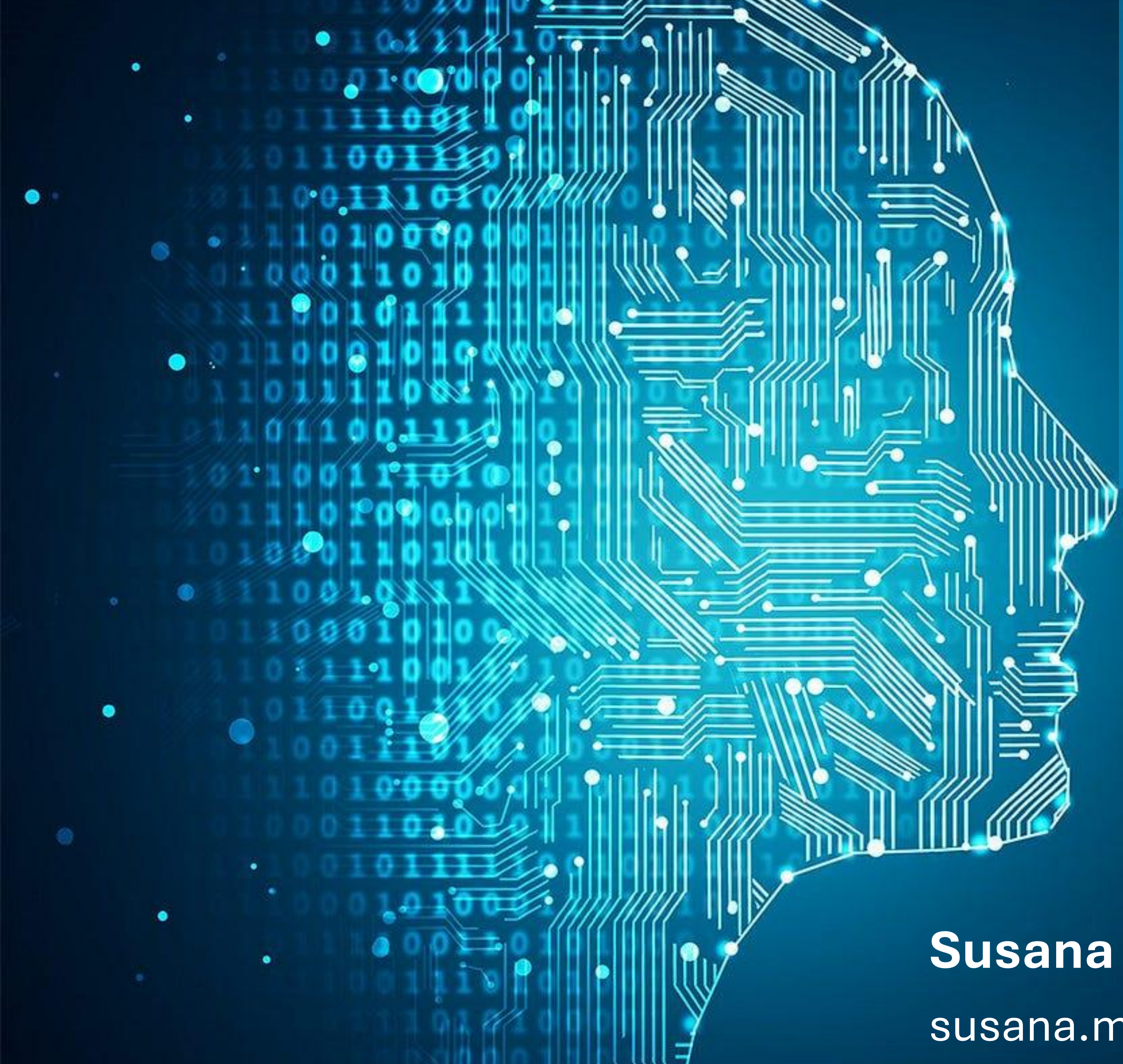


Machine Learning

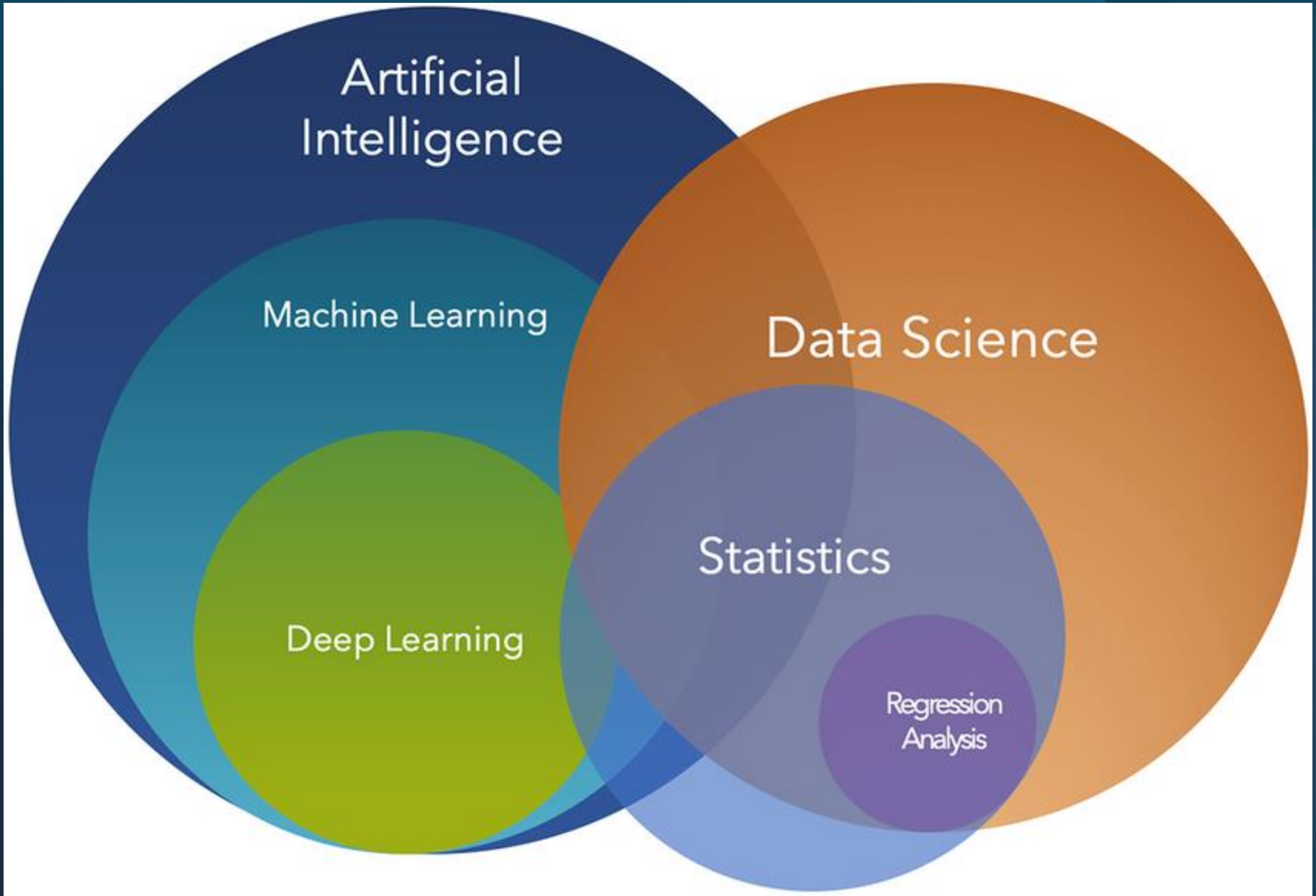
Susana Medina Gordillo

susana.medina@correounivalle.edu.co

Fundamentos Matemáticos y Estadísticos para Machine Learning

Fundamentos Matemáticos y Estadísticos

- Álgebra Lineal
- Probabilidad Aplicada



https://www.researchgate.net/figure/Venn-diagram-of-artificial-intelligence-and-data-science-Artificial-intelligence-AI-is_fig1_380790362

¿Por qué matemáticas y estadística en ML?

- El ML se basa en modelos matemáticos y estadísticos.

- Comprensión profunda = mejor diseño, ajuste y evaluación de modelos.

- Habilidades transferibles a otros campos (análisis de datos, etc.).

Álgebra Lineal

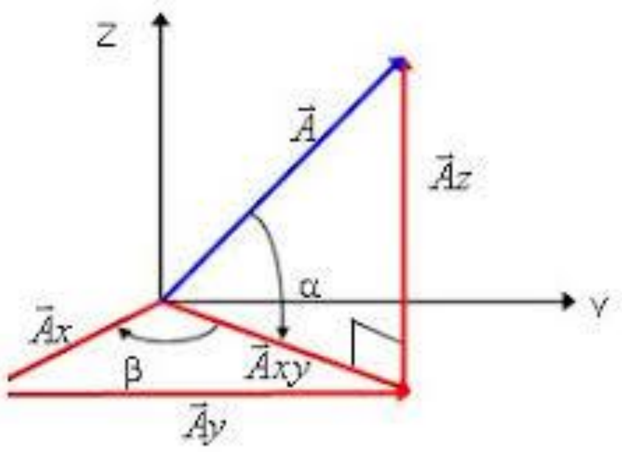
Álgebra Lineal: Vectores

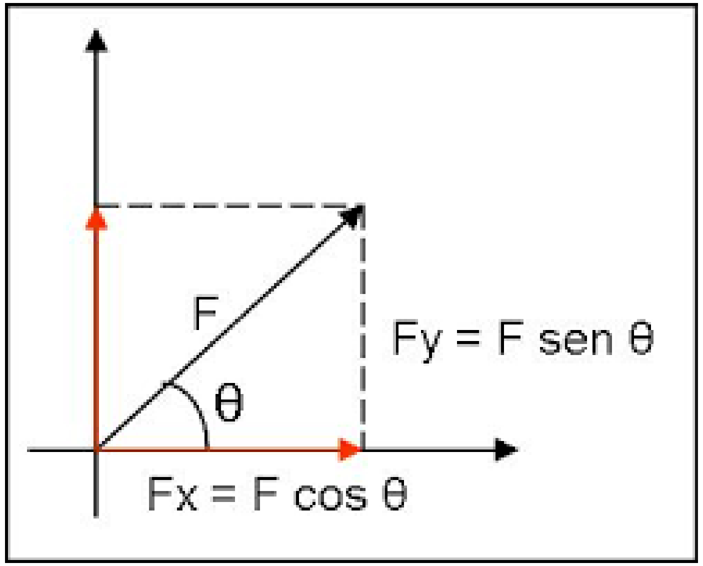
Vectores: Direcciones y magnitudes

Representación de puntos en el espacio, flechas con dirección y magnitud.

$$\vec{V} = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ \vdots \\ V_n \end{bmatrix} ; \vec{V} = [V_1, V_2, V_3, \cdots, V_n]$$

$\vec{A}xy = |A| * Cos\,\alpha$ y $\vec{A}z = |A| * Sen\,\alpha$





Álgebra Lineal: Matrices

Matrices: Organizaciones de datos

Arreglos rectangulares de números.

- Notación: Mayúsculas (ej. **A**).
- Dimensiones: Filas x columnas.

$$\underbrace{\begin{pmatrix}1 & 2\\ 1 & -1\\ 0 & 1\end{pmatrix}}_{3\times 2}\cdot\underbrace{\begin{pmatrix}x\\ y\end{pmatrix}}_{2\times 1}=\underbrace{\begin{pmatrix}x+2y\\ x-y\\ y\end{pmatrix}}_{3\times 1}$$

Elementos de una matriz

Columnas

Letra mayúscula

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13}\dots & a_{1n}\\ a_{21} & a_{22} & a_{23}\dots & a_{2n}\\ a_{31} & a_{32} & a_{33}\dots & a_{3n}\\ & \vdots & & \vdots\\ a_{m1} & a_{m2} & a_{m3}\dots & a_{mn}\end{bmatrix}$$

Filas

Dimensión: mxn

mn Términos

Álgebra lineal en Machine Learning

Representación de datos	Transformaciones lineales	Cálculo de modelos
• Matrices de características	• Rotación • Escalado	• Optimización de funciones objetivo

Estadística y Probabilidad

Probabilidad: Incertidumbre cuantificada

Probabilidad es.. . la medida de la posibilidad de un evento.

Espacio muestral (Ω): Conjunto de todos los resultados posibles.

Eventos: Subconjuntos del espacio muestral.

Probabilidad condicional: Probabilidad de un evento dado otro.

Eventos Excluyentes (Disyuntos): Sean A y B dos eventos de Ω si $A \cap B = \emptyset$, estos dos eventos se dicen excluyentes

Probabilidad: Teorema de Bayes

Sea Ω y $A \subset \Omega$. Si $\Omega = B_1 \cup B_2 \cup \ldots \cup B_k$ con $B_i \cap B_j = \emptyset \forall i \neq j$ y $P(B_i) > 0$, entonces:

Ley multiplicativa de la probabilidad

Probabilidad condicional

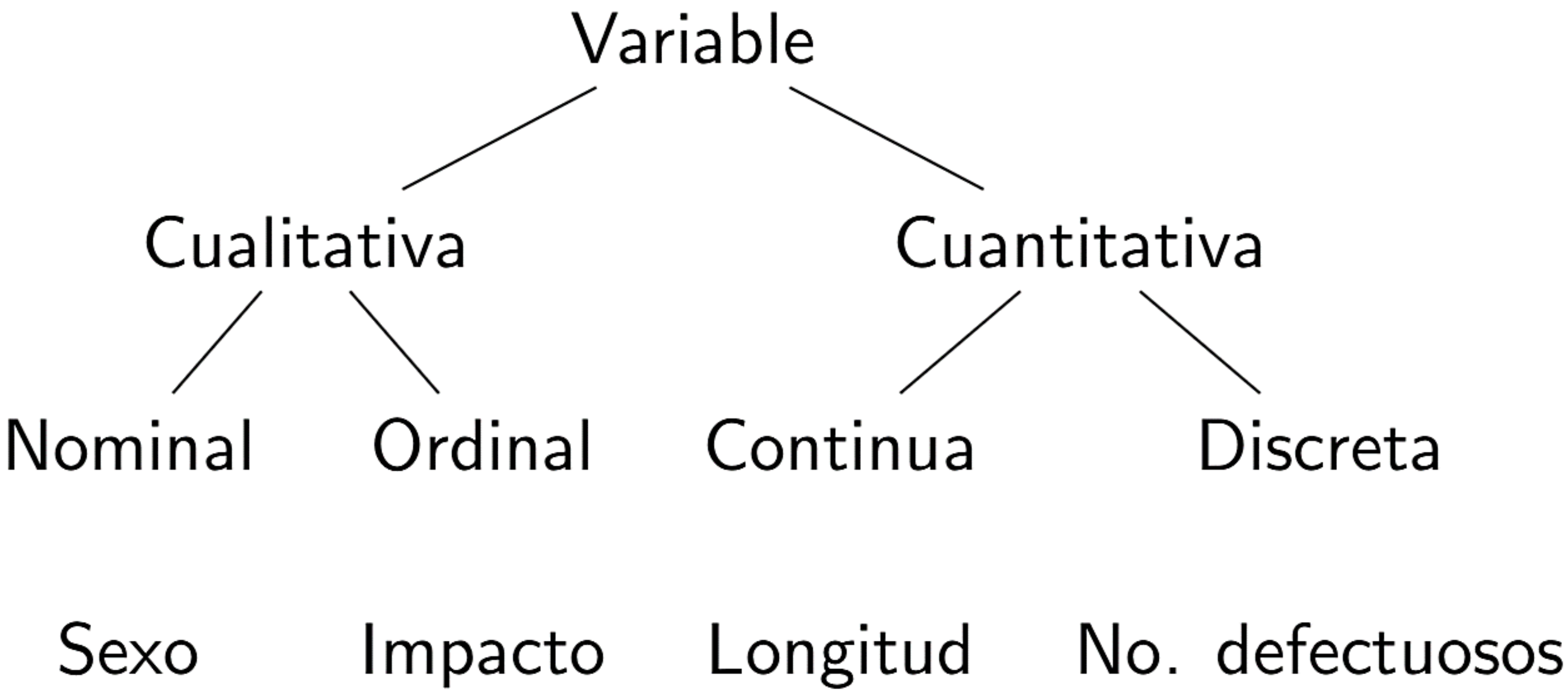
$$P(B_j/A) = \frac{P(B_j)P(A/B_j)}{\sum_{i=1}^{k} P(B_i)P(A/B_i)}$$

Ley de probabilidad total

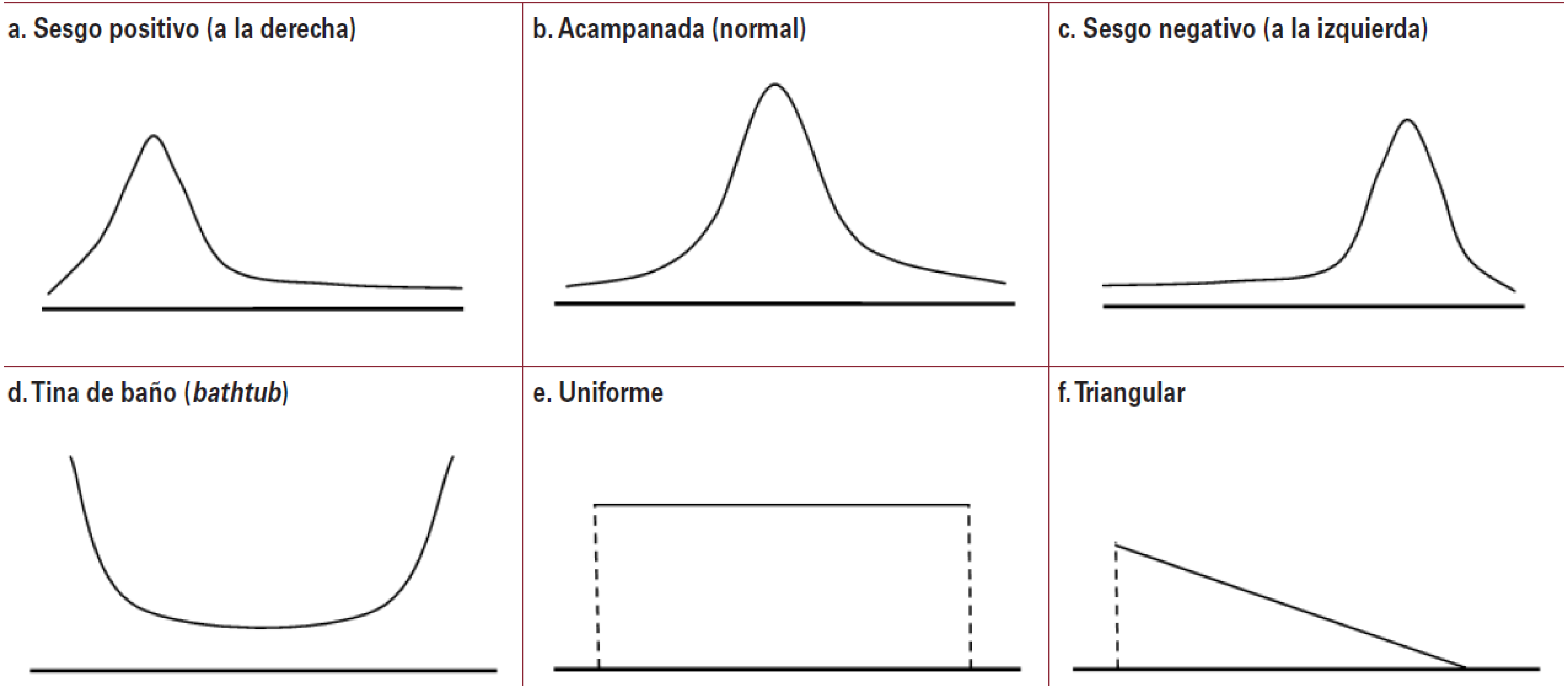
Distribuciones: Modelos de probabilidad

- **Discretas**: Bernoulli, binomial, Poisson.
- **Continuas**: Normal, uniforme, exponencial.
- **Visualizaciones**: Histogramas, curvas de densidad.

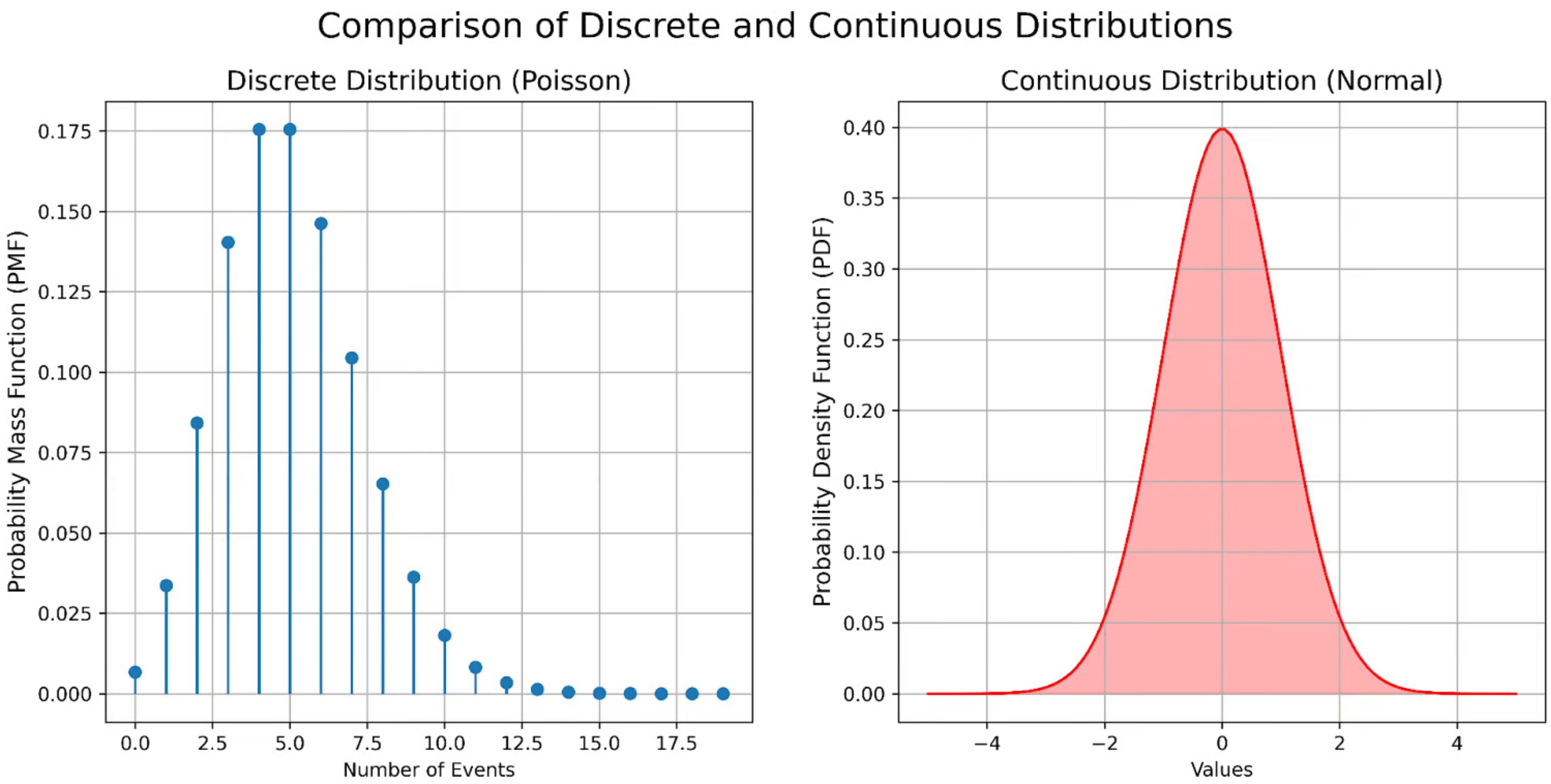
Variables


Variable
Cualitativa
Cuantitativa
Nominal
Ordinal
Continua
Discreta
Sexo
Impacto
Longitud
No. defectuosos

Distribuciones comunes



Distribuciones: Discretas vs Continuas



Distribución de Binomial

- Consta de n pruebas (repeticiones)
- Cada prueba: E: éxito o F: fracaso
- $P(E) = p$ (constante). $P(F) = q = 1 - p$
- Las pruebas son independientes
- X: número de éxitos observados en los n ensayos
  $\Omega_X = \{0, 1, 2, 3, \ldots, n\}$

Función de masa de probabilidad Binomial

$$P(X = x) = P(x) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, \ldots, n$$

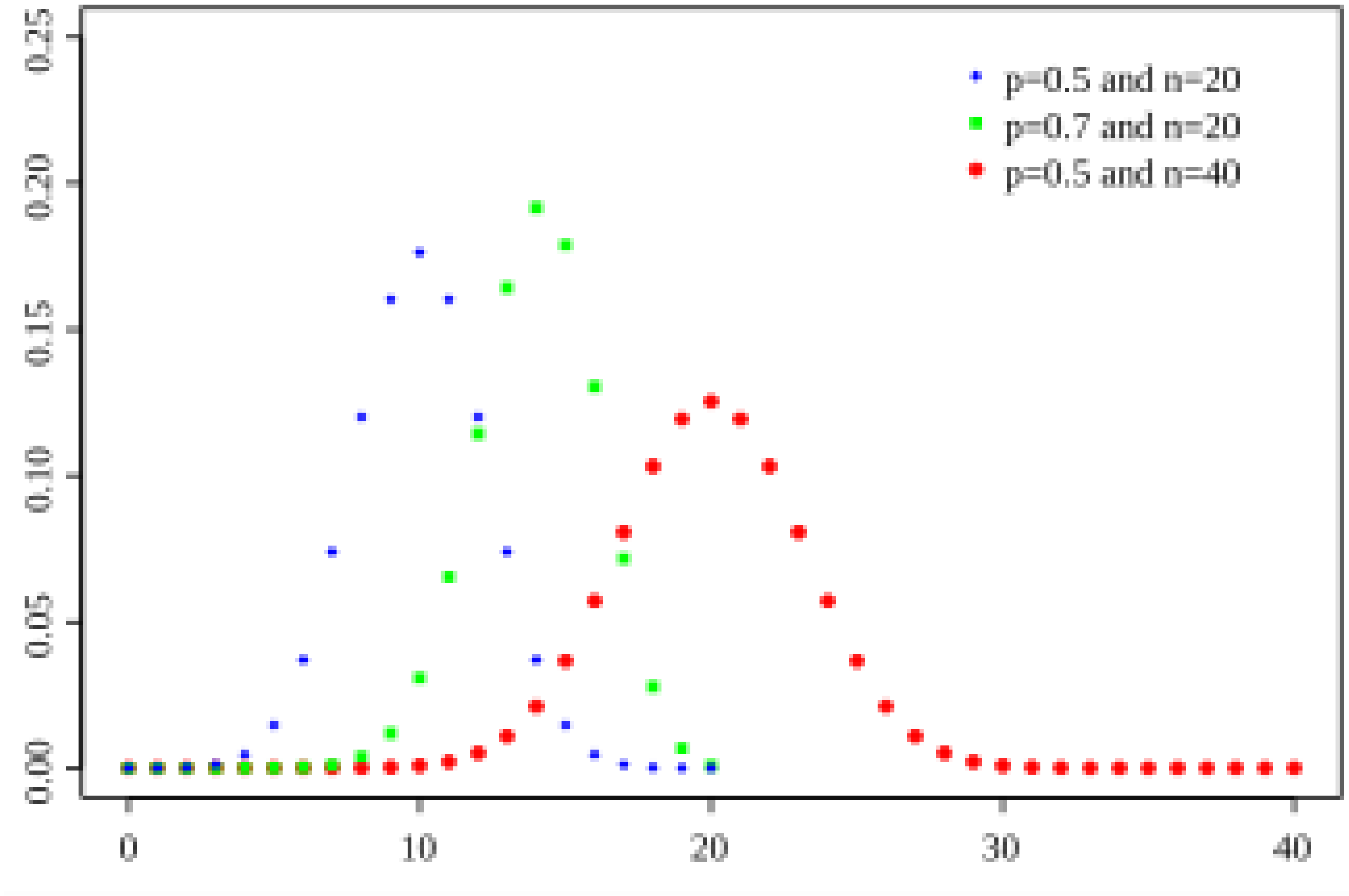
$$\binom{n}{x} = \frac{n!}{(n-x)!x!}$$

$$X \sim B(n; p)$$

Distribución de Binomial

El número de veces que la **moneda** aparece en la cara de la cruz sigue la distribución binomial donde el número de experimentos (**N**) realizados es n y la probabilidad de éxito es **`p=0.5`**

Distribución de Binomial



Fenómenos Binommiales

Distribución de Bernoulli

Si X es una variable aleatoria discreta que mide el "número de éxitos" y se realiza **un único experimento** con **dos posibles resultados** denominados éxito y fracaso, se dice que la variable aleatoria X se distribuye como una Bernoulli de parámetro p con $0 < p < 1$ y escribimos $X \sim \mathrm{Bernoulli}(p)$.

Su función de probabilidad es

$$\mathrm{P}[X = x] = p^x (1-p)^{1-x} \qquad x = 0, 1$$

lo anterior es equivalente a escribir

$$\mathrm{P}[X = x] = \begin{cases} 1 - p & \text{si } x = 0 \\ p & \text{si } x = 1 \end{cases}$$

en ocasiones también suele escribirse como

$$\mathrm{P}[X = x] = px + (1-p)(1-x) \qquad x = 0, 1$$

Bernoulli es un caso especial de la distribución Binomial en el que se realiza un ensayo

Distribución de Poisson

Es una distribución de probabilidad discreta.

La variable en una distribución de Poisson cuenta el número de eventos raros que ocurren en alguna unidad dimensional (espacio, tiempo, área, volumen, etc.)

Se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos raros.

- X: número de eventos que ocurren por unidad
- μ: promedio de ocurrencias por unidad

$$p(x) = \frac{e^{-\mu}\mu^x}{x!}, x = 0, 1, 2, \ldots$$

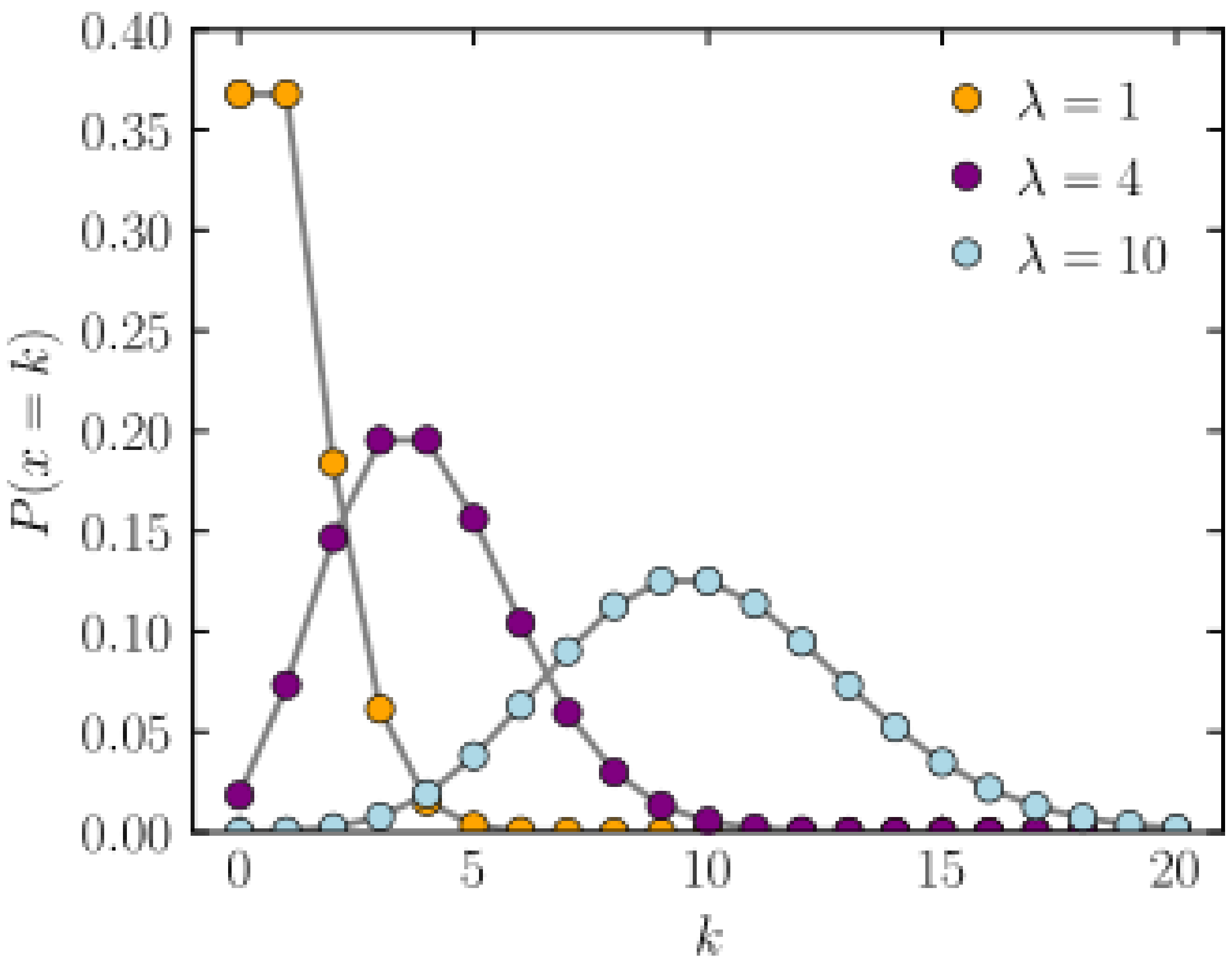
- Parámetro de la distribución: μ (también se denota con λ)
- Notación: $X \sim P(\mu)$

$$E(X) = \mu$$

$$\sigma^2 = V(X) = \mu$$

promedio y varianza son iguales

Distribución de Poisson



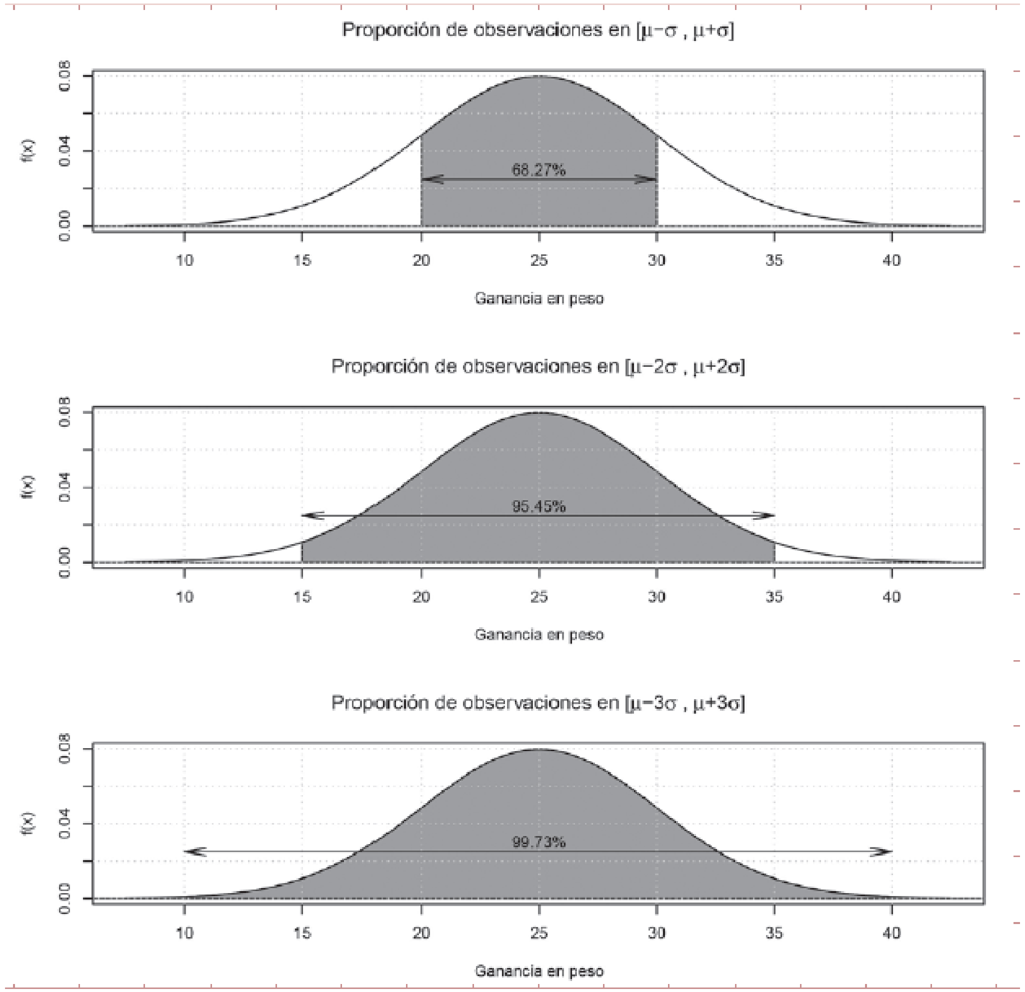
Situaciones con distribución Poisson

- ➢Número de avistamientos de ballenas/semana.
- ➢Clientes que llegan a un banco/hora.
- ➢Árboles caídos/km2
- ➢Serpientes arrolladas /sección de carretera de 500 m.
- ➢Partículas en suspensión /L.
- ➢Animales enfermos/200 especímenes.

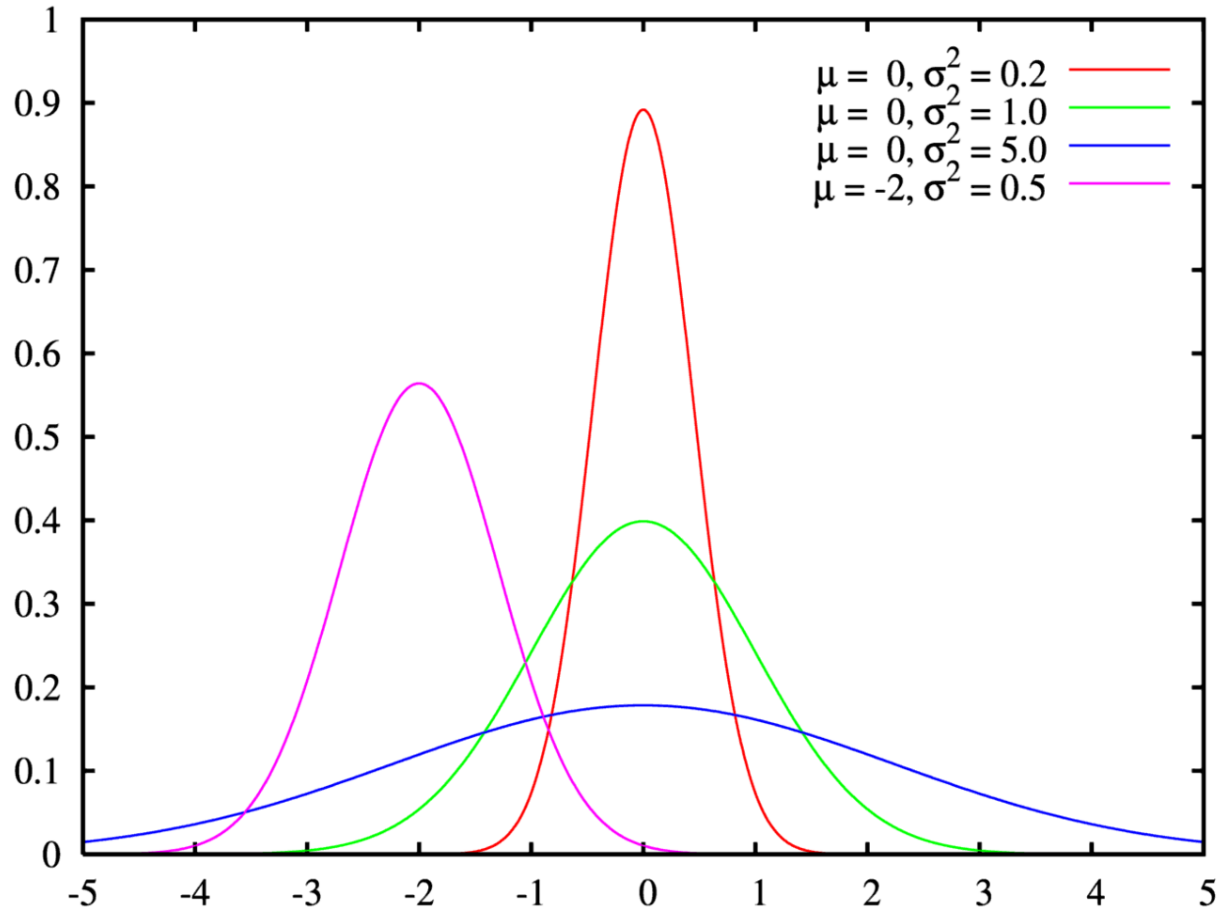
Distribución Normal

Es la más importante y también la más ampliamente referida en la probabilidad y la estadística.

Una característica muy distintiva es su gráfica en forma de campana, que implica una gran parte de las observaciones de la variable ubicada en las cercanías del promedio y *a medida que nos alejamos del centro van disminuyendo las observaciones*.



Distribución Normal



Función de probabilidad

$$f(X) = \frac{1}{\sqrt{2\pi\sigma}} \mathrm{e}^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$-\infty < x < \infty$$

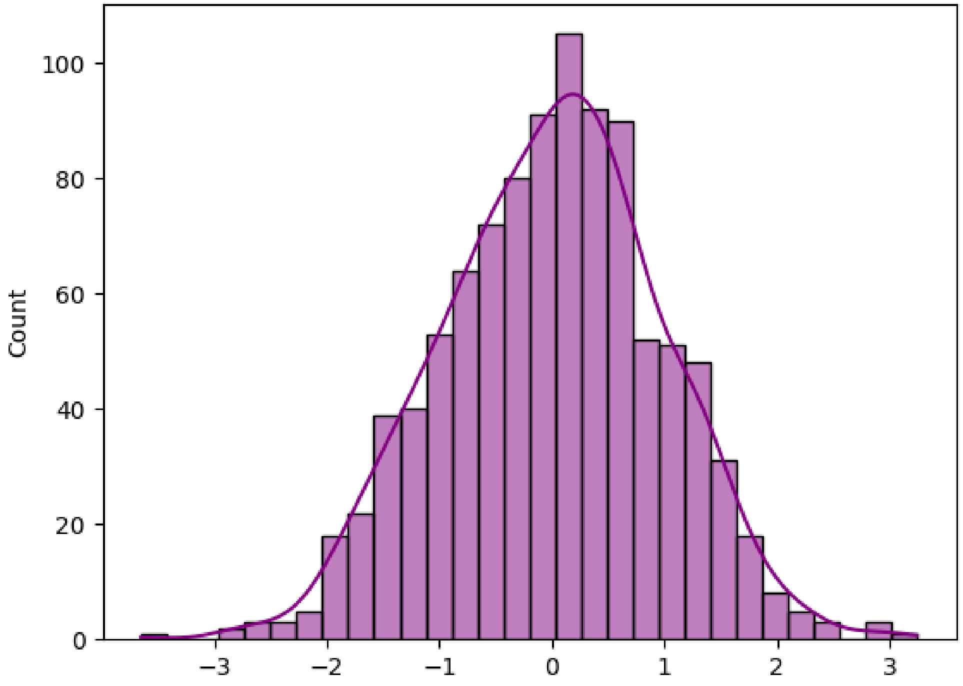
$$-\infty < \mu < \infty$$

$$\sigma > 0$$

Histogramas

Los histogramas permiten mostrar la distribución de un conjunto de datos en diferentes segmentos o rangos.

- ✓ La altura de cada barra representa el recuento de valores de cada rango.
- ✓ Las barras no están separadas entre sí.
- ✓ La altura de cada barra representa la frecuencia con la que aparece la variable.
- ✓ Los datos se agrupan en rangos llamados intervalos.
- ✓ Los histogramas son útiles para mostrar la distribución de una sola variable.



Medidas de tendencia central

- **Media** (promedio): sumatoria de números de **i** hasta **n** divididos por la cantidad total (**N**).

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{N}$$

- **Mediana**: es el número que queda en la posición exactamente central

$$\tilde{X} = \begin{cases} X\left(\frac{n+1}{2}\right) & n\ impar \\ \frac{X\left(\frac{n}{2}\right)+X\left(\frac{n}{2}+1\right)}{2} & n\ par \end{cases}$$

- **Moda**: el dato que más se repite.
  - Variables que no tienen moda
  - Variables con múltiples modas (multimodal)

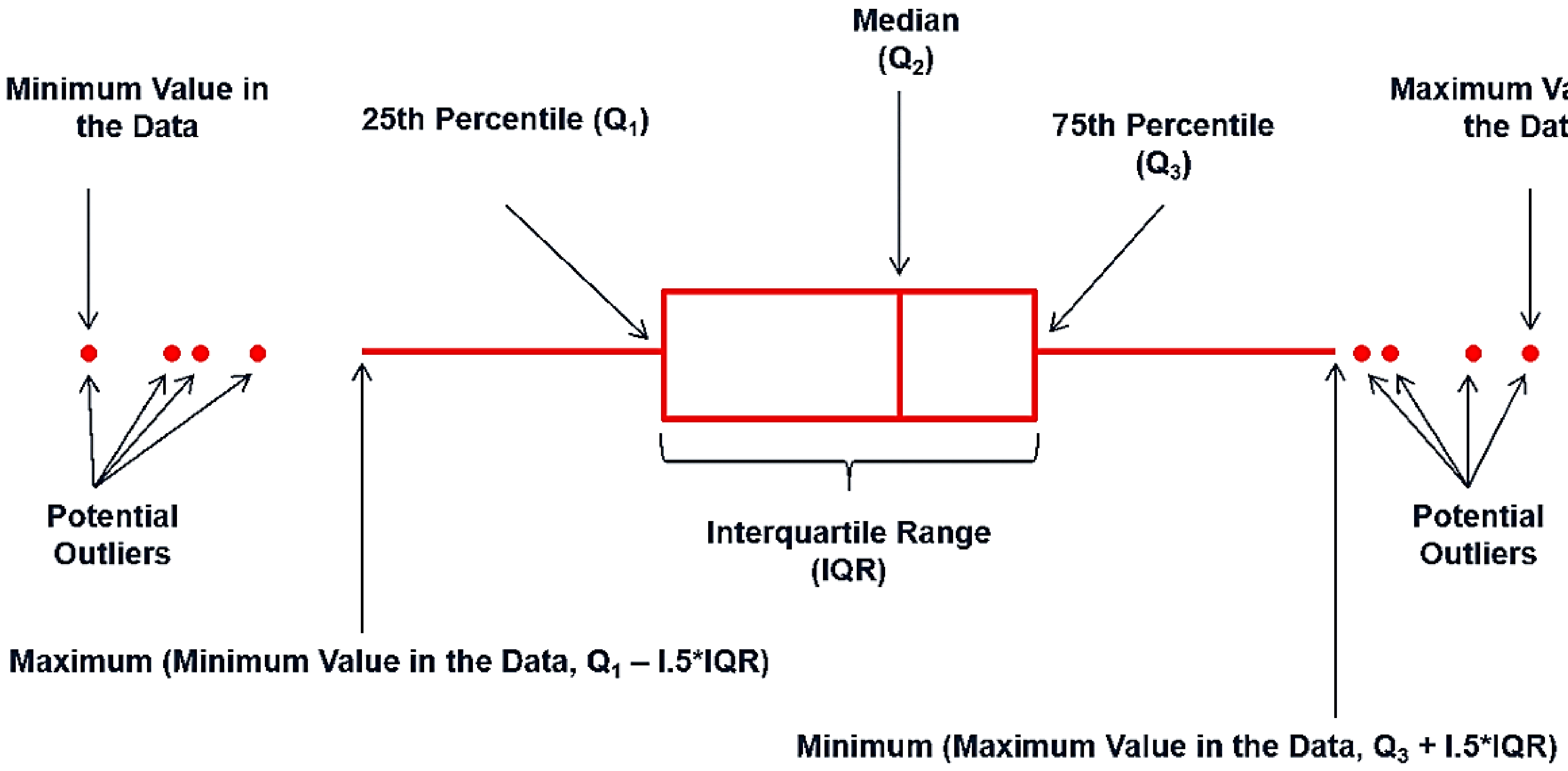
Estadística descriptiva: Medidas de dispersión

- **Varianza**: representa la variabilidad de un grupo de observaciones
  - Siempre es positiva!
- **Desviación estándar**: es la raíz cuadrada de la varianza.

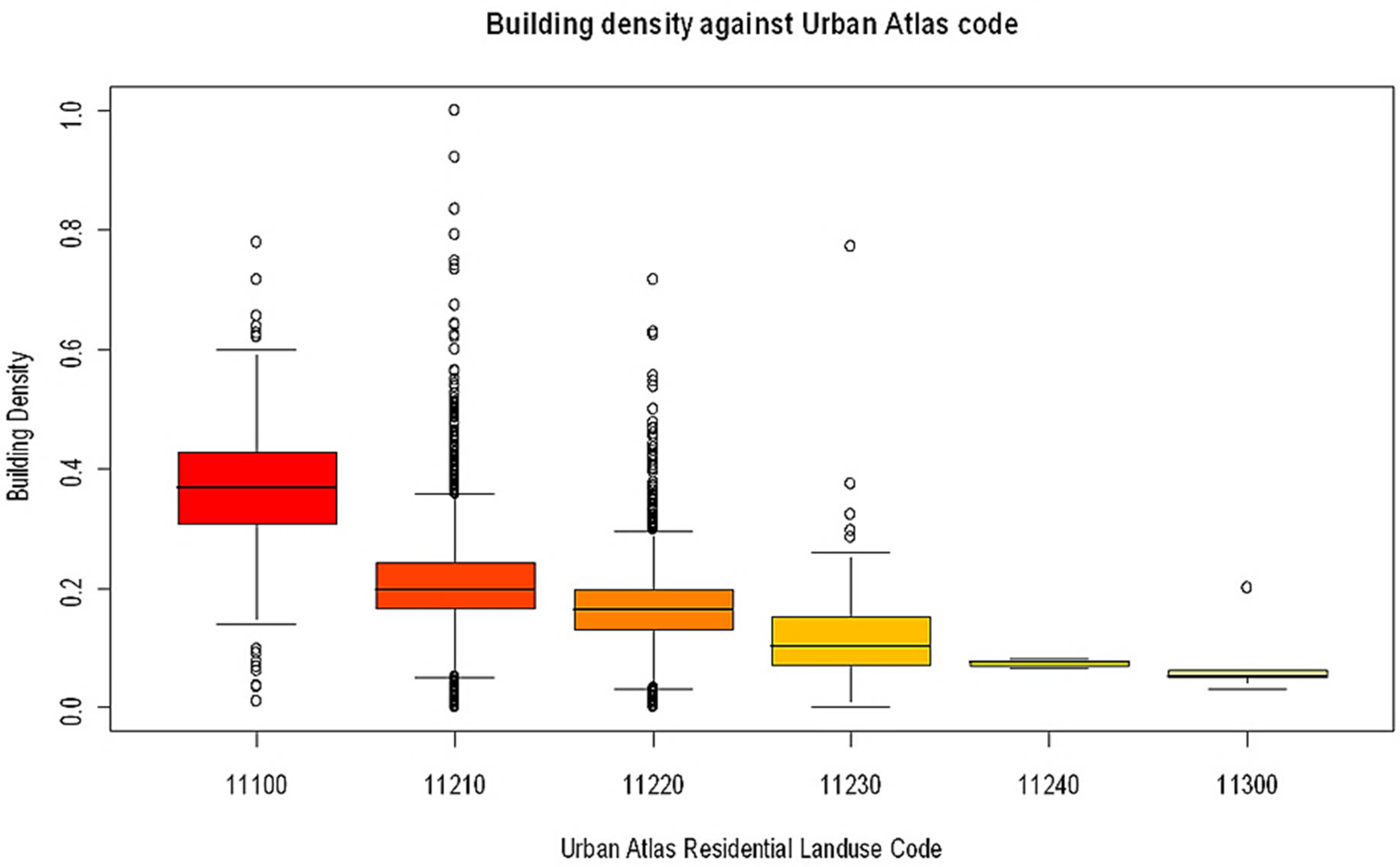
$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

Observaciones	Promedio	Diferencias	Diferencia^2
6	5.4	0.6	0.36
8	5.4	2.6	6.76
9	5.4	3.6	12.96
1	5.4	-4.4	19.36
3	5.4	-2.4	5.76
		Suma	45.2
		Varianza	9.04

Visualizaciones: Diagramas de caja (boxplot)



Visualizaciones: Diagramas de caja (boxplot)



Inferencia: Aprendiendo de los datos

- **Estimación de parámetros**: Encontrar valores de modelos.
- **Pruebas de hipótesis**: Evaluar la evidencia contra una afirmación.
- **Intervalos de confianza**: Rango de valores plausibles para un parámetro.

Pruebas de hipótesis

Una vez realizado el contraste de hipótesis, se habrá optado por una de las dos hipótesis, la **hipótesis nula** o base H_0 o la **hipótesis alternativa** H_1 , y la decisión escogida coincidirá o no con la que en realidad es cierta. Se pueden dar los cuatro casos que se exponen en el siguiente cuadro:

	H_0 **es cierta**	H_1 **es cierta**
Se escogió H_0	No hay error (verdadero negativo)	Error de tipo II (β o falso negativo)
Se escogió H_1	Error de tipo I (α o falso positivo)	No hay error (verdadero positivo)

Si la probabilidad de cometer un error de tipo I está unívocamente determinada, su valor se suele denotar por la letra griega α, y en las mismas condiciones, se denota por β la probabilidad de cometer el error de tipo II, esto es:

$$P(\text{escoger } H_1 | H_0 \text{ es cierta}) = \alpha$$
$$P(\text{escoger } H_0 | H_1 \text{ es cierta}) = \beta$$

En este caso, se denomina **Potencia del contraste** al valor 1-β, esto es, a la probabilidad de escoger H_1 cuando esta es cierta

$$P(\text{escoger } H_1 | H_1 \text{ es cierta}) = 1 - \beta \ .$$

Probabilidad y estadística en ML

- **Modelos probabilísticos**: Bayesiano, redes neuronales.
- **Evaluación de modelos**: Métricas de rendimiento.
- **Manejo de incertidumbre**: Datos faltantes, ruido.

Ejemplo: Clasificación de correos electrónicos

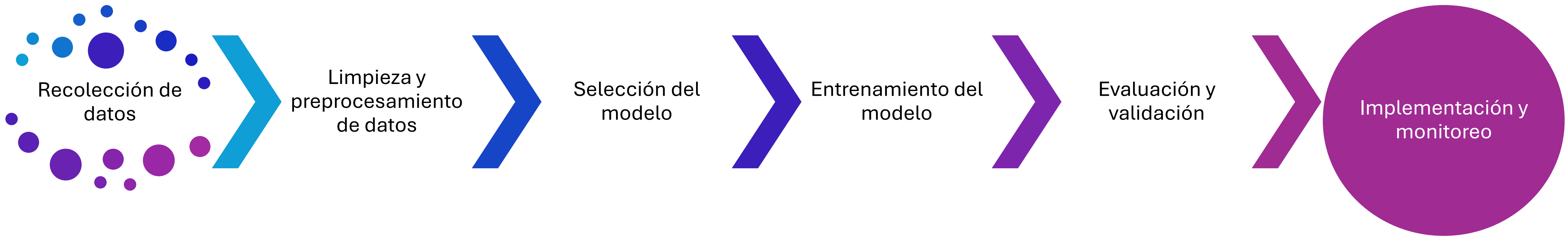
Problema: Distinguir spam de no spam.

Características: Frecuencia de palabras, remitente.

Modelo: Bayesiano.

Probabilidad condicional: Clave para la clasificación.

Flujo de trabajo en Machine Learning



Resumen...

Álgebra lineal

- Herramienta fundamental para representar y manipular datos.

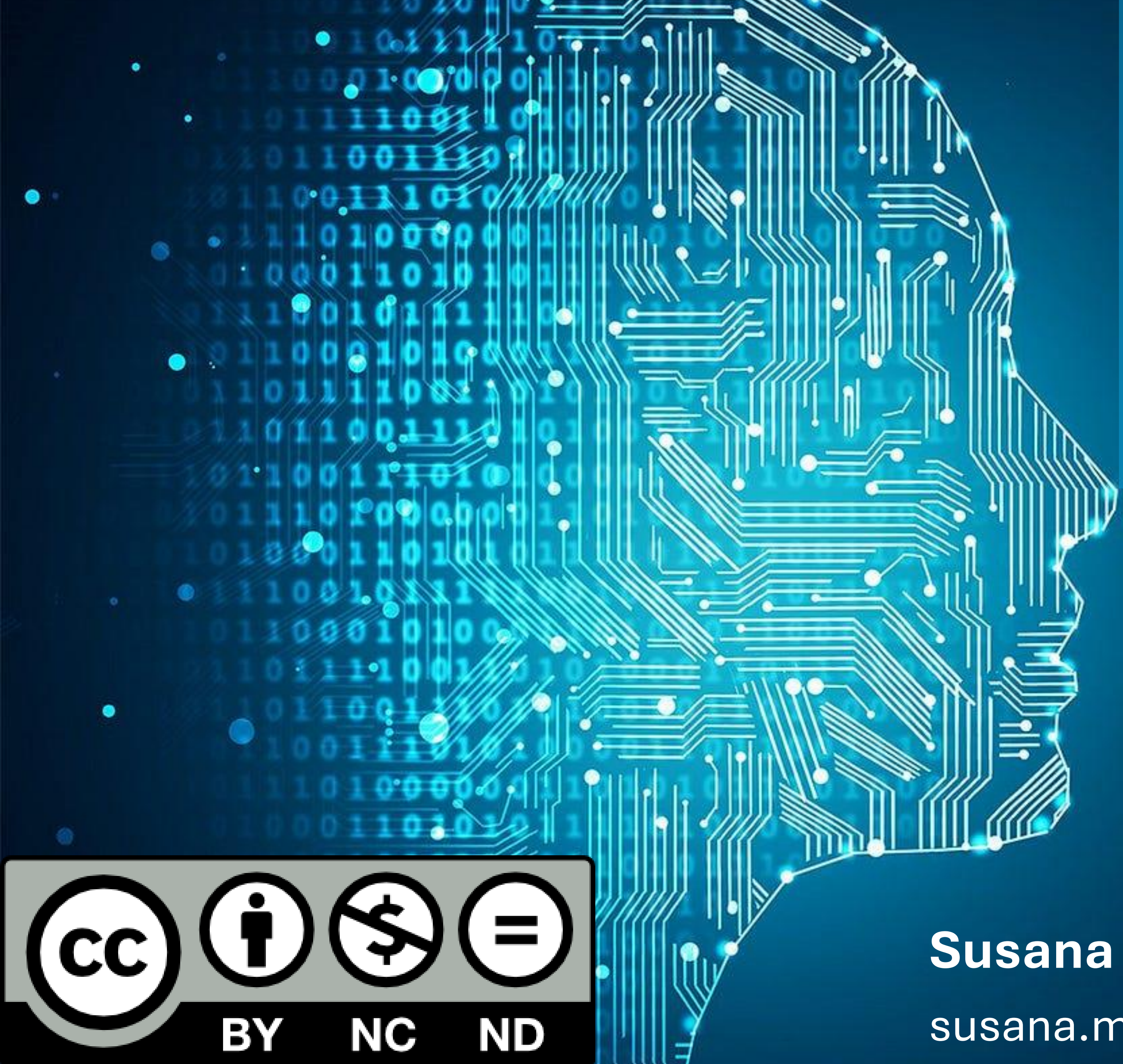
Probabilidad y estadística

- Esenciales para modelar la incertidumbre y aprender de los datos

Dominar estos fundamentos = base sólida para el éxito en ML

Referencias

- “Distribución de Poisson: Guía completa”. Consultado: el 6 de febrero de 2025. [En línea]. Disponible en: https://www.datacamp.com/tutorial/poisson-distribution
- Imagen Holistic BI https://halobi.com/blog/data-science-for-the-business-intelligence-professional/
- Imágenes de Datacamp slides: “Scaling Data Science at your Organization”.
- Imagen Data visualization: https://venngage.com/blog/venn-diagram-powerpoint/



Machine Learning

Susana Medina Gordillo

susana.medina@correounivalle.edu.co