

ACÀMICA

TEMA DEL DÍA

Probabilidad

Seguimos en el mundo de la Probabilidad y Estadística. Hoy, dos conceptos muy importantes: Distribución de Probabilidad y Correlación.



Agenda

Daily

Explicación: Probabilidad y distribución

Break

Hands-on training

Bonus Track: Relación entre Probabilidad y Estadística.

Cierre.



Daily



Daily



Sincronizando...

Toolbox



¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

Challenge



¿Cómo te ha ido?
¿Obstáculos?
¿Cómo seguimos?

Variables



Probabilidad: Variables aleatorias

PROBABILIDAD

Variables **discretas**

- Son aquellas que se *cuentan*
- Pueden estar acotadas o no

Ejemplo: cantidad de países, número de hijos, cantidad de dormitorios en una casa, etc.

Variables **continuas**

- Son aquellas que se *miden*
- Pueden estar acotadas o no

Ejemplo: altura de una persona, temperaturas, edades (medidas en tiempo transcurrido desde el nacimiento), etc.



La edad: ¿es una variable discreta o continua?



La edad: ¿es una variable discreta o continua?

DEPENDE



Veámoslo gráficamente:

Edad contada en años (variable **discreta**):

Edades: [1, 2, 3, 4, 5]



Edad contada en tiempo transcurrido (variable **continua**):

Edades: [1 a 5]



DISTRIBUCIONES



Variables discretas: Distribución

La distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra.



Variables discretas: Distribución **uniforme**

La distribución de probabilidad **uniforme** asigna la misma probabilidad para todo un rango de valores.

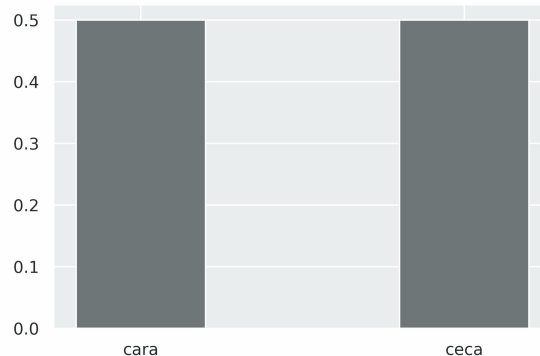
Ejemplos: moneda, dado.

$$X_{\text{moneda}}: \{\text{cara, ceca}\}$$

$$P(X = \text{cara, ceca}) = 1/2$$



Distribución de probabilidad uniforme: lanzamiento de una moneda



Variables discretas: Distribución **uniforme**

La distribución de probabilidad **uniforme** asigna la misma probabilidad para todo un rango de valores.



Ejemplos: moneda, dado.

$$X_{\text{dado}}: \{1, 2, 3, 4, 5, 6\}$$

$$P(X = 1, 2, 3, 4, 5, 6) = 1/6$$

Distribución de probabilidad uniforme: lanzamiento de un dado

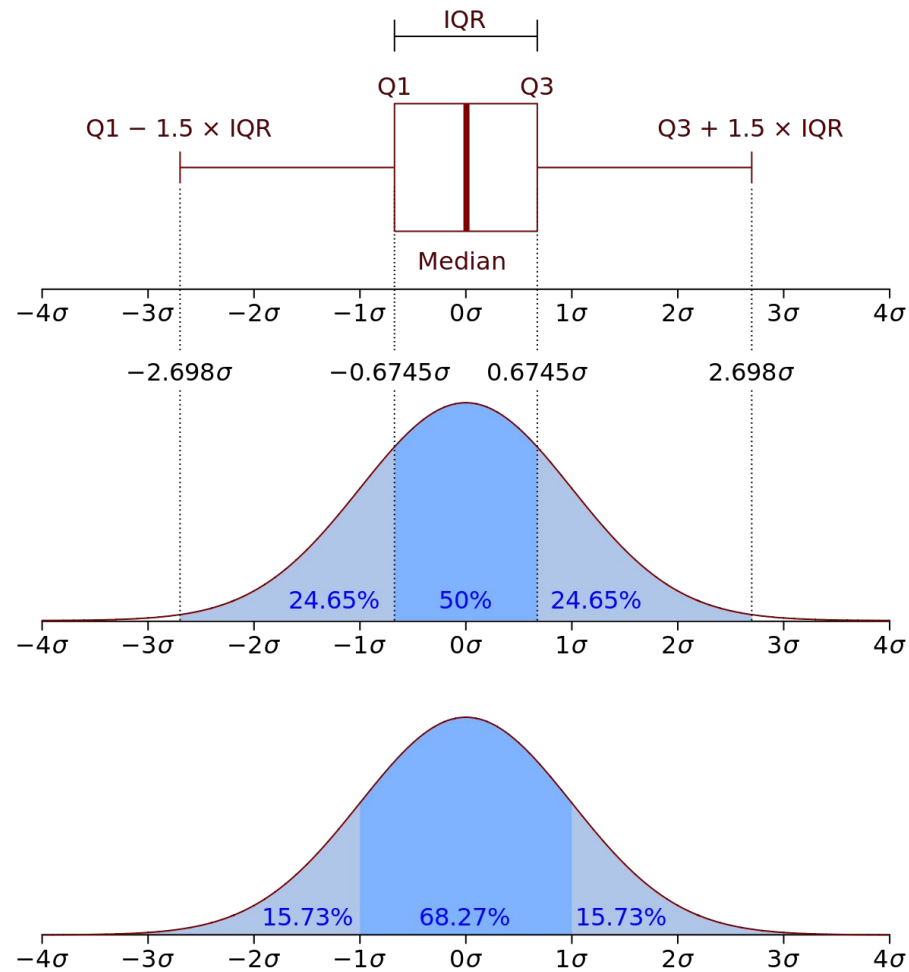


Para variables **continuas...**
¿qué concepto de probabilidad usamos?



Para variables **continuas**,
¡usamos el concepto de
densidad de probabilidad!





Probabilidad: Densidad normal o Gaussiana

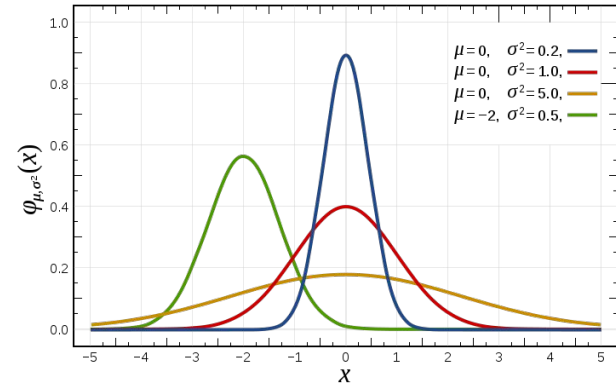
¡La más famosa de las distribuciones!

$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Parámetros:

μ : valor medio

σ : desviación estándar

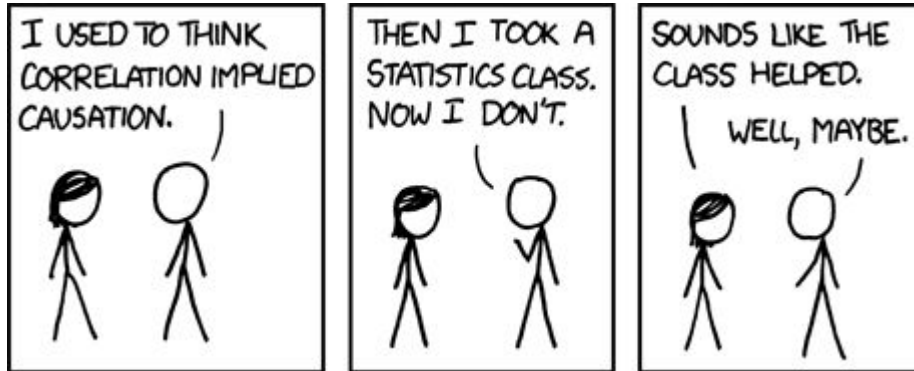


Causalidad y Correlación.



Correlación, Causalidad e Independencia estadística

Hoy hablaremos de correlación
(y un poco de Causalidad)



¡No podía faltar!

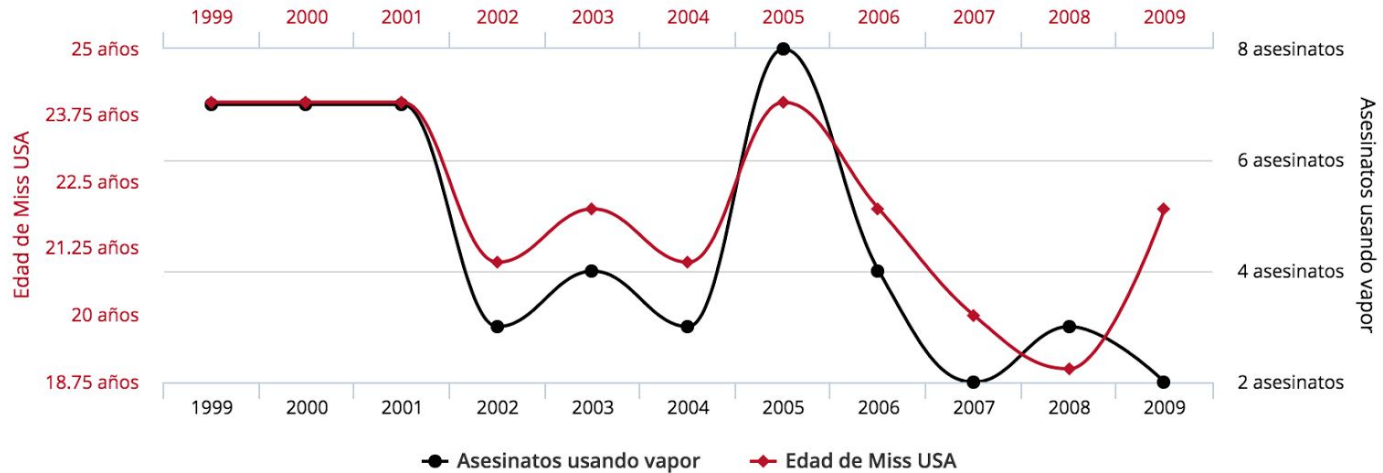


¿Correlación **implica
causalidad?**



La edad de Miss USA está correlacionada con Asesinatos usando vapor hirviendo como arma letal

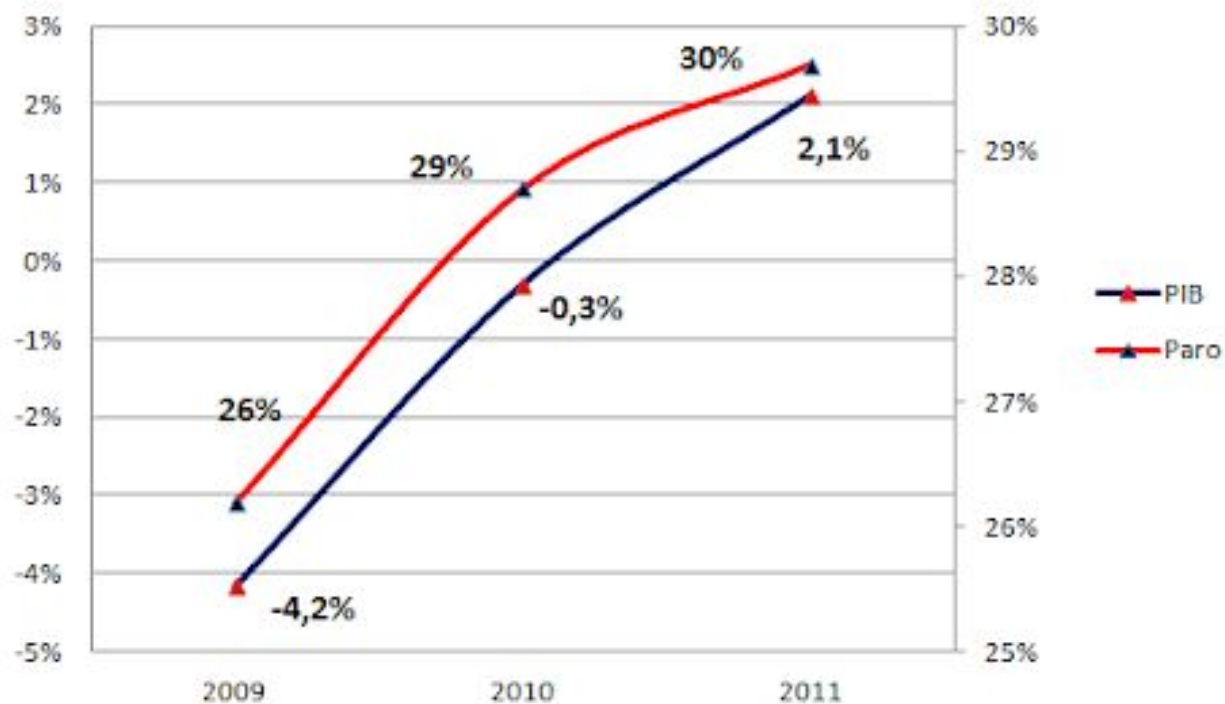
Correlación: 87.01% ($r=0.870127$)



tylervigen.com

Data sources: Wikipedia and Centers for Disease Control & Prevention

Canarias. Tasa de paro y variación del PIB % Años 2009-2011



Llamamos probabilidad de X ,
 $P(X)$ a un número entre 0 y 1
que le asignamos a cada
posible valor que puede
tomar esa variable aleatoria



Llamamos probabilidad de X ,
 $P(X)$ a un número entre 0 y 1
que le asignamos a cada
posible valor que puede
tomar esa variable aleatoria

También dijimos que la **mejor descripción** de $P(X)$ está dada por la distribución **(densidad) de probabilidades.**



Hasta ahora, consideramos **una sola variable aleatoria $X...$**

- Resultados obtenidos al tirar una moneda
- Resultados obtenidos al tirar un dado
- Altura de una persona
- Etc.



¿Pero qué ocurre si tengo más de una variable aleatoria?



¿Pero qué ocurre si tengo más de una variable aleatoria?

2 variables

Peso y altura de una persona
Temperatura y humedad

3 variables

Temperatura, humedad y velocidad del viento

4 variables

Iris Dataset: ancho y largo del sépalos y pétalo



Podemos pensar a un
Dataset como
una “**colección**” de
variables aleatorias



Probabilidad Conjunta

Dadas dos variables aleatorias X e Y

- **$P(X)$** es la distribución (o densidad) de probabilidades de X
- **$P(Y)$** es la distribución (o densidad) de probabilidades de Y
- **$P(X,Y)$** es la probabilidad CONJUNTA de X y de Y



Probabilidad Conjunta

$P(X,Y)$ es la probabilidad CONJUNTA de X y de Y

Es la distribución de probabilidad de los pares (x,y) , es decir, todos los posibles valores que pueden tomar las dos variables.



Probabilidad Conjunta

$P(X,Y)$ es la probabilidad CONJUNTA de X y de Y

Es la distribución de probabilidad de los pares (x,y) , es decir, todos los posibles valores que pueden tomar las dos variables.



Probabilidad Conjunta

$P(X,Y)$ es la probabilidad CONJUNTA de X y de Y

¡Veamos un ejemplo!

Medimos para muchas personas su peso y altura. Cada par (peso, altura) es una “muestra” de la distribución conjunta $P(X,Y)$.



Correlación, Causalidad e Independencia estadística

- Tres conceptos que tratan sobre la relación entre dos variables aleatorias.
- Muy fácil confundirlos entre ellos
- En su uso cotidiano tienen un significado un poco más “laxo” que en su uso estadístico



A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver spoon are visible, though they are out of focus. The overall lighting is soft and even.

¡BREAK!

Ph. Credit: Drew Coffmann



Hands-on training



DS_Bitácora_05_Probabilidad.ipynb

Sección 2 y 3



Relación entre Probabilidad y Estadística



Probabilidad y Estadística

PROBABILIDAD

Qué espero ver.

Modelos sobre la naturaleza o nuestro problema

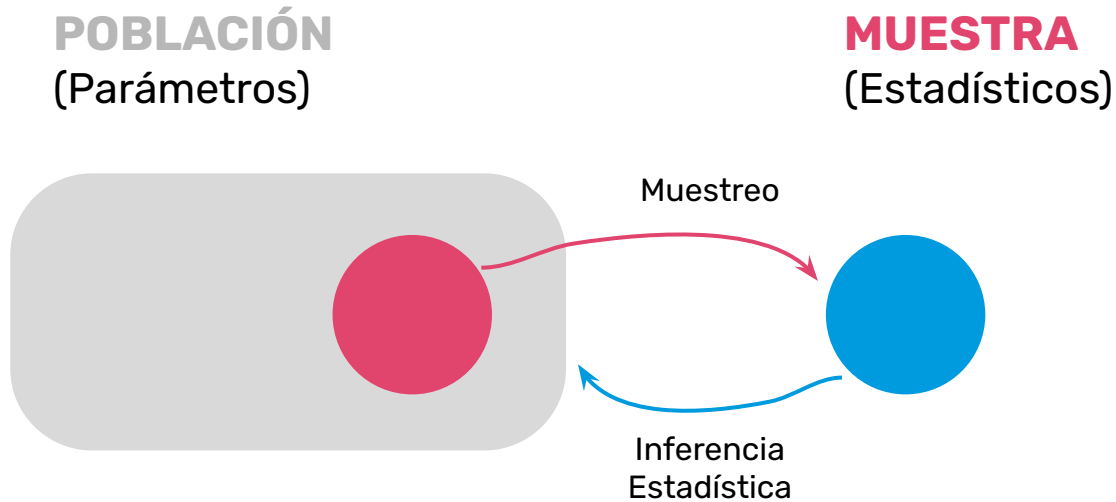
ESTADÍSTICA

Lo que vi. **Preguntas:** ¿tiene sentido con mi modelo? ¿Qué puedo aprender?

(Lo que mido en el laboratorio)



Población y Muestra



Si conozco la distribución de probabilidad
(o densidad de probabilidad)
con las fórmulas que mostramos,
podemos calcular su esperanza, varianza y más.

Pero **en general** no conocemos las distribuciones,
sino que tenemos **datos**.

Ahí es donde entra la **Estadística**.

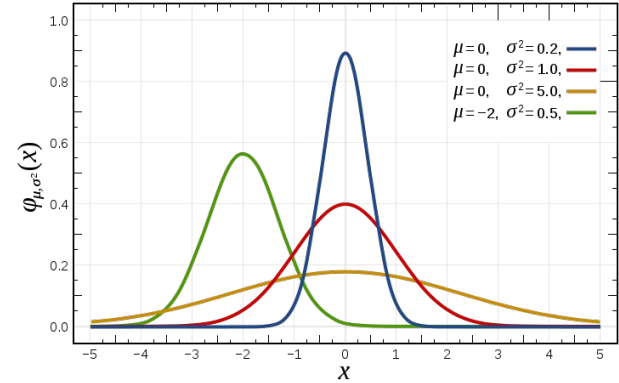
Relación entre estadísticos y parámetros

$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Parámetros:

μ : valor medio

σ : desviación estándar



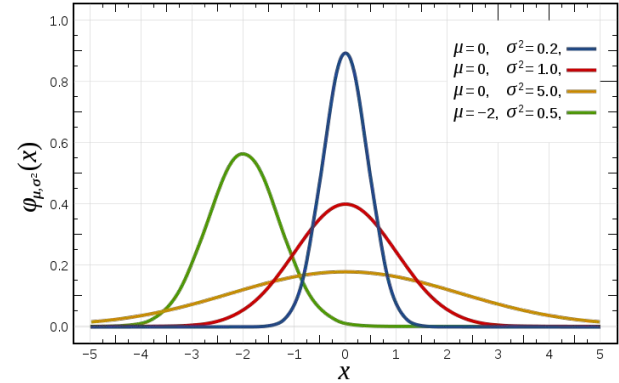
Relación entre estadísticos y parámetros

$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Parámetros:

μ : valor medio

σ : desviación estándar



Si nuestros datos tienen una distribución Gaussiana

Parámetro	Estadístico
μ	Promedio de los datos
σ	Desviación Estándar Calculada de los datos



Recursos



Probabilidad y Estadística

- [Correlaciones Espurias](#) - Visita obligada y divertida.
- [Statistics is the Grammar of Data Science](#) - Serie de cinco artículos sobre sobre Probabilidad y Estadística. Recomendamos particularmente leer el cuarto artículo sobre correlación.



Para la próxima

- Termina el notebook de hoy
- Lee la bitácora 06
- Resuelve el Challenge.

ACÀMICA