivel educacional de sus padres sea ‘high school”, si se sabe que pertenece al grupo étnico A o B
6. Que su preparación para el examen sea completa, si se sabe que su puntaje en matemáticas es superior o igual a 50

## Preguntas de cierre

- ¿Cómo puedo aplicar el concepto de probabilidad en mi vida cotidiana? ¿Hay situaciones en las que tomar decisiones basadas en la probabilidad podría ser útil?
- ¿Puedes pensar en ejemplos adicionales de eventos independientes y dependientes en tu entorno personal o profesional? ¿Cómo influye la dependencia de eventos en la toma de decisiones o en la planificación?
- ¿Qué otros nombres o términos se utilizan comúnmente para describir la intersección de eventos? ¿Cómo se diferencian estos términos y en qué contextos se aplican?
- ¿Cuáles son algunos ejemplos de problemas reales en los que se pueda utilizar el teorema de Bayes? ¿Cómo podría aplicarse este teorema para mejorar la toma de decisiones basada en evidencia?
- ¿Hay casos en los que la probabilidad condicional o el teorema de Bayes puedan ser malinterpretados o aplicados incorrectamente? ¿Cómo se pueden evitar estos errores?
- ¿En qué áreas de tu vida o trabajo podrías utilizar la probabilidad y los conceptos relacionados para mejorar la toma de decisiones, realizar análisis más precisos o comprender mejor los fenómenos que te rodean?
- ¿Qué ejemplos prácticos puedes encontrar en los que la independencia o dependencia de eventos afecten directamente los resultados o las probabilidades asociadas a un evento en particular?

- ¿Cuáles son algunas áreas de investigación o disciplinas científicas en las que la probabilidad y la teoría de eventos desempeñan un papel fundamental? ¿Cómo se aplican estos conceptos en esos campos?
- ¿De qué manera el conocimiento de la probabilidad y los conceptos relacionados pueden ayudarte a evaluar críticamente información, datos o afirmaciones que encuentres en tu vida diaria o en los medios de comunicación?