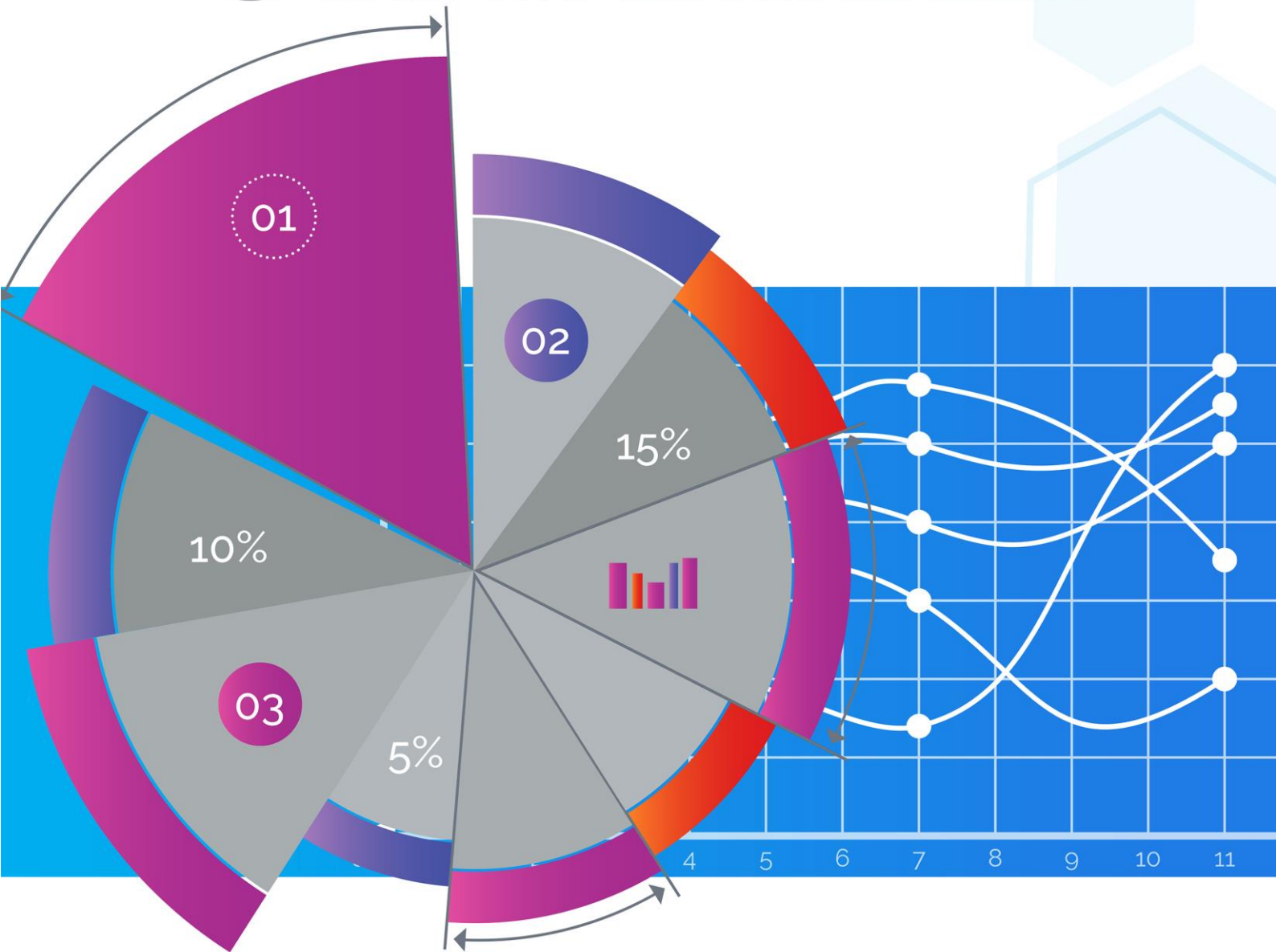
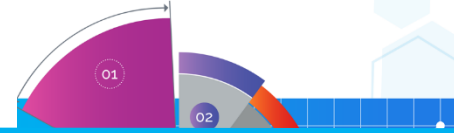


6 TÉCNICAS ESTADÍSTICAS INDISPENSABLES



Si de verdad quieres conseguir Analizar los Datos de tu proyecto investigador o Tesis doctoral con éxito y garantías



¡HOLA! Bienvenid@ a Conceptos Claros

¡Hola! Gracias por estar aquí conmigo y entrar en la comunidad. ¡Es un orgullo para mí tenerte aquí!

A lo mejor no me conoces. Así que te voy a contar algo sobre mí antes de desvelarte esta guía.

Soy Jordi Ollé el autor de conceptosclaros.com y quiero ayudarte a analizar los datos de tu investigación, tesis doctoral o tu proyecto profesional para que puedas presentarlo con éxito y consigas el reconocimiento y el trabajo investigador de tus sueños.

Mi recorrido vital

Soy Jordi. Amante de las cosas “complejas”.

Vivo en Barcelona con mi compañera de viaje Anna. Soy una mente inquieta y me encanta ser un sherpa de los datos.

Después de 7 años en el departamento de I+D de una multinacional, entendí que aplicar estadística y matemáticas en los datos es apasionante.

Entender con una visión práctica estas ciencias es especial. Me encanta formarme día tras día y buscar un sentido práctico.

Estudié Ingeniería Industrial y el Máster Oficial en Ingeniería Biomédica en la especialidad de señales y Análisis de Datos.

Tengo una capacidad innata para resumir y hacer fácil lo difícil. No temas, estás en buenas manos ;)



¿Por qué nació Conceptos Claros?

Por otro lado descubrí que me encanta ser un guía, un mentor, un sherpa de otros que están un poco más atrás que yo. Quiero hacer crecer a los demás con mucho cariño y dedicación.

Comprobé este hecho trabajando con estudiantes de doctorado codo con codo. ¡Ellos quedaron encantados con mi compañía y ayuda!

Las 6 Técnicas Estadísticas Indispensables

Me di cuenta que puedo ayudar a muchas más personas. Quiero que tú aprendas como ellos lo hicieron.

¿Por qué puedes confiar en C2?

Es un poco extraño que lo diga yo mismo. Pero voy a dejar a los datos que hablen por si solos:

En poco más de un año de vida del blog he conseguir más de 15.000 visitas al mes, 4222 suscriptores y cada día se amplía la familia entre 5-15 nuevos integrantes como tú.

He formado y ayudado con mis contenidos Premium a 256 investigadores y profesionales técnicos en sus tesis doctorales o proyectos de investigación. Cada día el blog tiene más impacto.

¡Y me hace muy feliz!

No digo todo esto porque veas lo “guapo que soy” ni lo “bonito que es el blog”.

Si no para que veas realmente que este blog ayuda a los demás y quiero con todas mis fuerzas ayudarte en el duro camino del investigador y el estudiante de doctorado.

En nada, recibirás acceso al curso Pro Statistics, un curso gratuito que he ideado para que situes la estadística dónde realmente es interesante ;)

¡Espero que lo disfrutes!

...

Pero antes, puedes disfrutar de una guía con las 6 técnicas estadísticas indispensables para analizar los datos de cualquier tesis doctoral y cómo empezar a utilizarlas sin morir en el intento

¡Espero que te ayuden y las puedas aplicar muy hoy mismo!



Jordi Ollé

Autor de *ConceptosClaros.com*



LA TABLA DE DATOS Y LAS VARIABLES

¿Cómo ordenar y clasificar tus datos antes de nada?

Empezamos la guía con algo esencial. Saber ordenar y clasificar las variables de una base de datos.

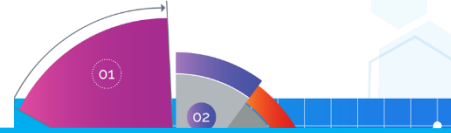
El análisis de datos es una ciencia que te permite interpretar y manipular datos.

- Los datos son observaciones de la realidad.
- Los datos se ordenan en forma de tabla de datos.

Fíjate en este ejemplo:

Imagínate que eres un biólogo muy interesado en estudiar tipos de flores y sus diferencias. Has recogido características de las dimensiones y la especie de 150 flores. Aquí tienes la tabla de datos:

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>I. versicolor</i>
6.0	2.2	4.0	1.0	<i>I. versicolor</i>
6.0	2.2	5.0	1.5	<i>I. virginica</i>
6.2	2.2	4.5	1.5	<i>I. versicolor</i>
4.5	2.3	1.3	0.3	<i>I. setosa</i>
5.0	2.3	3.3	1.0	<i>I. versicolor</i>
5.5	2.3	4.0	1.3	<i>I. versicolor</i>
6.3	2.3	4.4	1.3	<i>I. versicolor</i>
4.9	2.4	3.3	1.0	<i>I. versicolor</i>
5.5	2.4	3.7	1.0	<i>I. versicolor</i>
5.5	2.4	3.8	1.1	<i>I. versicolor</i>
4.9	2.5	4.5	1.7	<i>I. virginica</i>
5.1	2.5	3.0	1.1	<i>I. versicolor</i>
5.5	2.5	4.0	1.3	<i>I. versicolor</i>
5.6	2.5	3.9	1.1	<i>I. versicolor</i>
5.7	2.5	5.0	2.0	<i>I. virginica</i>
6.3	2.5	4.9	1.5	<i>I. versicolor</i>
6.3	2.5	5.0	1.9	<i>I. virginica</i>
6.7	2.5	5.8	1.8	<i>I. virginica</i>
5.5	2.6	4.4	1.2	<i>I. versicolor</i>
5.7	2.6	3.5	1.0	<i>I. versicolor</i>
5.8	2.6	4.0	1.2	<i>I. versicolor</i>
6.1	2.6	5.6	1.4	<i>I. virginica</i>



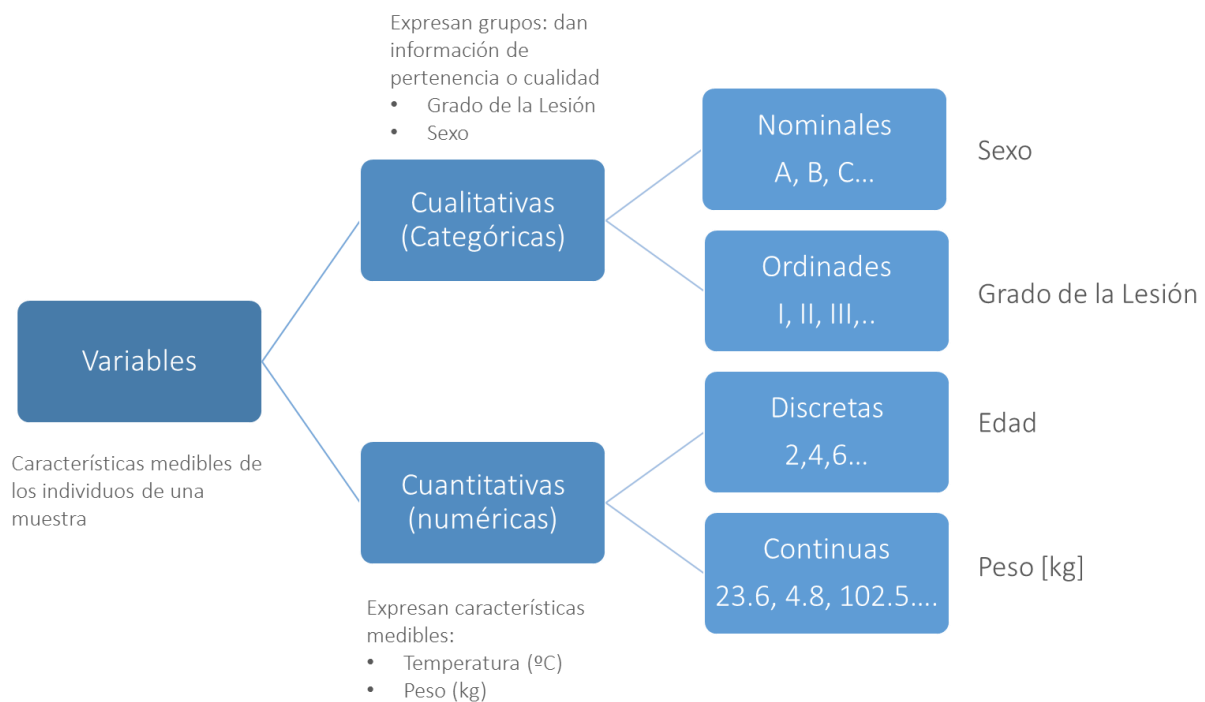
- Las **VARIABLES** son las **características** que observamos (largo, ancho del pétalo y del sépalo)
- Las **OBSERVACIONES** son las **veces** que **medimos** las variables (las flores que medimos. En este caso tenemos 150)

LA TABLA DE DATOS SE ORDENA EN FORMA DE MATRIZ (FILAS y COLUMNAS)

- Variables en columnas
- Las observaciones en filas

LAS VARIABLES SE PUEDEN CLASIFICAR EN 4 TIPOS

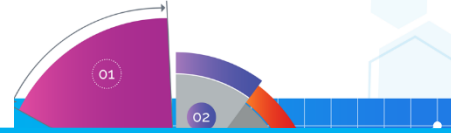
- **Cuantitativas:** expresan características medibles
 - **Discretas:** no tienen decimales
 - **Continuas:** tienen decimales y tienen unidades
- **Cualitativas:** expresan etiquetas y dan información de pertenencia a un grupo
 - **Nominales:** grupos
 - **Ordinales:** los grupos tienen orden de escala





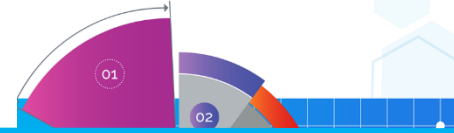
Las 6 Técnicas Estadísticas Indispensables

6 TÉCNICAS ESTADÍSTICAS
INDISPENSABLES



En el Excel adjunto verás esta información en la primea hoja.

¡Asigna el tipo de variable con la ayuda de esta clasificación y estructura tus datos en forma de tabla como la que has visto!



LA CENTRALIDAD

¿Cómo cuantificar el centro de tus datos?

En el 99.9% de los proyectos de investigación utiliza medidas de centralidad. En un momento u otro te tocará a ti utilizarlas ¡Seguro!

¡Allá va su utilidad!

Para las **variables cuantitativas** puedes calcular el centro de estas variables.

El **centro o centralidad** nos indica **dónde se concentra** la mayoría de observaciones o individuos de tu muestra. (¡Perdona! La muestra es la tabla de datos que has recogido. En el ejemplo del biólogo, son las 150 flores que hemos recogido en la tabla de datos, ¿se ve?)

La centralidad básicamente se puede calcular de dos maneras: MEDIA y MEDIANA

MEDIA o PROMEDIO

Es una **fórmula matemática**.

Es la **suma de todas las observaciones de la variable repartida por el total de observaciones** que tenemos.

Si la variable es el peso de 150 individuos. Calculamos el peso total de los 150 individuos y lo dividimos entre 150. (Dividimos el peso total en partes iguales para los 150 pacientes)

Este peso es el peso promedio.

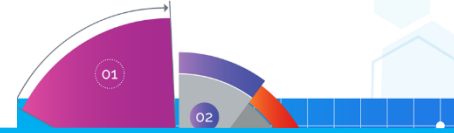
Si el peso de **10** pacientes en kg es 59, 65, 57, 49, 85, 95, 105, 78, 64, 101 la media es:

$$\frac{59 + 65 + 57 + 49 + 85 + 95 + 105 + 78 + 64 + 101}{10} = 75.8 \text{ kg}$$

(Más detalle en la hoja Excel)

Es el **centro de gravedad** de la variable. El **punto de equilibrio** si lo quieres ver así.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N}$$

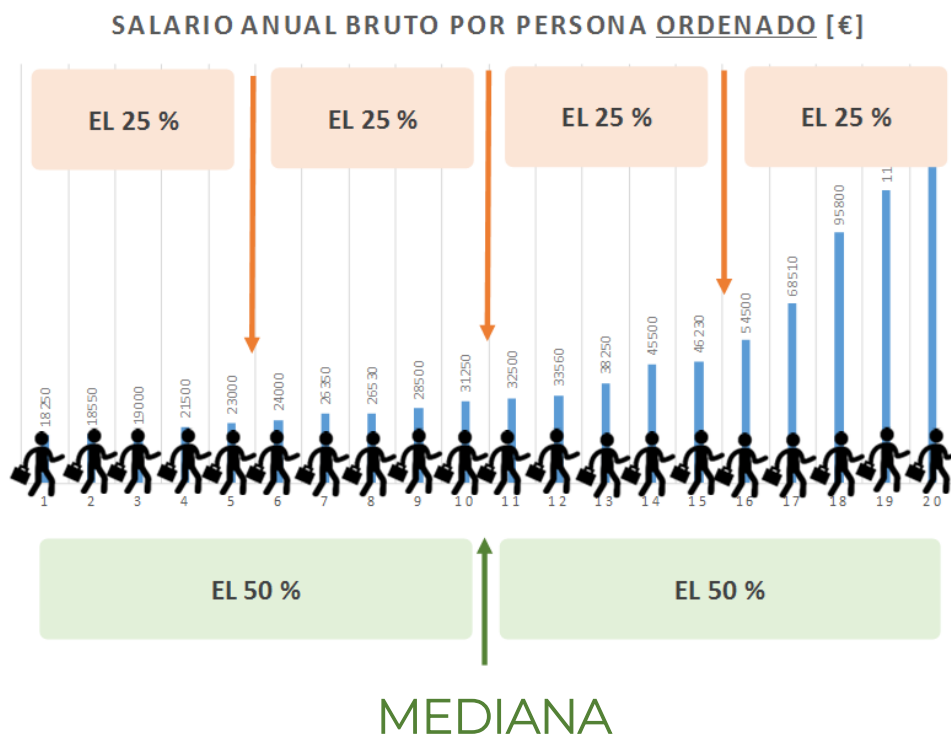


MEDIANA

La **mediana** es el valor frontera que deja el 50% de las observaciones más pequeños a un lado y el 50% de las observaciones más grandes al otro lado. Un poco extraño, ¿verdad?

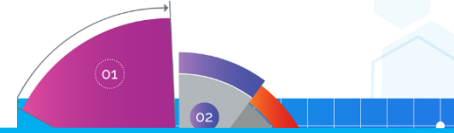
Por ejemplo, hemos recogido la variable sueldo anual bruto de 20 trabajadores. El centro del sueldo anual de los trabajadores de la empresa se calcula siguiendo estos puntos:

1. Ordenando el sueldo de menor a mayor
2. Cortamos por el centro para dejar el 50% de trabajadores con el sueldo menor a un lado y el 50% de los trabajadores con el sueldo mayor al otro lado.
3. Esa frontera es la mediana



En el Excel te explico cómo calcular la centralidad de una variable con el promedio y la mediana.

¡Calcula la centralidad a tus variables cuantitativas!



LA DISPERSIÓN

¿Cómo cuantificar la variabilidad?

Otro aspecto muy importante que se repite una y otra vez en tesis doctorales o proyectos de investigación es el concepto de dispersión y cómo cuantificarlo.

¡En esta sección te explico cómo!

La **variabilidad** o la **dispersión** mide **cuánto de alejados** están las observaciones de la variable **al valor de centralidad** que has calculado antes.

Hay dos formas de calcular la dispersión:

- Con la **desviación estándar**
- Con el **rango intercuartílico**

LA DESVIACIÓN ESTÁNDAR

La **desviación estándar** se puede definir como “el promedio de las distancias a la media”

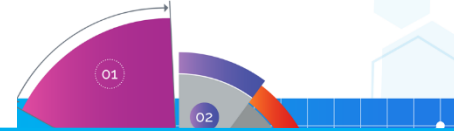
Si te gusta entender la fórmula matemática sigue estos puntos (sino pasa de ellos):

- Se calculan las distancias de las observaciones a la media. $(x_i - \bar{x})$
- Se elevan al cuadrado para que sean positivas $(x_i - \bar{x})^2$
- Después se hace la media de estas distancias al cuadrado. (Se suman y se dividen por el número de observaciones)
- Y finalmente se aplica la raíz cuadrada para quitar el cuadrado que hemos puesto antes.

Sí, son matemáticas. ¡Lo sé! Pero quédate con la idea que la **desviación estándar mide cuánto de lejos están las observaciones a la media**. Mide cuánto está variando la variable en promedio ☺

$$S_x = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}}$$

No te preocupes por la fórmula. No hace falta que la entiendas para utilizar la desviación estándar ;)



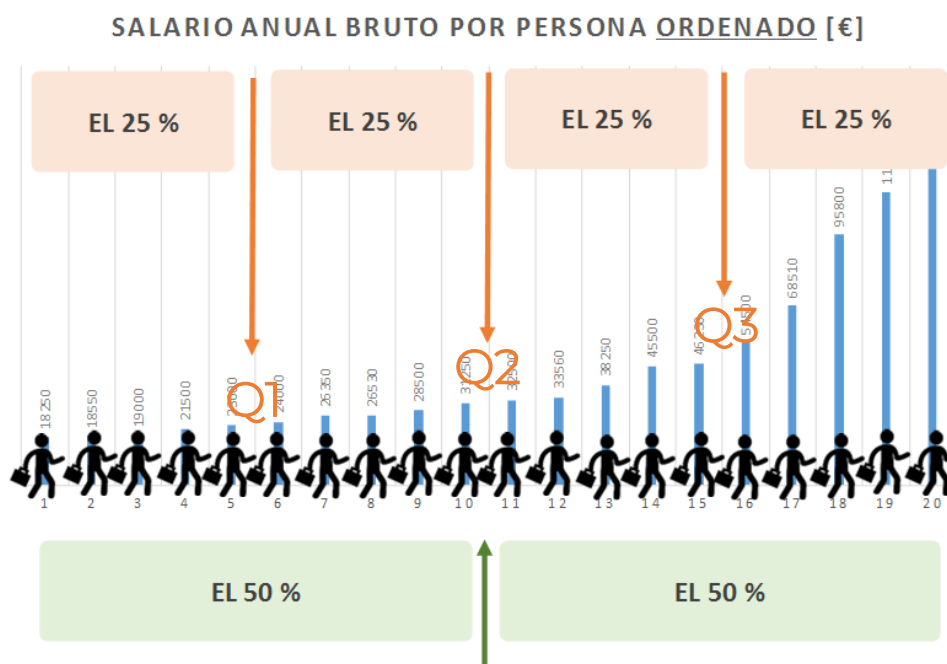
EL RANGO INTERCUARTÍLICO

Es la distancia entre el cuartil 3 y 1.

¿Pero qué son los cuartiles? Son medidas de posición y se calculan de manera similar a la mediana

Vuelvo al ejemplo del sueldo anual bruto, hemos recogido la variable sueldo anual de 20 trabajadores. Los cuartiles se calculan:

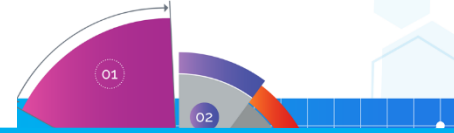
1. Ordenando el sueldo de menor a mayor
2. Cortamos en 4 partes iguales en número. Con el 25 % de los trabajadores (5 trabajadores por parte)
3. Las tres fronteras de los 4 grupos son los cuartiles 1, 2 y 3 (Q1, Q2, Q3)
4. El cuartil 2 coincide con la mediana, ¿se ve? Porque parte en el 50% de la variable. ¡Voilà!



El rango intercuartilico, IQR es igual al cuartil 3 menos el cuartil 1. Se representa así.

$$IQR = Q3 - Q1$$

¡Calcula la dispersión a tus variables cuantitativas utilizando la desviación estándar y la mediana!



DESCRIPCIÓN NUMÉRICA

¿Cómo caracterizar tus datos con el centro y la dispersión?

Ahora te explicaré cómo juntar la centralidad y la dispersión. En el 99.9% de los proyectos de investigación se utiliza esta receta que te desvelaré hoy.

En la práctica para caracterizar una variable cuantitativa se utiliza la centralidad y la dispersión al mismo tiempo. Sigue esta receta:

CENTRALIDAD +/- DISPERSIÓN

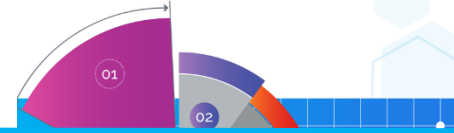
OPCIÓN 1: MEDIA y DESVIACIÓN ESTÁNDAR

Se utiliza para variables cuantitativas cuando no hay valores muy muy alejados del resto. Si pasa esto utiliza la opción 2.

OPCIÓN 2: MEDIANA y RANGO INTERCUARTÍLICO

Cuando hay bastantes valores extremos que puedan influir en el cálculo de la media se utiliza la mediana y el rango intercuartílico.

¡Utiliza las 2 opciones con la receta CENTRALIDAD +/- DISPERSIÓN para caracterizar las variables cuantitativas!



EL TEST DE HIPÓTESIS (Ejemplo: T-test)

¿Cómo responder a las preguntas de investigación?

Una de las técnicas más utilizadas son los contrastes de hipótesis.

El test estadístico te permite contrastar una hipótesis de investigación. Y te preguntará, ¿qué es exactamente una hipótesis de investigación?

Volviendo al ejemplo de antes. Una hipótesis de investigación podría ser:

“La especie de versicolor tiene un ancho del pétalo diferente que la especie de virginica”

O por ejemplo si eres fisioterapeuta puedes preguntarte si el nuevo tratamiento que propones tiene mejor efecto en la recuperación de dolores lumbares que el tratamiento convencional.

Las dos preguntas sólo permiten recibir dos posibles respuestas: o un SÍ o un NO.

Para contestar estas preguntas de investigación o hipótesis de investigación es necesario un método. O mejor dicho, una estrategia paso a paso para lograrlo.

Y es precisamente lo que te voy a desvelar en este apartado

Pero antes déjame contarte qué es un test estadístico y cómo se relaciona con las hipótesis de investigación:

UN TEST ESTADÍSTICO TE PERMITE CONTESTAR A UNA PREGUNTA DE INVESTIGACIÓN O HIPÓTESIS DE INVESTIGACIÓN

UNA HIPÓTESIS ES UNA PREGUNTA QUE PERMITE DOS RESPUESTAS: VERDADERO O FALSO

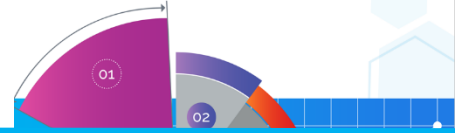
A continuación te muestro la plantilla para contestar a cualquier hipótesis o pregunta de investigación que permite un verdadero o falso como respuesta.

¡Allá va!



Imagínate que eres un biólogo interesado en flores. El objetivo es afirmar si hay diferencias en el ancho del pétalo entre dos especies de flores: virginica y versicolor.

PLANTILLA DE CONTRASTE DE HIPÓTESIS	EJEMPLO
1 – DEFINIR H1: PREGUNTA DE INVESTIGACIÓN O HIPÓTESIS ALTERNATIVA (CASO EXTRAÑO O DE DIFERENCIAS)	El grupo de versicolor tiene un ancho del pétalo diferente que el grupo de virginica
2 – DEFINIR H0: HIPÓTESIS NULA (CONTRARIA A LA HIPÓTESIS DE INVESTIGACIÓN) (CASO NORMAL O DE IGUALDAD)	El grupo de versicolor tiene un ancho del pétalo igual que el grupo de virginica
3 – ESCOGER EL TIPO DE TEST – Para este pregunta estamos comparando 2 grupos	<ul style="list-style-type: none"> • T-TEST (si los grupos son normales y tenemos 2 grupos) • ANOVA (si los grupos son normales y tenemos 3 o más grupos)
4 – NIVEL DE SIGNIFICACIÓN (LA PROBABILIDAD DE ERROR QUE ADMITIMOS A LA PREGUNTA DE INVESTIGACIÓN)	Normalmente el 5 % (con muestras muy pequeñas se utiliza el 1%)
5 – CÁLCULO DEL P-VALOR (se calcula el test estadístico con un software) (EL P-VALOR ES EL NIVEL DE ERROR DE LA HIPÓTESIS DE INVESTIGACIÓN O H1)	0.012 o 1.2 %
6 – RESPUESTA (APLICA LA SIGUIENTE FÓRMULA = H1 + “con un” + P_VALOR + “de probabilidad de error”)	El grupo de versicolor tiene un ancho del pétalo diferente que el grupo de virginica con un 1.2% de probabilidad de error
7 – CONCLUSIÓN <ul style="list-style-type: none"> • El p-valor calculado es menor que 0.05 --> La H1 es cierta --> Afirmamos la hipótesis • El p-valor calculado es mayor que 0.05 --> La H1 NO es cierta --> NO Afirmamos la hipótesis 	En este caso la H1 es cierta y puedes afirmar a la hipótesis de investigación



Grábate esta plantilla para siempre.

Te va a servir para contestar cualquier hipótesis de investigación.

Es decir, cualquier pregunta de tu proyecto que admita la respuesta: CIERTO o FALSO

¡En el Excel te la he dejado preparada para ser utilizada!

¡Utiliza la plantilla del contraste de hipótesis siempre que quieras contestar preguntas de investigación de verdadero o falso!

LA CORRELACIÓN

¿Cómo asociar medidas en tus datos?

Y por último te quiero aportar una técnica que seguramente utilizarás alguna vez.

¡La correlación!

CORRELACIÓN

La correlación es la asociación entre dos mediciones.

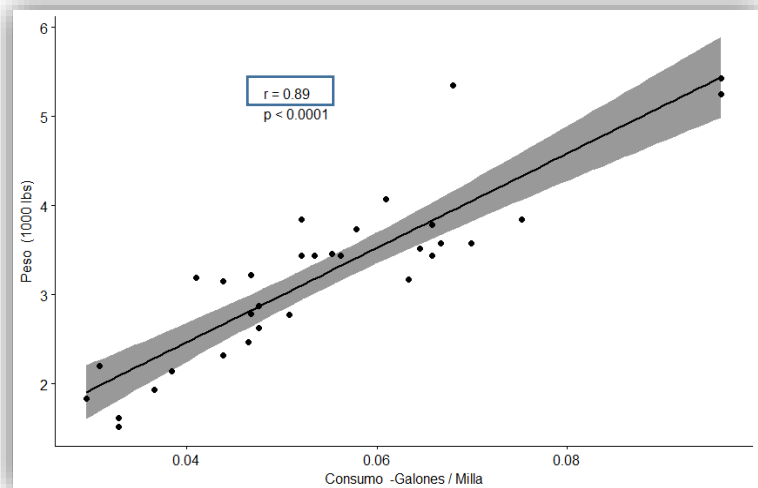
Por ejemplo:

- El peso del coche
- El consumo del coche

¿El peso del coche influye en el consumo del mismo?

¿Cuánto?

Esta pregunta la puedes responder con el coeficiente de correlación.

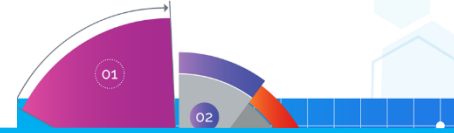


Antes de calcular este índice lo mejor es hacer un **gráfico de dispersión** poniendo los modelos de coche que conoces de diferentes marcas en un plano. Las características que dibujarás son el peso del coche y el consumo.

¿Resultado? Una nube de puntos. Dónde cada punto es un modelo de coche con su correspondiente peso y su consumo. ¡Tachán! (mira el gráfico de dispersión)

¿Cómo sé si consumo y peso están asociados? Solamente necesitas mirar hacia dónde apuntan los puntos más o menos. Imagínate una línea recta que sea el compromiso de los puntos.

En este caso van hacia arriba. La recta imaginaria va hacia arriba. Es decir que a mayor peso mayor consumo y viceversa. ¿Se ve?



Ahora sólo falta entender el coeficiente de correlación $r = 0.89$ que ves en el gráfico. ¡A por ello!

EL COEFICIENTE DE CORRELACIÓN

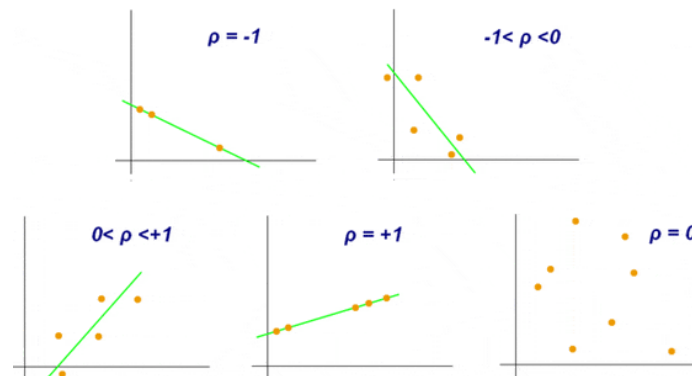
Para llevar a la práctica la correlación utilizarás el **coeficiente de correlación**.

Este te permitirá **cuantificar cómo de relacionadas o asociadas están dos mediciones**.

En el caso del ejemplo: peso y altura

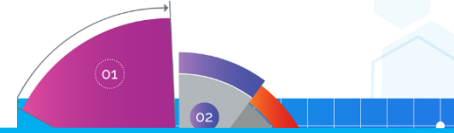
Este coeficiente va de -1 a 1.

- COEFICIENTE DE CORRELACIÓN POSITIVO → RELACIÓN CRECIENTE (DIRECTA)
- COEFICIENTE DE CORRELACIÓN NULO → RELACIÓN NULA
- COEFICIENTE DE CORRELACIÓN NEGATIVO → RELACIÓN DECRECIENTE (INVERSA)



Utiliza la siguiente tabla para saber el grado de correlación entre tus variables:

- -1.0 y -0.8 → la correlación es muy alta y es inversa
- -0.8 y -0.6 → la correlación es alta y es inversa
- -0.6 y -0.4 → la correlación es normal y es inversa
- -0.4 y -0.2 → la correlación es baja y es inversa
- -0.2 y 0.2 → la correlación es muy baja o nula. No hay correlación
- 0.2 y 0.4 → la correlación es baja y es directa
- 0.4 y 0.6 → la correlación es normal y es directa
- 0.6 y 0.8 → la correlación es alta y es directa
- 0.8 y 1.0 → la correlación es muy alta y es directa



¡Calcula el coeficiente de correlación para poder descubrir la asociación entre variables cuantitativas!

EN LA APLICACIÓN ESTÁ LA CLARIDAD

Ahora ya conoces técnicas muy utilizadas en tesis doctorales o proyectos de investigación. En la hoja de cálculo te he puesto ejemplos de cómo aplicarlas y cómo llevarlas a cabo con la ayuda de Excel.

Ahora sólo debes implementar estas técnicas en tu tabla de datos.

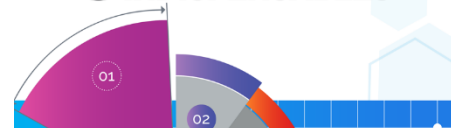
Sigue las instrucciones de la Hoja de Cálculo que te he preparado y ya puedes ir trabajando con tu proyecto de investigación.

¡Seguro que puedes sacarle mucho jugo a estas técnicas!

PD: cuando implementes alguna de ellas tómate una caña, un zumo de zanahorias o lo que más te apetezca.

Has logrado algo importante para tu tesis ;) ¡Hay que celebrarlo!

Has activado tu acción para encontrar la claridad que necesitas 😊



¿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 e 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*