

---

# ESTADÍSTICA DESCRIPTIVA EN INVESTIGACIÓN SOCIAL CON EXCEL

---



Universidad  
Zaragoza

PROF. DR. MIGUEL LAFUENTE BLASCO

DEPARTAMENTO DE MÉTODOS ESTADÍSTICOS

INSTITUTO DE BIOCOMPUTACIÓN Y FÍSICA DE SISTEMAS COMPLEJOS (BIFI)



Departamento de  
Métodos Estadísticos  
**Universidad** Zaragoza



Instituto Universitario de Investigación  
**Biocomputación y Física  
de Sistemas Complejos**  
**Universidad** Zaragoza

# Estadística Descriptiva en Investigación Social con Excel

Miguel Lafuente Blasco

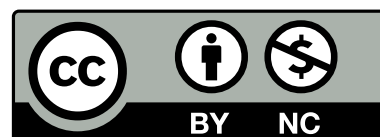
1ª edición. Zaragoza, 2025.

Edita: Servicio de Publicaciones. Universidad de Zaragoza.

ISBN: 978-84-10169-42-5



Servicio de Publicaciones  
**Universidad Zaragoza**



## Licencia Creative Commons

Este libro se distribuye bajo una licencia **Creative Commons Reconocimiento-NoComercial 4.0 Internacional (CC BY-NC 4.0)**, lo que permite compartir y adaptar el material siempre que se reconozca la autoría y no se use con fines comerciales.

Para conocer los detalles legales completos, consulta la licencia en:

<https://creativecommons.org/licenses/by-nc/4.0/>

# Prefacio

Este libro está concebido para guiar al estudiantado a través de las técnicas fundamentales de un curso introductorio de Estadística Descriptiva. La aplicación de los conceptos se ilustra con datos reales del ámbito social, proporcionando una guía paso a paso para su resolución mediante ordenador.

Aunque el texto está dirigido principalmente a estudiantes de ciencias sociales y trabajo social, las operaciones se describen con suficiente detalle para que cualquier persona interesada en el análisis de datos, sin conocimientos previos, pueda replicar los resultados de manera autónoma.

Las técnicas se han implementado utilizando la aplicación de escritorio *Microsoft Excel* de la suite informática Microsoft 365 [6]. Sin embargo, la mayoría de los procedimientos descritos, si no todos, son aplicables a versiones recientes de este software.

Como complemento, especialmente para la interpretación de los resultados, se recomienda encarecidamente consultar el libro [5], donde se ofrece una explicación exhaustiva de los conceptos estadísticos abordados en este texto y su papel en la investigación social en el siglo XXI.

Los datos empleados a lo largo de este libro están disponibles en acceso abierto en el enlace de la referencia [1], donde pueden consultarse y descargarse para replicar los ejercicios y ejemplos presentados.



# Índice general

<b>Prefacio</b>	<b>III</b>
<b>1. Introducción a Excel</b>	<b>1</b>
1.1. Libros y hojas . . . . .	1
1.2. Abrir y Guardar un Archivo . . . . .	2
1.3. Escribir en celdas . . . . .	2
1.4. Operaciones Básicas . . . . .	2
1.5. Funciones Básicas en Excel . . . . .	3
1.6. Cálculos Dinámicos . . . . .	4
1.7. Copiar y Arrastrar Fórmulas . . . . .	4
1.8. Completar Series . . . . .	5
1.9. Estética de una celda en Excel . . . . .	5
1.10. Ordenar datos cuantitativos . . . . .	6
1.10.1. Orden personalizado . . . . .	8
1.11. Filtrar datos por una variable cualitativa . . . . .	10
<b>2. Tabulación de variables</b>	<b>13</b>
2.1. Variable cualitativa nominal tabulada . . . . .	13
2.1.1. Cálculo del Tamaño de la Población (N) . . . . .	13
2.1.2. Calculando las Frecuencias Relativas . . . . .	14

2.1.3. Representando las Frecuencias Relativas como Porcentajes . . . . .	15
2.2. Variable Cualitativa Ordinal y Discretas con pocos valores distintos Tabuladas	16
2.3. Variable Cuantitativa Continua Tabulada . . . . .	18
2.4. Tabulando Variables Cualitativas a partir de microdatos . . . . .	19
2.4.1. Usando tablas dinámicas . . . . .	19
2.5. Tabulando Variables Cuantitativas a partir de microdatos . . . . .	24
<b>3. Representación Gráfica</b>	<b>29</b>
3.1. Representación de una Variable Nominal . . . . .	30
3.1.1. Diagrama de Sectores . . . . .	30
3.1.2. Diagrama de Barras . . . . .	32
3.2. Representación de una Variable Ordinal . . . . .	33
3.3. Representación de una Variable Cuantitativa Discreta . . . . .	33
3.4. Representación de una variable cuantitativa continua . . . . .	35
3.4.1. Histograma para representar una variable cuantitativa agrupada en intervalos de amplitudes iguales . . . . .	36
3.4.2. Polígonos de frecuencias como sustituto de un histograma cuando las amplitudes son desiguales: Marca de Clase, Amplitud y Densidad de Intervalos . . . . .	37
3.5. Representación de series temporales: Polígono de frecuencias . . . . .	40
3.5.1. Invertir el orden en el eje horizontal . . . . .	41
3.5.2. Modificar el lugar de las etiquetas en el caso de que haya valores negativos . . . . .	42
<b>4. Medidas de posición</b>	<b>45</b>
4.1. Medidas de Posición en tablas . . . . .	45
4.1.1. Moda . . . . .	45

4.1.2. Mediana y otros percentiles . . . . .	46
4.1.3. Media . . . . .	46
4.2. Medidas de Posición en microdatos . . . . .	48
4.2.1. Variables Cualitativas . . . . .	48
4.2.2. Variables Cuantitativas . . . . .	48
<b>5. Medidas de dispersión y forma</b>	<b>51</b>
5.1. Recorrido o Rango y rango intercuartílico . . . . .	51
5.2. Varianza, Desviación típica y CV: Datos Tabulados . . . . .	51
5.3. Cálculo de medidas de dispersión y forma sobre microdatos . . . . .	55
5.4. Cálculo de puntuaciones tipificadas . . . . .	56
<b>6. Análisis bivalente: una variable cualitativa y otra cuantitativa</b>	<b>57</b>
6.1. Representación gráfica: Diagramas de caja . . . . .	57
6.1.1. Una variable cuantitativa . . . . .	57
6.1.2. Varias variables cuantitativas la vez . . . . .	58
6.1.3. Una variable cuantitativa en función de las clases de una variable cualitativa nominal . . . . .	59
6.1.4. Una variable en función de las clases de una variable cualitativa ordinal . . . . .	60
6.2. Cálculo de las medidas de síntesis por grupo . . . . .	61
6.2.1. Media, desviación típica y número de observaciones . . . . .	61
6.2.2. Mínimo, Q1, Mediana, Q3, Máximo . . . . .	63
<b>7. Análisis bivalente de variables cualitativas</b>	<b>65</b>
7.1. A partir de microdatos . . . . .	65
7.1.1. Tablas de frecuencias absolutas conjuntas . . . . .	65

7.1.2. Frecuencias porcentuales conjuntas . . . . .	67
7.1.3. Frecuencias condicionadas por filas . . . . .	68
7.1.4. Frecuencias condicionadas por columnas . . . . .	69
7.2. A partir de datos tabulados . . . . .	70
7.2.1. Distribución marginal absoluta y tamaño muestral $N$ . . . . .	70
7.2.2. Frecuencias conjuntas relativas o porcentuales . . . . .	71
7.2.3. Frecuencias condicionadas por filas . . . . .	72
7.2.4. Frecuencias condicionadas por columnas . . . . .	74
7.3. Representación gráfica . . . . .	75
<b>8. Análisis bivalente de dos variables cuantitativas</b>	<b>79</b>
8.1. Diagrama de dispersión . . . . .	79
8.2. Covarianza y Correlación . . . . .	80
8.3. Estimación del modelo lineal: ajuste de mínimos cuadrados . . . . .	82
8.3.1. Estimación de la recta . . . . .	82
8.3.2. Representación de la línea de tendencia . . . . .	83
8.3.3. Bondad de ajuste: $R^2$ . . . . .	85
8.3.4. Predicción . . . . .	86
<b>Anexo: Datos Pisa</b>	<b>87</b>
<b>Bibliografía</b>	<b>89</b>



# 1

## Introducción a Excel

**Excel** es una herramienta de Microsoft Office diseñada para trabajar con datos en formato de tablas. Facilita la organización eficiente de datos numéricos y textuales, permite su análisis mediante gráficos y la realización de cálculos complejos. Además, su interfaz posibilita la interacción simultánea con múltiples hojas de trabajo, vinculando datos entre ellas y haciendo más eficaz el manejo de grandes volúmenes de información.

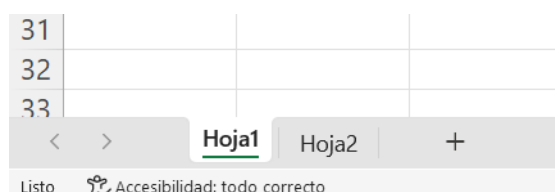
### 1.1. Libros y hojas

#### 1. Libro de Excel:

Un **libro de Excel** se refiere al archivo en sí que creamos con esta herramienta. Es lo que guardamos con una extensión **.xlsx** o **.xls**. Dentro de un libro, podemos tener múltiples hojas de trabajo. Imagina el libro como un archivador y cada hoja de trabajo como una de las fichas dentro de este archivador.

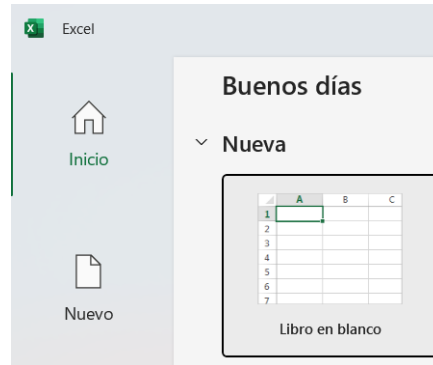
#### 2. Hoja de Excel:

Una **hoja de trabajo** (o simplemente “hoja”) en Excel se refiere a la cuadrícula individual que vemos al abrir Excel. Es donde introducimos y trabajamos con nuestros datos. Cada libro de Excel puede contener múltiples hojas de trabajo y podemos navegar entre ellas mediante las pestañas ubicadas en la parte inferior de la ventana.



## 1.2. Abrir y Guardar un Archivo

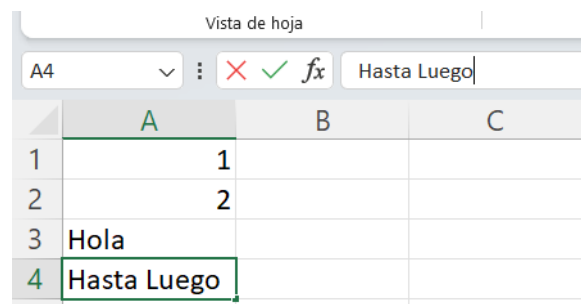
1. Al abrir Excel, se presenta una pantalla de inicio. Seleccionamos *Libro en blanco* para crear un documento nuevo o elegir una plantilla.



2. Para abrir un archivo existente, seleccionamos *Archivo* y luego *Abrir*.
3. Después de trabajar en el archivo, es importante guardarlo. Para hacerlo, seleccionamos *Archivo*, luego *Guardar* o *Guardar como* si queremos elegir un nuevo nombre o ubicación.

## 1.3. Escribir en celdas

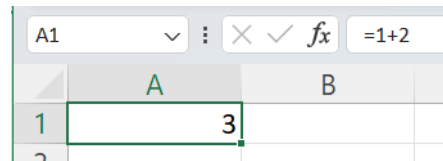
1. Hacemos clic en la celda donde queremos escribir.
2. Ingresamos el contenido deseado tecleando.
3. Presionamos la tecla *Enter* para confirmar el contenido ingresado.



## 1.4. Operaciones Básicas

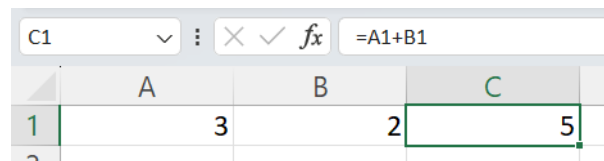
1. En Excel, las operaciones matemáticas básicas se ejecutan directamente en las celdas, pero siempre deben comenzar con un signo “=”. Si no se incluye este signo, Excel

mostrará el texto “1+2” en lugar de realizar la operación. En el siguiente ejemplo, el uso del signo “=” permite que se realice el cálculo:



	A	B
1	3	

2. También podemos utilizar los valores de las celdas en operaciones. Para sumar los números de las celdas A1 y B1, simplemente escribimos `=A1+B1` en la celda C1.



	A	B	C
1	3	2	5

**Nota 1.1.** En lugar de escribir A1 y B1, se pueden seleccionar las celdas haciendo clic sobre ellas. En este ejemplo, haríamos:

- =
  - Clic sobre la celda A1
  - +
  - Clic sobre la celda B1
  - Enter
3. Del mismo modo, podemos usar -, \* y / para realizar restas, multiplicaciones y divisiones.

## 1.5. Funciones Básicas en Excel

Excel ofrece una amplia variedad de funciones que permiten realizar cálculos y análisis de datos. A continuación, se presentan algunas funciones esenciales para usuarios principiantes. Para utilizarlas, seleccionamos una celda en blanco y escribimos la fórmula correspondiente:

### 1. MAX y MIN:

- `=MAX(A1:A10)`: Retorna el valor máximo en el rango A1 a A10.
- `=MIN(A1:A10)`: Retorna el valor mínimo en el rango especificado.

### 2. CONTAR y CONTARA:

- `=CONTAR(A1:A10)`: Cuenta las celdas que contienen valores numéricos en el rango.
- `=CONTARA(A1:A10)`: Cuenta todas las celdas no vacías, sean de texto o números.

### 3. PROMEDIO:

- **=PROMEDIO(B1:B10):** Calcula la media aritmética de los valores numéricos en el rango.

### 4. SUMA:

- **=SUMA(C1:C10):** Suma todos los valores numéricos dentro del rango.

### 5. SI:

- **=SI(B7 >10; "Mayor"; "Menor o igual"):** Evalúa si B7 es mayor que 10 y devuelve “Mayor” si es verdadero, o “Menor o igual” si es falso.

	A	B	C	D
1	3		maximo	6
2	5		minimo	3
3	4		contar	4
4	6		contara	4
5			promedio	4,5
6			suma	18
7		12	si mayor 10	Mayor

## 1.6. Cálculos Dinámicos

En Excel, los cálculos son dinámicos, lo que significa que se actualizan automáticamente. Si modificamos el valor de una celda implicada en una fórmula, el resultado reflejará dicho cambio de forma inmediata.

1. Supongamos que la celda A1 tiene el valor 5 y la celda B1 tiene el valor 10. Si en C1 escribimos la fórmula **=A1+B1**, mostrará 15.
2. Si tras la operación anterior cambiamos el valor de A1 a 7, C1 se actualizará automáticamente para mostrar 17. Compruébalo.

## 1.7. Copiar y Arrastrar Fórmulas

1. Para copiar o ampliar una fórmula en Excel, primero seleccionamos la celda que la contiene.
2. Situamos el cursor sobre el pequeño cuadrado en la esquina inferior derecha de la celda (denominado “controlador de relleno” o “manejador”) y lo arrastramos en la dirección deseada.

3. La fórmula se replicará automáticamente, ajustando las referencias de celda según sea necesario.

Este proceso es una de las principales ventajas de Excel, ya que facilita la ejecución de cálculos complejos en grandes conjuntos de datos y se utilizará de forma habitual a lo largo de este texto.

## 1.8. Completar Series

Excel puede identificar patrones en secuencias numéricas o textuales y completar automáticamente la serie en función del patrón detectado.

1. **Series Numéricas:** Supongamos que se pretende escribir una serie de números que aumente de 2 en 2. Para ello:
  - Escribe 2 en la celda A1 y 4 en la celda A2.
  - Selecciona **ambas** celdas y arrastra hacia abajo, tal como se describe en la Sección 1.7.
  - Excel continuará la serie automáticamente: 6, 8, 10, etc.
2. **Series con Letras:** Ahora pretendemos escribir los días de la semana (también funciona con fechas, meses...)
  - Si escribes **Lunes** en la celda B1 y **Martes** en la celda B2.
  - Al seleccionar ambas celdas y arrastrar con el “manejador de relleno”, Excel detectará el patrón y completará la serie con Miércoles, Jueves, Viernes, etc.

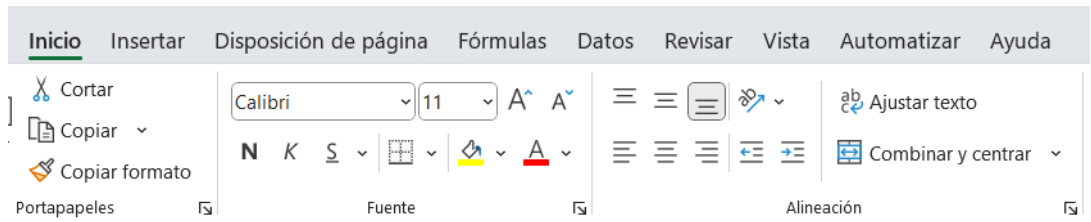
## 1.9. Estética de una celda en Excel

La estética y el diseño en Excel no son simplemente cuestiones cosméticas; pueden mejorar la legibilidad y comprensión de nuestros datos, especialmente cuando presentamos la información a otras personas.

### 1. Formato de Celda:

Para cambiar el formato de una celda:

- a) **Fuente y tamaño de letra:** Selecciona la celda o rango de celdas que desees modificar. Ve a la pestaña “Inicio” y en el grupo “Fuente”, podrás cambiar la tipografía, tamaño, color de fuente, etc.

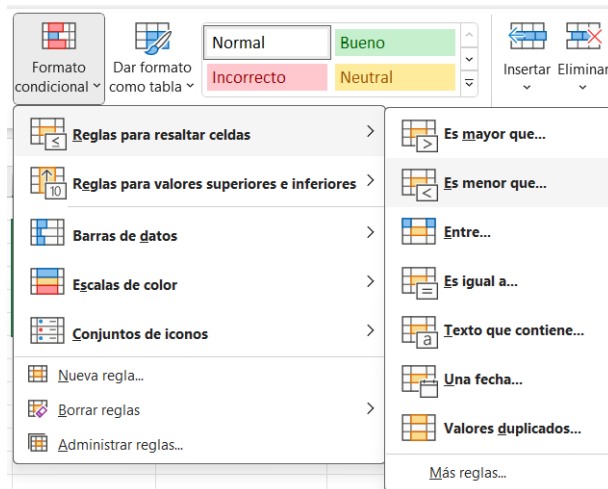


- b) **Alineación:** En el grupo “Alineación”, podrás cambiar la orientación del texto, así como la alineación horizontal y vertical.
- c) **Bordes y relleno:** En la pestaña “Inicio”, dentro del grupo “Fuente”, puedes agregar o modificar bordes y cambiar el color de relleno de las celdas.

## 2. Formato condicional:

La función “Formato condicional” permite cambiar la apariencia de una celda según su contenido. Por ejemplo, puedes hacer que las celdas con valores por debajo de un cierto umbral se resalten en rojo. Esto es muy útil en estadística para detectar valores atípicos o resaltar información importante de un vistazo.

- a) Selecciona las celdas donde quieras aplicar el formato.
- b) Ve a la pestaña “Inicio” y elige “Formato condicional” en el grupo “Estilos”.



- c) Selecciona una de las opciones disponibles o crea una nueva regla personalizada.

## 1.10. Ordenar datos cuantitativos

A menudo, nos encontramos con grandes cantidades de datos que necesitan ser organizados u ordenados. Una herramienta útil para ello es la función de ordenación, que permite organizar datos cuantitativos de menor a mayor o viceversa.

**Nota 1.2.** La ordenación es una de las múltiples operaciones en Excel que modifican los datos, o su presentación.

**Siempre es recomendable hacer una copia de seguridad antes de realizar operaciones que puedan alterar los datos originales**, como la ordenación, en previsión de posibles errores.

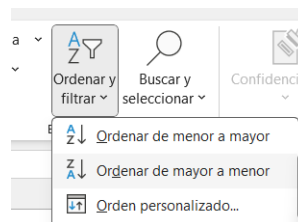
En este apartado, y en gran parte de los ejemplos posteriores, se va a trabajar con datos provenientes del Informe Pisa [8] en España. Esta submuestra está descrita en el [Anexo: Datos Pisa](#).

### Pasos para ordenar datos

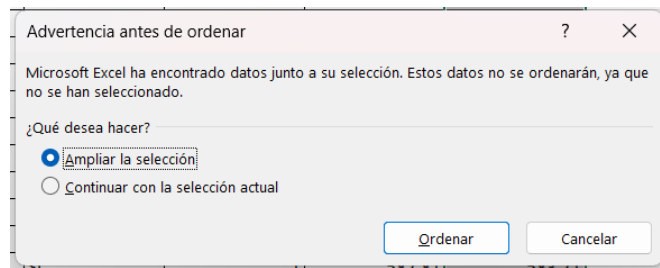
1. **Seleccionar la columna:** Haz clic en la letra correspondiente a la columna que contiene los datos cuantitativos que deseas ordenar. Esto resaltará toda la columna.

G	H
emáticas	Ciencia
346,49	310,68
372,5	448,6
405,25	431,24
441,14	444,65
451,53	439,13

2. **Acceder al menú de ordenación:** Busca en la cinta de opciones de Excel la pestaña “Inicio” y dentro de ella “Ordenar y filtrar”, donde verás diferentes opciones para ordenar.



3. **Elegir la dirección de ordenación:**
  - **De menor a mayor:** Esto ordenará tus datos cuantitativos en orden ascendente.
  - **De mayor a menor:** Esto hará lo contrario, ordenando tus datos en orden descendente.
4. **Confirmar y verificar:** Una vez hayas seleccionado la opción de ordenación deseada, Excel preguntará si queremos reorganizar solo los datos de la columna o también los del resto de columnas en función del orden de la columna seleccionada. Habitualmente, queremos ordenar individuos a partir de los valores de una variable, por lo que seleccionaremos *Ampliar la selección*.



**¡Precaución al ordenar datos!** Si no seleccionas *Ampliar la selección*, solo se reordenará la columna elegida, dejando el resto sin modificar, lo que puede provocar que los datos de diferentes individuos queden desorganizados y pierdan su coherencia.

Si se ha realizado la operación correctamente, ahora se dispondrán de los datos ordenados. Como recomendación, suele ser buena idea revisar los primeros y últimos datos de la columna para asegurarse de que no se han cometido errores.

### 1.10.1. Orden personalizado

En la pestaña de *Ordenar y filtrar*, observamos que también aparece la opción de *Orden personalizado*.

Esta opción es especialmente útil para ordenar variables cualitativas ordinales. Por ejemplo, permite ordenar los distintos niveles de estudios. Esto será clave para algunas representaciones como los diagramas de cajas.

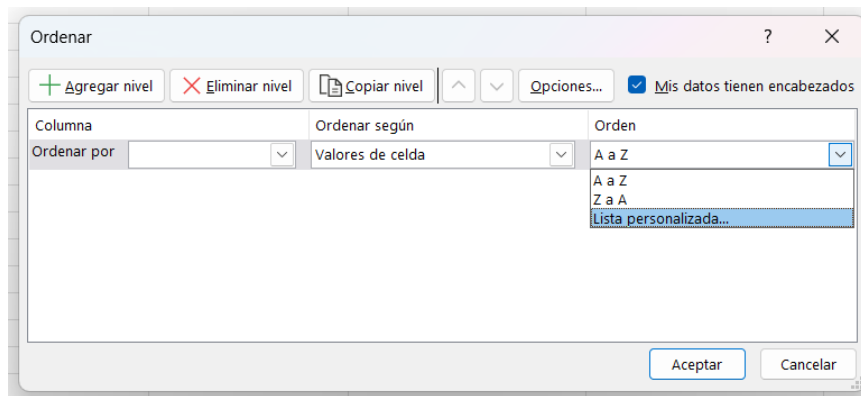
Para introducir un orden personalizado, vamos a trabajar con los datos de una submuestra de 30 estudiantes españoles obtenida del Informe PISA 2018 [8]. Para ello, seguimos las instrucciones del apartado anterior.

1. **Selecciona la columna del Estudios Padre en los datos PISA.**
2. **Accede al menú de ordenación y selecciona *Orden Personalizado*.**
3. **Elige *Ampliar la selección*.**

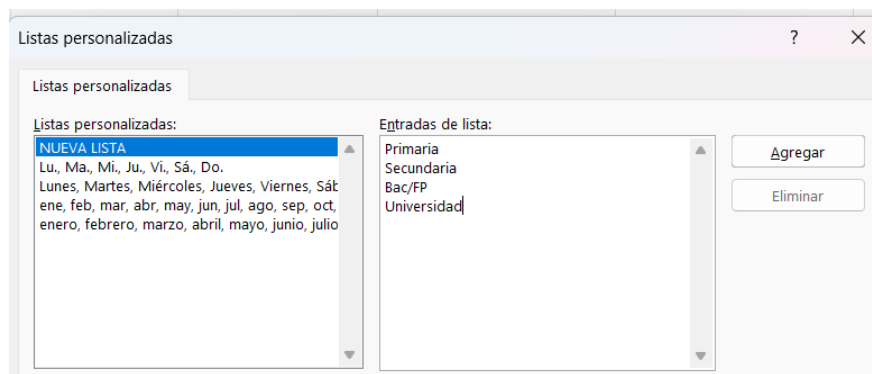
A partir de aquí, realizamos los siguientes pasos para introducir y ejecutar nuestro orden personalizado.

4. **En la pestaña de orden, selecciona *Lista personalizada*.**

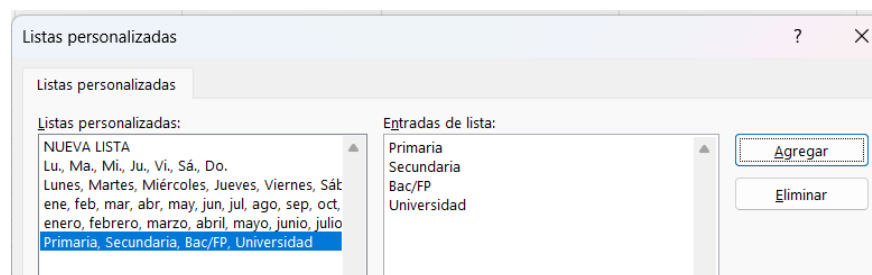




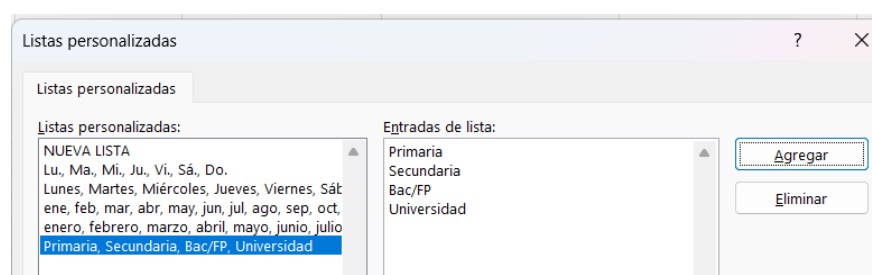
5. **Introducir el orden.** Hacemos clic en *NUEVA LISTA* en el recuadro de la izquierda. En el de la derecha, escribimos los niveles, ordenados de menor a mayor, de la variable. Es importante escribir los niveles de la variable exactamente como aparecen en el conjunto de datos.



6. Pulsar *Agregar*.



7. Selecciona *Ordenar por* la variable de interés. En este caso, seleccionamos ordenar por *Estudios Padre* y pulsamos *Aceptar*.



Deberíamos observar ahora los datos de los alumnos ordenados por el nivel de estudios de su padre.

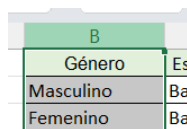
### 1.11. Filtrar datos por una variable cualitativa

En el análisis estadístico, es común trabajar con datos cualitativos, como categorías, estados o grupos. Excel ofrece herramientas de filtrado que facilitan la selección y análisis de datos específicos dentro de estas categorías. A modo de ejemplo, vamos a filtrar los datos para visualizar únicamente a los individuos de género *Femenino* en la submuestra del Informe Pisa.

**Nota 1.3.** Por precaución, es recomendable hacer una copia de seguridad de los datos antes de seguir estas Como se ha indicado anteriormente, al alterar la estructura de los datos, es recomendable guardar los datos antes como copia de seguridad para asegurarse de que no se pierde información.

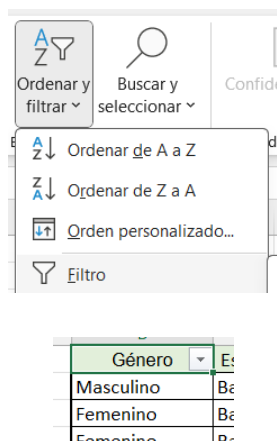
#### Pasos para filtrar datos cualitativos

1. **Activar los filtros:** Selecciona la celda del encabezado de la columna que contiene la variable cualitativa a filtrar.



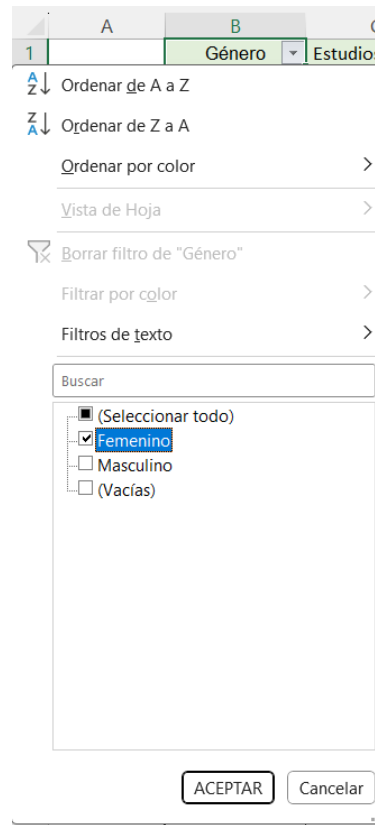
B	Es
Género	Ba
Masculino	Ba
Femenino	Ba

2. Accede a la pestaña “Inicio” arriba en las opciones de Excel. En el grupo “Ordenar y filtrar”, haz clic en “Filtrar”. Un pequeño triángulo aparecerá en el encabezado de la columna seleccionada.



3. **Acceder al menú de filtrado:** Haz clic en el triángulo del encabezado. Se desplegará un menú con varias opciones, incluyendo una lista de los valores únicos en esa columna.

4. **Seleccionar la categoría deseada:** Desmarca “(Seleccionar todo)” y elige solo las categorías que te interesan. Por ejemplo, en la columna “Género”, podrías marcar “Femenino” para ver solo esos registros.



5. **Aplicar el filtro:** Una vez seleccionadas las categorías, haz clic en “Aceptar”. Excel mostrará solo las filas que coincidan con lo que has elegido.

#### Consideraciones adicionales

- Para desactivar el filtro y visualizar todos los datos nuevamente, haz clic en el triángulo del encabezado de la columna y selecciona “Borrar filtro” o marca “(Seleccionar todo)”.
- Si se trabaja con múltiples columnas de variables cualitativas, se puede aplicar filtros a varias columnas simultáneamente para un análisis más detallado.



# 2

## Tabulación de variables

### 2.1. Variable cualitativa nominal tabulada

Las variables cualitativas nominales son aquellas que categorizan datos sin un orden específico. Un ejemplo es el tipo de ruptura matrimonial, que puede ser divorcio, separación o nulidad; no existe una relación jerárquica o de orden entre estos valores. En esta sección vamos a trabajar con los datos de las Estadísticas sobre rupturas matrimoniales 2019 del INE [3].

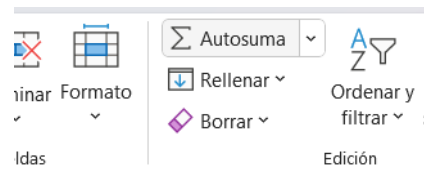
#### 2.1.1. Cálculo del Tamaño de la Población (N)

Para calcular el total de rupturas matrimoniales del año 2019, se deben sumar las frecuencias absolutas de divorcios, separaciones y nulidades.

1. En Excel, selecciona la celda donde deseas mostrar la suma, habitualmente debajo de los valores de las frecuencias absolutas o del número de individuos en cada categoría.

	A	B
1	Tipo de ruptura (xi)	Frecuencia Absoluta (ni)
2	Divorcio	91645
3	Separación	3600
4	Nulidad	75
5		

2. Usa el botón **Autosuma** ( $\Sigma$ ) para sumar automáticamente las celdas anteriores. Haz clic en él y luego pulsa *Enter*.



### 2.1.2. Calculando las Frecuencias Relativas

Las frecuencias relativas muestran la proporción de cada categoría en relación con el total.

La fórmula general es:

$$\text{Frecuencia Relativa} = \frac{\text{Frecuencia Absoluta de una categoría}}{N}$$

En Excel:

1. Selecciona la celda donde quieras calcular la frecuencia relativa para los divorcios, ya que es la primera categoría.

	A	B	C
1	Tipo de ruptura (xi)	Frecuencia Absoluta (ni)	Porcentajes
2	Divorcio	91645	
3	Separación	3600	
4	Nulidad	75	
5	Total	95320	

2. Introduce la fórmula para calcular: `=Celda_Divorcios/Celda_N`. Para ello, después del `=`, selecciona la celda de la frecuencia absoluta de los divorcios, luego usa `/` para dividir, y selecciona la celda con el total ( $N$ ).
3. Usa referencias absolutas para asegurarte de que la celda del denominador ( $N$ ) no cambie al copiar la fórmula. Esto se logra pulsando **F4** al seleccionar la celda de  $N$  en la fórmula.

	A	B	C
1	Tipo de ruptura (xi)	Frecuencia Absoluta (ni)	Porcentajes
2	Divorcio	91645	<code>=B2/\$B\$5</code>
3	Separación	3600	
4	Nulidad	75	
5	Total	95320	

**Nota 2.1.** Observa cómo antes y después de la letra de la columna en la fórmula aparece el símbolo `$`. Esto indica que la referencia es absoluta, es decir, está inmovilizada, y cuando arrastres después para generalizar la fórmula, ese valor no cambiará.

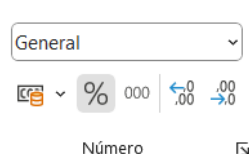
4. Generaliza la fórmula para separaciones y nulidades arrastrando con el manejador de relleno (esquina inferior derecha de la celda).

	A	B	C
1	Tipo de ruptura (xi)	Frecuencia Absoluta (ni)	Porcentajes
2	Divorcio	91645	0,96144566
3	Separación	3600	
4	Nulidad	75	
5	Total	95320	

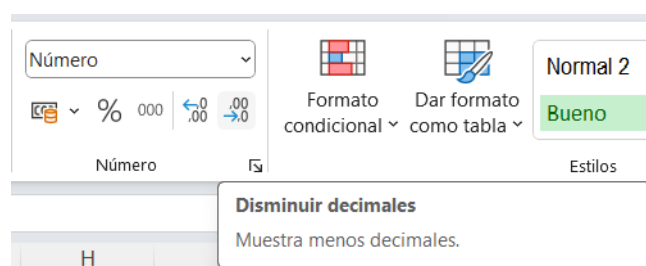
	A	B	C
1	Tipo de ruptura (xi)	Frecuencia Absoluta (ni)	Porcentajes
2	Divorcio	91645	0,96144566
3	Separación	3600	0,03776752
4	Nulidad	75	0,00078682
5	Total	95320	

### 2.1.3. Representando las Frecuencias Relativas como Porcentajes

1. Seleccionamos las celdas con las frecuencias relativas.
2. Usamos el botón **Porcentaje**, el cual se encuentra en la pestaña *Inicio*, dentro del grupo *Número* en la barra de herramientas, para convertir las cifras a formato porcentual.



3. Ajustamos a el número de decimales (normalmente no más de dos) utilizando los botones de **aumentar/disminuir decimales**.



	A	B	C
1	Tipo de ruptura (xi)	Frecuencia Absoluta (ni)	Porcentajes
2	Divorcio	91645	96,14%
3	Separación	3600	3,78%
4	Nulidad	75	0,08%
5	Total	95320	

**Nota 2.2.** Las variables cualitativas nominales, como el tipo de ruptura matrimonial, no tienen un orden inherente. Por tanto, no debemos calcular frecuencias acumuladas. Sin embargo, como hemos visto, podemos determinar las frecuencias absolutas y relativas para entender la distribución de datos en cada categoría.

## 2.2. Variable Cualitativa Ordinal y Discretas con pocos valores distintos Tabuladas

La variable de calificaciones en una asignatura es un ejemplo de una variable cualitativa ordinal. Esto significa que hay una relación de orden entre sus valores: suspenso, aprobado, notable, sobresaliente y matrícula de honor. Vamos a ilustrar esta sección con las calificaciones de la asignatura de Estadística Aplicada a la Investigación Social del curso 2022/2023.

**Nota 2.3.** Para variables ordinales o cuantitativas discretas, las tablas de distribución de frecuencias son similares, y se pueden obtener las mismas frecuencias de la misma manera.

En primer lugar, calculamos, como en la Sección 2.1, el total de observaciones y las frecuencias relativas porcentuales:

1. **Tamaño de la Población  $N$ :** Sumar todas las frecuencias absolutas.
2. **Frecuencias Relativas:** Dividir cada frecuencia absoluta por  $N$ , inmovilizando la celda del denominador.

Como es una variable que tiene un orden, podemos y debemos calcular las frecuencias acumuladas.

### 3. Frecuencias Acumuladas Absolutas:

- a) Para el primer valor (suspenso), la frecuencia absoluta acumulada es igual a su frecuencia absoluta.

	A	B	C	D
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)
2	Suspenso	26	22,81%	=B2
3	Aprobado	51	44,74%	
4	Notable	25	21,93%	
5	Sobresaliente	5	4,39%	
6	MH	7	6,14%	
7		114		

- b) Para los valores siguientes, sumamos la frecuencia absoluta acumulada anterior y la frecuencia absoluta de la clase.

	A	B	C	D
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)
2	Suspenso	26	22,81%	26
3	Aprobado	51	44,74%	=B3+D2
4	Notable	25	21,93%	
5	Sobresaliente	5	4,39%	
6	MH	7	6,14%	
7		114		

- c) Generalizamos la operación en **solo la segunda celda** para el resto de las clases, arrastrando usando el controlador de relleno.



	A	B	C	D
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)
2	Suspenso	26	22,81%	26
3	Aprobado	51	44,74%	77
4	Notable	25	21,93%	
5	Sobresaliente	5	4,39%	
6	MH	7	6,14%	
7		114		

	A	B	C	D
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)
2	Suspenso	26	22,81%	26
3	Aprobado	51	44,74%	77
4	Notable	25	21,93%	102
5	Sobresaliente	5	4,39%	107
6	MH	7	6,14%	114
7		114		

4. **Frecuencias Acumuladas Relativas:** Hay dos maneras equivalentes de calcular estas frecuencias.

a) **Utilizando las frecuencias absolutas:** Dividir cada frecuencia acumulada absoluta por  $N$ . Para ello, repetir el cálculo de las frecuencias relativas a partir de las frecuencias absolutas realizado en la Subsección 2.1.

1) Para la primera clase, se calcula la frecuencia absoluta acumulada dividida por el total de observaciones, inmovilizando este denominador.

	A	B	C	D	E
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)	Frec. Porcentual Acum. (Pi)
2	Suspenso	26	22,81%	26	=D2/\$B\$7
3	Aprobado	51	44,74%	77	
4	Notable	25	21,93%	102	
5	Sobresaliente	5	4,39%	107	
6	MH	7	6,14%	114	
7		114			

2) Arrastrar el resultado para generalizar la fórmula al resto de las clases.

	A	B	C	D	E
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)	Frec. Porcentual Acum. (Pi)
2	Suspenso	26	22,81%	26	0,228070175
3	Aprobado	51	44,74%	77	0,675438596
4	Notable	25	21,93%	102	0,894736842
5	Sobresaliente	5	4,39%	107	0,938596491
6	MH	7	6,14%	114	1
7		114			

3) **Expresar porcentajes:** Expresar las frecuencias relativas acumuladas en porcentajes pulsando el botón de % y ajustando el número de decimales como se ilustra en la Subsección 2.1.3.

b) **Utilizando las frecuencias relativas porcentuales:** Si no se dispone de las frecuencias absolutas, se pueden calcular las frecuencias porcentuales acumuladas a partir de las frecuencias porcentuales. Para ello, procedemos igual que en el paso anterior, pero utilizando las frecuencias porcentuales en lugar de las absolutas:

- 1) Para el primer valor (suspense), la frecuencia porcentual acumulada es igual a su frecuencia porcentual.

	A	B	C	D	E
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)	Frec. Porcentual Acum. (Pi)
2	Suspense	26	22,81%	26	=C2
3	Aprobado	51	44,74%	77	
4	Notable	25	21,93%	102	
5	Sobresaliente	5	4,39%	107	
6	MH	7	6,14%	114	
7		114			

- 2) Para los valores siguientes, sumamos la frecuencia porcentual acumulada anterior y la frecuencia porcentual de esa clase.

	A	B	C	D	E
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)	Frec. Porcentual Acum. (Pi)
2	Suspense	26	22,81%	26	22,81%
3	Aprobado	51	44,74%	77	=C3+E2
4	Notable	25	21,93%	102	
5	Sobresaliente	5	4,39%	107	
6	MH	7	6,14%	114	
7		114			

- 3) Generalizamos la operación en **solo la segunda celda** para el resto de las clases, arrastrando usando el controlador de relleno.

	A	B	C	D	E
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)	Frec. Porcentual Acum. (Pi)
2	Suspense	26	22,81%	26	22,81%
3	Aprobado	51	44,74%	77	67,54%
4	Notable	25	21,93%	102	
5	Sobresaliente	5	4,39%	107	
6	MH	7	6,14%	114	
7		114			

	A	B	C	D	E
1	Calificaciones	Número de estudiantes (ni)	Porcentaje (pi)	Frec. Absoluta Acum. (Ni)	Frec. Porcentual Acum. (Pi)
2	Suspense	26	22,81%	26	22,81%
3	Aprobado	51	44,74%	77	67,54%
4	Notable	25	21,93%	102	89,47%
5	Sobresaliente	5	4,39%	107	93,86%
6	MH	7	6,14%	114	100,00%
7		114			

## 2.3. Variable Cuantitativa Continua Tabulada

Una variable cuantitativa continua tabulada debe ser dada en intervalos. En este caso, el resumen tabular se calcula exactamente igual que en el caso de las variables cualitativas ordinales y cuantitativas discretas que se ilustran en la Sección 2.2. Por lo tanto, hay que calcular:

### 1. Frecuencia Absoluta.

## 2. Frecuencia Relativa.

## 3. Frecuencia Acumulada.

## 4. Frecuencia Relativa Acumulada.

**Nota 2.4.** Es posible que los intervalos tengan igual o distinta amplitud, lo cual, para efectos del cálculo de frecuencias, no tiene ninguna relevancia. Sí que la tendrá para la representación gráfica y el cálculo de la moda.

Para ilustrar el resultado, dado que el procedimiento es el mismo, se muestra el resumen tabular de los gastos mensuales en alimentación de una muestra de familias (datos ficticios)).

	A	B	C	D	E	F
1	Gasto mensual (euros)					
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)
3	150	250	10	5,0%	10	5,0%
4	250	350	25	12,5%	35	17,5%
5	350	450	50	25,0%	85	42,5%
6	450	750	80	40,0%	165	82,5%
7	750	1100	30	15,0%	195	97,5%
8	1100	1500	5	2,5%	200	100,0%
9			200			

## 2.4. Tabulando Variables Cualitativas a partir de microdatos

Para el desarrollo de esta sección, utilizaremos los datos de la muestra de 30 casos del Informe Pisa. El objetivo principal va a ser calcular las distintas frecuencias para la variable cualitativa nominal *Estudios Padre*.

### 2.4.1. Usando tablas dinámicas

Las tablas dinámicas son una herramienta poderosa y simple de Excel, además de ser menos propensa a errores del usuario. A continuación, mostramos cómo podemos hallar las frecuencias de una variable cualitativa nominal, ordinal, o cuantitativa discreta usando esta herramienta.

### Calculando las frecuencias absolutas

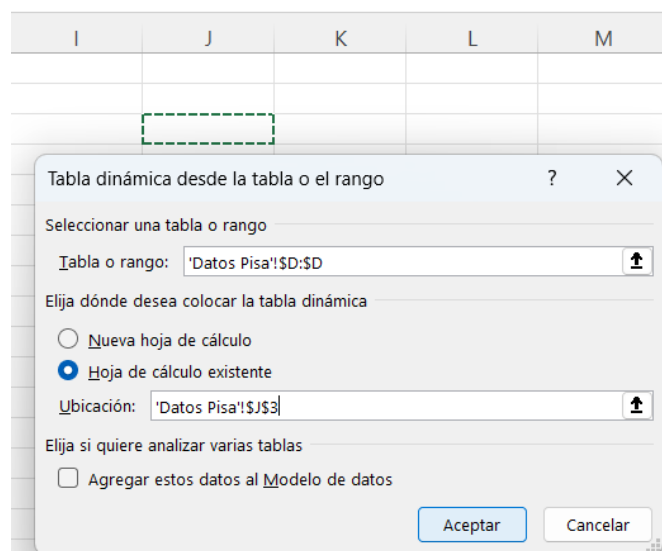
1. Seleccionamos los datos con su encabezado.

	D
Pre	Estudios Padre
	Secundaria
	Bac/FP

2. Pulsamos en 'Insertar' - 'Tabla dinámica'.



3. Seleccionamos dónde queremos poner la nueva tabla. Puede ser en otra celda (pulsar *Hoja de cálculo existente* y a continuación hacer clic en una celda de la hoja donde queramos poner la tabla), o en otra hoja (pulsar *Nueva hoja de cálculo*).



4. Arrastramos la variable de interés tanto al cuadro de *Filas* como al de *Valores*.

Arrastrar campos entre las áreas siguientes:

<b>Filtros</b>	<b>Columnas</b>
<b>Filas</b>	<b>Valores</b>
Estudios Padre ▾	Cuenta de Estudios Padre ▾

☐ Aplazar actualización del diseño Actualizar

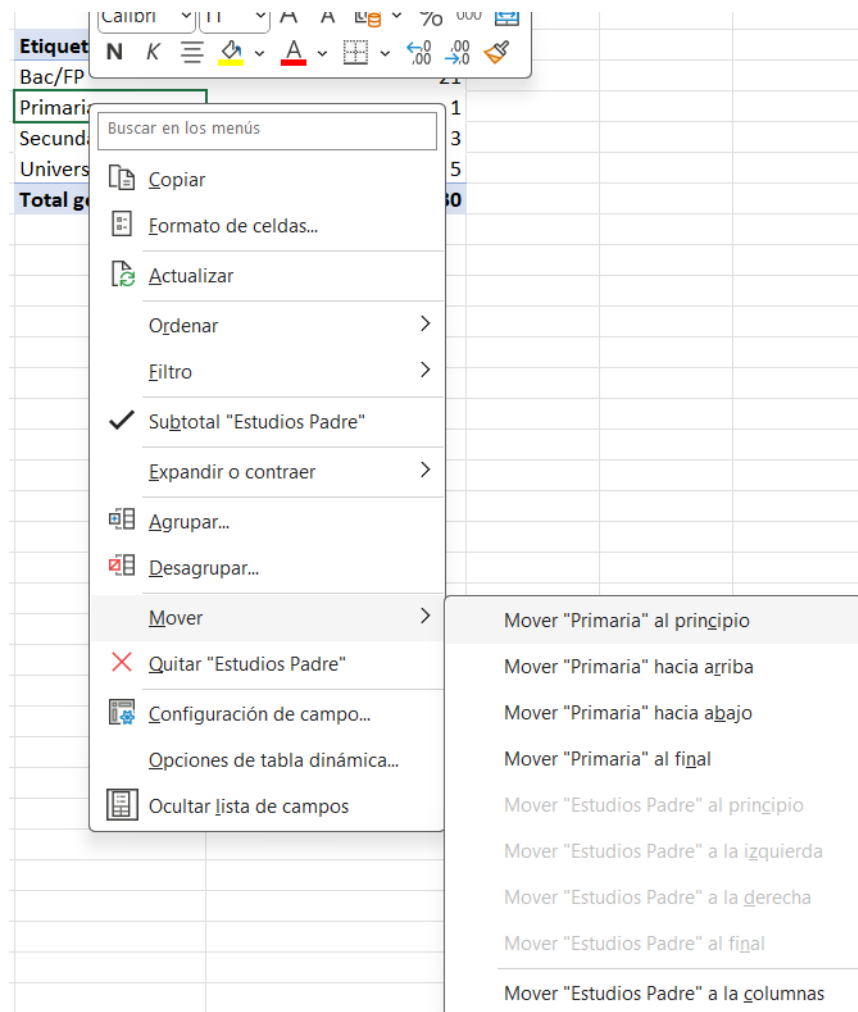
Obtenemos inmediatamente el conteo de las clases mediante las **Frecuencias absolutas**.

Si queremos quitar las observaciones en blanco, podemos hacer clic en el símbolo del triángulo en las *Etiquetas de fila* y desmarcar la opción *(en blanco)*.

448,6	Etiquetas de fila ▾	Cuenta de Estudios Padre
A ↓	Ordenar de A a Z	21
Z ↓	Ordenar de Z a A	1
	Más opciones de ordenación...	3
	Borrar filtro de "Estudios Padre"	5
	Filtros de etiqueta >	<b>30</b>
	Filtros de valor >	
<div> <div>Buscar</div> <div> <input checked="" type="checkbox"/> (Seleccionar todo)           <input checked="" type="checkbox"/> Bac/FP           <input checked="" type="checkbox"/> Primaria           <input checked="" type="checkbox"/> Secundaria           <input checked="" type="checkbox"/> Universidad           <input type="checkbox"/> (en blanco)         </div> </div>		
<div> <div>ACEPTAR</div> <div>Cancelar</div> </div>		

Etiquetas de fila	Cuenta de Estudios Padre
Bac/FP	21
Primaria	1
Secundaria	3
Universidad	5
<b>Total general</b>	<b>30</b>

5. Si la variable tiene un orden, hemos de reordenar las clases. Para ello, hacemos clic derecho en la clase de la variable que queremos mover, pulsamos *Mover* y escogemos la opción adecuada para acabar ordenando los posibles valores...



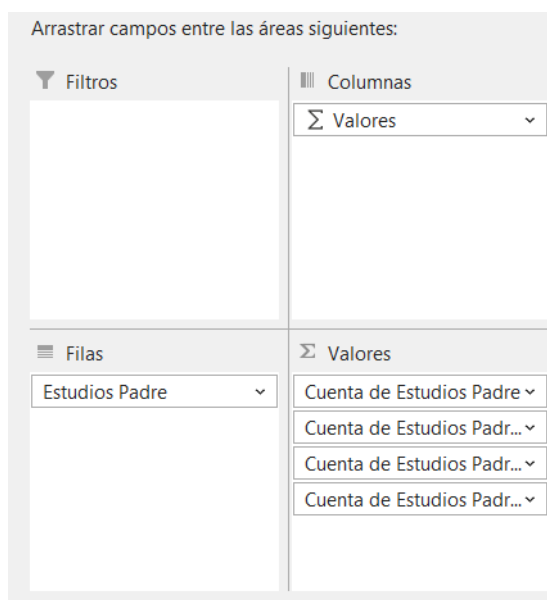
...hasta que tengamos el orden correcto:

Etiquetas de fila	Cuenta de Estudios Padre
Primaria	1
Secundaria	3
Bac/FP	21
Universidad	5
<b>Total general</b>	<b>30</b>

## Calculando las frecuencias relativas

Si bien, a partir de las frecuencia absolutas, se pueden calcular el resto de frecuencias como se ha hecho en la Sección 2.2, se recomienda aprovechar las herramientas que ofrecen las tablas dinámicas, ya que facilitan el trabajo y son menos propensas a errores.

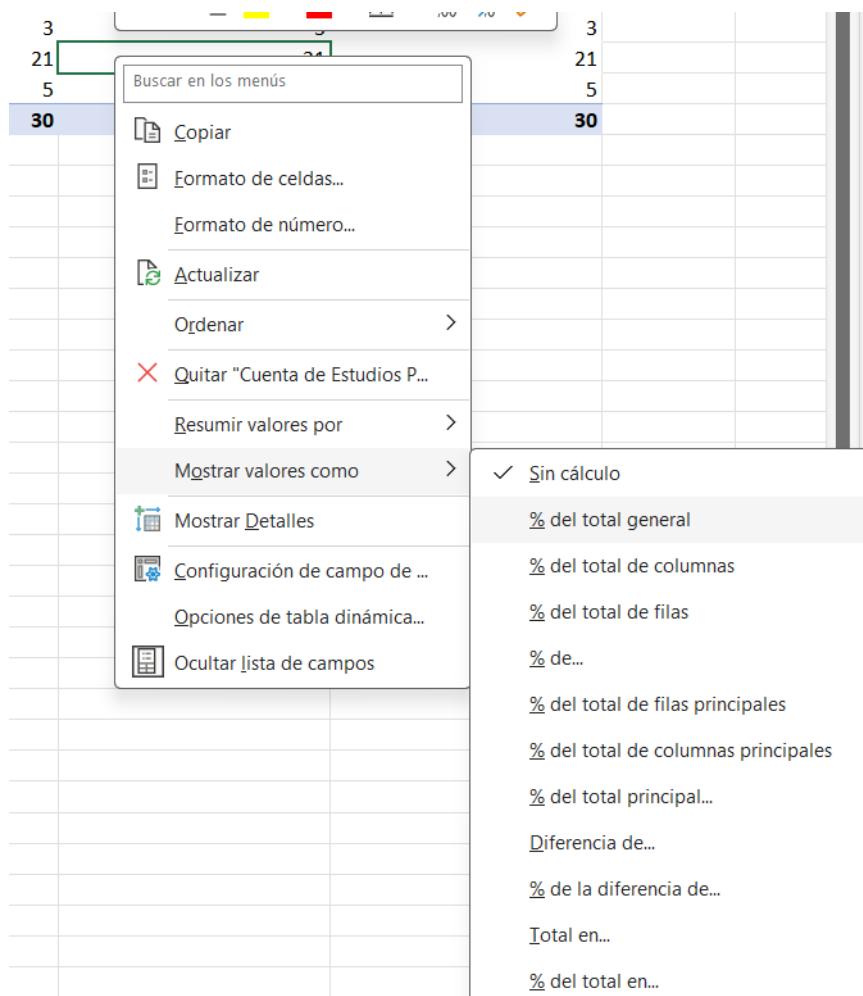
1. Seleccionando la tabla dinámica, arrastramos cuatro veces la variable al cuadro de *Valores* (cuatro veces para mostrar las cuatro columnas deseadas; si se trata de una variable nominal, bastará con poner dos columnas en lugar de cuatro).



2. Nos aparecerá la siguiente tabla:

Etiquetas de fila	Cuenta de Estudios Padre	Cuenta de Estudios Padre2	Cuenta de Estudios Padre3	Cuenta de Estudios Padre4
Primaria	1	1	1	1
Secundaria	3	3	3	3
Bac/FP	21	21	21	21
Universidad	5	5	5	5
<b>Total general</b>	<b>30</b>	<b>30</b>	<b>30</b>	<b>30</b>

3. Posteriormente, como la primera ya por defecto muestra las frecuencias absolutas, hacemos clic derecho en un valor de la segunda columna y seleccionamos *Mostrar valores como* y luego *% del total general* para mostrar las frecuencias porcentuales.



4. Repetimos la misma operación con la tercera y la cuarta aparición de *Cuenta de Estudios...*, seleccionando *Total en* para mostrar las Frecuencias Absolutas Acumuladas, y *% del total en* para mostrar las Frecuencias Porcentuales Acumuladas.

Etiquetas de fila	Cuenta de Estudios Padre	Cuenta de Estudios Padre2	Cuenta de Estudios Padre3	Cuenta de Estudios Padre4
Primaria	1	3,33%	1	3,33%
Secundaria	3	10,00%	4	13,33%
Bac/FP	21	70,00%	25	83,33%
Universidad	5	16,67%	30	100,00%
<b>Total general</b>	<b>30</b>	<b>100,00%</b>		

## 2.5. Tabulando Variables Cuantitativas a partir de micro-datos

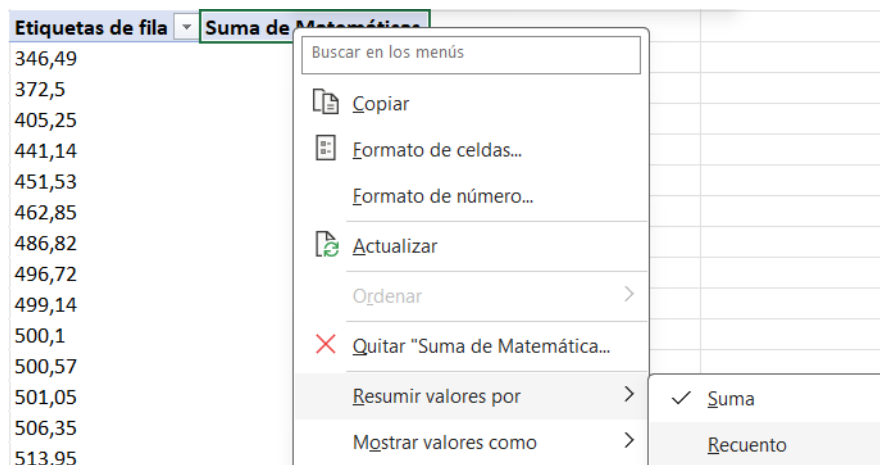
Seleccionando la variable cuantitativa de interés, procedemos a realizar una tabla dinámica igual que para cualquier otra variable, tal y como se ilustra en la Sección 2.4.1. En el caso de las puntuaciones de Matemáticas en el Informe Pisa, el resultado es el siguiente:



Etiquetas de fila	Suma de Matemáticas
346,49	346,49
372,5	372,5
405,25	405,25
441,14	441,14
451,53	451,53
462,85	462,85
486,82	486,82
496,72	496,72
499,14	499,14

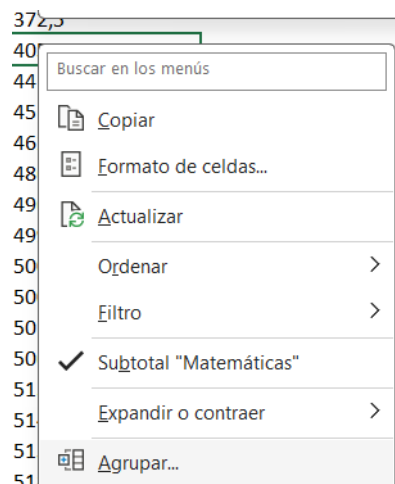
### Importante: la suma y el conteo.

En primer lugar, veremos que, en vez del conteo de observaciones de cada tipo, dependiendo de la versión de Excel, para datos cuantitativos el programa no pone la frecuencia absoluta, sino la suma. Lo observamos en el encabezado de la segunda columna. Para cambiarlo, hacemos clic derecho en el encabezado que pone *Suma de ...* (*Matemáticas* en este caso) y pulsamos *Resumir valores por*, luego seleccionamos *Recuento*.



### Para variables cuantitativas continuas

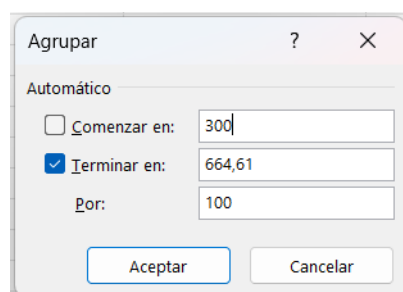
Si la variable es cuantitativa discreta, no hace falta agrupar los datos en intervalos. Sin embargo, cuando una variable cuantitativa toma muchos valores distintos, cada clase aparece una sola vez por lo general, por lo que **es necesario agrupar en clases** para elaborar un resumen. Para ello, hacemos clic izquierdo sobre cualquiera de las clases de la tabla y pulsamos la opción de *Agrupar*.



Por defecto, en los campos *Comenzar en:* y *Terminar en:* aparecerán el mínimo y el máximo de las observaciones, respectivamente. Esta información también podríamos haberla obtenido previamente utilizando las funciones MIN y MAX. Sin embargo, es importante reflexionar sobre cómo queremos hacer la agrupación, es decir, dónde establecer el inicio de los intervalos y cuál será su amplitud.

Para definir los intervalos de las clases agrupadas:

- En *Comenzar en:* seleccionamos un valor inferior al que aparece por defecto, de manera que el primer límite sea un número redondo.
- En *Terminar en:* generalmente mantenemos el valor máximo para asegurarnos de contar todas las observaciones.
- En *Por:* indicamos la amplitud de los intervalos que queremos crear.



Al darle a aceptar, ya tenemos en la tabla dinámica el resultado del conteo en clases de intervalos.

Etiquetas de fila ▾	Cuenta de Matemáticas
300-400	2
400-500	7
500-600	18
600-700	3
<b>Total general</b>	<b>30</b>

**Para calcular el resto de frecuencias:** procedemos como para el resto de variables explicadas anteriormente en la Subsección [2.4.1](#).

Etiquetas de fila ▾	Cuenta de Matemáticas	Cuenta de Matemáticas2	Cuenta de Matemáticas3	Cuenta de Matemáticas4
300-400	2	6,67%	2	6,67%
400-500	7	23,33%	9	30,00%
500-600	18	60,00%	27	90,00%
600-700	3	10,00%	30	100,00%
<b>Total general</b>	<b>30</b>	<b>100,00%</b>		



# 3

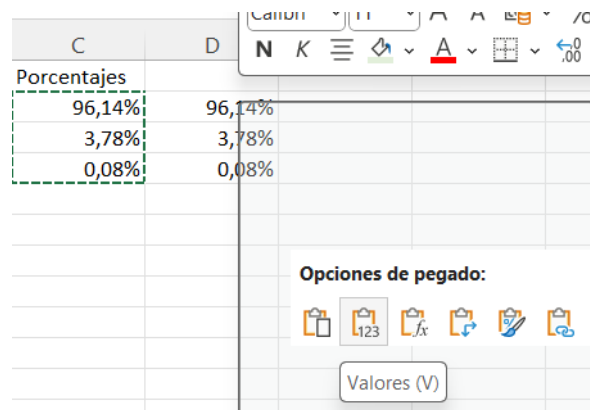
## Representación Gráfica

Como consideración general, no se recomienda realizar un gráfico directamente sobre los resultados de una tabla dinámica ni de un resumen tabular. Es menos propenso a errores copiar y pegar aparte los valores que se van a representar, generalmente los nombres de las clases y las frecuencias que se utilizarán.

**Copiar y pegar valores:** Si copiamos y pegamos resultados de valores calculados mediante fórmulas, frecuentemente obtendremos valores no deseados o un error. Esto ocurre porque, en realidad, cuando se copia y pega en Excel, no se copian los valores, sino las fórmulas o funciones utilizadas en cada celda. Por ejemplo, si copiamos los porcentajes del resumen tabular de la variable “tipo de ruptura”, observaremos que no se copian los valores, sino que aparece 0% en todas las clases.

C	D
Porcentajes	
96,14%	0,00%
3,78%	0,00%
0,08%	0,00%

La solución universal para resolver este problema es *copiar como valor*. Así pues, cuando se hace clic derecho para pegar lo copiado (los porcentajes), se puede hacer clic en el icono del portapapeles con los números 123, lo cual realiza el pegado solo de los valores.



De esta manera se obtiene el resultado deseado:

C	D
Porcentajes	
96,14%	96,14%
3,78%	3,78%
0,08%	0,08%

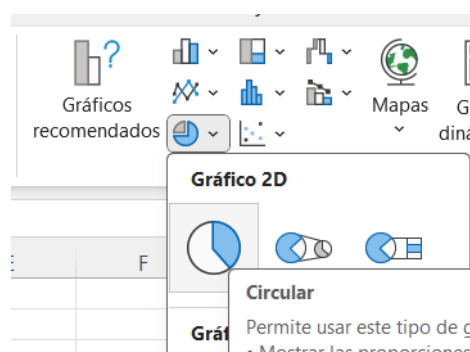
### 3.1. Representación de una Variable Nominal

#### 3.1.1. Diagrama de Sectores

1. Seleccionar los valores de la variable y sus frecuencias absolutas o relativas.

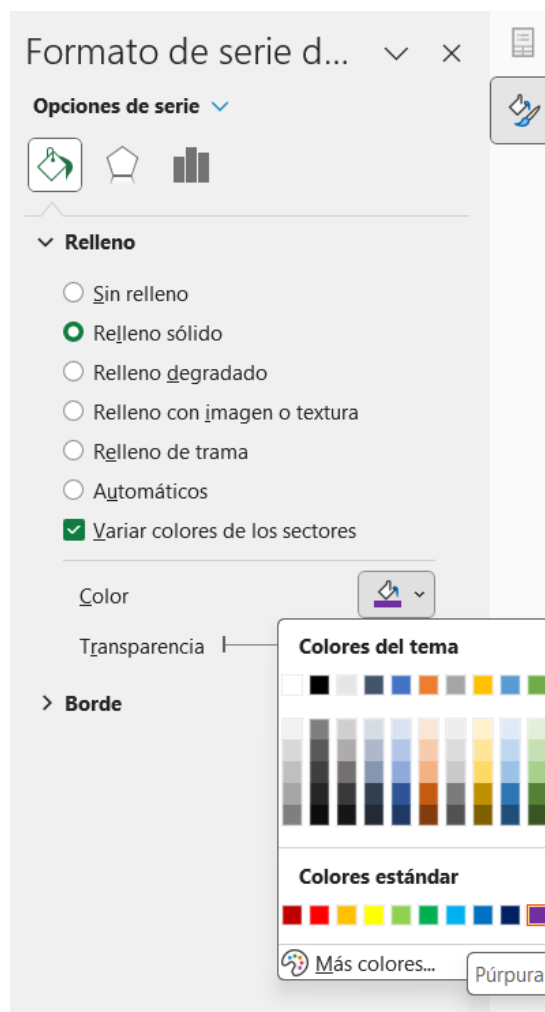
Femenino	60,00%
Masculino	40,00%

2. Ir al menú *Insertar*.
3. Seleccionar un gráfico circular o de anillos: es recomendable optar por el gráfico circular en 2D para una representación simple.

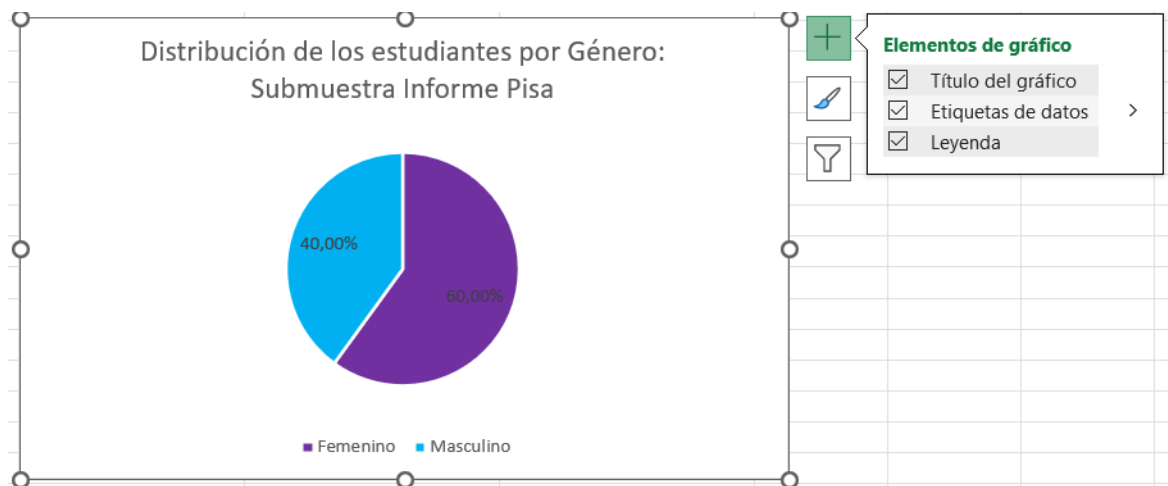


4. Etiquetar y formatear el gráfico según se desee. Para ello, hacer clic en el símbolo + que aparece en la esquina superior derecha del gráfico generado, y seleccionar la opción para mostrar etiquetas de los datos o títulos.

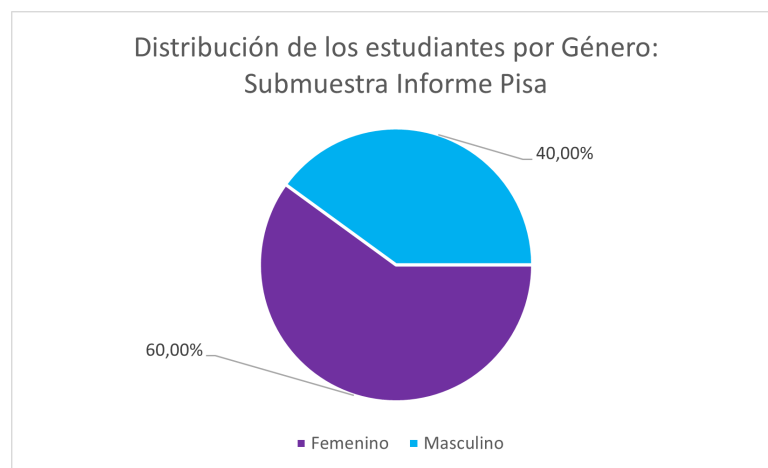
**Cambiar colores:** Para cambiar los colores, haz doble clic en la parte del sector circular que deseas modificar, y en el menú que aparece a la derecha, busca el icono del cubo de pintura dentro del campo “Relleno” para elegir el color deseado.



**Etiquetas de los datos:** Las etiquetas se pueden seleccionar y mover a la ubicación deseada para mejorar la visibilidad.



Además, existen múltiples opciones para modificar estéticamente el gráfico que el usuario puede explorar. Un posible resultado de un buen diagrama circular es el siguiente:



**Nota 3.1.** Se ha cambiado la orientación de los sectores respecto al penúltimo gráfico. No es necesario realizarlo, pero ¿puedes encontrar la opción para hacerlo? Busca la opción *Ángulo del primer sector* y experimenta con los números.

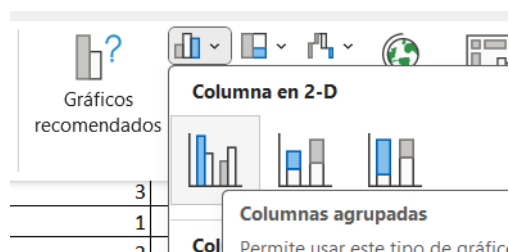
### 3.1.2. Diagrama de Barras

1. Seleccionar los valores de la variable y sus frecuencias absolutas o relativas.

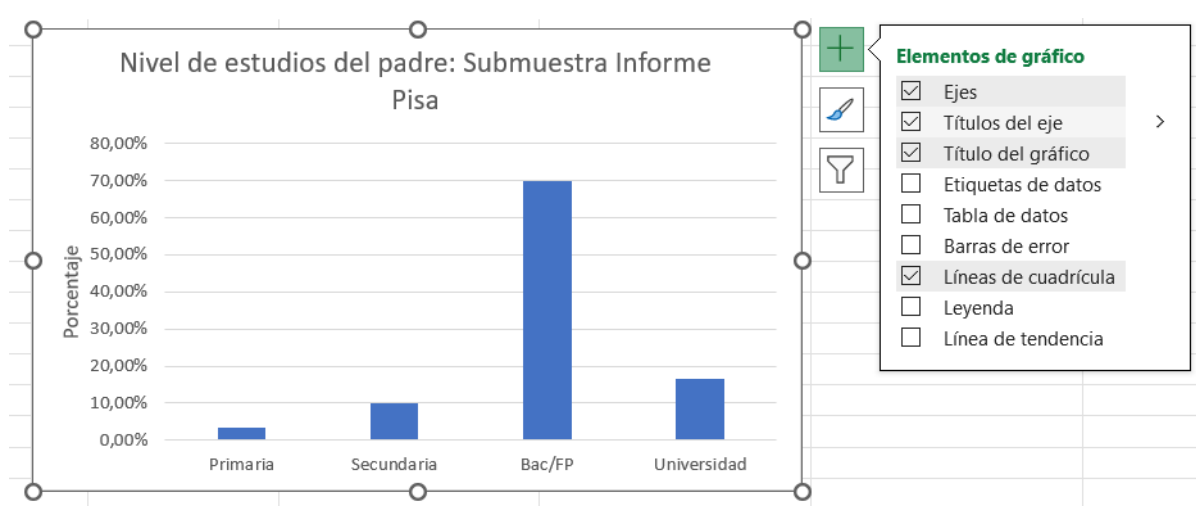
Primaria	3,33%
Secundaria	10,00%
Bac/FP	70,00%
Universidad	16,67%



2. Insertar un gráfico de columnas (vertical) o de barras (horizontal), según la preferencia.
3. Seleccionar un gráfico de columnas en 2D.



4. Etiquetar y formatear el gráfico según se desee: es necesario incluir los títulos de los ejes.



### 3.2. Representación de una Variable Ordinal

Para representar un diagrama de barras de una variable cualitativa ordinal, se procede como en la Subsección 3.1.2. Es imperativo ordenar correctamente la variable, puesto que, de lo contrario, la representación gráfica resultará engañosa.

### 3.3. Representación de una Variable Cuantitativa Discreta

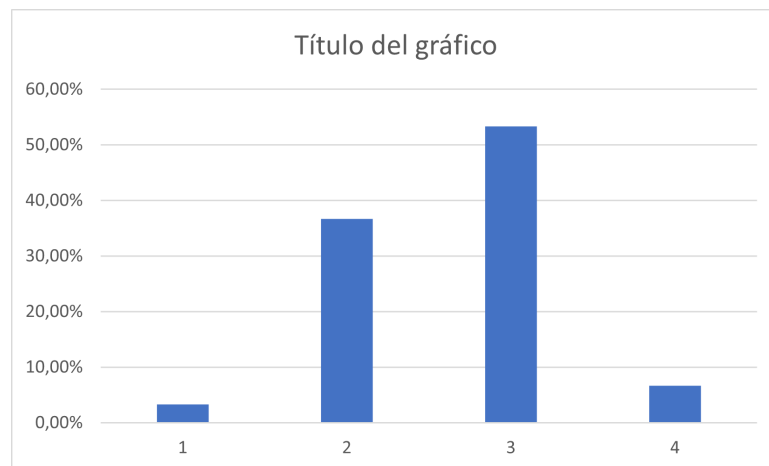
Para representar gráficamente una variable cuantitativa discreta, la técnica adecuada es la del diagrama de barras. Sin embargo, al ser las clases también números, en Excel es necesario realizar un paso adicional para obtener un resumen gráfico correcto.

A continuación, se muestra cómo representar gráficamente la variable *número de coches* en casa, utilizando la submuestra del Informe Pisa.

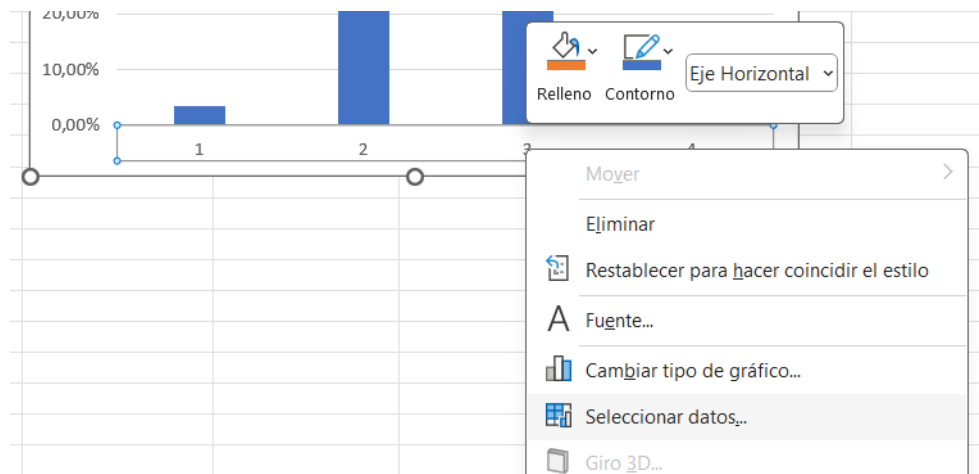
1. Seleccionar **únicamente** las frecuencias de la variable que se deseen representar. No se deben seleccionar las clases de la variable, que en este caso son 0, 1, 2 y 3.

	3,33%
	36,67%
	53,33%
	6,67%

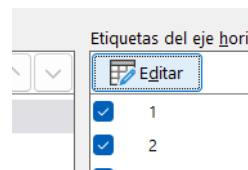
2. Insertar un gráfico de barras o columnas en 2D, como se explicó en secciones anteriores.



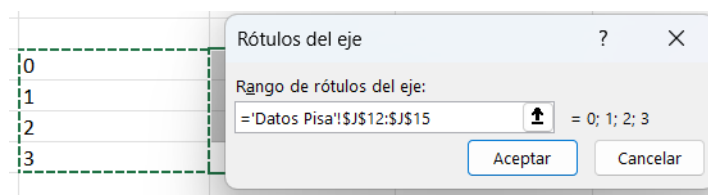
3. Nótese que, por defecto, las etiquetas del eje horizontal comienzan en 1. Si estas etiquetas automáticas no se corresponden con las clases que queremos representar, deben ser modificadas. Para ello, se debe seleccionar el lugar donde se encuentran los nombres de las clases en el eje horizontal y hacer clic derecho, donde aparecerá la opción *Seleccionar datos*.



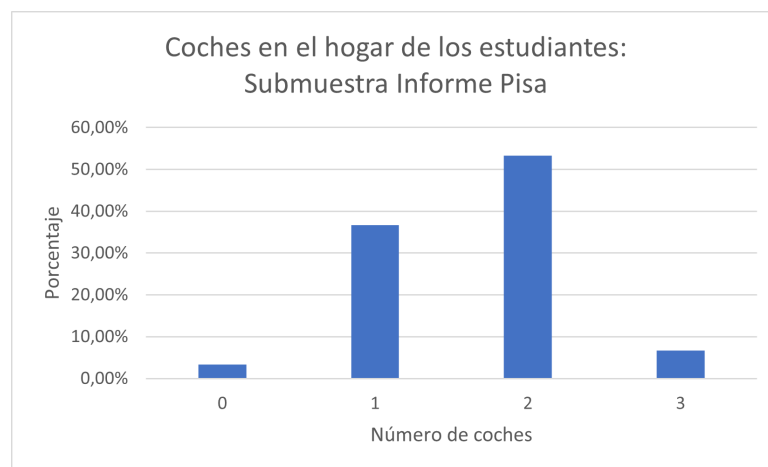
4. Pulsar *Editar* debajo de donde pone *Etiquetas del eje horizontal*.



5. En el menú que aparece, seleccionar el rango donde están las etiquetas de las clases a representar. En este caso, los números del 0 al 3. Pulsar *Aceptar*.



6. Etiquetar y formatear el gráfico según se desee: títulos de los ejes, colores. . .



### 3.4. Representación de una variable cuantitativa continua

Los histogramas y polígonos de frecuencias representan las densidades de cada intervalo. En situaciones donde los intervalos tienen la misma amplitud, las densidades son proporcionales a las frecuencias. En ese caso, representaremos las frecuencias directamente, sin calcular las densidades, ya que es más intuitivo y fácil de comprender.

**Nota 3.2.** El histograma que calcula Excel por defecto tiene, entre sus defectos, la falta de un control absoluto sobre las clases de intervalo que queremos construir, por lo que no lo utilizaremos. En su lugar, se propone representar un diagrama de barras, en el que habremos de juntar las barras para que tenga el aspecto de un histograma.

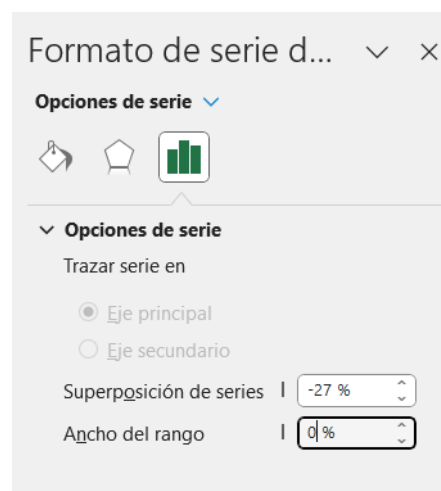
### 3.4.1. Histograma para representar una variable cuantitativa agrupada en intervalos de amplitudes iguales

Cuando las amplitudes son iguales, tan solo debemos representar un gráfico de barras como en el caso de las variables ordinales, tal y como se describe en la Sección 3.1.2, donde seleccionaremos las etiquetas de las clases y las frecuencias a representar.

300-400	6,67%
400-500	23,33%
500-600	60,00%
600-700	10,00%

El único cambio necesario respecto a esas instrucciones es asegurarse de que las barras del histograma no tengan separación, para dar la sensación de continuidad de la variable. Para ello:

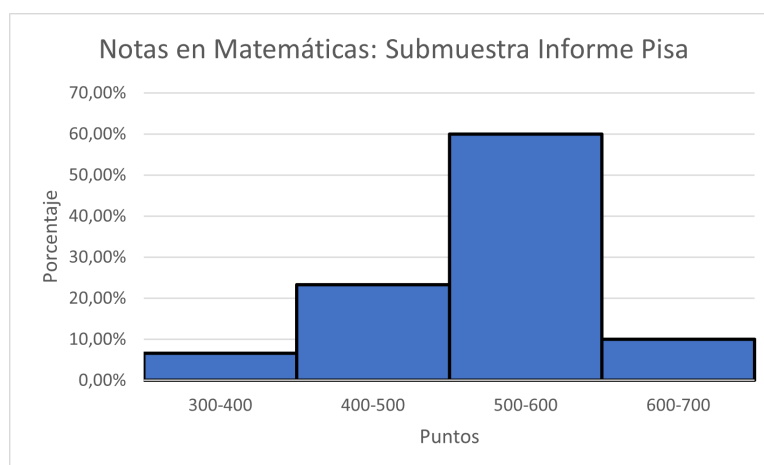
1. Haz doble clic en una barra del diagrama de barras. Aparecerá un menú a la derecha.



2. En las opciones de serie del menú se puede manipular el *Ancho del rango*, en el que introduciremos un 0.



Con estos dos pasos ya tendremos un histograma. Faltaría, por supuesto, formatear el gráfico con los títulos de los ejes y etiquetas si lo deseamos...



### 3.4.2. Polígonos de frecuencias como sustituto de un histograma cuando las amplitudes son desiguales: Marca de Clase, Amplitud y Densidad de Intervalos

Si queremos hacer un resumen tabular, no se deben representar la amplitud, la densidad o la marca de clase, ya que añaden información difícil de interpretar a simple vista o sobrecargan innecesariamente el resumen, haciéndolo menos legible. Sin embargo, estos cálculos nos permitirán realizar representaciones gráficas y calcular medidas de posición en secciones posteriores, por lo que mostramos cómo calcularlos. Estos cálculos se ilustran con los datos del gasto de las familias.

1. **Marca de Clase:** Es la semisuma (promedio) de los extremos del intervalo. Juega

el papel de *representante del intervalo* y se calcula como:

$$=(\text{extremo superior} + \text{extremo inferior})/2.$$

En Excel, si se dispone de los extremos inferiores y superiores, se puede utilizar la función *promedio*.

	A	B	C	D	E	F	G
1	Gasto mensual (euros)						
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)	Marca de clase (xi)
3	150	250	10	5,0%	10	5,0%	=promedio(A3:B3)
4	250	350	25	12,5%	35	17,5%	
5	350	450	50	25,0%	85	42,5%	
6	450	750	80	40,0%	165	82,5%	
7	750	1100	30	15,0%	195	97,5%	
8	1100	1500	5	2,5%	200	100,0%	
9			200				

Finalmente, se arrastra para generalizar la fórmula a todas las clases.

2. **Amplitud:** Es la diferencia entre los extremos del intervalo. En Excel, se calcula como  $=(\text{extremo superior}) - (\text{extremo inferior})$ .

	A	B	C	D	E	F	G	H
	Gasto mensual (euros)							
	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)	Marca de clase (xi)	Amplitud (Ai)
	150	250	10	5,0%	10	5,0%	200	=B3-A3
	250	350	25	12,5%	35	17,5%		300
	350	450	50	25,0%	85	42,5%		400
	450	750	80	40,0%	165	82,5%		600
	750	1100	30	15,0%	195	97,5%		925
	1100	1500	5	2,5%	200	100,0%		1300
			200					

De nuevo, se arrastra para generalizar la fórmula a todas las clases.

3. Podemos calcular la densidad que consideremos más conveniente (normalmente la absoluta si disponemos de un censo y la relativa o porcentual en caso contrario).

- **Densidad absoluta:** Se define como el número de observaciones dividido por la amplitud del intervalo. En Excel, se calcula como:

$$=(\text{frecuencia absoluta})/(\text{amplitud del intervalo}).$$

- **Densidad relativa o porcentual:** Se define como la proporción o porcentaje de observaciones dividido por la amplitud del intervalo. En Excel, se calcula como:

$$=(\text{frecuencia relativa o porcentual})/(\text{amplitud del intervalo}).$$

	A	B	C	D	E	F	G	H	I
1	Gasto mensual (euros)								
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)	Marca de clase (xi)	Amplitud (Ai)	Densidad (di)
3	150	250	10	5,0%	10	5,0%	200	100	=D3/H3
4	250	350	25	12,5%	35	17,5%	300	100	
5	350	450	50	25,0%	85	42,5%	400	100	
6	450	750	80	40,0%	165	82,5%	600	300	
7	750	1100	30	15,0%	195	97,5%	925	350	
8	1100	1500	5	2,5%	200	100,0%	1300	400	
9			200						

El resultado final del cálculo de estas medidas será:

i)	Marca de clase (xi)	Amplitud (Ai)	Densidad (di)
✓	200	100	0,0005
✓	300	100	0,00125
✓	400	100	0,0025
✓	600	300	0,00133333
✓	925	350	0,00042857
✓	1300	400	0,0000625

### Polígono de Frecuencias: clases de amplitud desigual

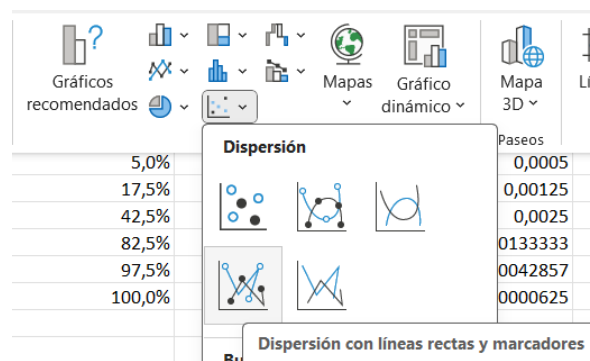
Cuando las amplitudes de los intervalos son desiguales, es indispensable representar las densidades en lugar de las frecuencias, ya que de esta manera se ponderan los intervalos en función de su amplitud y las comparaciones entre ellos son justas.

Partimos de la base de que ya se han calculado las marcas de clase y las densidades tal y como se ha hecho anteriormente. Luego:

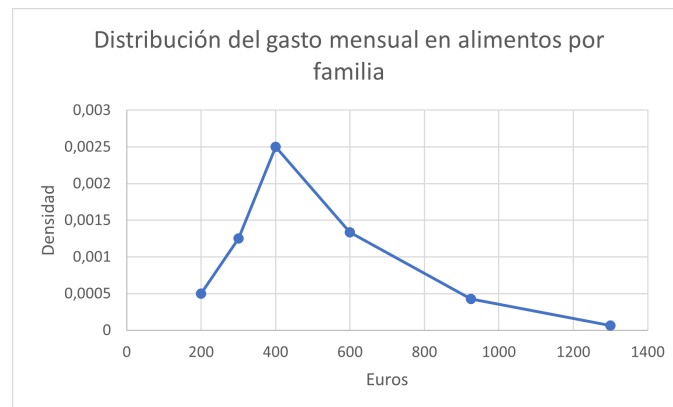
1. Seleccionar las marcas de clase y las densidades, para lo cual lo ideal es copiar y pegar (como valores) los resultados obtenidos aparte.

Marca de clas	Densidad (di)
200	0,0005
300	0,00125
400	0,0025
600	0,00133333
925	0,00042857
1300	0,0000625

2. Se seleccionan los datos que se acaban de copiar y, en el menú *Insertar*, se selecciona el icono del *Gráfico de Dispersión*, donde la opción recomendada es *Dispersión con líneas rectas y marcadores*.



3. Formatear el gráfico con los títulos de los ejes, etiquetas, etc.



### 3.5. Representación de series temporales: Polígono de frecuencias

Vamos a representar la serie de defunciones y nacimientos en España desde 1992 [4]. En la hoja *INE DefNac* puedes encontrar los datos. Si solo se quisiera representar una serie temporal, basta con replicar los pasos utilizando una columna menos. Para ello:

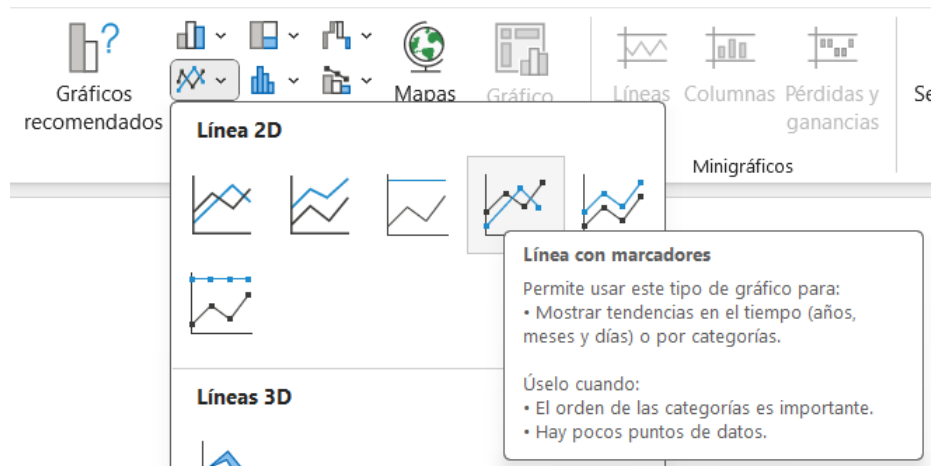
1. Selecciona las tres columnas, incluyendo la del año.

	A	B	C
1	Año	Defunciones	Nacimientos
2	2021	450744	337380
3	2020	493776	341315
4	2019	416602	360617

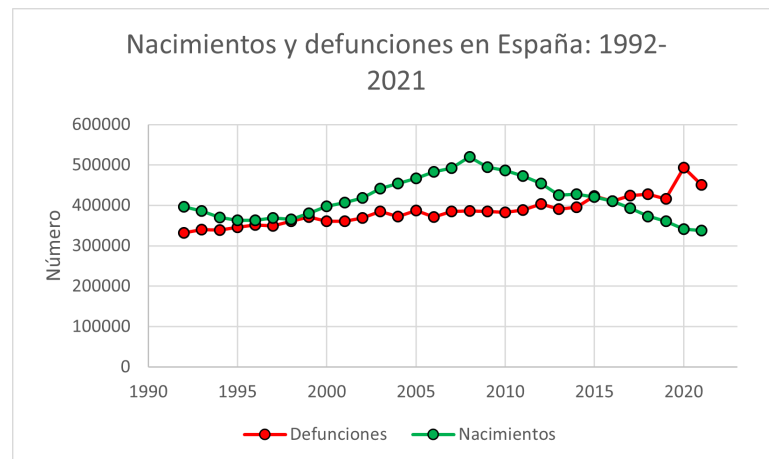
2. (a) Si la columna de la variable que indica la variable temporal es un **número** (por ejemplo un año, como en el ejemplo), entonces: Selecciona *Insertar*, después el icono del *Gráfico de Dispersión*, y elige la opción recomendada: *Dispersión con líneas rectas y marcadores*, igual que en la Subsección 3.4.2.



- (b) Si por el contrario la variable temporal no es un número (por ejemplo, son cadenas de caracteres como 2023T1 o ene-24) entonces: Selecciona *Insertar*, después el icono del *Línea 2D*, y elige la opción recomendada: *Línea con marcadores*.

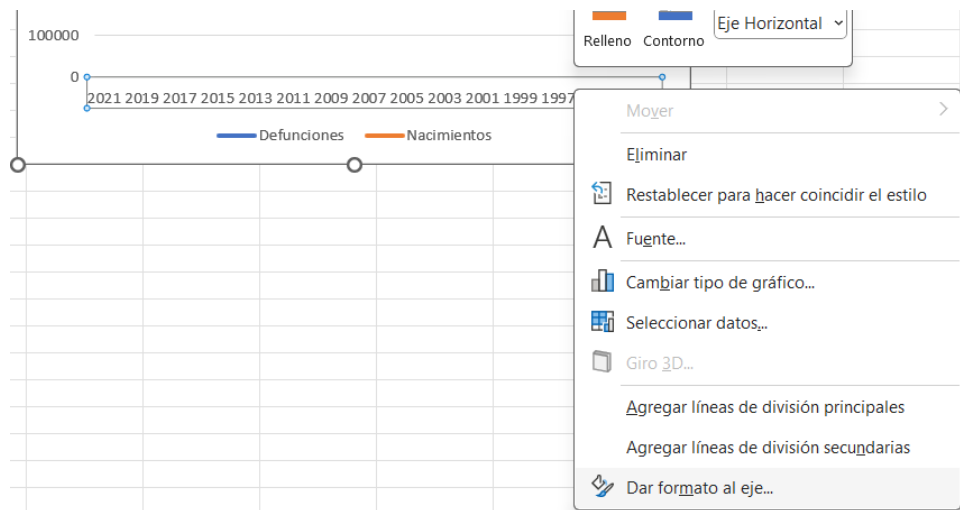


3. Obtendrás el resultado, que ahora debe formatearse como de costumbre con las unidades de los ejes...

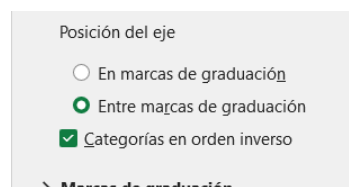


### 3.5.1. Invertir el orden en el eje horizontal

A veces, como en este caso, los datos están en orden contrario al cronológico. Para cambiarlo, haz clic con el botón izquierdo en el eje horizontal y selecciona *Dar formato al eje...*

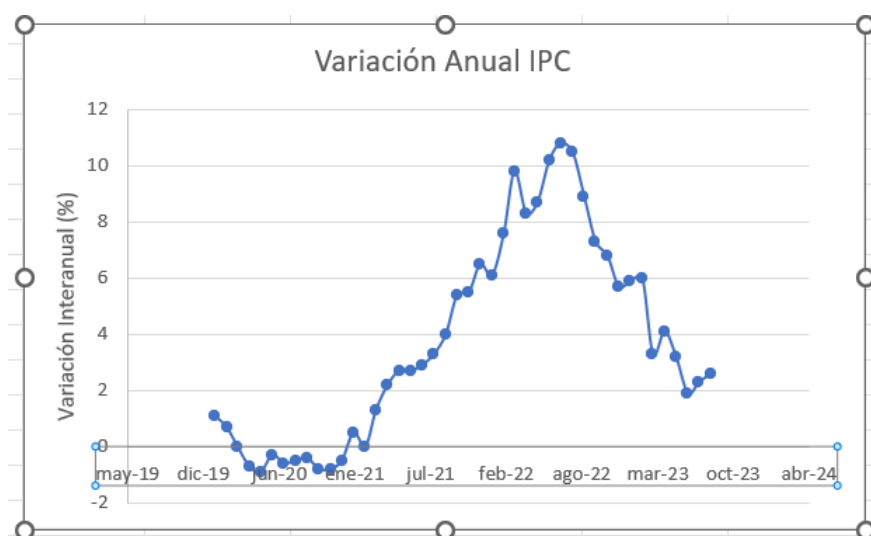


En el menú de la derecha, dentro del submenú *Opciones del eje*, selecciona *Categoría en orden inverso*.



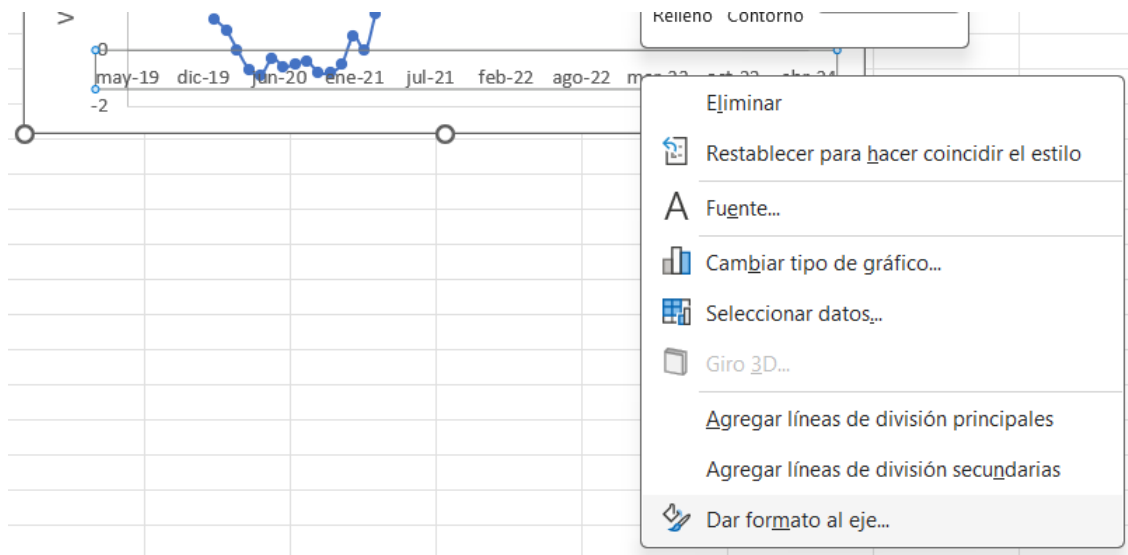
### 3.5.2. Modificar el lugar de las etiquetas en el caso de que haya valores negativos

Si hay valores negativos, las etiquetas del eje horizontal se pueden solapar con la representación gráfica. Véase como ejemplo la siguiente representación de la variación interanual del IPC (datos del INE).

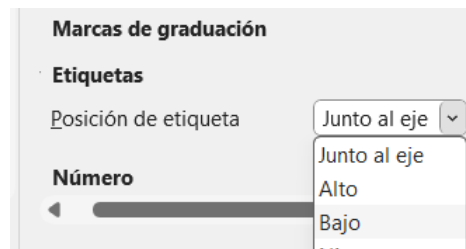


Para subsanar esta situación se puede proceder de la siguiente manera.

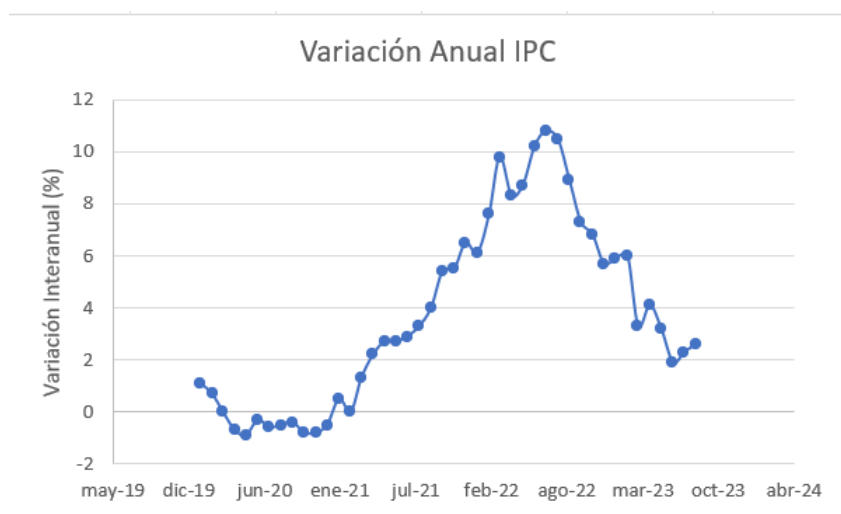
1. Hacer clic derecho sobre las etiquetas del eje horizontal y selecciona *Dar formato al eje...*



2. En la opciones, buscamos en el submenú de *Etiquetas* y seleccionamos otra posición, como por ejemplo *Bajo*.



El resultado final será el siguiente:





# 4

## Medidas de posición

### 4.1. Medidas de Posición en tablas

#### 4.1.1. Moda

- **Si las clases de la tabla no son intervalos de amplitud desigual:** Para identificar la moda, basta con señalar la clase más frecuente mediante alguna de las frecuencias no acumuladas. Si las clases son intervalos, hablaremos de un intervalo modal y no de un valor concreto.

**Nota 4.1.** Este método aplica a variables de cualquier tipo, incluyendo cualitativas. Se excluye solo el caso en que se tiene una variable cuantitativa agrupada en clases de intervalos desiguales.

- **Si las clases de la tabla sí son intervalos de amplitud desigual:** Para identificar la moda, señalaremos el intervalo con mayor densidad, y no necesariamente con mayor frecuencia absoluta o relativa. Para calcular la densidad, se remite a la Sección [3.4.2](#).

**Ejemplo:** en los datos de gastos de las familias, la densidad más alta es 0,0025 y corresponde a la clase de 350 a 450 euros.

Alternativamente, **si se dispone de un histograma o polígono de frecuencias**, se puede identificar la moda gráficamente utilizando las densidades (o frecuencias con amplitudes iguales), siendo esta la clase con una densidad (o frecuencia, en el caso de amplitudes iguales) mayor también en el gráfico.

### 4.1.2. Mediana y otros percentiles.

**Mediana:** Basta con identificar la primera clase para la cual la frecuencia porcentual acumulada iguala o supera el 50 %.

**Percentil  $k$ :** Identifica la primera clase para la cual la frecuencia porcentual acumulada iguala o supera el  $k$  %, considerándola como la clase en la que se encuentra el percentil  $k$ .

**Ejemplo:** En los datos de los gastos de la familia, la mediana está en el intervalo de 450 a 750 euros, mientras que el percentil 25 se encuentra en el intervalo de 350 a 450 euros.

	A	B	C	D	E	F
1	Gasto mensual (euros)					
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)
3	150	250	10	5,0%	10	5,0%
4	250	350	25	12,5%	35	17,5%
5	350	450	50	25,0%	85	42,5%
6	450	750	80	40,0%	165	82,5%
7	750	1100	30	15,0%	195	97,5%
8	1100	1500	5	2,5%	200	100,0%
9			200			

### 4.1.3. Media

Existen dos fórmulas alternativas y equivalentes, donde para la primera se necesitan las frecuencias absolutas, mientras que para la segunda se hace uso de las frecuencias relativas:

$$\text{Media} = \frac{\sum x_i n_i}{N} = \sum x_i f_i.$$

#### ■ Primera fórmula.

1. Calculamos una columna al final de la tabla con los valores  $x_i \times n_i$  (para  $x_i$  se utilizan las marcas de clase si los datos están agrupados en intervalos). Para ello, calculamos la primera celda multiplicando el valor de la frecuencia absoluta de la primera clase multiplicado por su marca de clase.

	A	B	C	D	E	F	G	H	I	J
1	Gasto mensual (euros)									
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)	Marca de clase (xi)	Amplitud (Ai)	Densidad (di)	xi ni
3	150	250	10	5,0%	10	5,0%	200	100	0,0005	=G3*C3
4	250	350	25	12,5%	35	17,5%	300	100	0,00125	
5	350	450	50	25,0%	85	42,5%	400	100	0,0025	
6	450	750	80	40,0%	165	82,5%	600	300	0,00133333	
7	750	1100	30	15,0%	195	97,5%	925	350	0,00042857	
8	1100	1500	5	2,5%	200	100,0%	1300	400	0,0000625	

2. Se arrastra para generalizar el cálculo al resto de clases.

	J
di)	xi ni
05	2000
125	
125	
133	
157	
125	

3. Se suman todos los valores de la columna (por ejemplo, con autosuma). El resultado obtenido será  $\sum x_i \times n_i$ .

	J
)	xi ni
5	2000
5	7500
5	20000
3	48000
7	27750
5	6500
	=SUMA(J3:J8)

4. Finalmente, se divide el resultado de la suma entre el tamaño muestral  $N$ .

	A	B	C	D	E	F	G	H	I	J
1	Gasto mensual (euros)									
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)	Marca de clase (xi)	Amplitud (Ai)	Densidad (di)	xi ni
3	150	250	10	5,0%	10	5,0%	200	100	0,0005	2000
4	250	350	25	12,5%	35	17,5%	300	100	0,00125	7500
5	350	450	50	25,0%	85	42,5%	400	100	0,0025	20000
6	450	750	80	40,0%	165	82,5%	600	300	0,00133333	48000
7	750	1100	30	15,0%	195	97,5%	925	350	0,00042857	27750
8	1100	1500	5	2,5%	200	100,0%	1300	400	0,0000625	6500
9			200							111750
10										
11	Marca de clas Densidad (di)									Media
										=J9/C9

En este caso, el resultado es de 558,75 euros mensuales de media.

Media	558,75
-------	--------

#### ■ Segunda fórmula.

1. Se calcula la columna al final de la tabla con los valores  $x_i \times f_i$ . Para ello, se multiplica la columna del *Porcentaje* por la de la *marca de clase*.

	A	B	C	D	E	F	G	H	I	J
1	Gasto mensual (euros)									
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)	Marca de clase (xi)	Amplitud (Ai)	Densidad (di)	xi fi
3	150	250	10	5,0%	10	5,0%	200	100	0,0005	=D3*G3
4	250	350	25	12,5%	35	17,5%	300	100	0,00125	
5	350	450	50	25,0%	85	42,5%	400	100	0,0025	
6	450	750	80	40,0%	165	82,5%	600	300	0,00133333	
7	750	1100	30	15,0%	195	97,5%	925	350	0,00042857	
8	1100	1500	5	2,5%	200	100,0%	1300	400	0,0000625	

**Nota 4.2. Importante:** Hemos multiplicado por la columna del porcentaje, porque está expresado en Excel como porcentaje (vemos el símbolo %). Si los porcentajes fueran números como 5; 12,5; ... sin el símbolo de porcentaje, entonces habría que dividirlos por 100 en otra columna para calcular las verdaderas frecuencias relativas  $f_i$ .

2. Se arrastra la primera celda para calcular  $x_i \times f_i$  para todas las clases.

xi	fi
5	10
5	
5	
3	
7	
5	

3. Se suma, normalmente con el botón de autosuma, para obtener  $\sum x_i \times f_i$ .

xi	fi
5	10
5	37,5
5	100
3	240
7	138,75
5	32,5
=SUMA(J3:J8)	

Como la fórmula ya está aplicada por completo, este será el resultado de la media: 558,75 euros mensuales, que como no puede ser de otra manera, coincide con el resultado calculado mediante la primera fórmula.

xi	fi
5	10
5	37,5
5	100
3	240
7	138,75
5	32,5
	558,75

## 4.2. Medidas de Posición en microdatos

### 4.2.1. Variables Cualitativas

Calculamos los resúmenes tabulares mediante tablas dinámicas como en la Sección 2 y hallamos las medidas deseadas siguiendo la Sección 4.1.

### 4.2.2. Variables Cuantitativas

Para calcular las medidas de posición, vamos a aplicar las siguientes funciones, cuyo argumento será el rango de valores donde se encuentren los datos para los que deseamos calcular dichas medidas:

- **Moda:** *MODA.VARIOS*.



**Nota 4.3.** La moda puede no ser única, y esta función puede devolver varios valores en consecuencia.

**Nota 4.4.** No se recomienda calcular la moda en variables cuantitativas con muchos valores distintos. Si disponemos de un número muy elevado de datos, el cálculo puede tomar demasiado tiempo al ordenador.

- Mediana: *MEDIANA*.
- Media: *PROMEDIO*.
- Percentiles: *PERCENTIL.INC*.

Debe indicarse de la siguiente manera:

$$=PERCENTIL.INC(rango\_celdas;k),$$

donde  $k$  es el valor del percentil dividido por 100. Por ejemplo, para calcular el percentil 37, debemos escribir  $=PERCENTIL.INC(rango\_celdas;0,37)$ .

**Ejemplo:** En los datos del Informe Pisa, la variable *Matemáticas* está en la columna *G*. Un truco habitual para calcular un resumen numérico de una variable consiste en escribir la fórmula deseada, y al seleccionar el rango de celdas sobre el que se aplica la función, hacer clic sobre el encabezado de la columna. De esta manera, aparecerá escrito *G:G*, lo que indica que se debe hacer el resumen con todos los valores de la columna *G*.

A continuación se muestran las fórmulas utilizadas para calcular todas las medidas de síntesis, excepto la moda, ya que se ha recomendado no utilizarla en variables cuantitativas continuas.

Mínimo	=MIN(G:G)
Q1 o P25	=PERCENTIL.INC(G:G;0,25)
Mediana	=MEDIANA(G:G)
Q3 o P75	=PERCENTIL.INC(G:G;0,75)
Máximo	=MAX(G:G)
Media	=PROMEDIO(G:G)

Mínimo	346,49
Q1 o P25	497,325
Mediana	514,59
Q3 o P75	568,9875
Máximo	664,61
Media	520,112



# 5

## Medidas de dispersión y forma

### 5.1. Recorrido o Rango y rango intercuartílico

Para calcular el **recorrido** o **rango** se debe aplicar la definición directamente:  $\text{máx} - \text{mín}$ .

Para obtener el **rango intercuartílico** es suficiente con calcular los cuartiles primero y tercero, y luego hallar su diferencia tal y como se indica en la definición.

**Nota 5.1.** Para calcular los rangos en variables agrupadas en intervalos, se necesitan los valores exactos de los extremos (mínimo y máximo) o de los cuartiles. Para el rango, el máximo y el mínimo pueden tomarse como el menor valor posible de los datos (límite inferior de la primera clase) y el máximo como el mayor valor posible (límite superior de la última clase). El rango intercuartílico no se podrá calcular en esta situación.

### 5.2. Varianza, Desviación típica y CV: Datos Tabulados

Para el cálculo de la **varianza** existen dos fórmulas equivalentes cuando se tienen datos tabulados. Al igual que ocurre con la media, la primera de ellas hace uso de las frecuencias absolutas, mientras que la segunda necesita de las frecuencias relativas:

$$\text{Var} = \frac{1}{N} \sum_{i=1}^k (n_i \times x_i^2) - \bar{x}^2 = \sum_{i=1}^k (f_i \times x_i^2) - \bar{x}^2.$$

**Nota 5.2.** Si las clases son intervalos en vez de números individuales, como valor de las clases se utilizará la marca de clase  $x_i$ , al igual que se hacía en el cálculo de la media.

### ■ Primera fórmula.

1. Calculamos una columna al final de la tabla con los valores  $n_i \times x_i^2$ . Para ello, calculamos la primera celda multiplicando el valor de la frecuencia absoluta de la primera clase **por la marca de clase al cuadrado**.

	A	B	C	D	E	F	G	H	I	J
1	Gasto mensual (euros)									
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)	Marca de clase (xi)	Amplitud (Ai)	Densidad (di)	ni xi^2
3	150	250	10	5,0%	10	5,0%	200	100	0,0005	=C3*G3^2
4	250	350	25	12,5%	35	17,5%	300	100	0,00125	
5	350	450	50	25,0%	85	42,5%	400	100	0,0025	
6	450	750	80	40,0%	165	82,5%	600	300	0,00133333	
7	750	1100	30	15,0%	195	97,5%	925	350	0,00042857	
8	1100	1500	5	2,5%	200	100,0%	1300	400	0,0000625	

2. Se arrastra para generalizar el cálculo al resto de clases.

ni xi^2
400000

3. Se suman todos los valores de la columna. El resultado obtenido será  $\sum n_i \times x_i^2$ .

ni xi^2
400000
2250000
8000000
28800000
25668750
8450000
=SUMA(I3:I8)

4. En una celda en blanco se divide el resultado de la suma entre el tamaño muestral  $N$ . De esta manera, tendremos calculado  $\frac{1}{N} \sum n_i \times x_i^2$ .

	A	B	C	D	E	F	G	H	I	J
1	Gasto mensual (euros)									
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)	Marca de clase (xi)	Amplitud (Ai)	Densidad (di)	ni xi^2
3	150	250	10	5,0%	10	5,0%	200	100	0,0005	400000
4	250	350	25	12,5%	35	17,5%	300	100	0,00125	2250000
5	350	450	50	25,0%	85	42,5%	400	100	0,0025	8000000
6	450	750	80	40,0%	165	82,5%	600	300	0,00133333	28800000
7	750	1100	30	15,0%	195	97,5%	925	350	0,00042857	25668750
8	1100	1500	5	2,5%	200	100,0%	1300	400	0,0000625	8450000
9			200							73568750
10										
11								Media	558,75	
12								Suma/N	=J9/C9	
13								Varianza		

5. Para calcular la varianza, el paso final es **restar la media al cuadrado**. Calculamos la resta de la cantidad obtenida en el paso anterior, que es la suma dividida por  $N$ , y le restamos la media al cuadrado.

Media	558,75
Suma/N	367843,75
Varianza	=I12-I11^2

En este caso, el resultado es de 55642,1875 euros al cuadrado mensuales.

Media	558,75
Suma/N	367843,75
Varianza	55642,1875

### ■ Segunda fórmula.

1. Calculamos una columna al final de la tabla con los valores  $f_i \times x_i^2$ . Nuevamente, calculamos la primera celda multiplicando el valor de la frecuencia relativa porcentual de la primera clase **por la marca de clase al cuadrado**.

	A	B	C	D	E	F	G	H	I	J
1	Gasto mensual (euros)									
2	Lím. Inferior	Lím. Superior	Núm. Familias (ni)	Porcentaje (pi)	Frec. Abs. Acum. (Ni)	Frec. Por. Acum. (Pi)	Marca de clase (xi)	Amplitud (Ai)	Densidad (di)	fi xi^2
3	150	250	10	5,0%	10	5,0%	200	100	0,0005	=D3*G3^2
4	250	350	25	12,5%	35	17,5%	300	100	0,00125	
5	350	450	50	25,0%	85	42,5%	400	100	0,0025	
6	450	750	80	40,0%	165	82,5%	600	300	0,00133333	
7	750	1100	30	15,0%	195	97,5%	925	350	0,00042857	
8	1100	1500	5	2,5%	200	100,0%	1300	400	0,0000625	

**Nota importante:** Igual que en la media, hemos multiplicado por la columna del porcentaje, ya que está expresado en Excel como porcentaje (vemos el símbolo %). Si los porcentajes fueran números como 6; 4,5; ... sin el símbolo, entonces habría que dividirlos por 100 en otra columna para calcular las verdaderas frecuencias relativas  $f_i$ .

2. Se arrastra para generalizar el cálculo al resto de clases.

fi xi^2
2000

3. Se suman todos los valores de la columna. El resultado obtenido será  $\sum f_i \times x_i^2$ .

fi xi^2
2000
11250
40000
144000
128343,75
42250
=SUMA(J3:J8)

4. Para calcular la varianza, el paso final es **restar la media al cuadrado**. Calculamos la resta utilizando la cantidad obtenida en el paso anterior, que es la suma dividida por  $N$ , y le restamos la media al cuadrado.

H	I	J
Amplitud (Ai)	Densidad (di)	fi xi^2
) 100	0,0005	2000
) 100	0,00125	11250
) 100	0,0025	40000
) 300	0,00133333	144000
) 350	0,00042857	128343,75
) 400	0,0000625	42250
		367843,75
Media	558,75	
Varianza	=J9-I11^2	

De nuevo, el resultado es de 55642,1875 euros al cuadrado mensuales. Como no puede ser de otra manera, este resultado coincide con el calculado con la primera fórmula basada en las frecuencias absolutas.

Media	558,75
Varianza	55642,1875

Para el cálculo de la **desviación típica** solo hay que calcular la raíz cuadrada de la varianza. Para ello, utilizamos la función *RAIZ*. Por ejemplo *=RAIZ(C10)* si la varianza anteriormente calculada estuviera en la celda *C10*.

Finalmente, el cálculo del **Coefficiente de Variación** o **CV** se realiza dividiendo la desviación típica entre la media.

**Ejemplo:** Con el resultado de la media y varianza anterior, podemos calcular la desviación típica y el coeficiente de variación.

H	I
Amplitud (Ai)	Densidad (di)
=B3-A3	=D3/H3
=B4-A4	=D4/H4
=B5-A5	=D5/H5
=B6-A6	=D6/H6
=B7-A7	=D7/H7
=B8-A8	=D8/H8
Media	558,75
Varianza	=J9-I11^2
Desv. Típica	=RAIZ(I12)
CV	=I13/I11

Observamos que el valor de la desviación típica es de 235,89 euros y el Coeficiente de Variación es del 42 %.

Media	558,75
Varianza	55642,1875
Desv. Típica	235,885963
CV	42%

### 5.3. Cálculo de medidas de dispersión y forma sobre micro-datos

- **Recorrido o rango:** Calcular *máximo - mínimo*.
- **Recorrido intercuartílico:** Calcular la diferencia entre los cuartiles  $Q_3 - Q_1$ .
- **Varianza:** Aplicar la función *VAR.P* sobre el rango de celdas deseado.
- **Desviación típica:** Dos opciones. La primera, calcular la raíz cuadrada de la varianza. La segunda, utilizar la función *DESVEST.P* sobre el rango de celdas.
- **Coefficiente de Variación:** Dividir la desviación típica entre la media. Habitualmente se expresa como porcentaje.
- **Coefficiente de Asimetría de Fisher:** Utilizar la función *COEFICIENTE.ASIMETRIA.P* sobre el rango de las celdas con los valores numéricos de la variable.

**Nota 5.3.** En varias de estas medidas, el nombre de la función acaba en *.P*. El motivo de ello es que existen estimadores alternativos para estas cantidades. En principio, utilizando otras versiones, no debería haber una diferencia sustancial entre los resultados, especialmente conforme el tamaño de la muestra es mayor.

**Ejemplo:** Se ilustra nuevamente el cálculo de estas medidas utilizando los datos de la puntuación en *Matemáticas* en la submuestra del Informe Pisa, que se encuentra en la columna *G*.

J	K
Mínimo	=MIN(G:G)
Q1 o P25	=PERCENTIL.INC(G:G;0,25)
Mediana	=MEDIANA(G:G)
Q3 o P75	=PERCENTIL.INC(G:G;0,75)
Máximo	=MAX(G:G)
Media	=PROMEDIO(G:G)
Rango	=K6-K2
Rango Intercuartílico	=K5-K3
Desviación Típica	=DESVEST.P(G:G)
Coefficiente de Variación	=K13/K8
Coefficiente Asimetría de Fisher	=COEFICIENTE.ASIMETRIA.P(G:G)

Los resultados de estas operaciones son:

J	K
Mínimo	346,49
Q1 o P25	497,325
Mediana	514,59
Q3 o P75	568,9875
Máximo	664,61
Media	520,112
Rango	318,12
Rango Intercuartílico	71,6625
Desviación Típica	70,38797901
Coefficiente de Variación	14%
Coefficiente Asimetría de Fisher	-0,434855478

#### 5.4. Cálculo de puntuaciones tipificadas

Aplicamos la fórmula directamente, es decir, calculamos

$$\frac{x_i - \bar{x}}{s_x}.$$

Para ello, necesitamos el valor sobre el que queremos calcular esta puntuación, la media y la desviación típica de la variable.

Con estos valores, seleccionamos una celda en blanco y realizamos la operación:  $=(valor-media)/Desv.Típica$ .

**Ejemplo:** En el conjunto de datos, anterior, ¿cuál es la puntuación tipificada de una familia que gasta 200 euros?

Debemos calcular

$$\frac{200 - 558,75}{235,86}$$

cuyo resultado es  $-1,52$ .

Dado que ya tenemos la media y la desviación típica calculadas en Excel, también podemos realizar el cálculo reconstruyendo la operación deseada haciendo clic en la media y la desviación típica calculadas, obteniendo el mismo resultado:

Media	558,75
Varianza	55642,1875
Desv. Típica	235,885963
CV	42%
Z_200	$=(200-I11)/I13$
Z_200	-1,52086201



# 6

## Análisis bivalente: una variable cualitativa y otra cuantitativa

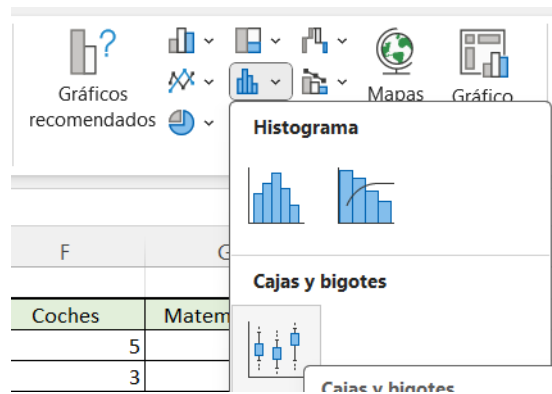
### 6.1. Representación gráfica: Diagramas de caja

Los diagramas de caja, también conocidos como *boxplots*, son herramientas gráficas que ofrecen una visión de la distribución y variabilidad de un conjunto de datos. Son especialmente útiles para comparar la distribución de una variable cuantitativa en función de dos o más grupos. A continuación, se muestra cómo realizarlos en Excel utilizando el conjunto de datos de la submuestra del informe Pisa.

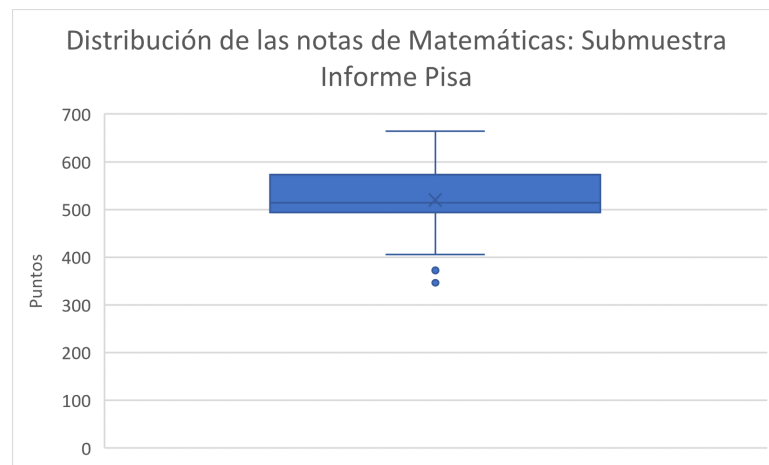
#### 6.1.1. Una variable cuantitativa

Para ilustrar la distribución de las notas en matemáticas, sigue estos pasos:

1. Selecciona la columna que contiene las puntuaciones de Matemáticas.
2. Ve a la pestaña “Insertar” y haz clic en el icono del histograma pequeño que se muestra en la imagen.



3. En las opciones, selecciona “Diagrama de caja y bigotes”. Excel generará automáticamente el diagrama de caja para las puntuaciones en matemáticas.
4. Formatea el gráfico adecuadamente: colores, títulos de los ejes, etc.



### 6.1.2. Varias variables cuantitativas la vez

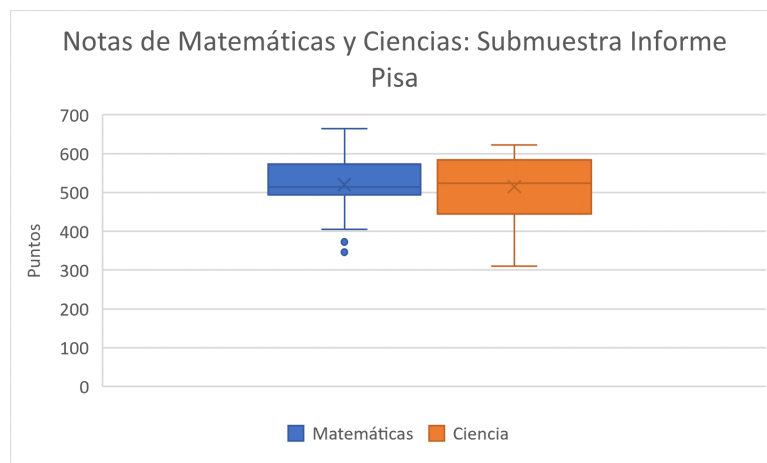
Para comparar la distribución de las notas en matemáticas y ciencias en un solo gráfico, sigue estos pasos:

1. Selecciona ambas columnas que contienen las puntuaciones de matemáticas y ciencias. *Nota: si las columnas no están una al lado de la otra, puedes seleccionar las dos columnas de interés pulsando la tecla **Ctrl** y, sin dejar de pulsarla, haciendo clic en los encabezados de cada una de las columnas de interés.*

	G	H
	Matemáticas	Ciencia
1	346,49	310,68
2	372,5	448,6

2. Sigue el mismo proceso de la subsección anterior (pestaña “Insertar” → “Gráficos” → “Diagrama de caja y bigotes”). Excel mostrará dos diagramas de caja en un mismo gráfico y en la misma escala, uno para matemáticas y otro para ciencias.

3. Formatea el gráfico adecuadamente: colores, títulos de los ejes... Es fundamental mostrar la *Leyenda* en las opciones del gráfico para conocer a qué variable pertenece cada color.



**Nota 6.1.** Este esquema, de representar distintas variables (de la misma naturaleza como son las notas del examen estandarizado) sobre la misma muestra, responde de manera a lo que se conoce como muestras emparejadas. Esto es así porque, en los mismos individuos, se han tomado distintas medidas de la misma naturaleza.

### 6.1.3. Una variable cuantitativa en función de las clases de una variable cualitativa nominal

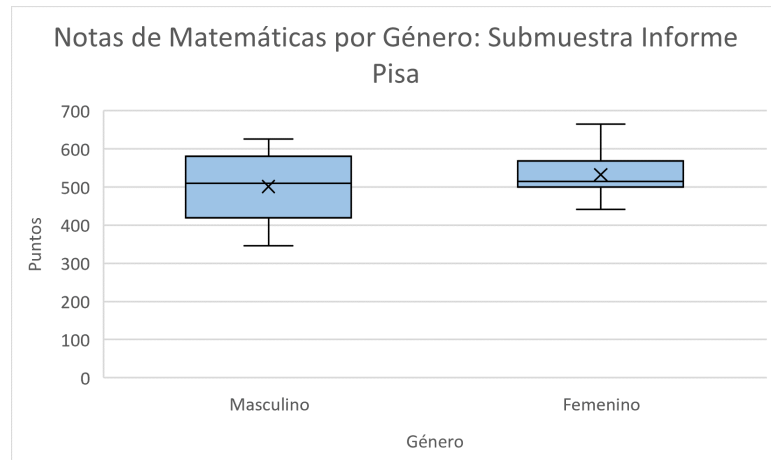
Para visualizar las diferencias en las puntuaciones de matemáticas según el género, realiza los siguientes pasos:

1. Se han de seleccionar las dos columnas donde estén las dos variables de interés, en este caso *Género* y *Matemáticas*.
  - a) Presiona la tecla *Ctrl* y mantenla pulsada.
  - b) Seleccionar los datos de la primera columna haciendo clic sobre la letra de la columna.
  - c) Repetir la operación con la segunda columna.
  - d) Soltar la tecla *Ctrl*.

B	C	D	E	F	G	H
Género	Estudios Madre	Estudios Padre	Ordenador	Coches	Matemáticas	Ciencia
Masculino	Secundaria	Secundaria	Sí	1	346,49	310,68
Masculino	Secundaria	Bac/FP	Sí	3	372,5	448,6

2. Sigue el mismo proceso de las subsecciones anteriores (pestaña “Insertar” → “Gráficos” → “Diagrama de caja y bigotes”). Excel mostrará dos diagramas de caja en un mismo gráfico, uno para las notas de cada género en la misma escala.

3. Formatea el gráfico adecuadamente: colores, títulos de los ejes, etc.

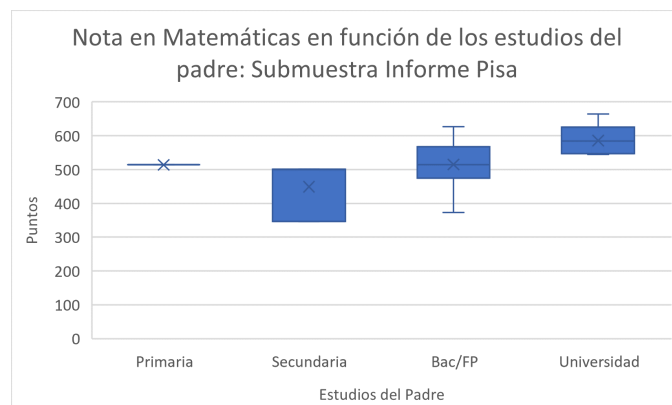


#### 6.1.4. Una variable en función de las clases de una variable cualitativa ordinal

Para visualizar las diferencias en las puntuaciones de matemáticas según los estudios del padre, que es una variable cualitativa ordinal, se deben seguir los mismos pasos que para el caso donde la variable cualitativa sea nominal, como en la Sección 6.1.3. Sin embargo, debemos establecer un orden personalizado para indicarle a Excel cómo tratar el nivel de estudios. Para ello, basta con seguir los pasos de la Sección 1.10.1.

1. Selecciona las columnas *Estudios Padre* y *Matemáticas* para obtener el diagrama de cajas deseado, el cual deberá formatearse correctamente como de costumbre.
2. Introduce un orden en la variable “nivel de estudios del padre” siguiendo los pasos de la Sección 1.10.1.

El resultado final puede verse a continuación, siendo un diagrama de cajar pero con las clases ordenadas según su orden natural.



## 6.2. Cálculo de las medidas de síntesis por grupo

No existe una manera rápida única de obtener las medidas de síntesis por grupos. A continuación, mostramos el procedimiento recomendado para calcular algunas de esas medidas.

### 6.2.1. Media, desviación típica y número de observaciones

Para calcular estas medidas de síntesis, así como el número de observaciones de cada tipo, podemos emplear las opciones de las tablas dinámicas.

1. Seleccionamos todas las filas de la tabla.

	B	C	D	E	F	G	H
	Género	Estudios Madre	Estudios Padre	Ordenador	Coches	Matemáticas	Ciencia
	Masculino	Secundaria	Secundaria	Sí	1	346,49	310,68

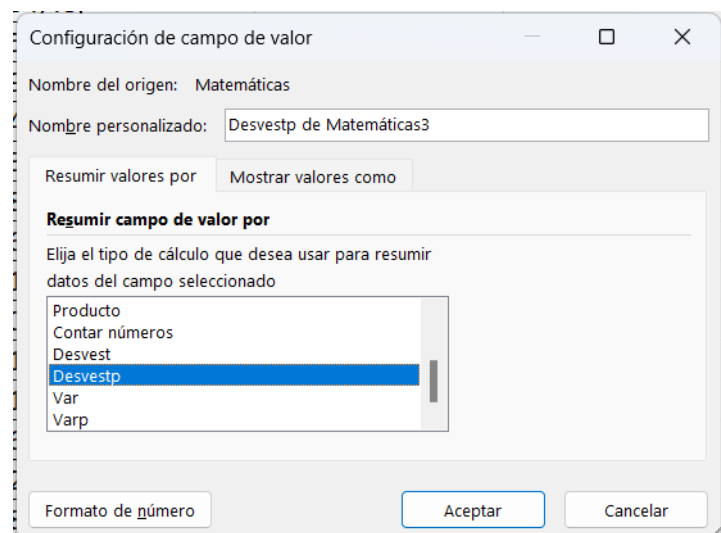
2. Insertamos una tabla dinámica.
3. Llevamos la variable cualitativa a las filas y la cuantitativa a las columnas tres veces, quedando:

Filas	Valores
Género	Suma de Matemáticas
	Suma de Matemáticas2
	Suma de Matemáticas3

4. **Número de observaciones:** Hacemos clic en el encabezado de la primera columna de la tabla dinámica y mostramos el recuento haciendo *clic derecho*, *Resumir valores por* y finalmente *Recuento*.

Etiquetas de fila	Suma de Matemáticas	Suma de Matemáticas2	Suma de Matemáticas3
Femenino	958	958	958
Masculino	601	601	601
(en blanco)			
<b>Total general</b>	<b>1560</b>	<b>1560</b>	<b>1560</b>

5. **Media:** Igual que en el paso anterior, pero en vez de seleccionar recuento, pulsamos *Promedio*.
6. **Desviación Típica:** Igual que en los dos pasos anteriores, pero pulsamos ahora *Más opciones* y seleccionamos *Desvestp*.

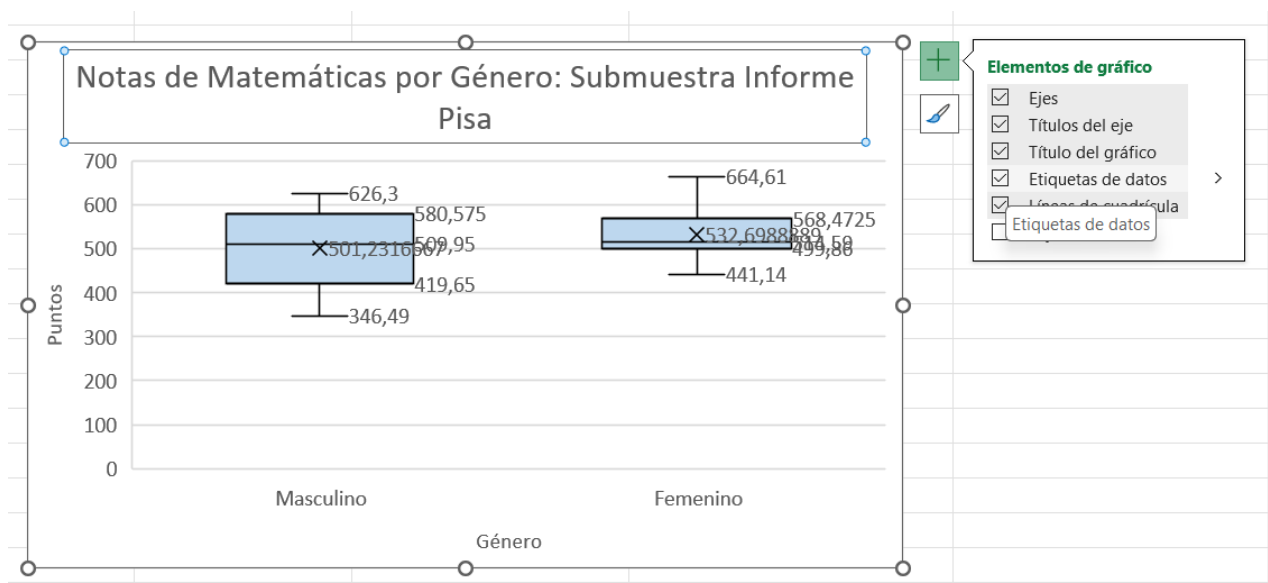


El resultado final puede expresarse con menos decimales para mayor claridad, si así lo deseamos:

Etiquetas de fila	Cuenta de Matemáticas	Promedio de Matemáticas2	Desvestp de Matemáticas3
Femenino	18	532,7	53,9
Masculino	12	501,2	86,2
<b>Total general</b>	<b>30</b>	<b>520,1</b>	<b>70,4</b>

### 6.2.2. Mínimo, Q1, Mediana, Q3, Máximo

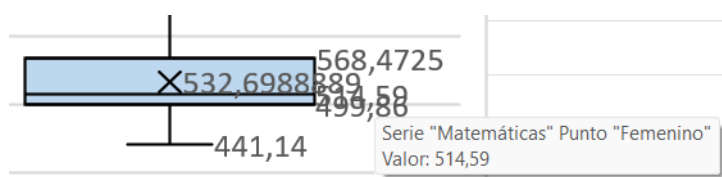
Calcular estas medidas de síntesis es sencillo usando un truco sobre el diagrama de cajas. Una vez realizado el diagrama de cajas como se explica en la Sección 6.1.3, podemos ir a las opciones del gráfico (en el símbolo  $+$  que aparece en la esquina superior derecha del gráfico) y seleccionar *Etiquetas de los datos*.



En cada una de las cajas podemos ver los resultados. Por ejemplo, para el género masculino tenemos:

- Mínimo: 346,49 puntos.
- Q1: 419,65 puntos.
- Media (hay que fijarse en la  $x$ ): 501,23 puntos.
- Mediana: 509,95 puntos.
- Q3: 580,575 puntos.
- Máximo: 626,3 puntos.

**Nota 6.2.** Si alguno de los números resulta confuso, podemos dejar el cursor sobre el valor, hasta que aparezca una caja de texto que nos informe de su valor.







# 7

## Análisis bivalente de variables cualitativas

### 7.1. A partir de microdatos

Para esta sección, utilizaremos los datos de la muestra de la Encuesta de Estructura Salarial del INE de 2018 [2]. En particular, analizaremos el nivel de estudios de los trabajadores en función de su sexo.

#### 7.1.1. Tablas de frecuencias absolutas conjuntas

1. **Seleccionar todos los datos:** Seleccionamos todas las columnas donde están los datos pulsando sobre las letras de las columnas. Al hacer esto, especificamos qué datos se usarán para la tabla de contingencia.

B	C	D	E	F	G	H	I
Edad	Estudios	Zona	Responsable	Sexo	DuracionContrato	TipoJornada	Salario
De 50 a 59	ESO	Comunida	No	Mujer	Indefinida	Parcial	7216,44
De 50 a 59	Primaria	Sur	No	Hombre	Indefinida	Completa	18504,92

2. **Insertar una Tabla Dinámica:** Ve a la pestaña *Insertar* y luego a *Tabla Dinámica*. Elige si quieres que la tabla dinámica se coloque en una nueva hoja de cálculo o en la existente, y haz clic en “Aceptar”.
3. **Construir la Tabla de Contingencia:** Arrastra el campo *Sexo* al área de *Filas* y el campo *Estudios* al área de *Columnas*. Además, arrastra cualquiera de los dos

campos al área de *Valores* para contar la cantidad de registros en cada categoría. No importa qué variable utilices para el campo de *Valores*.

Arrastrar campos entre las áreas siguientes:

<b>Filtros</b> 	<b>Columnas</b> Estudios ▼
<b>Filas</b> Sexo ▼	<b>Valores</b> Σ Cuenta de Sexo ▼

El resultado de la operación anterior debería ser el siguiente:

Cuenta de Sexo	Etiquetas de columna ▼								
Etiquetas de fila ▼	Bachillerato o grado medio	Diplomatura	ESO	Grado superior	Licenciatura y superior	Menos que primaria	Primaria (en blanco)	Total general	
Hombre	514	226	713	243	331	28	560	2615	
Mujer	517	320	541	181	426	22	378	2385	
(en blanco)									
<b>Total general</b>	<b>1031</b>	<b>546</b>	<b>1254</b>	<b>424</b>	<b>757</b>	<b>50</b>	<b>938</b>	<b>5000</b>	

4. **Orden de las clases y categorías en blanco:** Si no hay individuos en las categorías (*en blanco*), es recomendable quitar los datos de la tabla. Para ello, como en secciones anteriores, vamos al triángulo de las etiquetas de la fila/columna y desmarcamos la opción (*en blanco*).

18504,92	Cuenta de Sexo	E
51578,85	Etiquetas de fila ▼	B

A ↓ Ordenar de A a Z  
 Z ↓ Ordenar de Z a A  
 Más opciones de ordenación...

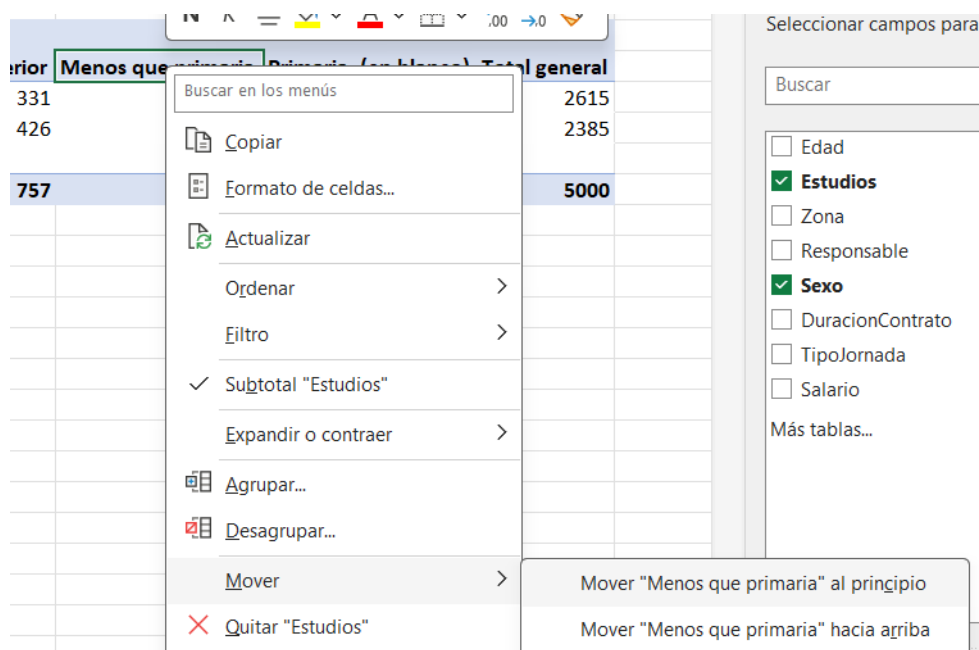
Borrar filtro de "Sexo"

Filtros de etiqueta >  
 Filtros de valor >

Buscar

☐ (Seleccionar todo)  
☒ Hombre  
☒ Mujer  
☐ (en blanco)

Las variables nominales no tienen un orden, por lo que podemos utilizar el orden que deseemos. Sin embargo, las variables ordinales deben estar ordenadas de menor a mayor clase o de peor a mejor. Igual que en secciones anteriores, hacemos clic derecho sobre la clase que queremos mover y la colocamos hacia arriba/abajo/principio/final hasta que todas las clases estén ordenadas.



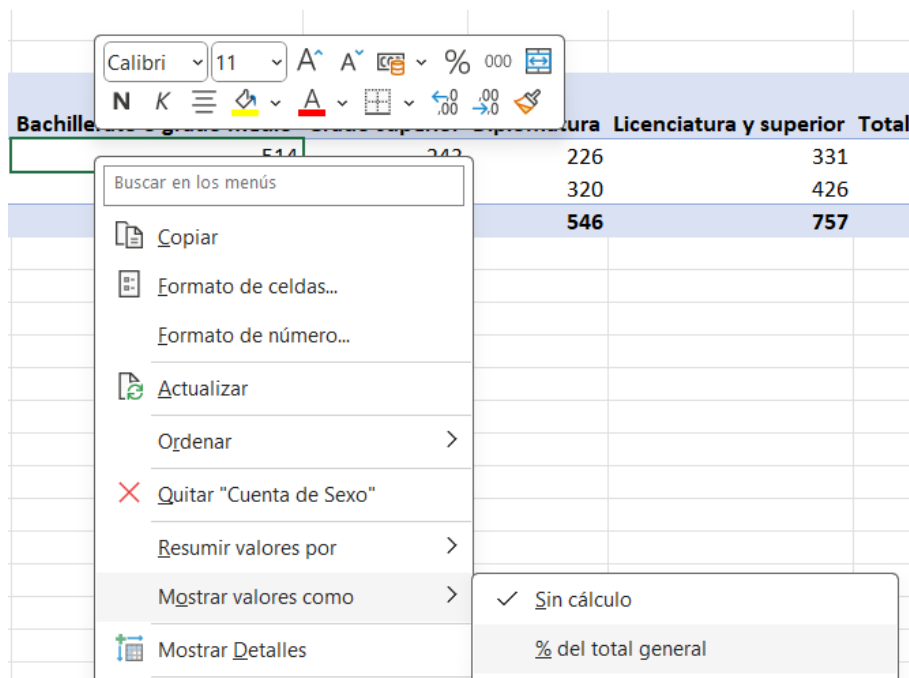
Tras realizar las operaciones anteriores, llegamos al siguiente resultado. Nótese que la variable *Sexo* es nominal y se puede representar en cualquier orden, pero la variable *Estudios* debe representarse ordenada.

Cuenta de Sexo	Etiquetas de columna	Menos que primaria	Primaria	ESO	Bachillerato o grado medio	Grado superior	Diplomatura	Licenciatura y superior	Total general
Hombre		28	560	713	514	243	226	331	2615
Mujer		22	378	541	517	181	320	426	2385
<b>Total general</b>		<b>50</b>	<b>938</b>	<b>1254</b>	<b>1031</b>	<b>424</b>	<b>546</b>	<b>757</b>	<b>5000</b>

### 7.1.2. Frecuencias porcentuales conjuntas

Para calcular la tabla de frecuencias porcentuales conjuntas, partimos de la tabla anterior de frecuencias absolutas conjuntas.

1. A partir de las frecuencias absolutas conjuntas, haz clic derecho en un número cualquiera de la tabla, y selecciona *Mostrar valores como* y después *% del total general*.



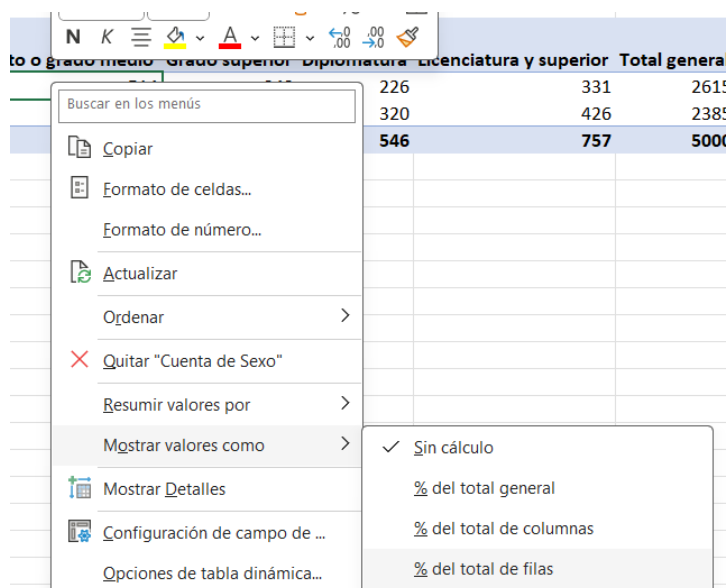
Con esta simple operación obtenemos la tabla deseada.

Cuenta de Sexo	Etiquetas de columna								
Etiquetas de fila	Menos que primaria	Primaria	ESO	Bachillerato o grado medio	Grado superior	Diplomatura	Licenciatura y superior	Total general	
Hombre	0,56%	11,20%	14,26%	10,28%	4,86%	4,52%	6,62%	52,30%	
Mujer	0,44%	7,56%	10,82%	10,34%	3,62%	6,40%	8,52%	47,70%	
<b>Total general</b>	<b>1,00%</b>	<b>18,76%</b>	<b>25,08%</b>	<b>20,62%</b>	<b>8,48%</b>	<b>10,92%</b>	<b>15,14%</b>	<b>100,00%</b>	

### 7.1.3. Frecuencias condicionadas por filas

Siguiendo las instrucciones de la sección anterior asumimos que se parte de la tabla de frecuencias absolutas conjuntas o de la de porcentajes conjuntos. Lo importante es tener una tabla dinámica con los niveles ordenados de las variables ordinales.

1. Haz clic derecho en un número cualquiera de la tabla y selecciona *Mostrar valores como* y después *% del total de filas*.



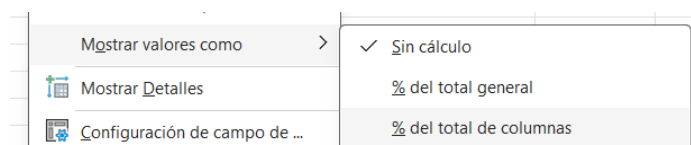
El resultado es una tabla cuya suma por filas es 100 %, tal y como se desea, y describe la distribución de la variable *Estudios* para cada posible valor de *Sexo*.

Cuenta de Sexo	Etiquetas de columna								
Etiquetas de fila	Menos que primaria	Primaria	ESO	Bachillerato o grado medio	Grado superior	Diplomatura	Licenciatura y superior	Total general	
Hombre	1,07%	21,41%	27,27%	19,66%	9,29%	8,64%	12,66%	100,00%	
Mujer	0,92%	15,85%	22,68%	21,68%	7,59%	13,42%	17,86%	100,00%	
<b>Total general</b>	<b>1,00%</b>	<b>18,76%</b>	<b>25,08%</b>	<b>20,62%</b>	<b>8,48%</b>	<b>10,92%</b>	<b>15,14%</b>	<b>100,00%</b>	

#### 7.1.4. Frecuencias condicionadas por columnas

De nuevo, partimos de una tabla dinámica con las frecuencias como en la sección anterior:

1. Haz clic derecho en un número cualquiera de la tabla, selecciona *Mostrar valores como* y después *% del total de columnas*.



El resultado es una tabla cuya suma por columnas es del 100 %, tal y como se desea, que describe la distribución de la variable *Sexo* para cada posible valor de los *Estudios*.

Cuenta de Sexo	Etiquetas de columna							
Etiquetas de fila	Menos que primaria	Primaria	ESO	Bachillerato o grado medio	Grado superior	Diplomatura	Licenciatura y superior	Total general
Hombre	56,00%	59,70%	56,86%	49,85%	57,31%	41,39%	43,73%	52,30%
Mujer	44,00%	40,30%	43,14%	50,15%	42,69%	58,61%	56,27%	47,70%
<b>Total general</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>

## 7.2. A partir de datos tabulados

Es frecuente encontrarse con datos que ya han sido tabulados. Además, es habitual que esas tablas sean de frecuencias absolutas, en las que se indica cuántos individuos hay de cada tipo.

En esta sección, trabajaremos con unos datos ficticios de género y estudios, con solo 10 individuos. Las dos variables son cualitativas, aunque una es nominal y la otra es ordinal, por lo que respetaremos el orden natural de esta última, como es habitual.

	A	B	C	D
1		Primaria	Secundaria	Universidad
2	Hombre	2	1	1
3	Mujer	0	3	3

### 7.2.1. Distribución marginal absoluta y tamaño muestral $N$

El primer paso para calcular las frecuencias restantes es hallar las distribuciones marginales. Para ello, debemos sumar los valores por filas y por columnas.

Empezaremos por las filas, aunque también es posible comenzar por las columnas.

1. Sumamos, por ejemplo utilizando el botón de autosuma, las frecuencias de la primera fila, pulsando en la casilla a la derecha de la primera fila como se observa en la imagen.

	A	B	C	D	E
1		Primaria	Secundaria	Universidad	
2	Hombre	2	1	1	=SUMA(B2:D2)
3	Mujer	0	3	3	SUMA(númer
4					

2. Generalizamos el cálculo al resto de filas arrastrando el resultado hacia abajo.

	A	B	C	D	E
1		Primaria	Secundaria	Universidad	
2	Hombre	2	1	1	4
3	Mujer	0	3	3	

3. Repetimos la operación para las columnas. Nos colocamos debajo de la primera columna y sumamos con el botón de autosuma.

	A	B	C	D	E
1		Primaria	Secundaria	Universidad	
2	Hombre	2	1	1	4
3	Mujer	0	3	3	6
4		=SUMA(B2:B3)			
5		SUMA(número1; [número2]; ...)			

4. Generalizamos el cálculo arrastrando, pero ahora hasta la fila de las sumas anteriores.

	A	B	C	D	E
1		Primaria	Secundaria	Universidad	
2	Hombre	2	1	1	4
3	Mujer	0	3	3	6
4		2			

De esta manera, podemos observar, además de las dos distribuciones marginales absolutas, el tamaño muestral  $N$ , que en este caso es 10, en la esquina inferior derecha de la tabla.

	A	B	C	D	E
1		Primaria	Secundaria	Universidad	
2	Hombre	2	1	1	4
3	Mujer	0	3	3	6
4		2	4	4	10

### 7.2.2. Frecuencias conjuntas relativas o porcentuales

Para calcular las frecuencias conjuntas relativas o porcentuales, debemos partir de la tabla de frecuencias absolutas conjuntas y el tamaño muestral, tal y como se ha mostrado en la Subsección 7.2.1.

1. Copia y pega la tabla de las frecuencias absolutas. Después, borra los números de las frecuencias para así tener el esqueleto de la tabla. El resultado final ha de ser:

8		Primaria	Secundaria	Universidad	
9	Hombre				
10	Mujer				
11					

2. Para la primera celda de la nueva tabla, divide la frecuencia absoluta de la primera tabla entre el tamaño muestral. Como siempre, queremos dividir para el mismo denominador por lo que lo inmovilizaremos pulsando **F4**. Fíjate en cómo ha aparecido un símbolo \$ antes y después de la letra de la columna en la última casilla que aparece en la fórmula a calcular.

FRECUENCIA $\times$ $\checkmark$ $f_x$ =B2/\$E\$4					
	A	B	C	D	E
1		Primaria	Secundaria	Universidad	
2	Hombre	2	1	1	4
3	Mujer	0	3	3	6
4		2	4	4	10
5					
6					
7		Primaria	Secundaria	Universidad	
8	Hombre	=B2/\$E\$4			
9	Mujer				
10					

3. Arrastra hacia la derecha para generalizar el cálculo, incluyendo la columna donde antes estaban las frecuencias marginales.

7		Primaria	Secundaria	Universidad	
8	Hombre	0,2			
9	Mujer				
10					

4. Repite ahora con el resultado de toda la primera fila, arrastrando hacia abajo, incluyendo de nuevo la fila donde antes estaban las marginales.

7		Primaria	Secundaria	Universidad	
8	Hombre	0,2	0,1	0,1	0,4
9	Mujer				
10					

5. Finalmente, selecciona todas estas frecuencias relativas y pulsa el botón % para expresarlas como porcentajes.

7		Primaria	Secundaria	Universidad	
8	Hombre	20%	10%	10%	40%
9	Mujer	0%	30%	30%	60%
10		20%	40%	40%	100%

### 7.2.3. Frecuencias condicionadas por filas

Para calcular las frecuencias condicionadas por filas, también conocidas como *perfil fila*, copiamos y pegamos un esqueleto vacío de la tabla original. Sin embargo, ahora no necesitaremos la última fila de las marginales, pero sí mostraremos la última columna que, dado que estamos condicionando a los valores de las filas, si hacemos bien el proceso, acabará mostrando un valor de 100 %.

**Nota 7.1.** En este ejemplo se ilustrará el cálculo a partir de las frecuencias absolutas conjuntas, pero el proceso es equivalente si se realiza a partir de las frecuencias porcentuales conjuntas que se han hallado en la Subsección 7.2.2.

1. Partimos de un esqueleto de la tabla sin la última fila, puesto que no la necesitaremos.



7		Primaria	Secundaria	Universidad	
8	Hombre				
9	Mujer				

2. Calculamos la primera celda dividiendo la primera frecuencia conjunta por el valor de la primera frecuencia marginal de las filas.

**Importante:** Como se va a dividir siempre por el valor de la marginal de la fila, pero este valor cambiará cuando cambiemos de fila, debemos inmovilizar solo la columna, no la fila.

Esto se realiza poniendo el símbolo \$ antes de la letra de la columna en el denominador. Un atajo para inmovilizar solo la columna consiste en pulsar exactamente tres veces la tecla F4. Otra opción es escribir manualmente el símbolo de dólar antes de la letra de la columna.

E2					
	A	B	C	D	E
1		Primaria	Secundaria	Universidad	
2	Hombre	2	1	1	4
3	Mujer	0	3	3	6
4		2	4	4	10
5					
6					
7		Primaria	Secundaria	Universidad	
8	Hombre	=B2/\$E2			
9	Mujer				

3. Ahora arrastramos hacia la derecha la primera celda.

6		Primaria	Secundaria	Universidad	
7	Hombre	0,5			
8	Mujer				

4. Finalmente arrastramos hacia abajo el resultado de la primera fila.

6		Primaria	Secundaria	Universidad	
7	Hombre	0,5	0,25	0,25	1
8	Mujer				

5. Si fuera necesario, seleccionando todas las frecuencias y pulsando la opción %, mostramos las frecuencias condicionadas ya calculadas como porcentajes.

7		Primaria	Secundaria	Universidad	
8	Hombre	50%	25%	25%	100%
9	Mujer	0%	50%	50%	100%

Observa cómo las filas suman un 100 %.

### 7.2.4. Frecuencias condicionadas por columnas

Para calcular las frecuencias condicionadas por columnas, también conocidas como *perfil columna*, copiamos y pegamos un esqueleto vacío de la tabla original. Ahora no necesitaremos la última columna de las marginales, pero sí mostraremos la última fila, que acabará mostrando un valor de 100 % en todas sus celdas.

**Nota 7.2.** De nuevo, este proceso se puede realizar indistintamente a partir de las frecuencias absolutas conjuntas o de las frecuencias porcentuales conjuntas.

1. Partimos de un esqueleto de la tabla sin la última columna donde aparecía anteriormente la distribución marginal.

7		Primaria	Secundaria	Universidad
8	Hombre			
9	Mujer			
10				

2. Calculamos la primera celda como la primera frecuencia conjunta dividida por el valor de la primera frecuencia marginal de las columnas.

**Importante:** análogamente al caso anterior pero cambiando roles de filas y columna, se quiere dividir por el valor de la frecuencia marginal de la columna, pero este valor cambiará cuando cambiemos de columna, debemos inmovilizar solo la fila, no la columna.

Esto se realiza poniendo el símbolo \$ después de la letra de la columna en el denominador. Un atajo para inmovilizar solo la fila consiste en pulsar exactamente dos veces la tecla F4. Otra opción es escribir manualmente el símbolo del dólar después de la letra de la columna.

B8				
	A	B	C	D
1		Primaria	Secundaria	Universidad
2	Hombre	2	1	1
3	Mujer	0	3	3
4		2	4	4
5				
6				
7		Primaria	Secundaria	Universidad
8	Hombre	=B2/B\$4		
9	Mujer			
10				

3. Ahora arrastramos hacia abajo la primera celda.

7		Primaria	Secundaria	Universidad
8	Hombre	1		
9	Mujer			
10				

4. Finalmente, arrastramos hacia la derecha el resultado de la primera columna.

7		Primaria	Secundaria	Universidad
8	Hombre	1		
9	Mujer	0		
10		1		

5. Si fuera necesario, seleccionamos todas las frecuencias y pulsamos la opción % para mostrar las frecuencias condicionadas ya calculadas como porcentajes.

7		Primaria	Secundaria	Universidad
8	Hombre	100%	25%	25%
9	Mujer	0%	75%	75%
10		100%	100%	100%

Observa cómo las columnas suman un 100 %.

### 7.3. Representación gráfica

Para hacer una representación, como en secciones anteriores, se recomienda copiar y pegar los valores en otro lugar. En el caso del análisis bivalente de dos variables cualitativas, se debe respetar el orden de las variables ordinales, si las hubiera.

Además, tenemos la opción de mostrar cualquier tipo de frecuencias: absolutas conjuntas, porcentuales conjuntas, condicionadas por filas o condicionadas por columnas. Todas las frecuencias se representan de igual manera en Excel, variando únicamente los datos de entrada de la tabla, que ya han sido calculados en las secciones anteriores.

Por lo general, se desean representar las frecuencias conjuntas cuando solo queremos describir cómo son los individuos, prefiriendo las absolutas cuando se trabaja con un censo y las porcentuales cuando se trabaja con una muestra. Esta representación, sin embargo, si las clases son desiguales en cuanto al número de individuos, no permitirá ver las posibles dependencias estadísticas que puedan existir entre las dos variables.

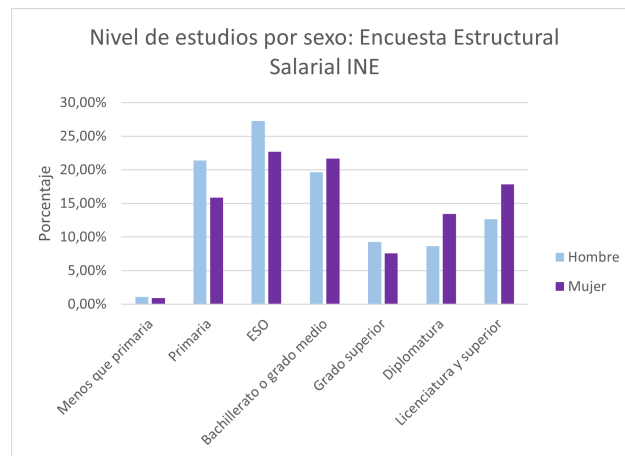
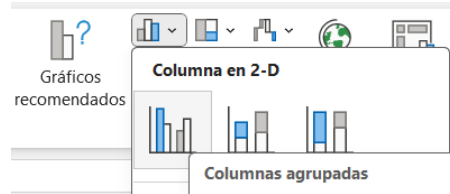
Por lo tanto, la práctica habitual es representar las frecuencias de la variable explicada condicionadas por los valores de la variable explicativa.

En el caso del nivel de estudios de los trabajadores en función del sexo, es obvio que la variable explicativa es el sexo y la variable explicada es el nivel de estudios. Por ello, dado que teníamos el sexo en las filas, representaremos las frecuencias condicionadas a las filas.

1. **Partir de la tabla de frecuencias deseada:** Como se ha argumentado, partimos de la tabla copiada en la que se representan las frecuencias condicionadas a las filas (en este caso, se pueden representar otras frecuencias). Es importante notar que no se utiliza la columna del total, ya que si lo hacemos, representará una columna con el 100 % que alterará la escala del gráfico y entorpecerá la visualización.

Etiquetas de fila	Menos que primaria	Primaria	ESO	Bachillerato o grado medio	Grado superior	Diplomatura	Licenciatura y superior
Hombre	1,07%	21,41%	27,27%	19,66%	9,29%	8,64%	12,66%
Mujer	0,92%	15,85%	22,68%	21,68%	7,59%	13,42%	17,86%

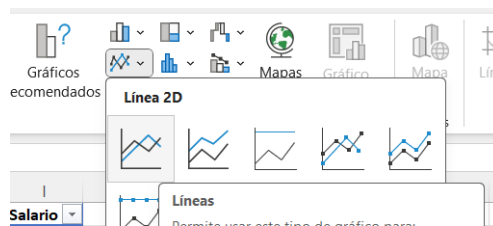
2. **A. Insertar un gráfico de barras:** Igual que en secciones anteriores. Se debe formatear con colores, nombres de los ejes y título como es habitual. En este caso será necesario incluir una leyenda en el gráfico, aunque esta aparecerá por defecto.



**Nota 7.3.** A veces deseamos mostrar, no todas las clases, sino solo una de ellas. Por ejemplo, en el nivel de estudios por sexo, podría haberse representado únicamente el porcentaje de hombres y de mujeres que tienen estudios de *Licenciatura y superior* mediante un diagrama de barras con solo esas dos barras. Este tipo de representación facilitaría la transmisión de información, al simplificar el gráfico, si nuestro objetivo fuera ver únicamente como son estas personas en cuanto a su acceso a dichos estudios.

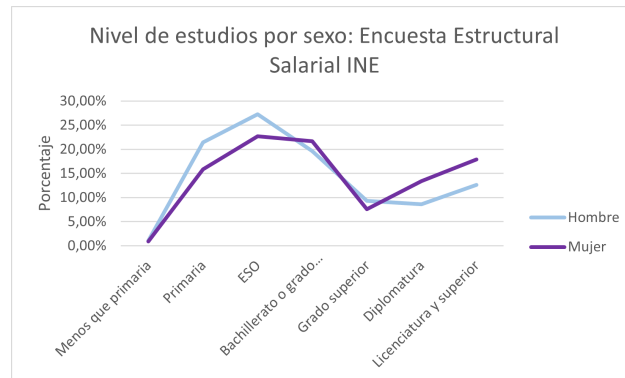
2. **B. Líneas 2D como gráfico alternativo cuando la variable explicada es ordinal:** En este caso, también se puede representar un *Gráfico de Líneas 2D*.

**Importante:** Es absolutamente necesario seleccionar el primer gráfico 2D que ofrece Excel, en el que las dos líneas se cruzan. El gráfico de la derecha puede parecer similar, pero no representa exactamente lo que buscamos, que son las frecuencias ya calculadas.



El resultado es muy intuitivo, ya que el orden de las variables se refleja bien con las líneas continuas del gráfico. Observa cómo es fácil identificar la (ligera) tendencia

de que las mujeres, en general, tienen más estudios superiores y menos primarios y secundarios que los hombres. En otras palabras, tienden a alcanzar un nivel de estudios mayor.





# 8

## Análisis bivalente de dos variables cuantitativas

El análisis de la relación entre dos variables cuantitativas es una de las ramas más extensas de la Estadística. En esta sección, se ilustrarán las técnicas básicas con Excel.

Para ello, utilizaremos un conjunto de datos que representa la temperatura de la superficie del mar (en grados Celsius) y el crecimiento del coral (en milímetros anuales) en localizaciones del Golfo de México y el Mar Caribe a lo largo de varios años [7].

B	C
Temperatura	Crecimiento
26,7	8,5
26,6	8,5
26,6	7,9
26,5	8,6
26,3	8,9
26,1	9,2

### 8.1. Diagrama de dispersión

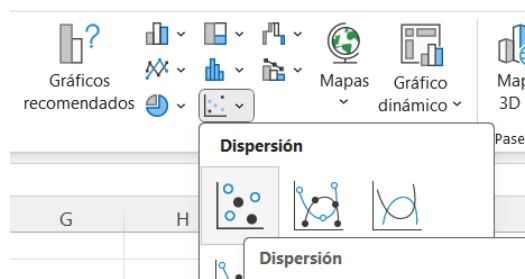
En el escenario de dos variables cuantitativas, siempre es necesario realizar una representación gráfica, siendo la más habitual el diagrama de dispersión, también conocido como nube de puntos.

1. Seleccionamos las dos columnas donde se encuentran las dos variables.

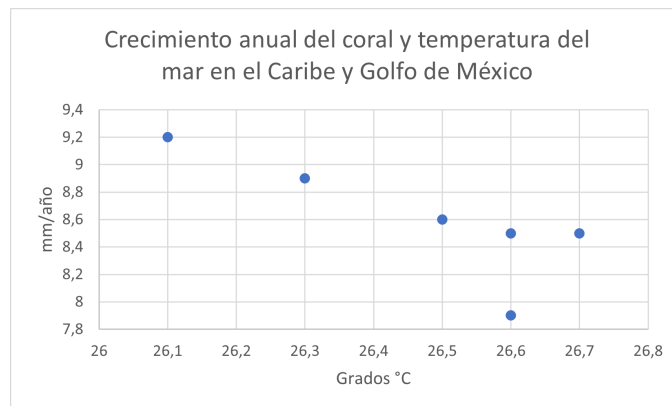
B	C
Temperatura	Crecimiento
26,7	8,5
26,6	8,5

**Nota 8.1.** La variable que queramos representar en el eje X será la que asuma el rol de variable explicativa, mientras que la que se represente en el eje Y será la variable explicada. Es fundamental respetar este orden. Excel siempre representa en el eje X la variable que está a la izquierda. Si quisiéramos representar en el eje X la otra variable, podemos hacerlo editando el diagrama, aunque es una opción tediosa. La recomendación es crear una nueva hoja de Excel donde copiemos y peguemos las columnas, procurando que la variable explicativa esté situada a la izquierda de la variable explicada.

2. Insertamos un gráfico de dispersión que represente solo los puntos.



3. Formateamos el gráfico: colores, título, nombre de los ejes...



## 8.2. Covarianza y Correlación

Para calcular la **covarianza** se debe utilizar la función *COVARIANCE.P*.

1. En una celda vacía, escribimos = seguido del nombre de la función *COVARIANCE.P*, y para el primer argumento, después del paréntesis, seleccionamos la primera variable (no importa el orden).



B	C	D	E	F
Temperatura	Crecimiento			
26,7	8,5			
26,6	8,5	Covarianza	=COVARIANCE.P(B:B	
26,6	7,9		COVARIANCE.P(matriz1; ma	
26,5	8,6			
26,3	8,9			
26,1	9,2			

2. Para separar el primer argumento de la función del segundo, escribimos ; y seleccionamos la columna donde se encuentran los datos de la segunda variable.

=COVARIANCE.P(B:B;C:C)				
B	C	D	E	F
Temperatura	Crecimiento			
26,7	8,5			
26,6	8,5	Covarianza	=C	
26,6	7,9			
26,5	8,6			
26,3	8,9			
26,1	9,2			

3. Tras pulsar *Enter*, se obtiene el resultado de la **covarianza**.

Covarianza	-0,06666667
------------	-------------

El cálculo del **coeficiente de correlación** se realiza de manera análoga al de la covarianza, pero utilizando la función de Excel *COEF.DE.CORREL*.

1. Escribimos en una celda vacía, tras un signo =, la función *COEF.DE.CORREL*, y seleccionamos la primera columna como primer argumento.

B	C	D	E	F
Temperatura	Crecimiento			
26,7	8,5			
26,6	8,5			
26,6	7,9			
26,5	8,6			
26,3	8,9			
26,1	9,2			

2. Separamos ahora de nuevo usando ; para proceder a seleccionar la segunda columna de datos.

=COEF.DE.CORREL(B:B;C:C)				
B	C	D	E	F
Temperatura	Crecimiento			
26,7	8,5			
26,6	8,5			
26,6	7,9			
26,5	8,6			
26,3	8,9			
26,1	9,2			

3. Pulsamos la tecla *Enter* y obtenemos el resultado.

B	C	D	E
Temperatura	Crecimiento		
26,7	8,5		
26,6	8,5	r	-0,81110711
26,6	7,9		
26,5	8,6		
26,3	8,9		
26,1	9,2		

### 8.3. Estimación del modelo lineal: ajuste de mínimos cuadrados

Estimar los coeficientes de la recta de mínimos cuadrados es una tarea fácilmente realizable con Excel. En las siguientes subsecciones se ilustra cómo estimar el modelo, su bondad de ajuste, representación y utilización en la predicción.

#### 8.3.1. Estimación de la recta

Para estimar la pendiente y el término independiente de la recta de mínimos cuadrados, tan solo hemos de aplicar las funciones *PENDIENTE* e *INTERSECCION.EJE* adecuadamente.

Para la **pendiente**, en una celda en blanco, realizamos los siguientes pasos:

1. Tras un signo =, escribimos la función *PENDIENTE*, y después del primer paréntesis seleccionamos la columna donde se halla la variable Y, es decir, la variable explicada.

B	C	D	E
Temperatura	Crecimiento		
26,7	8,5	Pendiente	=pendiente(C:C
26,6	8,5	T. Independiente	PENDIENTE(con
26,6	7,9		
26,5	8,6		
26,3	8,9		
26,1	9,2		

2. Tras separar con ; seleccionamos la otra variable, la X, la que juega el papel de variable explicativa.

✖	✓	$f_x$	=pendiente(C:C;B:B
B	C	D	E
Temperatura	Crecimiento		
26,7	8,5	Pendiente	=pendiente(C:C;B:B
26,6	8,5	T. Independiente	
26,6	7,9		
26,5	8,6		
26,3	8,9		
26,1	9,2		

3. Pulsamos *Enter* y obtenemos el resultado.

Para el cálculo del **término independiente**, se siguen los mismos pasos que para la pendiente, pero usando la función *INTERSECCION.EJE* en una celda en blanco:

1. Tras un signo =, escribimos la función *INTERSECCION.EJE*, y después del primer paréntesis seleccionamos la columna donde se halla la variable Y, es decir, la variable explicada.

B	C	D	E	F
Temperatura	Crecimiento			
26,7	8,5	Pendiente	-1,57894737	
26,6	8,5	T. Independiente	=INTERSECCION.EJE(C:C)	
26,6	7,9			
26,5	8,6			
26,3	8,9			
26,1	9,2			

2. Tras separar con ; seleccionamos la otra variable, la X, la que juega el papel de variable explicativa.

B	C	D	E	F	G
Temperatura	Crecimiento				
26,7	8,5	Pendiente	-1,57894737		
26,6	8,5	T. Independiente	=INTERSECCION.EJE(C:C;B:B)		
26,6	7,9				
26,5	8,6				
26,3	8,9				
26,1	9,2				

3. Pulsamos *Enter* y obtenemos el resultado.

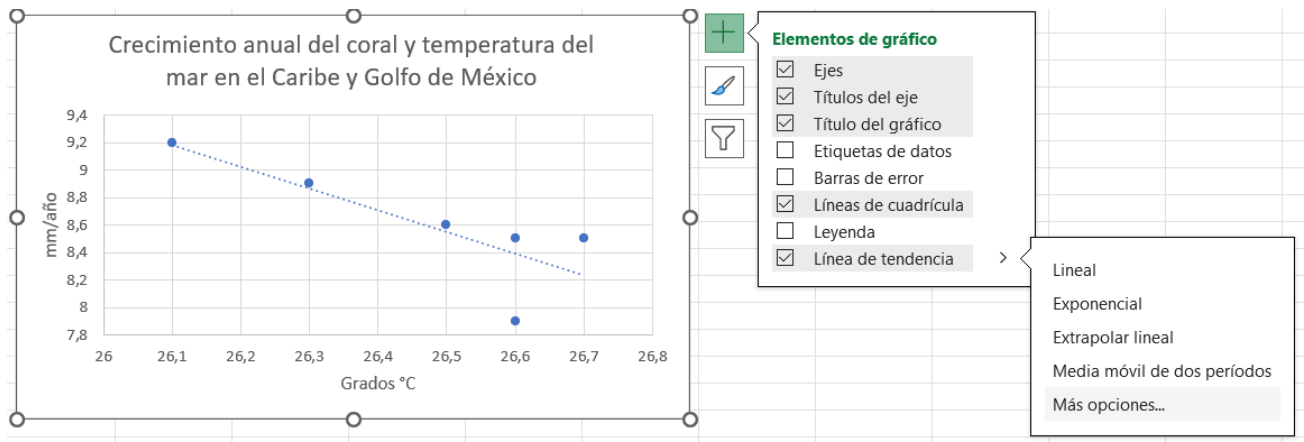
Pendiente	-1,57894737
T. Independiente	50,3894737

### 8.3.2. Representación de la línea de tendencia

Muchas veces se sobreimprime la línea de tendencia ajustada mediante mínimos cuadrados al diagrama de dispersión. A veces se hace para mostrar la posible dependencia estadística de los datos, aunque en otras ocasiones también se realiza para remarcar la ausencia de dicha relación.

Para ello, a partir del diagrama de dispersión calculado en la Subsección 8.1 procedemos con los siguientes pasos:

- Se inserta la Línea de tendencia en las opciones del eje que se despliegan en el icono +.



- En “Más opciones” (ver imagen anterior), podemos modificar la línea de tendencia. Además, podemos señalar *Presentar ecuación en el gráfico* y *Presentar el valor  $R$  cuadrado en el gráfico*, para sobreimprimir sobre el gráfico el modelo y el coeficiente de bondad de ajuste que se calculan en las Subsecciones 8.3.1 y 8.3.3.

Formato de línea d... ▾ ×

Opciones de línea de tendencia ▾

Formato de línea de tendencia

▼ Opciones de línea de tendencia

Exponencial ☐

Lineal ☒

Logarítmica ☐

Polinómica ☐ Grado 2

Potencial ☐

Media móvil ☐ Período 2

Nombre de la línea de tendencia

Automático ☒ Lineal (Crecimien

Personalizado ☐

Extrapolar

En el futuro 0,0 período

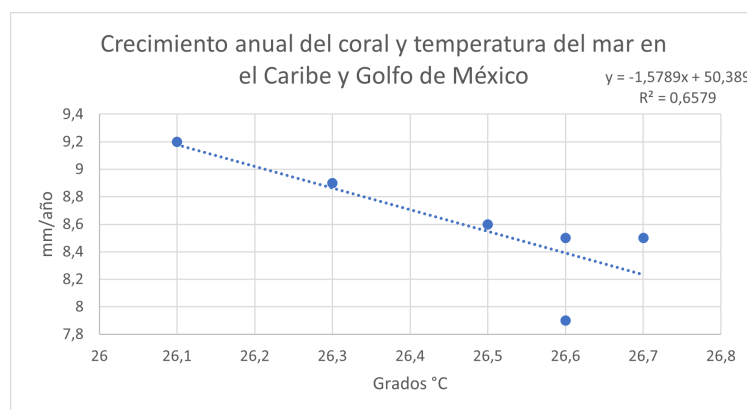
En el pasado 0,0 período

Señalar intersección 0,0

Presentar ecuación en el gráfico ☐

Presentar el valor  $R$  cuadrado en el gráfico ☐

- El resultado es el siguiente, donde por supuesto se pueden modificar las opciones estéticas como de costumbre.



### 8.3.3. Bondad de ajuste: $R^2$

1. Tras un signo =, escribimos la función *COEFICIENTE.R2*, y después del primer paréntesis seleccionamos la columna donde se encuentra la variable Y, es decir, la variable explicada (aunque en este caso, el resultado no varía si se altera el orden).

B	C	D	E	F	G
Temperatura	Crecimiento				
26,7	8,5				
26,6	8,5				
26,6	7,9				
26,5	8,6				
26,3	8,9				
26,1	9,2				

r -0,81110711  
R<sup>2</sup> =COEFICIENTE.R2(C:C

COEFICIENTE.R2(conocido y, conocido x)

CAMBIAR  
CANTIDAD.RECIBIDA  
CARACTER

2. Tras separar con ; seleccionamos la otra variable, la X, la que juega el papel de variable explicativa.

B	C	D	E	F
Temperatura	Crecimiento			
26,7	8,5			
26,6	8,5			
26,6	7,9			
26,5	8,6			
26,3	8,9			
26,1	9,2			

r -0,81110711  
R<sup>2</sup> =COEFICIENTE.R2(C:C;B:B)

3. Pulsamos *Enter* y obtenemos el resultado.

B	C	D	E
Temperatura	Crecimiento		
26,7	8,5		
26,6	8,5		
26,6	7,9		
26,5	8,6		
26,3	8,9		
26,1	9,2		

r -0,81110711  
R<sup>2</sup> 0,65789474

### 8.3.4. Predicción

Para realizar una predicción podemos aplicar la ecuación del modelo en una calculadora o en una hoja de Excel directamente.

Sin embargo, ya que tenemos calculados los coeficientes del modelo, es frecuentemente más fácil aplicar la ecuación del modelo haciendo clic en los coeficientes calculados.

A modo de ejemplo, supongamos que queremos predecir el crecimiento del coral en un año en el que la temperatura de la superficie del mar ha sido de 27 grados Celsius.

La predicción será  $50,389 - 1,578 \times 27 = 7,76$ , puesto que aplicamos la ecuación de la recta que se ha obtenido en secciones anteriores sustituyendo la  $x$  por el valor 27.

Para aplicar esto en Excel se realizan los siguientes pasos:

1. En una celda en blanco escribimos 27, que es el valor para el cual queremos hacer la predicción.
2. En otra celda en blanco, tras un signo =, pulsamos primero la celda del término independiente, después + para sumar, hacemos clic sobre la celda de la pendiente y finalmente \* para multiplicar por la celda donde se encuentra el valor 27.

D	E	F
Pendiente	-1,57894737	
T. Independiente	50,3894737	
Predicción	27	grados
	=E3+E2*E5	

3. El resultado en este caso es 7,76.

Pendiente	-1,57894737
T. Independiente	50,3894737
Predicción	27 grados
	7,75789474

## Anexo: Datos Pisa

El Programa Internacional para la Evaluación de Estudiantes, o **Informe PISA**, llevado a cabo por la OCDE, evalúa las competencias de los alumnos de 15 años. Para ello, se seleccionan institutos de secundaria en los que los estudiantes son examinados en diversas disciplinas. En particular, en la evaluación de 2018 participaron 600.000 estudiantes, representando a cerca de 32 millones de alumnos de 15 años de centros educativos de 79 países y economías.

A continuación, se presentan datos seleccionados al azar de 30 estudiantes españoles, donde se recoge su género, los estudios de sus padres y madres, el acceso a un ordenador en casa, el número de coches que posee su núcleo familiar y la puntuación en el test normalizado de PISA en las áreas de Matemáticas y Ciencias.

Las puntuaciones de los test de PISA se diseñan para proporcionar una medida estandarizada del rendimiento académico de los estudiantes en las distintas áreas evaluadas: Matemáticas, Ciencias y Lectura.

Estas puntuaciones se reportan en una escala continua que generalmente varía entre 0 y 1000 puntos. Sin embargo, en la práctica, las puntuaciones suelen estar en el rango de 300 a 700 puntos.

Estos datos están disponibles en abierto en el enlace de la referencia [1].

	Género	Estudios Madre	Estudios Padre	Ordenador	Coches	Matemáticas	Ciencia
1	Masculino	Secundaria	Secundaria	Sí	1	346.49	310.68
2	Masculino	Secundaria	Bac/FP	Sí	3	372.50	448.60
3	Masculino	Universidad	Bac/FP	Sí	1	405.25	431.24
4	Femenino	Secundaria	Bac/FP	Sí	2	441.14	444.65
5	Femenino	Bac/FP	Bac/FP	No	3	451.53	439.13
6	Masculino	Bac/FP	Bac/FP	Sí	1	462.85	382.89
7	Masculino	Bac/FP	Bac/FP	Sí	2	486.82	566.59
8	Femenino	Bac/FP	Bac/FP	Sí	1	496.72	520.86
9	Femenino	Bac/FP	Bac/FP	Sí	2	499.14	412.93
10	Femenino	Secundaria	Secundaria	Sí	0	500.10	504.48
11	Femenino	Universidad	Secundaria	Sí	1	500.57	574.30
12	Masculino	Bac/FP	Bac/FP	Sí	2	501.05	599.07
13	Femenino	Bac/FP	Bac/FP	Sí	2	506.35	445.49
14	Femenino	Bac/FP	Bac/FP	Sí	2	513.95	437.41
15	Femenino	Bac/FP	Primaria	Sí	2	514.02	487.41
16	Femenino	Bac/FP	Bac/FP	Sí	2	515.16	462.54
17	Masculino	Bac/FP	Bac/FP	Sí	2	518.85	519.37
18	Femenino	Bac/FP	Bac/FP	Sí	1	543.83	608.26
19	Femenino	Bac/FP	Universidad	Sí	1	544.32	513.72
20	Femenino	Bac/FP	Universidad	Sí	2	548.63	619.05
21	Masculino	Bac/FP	Bac/FP	Sí	2	552.42	622.90
22	Femenino	Bac/FP	Bac/FP	Sí	2	564.42	527.11
23	Masculino	Bac/FP	Bac/FP	Sí	2	570.51	588.83
24	Femenino	Bac/FP	Bac/FP	Sí	1	580.63	553.36
25	Masculino	Universidad	Universidad	Sí	2	583.93	543.95
26	Femenino	Bac/FP	Universidad	Sí	2	586.72	586.45
27	Masculino	Bac/FP	Bac/FP	Sí	1	587.81	583.71
28	Femenino	Bac/FP	Bac/FP	Sí	1	616.74	579.95
29	Masculino	Bac/FP	Bac/FP	Sí	2	626.30	583.36
30	Femenino	Bac/FP	Universidad	Sí	1	664.61	553.09



## Bibliografía

- [1] Álvarez Tena, B., Anglada Salvanés, S., Asín, J., Badía, F. G., Berrade, M. D., Castillo Mateo, J., Castillo Salazar, I. P., Galé C., Gracia Tabuenca, Z., Hernández González, A., Iranzo, J. A., Isidoro Ramírez, D., Lafuente Blasco, M., Lahoz Arnedo, D., Plo Alastrué, B. F., Sánchez-Valverde García, M. B., Sangüesa, C., & Tejel Altarriba, F. J. (2025). *Prácticas de estadística básica con software específico y de propósito general* [Curso abierto]. Universidad de Zaragoza. <https://ocw.unizar.es>
- [2] Instituto Nacional de Estadística. (2020). *Encuesta de Estructura Salarial (EES). Año 2018*. <https://www.ine.es>
- [3] Instituto Nacional de Estadística. (2019). *Estadística de Nulidades, Separaciones y Divorcios (ENSD). Año 2020*. <https://www.ine.es>
- [4] Instituto Nacional de Estadística. (2023). *Movimiento natural de la población: Nacimientos en España 2022*. <https://www.ine.es>
- [5] Lafuente Blasco, M. (2025). *Estadística Aplicada a la Investigación Social*. Prensas de la Universidad de Zaragoza. ISBN: 978-84-1340-977-1.
- [6] Microsoft Corporation. (2024). *Microsoft 365: Excel* [Software]. <https://www.microsoft.com>
- [7] Moore, D. S., Notz, W. I., & Fligner, M. A. (2021). *The basic practice of statistics* (9.<sup>a</sup> ed.). Macmillan Learning.
- [8] OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. PISA, OECD Publishing. <https://doi.org/10.1787/b25efab8-en>