Estadística descriptiva y probabilidades

Tutoría - Parte II

{desafío}
latam_

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

¿Sabrías dar un grado de certeza a dicha afirmación?

{desafío} latam_

Autoaprendizaje
Recursos asincrónicos

- ¿Revisaste los recursos de la semana 1 (Guía y desafío)?
- ¿Tienes dudas sobre alguno de ellos?

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.*

/* Definiciones de probabilidad */

{desafío}
latam_

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

- 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:

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

{desafío} latam_

/*Probabilidad teórica y experimental*/

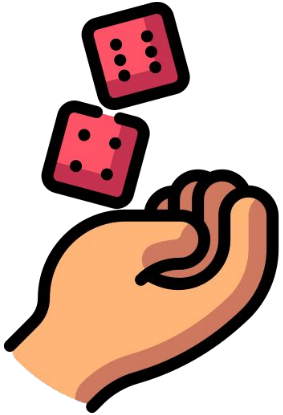
{desafío}
latam_

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**.



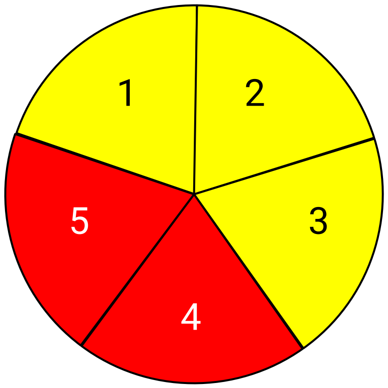
{desafío} latam_

/*Probabilidad condicional*/

Probabilidad condicional

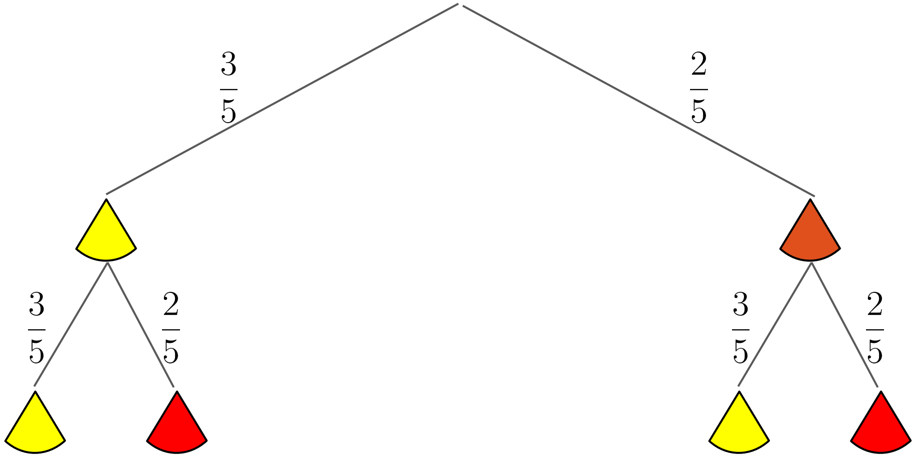
Consideremos nuevamente nuestra ruleta, y el experimento correspondiente a lanzarla dos veces seguidas y anotar el color en cada ocasión.

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.



Probabilidad condicional

Diagrama de árbol simplificado



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

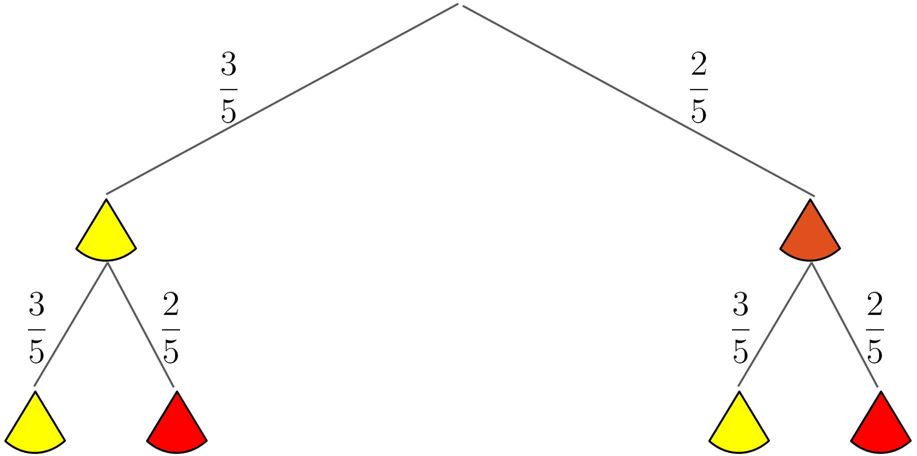
$$P\left(AyR\right) = \frac{3}{5}\cdot\frac{2}{5} = \frac{6}{25}$$

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

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

Probabilidad condicional

Diagrama de árbol simplificado



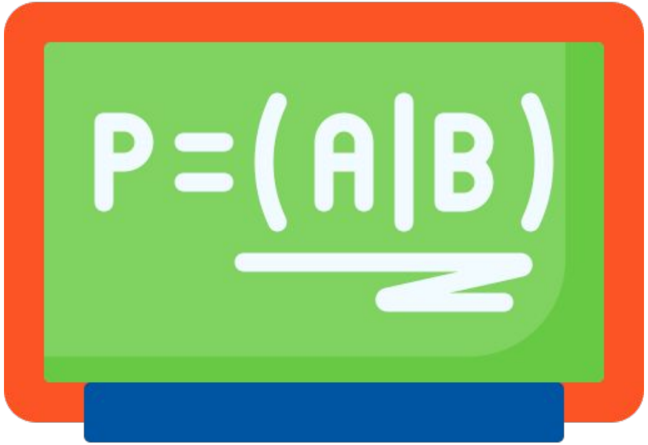
- ¿Cuál es la probabilidad de que salga un número en amarillo?
- ¿Cuál es la probabilidad de que salga un número en amarillo en la segunda tirada, si se sabe que en la primera salió amarillo?

Probabilidad 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:

$$P(A|B) = \frac{P(A y B)}{P(B)}$$



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

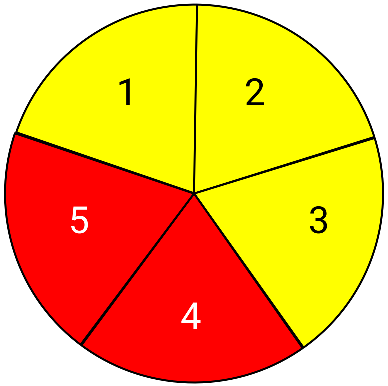
Caso inverso

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



{desafío} latam_

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*/

{desafío}
latam_

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

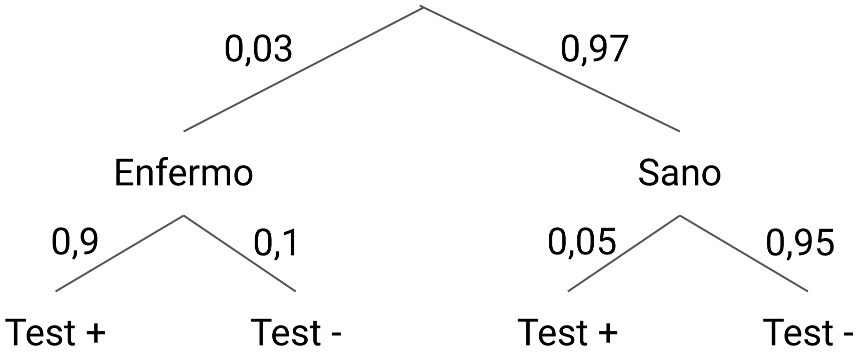
B: test positivo — B^C: test negativo

y tenemos:

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

Teorema de Bayes

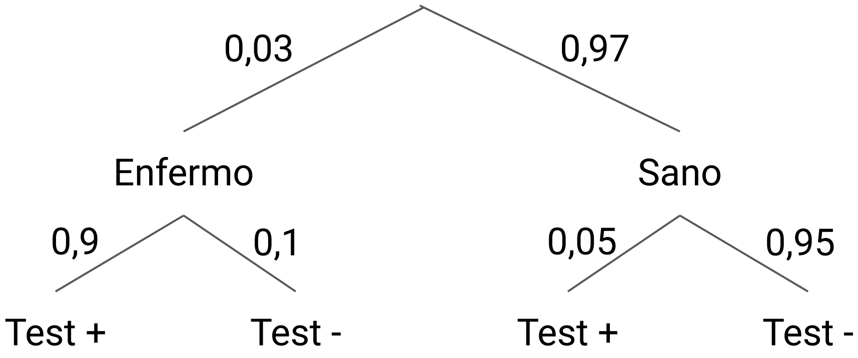
Falsos positivos y negativos



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

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

{desafío} latam_

¿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.

¡Manos a la obra! - Probabilidades con Python

{desafío} latam_

Probabilidades con Python

Vamos a calcular algunas probabilidades utilizando datos de DataFrames, en Python Para esto, abre tu propio archivo de Jupyter y sigue las instrucciones de tu profesor. A continuación, aprenderemos:

1. Cálculo de probabilidades con Pyhton - caso general
2. Probabilidad de sucesos compuestos
3. Probabilidad condicional

Desafío
Estadística descriptiva y probabilidades (Parte II)

Desafío
“Estadística descriptiva y probabilidades (parte II)”

- ¿Leíste el desafío de esta semana? ¿Comprendes bien lo que se solicita en cada caso?
- ¿Hay contenidos que necesitas repasar antes de comenzar este desafío?
- ¿Necesitas algún ejemplo o indicación para alguna pregunta o requerimiento específico?

Próxima sesión…

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

{desafío}
latam_

{desafío} latam_

Academia de talentos digitales