

Clase 3. Estadística con Excel

Introducción

- La **estadística** es la rama de las matemáticas que se dedica al análisis e interpretación de series de datos, generando unos resultados que se utilizan básicamente en dos contextos: la toma de decisiones y la proyección de situaciones futuras.
- Tradicionalmente la estadística se ha dividido en dos ramas diferentes:
 - la **estadística descriptiva** y,
 - la **inferencia estadística**.
- La **estadística descriptiva** sirve para recoger, analizar e interpretar los datos. Mediante la **inferencia estadística** se intenta determinar una situación futura basándose en información pasada.

Variables, muestras y tablas de datos

- La parte más conocida de la estadística es aquella en la que se **estudian una o más características** de una cierta **población**, generando una tabla de datos sobre la que se realizan cálculos para obtener diversas medidas. De esta forma, se obtiene por ejemplo la altura media de los alumnos de una clase.
- Una **hoja de cálculo** es una de las herramientas más adecuadas para introducir tablas de valores y obteniendo resultados y efectuando representaciones **gráficas** que faciliten su representación.

Población, muestras y variables

- Se llama **población** al conjunto de los individuos sobre los que se lleva a cabo un estudio estadístico.
- Los individuos de una **población** no tienen que ser necesariamente personas, pueden ser un conjunto de personas, o de objetos, o de medidas... que puede ser muy grande, infinita, cambiante con el tiempo...
- Cuando la población es muy grande, se suele elegir para el estudio estadístico una parte de la misma.
- Se llama **muestra** a una parte de la población elegida mediante algún criterio.

Tipos de variables

- Dependiendo de cómo sea la característica que vamos a estudiar podemos encontrarnos con dos tipos distintos de variables estadísticas:
- **Variables cualitativas**, si los valores de la variable no se pueden medir, por ejemplo sexo, estado civil, nivel de estudios, color de ojos,...
- **Variable cuantitativas**, si los valores se pueden medir, por ejemplo, altura, edad, peso,...
- A su vez las variables cuantitativas pueden ser:
- **discretas**, si los valores que toma la variable son aislados, por ejemplo edad, número de hermanos,...
- **continua**, si la variable puede tomar todos los valores de un intervalo, por ejemplo peso, altura, temperatura,...

Tablas estadísticas

- Una vez determinada la población, las características que quieren analizarse y seleccionada la muestra, llega el momento de **recoger los datos** y de **organizarlos en tablas**.
- Las **tablas de frecuencias** resumen numéricamente, la información sobre el carácter estadístico que queremos estudiar.
- Antes de construir una tabla de frecuencias, vamos a definir los elementos que suelen aparecer en ella:

...Tablas estadísticas

La **frecuencia absoluta** f_i , de un valor x_i es el número de veces que se repite dicho valor.

La **frecuencia relativa** h_i del valor x_i es el cociente entre la frecuencia absoluta del x_i y el número total de valores, N .

$$h_i = \frac{f_i}{N}$$

La **frecuencia absoluta acumulada** F_i del valor x_i , es la suma de todas las frecuencias absolutas de todos los valores anteriores a x_i , más la frecuencia absoluta de x_i .

$$F_i = f_1 + f_2 + \dots + f_i$$

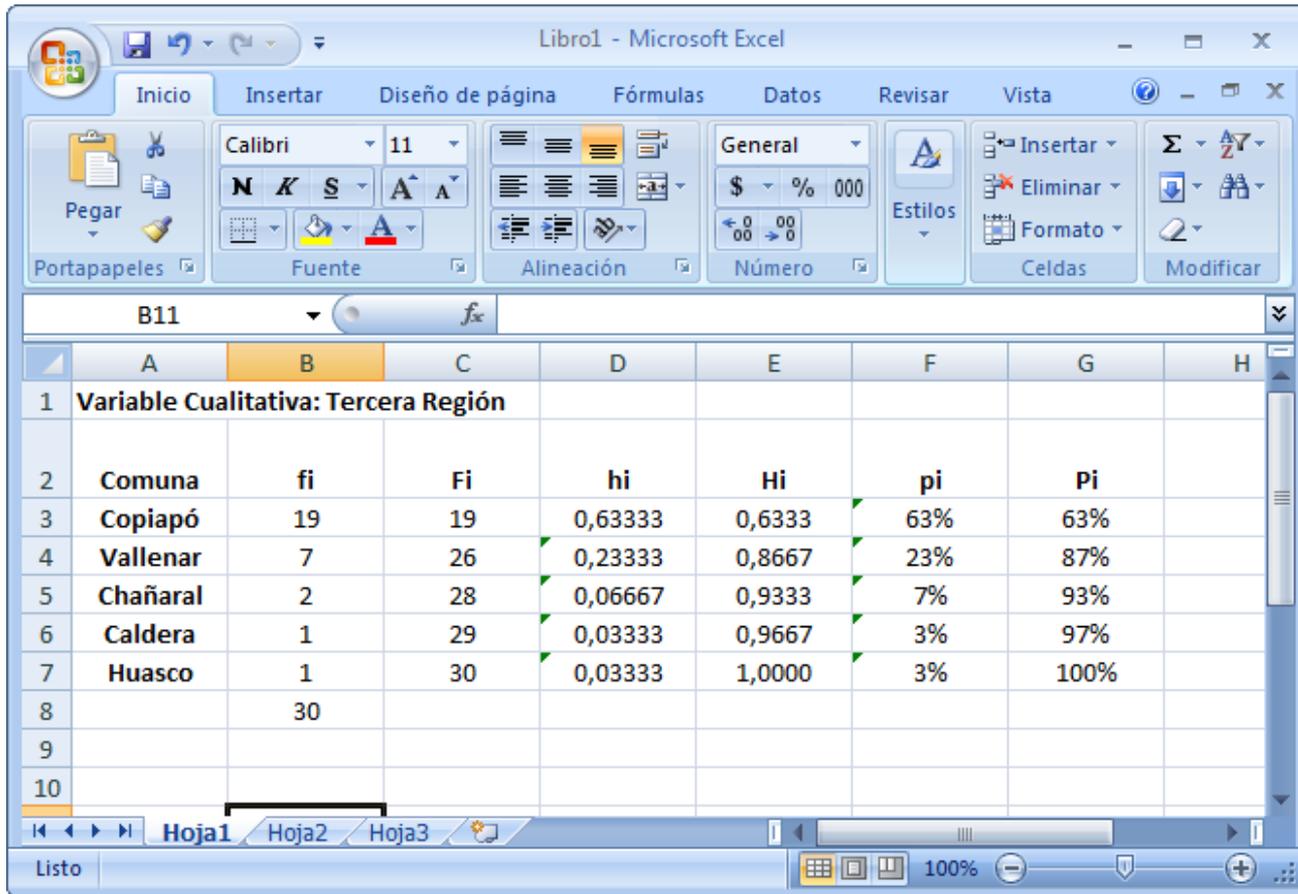
La **frecuencia relativa acumulada** H_i del valor x_i es la suma de todas las frecuencias relativas de todos los valores anteriores a x_i , más la frecuencia relativa de x_i .

$$H_i = h_1 + h_2 + \dots + h_i$$

El **porcentaje** p_i de un valor x_i se obtiene multiplicando por 100 la frecuencia relativa del valor x_i .

...Tablas estadísticas

- Así es como se construye con Excel una tabla de frecuencias:
- Introducimos en la primera columna (**A**) las distintas modalidades si el carácter es cualitativo:



The screenshot shows the Microsoft Excel interface with a frequency table for a qualitative variable. The table is located in the following cells:

	A	B	C	D	E	F	G	H
1	Variable Cualitativa: Tercera Región							
2	Comuna	fi	Fi	hi	Hi	pi	Pi	
3	Copiapó	19	19	0,63333	0,6333	63%	63%	
4	Vallenar	7	26	0,23333	0,8667	23%	87%	
5	Chañaral	2	28	0,06667	0,9333	7%	93%	
6	Caldera	1	29	0,03333	0,9667	3%	97%	
7	Huasco	1	30	0,03333	1,0000	3%	100%	
8		30						
9								
10								

...Tablas estadísticas

- o bien, los valores de la variable estadística discreta:

The screenshot shows the Microsoft Excel interface with a spreadsheet titled 'Libro1 - Microsoft Excel'. The ribbon is set to 'Inicio'. The spreadsheet contains the following data:

	A	B	C	D	E	F	G	H
1	Variable cuantitativa discreta: número de hijos							
2	Hijos	fi	Fi	hi	Hi	pi	Pi	
3	0	31						
4	1	54						
5	2	27						
6	3	5						
7	4	2						
8	5	1						
9		120						
10								
11								

...Tablas estadísticas

- En la segunda columna (**B**) introducimos los valores de la **frecuencia absoluta fi**.
- En la tercera columna (**C**) vamos a colocar la **frecuencia absoluta acumulada** (F_i), pero en lugar de hacer nosotros los cálculos, será el programa el que se encargue de hacerlos.
- **¿Cómo?**
- En la celda C3 escribimos **= B3** y en la celda C4 escribimos **=C3 + B4**. A continuación copiamos la fórmula, situando el puntero del ratón en la esquina inferior derecha de esta celda y cuando el puntero del ratón se convierta en **+** y arrastramos hasta la casilla última casilla.



...Tablas estadísticas

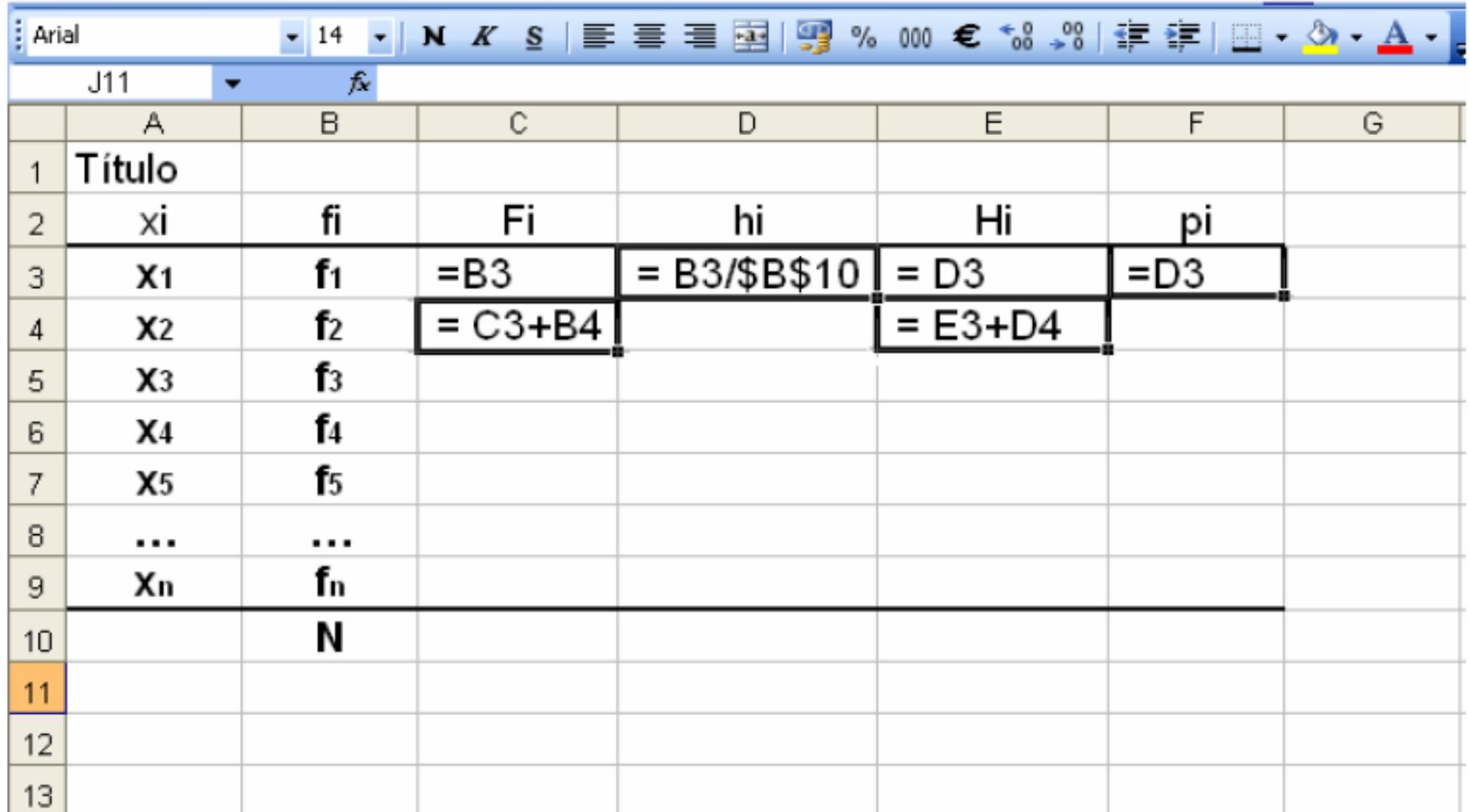
- Para completar la columna de la **frecuencia relativa (hi)**, basta con escribir en la celda **D3=B3/\$B\$8**. (Con el símbolo \$, lo que hacemos es fijar el valor de la celda que no varía).
- En la columna de la **frecuencia relativa acumulada (Hi)**, en E3, escribimos **=D3**; en E4, **=E3 + D4** y copiamos la fórmula.
- Para el porcentaje, en F3, se escribe **=D3** y pulsamos el botón .



- El paso siguiente es copiar la expresión de la celda anterior.

...Tablas estadísticas

- En resumen, la tabla de frecuencias se construye así:



The image shows a screenshot of an Excel spreadsheet with a frequency table. The spreadsheet has a toolbar at the top with various icons and a formula bar. The table is structured as follows:

	A	B	C	D	E	F	G
1	Título						
2	x_i	f_i	F_i	h_i	H_i	p_i	
3	X1	f ₁	=B3	= B3/\$B\$10	= D3	=D3	
4	X2	f ₂	= C3+B4		= E3+D4		
5	X3	f ₃					
6	X4	f ₄					
7	X5	f ₅					
8					
9	X _n	f _n					
10		N					
11							
12							
13							

...Tablas estadísticas

- Veamos ahora cómo podemos construir una tabla de frecuencias cuando la variable es continua o está agrupada en intervalos o clases.
- En la primera columna (**A**) escribimos los **intervalos o clases [a, b)**, en la columna **B** el extremo de la **izquierda a** y en la **C** el extremo de la **derecha b**. En la columna **D**, vamos a calcular la marca de clase, escribimos la fórmula $=\frac{b^3+c^3}{2}$ y la copiamos.
- La primera columna, la utilizamos para la representación gráfica y las dos siguientes **B y C**, para calcular la marca de clase.
- En la siguiente columna **E**, introducimos la **frecuencia absoluta (fi)**, en la siguiente introducimos la fórmula para el cálculo de la **frecuencia absoluta acumulada** de forma análoga a los ejemplos anteriores y así sucesivamente hasta terminar de construir la tabla.

...Tablas estadísticas

- El resultado debe ser algo así:

The screenshot shows a Microsoft Excel spreadsheet titled "Microsoft Excel - gráficos estadísticos". The menu bar includes "Archivo", "Edición", "Ver", "Insertar", "Formato", "Herramientas", "Datos", and "Ventana". The toolbar contains various icons for file operations, editing, and formatting. The active cell is B12, and the formula bar shows a function symbol (fx). The spreadsheet contains a table with the following data:

	A	B	C	D	E	F	G	H	I	J
1			Peso de los estudiantes							
2	[a, b)	a	b	xi	fi	Fi	hi	Hi	pi	Pi
3	[41, 47)	41	47	44	4	4	0,16	0,16	16%	16%
4	[47, 53)	47	53	50	7	11	0,28	0,44	28%	44%
5	[53, 59)	53	59	56	4	15	0,16	0,6	16%	60%
6	[59, 65)	59	65	62	3	18	0,12	0,72	12%	72%
7	[65, 71)	65	71	68	4	22	0,16	0,88	16%	88%
8	[71, 77]	71	77	74	3	25	0,12	1	12%	100%
9					25		1		100%	
10										