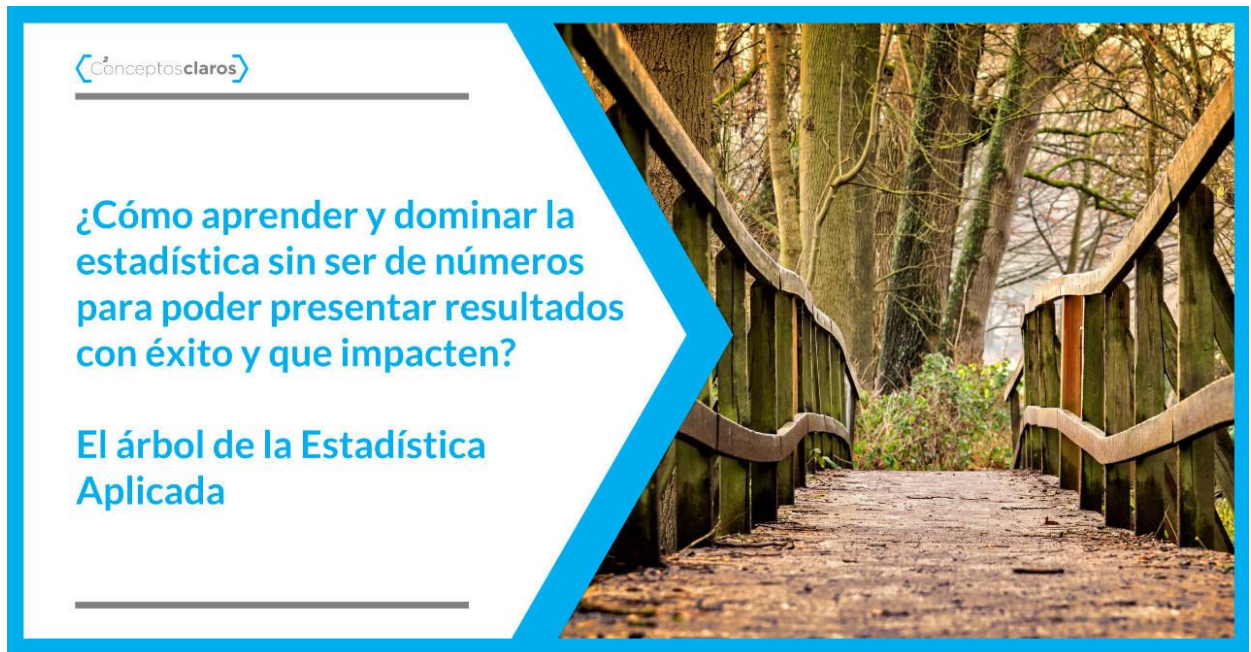


# ¿Cómo aprender y dominar la estadística sin ser de números para poder presentar resultados con éxito y que impacten?

## Descubre El árbol de la Estadística Aplicada



Ya lo sé.

No eres de números.

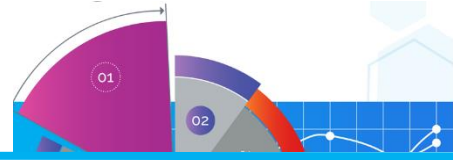
O eso crees.

Puede que alguna vez hayas odiado las mates y más la estadística.

En la universidad fue la asignatura que no te gustó y te costó más. El profe era terriblemente aburrido. Y lo que contaba también.

¡Vale! Acepto tu malestar.

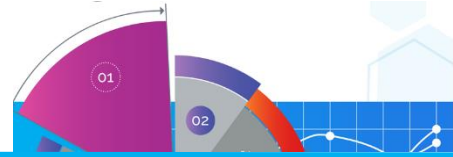
Déjame con este artículo ayudarte a ver los números desde otra perspectiva.



Y entender cómo ordenar tu mente para que puedas superar la estadística y la puedas utilizar en tu vida profesional divirtiéndote por el camino 😊

## Tabla de contenidos

- [# El síndrome del extraterrestre](#)
- [# Cómo ordenar la mente para lograr aprender con seguridad y disfrutando del camino](#)
- [# El árbol de la Estadística Aplicada](#)
- [# Las raíces: los conceptos Base](#)
  - [La tabla de datos y variables: tu materia prima](#)
- [Las dos estadísticas: descriptiva e inferencial](#)
- [# Las ramas: las técnicas estadísticas. Descubre el mapa de técnicas](#)
  - [Comparación de medias \(grupos\)](#)
  - [Comparación de proporciones](#)
  - [Relación de proporciones](#)
  - [Análisis de correlación](#)
  - [Modelos predictivos](#)
  - [Técnicas de Machine Learning](#)
- [# El tronco: la metodología y la herramienta](#)
- [# Con claridad es más fácil avanzar. ¡A por ello!](#)



## # El síndrome del extraterrestre

¡Si! Probablemente no sepas muy bien cómo aprender estadística y superarla de verdad.

Crees que no eres de números y esto no es para ti.

Te sientes un extraterrestre en el universo numérico. Sea cual sea. En mates, estadística o con datos.

Sabes que **necesitas entender los datos y la estadística** para ser mejor en tu trabajo, en tus proyectos.

Pero...

Se te hace un mundo.

Lo has intentando algunas veces. Pero **el tono que recuerdas en las clases de la universidad es aburrido y poco práctico.**

Esos recuerdos te alejan de poder superar la estadística de una vez por todas.

Algo te suena. Chi cuadrado, intervalos de confianza, p-valor, normalidad y alguna cosa más...

¡Tranquilo! No es tu culpa.

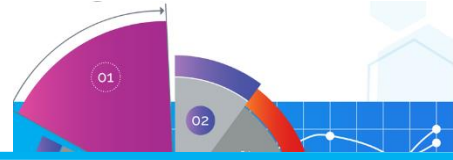
La estadística es una ciencia que acompaña una habilidad: el análisis de datos.

¡Eso es! Es la parte práctica que nunca te han contado.

Aprender e interpretar lo que observas. Esto es precisamente el análisis de datos. Y es lo que te enseñaré en el blog.

¡Bien!

Para aprender a analizar los datos y contribuir con grandes resultados en tus proyectos **necesitas superar la estadística.**



¿Por qué? Porque **es el lenguaje central para interpretar resultados.**

De nada sirve aprender a manejar herramientas de cálculo si no entiendas la esencia de los conceptos.

Y me da igual que pienses que no eres de números porque, en realidad, no es importante.

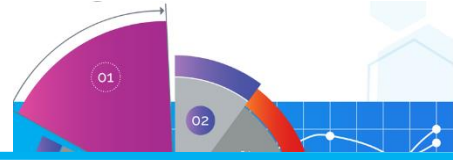
Es una etiqueta que hay en tu mente y que te han impuesto.

Crees que diferencia del que sabe leer y escribir muy bien, del que sabe contar y calcular muy bien.

¡Menuda tontería!

Ahora te voy a demostrar que lo que **necesitas es un camino para poder ordenar tu mente e ir superando la estadística paso a paso!**

¡Let's go!



## # Cómo ordenar la mente para lograr aprender con seguridad y disfrutando del camino

Es muy probable que tu inquietud sea hoy saber seleccionar el test o técnica para poder analizar los datos de tus investigaciones o proyectos.

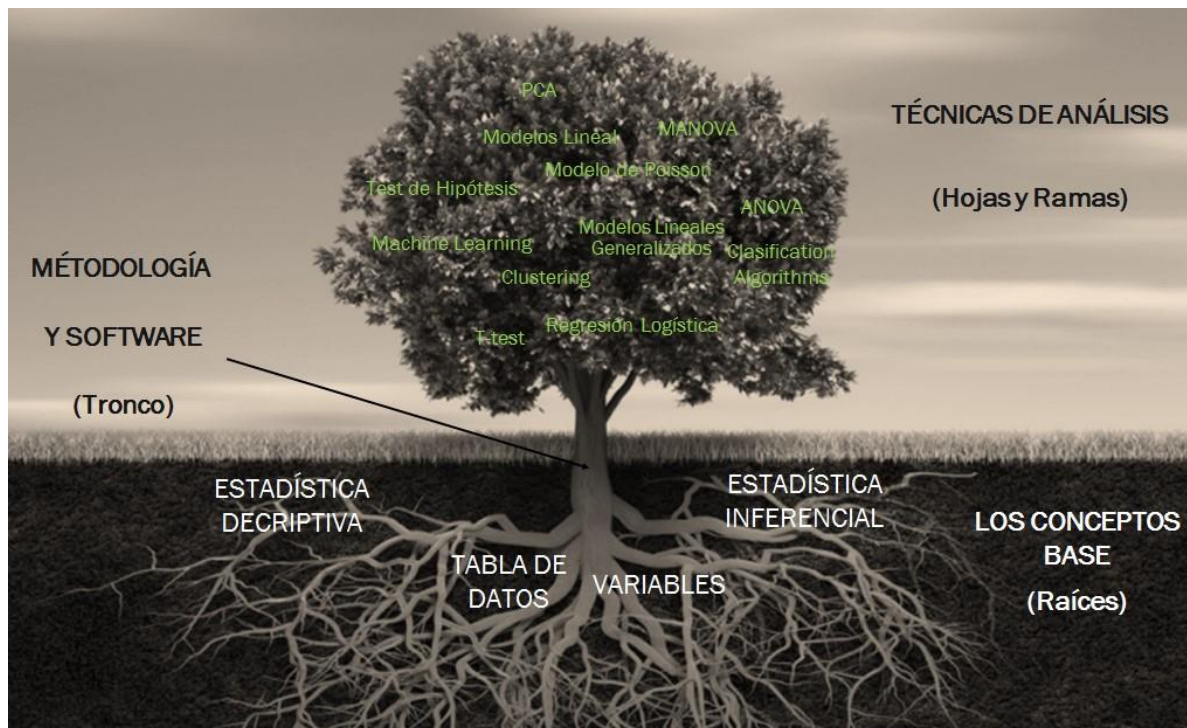
Sabes que existen algunas técnicas que podrías utilizarla pero no sabes cómo hacerlo. Te sentirías inseguro cuando lo intentes. Y más, pensando en la cantidad de técnicas que existen hoy.

Es normal, al principio puedes percibir que no tienes el control del conocimiento y todo te suene un poco a lo mismo.

Bien.

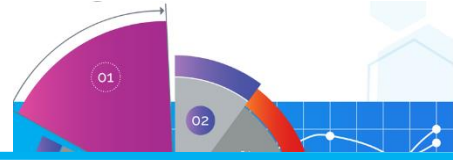
Para ayudarte, he decidido aportarte claridad.

¡Visualiza este árbol!



En él, está la respuesta a lo que andas buscando.

Seguramente estás pensando que soy un flipado pero, de verdad, creo que **está** es la clave del éxito y disfrute de tu aprendizaje.



Y fíjate muy bien en la imagen. Es diferente al árbol que hubieses dibujado 😊

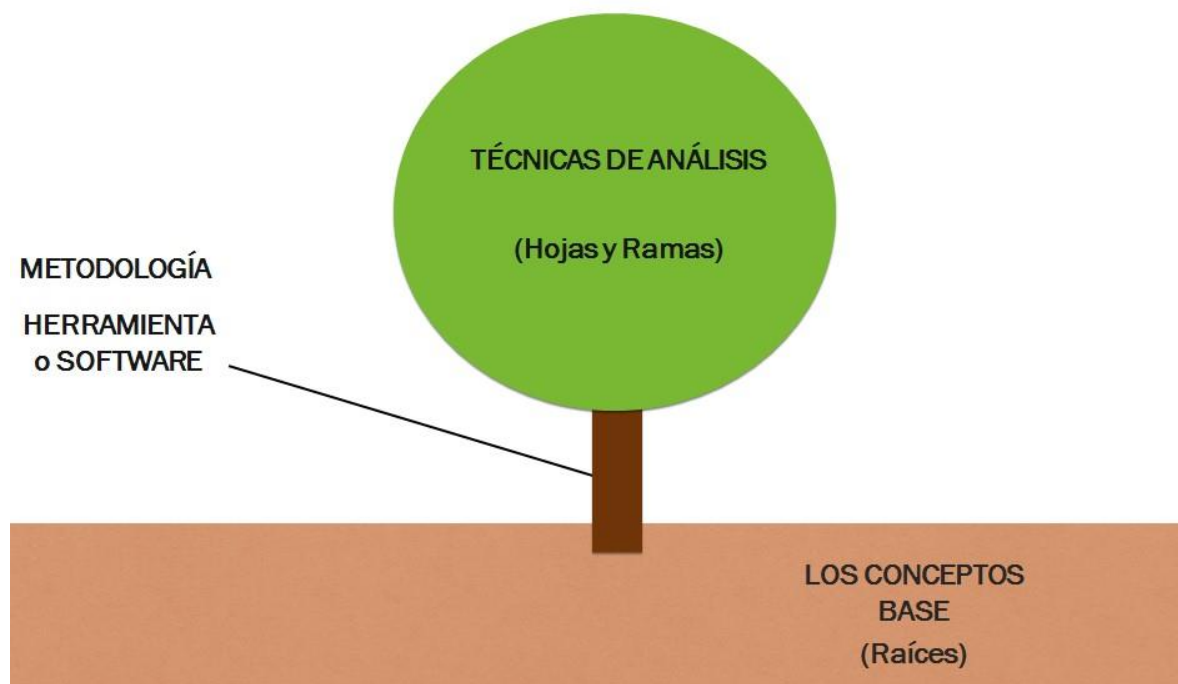
Aparece una parte muy interesante que seguramente no habías pensado.

Sabes de lo que hablo, ¿verdad?

## # El árbol de la Estadística Aplicada

El árbol que ves lo llamo **el árbol de la estadística Aplicada** (no soy muy original con los nombres, lo reconozco)

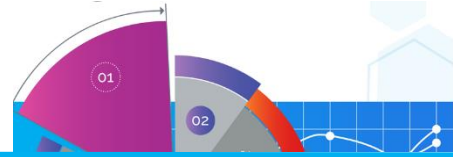
Déjame simplificar el árbol de conocimiento de una manera esquemática y sencilla:



¡Eso es! Tres partes fundamentales:

1. Las raíces: **los conceptos base**
2. El tronco: **la metodología y el software**
3. Las ramas: **las técnicas de análisis**

Ahora mismo tienes el foco en **qué técnicas aplicar y qué test utilizar**.



Te estás centrando en la parte alta del árbol. Lo que se ve. Las ramas.

La verdad es que existen muchísimas técnicas que puedes utilizar y, al principio, puedes sentirte agobiado y superado por tantas cosas a aprender. ¡Pero tranquilo!

Para que tu árbol de la estadística sea fuerte y hermoso. Con una gran copa. Con grandes ramas **necesitas alimentar muy bien las bases** (las raíces)

Y hacer **un tronco** suficientemente **robusto** como para poder manejar los datos como nadie.

Déjame explicarte las partes fundamentales del árbol de la estadística.

## # Las raíces: los conceptos Base

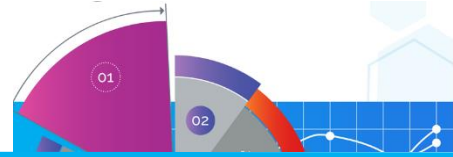
¡Esta parte es fundamental!

Aquí te encargarás de conocer con ejemplos prácticos los conceptos relacionados:

1. Con **las tablas de datos**. Cómo se organizan, qué son las variables y qué tipos existen.
2. Con **la estadística descriptiva**: conocer los conceptos fundamentales de descripción de datos y para qué se utilizan.
3. Con **la estadística inferencial**: los conceptos más esenciales que nos servirán para aprender y dominar las técnicas de las ramas del árbol



Voy a ayudarte un poco más a entender tus raíces.



## La tabla de datos y variables: tu materia prima

Te pongo un ejemplo. Imagínate que eres biólogo y quieres comparar especies de flores.

Lo primero que vas a hacer es recoger datos de flores que te interese estudiar, ¿verdad?

**Fisher's Iris Data**

Largo de sépalo	Ancho de sépalo	Largo de pétalo	Ancho de pétalo	Especies
5.0	2.0	3.5	1.0	I. versicolor
6.0	2.2	4.0	1.0	I. versicolor
6.0	2.2	5.0	1.5	I. virginica
6.2	2.2	4.5	1.5	I. versicolor
4.5	2.3	1.3	0.3	I. setosa
5.0	2.3	3.3	1.0	I. versicolor
5.5	2.3	4.0	1.3	I. versicolor
6.3	2.3	4.4	1.3	I. versicolor
4.9	2.4	3.3	1.0	I. versicolor
5.5	2.4	3.7	1.0	I. versicolor
5.5	2.4	3.8	1.1	I. versicolor
4.9	2.5	4.5	1.7	I. virginica
5.1	2.5	3.0	1.1	I. versicolor
5.5	2.5	4.0	1.3	I. versicolor
5.6	2.5	3.9	1.1	I. versicolor
5.7	2.5	5.0	2.0	I. virginica
6.3	2.5	4.9	1.5	I. versicolor
6.3	2.5	5.0	1.9	I. virginica
6.7	2.5	5.8	1.8	I. virginica
5.5	2.6	4.4	1.2	I. versicolor
5.7	2.6	3.5	1.0	I. versicolor
5.8	2.6	4.0	1.2	I. versicolor

Se trata de recoger las dimensiones (ancho y largo) de diferentes flores. Podemos distinguir dos aspectos en cuanto a la estructura:

- **Variables** = dimensiones de las flores  
= características = columnas
- **Observaciones** = cada flor que medimos = filas

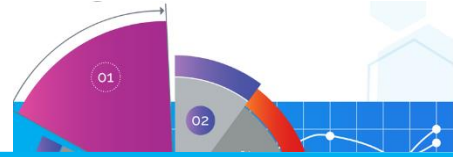
Nuestra muestra es el número de flores que vamos a medir. Imagínate N = 150, 150 filas con 5 variables. En total, en la imagen has recogido:

- Las dimensiones de las flores (4 variables)
- El tipo de especies (1 variable)

Como ves, la tabla de datos es una matriz de números y letras que engloban las características de lo que estás observando. Simplemente.

Y se organiza en forma de tabla. De matriz.

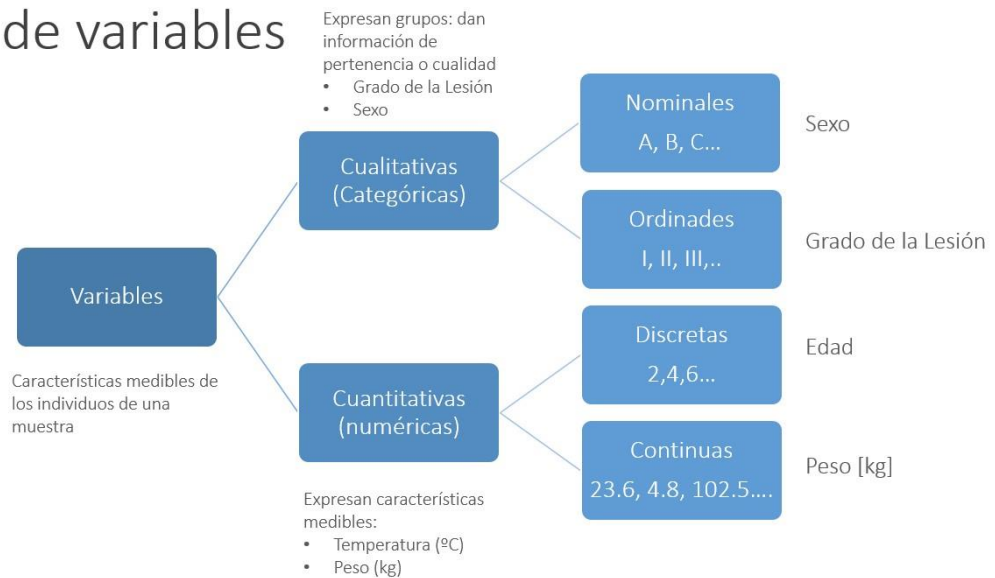
Vamos a entender un poquito mejor las variables. ¡Que es muy importante!



Ahora te ayudaré a entender qué tipos de variables existen.

Para no alargarme, con esta imagen tienes suficiente para conocer los tipos que hay. Fíjate con los ejemplos:

## Tipos de variables



Estamos avanzando con las raíces. ¡Vamos a seguir descubriendo los conceptos base!

## Las dos estadísticas: descriptiva e inferencial

Para construir unas buenas raíces necesitas entender las dos estadísticas. Es un paso fundamental antes de entrar en materia.

Una tabla de datos por sí sola no dice nada. Necesitas herramientas gráficas y pequeñas características para entenderla.

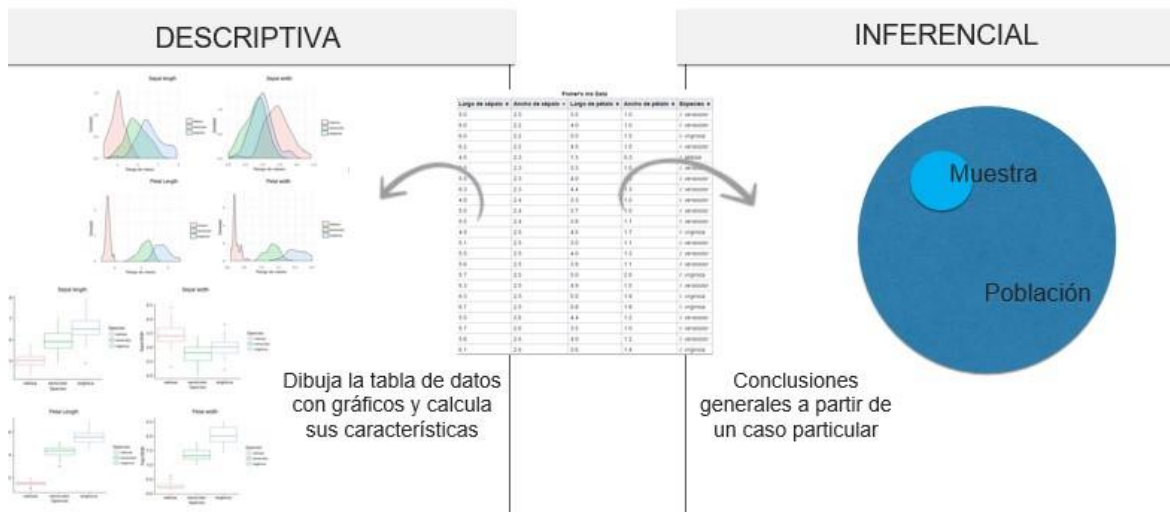
**La estadística descriptiva** dibuja y se utiliza para entender la información escondida de tu tabla de datos.

En cambio, **la estadística inferencial** pretende ir un paso más allá y aportar conclusiones generales a partir de los datos de tu muestra (de tu tabla de datos)

La población es el conjunto global y la muestra es la pequeña parte que has observado. Que has recogido datos.

Quédate con una idea maestra:

La **descriptiva** **dibuja** y te ayuda entender tu tabla de datos. La **inferencia** te ayuda a **responder preguntas globales** (de la población) a partir de tus datos (de tu muestra)



Para ir creciendo en estas dos estadísticas te voy a listar los conceptos que creo que son fundamentales para que tengas unas raíces sólidas y te permitan crecer muy rápido.

Simplemente tendrás que ir aprendiendo este listado de conceptos poco a poco para poder avanzar, más adelante, mucho más rápido en tu aprendizaje.

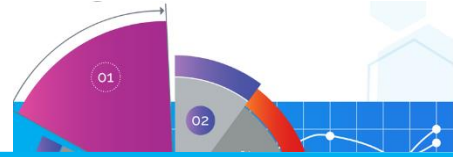
**Estadística descriptiva:** podemos clasificar los conceptos en función de qué variables describimos o dibujamos:

Variables cuantitativas:

1. Tabla de frecuencias
2. Valor central (media o mediana))
3. Dispersión (desviación estándar, rango intercuartílico)
4. Medidas de posición (cuartiles, percentiles)
5. Distribución (histograma, boxplot, densidad)
6. Correlación y covarianza
7. Scatterplot y matrixplot

## Variables cualitativas

- ## 1. Tablas de contingencias



2. Concepto de proporciones
3. Distribución
4. Diagrama de barras y sectores

En la **estadística inferencial** es vital entender:

1. El concepto de muestra y población
2. Qué son las distribuciones de variables
3. Qué es y para qué sirve la distribución de densidad de probabilidad y las probabilidades
4. El poder de la distribución normal
5. Qué son los intervalos de confianza y su gran poder
6. Tipos importantes de distribuciones de probabilidad
7. El contraste de hipótesis y el p-valor
8. La esencia de los modelos predictivos y sus dos super poderes

Viendo estos conceptos con ejemplos puedes nutrir las raíces de gran alimento para seguir creciendo y aprender técnicas avanzadas de análisis.

¡Let's go!

## # Las ramas: las técnicas estadísticas. Descubre el mapa de técnicas

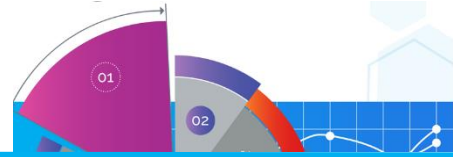
Tu árbol crecerá y cada vez entenderás mejor las técnicas o test estadísticos para obtener grandes resultados de tus datos.

Las ramas de tu árbol pueden ser increíblemente grandes.

El tema es que ¡existen muchísimas técnicas! Y al principio, puede parecer una locura aprender todo.

Para ayudarte a reducir la complejidad he clasificado las técnicas en 6 problemas tipo para que de esta forma te quede mucho más claro en qué situación te encuentras.

Ya verás que te sentirás mucho más tranquilo 😊



Te presento el mapa de técnicas de Análisis



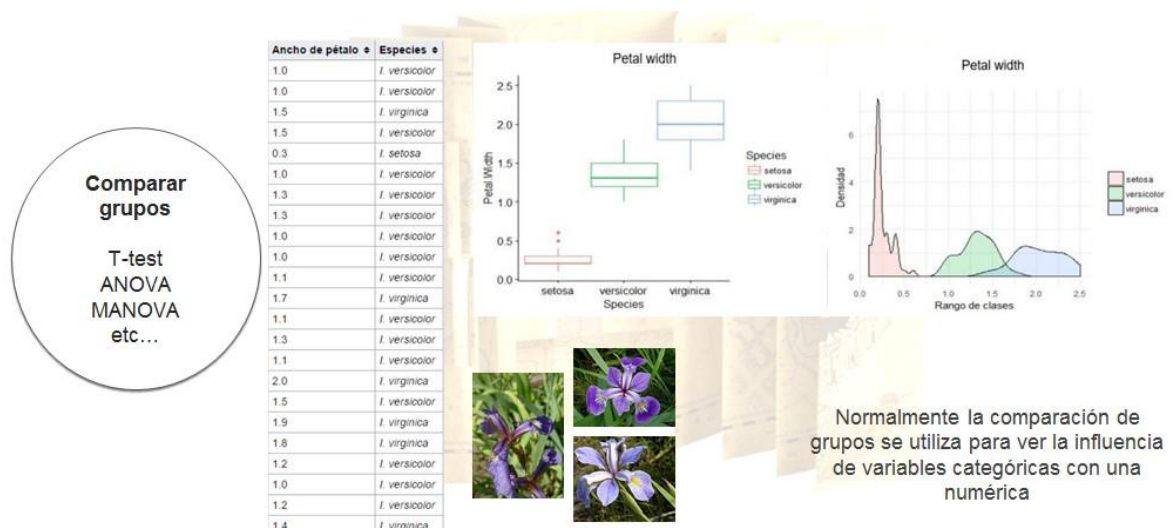
La figura que ves a continuación están representados los problemas tipos que nos servirán para aprender las técnicas 😊

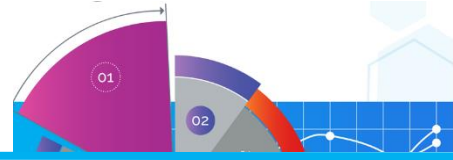
Voy a tratar de ayudarte con varios ejemplos de cada problema tipo 😊

## Comparación de medias (grupos)

En este ejemplo tratamos de comparar tres especies de flores.

La característica que nos interesa comparar son las dimensiones de las flores. (cuantitativa)





Se trata de observar si tenemos diferencias significativas entre especies y en qué característica tenemos más diferencias.

¿Se ve?

- Una variable cuantitativa que ejerce de variable respuesta. Es la variable que quieres analizar.
- Y los grupos es la variables de estudio. En este caso los grupos de especies. Tenemos 3 tipos.

En los gráficos que ves en la imagen puedes observar que la variable respuesta es el ancho del pétalo. Y comparamos en función de la especie.

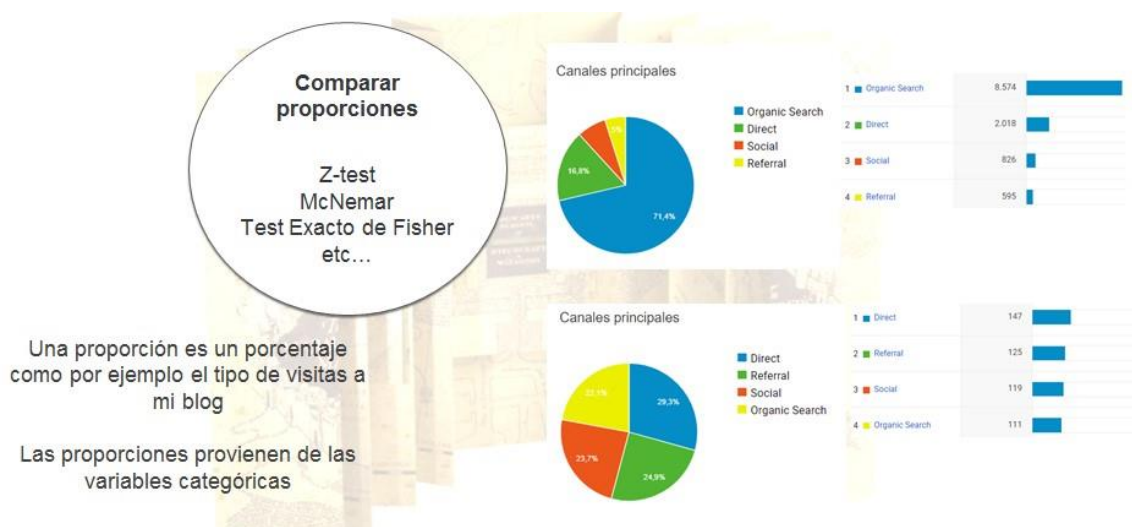
¡Voilà! Esta es la esencia de la comparación de media 😊

### Comparación de proporciones

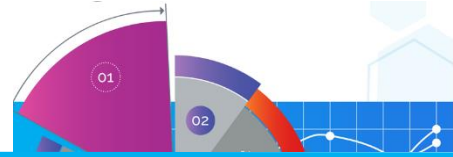
Antes hemos comparado distribuciones (medias) siempre teniendo en cuenta variables cuantitativas.

Podemos también comparar proporciones o porcentajes que viene a ser lo mismo.

Tienes algunos ejemplos de proporciones en esta imagen. Los porcentajes de los quesitos son proporciones 😊



El objetivo de este problema tipo es comparar porcentajes entre grupos.



Aquí debajo, un ejemplo bastante típico.

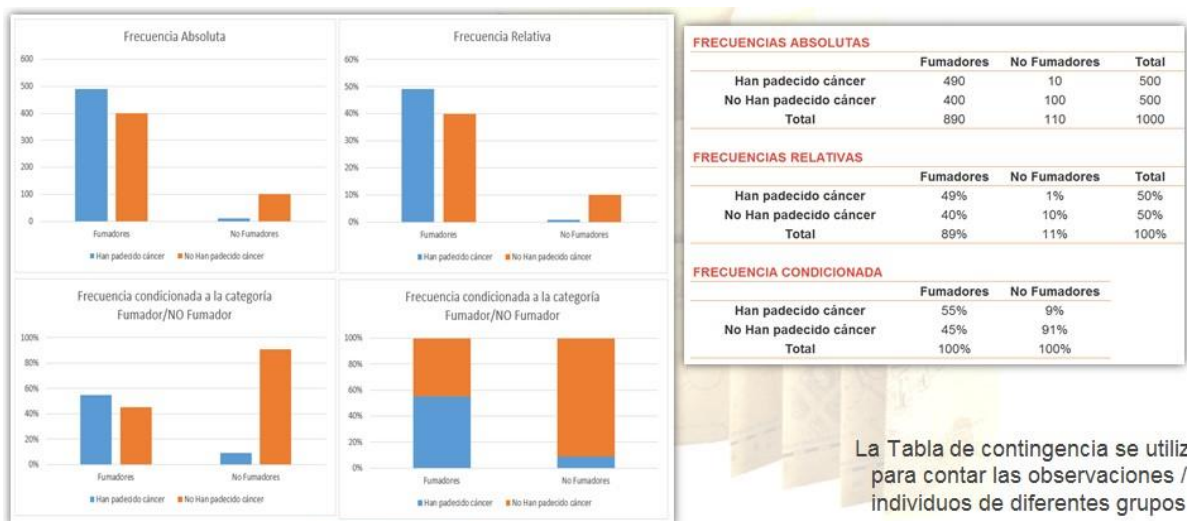
Queremos demostrar si fumar es una práctica que favorece a la aparición del cáncer de esófago. ¿Cómo?

Por ejemplo ver si las proporciones de fumadores y no fumadores es la misma para el grupo que han tenido cáncer. ¿Se ve?

Aquí puedes ver dos variables cualitativas:

- Variable cualitativa 1: si han tenido cáncer o no
- Variable cualitativa 2: si son fumadores o no

Intento buscar las proporciones en la tabla que te muestro a continuación 😊



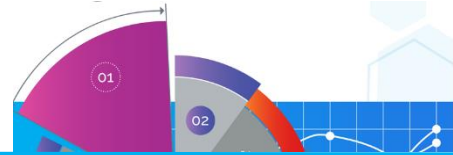
## Relación de proporciones

A parte de comparar puedes relacionar o asociar.

Por ejemplo, se trata de ver si el tipo de tarea de casa está relacionado con quién hace la tarea.

Dos variables cualitativas:

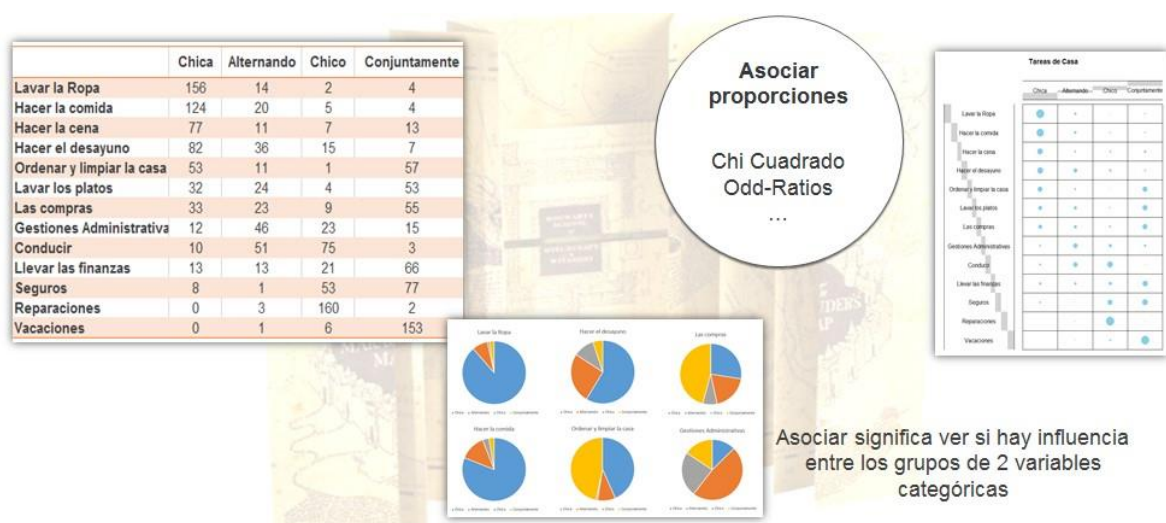
- Variable cualitativa 1: el tipo de tarea
- Variable cualitativa 2: el encargado de hacer la tarea



Puedes resumir en una tabla de contingencias (tabla de la imagen) para observar la cantidad de familias por parejas de grupos. Por ejemplo, tenemos 156 familias que lavar ropa se encarga la chica de la relación.

En este ejemplo estás tratando de ver si lavar ropa tiene que ver con si es la chica encargada de hacer la tarea o no. Si las vacaciones se planean conjuntamente o si el encargado de las reparaciones es el chico de la familia.

O si hacer la compra se encarga el chico o conjuntamente.



Buscamos relaciones entre grupos. ¡Eso es!

En este tipo de problemas de relación trabajamos con dos variables cualitativas

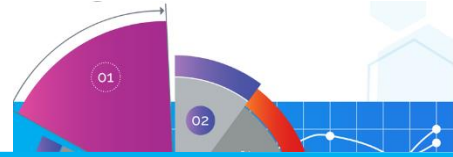
## Análisis de correlación

También podemos relacionar medidas. Parecido al problema anterior.

Aquí se trata de ver si el consumo del coche en combustible tiene relación con el peso.

Para ello se han recogido diferentes modelos de coches y se ha tabulado el consumo y el peso de cada modelo.

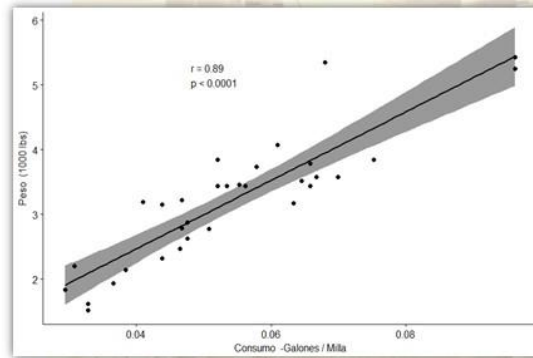
Podemos crear un **diagrama de puntos o dispersión** como el de la imagen. Cada punto representa un modelo de coche.



Si te fijas, estos puntos siguen una “tendencia” hacia arriba. (como la recta que observas)

¡Voilà! Estás entendiendo qué es una relación entre medidas. La **correlación**.

consumo	peso
0.04761905	2.620
0.04761905	2.875
0.04385965	2.320
0.04672897	3.215
0.05347594	3.440
0.05524862	3.460
0.06993007	3.570
0.04098361	3.190
0.04385965	3.150
0.05208333	3.440
0.05617978	3.440
0.06097561	4.070
0.05780347	3.730
0.06578947	3.780
0.09615385	5.250
0.09615385	5.424
0.06802721	5.345
0.03086420	2.200



**Asociar var. numéricas**

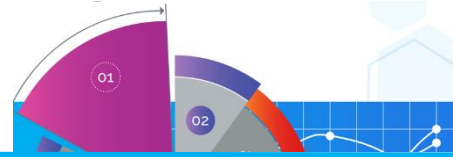
Análisis de Correlación

Asociar variables numéricas es encontrar numéricamente cómo de relacionadas están 2 variables

En este caso se intuye que a mayor peso de los coches, mayor es el consumo como es lógico pensarlo.

En muy chico, puedes ver el  $r=0.89$ . Este coeficiente nos está indicando que la relación es muy alta porque está cercana a 1.

Y es **positivo**, es decir una relación directa o creciente (la recta va hacia arriba)



## Modelos predictivos

¡Me encantan los modelos! Y su potencia es brutal. Pero como dice el tópico, la potencia sin control no sirve de nada 😊

¿Pero qué son y para qué sirven los modelos predictivos?

Déjame ponerte un ejemplo que será mucho más fácil para explicártelo.

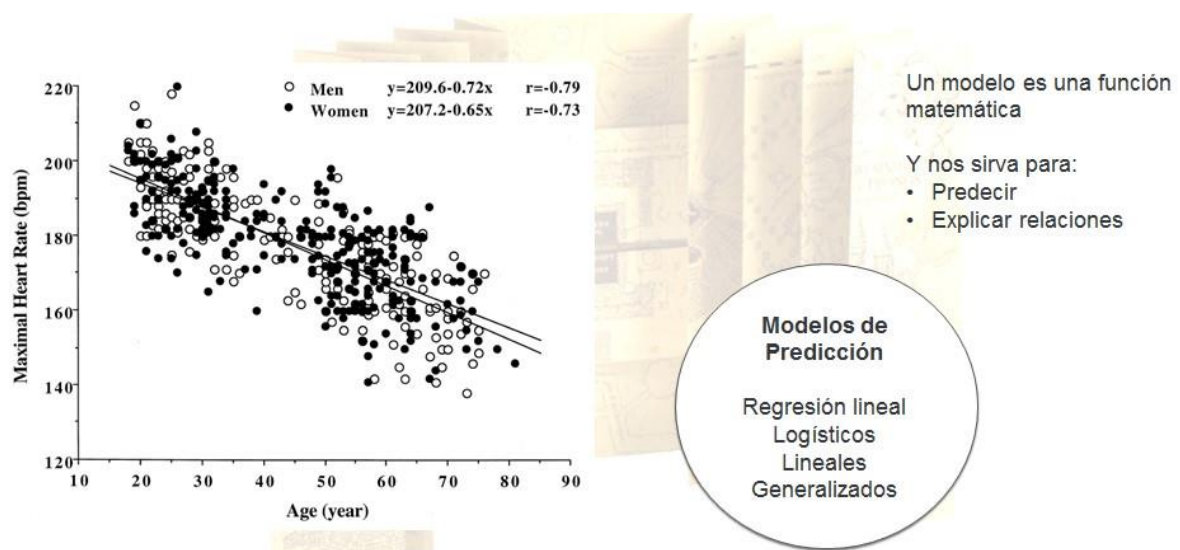
Lo que queremos es poder **predecir** la frecuencia cardíaca máxima sabiendo únicamente la edad de los pacientes.

Encontrar una relación matemática que me permita “más o menos” saber qué frecuencia cardíaca máxima sabiendo solo la edad.

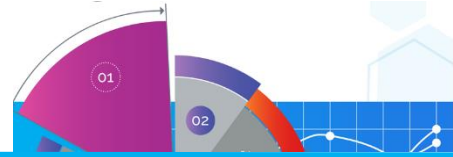
Para construir esta función matemática necesitas datos de diferentes pacientes. Pacientes que se les han medido la frecuencia cardíaca máxima y recogido la edad.

Montamos el **diagrama de dispersión** y calculamos una función **matemática**. La más simple. La recta.

Fíjate que esta recta sigue muy bien a los puntos. O al menos, es la recta que dibujarías siguiendo los puntos.



Fíjate que la **pendiente es negativa**. Quiere decir que a mayor edad menor es la frecuencia cardíaca máxima. Es lógico pensarlo. A mayor edad, el corazón es más débil.



Acabas de ver los **dos grandes super poderes** de los modelos.

- **Explicar relaciones causa-efecto** (a mayor edad, menor es la frec. Cardíaca máxima)
- **Predecir resultados sin necesidad de medir** (con la ayuda de la función matemática)

Existen tantos modelos como funciones matemáticas te puedas imaginar. Es decir muuuuchoos! Como siempre hay que conocer los que mejor funcionan.

- Regresiones lineales
- Modelos logísticos
- Modelos lineales generalizados
- ...

Lo más importante es entender **qué te están aportando y para qué se utilizan**, Y creo que te he ayudado a hacerlo 😊

### Técnicas de Machine Learning

Y llegamos a las técnicas más de moda hoy en día.

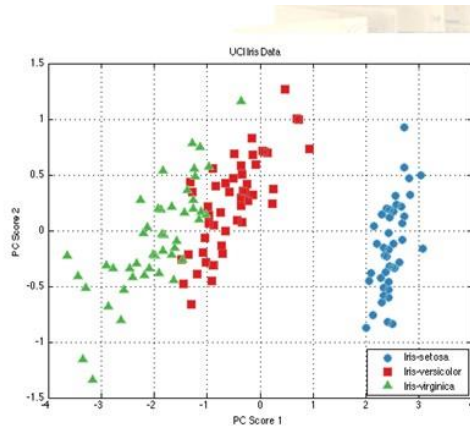
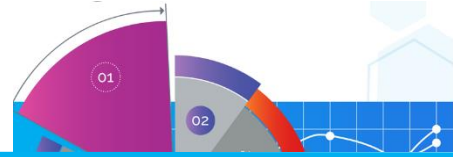
Es el siguiente paso. Y está muy relacionado con los modelos predictivos en muchas ocasiones.

Son técnicas que te permiten encontrar **patrones** que se repiten en tus datos.

Por ejemplo, el **clustering** o agrupación nos permite agrupar la muestra en diferentes grupos teniendo en cuenta únicamente las características que hemos observado.

Sin saber, a priori, en qué grupo pertenecen.

La **clasificación**, en cambio, te permite a partir de unos datos observados saber en qué grupo se encuentra **entrenando un modelo matemático** como puede ser una red neuronal.



Básicamente los más comunes son:

- Clustering - agrupamiento
- Clasificación
- Regresión

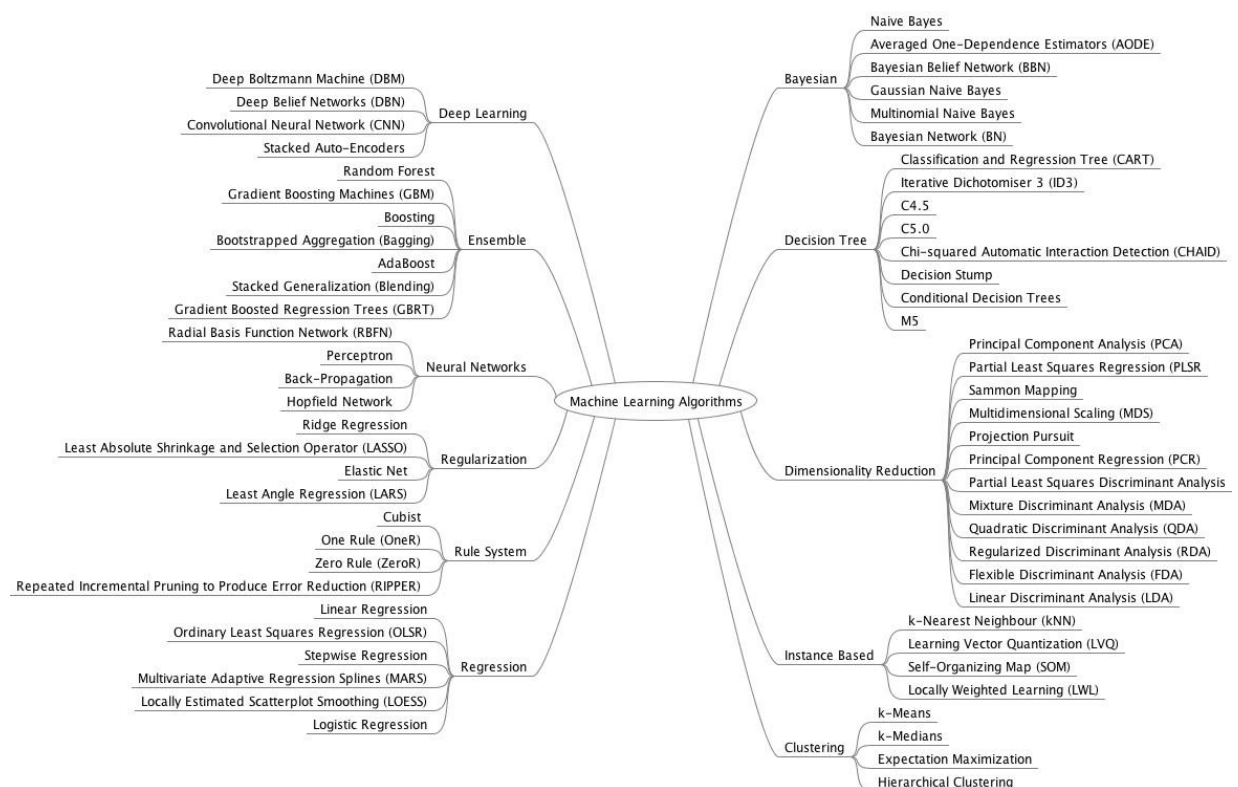
Si no sabes a qué grupo pertenece trabajas a priori  
→ **Clustering**

Si sabes el grupo a que pertenece lo que vas a entrenar a un modelo para que clasifique con nuevos datos  
→ **Clasificación** (redes neuronales)

Aquí el mapa global de las técnicas de machine learning.

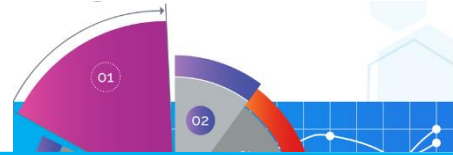
El mundo del análisis de datos se abre a tus pies dando lugar a un nuevo universo.

¡Sencillamente increíble!



Fuente: [Machine Learning Mastery](#)

Hasta aquí te he ayudado a entender las raíces de tu arbol y las ramas.



Ahora voy a citarte algo más sobre la metodología y el software.

¡Seguimos!

## # El tronco: la metodología y la herramienta

Para poner en práctica y aprender las técnicas de análisis es muy importante que conozcas la metodología necesaria para tratar de llevar a cabo todo este conocimiento.

Y además, una calculadora (o software) que te permita calcular los análisis que necesitas.

En resumen.

Necesitas un software y un método.

Hoy no pretendo hablarte de los detalles pero si quiero listarte los pasos de la metodología que te explicaré a fondo en otro post. Estos son los 4 pasos:

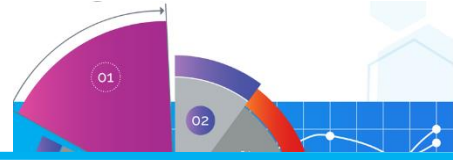
1. **Objetivo:** definir qué es lo que quieres encontrar en tus datos
2. **Descripción:** ¿emplearás la estadística descriptiva
3. **Análisis:** emplearás la estadística inferencial
4. **Conclusión:** dar respuestas a tus preguntas a partir de los resultados que has obtenido

En cuanto al software. Lo más importante es seleccionar la herramienta del mercado que mejor se adapte a tu caso.

Tus necesidades te ayudarán a decidirte. En otro artículo te ayudo un poco más en este tema. Pero déjame darte la visión global de las tres aplicaciones:

### Aplicación 1: estudios científicos

Se trata de poder analizar los datos de estudios del mundo de la salud, biología, análisis de mercados. Disciplinas científicas que requieran solucionar problemas como por ejemplo la comparación de tratamientos, de grupos de pacientes etc. Requiere de programas a base de clicks con el ratón.



### Aplicación 2: ciencia de datos

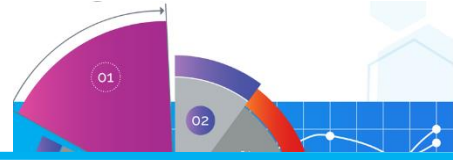
Son aplicaciones con necesidad alta de técnicas avanzadas y generalmente grandes volúmenes de datos. Requiere que seas más flexible y con capacidad para programar.

### Aplicación 3: reporting y pequeños cálculos

Podemos utilizar hojas de cálculo para poder tabular los resultados y presentar sencillos gráficos descriptivos y ordenar tablas de resultados.

...

Hasta aquí el árbol de la estadística. En otros artículos te ayudaré a comprender mejor la metodología y el tipo de softwares que puedes utilizar.



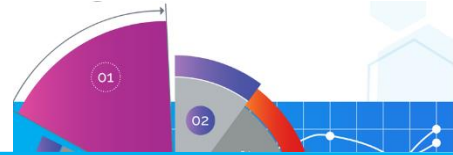
## # Con claridad es más fácil avanzar. ¡A por ello!

Ahora que tienes la foto grande y te he relajado el agobio que tenías (o eso espero) te propongo que te centres en aprender los conceptos base (las raíces).

Te recuerdo los tres grandes pilares. Quiero que entiendas muy bien:

1. Cómo se organiza una tabla de datos y sus variables
2. Los principales conceptos de la Estadística Descriptiva
3. Los principales conceptos de la Estadística Inferencial

Antes te he listado los conceptos que creo que son clave para poder crecer en la habilidad de análisis de datos.



## ¿Y AHORA QUÉ?

Durante varios días te iré enviando información interesante para que puedas ir progresando en tu proyecto de tesis y de investigación. (O simplemente quieras mejorar tu capacidad de jugar con datos) ;)

## ¿ME AYUDAS A COMPARTIR LA GUÍA?

Si te ha parecido útil esta guía y quieres compartirla con tus colegas de profesión o colegas investigadores puedes utilizar los botones de las redes sociales.

¡Haz click en el botón de tu red social favorita para compartir la guía!



¡Muchas gracias!

¡Ciao ciao!

Ayudo a estudiantes de doctorado e investigadores a analizar los datos de sus tesis para que puedan presentarlas con éxito ante el tribunal y consigan el reconocimiento y el trabajo investigador de sus sueños



Jordi Ollé

Autor de *ConceptosClaros.com*