Estadística descriptiva y probabilidades (Parte II)

Clase sincrónica

{desafío} latam_

Aplica herramientas de Python para calcular indicadores estadísticos y probabilidades.

- **Unidad 1: Estadística descriptiva y probabilidades**
  (Parte I)

  (Parte II)

  Te encuentras aquí

- **Unidad 2: Variable aleatoria**
  (Parte I)

  (Parte II)

- **Unidad 3:Estadística inferencial**

- **Unidad 4: Regresión**
  (Parte I)

  (Parte II)

{desafío} latam_

¿Qué aprenderás en esta sesión?

Las definiciones de básicas de probabilidad: sucesos, eventos, definición, probabilidad condicional, dependencia e independencia, y aplicar herramientas de Python para el análisis de estas.

¿Cuál crees que es la probabilidad de que llueva mañana?

¿Sabrías dar un grado de certeza a dicha afirmación?

/* Definiciones de probabilidad */

Introducción a la probabilidad

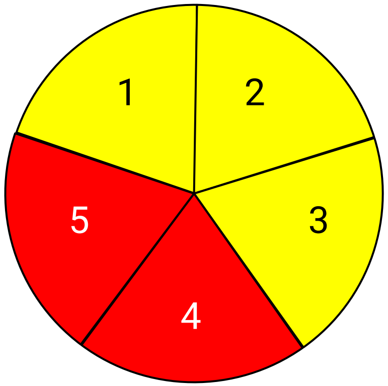
Definiciones y ejemplos

01	**Experimento aleatorio**	experimento en el que influye el azar. Es decir, no es posible determinar a priori un resultado en particular.
02	**Espacio muestral**	conjunto de posibles resultados individuales de un experimento aleatorio.
03	**Suceso o evento**	subconjunto del espacio muestral.
04	**Probabilidad de un suceso**	corresponde al cociente entre la cardinalidad del suceso, y la cardinalidad del espacio muestral

Introducción a la probabilidad
Definiciones y ejemplos

Consideremos una ruleta como la que se muestra. Podemos definir algunos sucesos:

A. que salga el 3
B. que salga un número menor o igual que 5
C. que salga un número mayor que 10
D. que salga un número par
E. que salga un número en rojo
F. que salga un número par o en rojo
G. que no salga el 2



¿Cuál es la probabilidad de cada uno?

Introducción a la probabilidad
Definiciones y ejemplos

- La probabilidad de un suceso es un número entre 0 y 1.
- Un **suceso seguro** tiene probabilidad igual a 1; un **suceso imposible** tiene probabilidad 0
- Se define la **unión** de dos sucesos como la ocurrencia de uno **o** el otro. Podemos calcular su probabilidad como:

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

  Si los sucesos son **disjuntos**, la probabilidad de la intersección es igual a cero y pueden simplemente sumarse.
- Se llama **complemento** de un suceso a todo lo que no pertenece a él (“no A”). Se tiene que:

$$P\left(\overline{A}\right) = P\left(A^C\right) = 1 - P\left(A\right)$$

/*Probabilidad teórica y experimental*/

{desafío}
latam_

Probabilidad teórica y experimental

¿Cómo calculamos?

- ¿Es más probable tener un accidente de tránsito en automóvil particular o en bus?
- En un día de invierno, ¿es más probable que llueva en una ciudad o en otra?

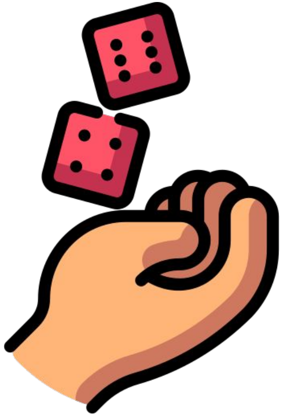
Naturalmente, no podemos “calcular” la probabilidad de la misma manera como lo hacemos con un dado. La alternativa es utilizar la estadística.

Probabilidad teórica y experimental

Diferencias

Utilizamos la **probabilidad teórica** cuando conocemos exactamente un experimento, es decir, conocemos perfectamente su espacio muestral.

En ocasiones, solo podemos tener un número limitado de registros y sin posibilidad de saber si son todos. Empleamos, en este caso, la **probabilidad experimental** o **frecuentista**.



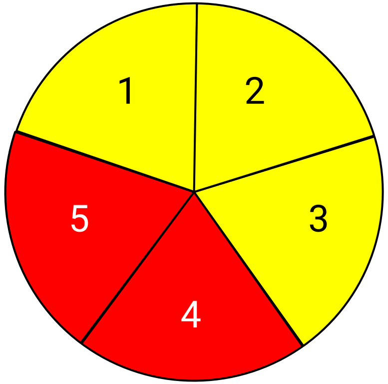
/*Probabilidad condicional*/

{desafío} latam_

Probabilidad condicional

Consideremos nuevamente nuestra ruleta, y el experimento correspondiente a lanzarla dos veces seguidas y anotar el color en cada ocasión.

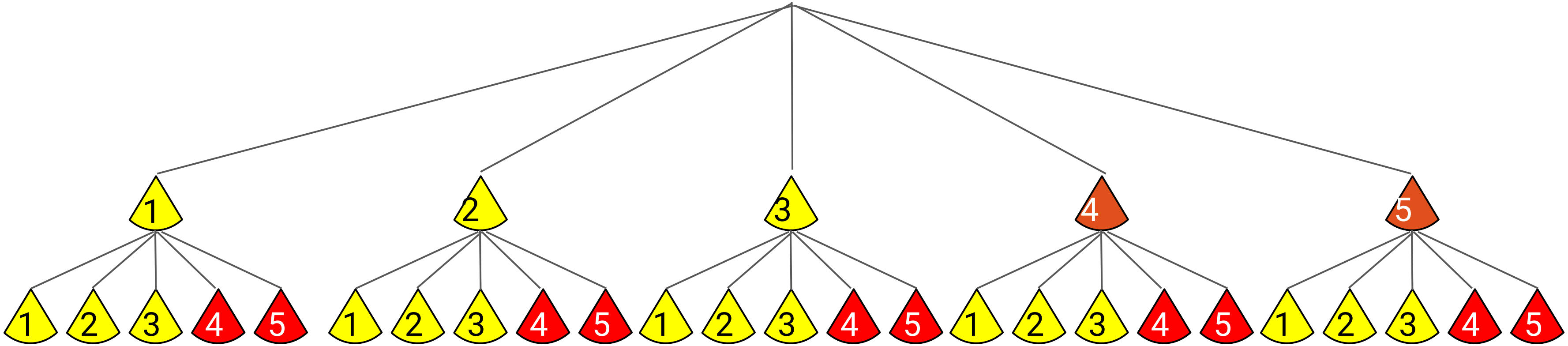
Vamos a representar este experimento en un **diagrama de árbol**: a partir de un punto inicial se establecen **ramas**, que finalizan en **nodos** que corresponden a resultados posibles.



{desafío} latam_

Probabilidad condicional

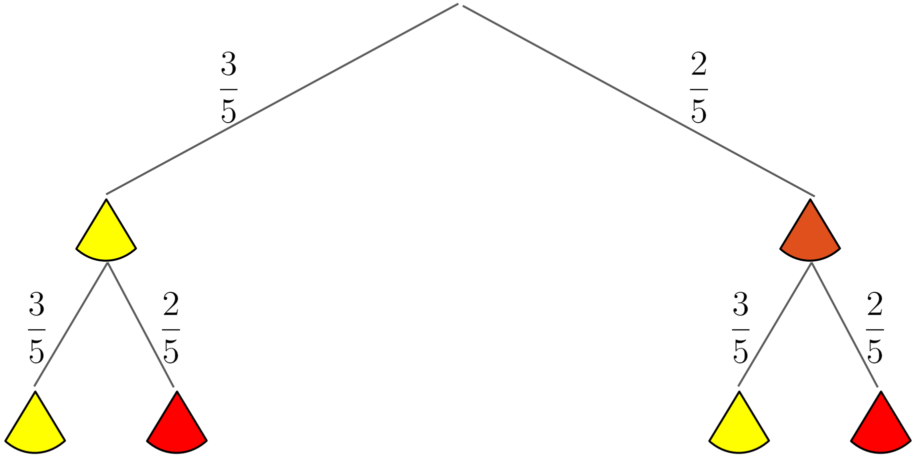
Diagrama de árbol



- ¿Cuántos casos totales hay? ¿Cómo se pueden calcular?
- ¿Qué combinaciones de colores hay? ¿Cuántos casos favorables tiene cada una? ¿Cómo se pueden calcular?

Probabilidad condicional

Diagrama de árbol simplificado



$$P\left(AyA\right)=\frac{3}{5}\cdot\frac{3}{5}=\frac{9}{25}$$

$$P\left(AyR\right)=\frac{3}{5}\cdot\frac{2}{5}=\frac{6}{25}$$

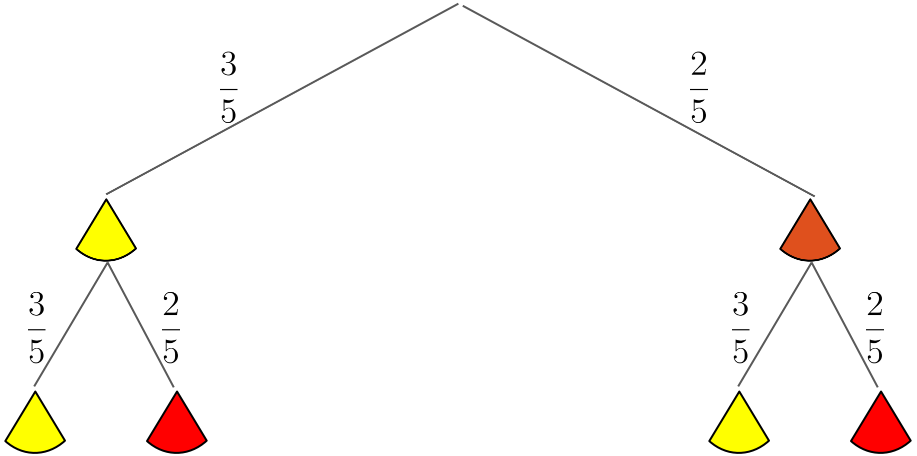
$$P\left(RyA\right)=\frac{2}{5}\cdot\frac{3}{5}=\frac{6}{25}$$

$$P\left(RyR\right)=\frac{2}{5}\cdot\frac{2}{5}=\frac{4}{25}$$

{desafío} latam_

Probabilidad condicional

Diagrama de árbol simplificado



- ¿Cuál es la probabilidad de que salga un número en amarillo?
- ¿Cuál es la probabilidad de que salga un número en amarillo en la segunda tirada, si se sabe que en la primera salió amarillo?

Probabilidad condicional

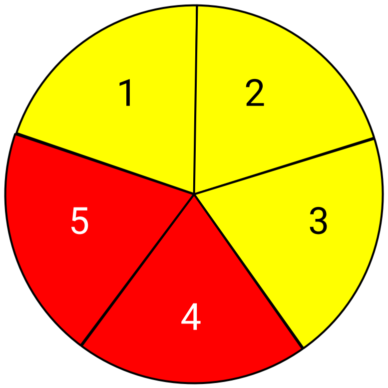
Relacionando sucesos

Consideremos nuevamente nuestra ruleta. La lanzamos una vez y observamos el color, y si el número es par o impar. Observa que:

P(A) = 3/5 P(R) = 2/5 P(P) = 2/5 P(I) = 3/5

P(A y P) = 1/5   P(A y I) = 2/5

P(R y P) = 1/5   P(A y I) =1/5



{desafío} latam_

Dependencia e independencia

Relacionando sucesos

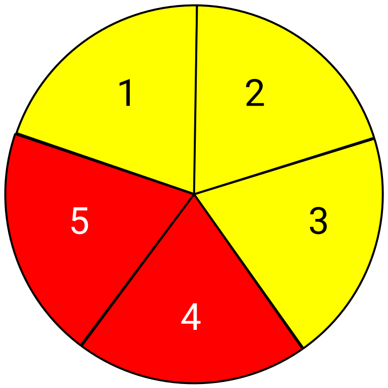
$P(A) = 3/5$ $P(R) = 2/5$ $P(P) = 2/5$ $P(I) = 3/5$

$P(A \text{ y } P) = 1/5$ $P(A \text{ y } I) = 2/5$

$P(R \text{ y } P) = 1/5$ $P(A \text{ y } I) = 1/5$

Si sabemos que salió un número del sector amarillo, ¿cuál es la probabilidad de que sea un número par?

$P(A \text{ y } P) = 1/5, P(A) = 3/5 \rightarrow P(P / A) = 1/3$

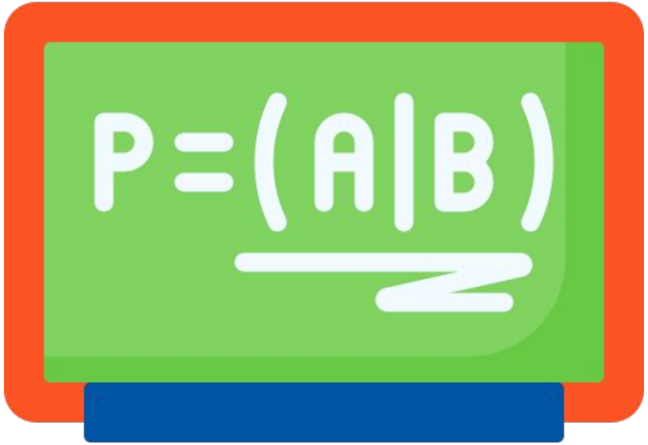


Probabilidad condicionada

Definición

Dados dos sucesos A y B, se llama **probabilidad condicionada de A, dado B**, a la probabilidad de ocurrencia del suceso A si se sabe que ya ha ocurrido el suceso B. Se anota P(A / B), y tenemos que:

$$P\left(A|B\right) = \frac{P\left(AyB\right)}{P\left(B\right)}$$



Probabilidad condicionada

Dependencia e independencia

Considerando la fórmula anterior, podemos reescribir como:

$$P(A / B) * P(B) = P(A \text{ y } B)$$

- Decimos que A es **independiente** de B si la ocurrencia de B no modifica la probabilidad de A, o no influye en ella. Por ende, $P(A / B) = PA$ y con ello $P(A \text{ y } B) = P(A) * P(B)$
- En caso contrario, A y B son **dependientes** entre sí.

Probabilidad condicionada

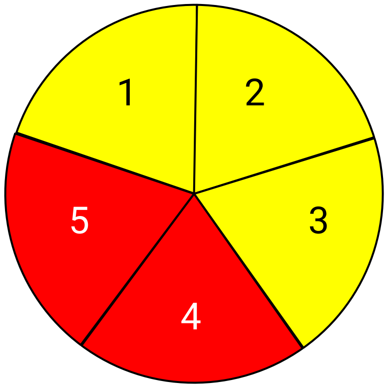
Caso inverso

Podemos hacernos ahora la pregunta inversa: si se sabe que el número que salió es par, ¿cuál es la probabilidad de que sea en un sector amarillo?

P(A) = 3/5 P(R) = 2/5 P(P) = 2/5 P(I) = 3/5

P(A y P) = 1/5 P(A y I) = 2/5 P(R y P) = 1/5 P(A y I) = 1/5

P(A y P) = 1/5 P(P) = 2/5



Probabilidad condicionada

Caso inverso

Podemos hacernos ahora la pregunta inversa: si se sabe que el número que salió es par, ¿cuál es la probabilidad de que sea en un sector amarillo?

P(A y P) = 1/5 P(P) = 2/5

$$P\left(A|P\right) = \frac{P\left(AyP\right)}{P\left(P\right)}$$

$$= \frac{\frac{1}{5}}{\frac{2}{5}}$$

$$= \frac{1}{2}$$

/* Teorema de Bayes*/

Teorema de Bayes

Probabilidad "a posteriori"

Supongamos que tenemos una prueba médica para detectar una enfermedad. De acuerdo a estudios, la probabilidad de que la prueba dé resultado positivo al ser aplicada a una persona enferma es del 90%, y del 5% si se aplica a personas sanas. Por otra parte, la probabilidad de que un individuo seleccionado al azar tenga la enfermedad es del 3%

Teorema de Bayes
Probabilidad "a posteriori"

- la probabilidad de que la prueba dé resultado positivo al ser aplicada a una persona enferma es del 90%, y del 5% si se aplica a personas sanas.
- la probabilidad de que un individuo seleccionado al azar tenga la enfermedad es del 3%

Definimos los sucesos

A: estar enfermo — A^C: estar sano

B: test positivo — B^C: test negativo

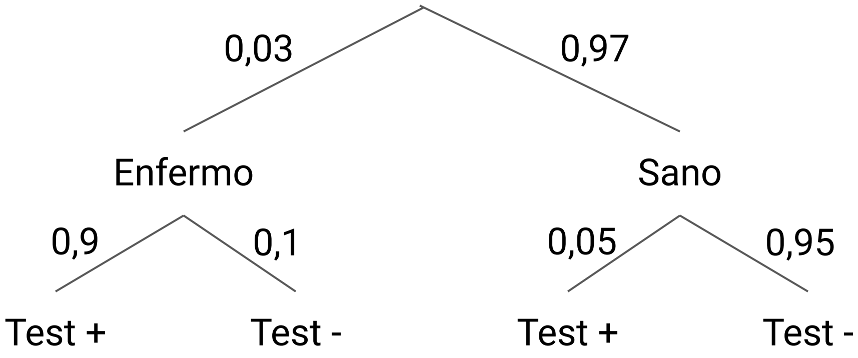
y tenemos:

$$P(B \mid A) = 0,9$$
$$P(B \mid A^C) = 0,05$$
$$P(A) = 0,97 \qquad P(A^C) = 0,03$$

{desafío} latam_

Teorema de Bayes

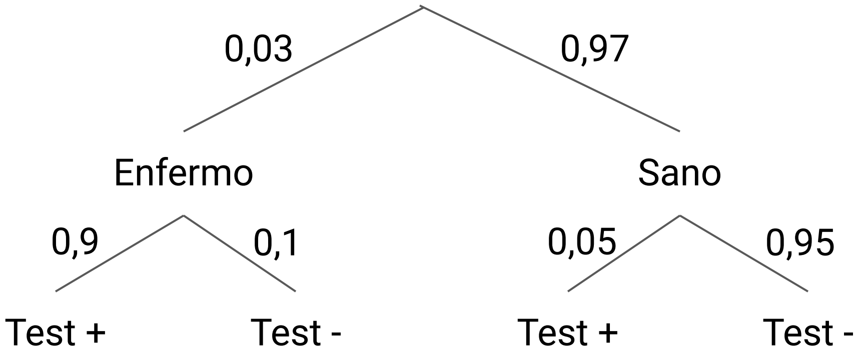
Falsos positivos y negativos



- **Falso positivo:** Si un paciente da positivo en el test, ¿cuál es la probabilidad de que esté sano, realmente? → $P(A^C / B)$
- **Falso negativo:** Si un paciente da negativo en el test, ¿cuál es la probabilidad de que esté enfermo, realmente? →$P(A / B^C)$

Teorema de Bayes

Falsos positivos y negativos



$P(B \text{ y } A) = 0{,}03 * 0{,}9$
$= 0{,}027$

$P(B^C \text{ y } A) = 0{,}03 * 0{,}1$
$= 0{,}003$

$P(B \text{ y } A^C) = 0{,}05 * 0{,}97$
$= 0{,}0485$

$P(B^C \text{ y } A^C) = 0{,}95 * 0{,}97$
$= 0{,}9215$

$P(B) = P(B \text{ y } A) + P(B \text{ y } A^C)$
$= 0{,}03 * 0{,}9 + 0{,}05 * 0{,}97$
$= 0{,}027 + 0{,}0485$
$= 0{,}0755$

Teorema de Bayes

$P(B \mid A) = 0{,}9$ $\quad$ $P(B \mid A^C) = 0{,}05$

$P(A) = 0{,}97$ $\quad$ $P(A^C) = 0{,}03$

$P(B) = 0{,}0755$ $\quad$ $P(B^C) = 0{,}9245$

$P(A \text{ y } B) = 0{,}027$ $\quad$ $P(B^C \text{ y } A) = 0{,}003$ $\quad$ $P(B \text{ y } A^C) = 0{,}0485$ $\quad$ $P(B^C \text{ y } A^C) = 0{,}9215$

$P(B) = 0{,}0755$ $\quad$ $P(B^C) = 0{,}9245$

- **Falso positivo:**

$$P(A^C / B) = P(A^C \text{ y } B)/P(B)$$
$$= 0{,}0485 / 0{,}0755$$
$$= 0{,}6424$$

- **Falso negativo:**

$$P(A / B^C) = P(A \text{ y } B^C)/P(B^C)$$
$$= 0{,}003 / 0{,}9245$$
$$= 0{,}003244997$$

¿Qué es más grave: un falso positivo, o un falso negativo?

El falso positivo puede ser verificado (de hecho, lo es), mientras que un falso negativo es una situación de evidente riesgo.

Desafío
Estadística descriptiva y probabilidades (Parte II)

{desafío}
latam_

Desafío

“Estadística descriptiva y probabilidades (parte II)”

- Descarga el archivo “Desafío”.
- Tiempo de desarrollo asincrónico: desde 2 horas.
- Tipo de desafío: individual.

¡AHORA TE TOCA A TI! 💪

Ideas fuerza

*La **probabilidad** nos permite cuantificar la posibilidad **ocurrencia de un suceso,** es decir, **el resultado de un experimento aleatorio***

*Podemos **calcular la probabilidad** de un suceso dividiendo la cantidad de casos favorables por la cantidad de casos posibles totales del experimento, lo que se conoce como **Regla de Laplace***

*Llamamos **probabilidad condicional** de un suceso a la probabilidad de que este ocurra, si e sabe que además ocurre otro. Si las probabilidades no se modifican, los sucesos son **independientes.***

*El **teorema de Bayes** nos permite calcular la probabilidad de un suceso **a posteriori,** es decir, la probabilidad de que haya ocurrido si se conoce otro resultado posterior a él.*

{desafío} latam_

¿En qué otras áreas puede utilizarse la probabilidad a posteriori?

Recursos asincrónicos

¡No olvides revisarlos!

Para esta semana deberás revisar:

- Guía de estudio.
- Desafío “Estadística descriptiva y probabilidades (Parte II)”.

Próxima sesión...

- *Variables Aleatorias, Discretas y Continuas.*
- *Ley de los grandes números.*

{desafío}
latam_

{desafío}
latam_

Academia de
talentos digitales