

MODULO I – INTRODUCCIÓN Y REPASO GENERAL DE CONCEPTOS

- DATA SCIENCE CON PYTHON
- 
- A large yellow triangle is positioned in the bottom right corner of the slide, pointing towards the top right.

Que es el Data Science

Es un campo **interdisciplinario** que utiliza técnicas, procesos, algoritmos y sistemas científicos para extraer conocimiento y percepciones a partir de datos estructurados y no estructurados.

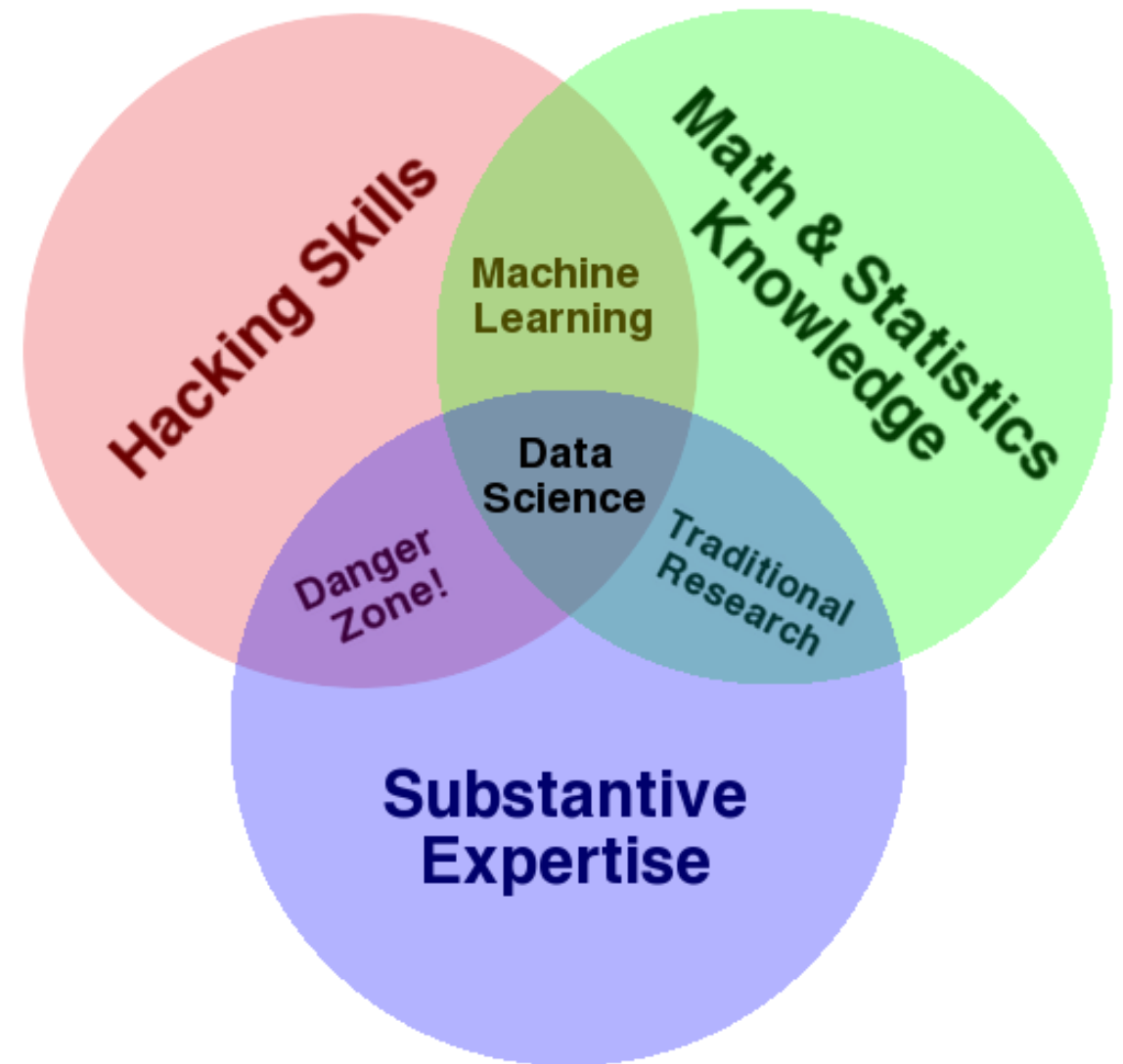


Diagrama de Venn propuesto por O'Reilly para Data Science

Incluye tres áreas clave:

1. "Hacking Skills" (Habilidades de Hacking),
2. "Math & Statistics Knowledge" (Conocimiento en Matemáticas y Estadísticas), y
3. "Substantive Expertise" (Conocimiento especializado en el Dominio).

Cada una de estas áreas se superpone con otras, creando intersecciones que representan las habilidades y conocimientos necesarios en el campo de Data Science.

Diagrama de Venn propuesto por O'Reilly para Data Science. Areas.

1. **Hacking Skills (Habilidades de Hacking)**: Esta área se refiere a la capacidad de trabajar con datos de manera práctica y técnica. Incluye habilidades de programación, manejo de herramientas y software, y la capacidad de extraer, limpiar y manipular datos de diversas fuentes. Las habilidades de hacking son esenciales para la **recopilación y preprocesamiento de datos**.
2. **Math & Statistics Knowledge (Conocimiento en Matemáticas y Estadísticas)**: Esta área engloba la comprensión profunda de conceptos matemáticos y estadísticos. Incluye la capacidad de aplicar técnicas estadísticas avanzadas para analizar datos y construir modelos predictivos. Las matemáticas y estadísticas son fundamentales para **el análisis de datos** en Data Science.
3. **Substantive Expertise (Conocimiento especializado en el Dominio)**: Esta área se refiere al conocimiento específico en un dominio particular, como la medicina, las finanzas, la biología, etc. Tener un **entendimiento sustantivo en un campo específico** es importante para contextualizar los resultados de Data Science y tomar decisiones basadas en datos en ese dominio.

Diagrama de Venn propuesto por O'Reilly para Data Science. Intersecciones

- **Machine Learning:** Representa la combinación de habilidades de programación y conocimientos en matemáticas y estadísticas. En este espacio, se desarrollan **modelos de Machine Learning para el análisis y la predicción de datos**.
- **Danger Zone (Zona de Peligro):** Esta área señala que el conocimiento de programación y habilidades técnicas sin matemáticas y estadísticas sólidas puede ser arriesgado, ya que podría llevar a **conclusiones incorrectas o sesgadas en el análisis de datos**.
- **Traditional Research (Investigación Tradicional):** Muestra que el conocimiento del sujeto y las habilidades de hacking pueden ser útiles en **investigaciones tradicionales** que no requieren un enfoque cuantitativo, como la **recopilación de datos de campo**.
- **Data Science:** En esta intersección, se **combinan las tres áreas clave: habilidades de programación, conocimientos en matemáticas y estadísticas, y Conocimiento especializado en el Dominio**. Es donde se produce una verdadera **Data Science efectiva**, con la capacidad de abordar problemas complejos y extraer información valiosa de los datos.

Fases en un proyecto de Data Science

- **Comprensión del negocio:** En esta fase, se recopila información sobre el negocio y los objetivos del proyecto. Esto incluye entender las necesidades de los clientes, los procesos comerciales y los datos disponibles.
- **Preparación de los datos:** En esta fase, se limpian, transforman y preparan los datos para el análisis. Esto incluye identificar y eliminar los datos faltantes, duplicados o erróneos.
- **Análisis de los datos:** En esta fase, se utilizan técnicas de análisis de datos para explorar los datos y obtener insights. Esto puede incluir técnicas de aprendizaje automático, estadísticas y visualización de datos.
- **Presentación de los resultados:** En esta fase, se comunican los resultados del análisis a los stakeholders. Esto puede incluir la creación de informes, presentaciones o modelos predictivos.
- **Implementación de los resultados:** En esta fase, se implementan los resultados del análisis para mejorar el negocio. Esto puede incluir cambios en los procesos comerciales, productos o servicios.

Estas fases no siempre son secuenciales y pueden repetirse a medida que se avanza en el proyecto. Por ejemplo, es posible que sea necesario volver a la fase de preparación de los datos si se descubren nuevos problemas con los datos.

Fases en un proyecto de Data Science.

Comprensión del negocio

- La comprensión del negocio es una fase fundamental en cualquier proyecto de Data Science. En esta fase, se recopila información sobre el negocio y los objetivos del proyecto. Esto incluye entender las necesidades de los clientes, los procesos comerciales y los datos disponibles.
- Para comprender las necesidades de los clientes, se pueden realizar entrevistas, encuestas o grupos focales. Para entender los procesos comerciales, se pueden analizar los documentos y los procesos existentes. Para entender los datos disponibles, se pueden realizar análisis exploratorios de los datos.



Fases en un proyecto de Data Science.

Preparación de los datos

- La preparación de los datos es una fase importante para garantizar que los datos sean precisos y fiables. En esta fase, se limpian, transforman y preparan los datos para el análisis.
- La limpieza de datos consiste en identificar y eliminar los datos faltantes, duplicados o erróneos. La transformación de datos consiste en cambiar el formato o la estructura de los datos para que sean más adecuados para el análisis. La preparación de datos consiste en organizar los datos de forma que sean fáciles de analizar.



Fases en un proyecto de Data Science.

Análisis de los datos

- El análisis de los datos es la fase en la que se utilizan técnicas de análisis de datos para explorar los datos y obtener insights. Esto puede incluir técnicas de aprendizaje automático, estadísticas y visualización de datos.
 - El aprendizaje automático es un campo de la inteligencia artificial que permite a las máquinas aprender de los datos sin ser explícitamente programadas. Las estadísticas son un conjunto de herramientas y técnicas para recopilar, organizar, analizar e interpretar datos. La visualización de datos es el uso de gráficos y diagramas para comunicar información de los datos.
-



Fases en un proyecto de Data Science.

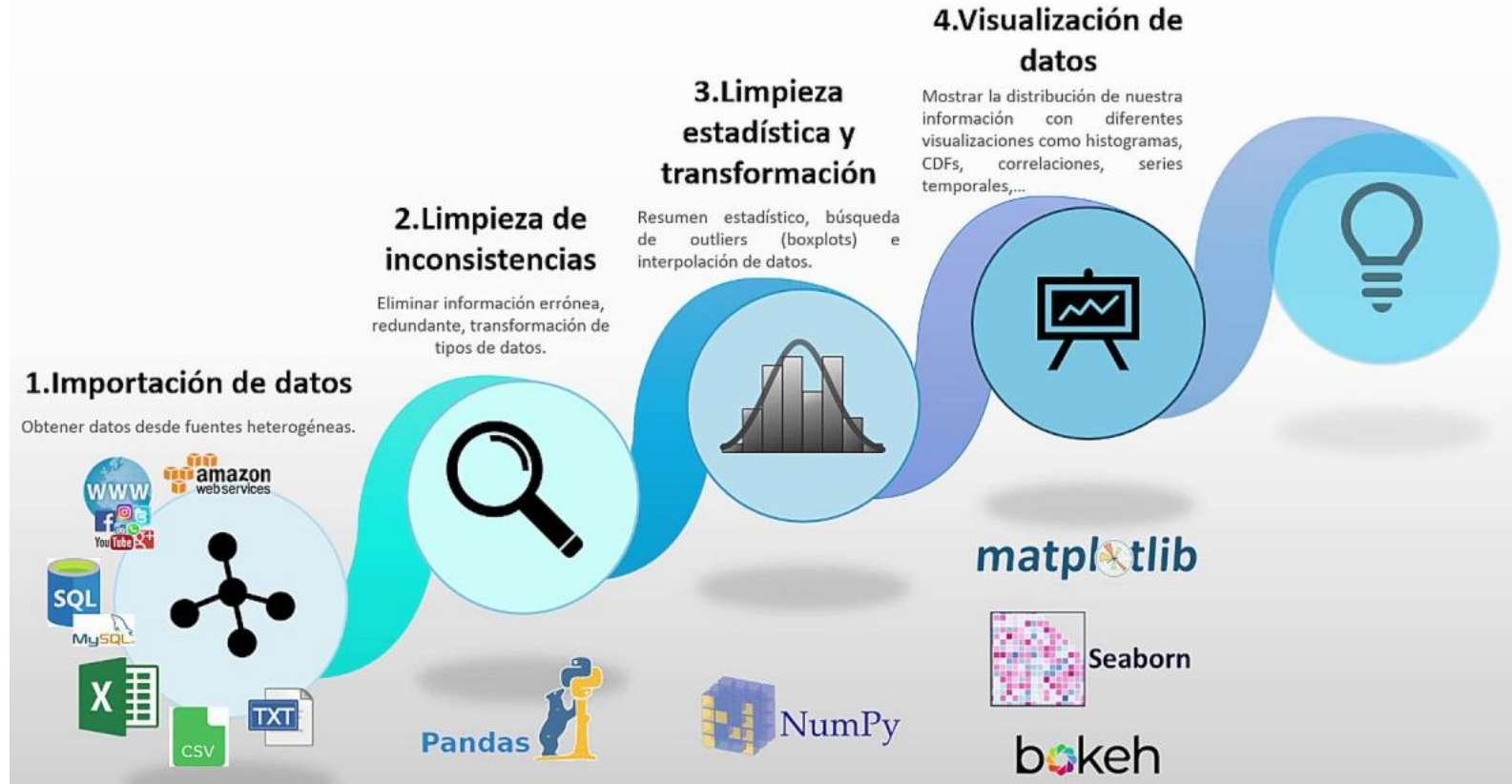
Presentación de los resultados

- La presentación de los resultados es una fase importante para comunicar los resultados del análisis a los stakeholders. Esto puede incluir la creación de informes, presentaciones o modelos predictivos.
 - Los informes son documentos que resumen los resultados del análisis. Las presentaciones son charlas o conferencias que explican los resultados del análisis. Los modelos predictivos son modelos que se pueden utilizar para predecir resultados futuros.
-

Fases en un proyecto de Data Science.

Implementación de los resultados

- La implementación de los resultados es la fase en la que se implementan los resultados del análisis para mejorar el negocio. Esto puede incluir cambios en los procesos comerciales, productos o servicios.
- Para implementar los resultados, es importante tener en cuenta las limitaciones del análisis y las necesidades del negocio. También es importante realizar pruebas para garantizar que los resultados se implementen correctamente.



Fases en un
proyecto de
Data Science.
Python

Herramienta de trabajo Google Colab (Google Colaboratory)

Google Colab es una herramienta para escribir y ejecutar código Python en la nube de Google.

También es posible incluir texto enriquecido, enlaces e imágenes. En caso de necesitar altas prestaciones de cómputo, el entorno permite configurar algunas propiedades del equipo sobre el que se ejecuta el código.

En definitiva, el uso de **Google Colab** permite disponer de un entorno para llevar a cabo tareas que serían difíciles de realizar en un equipo personal. Por otro lado, siguiendo la idea de **Google Drive**, **Google Colab** brinda la opción de compartir los códigos realizados lo que es ideal para trabajos en equipo.

Herramienta de trabajo Google Colab (Google Colaboratory)

Google es bastante agresivo en la investigación de IA. Durante muchos años, desarrolló un marco de inteligencia artificial llamado **TensorFlow** y una herramienta de desarrollo llamada Colaboratory. Hoy TensorFlow es de código abierto y, desde 2017, Google hizo que Colaboratory fuera gratuito para uso público. Colaboratory ahora se conoce como Google Colab o simplemente Colab.

Otra característica atractiva que ofrece Google a los desarrolladores es el uso de GPU y TPU.

Una **CPU**, o Unidad Central de Procesamiento, es un procesador de propósito general.

Herramienta de trabajo Google Colab (Google Colaboratory)

Una **GPU** es una **Unidad de Procesamiento Gráfico**, que es principalmente usada por los gamers o los desarrolladores de videojuegos. Una GPU es más poderosa que una CPU, pues su arquitectura está pensada para procesar de manera eficiente datos más complejos.

En 2013 Google comenzó a desarrollar la primera versión de lo que hoy se conoce como **TPU (Tensor Processing Unit**, o unidad de procesamiento de tensores). Esta es una arquitectura de computador pensada específicamente para el desarrollo de modelos de Machine Learning

Herramienta de trabajo Google Colab (Google Colaboratory)

Google Colab permite ejecutar código Python en la nube. Para ello usa el formato de Jupyter Notebook, que es simplemente como una versión de Python enriquecida: además del código Python, es posible agregar texto con diferentes formatos o imágenes.

Para utilizar Google Colab es necesario tener una cuenta de Google. En nuestro caso crearemos la cuenta en Google:

nombre.apellido.aepi@gmail.com

en el siguiente enlace:

<https://accounts.google.com/signup/v2/createaccount?continue=https%3A%2F%2Faccounts.google.com%2FManageAccount%3Fnc%3D1&biz=false&theme=glif&flowName=GlifWebSignIn&flowEntry=SignUp>



Crea una cuenta de Google

Introduce tu nombre

Nombre

Pedro

Apellidos (opcional)

Bonillo Ramos

Siguiente

Español (España)



[Ayuda](#)

[Privacidad](#)

[Términos](#)



Información básica

Introduce tu fecha de nacimiento y tu sexo.

Día	Mes	Año
<input type="text" value="2"/>	<input type="text" value="Septiemb"/>	<input type="text" value="1972"/>
Género		
<input type="text" value="Hombre"/>		

Siguiente

Español (España)



Ayuda

Privacidad

Términos



Elige tu dirección de Gmail

Elige una dirección de Gmail o crea otra

☐ pbonilloramos@gmail.com

☐ bonilloramosp@gmail.com

☒ Crear dirección de Gmail personalizada

Crea una dirección de Gmail

pedro.bonillo.aepi

@gmail.com

Puedes utilizar letras, números y puntos

[Usar tu correo electrónico](#)

[Siguiente](#)



Crea una clave segura

Crea una contraseña segura con una combinación de letras, números y símbolos

Contraseña

.....

Confirmación

.....

☐

Mostrar contraseña

Siguiente

Español (España)



Ayuda

Privacidad

Términos



Añade un correo de recuperación

Dirección en la que Google puede ponerse en contacto contigo si detecta actividad inusual en tu cuenta o no puedes iniciar sesión.

Dirección de correo electrónico de recuperaci...
pbonillo@gmail.com

Siguiente

Saltar

Español (España)



Ayuda

Privacidad

Términos



Añadir el teléfono



Número de teléfono

615702581

Google solo usará este número para mantener la seguridad de la cuenta. No lo mostrará a otros usuarios. Más tarde podrás elegir si quieres que se use con otros fines.

Siguiente

Saltar

Español (España)



Ayuda


Privacidad

Términos



Verifica tu teléfono

Para comprobar que el número sea tuyo, te enviaremos un SMS con un código de verificación de 6 dígitos. *SMS sujeto a tarifas estándar*

 615 70 25 81

[Atrás](#)

[Ahora no](#)

[Enviar](#)

Español (España)



[Ayuda](#)

[Privacidad](#)

[Términos](#)



Verifica tu teléfono

Para comprobar que este número es tuyo, Google te enviará un SMS con un código de verificación de 6 dígitos.

 615 70 25 81

Introduce el código de verificación

G- 322460

Se aplicarán tarifas estándar.

[Atrás](#)

[Verificar](#)

Español (España) ▼

[Ayuda](#)

[Privacidad](#)

[Términos](#)



Verifica tu teléfono

Para comprobar que este número es tuyo, Google te enviará un SMS con un código de verificación de 6 dígitos.

 615 70 25 81

Introduce el código de verificación

G- 322460

Se aplicarán tarifas estándar.

[Atrás](#)

[Verificar](#)

Español (España) ▼

[Ayuda](#)

[Privacidad](#)

[Términos](#)



Sácale el máximo partido a tu número

Si quieres, puedes añadir tu número de teléfono a tu cuenta para usarlo en los servicios de Google. [Más información](#)

Por ejemplo, tu número se utilizará para

☐ Recibir videollamadas y mensajes

☒ Hacer que los servicios de Google, incluidos los anuncios, te resulten más relevantes


[Más opciones](#)

[Atrás](#)



Revisa la información de tu cuenta

Puedes usar esta dirección de correo o este número de móvil para iniciar sesión en otro momento

 Pedro Bonillo Ramos
pedro.bonillo.aepi@gmail.com

Número de móvil de recuperación
615 70 25 81

Siguiente

Español (España)



[Ayuda](#)

[Privacidad](#)

[Términos](#)



Elige los ajustes de personalización

- ☒ **Personalización rápida (1 paso)**
Usa ajustes de personalización que ofrecen contenido y anuncios personalizados. Te recordaremos que revises tus ajustes en un par de semanas.
- ☐ **Personalización manual (5 pasos)**
Configura los ajustes de personalización paso a paso. Tú decides qué ajustes quieres activar y cuáles no para tener la experiencia de contenido y anuncios que prefieras.

Puedes cambiar la configuración en cualquier momento en account.google.com

Siguiente

funcion de tus ajustes, en Google y en la Web

Con respecto al contenido y los anuncios no personalizados, lo que veas puede estar influido por factores como lo que estés viendo en ese momento y tu ubicación (el servicio de anuncios se basa en la ubicación general). El contenido y los anuncios personalizados se pueden basar, además de en esos criterios, en tu actividad, como las búsquedas que hagas en Google y los vídeos que veas en YouTube. Los anuncios y el contenido personalizados incluyen, entre otras cosas, resultados y recomendaciones más relevantes, una página de inicio de YouTube personalizada y anuncios adaptados a tus intereses.

Puedes cambiar la configuración del navegador para rechazar algunas o todas las cookies.



Recordatorio de privacidad

Te enviaremos un recordatorio en un par de semanas para que revises estos ajustes

[Atrás](#)

[Confirmar](#)

Español (España) ▼

[Ayuda](#)

[Privacidad](#)

[Términos](#)

derechos y su seguridad o se dañen sus bienes en la medida exigida o permitida por la ley, lo que incluye la divulgación de información a autoridades gubernamentales.

- Llevar a cabo investigaciones que mejoren nuestros servicios para los usuarios y beneficien al público en general.
- Cumplir las obligaciones que hemos asumido frente a nuestros partners como, por ejemplo, desarrolladores y titulares de derechos.
- Hacer valer nuestros derechos legales, lo que incluye la investigación de posibles infracciones de los Términos del Servicio aplicables.

Puedes ir a tu cuenta de Google (account.google.com) para hacer una Revisión de Privacidad o ajustar los controles de privacidad.

¿Tienes alguna duda? [Contacta con nosotros](#)

[Cancelar](#)

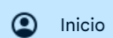
[Acepto](#)

Español (España) ▼

[Ayuda](#)

[Privacidad](#)

[Términos](#)



Inicio



Información personal



Datos y privacidad



Seguridad



Contactos y compartir



Pagos y suscripciones



Información general

[Privacidad](#) [Términos](#) [Ayuda](#)[Información](#)

Bienvenido, Pedro Bonillo Ramos

Gestiona tu información, la privacidad y la seguridad para mejorar tu experiencia en Google. [Más información](#)

Privacidad y personalización

Consulta los datos almacenados en tu cuenta de Google y elige qué actividad se debe guardar para personalizar tu experiencia en Google





[Gestionar tus datos y tu privacidad](#)


Hay consejos de seguridad para ti

Consejos de seguridad encontrados en la Revisión de Seguridad




[Revisar consejos de seguridad](#)

<https://colab.research.google.com/?hl=es>

← → ↻ colab.research.google.com/?hl=es     ⋮

 Te damos la bienvenida a Colaboratory

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda

 Compartir  


Conectar ▾ ^



as, la vista del historial

Te damos la bienvenida a Colaboratory

Abierto por última vez ▲

Abierto por primera vez ▼



 [Te damos la bienvenida a Colaboratory](#) 6:41 6:41 

+ Nuevo cuaderno Cancelar

con las siguientes

Índice

- Primeros pasos
- Ciencia de datos
- Aprendizaje automático
- Más recursos
- Ejemplos destacados
- Sección

Recientes

Google Drive

GitHub

Subir

colab.research.google.com/?hl=es#scrollTo=-Rh3-Vt9Nev9

☆

P

CO

Te damos la bienvenida a Colaboratory

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda

Compartir

P

+ Código

+ Texto

Copiar en Drive

✓

RAM

Disco

¿Qué es Colaboratory?

Colab, también conocido como "Colaboratory", te permite programar y ejecutar Python en tu navegador con las siguientes ventajas:

- No requiere configuración
- Acceso a GPUs sin coste adicional
- Permite compartir contenido fácilmente

Colab puede facilitar tu trabajo, ya seas **estudiante**, **científico de datos** o **investigador de IA**. No te pierdas el vídeo de [Introducción a Colab](#) para obtener más información. O simplemente empieza con los pasos descritos más abajo.

▼ Primeros pasos

El documento que estás leyendo no es una página web estática, sino un entorno interactivo denominado **cuaderno de Colab** que te permite escribir y ejecutar código.
Por ejemplo, a continuación se muestra una **celda de código** con una breve secuencia de comandos de Python que calcula un valor, lo almacena en una variable e imprime el resultado:

✓ Conectado a del backend de Google Compute Engine que utiliza Python 3

X

Herramienta de trabajo Google Colab (Google Colaboratory)

- Los cuadernos de Colab permiten combinar **código ejecutable y texto enriquecido** en un mismo documento, además de **imágenes, HTML, LaTeX** y mucho más.
- Los cuadernos que se crean en Colab se almacenan en la cuenta de Google Drive. Se puede compartir los cuadernos de Colab fácilmente con compañeros de trabajo o amigos, lo que permite comentarlos o incluso editarlos.
- Para crear un cuaderno de Colab, se debe usar el menú Archivo que aparece arriba o bien acceder al enlace para [crear un cuaderno de Colab](#).
- Los cuadernos de Colab son cuadernos de Jupyter alojados en Colab.



Nuevo cuaderno

Ctrl+O

Subir cuaderno

Guardar una copia en Drive

Guardar una copia como Gist de GitHub

Guardar una copia en GitHub

Guardar

Ctrl+S

Descargar

Imprimir

Ctrl+P

```
[ ] import numpy as np
    from matplotlib import pyplot as plt

    ys = 200 + np.random.randn(100)
```

o **ejecutable** y **texto enriquecido** en un mismo documento, además de **imágenes**, **HTML**,
Colab se almacenan en tu cuenta de Google Drive. Puedes compartir tus cuadernos de Colab
que les permite comentarlos o incluso editarlos. Consulta más información en [Información](#)
b, puedes usar el menú Archivo que aparece arriba o bien acceder al enlace para [crear un](#)
alojados en Colab. Para obtener más información sobre el proyecto Jupyter, visita

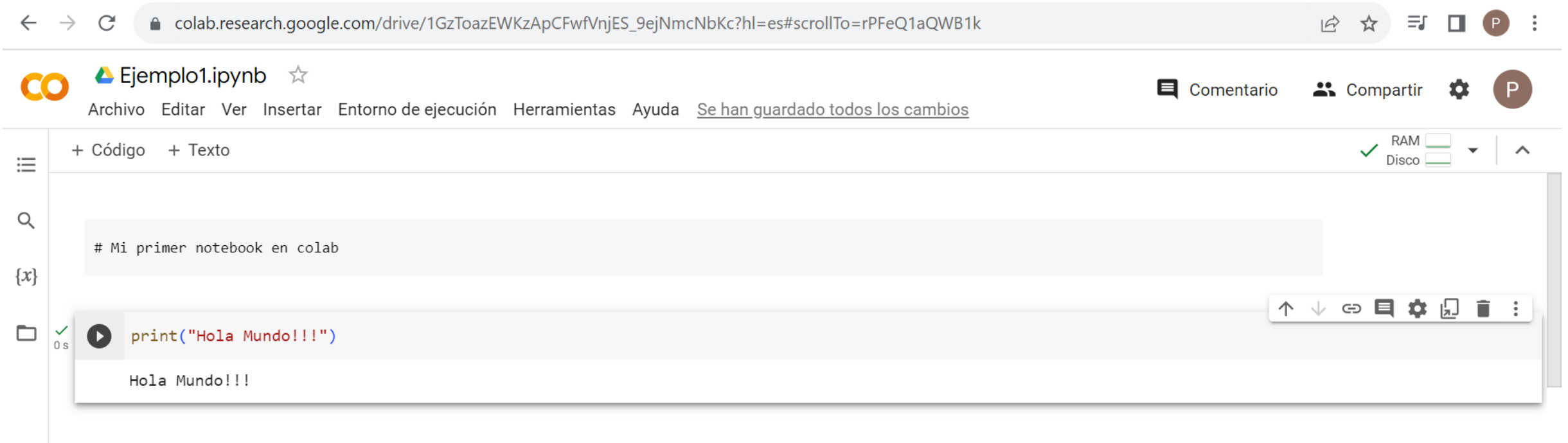
Imprimir Ctrl+P

Las bibliotecas más populares de Python para analizar y visualizar datos. La celda de código que abajo utiliza **Numpy** para generar datos aleatorios y **Matplotlib** para visualizarlos. Para editar el código, solo tienes que hacer clic en la celda.

✓ 0 s completado a las 7:20



Herramienta de trabajo Google Colab (Google Colaboratory)



The screenshot displays the Google Colab web interface. At the top, the browser address bar shows the URL `colab.research.google.com/drive/1GzToazEWKzApCFwfVnjES_9ejNmcNbKc?hl=es#scrollTo=rPFQ1aQWB1k`. Below the browser, the Colab logo is followed by the notebook title "Ejemplo1.ipynb" and a star icon. A menu bar contains the options: Archivo, Editar, Ver, Insertar, Entorno de ejecución, Herramientas, and Ayuda. A status message reads "Se han guardado todos los cambios". On the right side of the header, there are icons for "Comentario", "Compartir", settings, and a user profile icon labeled "P".

The main workspace shows a notebook with two cells. The first cell is a text cell containing the comment `# Mi primer notebook en colab`. The second cell is a code cell containing the Python code `print("Hola Mundo!!!")`. This code cell has a play button icon on its left and a status indicator showing "0 s" (0 seconds) and a green checkmark. The output of the code cell is displayed below the code: "Hola Mundo!!!".

On the far left, a sidebar contains icons for a menu, search, a variable `{x}`, and a file explorer. In the top right corner of the workspace, there are status indicators for "RAM" and "Disco" (Disk), both showing green checkmarks and progress bars. A small toolbar with icons for undo, redo, link, comment, settings, insert, delete, and a menu is located at the bottom right of the code cell.

Fundamentos generales de Python para la ciencia de datos

¿Por qué Python?

- **Legibilidad:**
 - Sintaxis limpia
 - Fácil de comprender
- **Rápida codificación:**
 - Estructuras de datos amigables.
 - Lenguaje listo para ejecutar (**interpretado**)
 - Interprete de python interactivo
- **Reusabilidad:**
 - Permite compartir funcionalidad entre programas usando módulos y paquetes
- **Portabilidad:**
 - Python tiene la misma interfaz en múltiples plataformas: Linux, Windows, MacOS, etc.
- **Código Abierto, Gratis, y Desarrollo Comunitario:**
 - La especificación y la implementación es open-source.
- **Orientado a Objeto:**
 - Conceptos y características disponibles pero no obligatorio.

Fundamentos generales de Python para la ciencia de datos

- **Tipado dinámico:**
 - Una misma variable puede tomar valores de distinto tipo en distintos momentos.
- **Resolución dinámica de nombres:**
 - Métodos, variables son enlazadas con su lógica durante la ejecución.
- **Biblioteca estándar de Python es extensa y bien documentada:**
 - Conocido como "batteries included"
- **Gran variedad de librerías de terceros:**
 - Más de 130.000 paquetes en un gran rango de funcionalidades, incluyendo: GUI, Web, Multimedia, Base de datos, Redes, Testing, Automatización, Web Scraping, Procesamiento de texto, Procesamiento de imágenes, etc.
- **Sponsorizado por grandes empresas:**
 - Facebook, Google, Amazon, Redhat, Microsoft, etc.

Legibilidad

- Más uso de palabras:
 - `!`, `||`, `&&` → not, or, y and
- **Indentación (sangrado):**
 - En lugar de usar llaves o palabras claves para delimitar bloques de código, usa espacios o tabuladores.

Bash

```
if [ $x -ge 18 ]; then
    echo "Es mayor"
else
    echo "Es menor"
fi
```

Python

```
if x >= 18: print("Es
            mayor")
else:
    print("Es      menor")
```

Comentarios y Variables

- **Comentarios:**

```
'''
```

```
Comentario más largo en una línea en Python '''
```

```
print("Hola mundo") # Al final de una línea de código
```

- **Variables:**

- Las variables no necesitan ser declaradas, no se tiene que especificar cuál es su tipo.
- Las variables son creadas cuando se les asigna valores. Se usa el símbolo = para asignar valores.
- Deben ser asignadas antes de ser usadas.
- Se les puede asignar cualquier tipo de dato.

```
x = 1
```

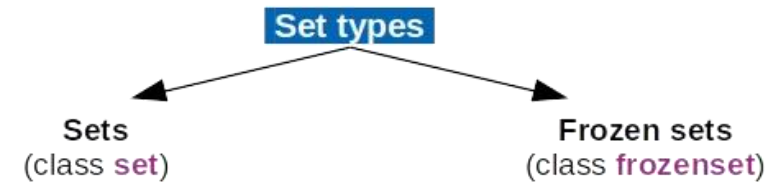
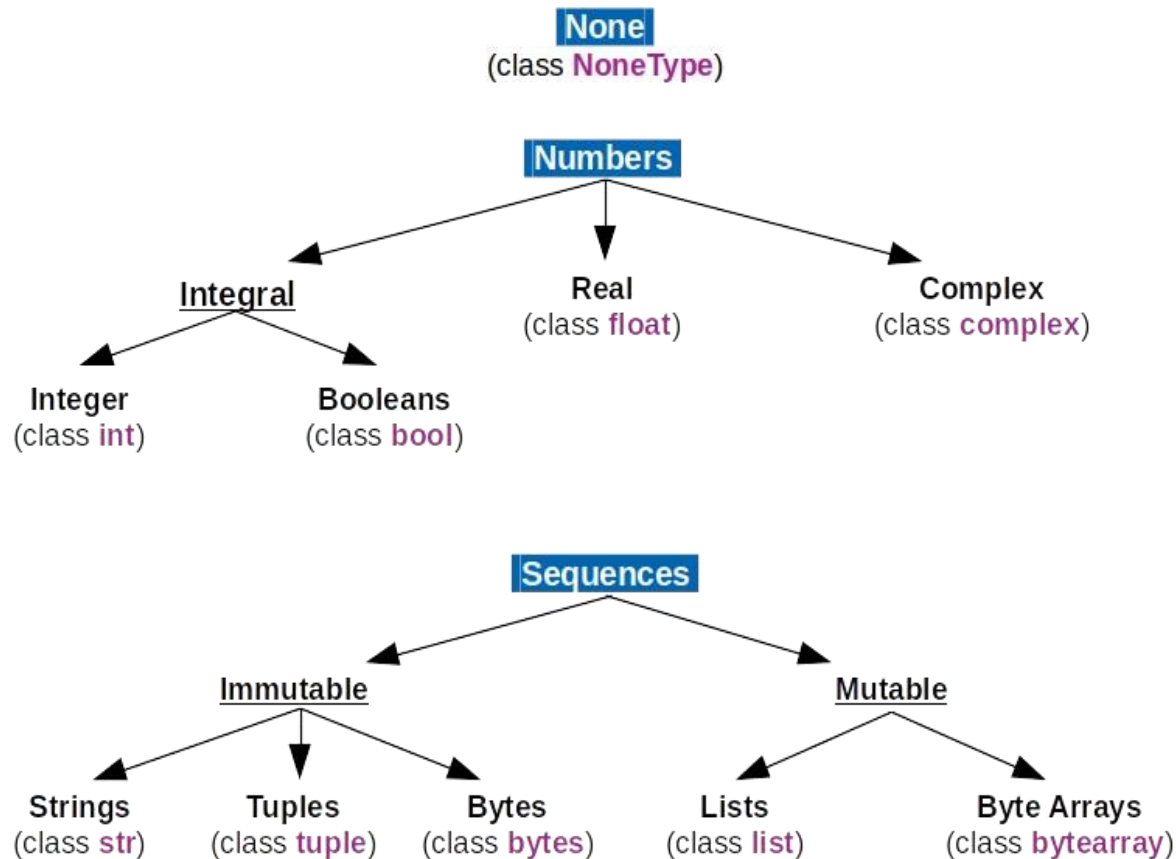
```
x = "texto" # Posible porque los tipos son  
# asignados dinámicamente
```


Tipo de datos

Tipo	Clase	Notas	Ejemplo
NoneType	None	Representa la ausencia de valor	None
bool	Numbers	Valor booleano verdadero o falso	True o False
int	Numbers	Número entero de tamaño ilimitado	42
float	Numbers	Número real; coma flotante	3.1415927
complex	Numbers	Número complejo con parte real y parte imaginaria j	4.5 + 3j
str	Sequences	Cadena en formato unicode. Inmutable	"Texto"
list	Sequences	Secuencia de datos, pueden ser de diversos tipos. Mutable.	[4.0, 'Cadena', True]
tuple	Sequences	Secuencia de datos, pueden ser de diversos tipos. Inmutable.	(4.0, 'Cadena', True)
set	Set Types	Conjunto de datos, sin orden, no contiene duplicados. Mutable	set([4.0, 'Cadena', True])
frozenset	Set Types	Conjunto de datos, sin orden, no contiene duplicados. Inmutable	frozenset([4.0, 'Cadena', True])
dict	Mappings	Diccionario de pares clave:valor (array asociativo)	{'key1': 1.0, 'key2': False}
bytearray	Binary Sequences	Secuencia de bytes. Mutable	bytearray([119, 105, 107, 105])
bytes	Binary Sequences	Secuencia de bytes. Inmutable	bytes([119, 105, 107, 105])

Jerarquía de Tipos de Datos

Python 3
The standard type hierarchy



Mappings

Dictionaries
(class dict)

Callable

< Functions, Methods, Classes >

Modules

Tipos Numéricos

- Números enteros (int):
 - Decimal: 24, 60
 - Binario: 0b010011, 0b1101
 - Hexadecimal: 0x18, 0x3cf4
 - Octal: 0o30, 0o74
- Números de punto flotante (float): números reales y la precisión depende del equipo.
 - 3.141595
 - 12.
 - -45.3556
 - 2,0/3,0
- Números complejos (complex):
 - $6.32 + 45j$
 - $0.117j$
 - $(2 + 0j)$
 - $1j$
- Valores booleanos (bool): Se usa para expresiones lógicas
 - False (equivale al número 0)
 - True (cualquier otro valor diferente de cero y 1 por defecto)

String y None

- **str** (cadena de caracteres):
 - 'Wikipedia'
 - "Wikipedia"
 - """Con
múltiples
líneas"""
- **None**:
 - El tipo None representa un valor "vacío".
 - a = None

Listas

- **Listas (array indexado):**

- Es la secuencia más general en python.
- Mutables; se puede cambiar su contenido en tiempo de ejecución.
- Para declarar una lista se usan los corchetes [] y los elementos se separan por comas.
- Pueden contener elementos de diferentes tipos.
- No tienen un tamaño fijo.
- Los elementos son ordenados por la posición.
- Para acceder a los elementos se utiliza un índice entero (empezando por "0", no por "1"). Se pueden utilizar índices negativos para acceder elementos a partir del final.

Listas

- Crear una lista:

```
lista = ["abc", 42, 3.1415]
```

- Acceder a un elemento por su índice:

```
lista[0]
```

```
'abc'
```

- Acceder a un elemento usando un índice negativo:

```
lista[-1]
```

```
3.1415
```

- Añadir un elemento al final de la lista:

```
lista.append(True)
```

```
lista
```

```
['abc', 42, 3.1415, True]
```

Listas

- Re-asignar el valor del primer elemento de la lista:

```
lista[0] = "xyz"
```

- Borrar un elemento de la lista:

```
lista.remove(True)
```

```
del lista[0]
```

- Mostrar una sublista:

```
lista[0:2]    # Del índice "0" al "2" (sin incluir este último)
```

```
['xyz', 42]
```

- Listas anidadas (una dentro de otra):

```
lista_anidada = [lista, [True, 42]]
```

```
lista_anidada
```

```
[['xyz', 42, 3.1415], [True, 42]]
```

```
lista_anidada[1][0]
```

```
True
```

Listas

```
lista = [22, True, "a list", [1, 2]]
```

```
lista[0] =>??
```

```
lista[2][4] =>??
```

```
lista[-1][-2] =>??
```

```
lista[0:3] =>??
```

```
lista[:3] =>??
```

```
lista.append('DevOps') =>??
```

```
lista.insert(0, "a list") =>??
```

```
lista.remove("a list") =>??
```


Listas

```
lista = [22, True, "a list", [1, 2]]
```

```
lista[0] => 22
```

```
lista[2][4] => 's'
```

```
lista[-1][-2] => 1
```

```
lista[0:3] => [22, True, 'a list']
```

```
lista[:3] => [22, True, 'a list']
```

```
lista.append('DevOps') => [22, True, 'a list', [1, 2], 'DevOps']
```

```
lista.insert(0, "a list") => ['a list', 22, True, 'a list', [1, 2], 'DevOps']
```

```
lista.remove("a list") => [22, True, 'a list', [1, 2], 'DevOps']
```

Tuplas

- **Tuplas:**

- Es otra secuencia en python como las listas.
- **Inmutables**; no se puede cambiar su contenido en tiempo de ejecución.
- Para declarar una lista se usan los **paréntesis ()** y los elementos se separan por comas. **Es necesario que tengan como mínimo una coma. También se pueden declarar sin los paréntesis.**
- Pueden contener elementos de diferentes tipos.
- Pueden definirse de cualquier tamaño.
- Los elementos son ordenados por la posición.
- Para acceder a los elementos se utiliza un índice entero (empezando por "0", no por "1"). Se pueden utilizar índices negativos para acceder elementos a partir del final.

Tuplas

- Crear una tupla:

```
tupla = ("abc", 42, 3.1415)
```

- Acceder a un elemento por su índice:

- `tupla[0]`

`'abc'`

- Acceder a un elemento usando un índice negativo:

- `tupla[-1]`

`3.1415`

- No es posible modificar la tupla:

- `del tupla[0]`

(Excepción)

- `tupla[0] = "xyz"`

(Excepción)

Tuplas

- Mostrar una sub-tupla:

```
tupla[0:2]  # Del índice "0" al "2" (sin incluir este último)
('abc', 42)
```

- Tuplas anidadas (una dentro de otra):

```
tupla_anidada = (tupla, (True, 3.1415))
(('abc', 42, 3.1415), (True, 3.1415))
tupla_anidada[1][0]
True
```

- También es una tupla:

```
1, 2, 3, "abc"
(1,) #Ojo (1) no es una tupla
(1, 2,)
```

Tuplas

- La inmutabilidad se puede omitir si una nueva estructura es enlazada a la tupla original
 - `>>> t = 10,15,20`
 - `>>> t = t[0],t[2]`
 - `>>> t`
 - `(10,20)`

Diccionarios

- **Diccionarios (array asociativo):**

- **Mutables**; se puede cambiar su contenido en tiempo de ejecución.
- Para declarar un diccionario se usan las llaves `{}`. Contienen elementos separados por comas, donde cada elemento está formado por un par **clave:valor** (el símbolo `:` separa la clave de su valor correspondiente).
- Las claves de un diccionario deben ser inmutables (strings, números, o tuplas)
- El valor asociado a una clave puede ser de cualquier tipo de dato, incluso un diccionario.
- No tienen un tamaño fijo.
- Indexados por la clave.

Diccionarios

- Crear un diccionario:

```
diccionario = {"cadena": "abc", "numero": 42, "lista": [True, 42]}
```

- Acceder a un elemento por su clave:

- `diccionario["cadena"]`

`'abc'`

- `diccionario["lista"][0]`

`True`

- Insertar un nuevo elemento clave:valor:

```
diccionario["decimal"] = 3.1415927
```

Diccionarios

- Re-asignar el valor del primer elemento de la lista:

```
diccionario["cadena"] = "xyz"
```

- Borrar un elemento de la lista:

- del diccionario["cadena"]

- También es posible que un valor sea un diccionario

```
diccionario_mixto = {"tupla": (True, 3.1415), "diccionario": diccionario}
```

```
diccionario_mixto["diccionario"]["lista"][1]
```


Diccionarios

```
dic = {'e': 2.718, 'pi': 3.141, 'fi': 1.618}
```

```
>>> dic['e']
```

```
?
```

```
# Actualizar el valor de pi a 3.141592
```

```
>>> ?
```

```
>>> dic.keys()
```

```
?
```

```
>>> dic.values()
```

```
?
```

```
>>> dic.items()
```

```
?
```

Diccionarios

```
dic = {'e': 2.718, 'pi': 3.141, 'fi': 1.618}
>>> dic['e']
2.718

# Actualizar el valor de pi a 3.141592
>>> dic['pi'] = 3.141592
>>> dic.keys()
dict_keys(['e', 'pi', 'fi'])
>>> dic.values()
dict_values([2.718, 3.141, 1.618])
>>> dic.items()
dict_items([('e', 2.718), ('pi', 3.141), ('fi', 1.618)])
```

Diccionarios

- Operaciones comunes con Diccionarios:
 - Agregar un elemento (update):
 - `dic.update({"d":4})`
 - Crear una copia del diccionario (copy):
 - `nuevodic = dic.copy()`
 - Eliminar todos los elementos de un diccionario:
 - `dic.clear()`

Conjuntos

- **Conjuntos:**

- Los conjuntos se construyen mediante **set(items) / frozenset(items)** donde items es cualquier objeto iterable, como listas o tuplas.
 - set para conjuntos mutables y
 - frozenset para conjuntos inmutables.
- Los conjuntos no mantienen el orden ni contienen elementos duplicados.
- Se suelen utilizar para eliminar duplicados de una secuencia, o para operaciones matemáticas como intersección, unión, diferencia y diferencia simétrica.

Conjuntos

- Crear conjuntos:

```
conjunto1 = set(["a", "b", "a"])
```

```
conjunto2 = set(["a", "b", "c", "d"])
```

```
conjunto_inmutable = frozenset(["a", "b", "a"])
```

- Intersección

```
conjunto1 & conjunto2
```

```
set(['a', 'b'])
```

- Unión

```
conjunto1 | conjunto2
```

```
set(['a', 'c', 'b', 'd'])
```

Conjuntos

- Diferencia (1)

conjunto1 - conjunto2

set([])

- Diferencia (2)

conjunto2 - conjunto1

set(['c', 'd'])

- Diferencia simétrica

conjunto1 ^ conjunto2

set(['c', 'd'])

Bytes, Bytearray

- Para manejar datos binarios python incluye los tipos bytes y bytearray.
- El tipo bytes es una secuencia inmutable de bytes, conceptualmente similar a una cadena. Y el tipo bytearray es una secuencia mutable de bytes.
- Representan a un carácter conforme a su número correspondiente en el código ASCII y se definen anteponiendo la letra b a los apostrofes o comillas.

b'<texto>'

b"<texto>"

Bytes

- Ejemplos:

```
palabra = b"Hola" // palabra = bytes([72,111,108,97])
```

```
palabra[0]
```

```
72
```

```
palabra[2:4]
```

```
b'la'
```

- str -> bytes.

```
bytes('hola', "utf-8")
```

- bytes -> str

```
str(b'hola'[1:3], 'ascii')
```


Bytearray

- Crear un bytearray desde un bytes:

```
x = bytearray(b"Python Bytes")
```

- Crear un bytearray desde un string:

- `x = bytearray("Python Bytes", "utf8")`

- Crear un bytearray desde una lista de enteros:

- `x = bytearray([94, 91, 101, 125, 111, 35, 120])`

Rebanadas de secuencias (slice)

- `secuencia[x:y:z]`
 - Desde x, hasta y sin incluir dicha posición y con incrementos de z
- Ejemplos de `t=(1,2,3,4,5)`
 - `t[2:4]` → Desde 2 hasta 3
 - `t[:3]` → Desde el comienzo hasta 2 `t[3:]` → Desde 3 hasta el final
 - `t[:]` → Desde el comienzo hasta el final `t[::-1]` → En sentido inverso
 - `t[::-2]` → En sentido inverso con incrementos de 2

Condicional: if

- If
 - ∴ if sintaxis:
 - Las clausulas elif y else son opcionales.
 - No existe el operador case en python; se puede usar la estructura if/elif/else.

```
if <condition>:  
    <statements>  
[elif <condition>:  
    <statements>]  
[elif <condition>:  
    • pass]  
...  
[else:  
    <statements>]
```

Condicional: if

- Ejemplos:

- `a=7`

- `if a>6:`

- `print("Es mayor que 6")`

- `if 1: # 1 significa verdadero`

- `print("sip")`

- `if a == 1:`

- `print("1")`

- `elif a == 2:`

- `print("2")`

- `else:`

- `print("Mayor a 2")`

Condicional: if

- **if anidados:**

```
a=1
```

```
b=2
```

```
if a==1:
```

```
    if b==2:
```

```
        print("a es 1 y b es 2")
```

- **if en una línea**

```
if a>4: print("Greater")
```

Bucle: for

```
for <variable> in <secuencia>:
```

```
    <sentencias>
```

```
[else:
```

```
    <sentencias>]
```

- La sentencia else es opcional y siempre se ejecuta al menos que se ejecute la sentencia break dentro del bucle.
- Se itera por cada valor en la secuencia, en cada iteración la variable tomando el valor correspondiente en la secuencia.
- Las secuencias pueden ser: una lista, una tupla, un diccionario, un conjunto o un string.

Bucle: for

- Ejemplos:
- #Iterando con una lista:

```
frutas = ["banana", "uva"]  
for x in frutas:  
    print(x)
```
- #Iterando con cadenas:

```
for x in "banana":  
    print(x)
```
- #usando la función range (genera una secuencia de números):

```
for i in range(6):  
    print(i, end=', ')
```

Bucle: while

while <condición>:

<sentencias>

[**else**:

<sentencias>]

- La sentencia else es opcional y siempre se ejecuta al menos que se ejecute la sentencia break dentro del bucle.
- Se itera mientras la condición sea verdadera.
- Es importante modificar dentro del bucle los elementos que forman la condición para finalizar las iteraciones.

Bucle: while

- Ejemplos:

- `i = 1`

```
while i < 6:
```

```
    print(i)
```

```
    i += 1
```

- #En una línea

```
a=3
```

```
while a>0: print(a); a-=1
```

- `a=3`

```
while(a>0):
```

```
    print(a)
```

```
    a-=1
```

```
else:
```

```
    print("a llego a 0")
```

Bucle anidados

- #Bucles **for** anidados:

```
for i in range(1,6):  
    for j in range(i):  
        print("*",end=' ')  
    print()
```

- #Bucles **while** anidados:

```
i=6  
while(i>0):  
    j=6  
    while(j>i):  
        print("*",end=' ')  
    j-=1  
    i-=1  
    print()
```

Control de bucles

- Modificar el comportamiento normal de los bucles en python: continue y break.
- **continue:**
 - detiene la iteración actual y continua con la siguiente.
- **break:**
 - detiene la iteración actual y todas las restantes.
- En python no se puede especificar que bucle anidado se pretende controlar.

Control de bucles

- #break

```
for i in 'break':  
    print(i)  
    if i=='a': break;
```

- # continue

```
i=0  
while(i<8):  
    i+=1  
    if(i==6): continue  
    print(i)
```

Control de bucles

#Con bucles anidados

```
for x in range(10):  
    for y in range(10):  
        print (x*y)  
        if x*y > 50:  
            break  
    else:  
        continue  
break
```

Funciones

```
def <NombreDeLafunción>(arg1, arg2, ...):  
    """<Texto>""" #Docstring: muy recomendado  
    <sentencias>  
    return <data>  
  
<NombreDeLafunción>(arg1, arg2, ...)
```

- Las funciones ayuda a dividir un programa en módulos. Hace al código más fácil de administrar, depurar y escalar. Reutilización del código.
- Se puede acceder el texto de documentación utilizando el atributo `__doc__` de la función.
- `return` nos permite retornar un valor.

Funciones

- `def hola():`

`"""`

`Esto es el docstring de hola`

`"""`

`print("Hola")`

`hola.__doc__`

`hola()`

- `#Con parámetros:`

`def suma(a,b):`

`print(a+b)`

`suma(1,2)`

- `# Con un valor de retorno`

`def func1(a):`

`if a%2==0:`

`return 0`

`else:`

`return 1`

`func1(7)`

Funciones

- Parámetros (argumentos):
 - No es necesario especificar el tipo de objeto de un argumento,
 - Los argumentos tienen un comportamiento posicional, pero ...
 - ... python también habilita pasar argumentos usando el nombre y su valor independientemente del orden.
 - Se puede especificar valores de los parámetros por defecto. Entonces si no se pasa un argumento, se usa el valor por defecto.
 - Los objetos mutables se pasan por referencia.
 - Los objetos inmutables se pasan por valor.
 - Los módulos, clases, instancias y otras funciones se pueden usar como argumentos y son examinados dinámicamente.
 - La cantidad de argumentos puede ser indefinido.

Funciones

- #Con valores por defecto:
- #Mutables x referencia:

```
def suma(a=1,b=3):
```

```
    print(a+b)
```

```
suma(1,2)
```

```
suma()
```

```
suma(a=7)
```

```
suma(b=10,a=5)
```

```
lista=[1,2]
```

```
def fun(a):
```

```
    a[0]=3
```

```
fun(lista)
```

```
lista[0]
```

. Funciones

- `def suma(a=1,b=3):`
 `print(a+b)`
- `def sumar(f):`
 `f(2,2)`
- # Función como argumento.
`sumar(suma)`

#Arg indeterminados - Posición

```
def ind_posicion(*names):  
    for name in names:  
        print("Hola " + name)  
ind_posicion("Carlos","Marta")
```

#Arg indeterminados –Nombre

```
def ind_nombre(**kwargs):  
    print (kwargs)  
ind_nombre(n=5, c="Hola", l=[1,2,3])
```

Funciones

- Retorno:
 - Detiene la ejecución de una función.
 - Puede ser una expresión.
 - Se pueden devolver múltiples valores usando tuplas.
 - Cuando una función no tiene ninguna sentencia de retorno, se devuelve implícitamente el valor “None”.

```
def sum(a,b):  
    return a+b
```

```
def test():  
    return 'abc', 100
```

Funciones

Built-in Functions				
<code>abs()</code>	<code>delattr()</code>	<code>hash()</code>	<code>memoryview()</code>	<code>set()</code>
<code>all()</code>	<code>dict()</code>	<code>help()</code>	<code>min()</code>	<code>setattr()</code>
<code>any()</code>	<code>dir()</code>	<code>hex()</code>	<code>next()</code>	<code>slice()</code>
<code>ascii()</code>	<code>divmod()</code>	<code>id()</code>	<code>object()</code>	<code>sorted()</code>
<code>bin()</code>	<code>enumerate()</code>	<code>input()</code>	<code>oct()</code>	<code>staticmethod()</code>
<code>bool()</code>	<code>eval()</code>	<code>int()</code>	<code>open()</code>	<code>str()</code>
<code>breakpoint()</code>	<code>exec()</code>	<code>isinstance()</code>	<code>ord()</code>	<code>sum()</code>
<code>bytearray()</code>	<code>filter()</code>	<code>issubclass()</code>	<code>pow()</code>	<code>super()</code>
<code>bytes()</code>	<code>float()</code>	<code>iter()</code>	<code>print()</code>	<code>tuple()</code>
<code>callable()</code>	<code>format()</code>	<code>len()</code>	<code>property()</code>	<code>type()</code>
<code>chr()</code>	<code>frozenset()</code>	<code>list()</code>	<code>range()</code>	<code>vars()</code>
<code>classmethod()</code>	<code>getattr()</code>	<code>locals()</code>	<code>repr()</code>	<code>zip()</code>
<code>compile()</code>	<code>globals()</code>	<code>map()</code>	<code>reversed()</code>	<code>__import__()</code>
<code>complex()</code>	<code>hasattr()</code>	<code>max()</code>	<code>round()</code>	

- Enlaces a más información:
 - [Liberia estándar](#)
 - [Doc en español](#)

Operadores Aritméticos

Operador	Descripción
+	Suma
-	Resta
-	Negativo
*	Multiplicación
**	Exponente
/	División
//	División entera
%	Residuo

Operadores de Asignación

Operador	Descripción	Ejemplo
=	Asignación simple	$x = y$
+=	Suma	$x += y$ equivale a $x = x + y$
-=	Resta	$x -= y$ equivale a $x = x - y$
*=	Multiplicación	$x *= y$ equivale a $x = x * y$
**=	Exponente	$x ** = y$ equivale a $x = x ** y$
/=	División	$x /= y$ equivale a $x = x / y$
//=	División entera	$x //= y$ equivale a $x = x // y$
%=	Residuo de división	$x \% = y$ equivale a $x = x \% y$

Operadores de Relación

Operador	Evalúa
<code>==</code>	$a == b$ ¿a igual a b?
<code>!=</code>	$a != b$ ¿a distinta de b?
<code>></code>	$a > b$ ¿a mayor que b?
<code><</code>	$a < b$ ¿a menor que b?
<code>>=</code>	$a >= b$ ¿a mayor o igual que b?
<code><=</code>	$a <= b$ ¿a menor o igual que b?

Operadores Lógicos.

Operador	Descripción	Ejemplo
and	¿se cumple a y b?	<pre>>>> True and False False</pre>
or	¿se cumple a o b?	<pre>>>> True or False True</pre>
not	No al valor	<pre>>>> not True False</pre>

Operadores de Identidad

Operador	Evalúa
<i>is</i>	<i>a is b</i> Equivale a $id(a) == id(b)$
<i>is not</i>	<i>a is not b</i> Equivale a $id(a) != id(b)$

Operadores de Pertenencia

- **In**: evalúan si un objeto se encuentra dentro de otro
"ni" in "Daniel"
True
- **not in** evalúan si un objeto **no** se encuentra dentro de otro.
"ni" not in "Daniel"
False

Operadores para objetos de tipo str

Operador	Descripción
+	Concatenación
*	Repetición

Operadores de bits

Operador	Descripción
&	AND
	OR
^	XOR
<<	Mover x bits a la izquierda
>>	Mover x bits a la derecha

Operadores de bits

Operador	Descripción
&	AND
	OR
^	XOR
<<	Mover x bits a la izquierda
>>	Mover x bits a la derecha

Identificadores

- "Un identificador es un nombre definido por el usuario para representar una variable, una función, una clase, un módulo o cualquier otro objeto". 5 reglas para nombrar identificadores en Python:
 - 1) Puede ser una combinación de letras minúsculas (a-z) / mayúsculas (A- Z), dígitos (0-9) o un guión bajo (_).
 - 2) No puede empezar con un dígito.
 - 3) No podemos usar símbolos especiales (! @ # \$ % .)
 - 4) No se puede usar las palabras reservadas.
 - 5) Puede tener cualquier largo.
- Nota: Python diferencia entre mayúsculas y minúsculas.

Identificadores

- Palabras reservadas:

```
>>> import keyword
```

```
>>> keyword.kwlist
```

and	def	False	import	not	True
as	del	finally	in	or	try
assert	elif	for	is	pass	while
break	else	from	lambda	print	with
class	except	global	None	raise	yield
continue	exec	if	nonlocal	return	

Manejo de variables

- Asignación simple (=)

```
>>> a=5
```

- Asignación multiple:

```
>>> edad,nombre = 21, 'Daniel'
```

- Mismo valor para múltiples variables:

- ```
>>> age=fav=7
```

- Intercambiando variables

```
>>> a,b='red','blue'
```

```
>>> a,b=b,a
```

- Borrando variables

```
>>> del a
```



# Conversión de tipos

- Para convertir entre tipos de datos se puede utilizar:
  - `bool(x)` Convierte x en un booleano.
  - `int (x)` Convierte x en un entero
  - `float (x)` Convierte x en un número real
  - `str (x)` Convierte x a una cadena.
  - `set(x)` Convierte x en un conjunto
  - `list(x)` Convierte x en una lista
  - `tuple(x)` Convierte x en una tupla
- Nota: hay conversiones que no se pueden realizar y retorna un error.

# Variables Locales y Globales

- Variables Locales:
  - Una variable creada dentro de una función pertenece al ámbito local de esa función y solo se puede usar dentro de esa función.
- Variables Globales:
  - Una variable creada en el cuerpo principal del código es una variable global y pertenece al ámbito global.
  - Se puede usar la palabra clave "global" cuando desee tratar una variable como global en un ámbito local.

# Variables Locales y Globales

- #con errores

```
a=0
```

```
def func():
```

```
 print(a)
```

```
 a=1
```

```
 print(a)
```

```
func()
```

- #ámbitos

```
def red():
```

```
 a=1
```

```
 def blue():
```

```
 b=2
```

```
 print(a)
```

```
 print(b)
```

```
 blue()
```

```
 print(a)
```

```
red()
```

- #global

```
a=1
```

```
def counter():
```

```
 global a
```

```
 a=2
```

```
 print(a)
```

```
counter()
```

```
a
```

- #nonlocal

```
def red():
```

```
 a=1
```

```
 def blue():
```

```
 nonlocal a
```

```
 a=2
```

```
 b=2
```

```
 print(a)
```

```
 print(b)
```

```
 blue()
```

```
 print(a)
```

```
red()
```

# Formateadores de strings

- Para imprimir variables junto con una cadena se puede usar comas o usar formateadores de cadenas.
- **Comas:**  

```
nombre="Carlos"
edad=45

print(nombre, "tiene", edad, "años de edad")
```
- **f-strings (nuevo formado, recomendado):**
  - ```
print(f"{nombre} tiene {edad} años de edad")
```
 - La letra "f" precede a la cadena, y las variables se mencionan entre llaves en sus lugares.

Formateadores de cadenas

- **Método format():**

- `print("{0} tiene {1} años de edad".format(nombre,edad))`
- `print("{a} tiene {b} años de edad".format(a=nombre,b=edad))`
- Sucede a la cadena y las variables van como argumentos separados por comas. Se usa llaves para colocar las variables. Se puede hacer referencia a la posición o al nombre si se definen.

Formateadores de cadenas

- **Operador %:**

- `print("%s tiene %s años de edad"%(nombre,edad))`
- Se colocan donde van las variables en una cadena. %s es para la cadena. Lo que sigue a la cadena es el operador % y las variables en una tupla.
- Otras opciones incluyen:
 - %d - para enteros
 - %f - para números de coma flotante

Entrada estándar

- La función `input()` permite obtener texto escrito por teclado. El programa se detiene esperando que se escriba algo y se pulse la tecla Intro.
 - `print("¿Cómo se llama? ", end="")`
`nombre = input()`
`print(f"Me alegro de conocerle, {nombre}")`
 - `nombre = input("¿Cómo se llama? ")`
`print(f"Me alegro de conocerle, {nombre}")`
- De forma predeterminada, la función `input()` convierte la entrada en una cadena, aunque escribamos un número. Si intentamos hacer operaciones, se producirá un error.

Entrada estándar

- Conversión de tipos

- `cantidad = input("Dígame una cantidad en euros:")`
`print(f"{cantidad} euros son {round(cantidad * 1.15, 2)} dólares")`
- `cantidad = int(input("Dígame una cantidad en euros:"))`
`print(f"{cantidad} euros son {round(cantidad * 1.15, 2)} dólares")`

Manejo de archivos

- Abrir / Crear un archivo:
 - Lo primero es abrir (o crear) un archivo.
 - `open(nombre, modo)`
 - Nombre: es el nombre del archivo.
 - Modos:
 - `r` para lectura (por defecto)
 - `w` para escritura
 - `r+` o `w+` lectura escritura
 - `a` para agregar al final
 - `T` modo texto (por defecto)
 - `b` modo binario.

Manejo de archivos

- Crear el archivo:

Crear una archivo en modo write

```
f = open("hola.txt", "w")
```

Escribir un texto al archivo

```
f.write("Hola Mundo!")
```

Cerrar el arhivo

```
f.close()
```

Manejo de archivos

- Leer un archivo:

Abrir el archivo en modo 'read'

```
f = open("hola.txt", "r")
```

Colocar el contenido del archivo en una variable

```
f_contenido= f.read()
```

Cerrar el archivo

```
f.close()
```

Imprimir el contenido del archivo

```
print(f_contenido)
```

Manejo de archivos

- Agregar a un archivo:

```
f = open("hola.txt", "a")
```

```
f_contenido = "\r\n" + "Agregando una linea!"
```

```
f.write(f_contenido)
```

```
f.close()
```

Manejo de archivos

- Leer una línea **del** archivo

```
f = open("hola.txt", "r")  
linea1 = f.readline()  
linea2 = f.readline()  
print("Línea 1:", linea1)  
print("Línea 2:", linea2)  
f.close()
```

Manejo de archivos

- Leer un archivo línea a línea:

```
f = open("hola.txt", "r")
```

```
for linea in f:
```

```
    print("Linea:", linea, end="")
```

```
f.close()
```

```
#Using 'list()'
```

```
f = open("hola.txt", "r")
```

```
f_content = list(f)
```

```
print(f_content)
```

```
f.close()
```

Manejo de archivos

- Sentencia **with**:

```
with open("hola.txt", "w") as f:  
    # Escribir un texto al archivo  
    f.write("Hola Mundo!")
```

- Nota: con **with** el close se hace automáticamente

Manejo de archivos

Method	Description
<u>close()</u>	Closes the file
<u>detach()</u>	Returns the separated raw stream from the buffer
<u>fileno()</u>	Returns a number that represents the stream, from the operating system's perspective
<u>flush()</u>	Flushes the internal buffer
<u>isatty()</u>	Returns whether the file stream is interactive or not
<u>read()</u>	Returns the file content
<u>readable()</u>	Returns whether the file stream can be read or not
<u>readline()</u>	Returns one line from the file
<u>readlines()</u>	Returns a list of lines from the file
<u>seek()</u>	Change the file position
<u>seekable()</u>	Returns whether the file allows us to change the file position
<u>tell()</u>	Returns the current file position
<u>truncate()</u>	Resizes the file to a specified size
<u>writable()</u>	Returns whether the file can be written to or not
<u>write()</u>	Writes the specified string to the file
<u>writelines()</u>	Writes a list of strings to the file

Programación Orientada a Objeto (POO)

- Los estilos de programación se les llama paradigmas:
 - Secuencial o lineal:
 - Las instrucciones van de arriba hacia abajo, no tenemos que abstraer cosas complejas, simplemente damos ordenes una tras otra.
 - Para aplicaciones sencillas suele ser muy directo y mantenible.
 - Estructurado:
 - Surge con la idea de mejorar la claridad, calidad y tiempo de desarrollo de una programación secuencial. Se basa en en subrutinas y estructuras básicas:
 - bloques de código
 - sentencias condicionales
 - y bucles.

Programación Orientada a Objeto (POO)

- Lineal y Estructurado:
 - Se centran en parte algorítmica y la lógica de programación más que en la representación de los datos y la descripción de la lógica del negocio. Los datos están separados y sin relación con los procedimientos o funciones.
 - Aunque es útil para muchos problemas para el desarrollo de grandes aplicaciones con una lógica de negocio extensa y dinámica es complejo de mantener y evolucionar.

Programación Orientada a Objeto (POO)

- POO:
 - Es la evolución natural de la programación estructurada. Se basa en una representación más cercana a como expresaríamos las cosas en la vida real.
 - Se basa en dividir el programa en pequeñas unidades lógicas de código. A estas **pequeñas unidades lógicas de código se les llama objetos**.
 - Los objetos son estructuras independientes que **combinan datos** o estados **(variables) con** acciones asociadas o **comportamiento (métodos)**.
 - La lógica de negocio se modela e implementa mediante una serie de objetos que **interactúan entre si**.
 - Este enfoque aumenta la capacidad para administrar la complejidad del software, lo cual resulta especialmente importante cuando se desarrollan y mantienen aplicaciones y estructuras de datos de gran tamaño.

Programación Orientada a Objeto (POO)

- Características principales:
 - **Abstracción:** Los objetos permiten modelar las características esenciales y el comportamiento de entidades reales sin revelar "cómo" se implementan.
 - **Encapsulamiento:** Consiste en agrupar en una clase las características y el comportamiento de un mismo nivel a abstracción.
 - **Ocultamiento:** Es la capacidad de ocultar los detalles internos de una clase y exponer sólo los detalles que sean necesarios para el resto del sistema.
 - **Polimorfismo:** Se refiere a la propiedad de invocar acciones sintácticamente iguales a objetos de tipos distintos.

Programación Orientada a Objeto (POO)

- Características principales:

- **Herencia:**

- Las clases no se encuentran aisladas, sino que se relacionan entre sí, formando una jerarquía de clasificación. Los objetos heredan las propiedades y el comportamiento de todas las clases a las que pertenecen.

- **Recolección de basura:**

- Es la técnica por la cual el entorno de objetos se encarga de destruir automáticamente los objetos que hayan quedado sin ninguna referencia a ellos.
 - Esto significa que el programador no debe preocuparse por la asignación o liberación de memoria, ya que el entorno asignará memoria al crear un nuevo objeto y la liberará la memoria cuando nadie esté usando el objeto.

POO en Python

- Características principales:
 - **Abstracción:** soportado
 - **Encapsulamiento:** soportado
 - **Ocultamiento:** brinda esta característica por convención.
 - **Polimorfismo:** natural por su enfoque dinámico (enlazado tardío).
 - **Herencia:** soporta herencia simple y múltiple
 - **Recolección de basura:** soportado

POO: Clases

- Una clase es un tipo de dato definido por el usuario. La clase es como el plano (la definición) de los objetos. Sintaxis:

- **# Clase base:**

```
class <NombreDeLaClase>:
```

```
    """documentación"""
```

```
    <Sentencias de la clase>
```

- **# Herencia (simple y multiple)**

```
class <NombreDeLaClase>(<ClaseBase1, ..., ClaseBaseN>):
```

```
    """documentación"""
```

```
    <Sentencias de la clase>
```

POO: Clases

- Crear una clase:

```
class MiClase:  
    x = 5
```

- Crear un objeto:

```
o = MiClase()  
print(o.x)
```

- Borrar un objeto:
del o

- Todo es un objeto:

```
[1]: a=5  
    type(a)
```

```
[1]: int
```

```
[4]: a=int(5)  
    type(a)
```

```
[4]: int
```


POO: Clases: `__init__()`

- Todas las clases tienen una función llamada `__init__()`, que siempre se ejecuta cuando se instancia la clase.
- Se le llama “constructor” y se usa para asignar valores a los atributos del objeto u otras operaciones que sean necesarias cuando se crea el objeto.
- Nota: El parámetro `self` es una referencia a la instancia actual de la clase y se usa para acceder a las variables que pertenecen a la clase.

```
class Persona:
```

```
    def __init__(self, nombre, edad):  
        self.nombre = nombre  
        self.edad = edad
```

```
p1 = Persona("Carlos", 36)
```

```
print(p1.nombre)
```

```
print(p1.edad)
```

POO: Clases: Métodos

- Los objetos pueden contener métodos (son funciones que pertenecen al objeto).
- Todos los métodos de clase deben tener el argumento adicional `self` como primer argumento.

```
class Persona:  
    def __init__(self, nombre, edad):  
        self.nombre = nombre  
        self.edad = edad  
  
    def mimetodo(self):  
        print("Hola, mi nombre es " + self.nombre)
```

```
p1 = Persona("Carlos", 36)  
p1.mimetodo()
```

POO: Clases: Variables

- Las variables del objeto se pueden modificar o borrar de la siguiente manera:

```
class Persona:
    def __init__(self, edad):
        self.edad = edad

p1 = Persona(36)
# Modificar
p1.edad = 40
# Borrar
del p1.edad
```

POO: Clases: Variables

- Dentro de una clase hay dos tipos principales de variables:
 - Variables de la clase:
 - Tienen el mismo valor en todas las instancias de la clase (es decir, variables estáticas)
 - Se accede a la variable de la clase usando:
<NombreDeLaClase>.<NombreDeLaVariable>
 - Variables de la instancia:
 - Suelen tener valores diferentes en todas las instancias de la clase.
 - Se accede a la variable de la instancia usando
<NombreDeLaInstancia>.<NombreDeLaVariable>

POO: Clases: Variables

```
class Auto:
```

```
    # Variable de la clase
```

```
    ruedas = 4
```

```
    def __init__(self, marca):
```

```
        # Variable de la instancia
```

```
        self.marca = marca
```

```
auto1= Auto("Renault")
```

```
auto2= Auto("Suzuki")
```

```
print(f"Marca: {auto1.marca}, Ruedas: {Auto.ruedas}")
```

```
print(f"Marca: {auto2.marca}, Ruedas: {Auto.ruedas}")
```

POO: Clases: Variables

- Cuando se crean variables de instancia realmente se generan dos, una de la clase y otra de instancia con el mismo nombre. Y cual se está usando depende de como se acceda.

```
class Auto:
    ruedas = 4
    def __init__(self, marca):
        self.marca = marca
```

```
instancia= Auto("Renault")
instancia.ruedas = 5
print(instancia.ruedas)
print(Auto.ruedas)
```

POO: Encapsulamiento

- Nada en Python es privado. Todos los atributos de Python (variables y métodos) son públicos.
- Por convención se usa un solo guión bajo antes del nombre para señalar que es un atributo interno y no debería usarse externamente.

```
class Auto:
```

```
    def __init__(self, marca):
```

```
        self._marca = marca
```

```
auto = Auto("Renault")
```

```
auto._marca
```

POO: Encapsulamiento

- Un doble guión bajo__al principio de una variable lo “hace privado”. Da una fuerte sugerencia de no tocarlo desde fuera de la clase. Cualquier intento de acceder dará como resultado un AttributeError ya que python cambia el nombre al siguiente formato:

<_NombreDeLaClase>__<NombreDeLaVariable>.

```
class Auto:
    def __init__(self, marca):
        self.__marca = marca
auto = Auto("Renault")
auto._Auto__marca
```


POO: Herencia

- La herencia nos permite definir una clase que hereda todos los métodos y propiedades de otra clase.
 - La clase padre es la clase de la que se hereda, también llamada clase base.
 - La clase hija es la clase que hereda de otra clase, también llamada clase derivada

```
class Persona:
    def __init__(self, nombre):
        self.nombre = nombre
    def printNombre(self):
        print(self.nombre)
class Estudiante(Persona):
    pass
x = Estudiante("Alejandro")
x.printNombre()
```

POO: Herencia

```
class Persona:
```

```
    def __init__(self, nombre):
```

```
        self.nombre = nombre
```

```
class Estudiante(Persona):
```

```
    def __init__(self, nombre, fecha): # Constructor propio
```

```
        super().__init__(nombre) # Para mantener la herencia
```

```
        self.fechagraduacion = fecha
```

```
    def welcome(self):
```

```
        print(f"Bienvenido {self.nombre} ({self.fechagraduacion})")
```

```
x = Estudiante("Alejandro", "2019")
```

```
x.welcome()
```

POO: Herencia Múltiple

```
class A:  
    def a(self):  
        print('b')
```

```
class B:  
    def b(self):  
        print('c')
```

```
class C(A, B):  
    def c(self):  
        print('d')
```

```
c = C()
```

```
c.a()
```

```
c.b()
```

```
c.c()
```

Manejo de Excepciones

- Hay dos tipos de errores durante la ejecución de un programa:
 - Errores de sintaxis:
 - Los errores de sintaxis ocurren cuando se escribe el código incorrectamente.
 - En tales casos, la línea errónea es repetida por el analizador con una flecha apuntando a la primera ubicación en donde el error fue detectado.
 - Excepciones:
 - Estos ocurren durante la ejecución de un programa cuando algo inesperado sucede. Por ejemplo, división por cero.
 - Cuando no estás manejando excepciones apropiadamente, el programa se cerrará de manera abrupta ya que no sabe que hacer en tales casos.

Manejo de Excepciones

try:

<sentencias>

except excepción1 [as variable1]:

<sentencias>

...

except excepciónN [as variableN]:

<sentencias>

except (exA, exB, ...) [as variable]:

<sentencias>

except:

<sentencias>

else:

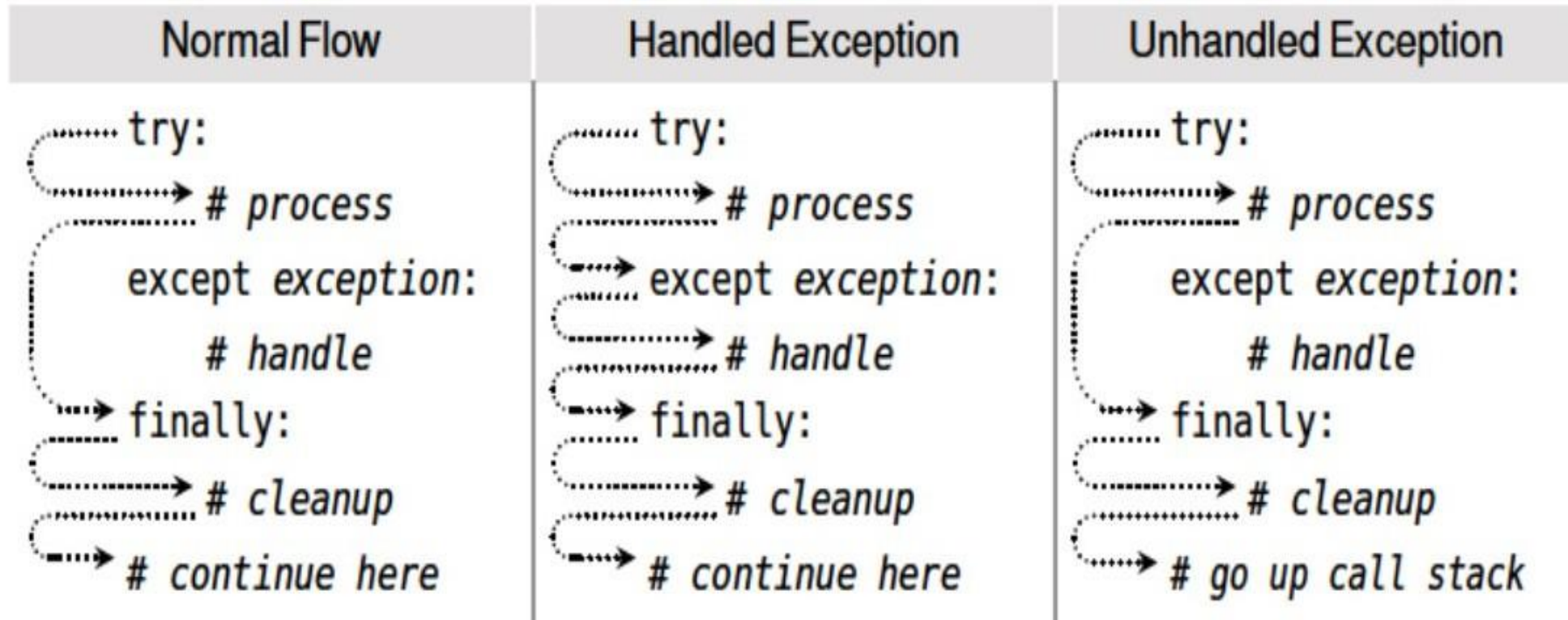
<sentencias>

finally:

<sentencias>

- **Try:** permite controlar las excepciones dentro de un bloque de código.
- **Except:** permite ejecutar código si ocurrió alguna excepción.
- **Else:** permite ejecutar código si no ocurrieron excepciones.
- **Finally:** permite ejecutar código independientemente de si ocurrieron o no excepciones.

Manejo de Excepciones



Manejo de Excepciones

```
while True:
```

```
    try:
```

```
        x = int(input("Ingrese un número: "))
```

```
    except ValueError:
```

```
        print("El valor ingresad no es un entero.")
```

```
    else:
```

```
        print("Calculando 50 /", x, "Resultado:", 50/x)
```

```
    finally:
```

```
        print("Ya hice todo lo necesario.")
```

Manejo de Excepciones

- En ciertas ocasiones es deseable generar una excepción:
 - Si estamos dentro de un bloque except, se puede lanzar una excepción sin especificar que excepción.
raise
 - Especificando un excepción (estándar o personalizada) y un argumento:
raise <TipoDeError>(<mensaje>)

```
x = -1
```

```
if x < 0:
```

```
    raise Exception("x tiene un valor negativo")
```


Manejo de Excepciones

- Las excepciones personalizadas son clases que heredan de otra excepción:

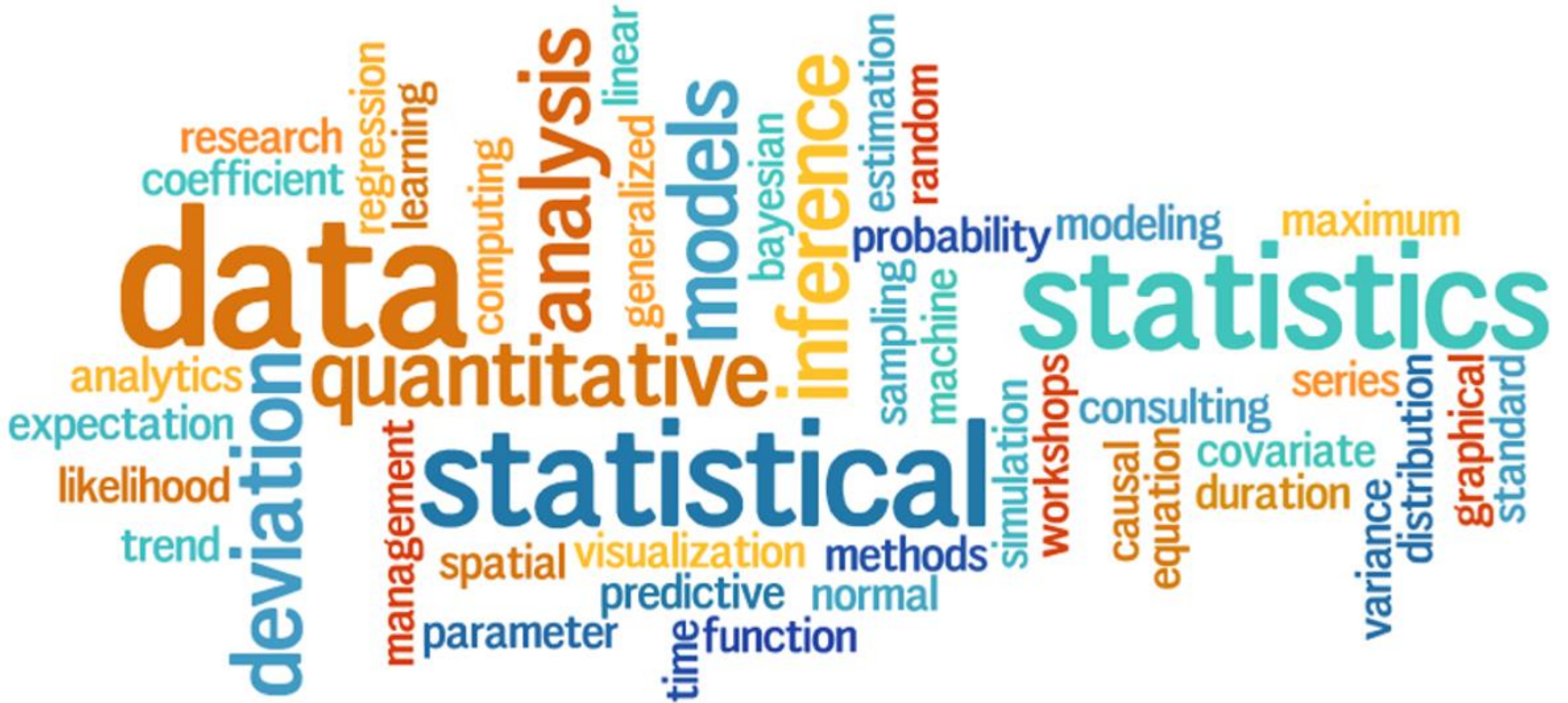
```
class exceptionName(baseException):  
    <sentencias>
```

- Ejemplo:

```
class MiException(Exception):  
    def __init__(self, mensaje):  
        super().__init__(mensaje)
```

```
raise MiException("Usando una excepción personalizada")
```

Fundamentos y conceptos de Probabilidad y Estadística para la ciencia de datos.



¿Qué es la Estadística?

- La estadística suele ser definida como la ciencia de aprender de los datos o como la ciencia de obtener conclusiones en la presencia de incertidumbre. Se relaciona principalmente con la recolección, análisis e interpretación de datos, así como también con la efectiva comunicación y presentación de los resultados basados en esos datos.
- La estadística puede ser muy importante para una efectiva toma de decisiones. Existe una gran cantidad de valiosa información escondida entre los datos, pero esta información no suele ser fácilmente accesible, la estadística nos brinda los principios fundamentales que nos permiten extraer y entender esa información; también nos proporciona las herramientas necesarias para verificar la calidad de nuestros datos y nuestra información.

Tipos de Estadísticas

- *Estadística Descriptiva*: La cual se dedica a recolectar, ordenar, analizar y representar a un conjunto de datos, con el fin de describir apropiadamente las características de este. Calcula los parámetros estadísticos que describen el conjunto estudiado. Algunas de las herramientas que utiliza son gráficos, medidas de frecuencias, medidas de centralización, medidas de posición, medidas de dispersión, entre otras.
- *Estadística Inferencial*: La cual estudia cómo sacar conclusiones generales para toda la población a partir del estudio de una muestra, y el grado de fiabilidad o significación de los resultados obtenidos. Sus principales herramientas son el muestreo, la estimación de parámetros y el contraste de hipótesis.

¿Qué es la Probabilidad?

- La [*probabilidad*](#) mide la mayor o menor posibilidad de que se dé un determinado resultado (suceso o evento) cuando se realiza un experimento aleatorio.
- Para calcular la probabilidad de un evento se toma en cuenta todos los casos posibles de ocurrencia del mismo; es decir, de cuántas formas puede ocurrir determinada situación.
- Los casos favorables de ocurrencia de un evento serán los que cumplan con la condición que estamos buscando. La [*probabilidad*](#) toma valores entre 0 y 1 (o expresados en tanto por ciento, entre 0% y 100%)



¿Qué es la Probabilidad?

- La probabilidad es a la vez el inverso y complemento para la estadística. Dónde la estadística nos ayuda a ir desde los datos observados hasta hacer generalizaciones sobre como funcionan las cosas; la probabilidad funciona en la dirección inversa: si asumimos que sabemos como las cosas funcionan, entonces podemos averiguar la clase de datos que vamos a ver y cuan probable es que los veamos.
 - La probabilidad también funciona como complemento de la estadística cuando nos proporciona una sólida base para la estadística inferencial. Cuando hay incertidumbre, no sabemos que puede pasar y hay alguna posibilidad de errores, utilizando probabilidades podemos aprender formas de controlar la tasa de errores para reducirlos.
-



Actividades básicas del análisis estadístico

Un análisis estadístico suele contener 5 actividades básicas:

1. **Diseño del análisis:** Esta actividad involucra el planeamiento de los detalles para obtener los datos que necesitamos y la generación de la hipótesis a ser evaluada.
 2. **Exploración de datos:** En esta actividad nos dedicamos a jugar con nuestros datos, los describimos, los resumimos, realizamos gráficos para mirarlos desde distintos ángulos. Esta exploración nos ayuda a asegurarnos que los datos que obtuvimos son completos y que la etapa de diseño fue correcta.
-



Actividades básicas del análisis estadístico

3. **Armado del modelo:** En esta actividad intentamos armar un modelo que explique el comportamiento de nuestros datos y pueda llegar a hacer predicciones sobre los mismos. La idea es que el modelo pueda describir las propiedades fundamentales de nuestros datos.
 4. **Realizar estimaciones:** Aquí vamos a intentar realizar estimaciones basadas en el modelo que armamos anteriormente. También vamos a intentar estimar el tamaño del error que nuestro modelo puede tener en sus predicciones.
 5. **Contraste de la hipótesis:** Esta actividad es la que va a producir la decisión final sobre si las predicciones del modelo son correctas y ayudarnos a concluir si los datos que poseemos confirman o rechazan la hipótesis que generamos en la actividad 1.
-

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

Conceptos básicos de la estadística descriptiva

La **media aritmética**, también conocida como promedio, es una medida estadística que se utiliza para representar el valor típico o central de un conjunto de datos. Se calcula sumando todos los valores en el conjunto y dividiendo esa suma por la cantidad de valores en el conjunto. La fórmula para calcular la media aritmética es la siguiente:

$$\text{Media} = (\text{Suma de todos los valores}) / (\text{Cantidad de valores})$$

Para ilustrar esto gráficamente, considere un conjunto de datos, por ejemplo, las edades de un grupo de personas. Si tenemos las siguientes edades: 25, 30, 35, 40 y 45, podemos calcular la media aritmética de la siguiente manera:

$$\text{Media} = (25 + 30 + 35 + 40 + 45) / 5 = 175 / 5 = 35$$

Entonces, la media aritmética de estas edades es 35 años.

$$D_i = |x_i - \mu|$$

Conceptos básicos de la estadística descriptiva

La **Desviación respecto a la media** se utiliza para cuantificar cuánto se aleja cada valor individual de un conjunto de datos de la media aritmética de ese conjunto. Se calcula siguiendo los siguientes tres pasos:

1. Resta el valor individual de la media aritmética del conjunto de datos.
2. Toma el valor absoluto del resultado obtenido en el paso 1 (esto se hace para que todas las desviaciones sean positivas).
3. Repite estos pasos para cada valor en el conjunto de datos.

Supongamos un conjunto de datos que representa las edades de un grupo de personas y la media de esas edades es de 40 años. Si una persona tiene 45 años, la desviación respecto a la media para esa persona sería:

$$D = |45 - 40| = 5 \text{ años}$$

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

La **varianza** se utiliza para cuantificar la dispersión o la variabilidad de un conjunto de datos. Se calcula siguiendo los siguientes tres pasos:

1. Resta la media aritmética de cada valor individual en el conjunto de datos.
2. Eleva al cuadrado cada una de esas diferencias.
3. Suma todos los valores resultantes.
4. Divide la suma por la cantidad de valores en el conjunto de datos (N).

Representa cuán alejados están los valores individuales de la media aritmética de esos datos.

Conceptos básicos de la estadística descriptiva

Conceptos básicos de la estadística descriptiva

Calificaciones: 85, 92, 88, 78, 90

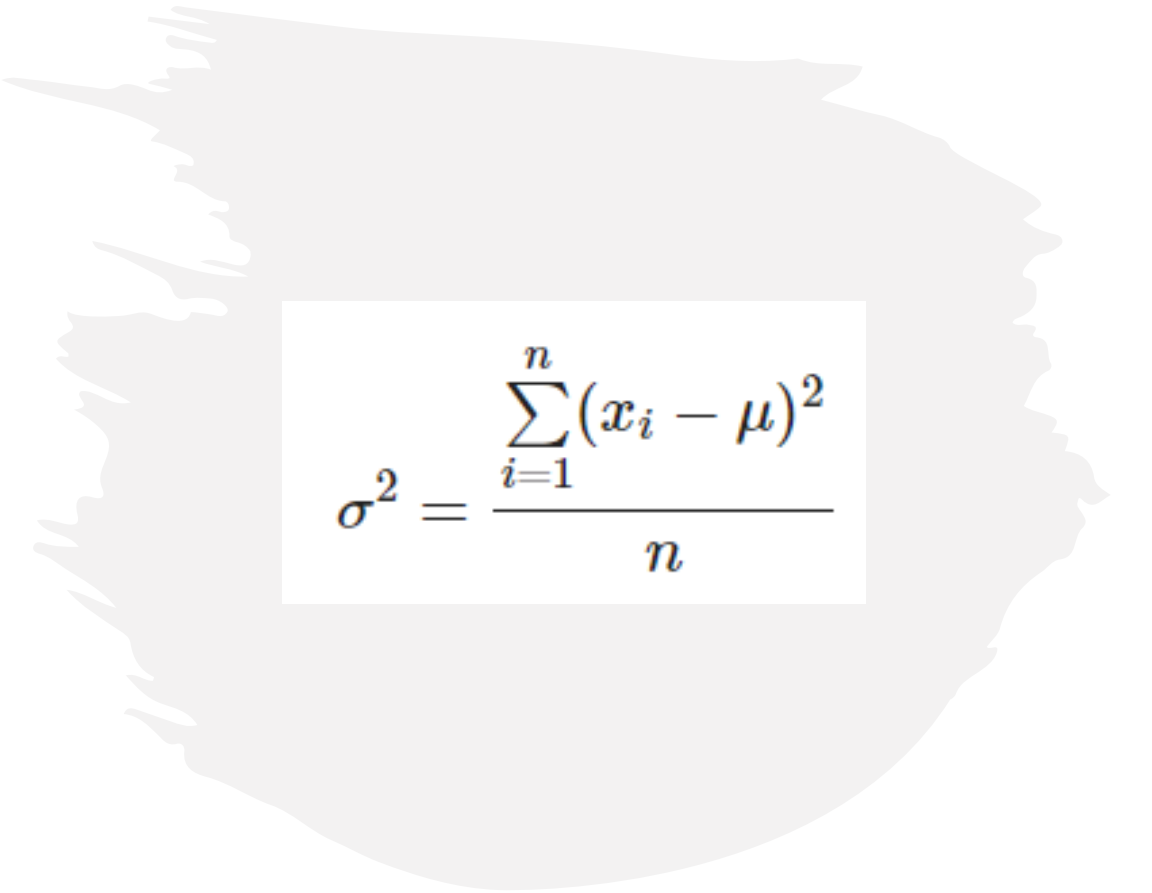
Paso 1: **Calcular la media**

Primero, calculamos la media aritmética de estos datos sumando todos los valores y dividiendo por la cantidad de valores:

$$\text{Media} = (85 + 92 + 88 + 78 + 90) / 5 = 443 / 5 = 88.6$$

La media es 88.6.

Paso 2: **Calcular la diferencia entre cada valor y la media**, elevar al cuadrado y sumar


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

Conceptos básicos de la estadística descriptiva

Ahora, restamos la media de cada valor, elevamos al cuadrado cada diferencia y sumamos estos valores al cuadrado. Luego, dividimos por la cantidad de valores (5) para obtener la varianza.

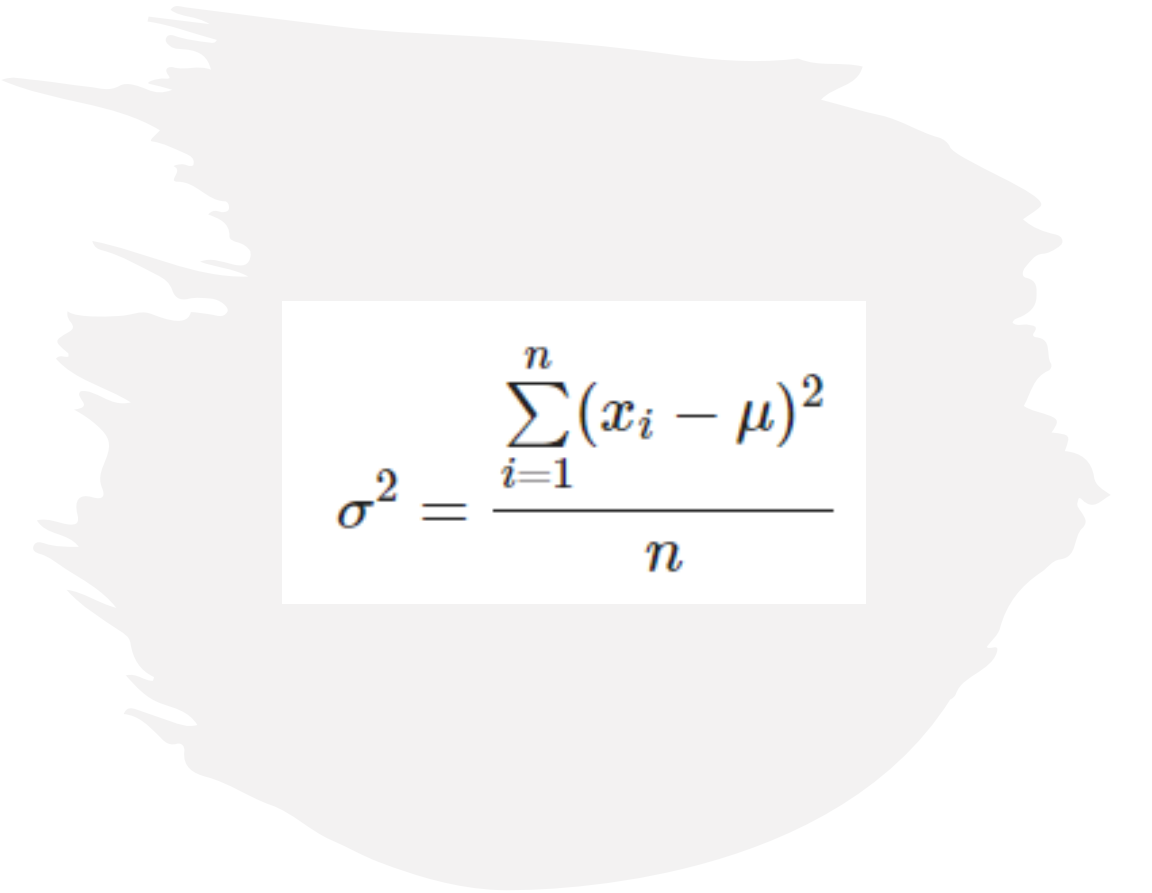
$$\text{Varianza} = [(85 - 88.6)^2 + (92 - 88.6)^2 + (88 - 88.6)^2 + (78 - 88.6)^2 + (90 - 88.6)^2] / 5$$

$$\text{Varianza} = [(10.36 + 10.36 + 0.36 + 92.16 + 1.96)] / 5$$

$$\text{Varianza} = (115.2) / 5$$

$$\text{Varianza} = 23.04$$

La varianza de las calificaciones es igual a 23.04.


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

Conceptos básicos de la estadística descriptiva

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

La **desviación típica**, también conocida como **desviación estándar**, cuantifica la dispersión o la variabilidad de un conjunto de datos.

- Es una medida de cuánto los valores individuales en un conjunto de datos tienden a alejarse de la media aritmética de esos datos.

En el ejemplo anterior, la Desviación estándar = $\sqrt{\text{Varianza}} = \sqrt{23.04} \approx 4.8$

La desviación estándar es aproximadamente 4.8.

Entonces, la **varianza de las calificaciones es 23.04** y la **desviación estándar es aproximadamente 4.8**. Estas medidas te dan una idea de cuánto varían las calificaciones con respecto a la media de 88.6

$$M = L_i + \left(\frac{D_1}{D_1 + D_2} \right) A_i$$

Conceptos básicos de la estadística descriptiva

La **moda** en estadística descriptiva se refiere al valor o valores que aparecen con mayor frecuencia en un conjunto de datos.

Por ejemplo, si tenemos las edades: 25, 30, 25, 35, 40, 30, 25, 28

En este caso, el valor "25" se repite con mayor frecuencia (tres veces), por lo que la moda de este conjunto de datos es "25".

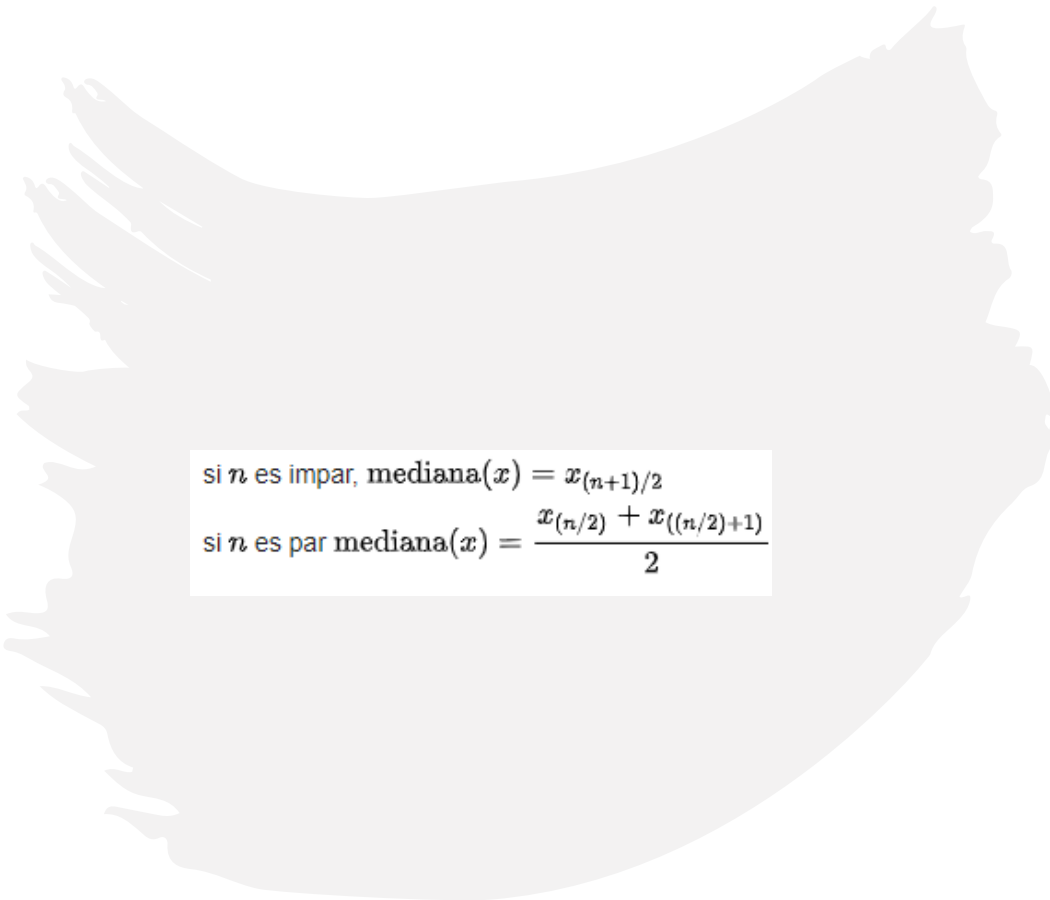
Puedes tener conjuntos de datos más complejos en los que haya múltiples modas o incluso casos en los que no haya una moda clara si todos los valores son diferentes.

La moda es especialmente útil cuando se trabaja con datos categóricos o discretos, como categorías de productos, nombres, colores, etc.

Conceptos básicos de la estadística descriptiva

La **mediana** representa el valor que ocupa la posición central en un conjunto de datos cuando estos se ordenan en orden ascendente o descendente. Para calcular la mediana, sigues estos pasos:

1. Ordena los datos en orden ascendente o descendente.
2. Si la cantidad de datos es impar, la mediana es el valor que se encuentra exactamente en el centro de la lista ordenada.
3. Si la cantidad de datos es par, la mediana es el promedio de los dos valores centrales.


$$\begin{aligned} \text{si } n \text{ es impar, } \text{mediana}(x) &= x_{(n+1)/2} \\ \text{si } n \text{ es par } \text{mediana}(x) &= \frac{x_{(n/2)} + x_{((n/2)+1)}}{2} \end{aligned}$$

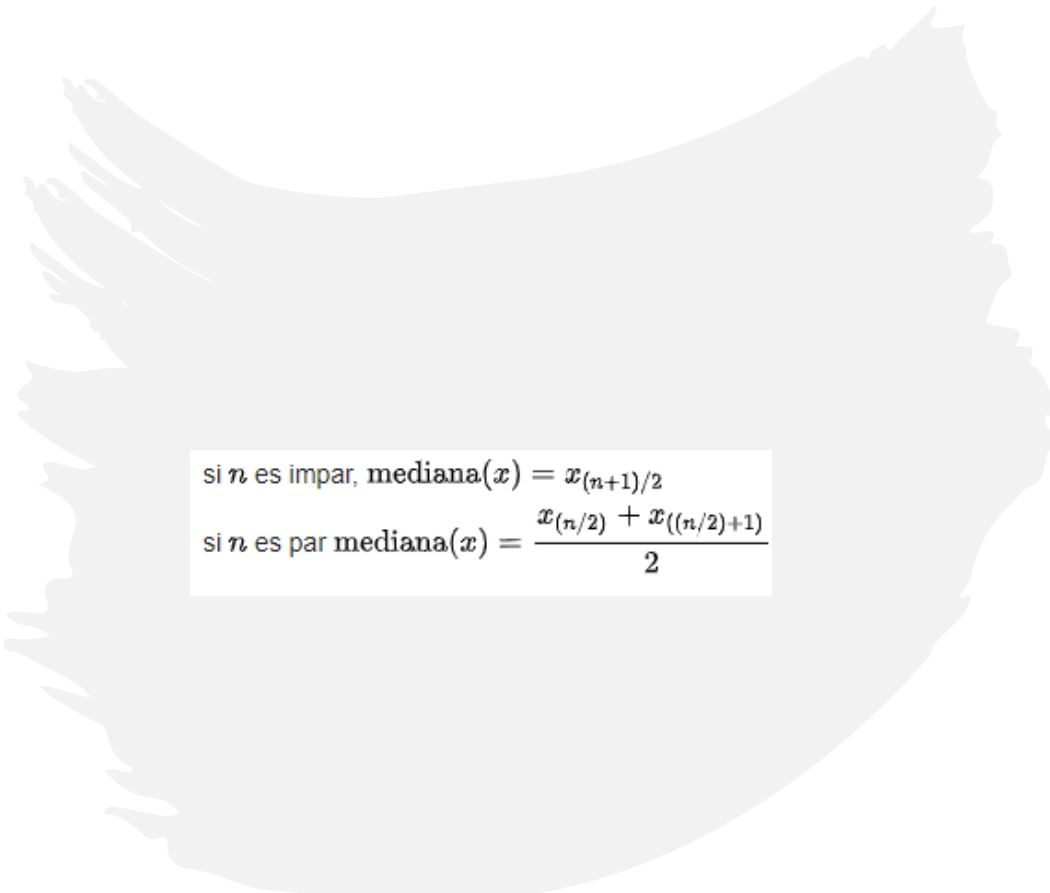
Conceptos básicos de la estadística descriptiva

Ejemplo mediana con datos impares:

Supongamos que tenemos el siguiente conjunto de datos que representan las edades de un grupo de personas: 20, 25, 30, 35, 40. La cantidad de datos es impar (en este caso, cinco), por lo **que la mediana es el valor del medio cuando se ordenan: 30.**

Ejemplo mediana con datos pares:

Si tenemos un conjunto de datos con una cantidad par de valores, como: 10, 15, 20, 25, **la mediana es el promedio de los dos valores centrales**, que en este caso son 15 y 20. La mediana sería $(15 + 20) / 2 = 17.5$.


$$\begin{aligned} \text{si } n \text{ es impar, } \text{mediana}(x) &= x_{(n+1)/2} \\ \text{si } n \text{ es par } \text{mediana}(x) &= \frac{x_{(n/2)} + x_{((n/2)+1)}}{2} \end{aligned}$$

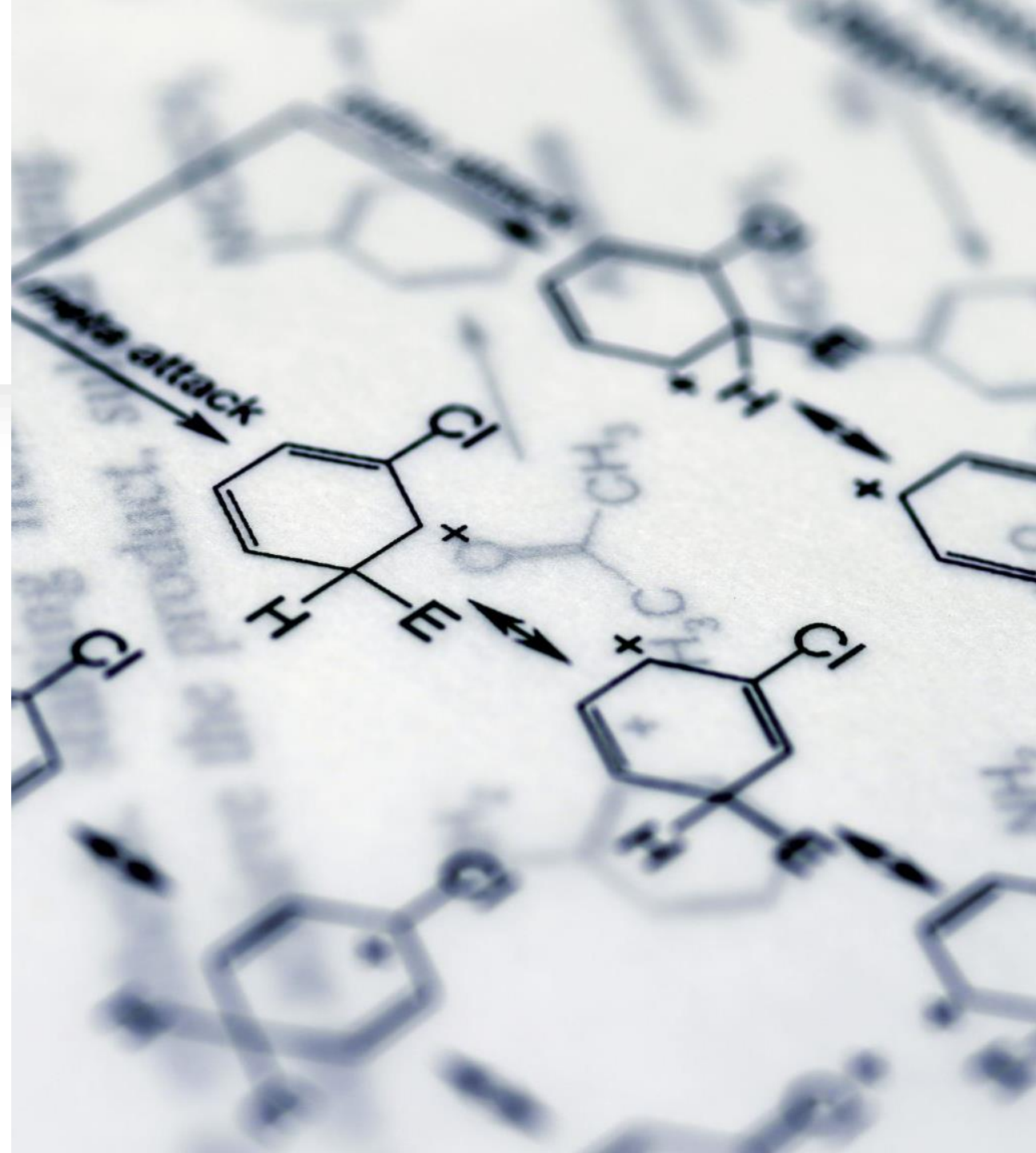
Conceptos básicos de la estadística descriptiva

La **correlación** trata de establecer la relación o dependencia que existe entre las dos variables que intervienen en una distribución bidimensional.

Es decir, determinar si los cambios en una de las variables influyen en los cambios de la otra.

En caso de que suceda, diremos que las variables están correlacionadas o que hay correlación entre ellas.

La correlación es positiva cuando los valores de las variables aumentan juntos; y es negativa cuando un valor de una variable se reduce cuando el valor de la otra variable aumenta.



$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad \text{if } \sigma_X \sigma_Y > 0.$$

Conceptos básicos de la estadística descriptiva

La **correlación** trata de establecer la relación o dependencia que existe entre las dos variables que intervienen en una distribución bidimensional.

Es decir, determinar si los cambios en una de las variables influyen en los cambios de la otra.

En caso de que suceda, diremos que las variables están correlacionadas o que hay correlación entre ellas.

La correlación es positiva cuando los valores de las variables aumentan juntos; y es negativa cuando un valor de una variable se reduce cuando el valor de la otra variable aumenta.

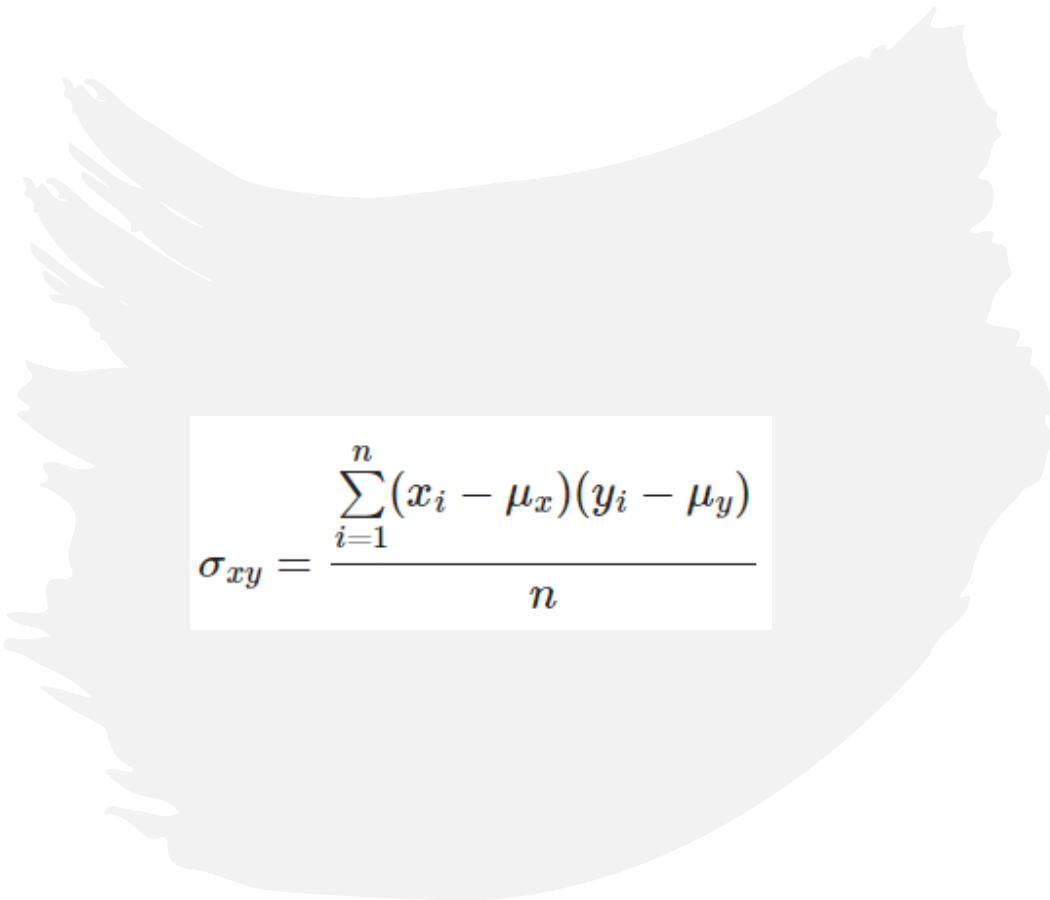
Conceptos básicos de la estadística descriptiva

La **covarianza** es el equivalente de la varianza aplicado a una variable bidimensional. Es la media aritmética de los productos de las desviaciones de cada una de las variables respecto a sus medias respectivas.

Si la covarianza es positiva, indica que cuando los valores de X aumentan, los valores de Y tienden a aumentar. Existe una relación positiva entre los conjuntos de datos.

Si la covarianza es negativa, indica que cuando los valores de X aumentan, los valores de Y tienden a disminuir. Existe una relación negativa entre los conjuntos de datos.

Si la covarianza es cercana a cero, indica que no hay una relación clara entre los conjuntos de datos, es decir, no están linealmente asociados.

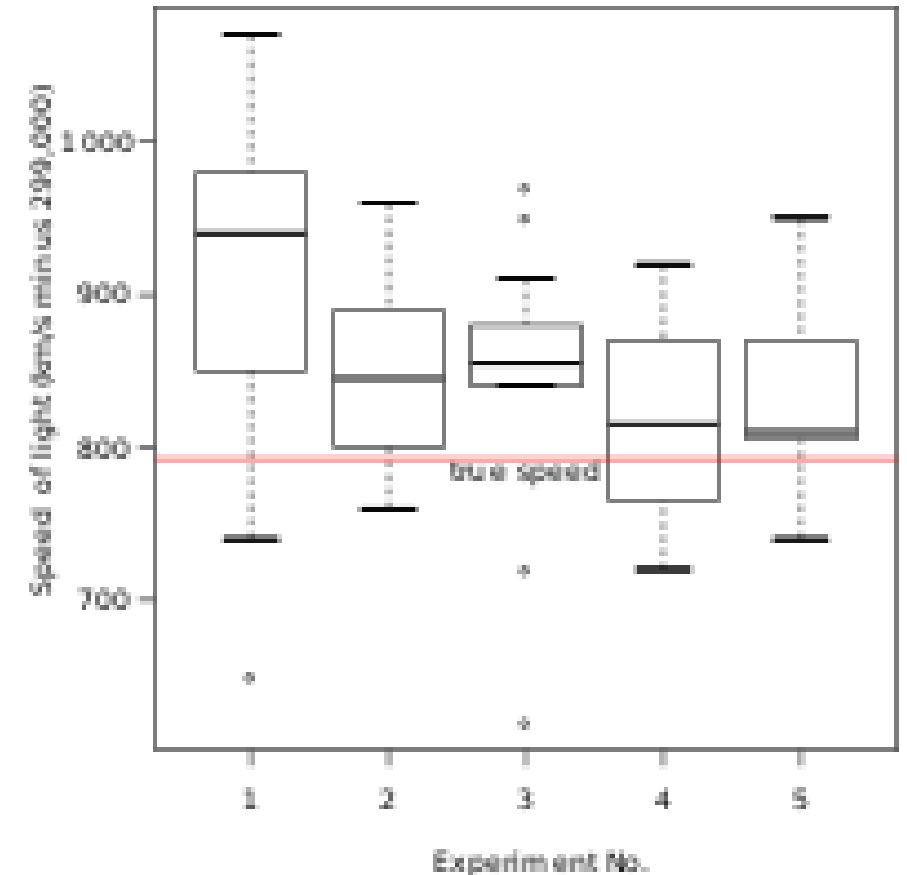

$$\sigma_{xy} = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{n}$$

Conceptos básicos de la estadística descriptiva

Un **valor atípico** es una observación que se aleja demasiado de la moda; está muy lejos de la tendencia principal del resto de los datos.

Pueden ser causados por errores en la recolección de datos o medidas inusuales.

Generalmente se recomienda eliminarlos del conjunto de datos. En un diagrama de caja se considera un valor atípico el que se encuentra 1,5 veces esa distancia de uno de esos cuartiles (atípico leve) o a 3 veces esa distancia (atípico extremo).



Conceptos básicos de la estadística descriptiva

Un **dato es categórico** cuando se puede clasificar en una categoría o grupo.

Los datos categóricos se diferencian de los **datos cuantitativos**, que se pueden medir en una escala numérica.

Los datos categóricos se pueden dividir en dos tipos principales:

1. **Datos nominales:** Los datos nominales son los datos categóricos más básicos. No hay orden lógico entre las categorías. Por ejemplo, el color de los ojos, el género o la afiliación política son datos nominales.
2. **Datos ordinales:** Los datos ordinales tienen un orden lógico, pero no es un orden preciso. Por ejemplo, la calificación de un estudiante en un examen, la clasificación de un equipo deportivo o el nivel de satisfacción de un cliente son datos ordinales.



Conceptos básicos de la estadística descriptiva

Algunos ejemplos de datos categóricos son:

- Género: Hombre, mujer
- Estado civil: Soltero, casado, divorciado, viudo
- Color: Rojo, verde, azul
- Tamaño: Pequeño, mediano, grande
- Puntuación: A, B, C, D, F
- Satisfacción: Muy satisfecho, satisfecho, insatisfecho, muy insatisfecho

Los datos categóricos se suelen representar con gráficos de barras o circulares.



Conceptos básicos de la estadística descriptiva

Aquí hay algunos ejemplos de cómo se pueden utilizar los datos categóricos en el análisis de datos:

- Un minorista puede utilizar datos categóricos sobre las compras de los clientes para determinar qué productos son más populares.
- Un político puede utilizar datos categóricos sobre los votantes para determinar qué temas son más importantes para ellos.
- Un investigador puede utilizar datos categóricos sobre los pacientes para determinar qué factores están asociados con una determinada enfermedad.