

UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# VISIÓN ARTIFICIAL

JOHN W. BRANCH

PROF. TITULAR

DEPARTAMENTO DE CIENCIAS DE COMPUTACIÓN Y DE LA DECISIÓN

DIRECTOR DEL GRUPO GIDIA

ALBERTO M. CEBALLOS

ASISTENTE DE DOCENCIA

**Nota:** Este material se ha adaptado con base en el material de los profesores Domingo Mery (U. de Chile), María Patricia Trujillo (Univalle), Ginés García (U. de Murcia) y Nicolas Fernández (U. de Córdoba)

# A CONTINUACIÓN...

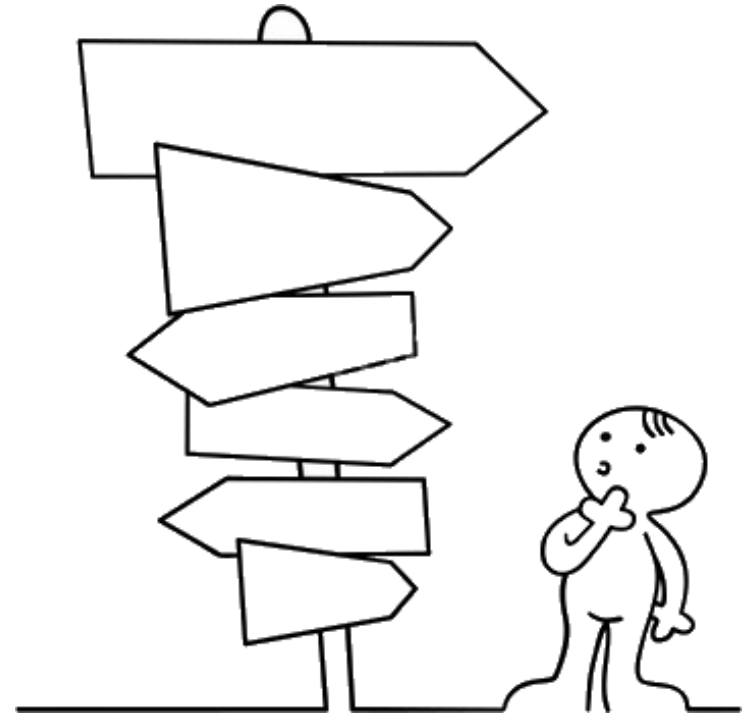
## 🚀 BASES DE ESTADÍSTICA

### 🎯 El histograma

### 🎯 Medidas estadísticas básicas

- 🌐 Media
- 🌐 Mediana
- 🌐 Moda
- 🌐 Desviación estándar

### 🎯 Métricas de error



# BASES DE ESTADÍSTICA

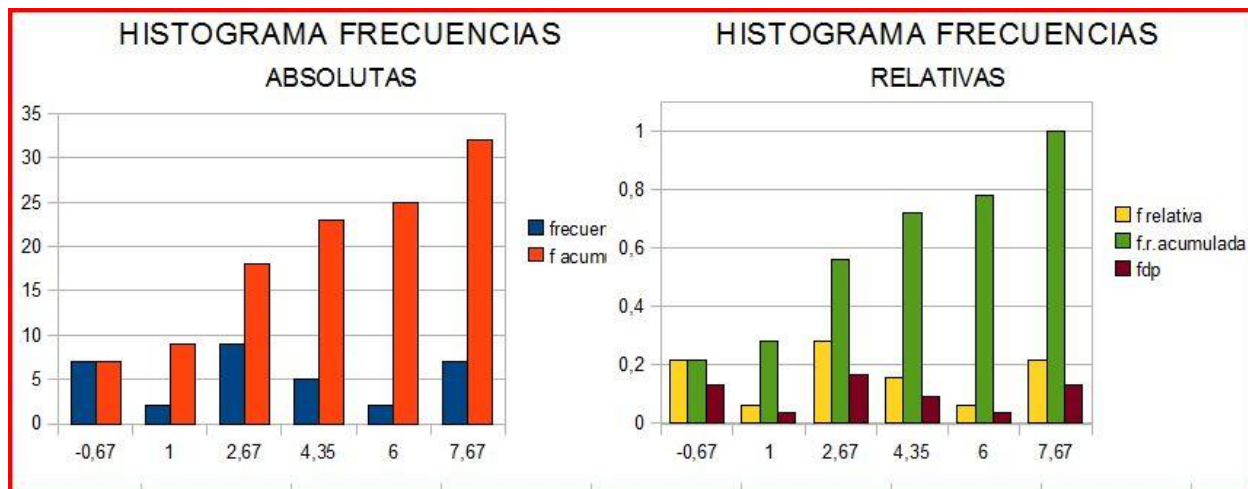
EL HISTOGRAMA



# BASES DE ESTADÍSTICA

## 🚀 EL HISTOGRAMA

- 🕒 Un histograma es una representación de la frecuencia de cada posible valor de una variable. Obtener el histograma de un conjunto de datos es bastante fácil a nivel computacional...
- 🕒 Un histograma de frecuencias relativas se obtiene al dividir la frecuencia de cada valor que toma la variable entre el número total de datos estudiados. Más adelante, trabajaremos con funciones de densidad de probabilidad preestablecidas...




# BASES DE ESTADÍSTICA

MEDIDAS ESTADÍSTICAS BÁSICAS



# BASES DE ESTADÍSTICA

## LA MEDIA

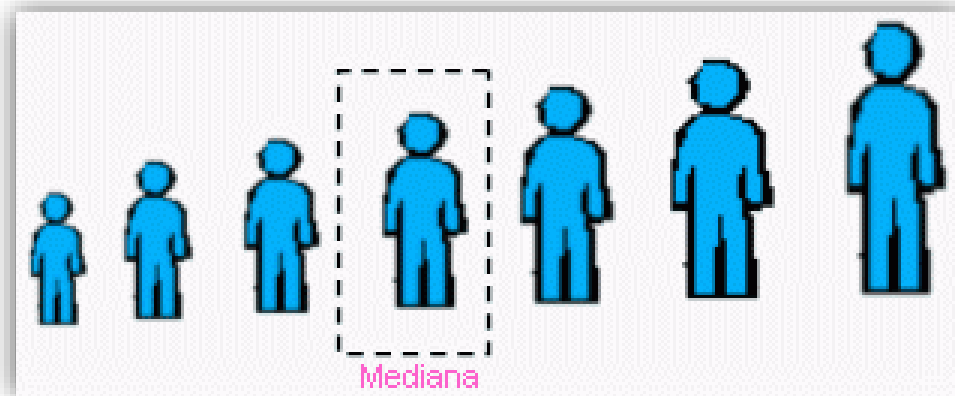
-  Dado un conjunto de datos numéricos, la media es la suma de todos los valores dividida entre la cantidad de elementos que hay en el conjunto de datos.

$$\bar{x} = \frac{\sum_{i=1}^n x[i]}{n}$$

# BASES DE ESTADÍSTICA

## 🚀 LA MEDIANA



- 🕒 La mediana es el valor intermedio en un conjunto de datos ordenado de tamaño impar.
- 🕒 Cuando se tiene una cantidad impar de datos, la media es el promedio de los dos valores centrales. Ej:  $X = [1, 2, 2, 3, 3, 3] \rightarrow \text{mediana}(X) = \frac{2+3}{2} = 2.5$
- 🕒 Es menos susceptible al ruido que la media, pero más costosa computacionalmente si se tienen que ordenar los datos.







# BASES DE ESTADÍSTICA

## LA MODA

-  La moda es el valor más frecuente en un conjunto de datos.
-  Cuando varios valores tienen la máxima frecuencia del conjunto de datos, la moda es el promedio de los susodichos.

# BASES DE ESTADÍSTICA


## LA DESVIACIÓN ESTÁNDAR

-  La desviación estándar es una medida de la dispersión en un conjunto de datos, y se calcula como la sumatoria del cuadrado de las diferencias entre cada valor del conjunto de datos y la media, dividida entre el total de elementos menos uno.
-  Para obtener la varianza de un conjunto de datos, basta con elevar la desv. estándar al cuadrado.

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

# BASES DE ESTADÍSTICA

## OTRAS MEDIDAS

-  Es conveniente hacer un repaso autónomo de otras medidas tales como la correlación y la covarianza, y de nociones de probabilidad, que aparecerán más adelante en el curso.

# BASES DE ESTADÍSTICA

MÉTRICAS DE ERROR



# BASES DE ESTADÍSTICA


## MÉTRICAS DE ERROR

- Las métricas de error nos permiten medir la desviación de una predicción o estimación con respecto a una referencia o 'ground-truth'. Son especialmente relevantes para los problemas de eliminación de ruido y regresión/clasificación.
- Abajo se relacionan algunas de las métricas más usadas para la regresión. Para procesamiento digital de imágenes se suele trabajar con base en otras métricas tales como el Error Cuadrático Medio, la Proporción Máxima de Señal a Ruido y el SSIM.


Definitions	Observation $y_t$	Forecast $\hat{y}_t$	Forecast error $e_t = y_t - \hat{y}_t$
Accuracy measure	Calculation		
Mean Absolute Error	$MAE = average( e_t )$		
Mean Squared Error	$MSE = average(e_t^2)$		
Mean Absolute Percentage Error	$MAPE = 100 \times average( \frac{e_t}{y_t} )$		

# BASES DE ESTADÍSTICA

## MÉTRICAS DE ERROR

-  Error Cuadrático Medio (ECM o MSE): Se define como el promedio de los cuadrados de los errores sobre el conjunto de datos.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

-  La Proporción Máxima de Señal a Ruido (PSNR): Es un término utilizado para definir la relación entre la máxima energía posible de una señal y el ruido que afecta a su representación fidedigna, medida en dB. Depende del MSE y del valor máximo de intensidad que puede tomar un píxel en una imagen,  $2^b - 1$ .

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right)$$

# BASES DE ESTADÍSTICA

## MÉTRICAS DE ERROR

- 🌀 En futuras sesiones trataremos métricas de error como el SSIM, y discutiremos las métricas de error manejadas en problemas de clasificación tales como las matrices de confusión.

# PREGUNTAS

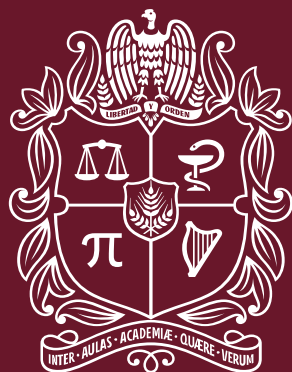




CARLOS ANDRÉS MERA BANGUERO, PH.D.

JOHN WILLIAN BRANCH BEDOYA, PH.D.

ALBERTO MARIO CEBALLOS ARROYO, ING.



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA