

Tutorial-01: Estadística descriptiva con Calc.

- Este documento pdf lleva adjuntos algunos de los ficheros de datos necesarios. Y está pensado para trabajar con él directamente en tu ordenador. Al usarlo en la pantalla, si es necesario, puedes aumentar alguna de las figuras para ver los detalles. Antes de imprimirlo, piensa si es necesario. Los árboles y nosotros te lo agradeceremos.
- Fecha: 6 de septiembre de 2017. Si este fichero tiene más de un año, puede resultar obsoleto. Busca si existe una versión más reciente.

Índice

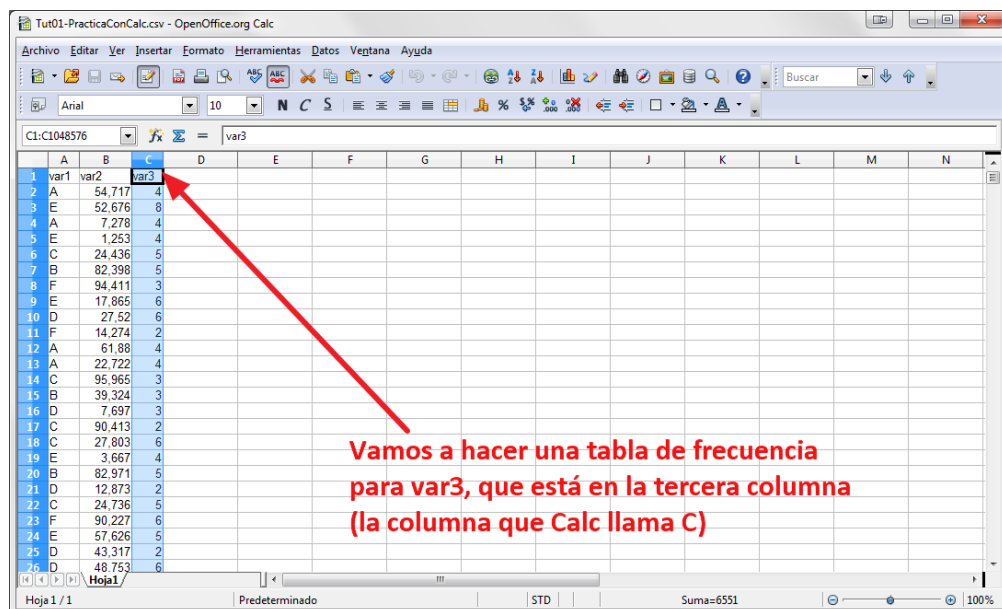
1. Tablas de frecuencia sencillas.	1
2. Gráficos de barras y sectores.	12
3. Cómo usar las referencias a celdas de la hoja de cálculo.	15
4. Media aritmética.	22
5. Medidas de posición: mediana, percentiles, moda.	34
6. Varianza y desviación típica.	38
7. Ejercicios adicionales y soluciones.	45

1. Tablas de frecuencia sencillas.

En el Tutorial-00 hemos aprendido a cargar un fichero de datos (de tipo `csv`) en la hoja de cálculo Calc. Ahora ha llegado el momento de empezar a hacer algo con esos datos. Para empezar, vamos a practicar obteniendo algunas tablas de frecuencia. Veremos en primer lugar tablas muy sencillas, aquellas en las que no es necesario agrupar los datos en clases o intervalos. Vamos a empezar abriendo, para ello, el fichero

[Tut01-PracticaConCalc.csv](#)

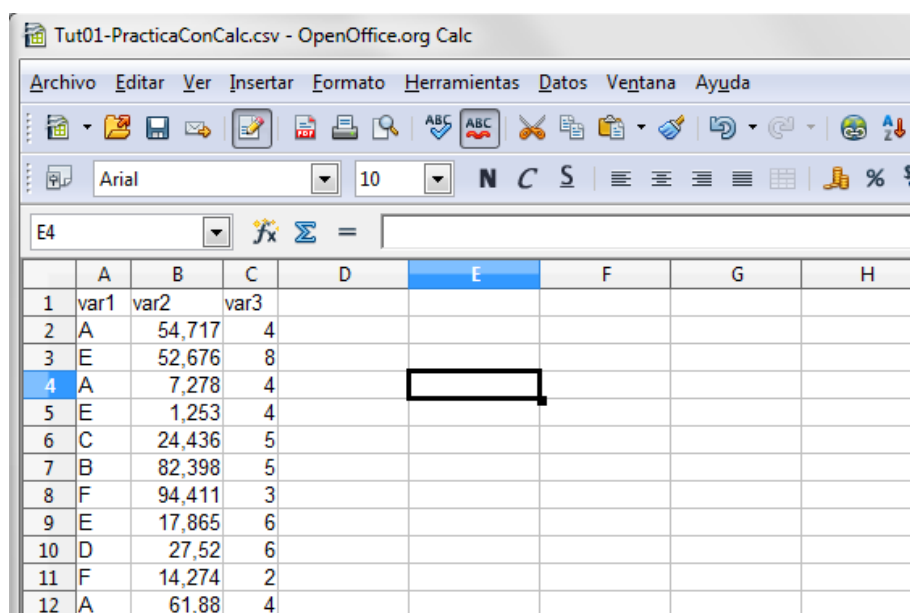
que hemos usado en el Tutorial-00 (lo hemos adjuntado aquí otra vez para facilitarte las cosas). Una vez abierto, el contenido aparecerá como en la siguiente figura. Nuestro primer objetivo es aprender a construir una tabla de frecuencia de la variable `var3`, cuyos valores ocupan la tercera columna de la tabla (la columna C en Calc).



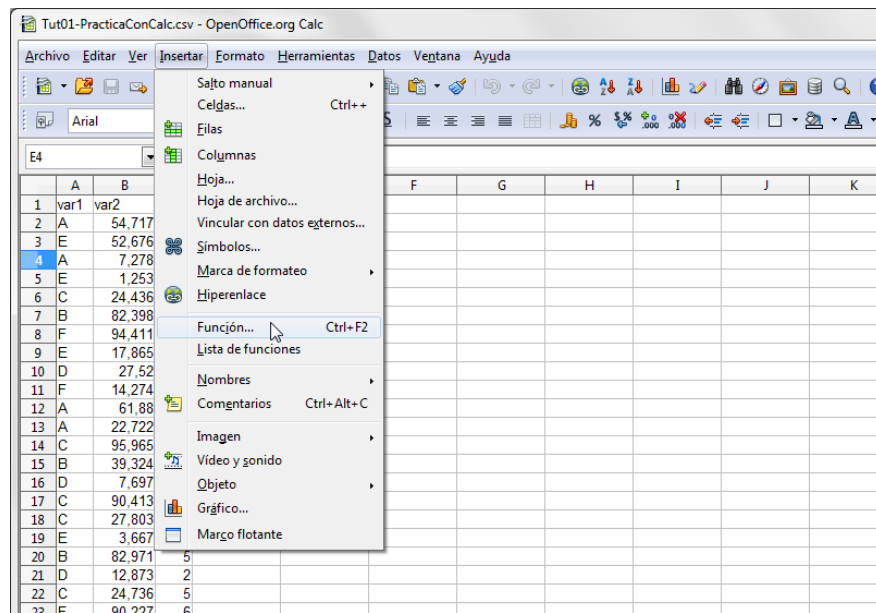
Si echamos un vistazo a los valores de esa columna veremos que se trata de números enteros. Pero es difícil saber, simplemente mirando, y teniendo en cuenta que hay 1300 filas, cuál es el valor máximo de esos números. Afortunadamente, Calc nos permite averiguar eso de una forma muy sencilla. Vamos a utilizar una función de la hoja de cálculo, la primera que encontramos. Veremos muchas más antes de que acabe el curso. La función que vamos a ver se llama MAX y sirve para encontrar el valor máximo en un conjunto de celdas ocupadas por números.

Una advertencia: en algunas versiones anteriores (pero recientes) de Calc, el nombre de esta función, y de algunas otras aparecía con acento, MÁX. Y así lo verás en algunas figuras de este tutorial, que se prepararon con esas versiones previas. Asegúrate de cuál es el nombre correcto en la versión de Calc que estés utilizando.

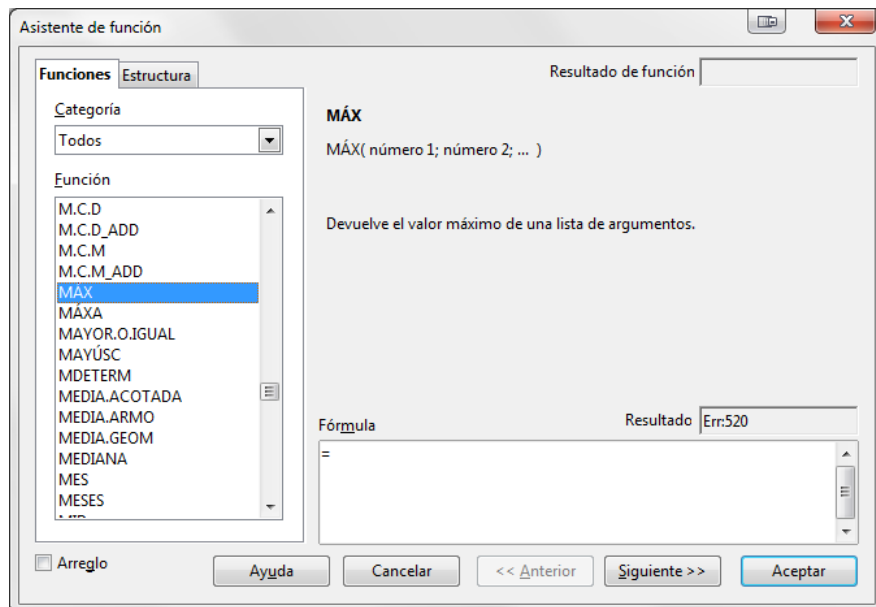
Empezamos por situarnos en una celda no ocupada de la hoja de cálculo. Yo he usado la celda E4, pero puedes usar otra celda libre. Por cierto, aprovechamos para indicar que las celdas de la hoja de cálculo se denotan así, con la letra de la columna seguida (sin espacio) del número de la fila, como en E4. Haz clic en esa celda y asegúrate de que está seleccionada, como en esta figura:



Ahora usa el menú Insertar de Calc, y selecciona Función



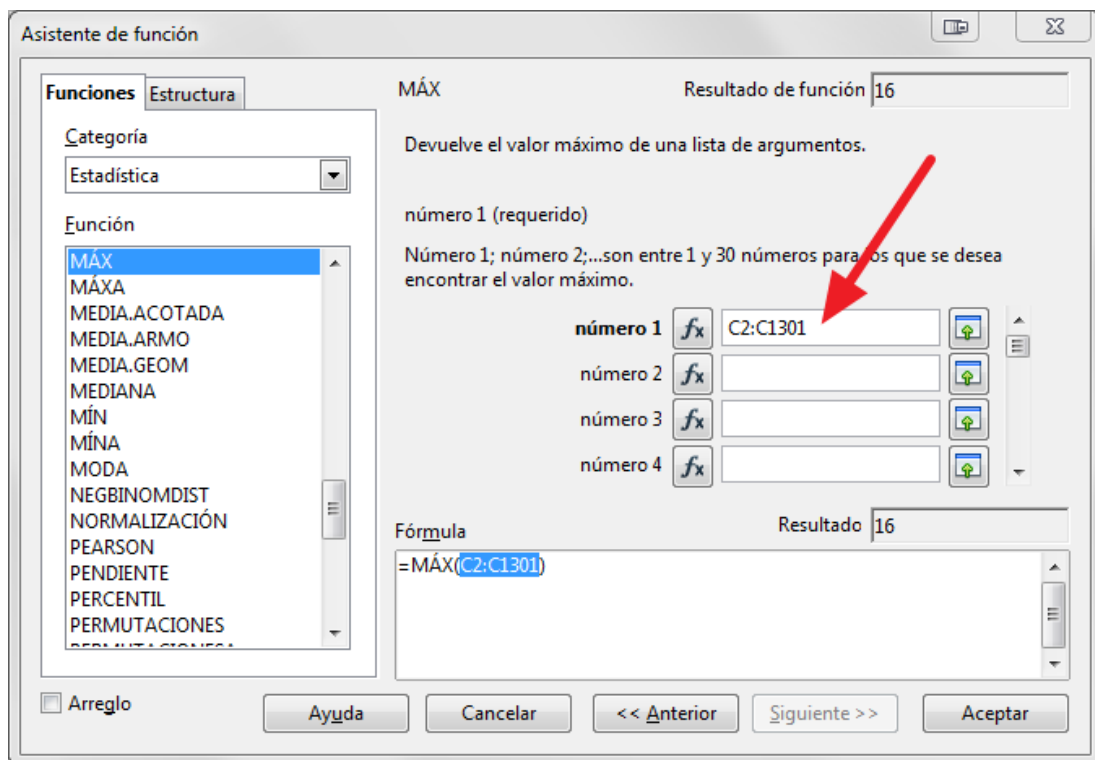
Aparecerá un cuadro de diálogo en el que tenemos que desplazarnos hacia abajo por la lista de funciones para buscar la función **MAX**, como se ve en la siguiente figura:



Una vez seleccionada esa función con un click, pulsamos en siguiente (o hacemos doble click en la función, es lo mismo). Aparece este diálogo, en el que debemos indicar cuáles son las celdas que contienen los números de los que queremos hallar el máximo. En nuestro caso, esas celdas están en la tercera columna, y van desde la C2 hasta la C1301. Ten en cuenta, para entender esto, que la primera celda de esa columna, la C1, está ocupada por el nombre de la variable. En Calc, para decir “desde la celda C2 hasta la celda C1301” escribimos

C2:C1301

separando los nombres de las dos celdas con dos puntos. Esto es lo que se llama un **rango** de celdas (no hay que confundirlo con el *rango* o *recorrido* en sentido estadístico, del que se habla en la Sección 2.3.1 del libro). Escribimos ese rango en el campo que se llama **número 1** (no te preocupes de los otros, puedes dejarlos vacíos).



Fíjate también en el campo llamado *Resultado*, que muestra una vista previa del valor que vamos a obtener. Esta información es especialmente útil para detectar errores anticipadamente.

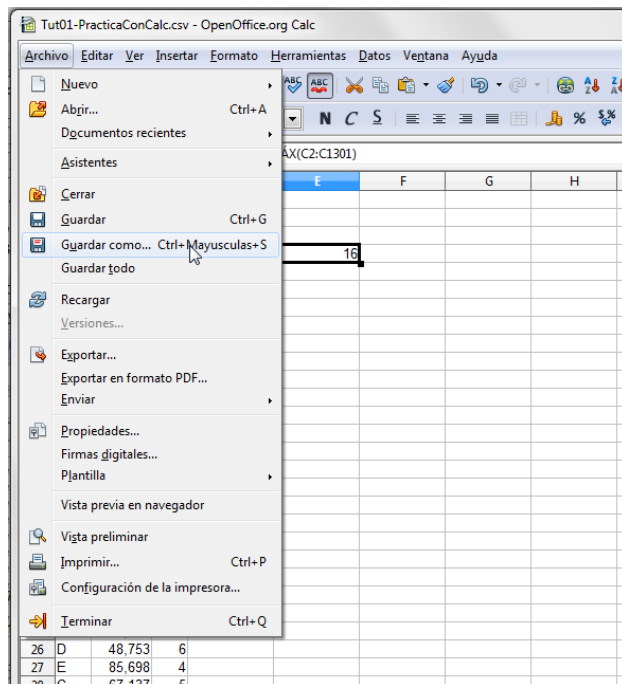
Ahora pulsamos en **Aceptar**, y el resultado aparece en la casilla que habíamos seleccionado.

	A	B	C	D	E	F	G
1	var1	var2	var3				
2	A	54,717	4				
3	E	52,676	8				
4	A	7,278	4		16		
5	E	1,253	4				
6	C	24,436	5				
7	B	82,398	5				
8	F	94,411	3				
9	E	17,865	6				
10	D	27,52	6				
11	F	14,274	2				
12	A	61,88	4				
13	A	22,722	4				
14	C	95,965	3				
15	B	39,324	3				
16	D	7,697	3				
17	C	90,413	2				

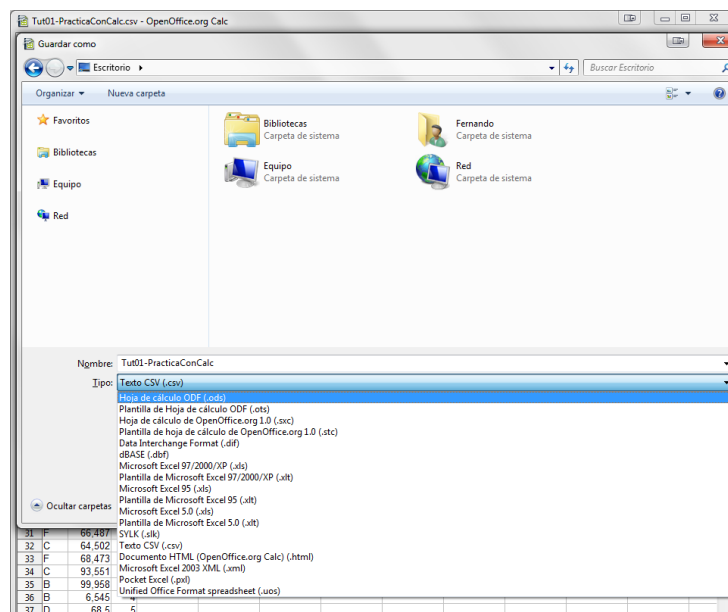
Pon a salvo tu trabajo. Ficheros binarios de tipo ods.

Antes de seguir adelante, vamos a hacer algo **muy importante**, y a aprender la diferencia entre los ficheros *csv* y otro tipo de ficheros, a los que llamaremos **ficheros binarios**. Vamos a recordar donde estamos: hemos empezado con un fichero *csv*, lo hemos abierto en Calc, y ahora le hemos añadido una “operación con los datos”, usando la función **MAX**. Los ficheros *csv* no sirven para almacenar ese tipo de operaciones, porque no están pensados para ello. Son ficheros muy simples, adecuados para *intercambiar* información, pero no para *procesarla*. Para almacenar las operaciones junto con los datos, tenemos que usar otro tipo de ficheros. Podemos seguir trabajando así un rato, pero si ocurre algo o nos equivocamos, perderemos todo el trabajo que llevemos hecho. Por eso

vamos a guardar ahora nuestro trabajo, usando un formato de fichero que nos permita almacenar las operaciones. Usamos el menú **Archivo**, y seleccionamos **Guardar como...**



Aparecerá un cuadro de diálogo, como el que se ve en la siguiente figura,



en el que debemos:

1. Elegir (y recordar) la carpeta en la que guardaremos el fichero. Recuerda que para los ficheros **csv** usamos la subcarpeta **datos** del *Directorio de Trabajo*.
2. Elegir un nombre para el fichero. Vamos a usar

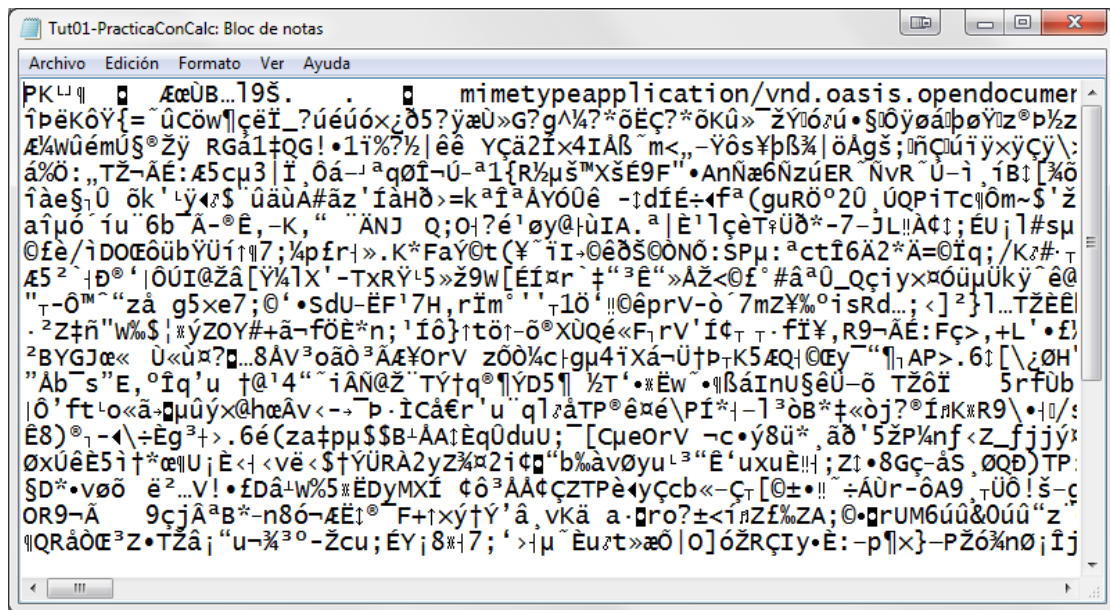
Tut01-PracticaConCalc.ods

Usa exactamente este nombre. No lo cambies, porque lo necesitarás en las próximas secciones.

3. Seleccionar el tipo de archivo **Hoja de cálculo ODF (.ods)**.

De esa forma, cuando pulsemos sobre el botón **Aceptar** guardaremos datos y fórmulas en un mismo fichero, distinto del **csv** con el que hemos empezado. En este paso puedes, si quieres, cambiar el nombre del fichero, aunque es recomendable que el nombre sea, si no igual, al menos parecido al del fichero **csv** con el que hemos empezado. Lo que sin duda habrá cambiado es la **extensión** del

fichero, que habrá pasado de .csv a .ods. Estos ficheros ods no son, desde luego, tan sencillos como los csv. Es un ejercicio saludable abrir uno de ellos con el Bloc de Notas, por ejemplo este que acabamos de crear. Verás algo como esto:



Es una jerigonza imcomprensible de código, en la que es básicamente imposible reconocer nuestros datos originales. La diferencia es que el csv es un fichero de texto plano (que también llamaremos un **fichero fuente**), mientras que este ods es un fichero binario. Simplificando un poco: los ficheros fuente los podemos escribir y entender las personas, mientras que los binarios están pensados para que los entienda el ordenador.

Obtener la tabla de frecuencias.

Volvamos al trabajo de obtener la tabla de frecuencia de la variable **var3**. Habíamos obtenido el valor máximo del rango C2:C1301, que es 16, y lo habíamos guardado en la celda E4.

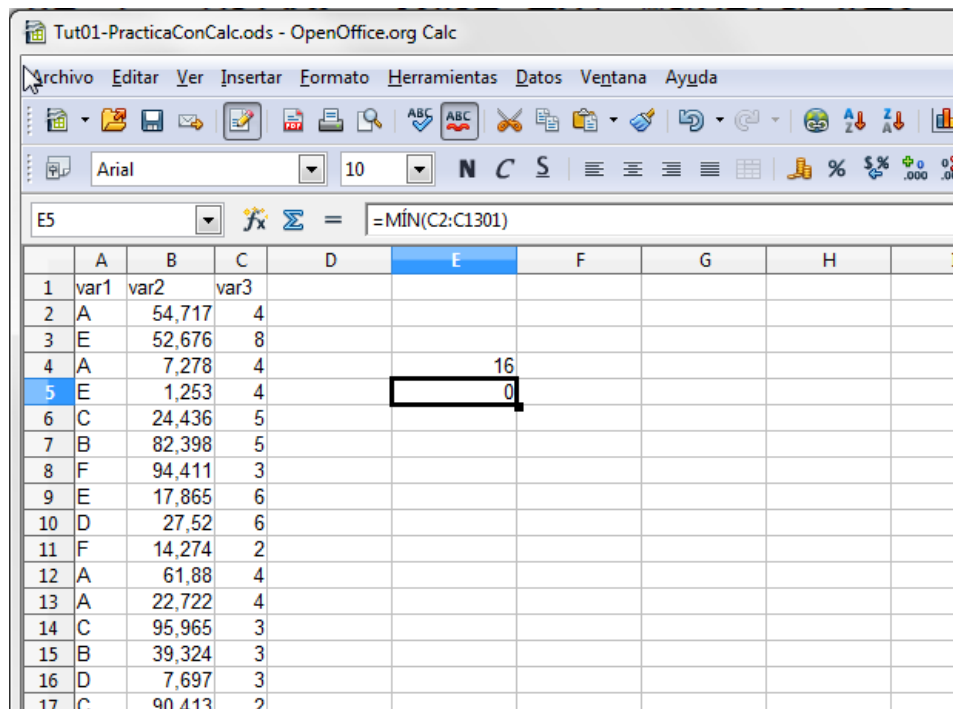
Ejercicio 1.

Busca el valor mínimo de ese rango y guárdalo en la celda E5.



¡No sigas, si no has hecho este ejercicio!

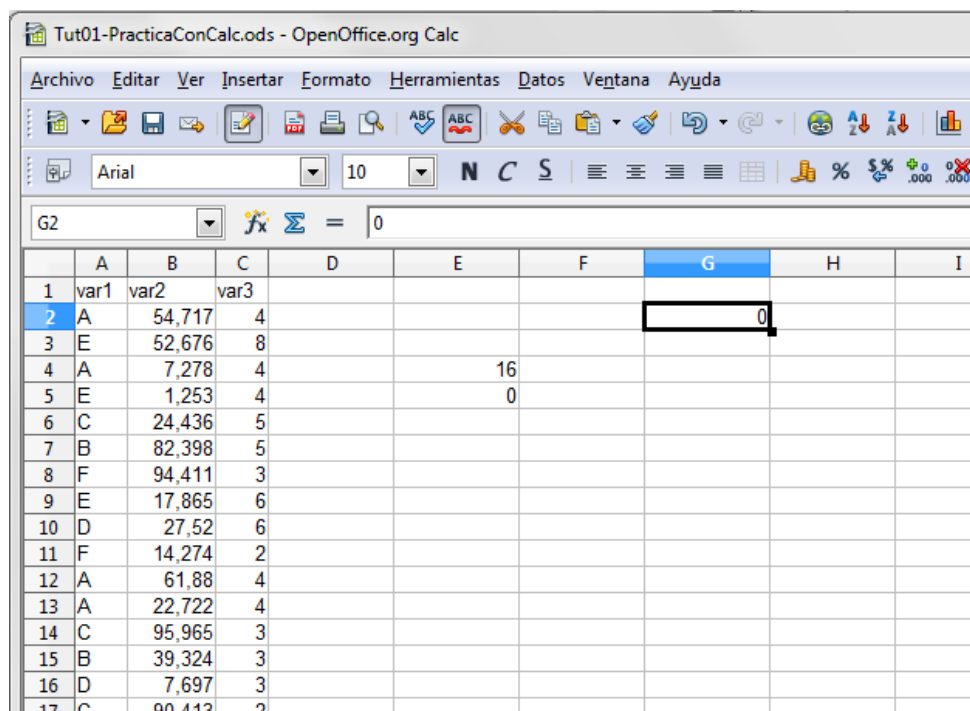
Como muestra la siguiente figura, deberías obtener un 0.



Ahora ya sabemos que ese rango contiene valores del 0 al 16. Así que tenemos que obtener una tabla de frecuencias como esta:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

En Calc, vamos a obtener esta tabla en vertical, en la columna G y H (puedes usar otras). Para ello, empieza por colocar un 0 en la celda G2:



Ahora queremos colocar, en el rango de celdas G3:G18, el resto de los números del 1 al 16, que forman la cabecera de la tabla de frecuencias. Si sabes algo sobre hojas de cálculo, sabrás que hay una forma muy rápida de hacer esto. Adelante, en ese caso. Si eres un recién llegado a este mundo, por el momento te toca escribir esos números “a mano”. Pero no te preocupes, porque es la última vez que te lo pedimos: en la Sección 3 de este tutorial te enseñaremos a ir mucho más rápido, y empezarás a entender cuál es el verdadero sentido y la utilidad de una hoja de cálculo como Calc.

En cualquiera de los dos casos, suponemos que ahora el estado de la hoja de cálculo es este:

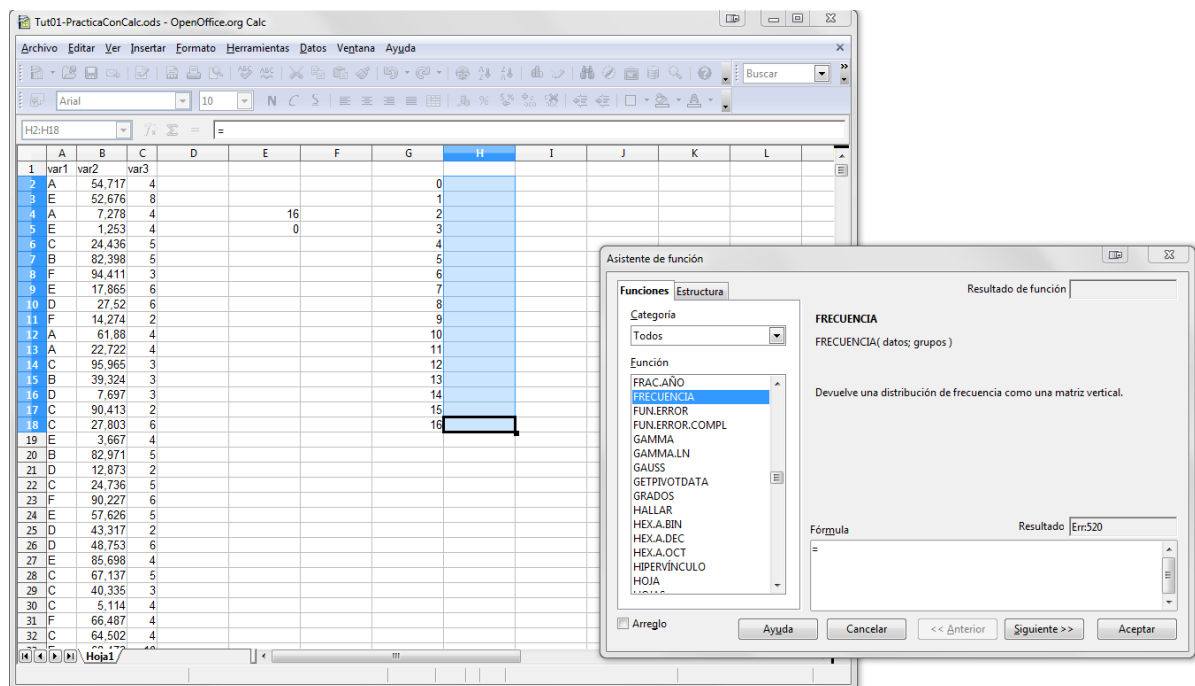
	A	B	C	D	E	F	G	H	I	J
1	var1	var2	var3							
2	A	54,717	4				0			
3	E	52,676	8				1			
4	A	7,278	4		16		2			
5	E	1,253	4		0		3			
6	C	24,436	5				4			
7	B	82,398	5				5			
8	F	94,411	3				6			
9	E	17,865	6				7			
10	D	27,52	6				8			
11	F	14,274	2				9			
12	A	61,88	4				10			
13	A	22,722	4				11			
14	C	95,965	3				12			
15	B	39,324	3				13			
16	D	7,697	3				14			
17	C	90,413	2				15			
18	C	27,803	6				16			
19	E	3,667	4							
20	B	82,971	5							
21	D	12,873	2							
22	C	24,736	5							
23	F	90,227	6							
24	E	67,626	6							

Y queremos que las frecuencias aparezcan justo a la derecha de estos valores, en el rango H2:H18. El primer paso consiste en marcar ese rango, como en esta figura:

D	E	F	G	H	I	J
				0		
				1		
		16		2		
		0		3		
				4		
				5		
				6		
				7		
				8		
				9		
				10		
				11		
				12		
				13		
				14		
				15		
				16		

Es *muy importante* que todo el rango (y sólo el rango) esté marcado, exactamente como aparece en esa figura. Puesto que es la primera tabla de frecuencias que hacemos, lo vamos a hacer con mucho cuidado y saldrá bien. Pero en el futuro, cuando haya problemas, recuerda que la mayoría de los errores al obtener tablas de frecuencia con Calc se deben a que no se ha seleccionado correctamente el rango que ocupa la tabla.

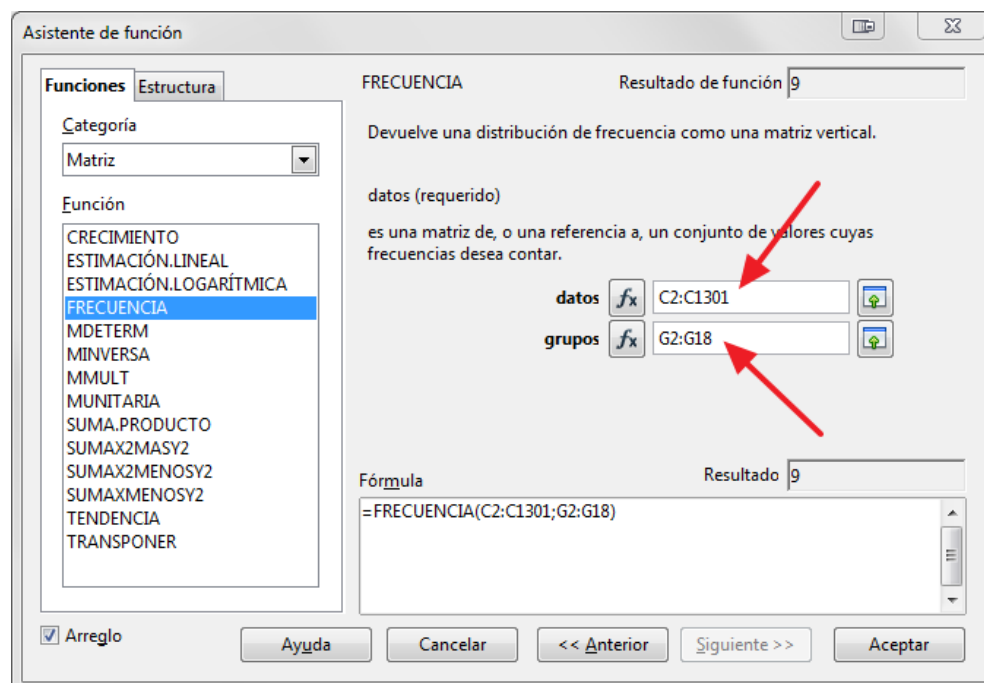
A continuación vamos a utilizar una nueva función de Calc, que se llama, adecuadamente, **FRECUENCIA**. Sin tocar nada (asegúrate de que el rango G2:G18 aparece marcado en azul) vamos al menú **Insertar**, opción **Función**, y en el cuadro de diálogo localizamos esa función **FRECUENCIA**. Es conveniente que coloques ese cuadro de diálogo de forma que no cubra las columnas G y H, como verás que hemos hecho nosotros:



Pulsa en **Siguiente** y asegúrate de rellenar los campos de este cuadro de diálogo de esta manera:

1. En **datos** indica los posiciones que ocupan los datos originales de **var3**, es decir **C2:C1301**,
2. En **grupos** debes indicar las posiciones donde está la lista de valores distintos. O sea, la que será la primera columna de la tabla de frecuencias. Es decir **G2:G18**.

También se pueden seleccionar esos rangos marcándolos con el ratón, y con la práctica, en muchos casos, decidirás si prefieres usar el ratón o el teclado. Haz experimentos, si quieres, y si algo no funciona, pulsa en **Cancelar** y vuelve al paso anterior. En cualquier caso, al final debe quedar como en esta figura:



Ahora puedes pulsar en **Aceptar** y verás aparecer tu nueva y flamante tabla de frecuencias:

	A	B	C	D	E	F	G	H	I	J	K	L
1	var1	var2	var3									
2	A	54,717	4				0	9				
3	E	52,676	8				1	25				
4	A	7,278	4		16		2	100				
5	E	1,253	4		0		3	188				
6	C	24,436	5				4	244				
7	B	82,398	5				5	246				
8	F	94,411	3				6	186				
9	E	17,865	6				7	131				
10	D	27,52	6				8	87				
11	F	14,274	2				9	43				
12	A	61,88	4				10	28				
13	A	22,722	4				11	6				
14	C	95,965	3				12	2				
15	B	39,324	3				13	2				
16	D	7,697	3				14	2				
17	C	90,413	2				15	0				
18	C	27,803	6				16	1				
19	E	3,667	4									
20	B	82,971	5									
21	D	12,873	2									
22	C	24,736	5									
23	F	90,227	6									
24	E	57,626	5									
25	D	43,317	2									
26	D	48,753	6									
27	E	85,698	4									
28	C	67,137	5									
29	C	40,335	3									
30	C	5,114	4									
31	F	66,487	4									
32	C	64,502	4									

¡Enhorabuena! Acabamos de dar el primer paso para convertirnos en gurús de la Estadística. Un vistazo rápido a la tabla que acabamos de obtener nos informa, por ejemplo, de que:

- el valor que más aparece es el 5, que se repite (su frecuencia es) 246 veces.
- el valor 15 no aparece en la tabla (su frecuencia es 0).
- la tabla de frecuencias tiene una “forma” curiosa, con valores que aumentan desde el 0 hasta el 5 y luego vuelven a disminuir.

Algunas observaciones adicionales.

- En nuestro blog *PostData* hay una entrada, a la que puedes llegar con este enlace:

<http://fernandosasegundo.wordpress.com/2012/09/07/tablas-de-frecuencias-en-hojas-de-calculo-calc-y-excel>

en la que se explica como hacer esto en la hoja de cálculo Excel. Es esencialmente igual, pero hay una pequeña diferencia al final del proceso, que te conviene conocer si vas a usar Excel. Además, en esa entrada hay un vídeo que resume el proceso.

- No hemos hecho tablas de frecuencias de las variables **var1** y **var2** del fichero de datos **Tut01-PracticaConCalc.csv**. La variable **var1** nos obligaría a aprender algunos trucos sobre el uso de la hoja de Cálculo que no vamos a necesitar (porque usaremos programas más avanzados, como R), así que dejaremos ese asunto pendiente hasta más adelante (pero si te intriga, busca información sobre la función **CONTAR.SI** de Calc). Por su parte, la variable **var2** requiere agrupar los datos en clases, y nos va a llevar a aprender más sobre hojas de cálculo. Lo haremos en la Sección 4.3.

Pero antes, en la próxima sección, vamos a avanzar en nuestra comprensión del funcionamiento de las hojas de cálculo.

De momento, y para practicar lo que acabamos de aprender, vamos a hacer varios ejercicios.

Ejercicio 2. *En este ejercicio vamos a usar otro fichero de datos, que tienes aquí adjunto:*

[Tut01-PracticaConCalc-01.csv](#).

Usando este fichero, haz lo siguiente:

- 1. Abrelo con Calc, verás que contiene una única columna de datos, de una variable llamada **x**.
¿Cuáles son sus valores mínimo y máximo?*
- 2. Antes de seguir, guarda tu trabajo en un fichero de tipo ods, llamado*

Tut01-PracticaConCalc-01.ods

¡No lo olvides! Acostúmbrate a grabar los ficheros, de lo contrario puedes perder tu trabajo. Además, vas a necesitar algunos de estos ficheros en futuros tutoriales.

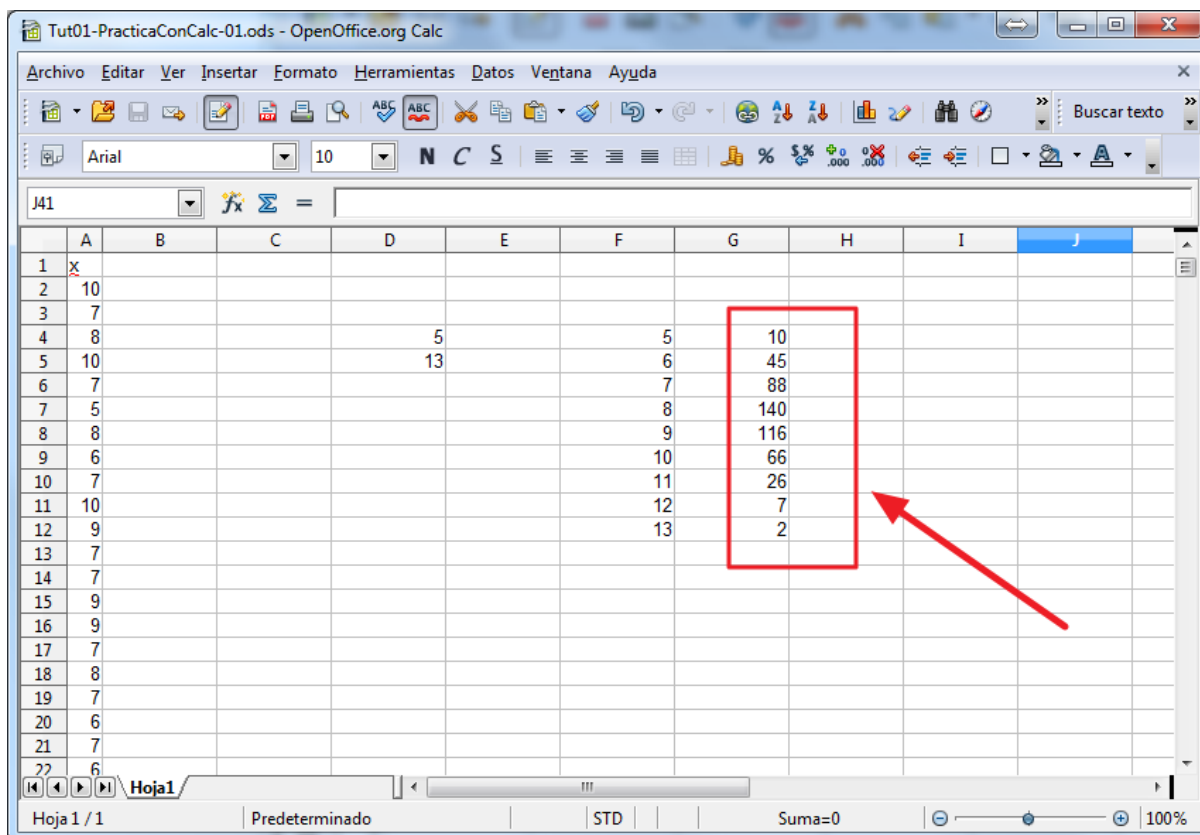
- 3. Construye la tabla de frecuencias de esa variable.*
- 4. ¿Cuál es el valor con mayor frecuencia? ¿Cuál el de menor? ¿Cuál es la frecuencia del valor 11?*
- 5. ¿Cuántos valores menores o iguales que 7 toma la variable **x** (sumando todas las repeticiones que aparezcan en la columna **A**)?*

Tienes las soluciones al comienzo de la siguiente sección. Pero no te apresures a miraras, es imprescindible que aprendas a hacer estas operaciones antes de seguir. □

¡No sigas, si no has hecho los ejercicios!

2. Gráficos de barras y sectores.

Para empezar a dibujar algunos gráficos con Calc, vamos a usar el fichero Tut01-PracticaConCalc-01.csv (con el que has hecho el Ejercicio 2 del final de la sección anterior). Con este fichero vamos a aprender a dibujar gráficos de sectores y columnas en Calc. Pero antes, veamos la solución de ese ejercicio. Si todo ha ido bien, has debido obtener este resultado:

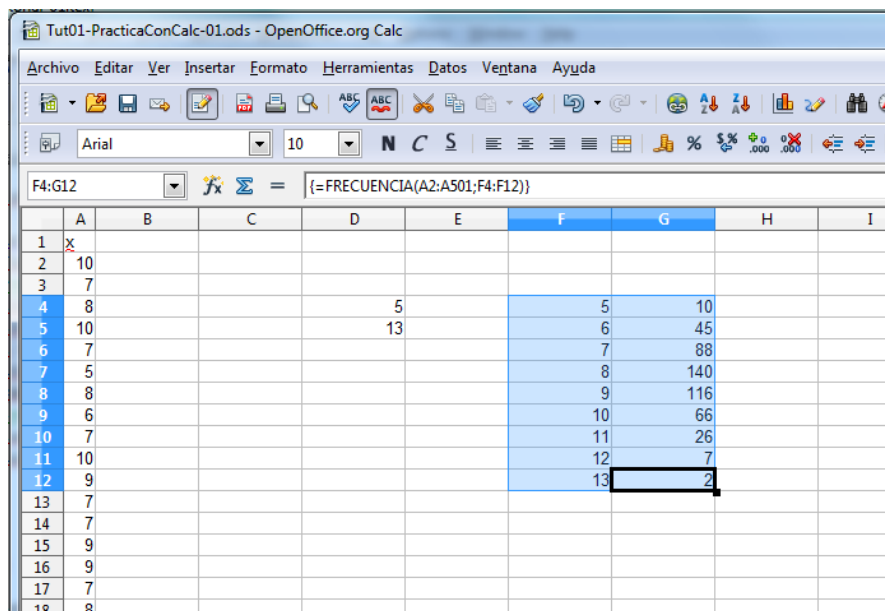


	A	B	C	D	E	F	G	H	I	J
1	x									
2	10									
3	7									
4	8			5		5	10			
5	10			13		6	45			
6	7					7	88			
7	5					8	140			
8	8					9	116			
9	6					10	66			
10	7					11	26			
11	10					12	7			
12	9					13	2			
13	7									
14	7									
15	9									
16	9									
17	7									
18	8									
19	7									
20	6									
21	7									
22	6									

aunque es posible que hayas colocado los valores en otras celdas. Como ves, los valores mínimo y máximo de x son, respectivamente, 5 y 13, el valor con mayor frecuencia es 8 (su frec. es 140), el de menor frecuencia es 13 (su frec. es 2), y la frecuencia de 11 es 26. Para calcular cuántos valores menores que o iguales a 7 toma x tienes que sumar las frecuencias de 5, 6 y 7. Se obtiene:

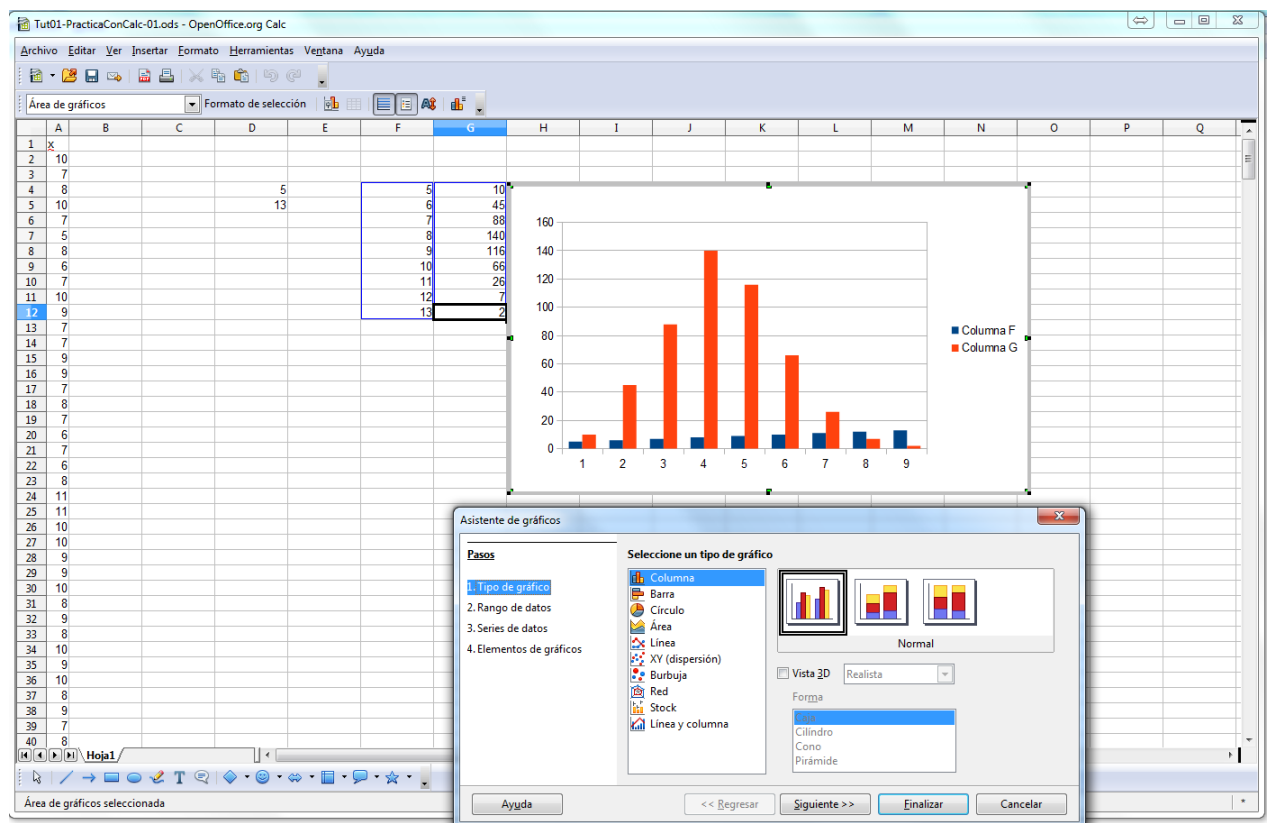
$$10 + 45 + 88 = 143.$$

Gráfico de sectores. Vamos a empezar dibujando uno de estos gráficos que corresponda a la tabla de frecuencias que acabamos de obtener. Seleccionamos toda la tabla,

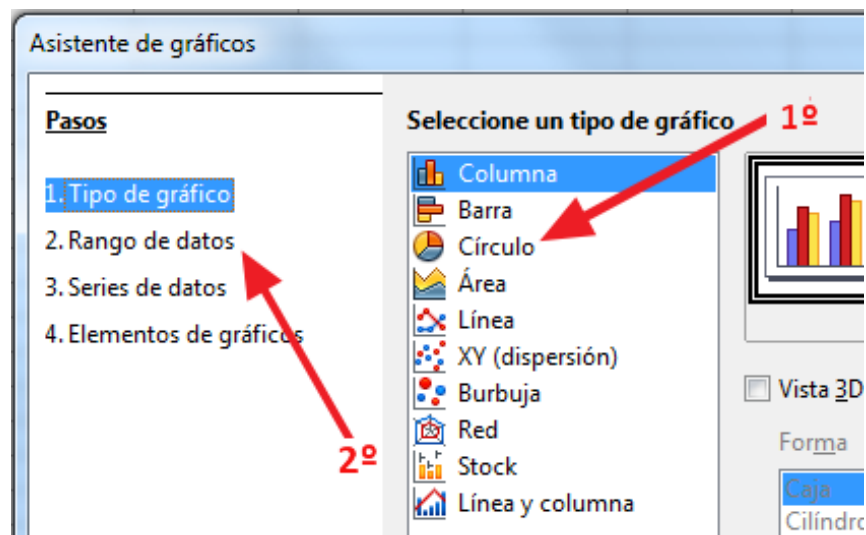


	A	B	C	D	E	F	G	H	I
1	x								
2	10								
3	7								
4	8			5		5	10		
5	10			13		6	45		
6	7					7	88		
7	5					8	140		
8	8					9	116		
9	6					10	66		
10	7					11	26		
11	10					12	7		
12	9					13	2		
13	7								
14	7								
15	9								
16	9								
17	7								
18	8								

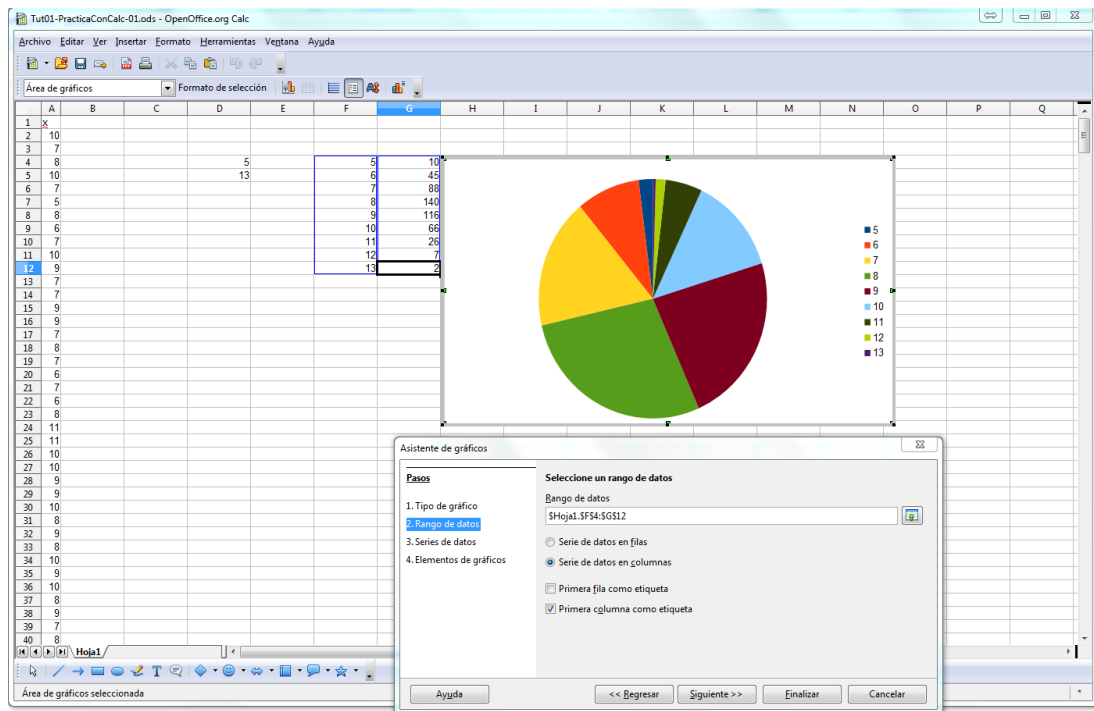
Y ahora vamos al menú **Insertar**, opción **Gráfico**. Aparecerá un gráfico y un cuadro de diálogo, como en la figura. No te preocupes por el gráfico, aún tenemos que ajustarlo.



En ese cuadro de diálogo selecciona **Círculo** y espera unos segundos.

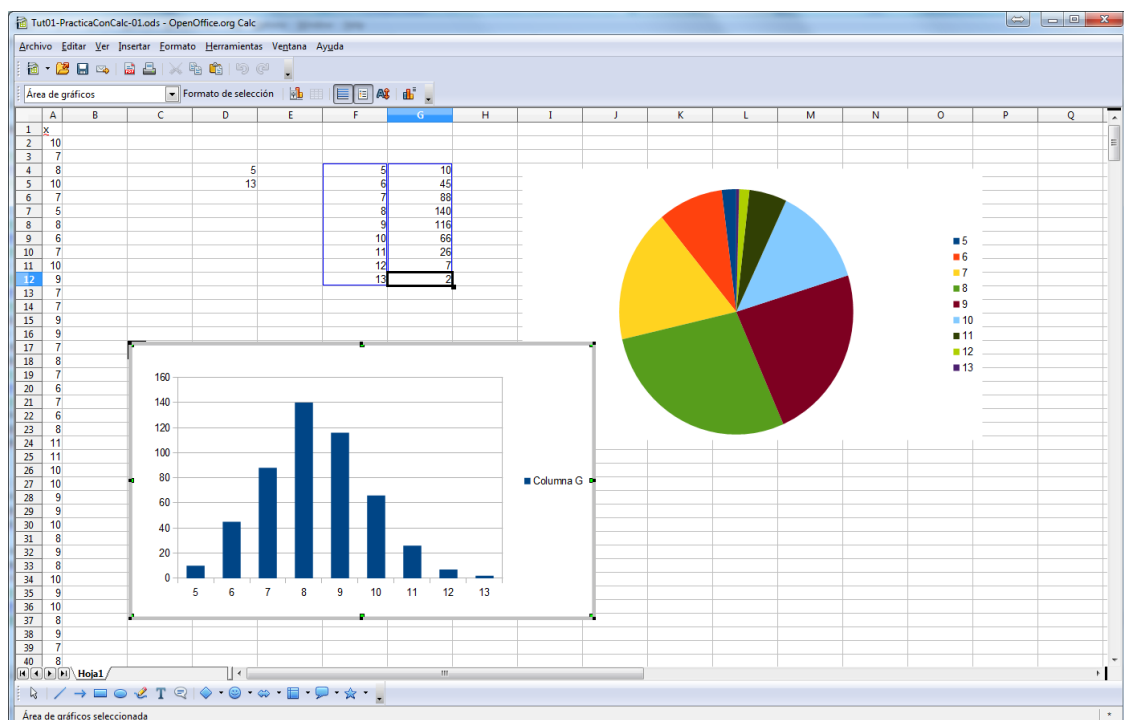


Verás aparecer un gráfico de sectores. ¡Pero es incorrecto, todavía debemos hacer otra cosa! En el cuadro de diálogo, en la ventana de la izquierda, ve a **2. Rango de datos** y asegúrate de marcar la casilla de la opción **Primera columna** como **etiqueta**. Al hacerlo verás que el gráfico cambia. Al hacer esto le estamos pidiendo a Calc que interprete las dos columnas de esa tabla como una tabla de valores y frecuencias.



y ya puedes pulsar en **Finalizar**. Puedes colocar ese gráfico dentro de la hoja de cálculo, donde más te convenga, pinchando sobre el pequeño marco gris que lo rodea y arrastrando con el ratón. No pinches en la zona blanca del gráfico (y si lo haces, usa **Ctrl+Z** para deshacer.) También puedes copiar y pegar el gráfico si lo necesitas en otro documento.

Gráfico de barras. Para terminar esta sección vamos a añadir un gráfico de barras (o columnas, la diferencia en todo caso es la orientación vertical u horizontal). El gráfico de sectores que hemos obtenido no es de los peores, pero en general desaconsejamos el uso de este tipo de gráficos. Queremos que entiendas cómo se hacen, y lo que significan, pero insistimos, demasiadas veces resultan confusos. Para obtener un gráfico de barras los pasos son muy parecidos. Vamos rápido: volvemos a seleccionar la tabla completa, menú **Insertar**, opción **Gráfico**, pero ahora elegimos columnas, de nuevo en **Rango de datos** marcamos la casilla de **Primera columna como etiqueta**, y veremos aparecer este gráfico:



Los valores aparecen como etiquetas al pie de cada columna. Fíjate en lo fácil que resulta, en este gráfico, localizar los valores de mayor y menor frecuencia.

Ejercicio 3.

Haz un gráfico de barras para la variable **var3** del fichero **Tut01-PracticaConCalc.csv**, que encontrarás adjunto en la página 1. □

En ese mismo fichero, la representación gráfica adecuada para la variable **var2**, una vez agrupada en clases, es un histograma. Todavía no hemos aprendido a agrupar por clases en Calc. Y, en cualquier caso, con Calc no es fácil dibujar histogramas correctamente. Pero no hay que preocuparse. En el próximo tutorial empezaremos a usar R, la herramienta a la que más tiempo vamos a dedicar, y veremos que con R, dibujar histogramas es muy fácil.

3. Cómo usar las referencias a celdas de la hoja de cálculo.

Este apartado va dirigido a aquellos lectores que tienen escasa o nula experiencia con una hoja de cálculo como Calc o Excel. Para ellos, es obligatorio una lectura atenta de lo que sigue, y además es necesario ir reproduciendo simultáneamente todos los pasos en un ordenador. Si te atascas en alguno de ellos, vuelve a leer, asegúrate de que estás haciendo exactamente lo que se describe en el texto. Y si aún así tienes problemas, pide ayuda a quien sepa más que tú (ya sabes...). Si, por el contrario, te mueves con soltura en este tipo de programas, seguramente no vas a aprender nada nuevo. Aún así te recomendamos, al menos, una lectura rápida para comprobar que no hay sorpresas, y para que te familiarices con la terminología que vamos a usar. Después, puedes pasar directamente al siguiente apartado.



En la Sección 1, al construir la tabla de frecuencias para la variable **var3**, teníamos que rellenar una rango de celdas de Calc con los números del 1 al 16. Y dijimos entonces que había una forma rápida de hacer esto, que vamos a ver a continuación. Empieza por abrir de nuevo con Calc el fichero **Tut01-PracticaConCalc.csv** (lo tienes en la página 1 de este tutorial), recordando los pasos necesarios para esa tarea. Si lo tienes a mano, también puedes usar el fichero de tipo ods, pero te recomendamos empezar a partir del csv.

Vamos a suponer, para empezar, que Calc está abierto, el fichero se ha cargado, y estamos en este estado:

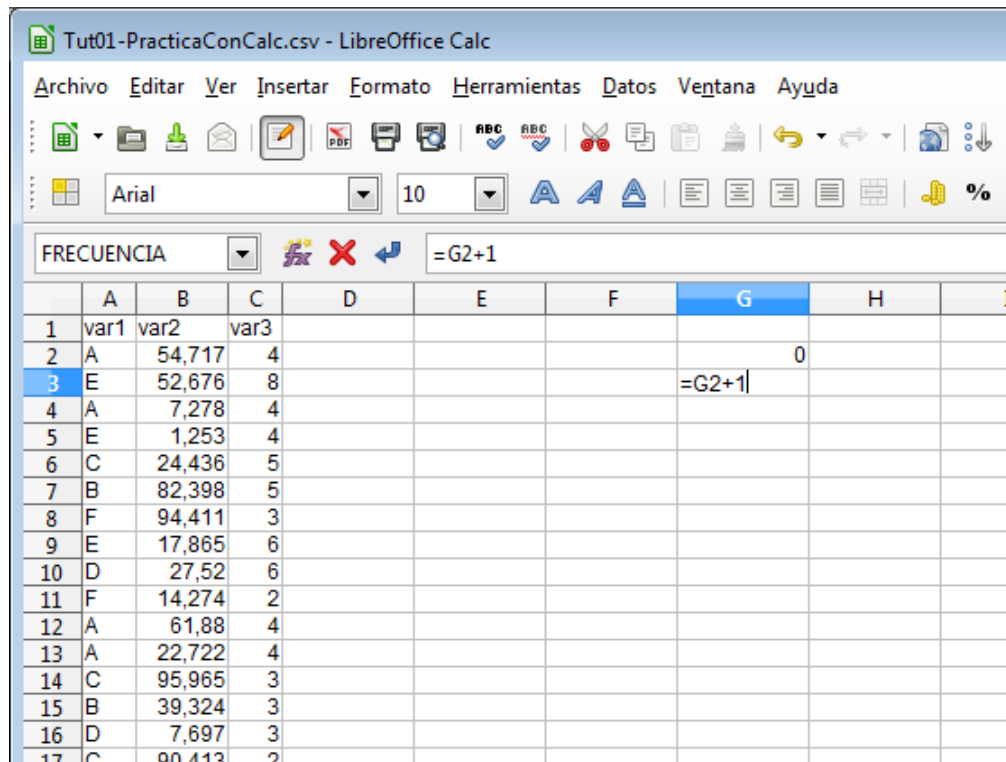
	A	B	C	D	E	F	G
1	var1	var2	var3				
2	A	54,717	4				0
3	E	52,676	8				
4	A	7,278	4				
5	E	1,253	4				
6	C	24,436	5				
7	B	82,398	5				
8	F	94,411	3				
9	E	17,865	6				

de manera que la celda **G2** contiene el número 0. Recuerda, **G2** significa: columna **G** y fila 2. Nuestro objetivo es que aparezcan los números 1 al 16 en las celdas **G3:G18** (recuerda que los dos puntos indican un rango o grupo de celdas contiguas en la hoja de cálculo). Vamos a ir paso a paso, en el futuro podrás ir más rápido. La idea básica es que:

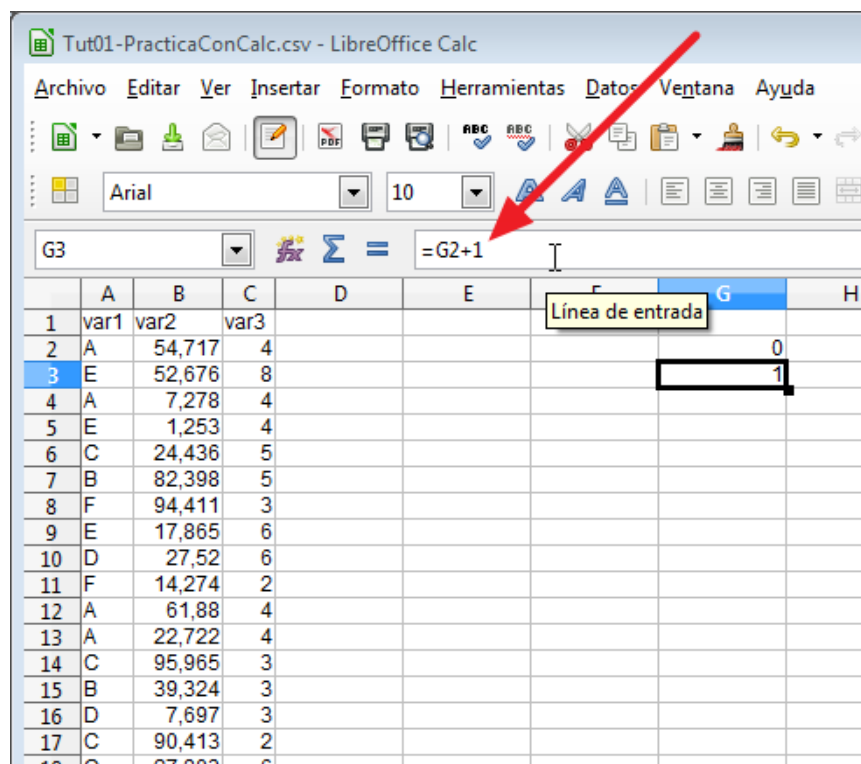
el contenido de cada celda se obtiene sumando 1 al de la celda que tiene encima.

La propiedad más importante de una hoja de cálculo es que es muy fácil utilizar descripciones como “la celda que está encima” o “la celda situada dos posiciones hacia la derecha”, etcétera. Y además, esas descripciones se pueden *copiar y pegar*, también muy fácilmente. No te preocupes si ahora mismo todo esto parece un poco confuso, con la práctica lo verás claro.

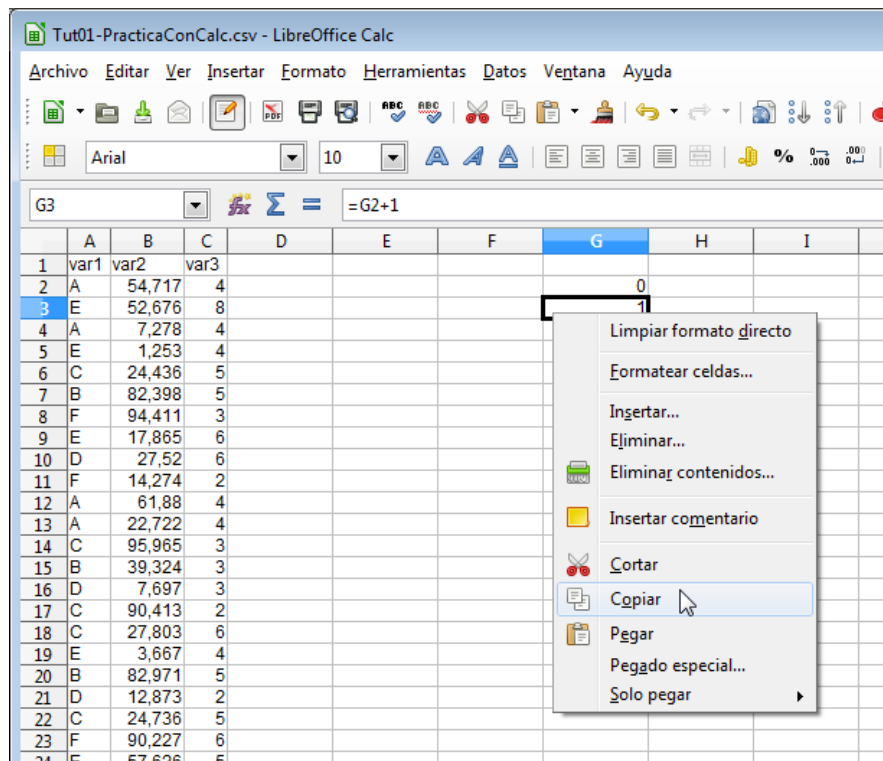
Ahora, haz clic con el ratón sobre la celda G3, y pulsa la tecla =. Al usar esta tecla le estamos diciendo a Calc que esa celda va a contener una fórmula, algo que la hoja de cálculo va a tener que calcular, en lugar de simplemente un valor que nosotros tecleamos directamente. A continuación del símbolo igual, teclea G2+1, como se ve en la figura. Esa es la expresión de la fórmula que Calc tiene que calcular. Y significa, como es fácil imaginar, “toma el contenido de G2 y súmale 1”.



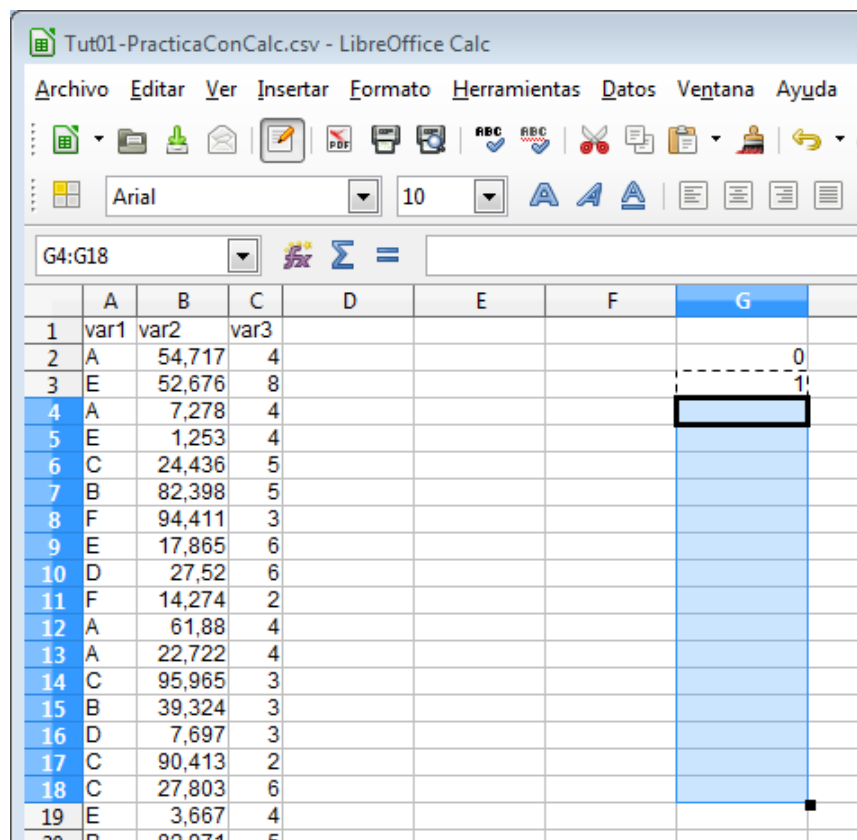
Cuando pulses **Entrar**, en la casilla G3 aparecerá un 1. Pero si haces clic sobre esa casilla, y miras en la *Línea de Entrada* (indicada por la flecha roja de la figura), verás la fórmula que hemos utilizado.



Hasta aquí, seguramente, no hemos hecho nada demasiado espectacular. Ahora empieza lo bueno: asegúrate de tener seleccionada la celda G3 (aparecerá resaltada, con un rectángulo negro más grueso). Vamos a copiar esa celda al Portapapeles. Puedes pulsar **Ctrl+C**, o usar el botón derecho del ratón como muestra la figura



Al hacerlo puede que veas que la celda G3 queda doblemente resaltada, con una línea de trazos (depende de la versión de Calc que uses). Ahora haz clic en la celda G4 y, manteniendo pulsado el ratón, arrástralo para marcar (seleccionar) las restantes celdas del rango, de la G4 a la G18. Si se hace correctamente, las celdas quedarán marcadas en azul, como en esta figura:



Ahora vamos a pegar en esas celdas la fórmula que hemos copiado al Portapapeles, Para ello puedes pulsar **Ctrl+V**, o puedes de nuevo usar el botón derecho del ratón (y seleccionar **Pegar**, claro). El resultado será la lista de números que queríamos:

Tut01-PracticaConCalc.csv - LibreOffice Calc

Archivo Editar Ver Insertar Formato Herramientas Datos Ventana Ayuda

Arial 10

G4:G18 =G3+1

	A	B	C	D	E	F	G	H
1	var1	var2	var3					
2	A	54,717	4				0	
3	E	52,676	8				1	
4	A	7,278	4				2	
5	E	1,253	4				3	
6	C	24,436	5				4	
7	B	82,398	5				5	
8	F	94,411	3				6	
9	E	17,865	6				7	
10	D	27,52	6				8	
11	F	14,274	2				9	
12	A	61,88	4				10	
13	A	22,722	4				11	
14	C	95,965	3				12	
15	B	39,324	3				13	
16	D	7,697	3				14	
17	C	90,413	2				15	
18	C	27,803	6				16	
19	E	3,667	4					
20	B	82,971	5					
21	D	12,873	2					
22	C	24,736	5					

Para entender mejor lo que ha sucedido, haz clic con el ratón en una de las celdas del rango (yo he usado la G13) y mira en la *Línea de Entrada*, para ver la fórmula que Calc está utilizando. Verás que allí aparece $=G12 + 1$. Lo que Calc hace, cuando copia una fórmula, es copiar la descripción de la fórmula en términos de posiciones relativas y no absolutas. Es decir, que al usar descripciones como “la celda de arriba”, si estamos en G13 eso significa la celda G12, y si estuviéramos en E7 significaría E6. Este tipo de descripciones relativas es lo que hace que sea muy fácil manejar y operar con rangos de celdas en la hoja de cálculo, para repetir operaciones con todos los datos de un conjunto. Más adelante en esta sesión, veremos que hay ocasiones en que, precisamente, lo que necesitamos es usar una posición absoluta, y veremos cómo se hace esto.

Deshacer operaciones

Antes de seguir adelante, vamos a aprender a deshacer operaciones; esto será muy útil si cometemos algún error, para evitarnos tener que retroceder hasta el principio. En la barra de herramientas de Calc verás un par de símbolos en forma de flechas curvadas, llamados respectivamente *deshacer* y *rehacer*.

Tut01-PracticaConCalc.csv - LibreOffice Calc

Archivo Editar Ver Insertar Formato Herramientas Datos Ventana Ayuda

Arial 10

G4:G18 =G3+1

	A	B	C	D	E	F	G	H
1	var1	var2	var3					
2	A	54,717	4				0	
3	E	52,676	8				1	
4	A	7,278	4				2	
5	E	1,253	4				3	
6	C	24,436	5				4	
7	B	82,398	5				5	

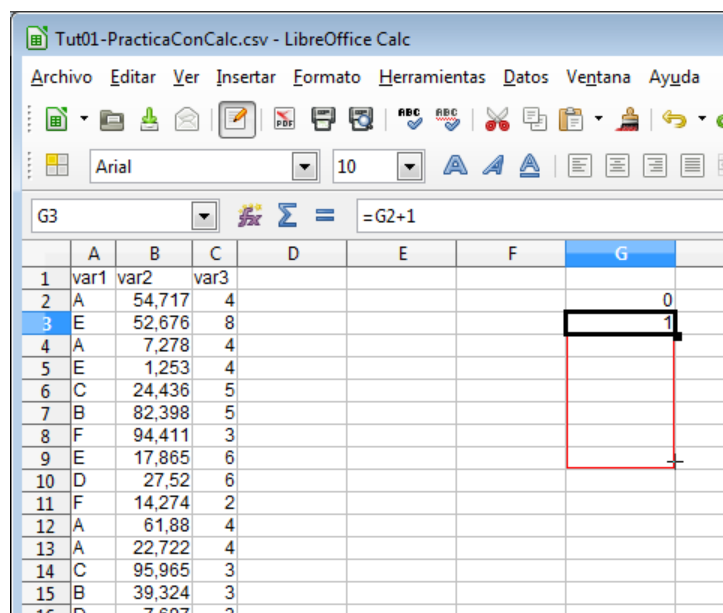
Pulsa el símbolo de deshacer y verás desaparecer la lista de números que obtuvimos en el último paso. Si vuelves a pulsarlo Calc deshace la anterior operación que hicimos, y así sucesivamente.

Haz clic unas cuantas veces sobre ese símbolo, y luego haz clic sobre el símbolo *rehacer*, hasta que hayas entendido como funcionan. Si eres más amigo del teclado, puedes usar **Ctrl+Z** y **Ctrl+Y** en lugar de deshacer y rehacer, respectivamente. Estas combinaciones de teclas funcionan en muchos otros programas, además de Calc, así que es bueno que las conozcas.

Usa estas dos herramientas para volver a la situación en la que sólo estaban ocupadas las celdas G2 (con un 0) y G3 (con un 1, resultado de la fórmula $=G2+1$). Ahora podemos seguir adelante.

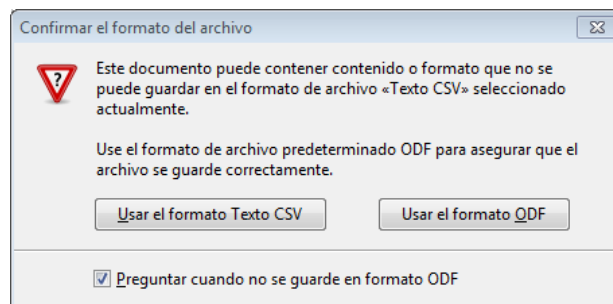
Otra manera de copiar las fórmulas

Haz clic en G3 y fíjate en que la esquina inferior derecha de esa celda aparece un pequeño cuadrado negro. Haz clic sobre ese cuadrado, y manteniendo el botón izquierdo del ratón pulsado, arrástralo para cubrir el resto del rango G3:G18. Si lo haces correctamente, mientras arrastras verás que esas celdas van quedando enmarcadas por un rectángulo rojo. En la figura puedes ver un momento intermedio del proceso:



Al llegar a G18 libera el ratón, y verás que aparecen de nuevo los números del 2 al 16 en esas aceldas. Este procedimiento de arrastrar es útil para copiar una fórmula rápidamente a un pequeño rango de celdas.

Si quieres guardar el trabajo que has hecho en esta hoja, recuerda que debes hacerlo en formato *ods*, y no en formato *csv*. Es importante entender que el formato *ods* es capaz de almacenar fórmulas y gráficos, mientras que el *csv* sólo almacena los datos. Es bueno, para practicar esto, que después del trabajo de esta sección guardes el fichero en los dos formatos, con nombres distintos si es preciso, y que después los abras para ver las diferencias. En cualquier caso, si tratas de guardar un fichero que contiene fórmulas o gráficos en formato *csv*, Calc te avisará con un mensaje de advertencia



Números (pesudo)aleatorios con Calc

Para aprender algo nuevo, y practicar un poco más con el manejo de fórmulas, vamos a usar otra función de Calc. A lo largo del curso vamos a necesitar muchas veces hacer experimentos con datos elegidos “al azar”. Tendremos sobradas ocasiones de discutir en detalle lo que queremos decir con esto, pero por el momento puedes pensar que es como si lanzáramos un dado muchas veces y fuéramos anotando los resultados. Naturalmente, un ordenador no es un dado, y no se pueden

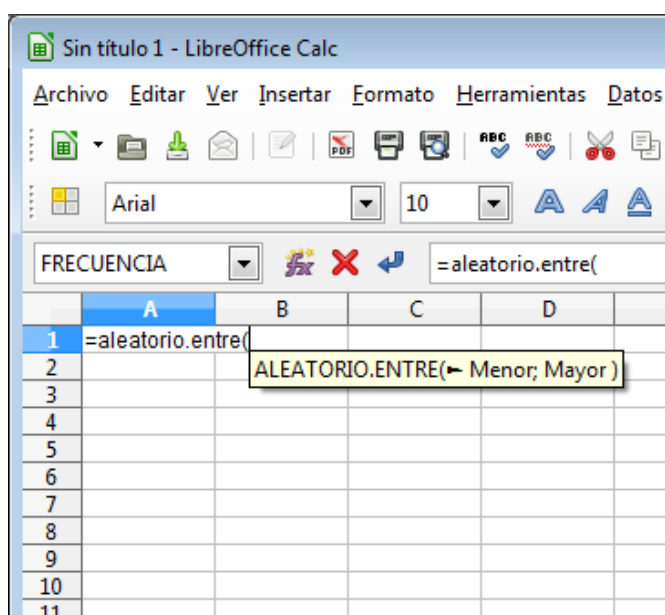
fabricar números verdaderamente aleatorios usando un programa de ordenador. Pero se pueden obtener números **pseudoaleatorios**, que son más que suficientes para nuestros propósitos. Veamos cómo hacerlo con Calc. Abrimos una hoja de cálculo nueva (Menú **Archivo** → **Nuevo** → **Hoja de cálculo**, o simplemente **Ctrl+N**).

Vamos a usar el rango **A1:A100** para simular 100 lanzamientos de un dado. Para conseguir esto vamos a usar una función de Calc, como las funciones **MIN**, **MAX** y **FRECUENCIA** que ya hemos visto. La función que necesitamos ahora se llama **ALEATORIO.ENTRE** (en inglés es **RANDBETWEEN**). Esta función sirve para obtener un número entero aleatorio entre dos valores que elegimos nosotros. Podríamos usar el menú **Insertar** → **Función**, como aprendimos a hacer, pero vamos a hacerlo de otra manera para ver que el resultado es el mismo (dejamos para el lector la tarea de comprobar el uso de los menús).

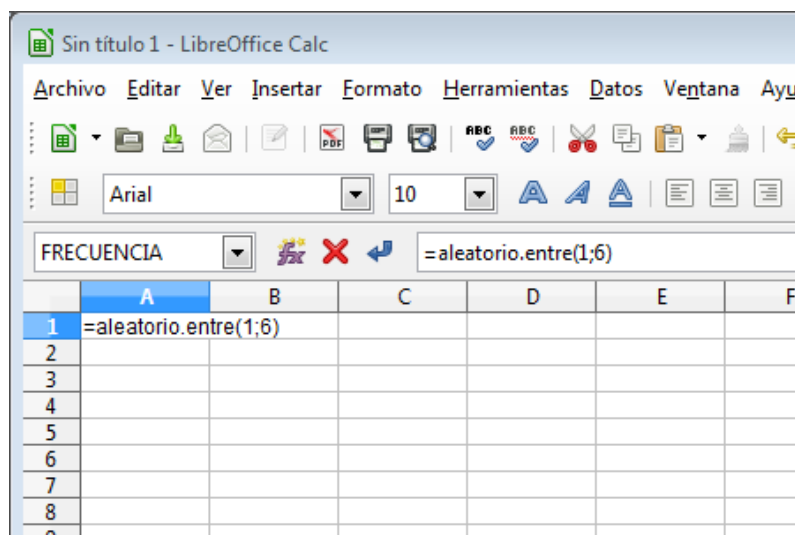
Empezamos haciendo click, por ejemplo, en la celda **A1**. Y ahora tecleamos:

=ALEATORIO.ENTRE(

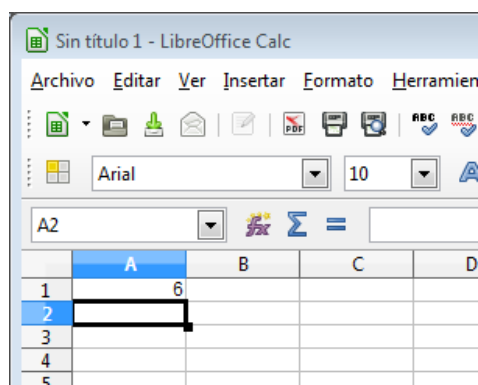
Da igual que uses mayúsculas o minúsculas, pero tienes que empezar con el **=** (que le dice a Calc que a continuación viene una fórmula) y abrir un paréntesis al final. Justo después de escribir el paréntesis verás aparecer un mensaje que indica que Calc ha reconocido la función que estamos usando, y nos da alguna pista sobre la forma correcta de usarla:



En particular, esta función necesita dos números, sus **argumentos**, que Calc representa como **Menor** y **Mayor**. En el caso de un dado, usaremos 1 y 6 como valores **Menor** y **Mayor**, respectivamente. Así que terminamos de escribir la función son esos valores, separados por punto y coma. ¡Esto es importante! En las hojas de cálculo los rangos se indican con dos puntos, como sabemos, y los argumentos de funciones se separan con punto y coma. Muchos de los errores que se cometen se deben a confundir cosas como estas. Una vez tecleado esto:

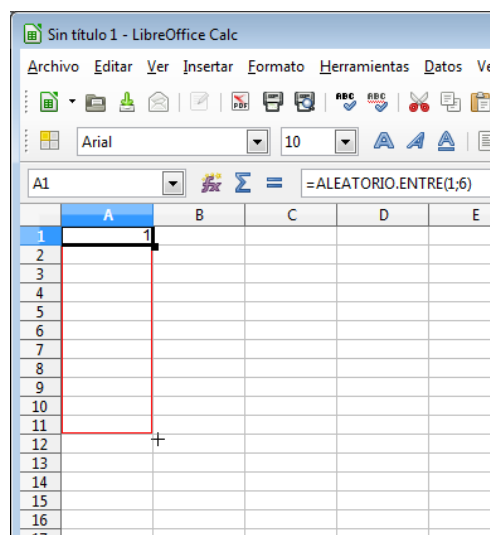


pulsamos **Entrar**, y obtenemos algo como esto:

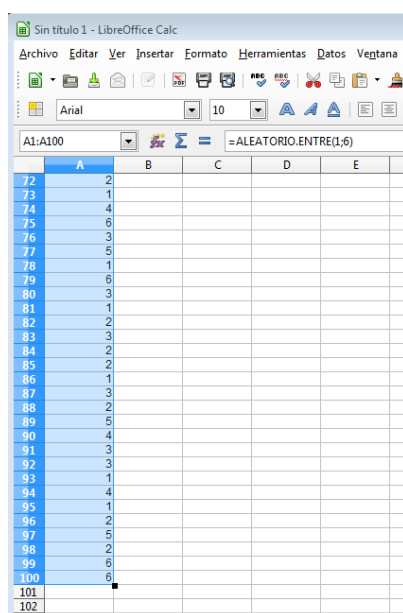


¡Atención! Si lo haces en tu ordenador, obtendrás probablemente otro número del uno al seis. Al fin y al cabo de eso se trata. Estamos simulando el lanzamiento de un dado, y es como si tú lanzaras un dado y yo otro.

Ahora queremos repetir esto cien veces. Y evidentemente, no se trata de que repitas los pasos anteriores cien veces en cada una de las casillas A2:A100. No, lo que vamos a hacer es copiar la fórmula de A1, como hemos aprendido a hacer. Hay varias formas de hacer esto. Mira la figura, que te da una pista del procedimiento que hemos usado nosotros.



El resultado final es una colección de cien números del uno al seis, de la que aquí mostramos el final (los tuyos serán distintos, claro).



Si guardas el fichero (en formato ods, claro), cada vez que lo abras Calc volverá a calcular esos números aleatorios y obtendrás cien distintos cada vez (¿Cuánta suerte crees que necesitas para que te salgan dos veces los mismos cien números? Más adelante en el curso contestaremos...) Si quieres volver a calcular los números sin tener que cerrar y abrir Calc, prueba a pulsar simultáneamente Ctrl + Mays + F9.

4. Media aritmética.

En esta sección, y en las siguientes, vamos a empezar a usar las características de la hoja de cálculo para explorar los conceptos que se discuten en el Capítulo 2 del libro. Empezaremos con la media aritmética, analizando como calcularla según la situación de partida.

4.1. El caso de valores no agrupados.

El fichero adjunto

[Tut01-mediaAritmeticaConCalc.csv](#)

contiene treinta valores de los que queremos calcular la media aritmética. Empieza por abrir el fichero con Calc. Al hacerlo, debes ver esto en tu pantalla:

	A	B	C	D	E	F	G	H
1	9							
2	12							
3	7							
4	12							
5	12							
6	7							
7	10							
8	16							
9	5							
10	4							
11	8							
12	6							
13	9							
14	12							
15	4							
16	11							
17	9							
18	7							
19	9							
20	7							
21	10							
22	6							
23	13							
24	12							
25	10							
26	5							
27	11							
28	11							
29	10							
30	10							
31								

Vamos a calcular la media aritmética de estos números, y lo haremos de varias maneras. Empezamos recordando la definición, que es:

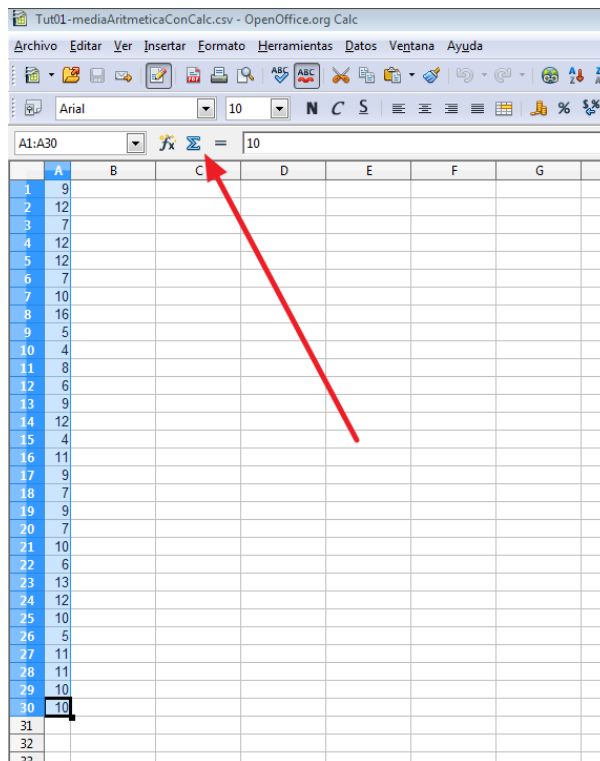
$$\bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}.$$

Así que la receta dice que tenemos que:

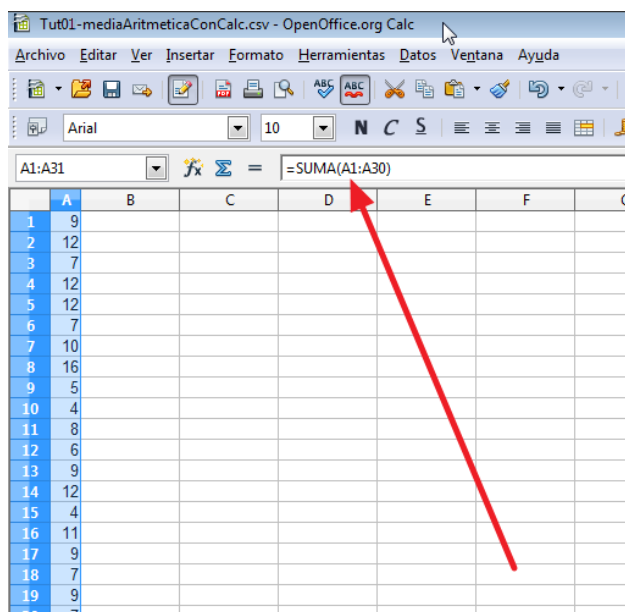
1. sumar todos los números
2. y dividir el resultado por n , que es la cantidad de números que tenemos (en este ejemplo, $n = 30$).

Podríamos hacerlo al revés, dividir primero todos los números por n y luego sumar, pero no hay ventajas en hacer esto y, en cambio, sí que hay inconvenientes. Cada división supone la posibilidad de cometer un error de redondeo. Así que, hablando en general, cuando hacemos cálculos es mucho mejor posponer la división, siempre que sea posible.

Naturalmente, sumar un grupo de números es una operación tan habitual (y no sólo en Estadística), que Calc ofrece herramientas para hacerlo cómodamente. La primera que vamos a ver es la más rápida. Selecciona con el ratón todas las celdas del rango que ocupan los valores, para que queden como en esta figura:



y después haz clic sobre el símbolo de sumatorio Σ de la barra de herramientas, al que señala la flecha roja de la figura. Al hacerlo aparecerá el número 274 en la celda A31, inmediatamente debajo del rango que ocupan los números que sumamos. Este es el procedimiento más rápido, pero no es el más flexible. En particular, a veces queremos más libertad a la hora de colocar el resultado de la suma en otras celdas, o sólo queremos sumar una parte del rango, etcétera. Para aprender a controlar más el proceso, fíjate en la fórmula que Calc ha introducido en la celda A31, que puedes ver en la línea de entrada.

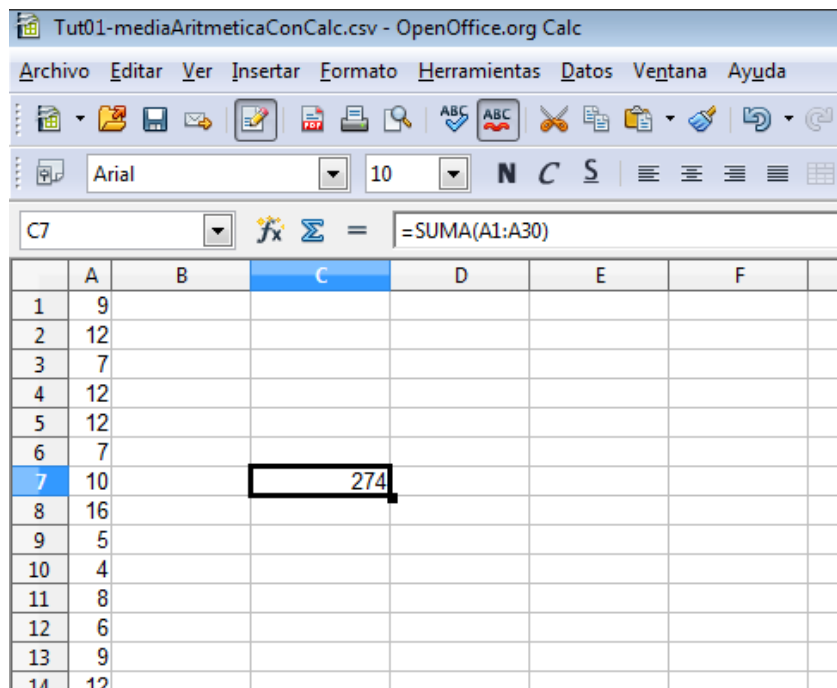


Verás que esa fórmula dice `=SUMA(A1:A30)`. A estas alturas, empieza a resultar muy fácil de entender. Calc está usando una función llamada SUMA, aplicada a todo el rango, de la misma forma que

hicimos con MIN o MAX. Para ver la diferencia con el método anterior, sitúate en cualquier celda de otra columna (yo voy a usar la C7) e introduce allí la misma fórmula:

=SUMA(A1:A30)

El resultado, naturalmente, vuelve a ser 274.



The screenshot shows the OpenOffice.org Calc interface. The title bar reads 'Tut01-mediaAritmeticaConCalc.csv - OpenOffice.org Calc'. The menu bar includes Archivo, Editar, Ver, Insertar, Formato, Herramientas, Datos, Ventana, and Ayuda. The toolbar contains various icons for file operations and editing. The formula bar shows the active cell is C7 and the formula is =SUMA(A1:A30). The spreadsheet grid has columns A through F and rows 1 through 14. Column A contains the values 9, 12, 7, 12, 12, 7, 10, 16, 5, 4, 8, 6, 9, and 12. Cell C7 is highlighted in blue and contains the value 274.

	A	B	C	D	E	F
1	9					
2	12					
3	7					
4	12					
5	12					
6	7					
7	10		274			
8	16					
9	5					
10	4					
11	8					
12	6					
13	9					
14	12					

Ahora, antes de acabar el cálculo de la media, vamos a hacer un pequeño experimento. Asegúrate de que la celda que hayas usado (para mi la C7) está seleccionada, y haz clic en la línea de entrada; en ese momento verás que el contenido de la celda se muestra de otra manera. A continuación, cambia la fórmula de la línea de entrada por esta:

=SUMA(A5:A8)

y pulsa **Entrar**. Debe aparecer 45, y si miras el contenido del rango A5:A8 verás que, en efecto, es:

$$12 + 7 + 10 + 16 = 45.$$

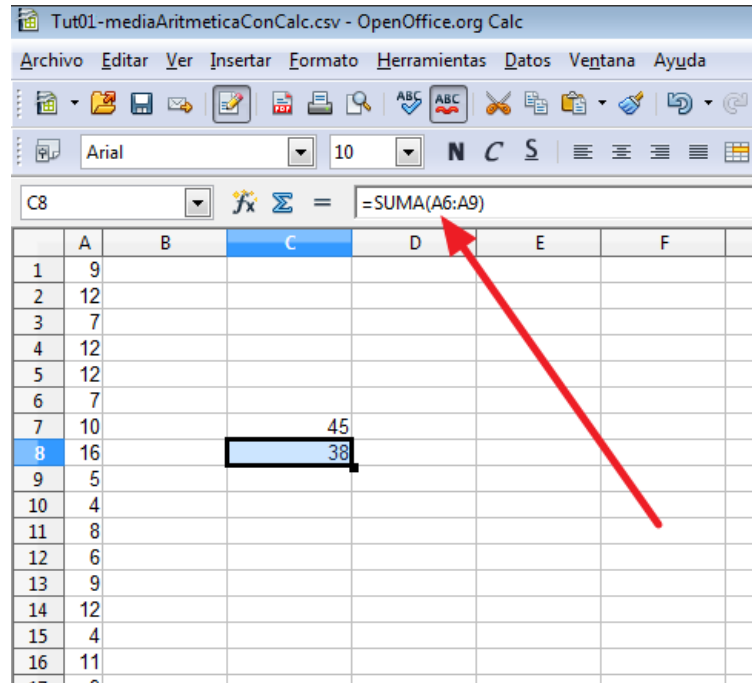
Como último paso del experimento, selecciona de nuevo la celda que hayas usado, cópiala y pégala (por ejemplo con CTRL+C, CTRL+V) en la de debajo (en mi caso, copio la C7 en C8). El resultado es 38.

Ejercicio 4. *¿Por qué ha sucedido esto?*

□

¡No sigas, hasta no haberlo intentado entender al menos!

Si no lo ves claro, la flecha roja de la figura te da la pista que necesitas para entender lo que ha pasado:



Volviendo al cálculo de la media, el resultado es, ahora, inmediato. Puedes usar esa cualquier celda, en la que introducimos esta fórmula:

`=SUMA(A1:A30)/30`

y al pulsar **Entrar** obtendrás 9.133 como valor de la media (con cuatro cifras significativas, recuerda el Capítulo 1).

Hay otra forma de calcular la media. Siendo una operación esencial en Estadística, Calc no puede sino incluir una función que calcula directamente la media de los valores de cierto rango. Esa función, lamentablemente, no se llama **MEDIA** o algo parecido, sino **PROMEDIO** (en inglés la situación es parecida; en lugar de **MEAN**, se llama **AVERAGE**). Para ver a esta función actuando, selecciona cualquier celda (yo he usado C10), y escribe la fórmula:

`=PROMEDIO(A1:A30)`

El resultado (sorpresa, sorpresa) es 9.133.

Nos imaginamos que puedes estar preguntándote, ¿y si puedo calcular la media en un sólo paso con **PROMEDIO**, a cuento de qué nos has estado enredando con sumas por aquí, sumas por allá? La respuesta, en el próximo apartado.

Debes guardar esta hoja de cálculo (¡cuidado! ¿qué debes recordar?), porque vamos a trabajar con ella varias veces en esta sesión, y usaremos el resultado que hemos obtenido. Pero te recomiendo que borres el 274 de A31 para evitar errores más adelante. Basta con que selecciones esa celda y pulses **Supr.** Recuerda grabar los cambios, nosotros hemos llamado al fichero **Tut01-mediaAritmeticaConCalc.ods**. Pero no cierres la hoja de cálculo, porque vamos a seguir usándola a continuación.

4.2. Media aritmética a partir de una tabla de frecuencias.

Vamos a calcular la media aritmética a partir de una tabla de frecuencias, así que empezamos fabricando una.

Ejercicio 5.

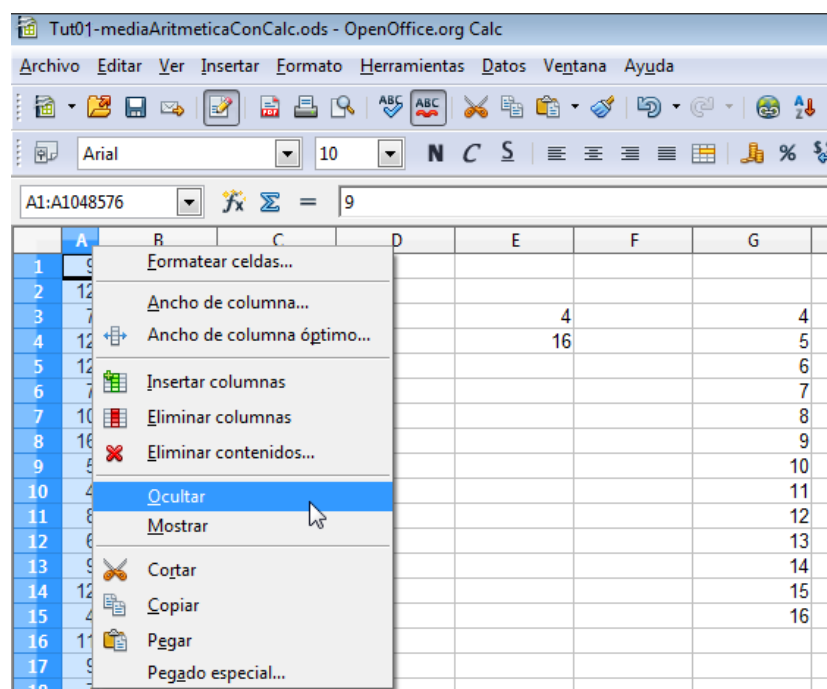
Recuerda lo que aprendimos en la Sección 1, para obtener una tabla de frecuencias de los treinta valores de la hoja **Tut01-mediaAritmeticaConCalc.csv** que estamos usando. ¡Manos a la obra, ya sabes hacerlo! □

¡No sigas, hasta que tengas esa tabla de frecuencias!

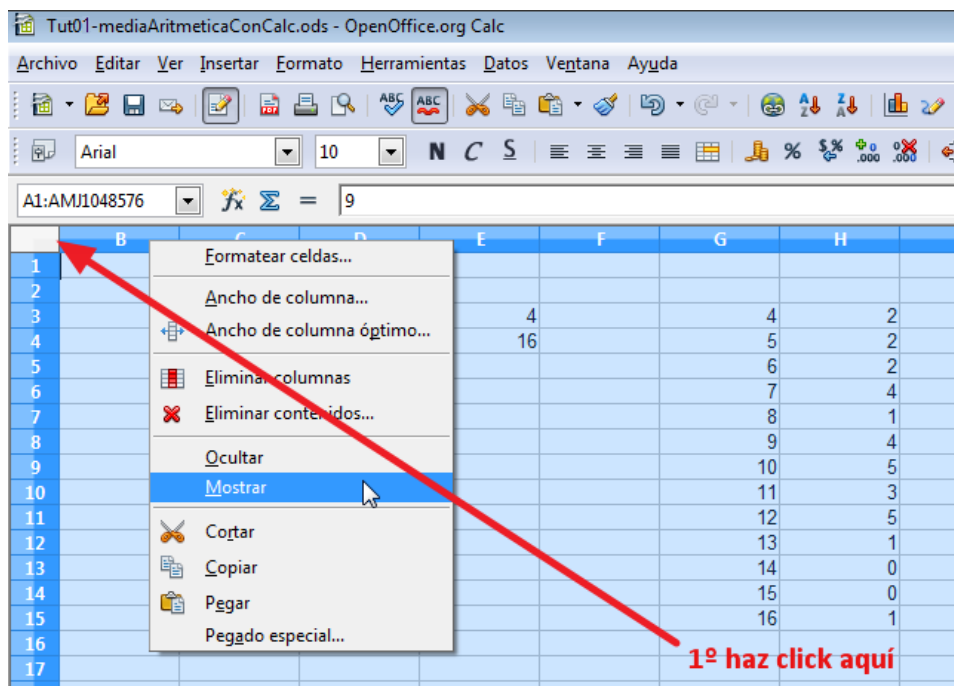
El resultado está en esta figura, en el rango G3:H15:

	A	B	C	D	E	F	G	H	I
1	9								
2	12								
3	7				4		4	2	
4	12				16		5	2	
5	12						6	2	
6	7						7	4	
7	10		9,133333333				8	1	
8	16						9	4	
9	5						10	5	
10	4		9,133333333				11	3	
11	8						12	5	
12	6						13	1	
13	9						14	0	
14	12						15	0	
15	4						16	1	
16	11								
17	a								

Ahora vamos a imaginar que no tenemos los datos originales, sólo esta tabla de frecuencias, y vamos a calcular la media a partir de esa información. Para aprender un truco nuevo, y hacer más realista el experimento, haz clic con el botón derecho sobre la letra A que encabeza esa columna, y luego haz clic en Ocultar:



Al hacerlo verás que la columna A parece haber desaparecido. En realidad sigue ahí (y las operaciones que dependen de ella no se ven afectadas), pero no es visible. Puedes hacerla aparecer ahora mismo con *deshacer* (recuerda Ctrl+Z, Ctrl+Y). O, más adelante, haciendo clic en la esquina superior izquierda (se seleccionará toda la hoja de cálculo), y haciendo después clic con el botón derecho en la columna B para seleccionar **mostrar** en el menú contextual que aparece.



Volvamos al tema de la media. La fórmula que vamos a usar, en este caso, es esta:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_k \cdot f_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^k x_i \cdot f_i}{\sum_{i=1}^k f_i}$$

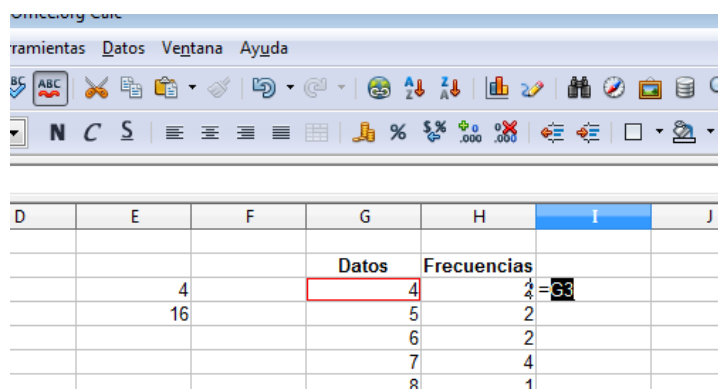
Es decir, que tenemos que dar estos cuatro pasos:

1. Multiplicar cada valor por la correspondiente frecuencia.
2. Sumar el resultado de todas esas multiplicaciones. Eso dará como resultado el numerador.
3. Sumar todas las frecuencias. Eso dará como resultado... ¿qué crees que saldrá? Lo haremos, sobre todo, para comprobar que no ha habido errores al hallar las frecuencias. Eso nos dará el denominador.
4. Dividir para obtener la media.

Pero antes, vamos a practicar la excelente costumbre de añadir comentarios a los datos, para evitar errores, y para que cuando volvamos a verlos (nosotros u otras personas), dentro de dos días, todavía podamos entender lo que nos traíamos entre manos. Insistiremos a lo largo del curso en la idea de que los análisis de datos no documentados son tan buenos como si no existieran. En mi caso, los datos están en la columna G y sus frecuencias en la H, así que añado rótulos descriptivos en la parte superior de la tabla. Basta hacer clic en cada celda y escribir la palabra o frase que queramos (en este caso sin el =, porque no es una fórmula).

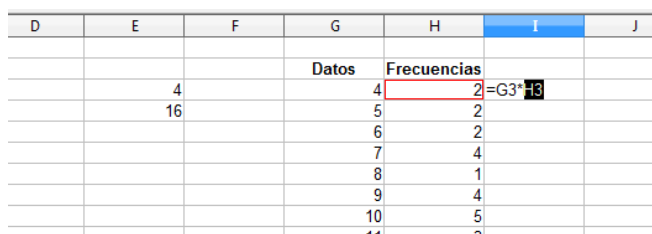
	A	B	C	D	E	F	G	H	I
1	9						Datos	Frecuencias	
2	12						4	2	
3	7				4		5	2	
4	12				16		6	2	
5	12						7	4	
6	7						8	1	
7	10		9,133333333				9	4	
8	16						10	5	
9	5						11	3	
10	4		9,133333333						

Ahora vamos con el primero de los cuatro pasos. En la columna I (conviene usar esta, para simplificar), hago clic en I3 y tecleo = para comenzar con una fórmula. Después de escribir = hago clic con el ratón en G3 (y veo que el nombre de esa celda aparece en H3, a continuación del =).



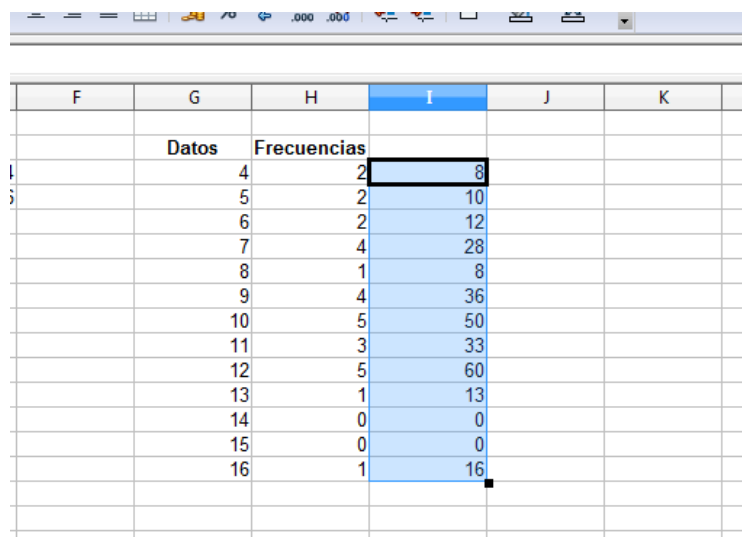
	D	E	F	G	H	I	J
				Datos	Frecuencias		
		4		4	2	=G3	
		16		5	2		
				6	2		
				7	4		
				8	1		

Sin hacer clic en ningún otro sitio, tecleo un asterisco (que representa la multiplicación), y acto seguido hago clic en H3, que aparece en I3.



	D	E	F	G	H	I	J
				Datos	Frecuencias		
		4		4	2	=G3*H3	
		16		5	2		
				6	2		
				7	4		
				8	1		
				9	4		
				10	5		
				11	3		
				12	5		
				13	1		
				14	0		
				15	0		
				16	1		

Lo que hemos hecho, en este caso, es introducir una fórmula seleccionando con el ratón las celdas que intervienen, en lugar de teclear sus nombres. Cuando te acostumbras, resulta una forma bastante cómoda de trabajar. Ya podemos pulsar **Entrar**, y ver el resultado, que es naturalmente 8. Lo que viene a continuación es algo ya conocido. Copiamos y pegamos la fórmula de I3 en las celdas del rango I4:I15 (por ejemplo, haciendo clic en el pequeño cuadrado negro de la esquina inferior derecha y arrastrando). El resultado debe ser este:



	F	G	H	I	J	K
		Datos	Frecuencias			
4		4	2	8		
5		5	2	10		
6		6	2	12		
7		7	4	28		
8		8	1	8		
9		9	4	36		
10		10	5	50		
11		11	3	33		
12		12	5	60		
13		13	1	13		
14		14	0	0		
15		15	0	0		
16		16	1	16		

El segundo paso consiste en sumar los valores que acabamos de obtener. Pero ya hemos aprendido varias formas de sumar los números de una columna. ¿Recuerda el lector que, al final del apartado anterior, le dijimos que ya veríamos la necesidad de aprender a hacer estas sumas? Ahora es el (primer) momento en que las necesitamos. En cualquier caso, el mejor sitio para colocar esa suma es, probablemente, la celda I17 debajo, pero no pegado a los datos, para distinguirlo y poder rotularlo. El resultado es 274, que es el numerador de la fórmula de la media. Hemos añadido a la figura varias flechas, para señalar varios aspectos en los que creemos que debes reparar.

	D	E	F	G	H	I	J
				Datos	Frecuencias	Dato*Frec	
		4		4	2	8	
		16		5	2	10	
				6	2	12	
				7	4	28	
9,13333333				8	1	8	
				9	4	36	
				10	5	50	
9,13333333				11	3	33	
				12	5	60	
				13	1	13	
				14	0	0	
				15	0	0	
				16	1	16	
					Suma:	274	

En el tercer paso sumamos las frecuencias. Después de la discusión previa, no sorprenderá saber que el resultado lo vamos a colocar en H17. Hacemos clic en esa celda e introducimos (de la manera que más te guste) la fórmula:

=SUMA(H3:H15)

El resultado (¡como no podía ser de otro modo! ¿por qué?) es 30, el denominador de la fórmula. Y, finalmente, tenemos que hacer la división para obtener la media. Hacemos clic, por ejemplo, en I21 e introducimos la fórmula:

=I17/H17

Al pulsar **Entrar** no hay sorpresa, el resultado es 9,133, como ya obtuvimos en el apartado anterior.

	A	B	C	D	E	F	G	H	I
1	9						Datos	Frecuencias	Dato*Frec
2	12						4	2	8
3	7				4		5	2	10
4	12				16		6	2	12
5	12						7	4	28
6	7						8	1	8
7	10		9,13333333				9	4	36
8	16						10	5	50
9	5						11	3	33
10	4		9,13333333				12	5	60
11	8						13	1	13
12	6						14	0	0
13	9						15	0	0
14	12						16	1	16
15	4						Suma:	30	Suma:
16	11								274
17	9								
18	7								
19	9								
20	7								
21	10								Media
22	6								9,13333333

Acuérdete de guardar el fichero Tut01-mediaAritmeticaConCalc.ods con estos cálculos, lo vamos a necesitar más adelante.

4.3. El caso de valores agrupados.

En las *Observaciones adicionales* de la página 10 hemos dejado pendiente la tarea de obtener la tabla de frecuencia de las variable **var2** en la tabla de datos Tut01-PracticaConCalc.csv. Ese va a ser el trabajo de la próxima sección. Pero, antes de hacer esto, vamos a pedir al lector que practique lo que ya hemos aprendido sobre el cálculo de la media aritmética.

Ejercicio 6. Carga el archivo Tut01-PracticaConCalc.ods, con la tabla de frecuencias que ya hicimos para var3, y calcula la media de esa variable por los dos métodos que hemos visto (con y sin tabla de frecuencia). Debes obtener 5.040 (con cuatro cifras significativas). Si quieres guardar el resultado, hazlo con otro nombre, porque volveremos a usar el archivo Tut01-PracticaConCalc.ods.

□

Si has hecho el ejercicio que se pedía al final de la anterior sesión, tendrás cargado en Calc el fichero Tut01-PracticaConCalc.ods. Y seguramente has ocupado algunas columnas con los cálculos de la media de var3. Nosotros vamos a empezar otra vez a partir del fichero csv original (que está adjunto en la página 1), así que, si quieres, haz lo mismo (y luego grabaremos nuestro trabajo de esta sección en formato ods, con otro nombre). Nuestro punto de partida, por tanto, es así:

	A	B	C	D	E	F	G	H	I	J
1	var1	var2	var3							
2	A	54,717	4							
3	E	52,676	8							
4	A	7,278	4							
5	E	1,253	4							
6	C	24,436	5							
7	B	82,398	5							
8	F	94,411	3							
9	E	17,865	6							
10	D	27,52	6							
11	F	14,274	2							
12	A	61,88	4							
13	A	22,722	4							

Lo primero que tenemos que hacer es calcular el rango de var2, es decir, el mínimo y máximo de esa variable. Puesto que los valores de var2 ocupan las celdas B2:B1301, es fácil obtener estos valores usando MIN y MAX:

	A	B	C	D	E	F	G	H	I	J
1	var1	var2	var3							
2	A	54,717	4							
3	E	52,676	8							
4	A	7,278	4							
5	E	1,253	4		Mínimo de var2 1,249					
6	C	24,436	5		Máximo de var2					
7	B	82,398	5		100,965					
8	F	94,411	3							
9	E	17,865	6							
10	D	27,52	6							
11	F	14,274	2							
12	A	61,88	4							
13	A	22,722	4							

Antes de seguir adelante, nosotros hemos grabado el fichero en formato ods, con nombre

Ahora tenemos que dividir el rango en intervalos (clases). ¿Cuántas clases? No vamos a dar, como hemos dicho en la teoría, reglas fijas para esto: ni muchas, ni pocas. En este caso en particular, vamos a usar diez clases. Concretando, los diez intervalos que vamos a usar son estos (cada desigualdad define un intervalo):

$$x \leq 11, \quad 11 < x \leq 21, \quad 21 < x \leq 31, \quad \dots \quad 91 < x \leq 101.$$

En este paso es fundamental asegurarse de que los intervalos cubren todo el rango de los valores. Por eso hemos llegado a 101.

¿Cómo le explicamos a Calc que queremos usar estos intervalos? Afortunadamente es muy fácil. Basta con colocar en una columna los extremos de los intervalos, así:

B	C	D	E	F	G	H
	var3					
1,717	4					
2,676	8					
7,278	4		Minimo de var2			
1,253	4		1,249		11	
1,436	5		Máximo de var2		21	
2,398	5		100,965		31	
1,411	3				41	
7,865	6				51	
27,52	6				61	
1,274	2				71	
31,88	4				81	
2,722	4				91	
5,965	3				101	
1,324	3					

Los extremos de los intervalos ocupan las celdas del rango G5:G14. A continuación, usamos la función FRECUENCIA, con la que ya estamos familiarizados, para obtener la tabla de frecuencias en el rango adyacente H5:H14. No olvides que antes de usar FRECUENCIA tienes que haber seleccionado *todas* las celdas del rango H5:H14. De lo contrario, se producirán errores. Las dos siguientes figuras resumen el proceso. El cuadro de diálogo de la función es:

The screenshot shows the 'Asistente de función' (Function Wizard) dialog box for the 'FRECUENCIA' (FREQUENCY) function. The 'Funciones' (Functions) tab is active, and 'FRECUENCIA' is selected from the list. The 'Categoría' (Category) is set to 'ESTADÍSTICAS' (STATISTICS). The 'datos' (data) range is 'B2:B1301' and the 'grupos' (bins) range is 'G5:G14'. The formula bar displays '=FRECUENCIA(B2:B1301;G5:G14)'. The background shows a spreadsheet with data in columns B and C, and a blue selection in column H.

Y el resultado es:

fx = {=FRECUENCIA(B2:B1301;G5:G14)}

B	C	D	E	F	G	H	I
var2	var3						
54,717	4						
52,676	8						
7,278	4		Minimo de var2				
1,253	4		1,249		11	117	
24,436	5		Máximo de var2		21	131	
82,398	5		100,965		31	137	
94,411	3				41	131	
17,865	6				51	125	
27,52	6				61	132	
14,274	2				71	137	
61,88	4				81	117	
22,722	4				91	117	
95,965	3				101	156	
39,324	3						
7,697	3						

Hay 131 valores x con $31 < x \leq 41$

Para entender el resultado lo mejor es un ejemplo: la celda G8 contiene el valor 40 (y, por supuesto, G7 contiene 30). Así que Calc coloca en la celda H8 el número de datos (del total de 1300) que caen en el intervalo:

$$31 < x \leq 41,$$

que, en este ejemplo, resultan ser 131. Y hace lo mismo con todas las demás celdas de la tabla de frecuencia. Sólo la primera, H5, es especial, porque en ese caso no hay límite inferior, y puesto que G5 contiene 10, Calc coloca ahí los valores del intervalo

$$x \leq 11.$$

Ya tenemos la tabla de frecuencias de var2. ¡No olvides rotular las columnas de la tabla! El siguiente paso, para calcular la media, es obtener las marcas de clase para cada uno de los intervalos. Las vamos a colocar en la siguiente columna, en el rango I5:I14. Vamos a empezar con un caso fácil, y dejamos la complicación para el final. En el Capítulo 2 hemos dicho que la marca de clase de un intervalo de la forma $(a, b]$ es el valor:

$$\frac{a + b}{2}.$$

Por ejemplo, para nuestro intervalo $(31, 41]$, es decir $31 < x \leq 41$, eso significa que la marca de clase es:

$$\frac{31 + 41}{2} = 36.$$

Para obtener las marcas de clase en Calc, en todos los casos salvo en el del primer intervalo, usamos la traducción directa de estas operaciones al lenguaje de Calc. Es decir, en I6 colocamos la fórmula:

$$=(G6+G5)/2$$

y copiamos esta fórmula en I7:I14. En la figura se muestra el resultado.

=(G7+G6)/2

	E	F	G	H	I	J
	Minimo de var2	Datos	Frecuencias	MarcaClase		
	1,249		11	117		
	Máximo de var2		21	131	16	
	100,965		31	137	26	
			41	131	36	
			51	125	46	
			61	132	56	
			71	137	66	
			81	117	76	
			91	117	86	
			101	156	96	

¿Y para el primer intervalo, que es $x \leq 11$? Bueno, si se observan nuestros datos, veremos que son todos positivos. Así que podríamos asumir que el intervalo es:

$$0 < x \leq 11.$$

Esto no está mal, aunque tiene el inconveniente de que todos los demás intervalos son de longitud 10, y este sería de longitud 11. No hay, en principio, ninguna regla que nos obligue a tomar todos los intervalos de la misma longitud. Pero, puesto que el mínimo de los datos es 1.249, no hay inconveniente alguno en usar como intervalo:

$$1 < x \leq 11,$$

y obtener como marca de clase 6. Para ello introducimos en I5 la fórmula:

$$=(G5+1)/2$$

A partir de aquí el resto de las operaciones son conocidas: multiplicamos las marcas de clase por las frecuencias, sumamos estos productos (la suma es el numerador de la media), después sumamos las frecuencias (para comprobar que en el denominador se obtiene 1300, ¡no se te ocurra saltarte esto y usar directamente 1300!), y finalmente hacemos la división. La figura muestra el resultado de todos estos pasos:

Σ = =J16/H16								
D	E	F	G	H	I	J	K	L
	<u>Mínimo de var2</u>		<u>Datos</u>	<u>Frecuencias</u>	<u>MarcaClase</u>	<u>Marca*Frec</u>		
	1,249		11	117	6	702		
	<u>Máximo de var2</u>		21	131	16	2096		
	100,965		31	137	26	3562		
			41	131	36	4716		
			51	125	46	5750		
			61	132	56	7392		
			71	137	66	9042		
			81	117	76	8892		
			91	117	86	10062		
			101	156	96	14976		
			Suma			Suma		
				1300		67190		
					Media	51.68461538		

Y la media que se obtiene es 51.68 (con cuatro cifras significativas).

Naturalmente, también podemos calcular la media sin agrupar, directamente a partir de los datos. Podemos sumar toda el rango B2:B1301 y dividir por 1300, o bien podemos aplicar directamente la función PROMEDIO a ese mismo rango.

Ejercicio 7.

Calcula la media aritmética por esos dos procedimientos.

□

¡No sigas, si no has hecho este ejercicio!

El resultado, el mismo en ambos casos, es 51.77 con cuatro cifras significativas. ¡Y no coincide con el que hemos obtenido a partir de la tabla de frecuencias! Y no es que nos hayamos equivocado, es así. Es el momento de pararse a pensar por qué sucede esto. Y si al cabo de un rato no llega la inspiración, hay que releer el Ejemplo 2.1.4 (pág. 24) del libro.

5. Medidas de posición: mediana, percentiles, moda.

Vamos a continuar nuestro trabajo, aprendiendo lo necesario sobre la forma de obtener, en Calc, las medidas de posición (mediana, cuantiles, percentiles) y las tablas de frecuencia relativas y acumuladas. Aparte de su interés en Estadística, estas operaciones nos van a brindar nuevas oportunidades de practicar las operaciones propias de la hoja de cálculo.

La mediana

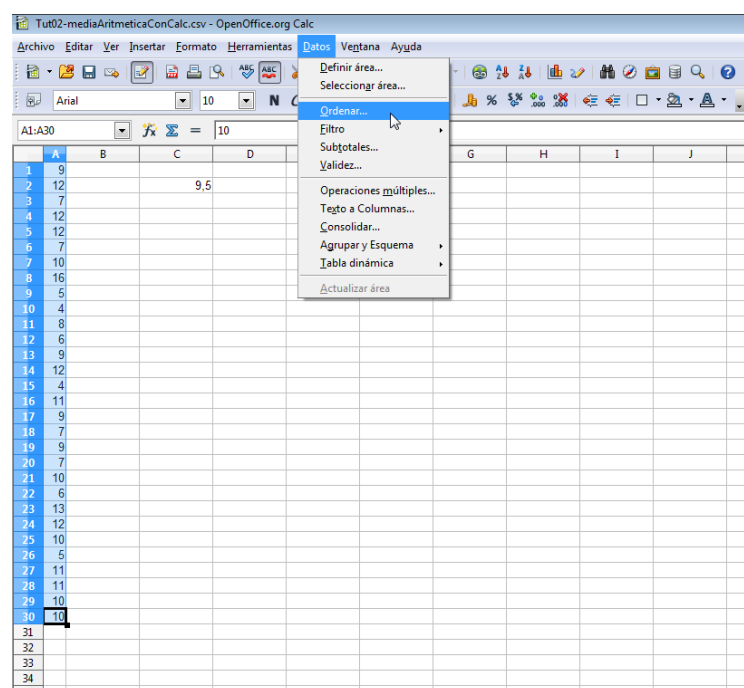
Vamos a volver a trabajar sobre el fichero (adjunto en la página 22)

Tut01-mediaAritmeticaConCalc.csv,

que hemos usado en las dos secciones anteriores. Lo cargamos en Calc, y en la celda C2 introducimos la fórmula

=MEDIANA(A1:A30)

Obtendrás 9.5 como resultado. Para ver por qué, es necesario ordenar los datos del rango A1:A30. Selecciona ese rango con el ratón y usa el menú **Datos** → **Ordenar**.



En el cuadro de diálogo que aparece, asegúrate de que aparece seleccionada la columna A y el sentido **ascendente** (de menor a mayor). Pulsa aceptar y obtendrás una lista de valores ordenados en el rango A1:A30. Puesto que tenemos $n = 30$ valores, debemos mirar a las celdas A15, que contiene 9, y A16, que contiene 10, para comprender porque la mediana es 9.5. Dos comentarios sobre la herramienta de ordenación de Calc:

- En el futuro puede que quieras ordenar una *tabla*, con varias columnas de datos, en la que todos los datos de una fila forman una cierta unidad, y por lo tanto las filas deben conservarse para no perder información. En ese caso, antes de usar **ordenar**, debes asegurarte de que toda la tabla está seleccionada, y después debes elegir cuál de las columnas de la tabla se usa para ordenarla.
- Recuerda que puedes deshacer (y rehacer) esta operación si es necesario.

Cuartiles y percentiles.

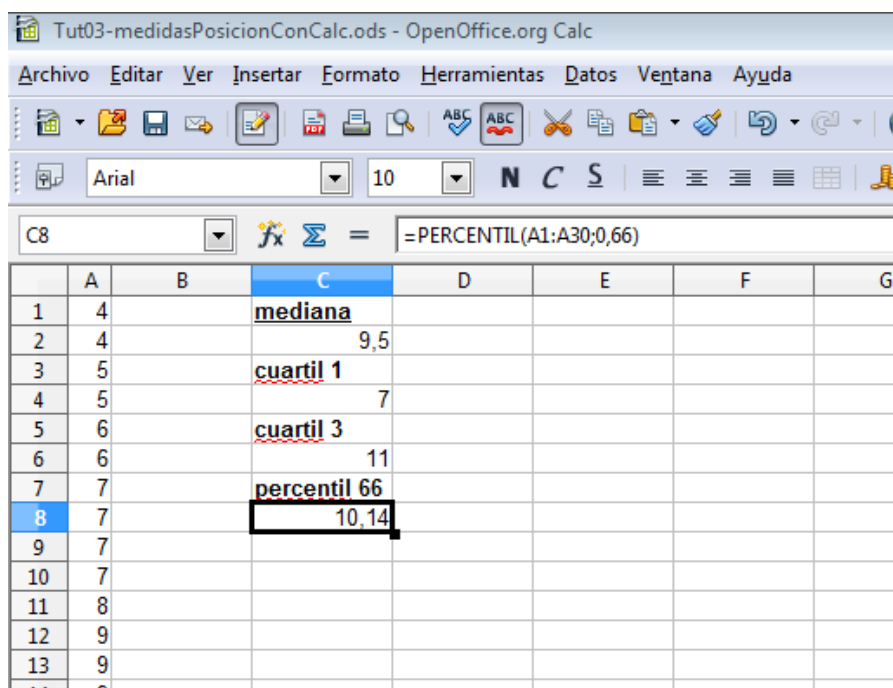
Ahora, con los datos todavía ordenados, vamos a preguntarnos cuál sería el primer cuartil. La idea informal es que, al igual que la mediana deja por debajo el 50 % de los datos, el primer cuartil debería dejar por debajo el 25 %. Y puesto que $30/4 = 7.5$, miramos las celdas A7 y A8. Ambas contienen un 7, así que es de esperar que ese primer cuartil valga 7. Y en efecto, si tecleamos en la celda C4 la fórmula:

`=CUARTIL(A1:A30;1)`

obtendremos como respuesta 7. Fíjate en que la fórmula tiene dos argumentos, separados por punto y coma. El primero es el rango del que calculamos el cuartil, y el segundo es el tipo de cuartil que calculamos. Usamos 1 para el primer cuartil, 3 para el tercer cuartil, y 2 para la mediana. De la misma forma se deduce que el tercer cuartil es 11. Ahora vamos a buscar un percentil, por ejemplo el percentil 66. Con la misma idea informal que antes, este valor deja por debajo el 66 % de los valores. Y puesto que $30 \cdot 0.66 = 19.8$, miramos las celdas A19 y A20, que contienen ambas 10. ¿Cuál crees que va a ser el percentil 66? Preguntemos a Calc. Introducimos, en Calc la fórmula

`=PERCENTIL(A1:A30;0,66)`

Ahora nos debemos fijar en que los porcentajes se indican como “tantos por uno” (y recuerda que desdichadamente Calc, en español, sigue la costumbre de usar la coma para los decimales). El resultado, como muestra la figura, es 10.14.



	A	B	C	D	E	F	G
1	4		mediana				
2	4		9,5				
3	5		cuartil 1				
4	5		7				
5	6		cuartil 3				
6	6		11				
7	7		percentil 66				
8	7		10,14				
9	7						
10	7						
11	8						
12	9						
13	9						
14	9						

¿De dónde sale un número como 10.14? Nos hemos entretenido en este ejemplo, sobre todo para hacer al lector consciente de esta situación. Queremos advertir, desde el principio, que el cálculo de los cuartiles y percentiles no es tan sencillo como parece sugerir la idea informal con la que hemos empezado el trabajo. Hay muchas formas distintas de definir los percentiles (por ejemplo, el programa R ofrece nueve formas distintas de calcularlos), dependiendo del fin que se les vaya a dar. En la Sección 2.2.3 (pág. 32) del libro puedes encontrar más referencias sobre este asunto.

Antes de abandonar este apartado, sólo queremos comentar que en Calc existe una función MODA, cuyo funcionamiento debería ser evidente.

5.1. Tablas de frecuencias relativas y acumuladas.

Para este apartado usaremos el fichero

`Tut01-PracticaConCalc.ods`,

que obtuvimos en la Sección 1. En ese fichero tiene que estar guardada la tabla de frecuencias de una variable `var3`, que ocupa el rango C2:C1301 de la hoja de cálculo. Por eso, al abrir el fichero nos encontraremos en una situación parecida a esta:

Tut01-PracticaConCalc.ods - OpenOffice.org Calc

Archivo Editar Ver Insertar Formato Herramientas Datos Ventana Ayuda

Arial 10

A1 Σ = var1

	A	B	C	D	E	F	G	H	I	J	K	L
1	var1	var2	var3				valor	frecuencia				
2	A	54.717	4				0	9				
3	E	52.676	8				1	25				
4	A	7.278	4	máx var 3=	16		2	100				
5	E	1.253	4	mín var 3=	0		3	188				
6	C	24.436	5				4	244				
7	B	82.398	5				5	246				
8	F	94.411	3				6	186				
9	E	17.865	6				7	131				
10	D	27.52	6				8	87				
11	F	14.274	2				9	43				
12	A	61.88	4				10	28				
13	A	22.722	4				11	6				
14	C	95.965	3				12	2				
15	B	39.324	3				13	2				
16	D	7.697	3				14	2				
17	C	90.413	2				15	0				
18	C	27.803	6				16	1				
19	E	3.667	4									

Vamos a empezar por calcular las frecuencias relativas en el rango I2:I18. Esto es muy fácil, porque cada una de esas frecuencias es igual a la correspondiente frecuencia absoluta, dividida por el número de datos (1300). Así que introducimos en I2 la fórmula

$$=H2/1300$$

y copiamos esa fórmula en el resto del rango I2:I18. No olvides rotular la columna I. El resultado será algo como:

Tut01-PracticaConCalc.ods - OpenOffice.org Calc

Archivo Editar Ver Insertar Formato Herramientas Datos Ventana Ayuda

Arial 10

I1 Σ = frec. relativa

	A	B	C	D	E	F	G	H	I	J	K	L
1	var1	var2	var3				valor	frecuencia	frec. relativa			
2	A	54.717	4				0	9	0,0069230769			
3	E	52.676	8				1	25	0,0192307692			
4	A	7.278	4	máx var 3=	16		2	100	0,0769230769			
5	E	1.253	4	mín var 3=	0		3	188	0,1446153846			
6	C	24.436	5				4	244	0,1876923077			
7	B	82.398	5				5	246	0,1892307692			
8	F	94.411	3				6	186	0,1430769231			
9	E	17.865	6				7	131	0,1007692308			
10	D	27.52	6				8	87	0,0669230769			
11	F	14.274	2				9	43	0,0330769231			
12	A	61.88	4				10	28	0,0215384615			
13	A	22.722	4				11	6	0,0046153846			
14	C	95.965	3				12	2	0,0015384615			
15	B	39.324	3				13	2	0,0015384615			
16	D	7.697	3				14	2	0,0015384615			
17	C	90.413	2				15	0	0			
18	C	27.803	6				16	1	0,0007692308			
19	E	3.667	4									
20	B	82.971	5									
21	D	12.873	2									
22	C	24.736	5									

Antes de seguir vamos a comprobar que las cosas parecen bien hechas. Sitúate en I20 y calcula la suma de las frecuencias relativas. Si lo hemos hecho bien, esa suma debe ser 1.

Ahora vamos con las frecuencias acumuladas, que son algo más difíciles. Las vamos a situar en el rango J2:J18. Vamos a recordar que si $f_1, \dots, f_k$ son las frecuencias absolutas, y $f'_1, \dots, f'_k$ son las acumuladas, se cumple esta relación (ver el Capítulo 2, página 27 del libro):

$$f'_1 = f_1, \quad f'_2 = f_2 + f'_1, \quad f'_3 = f_3 + f'_2, \dots, f'_k = f_k + f'_{k-1}.$$

Esta relación es la que vamos a usar en Calc. En la celda J2 escribimos la fórmula:

$$=H2$$

porque en H2 es donde está f_1 . Y ahora, en J3 vamos a usar la relación:

$$f'_2 = f_2 + f'_1$$

para obtener f'_2 . Ya sabemos que f_2 está en H3. ¿Dónde está f'_1 ? Lo acabamos de poner en J2. Así que la fórmula es:

$$=H3+J2$$

Lo bueno de proceder así es que ahora esa fórmula se puede copiar desde J3 a las restantes celdas del rango (es decir, a J4:J18) y se obtienen todas las frecuencias acumuladas. El resultado es:

Tut01-PracticaConCalc.ods - OpenOffice.org Calc

Archivo Editar Ver Insertar Formato Herramientas Datos Ventana Ayuda

Arial 10 N C S

J18 $\sum$ = =H18+J17

	A	B	C	D	E	F	G	H	I	J	K
1	var1	var2	var3				valor	frecuencia	frec. relativa		
2	A	54,717	4				0	9	0,0069230769	9	
3	E	52,676	8				1	25	0,0192307692	34	
4	A	7,278	4	máx var 3=	16		2	100	0,0769230769	134	
5	E	1,253	4	mín var 3=	0		3	188	0,1446153846	322	
6	C	24,436	5				4	244	0,1876923077	566	
7	B	82,398	5				5	246	0,1892307692	812	
8	F	94,411	3				6	186	0,1430769231	998	
9	E	17,865	6				7	131	0,1007692308	1129	
10	D	27,52	6				8	87	0,0669230769	1216	
11	F	14,274	2				9	43	0,0330769231	1259	
12	A	61,88	4				10	28	0,0215384615	1287	
13	A	22,722	4				11	6	0,0046153846	1293	
14	C	95,965	3				12	2	0,0015384615	1295	
15	B	39,324	3				13	2	0,0015384615	1297	
16	D	7,697	3				14	2	0,0015384615	1299	
17	C	90,413	2				15	0	0	1299	
18	C	27,803	6				16	1	0,0007692308	1300	
19	E	3,667	4						suma		
20	B	82,971	5						1		

y, naturalmente, la última de las frecuencias acumuladas es igual a $n = 1300$.

Ya sólo nos quedan las frecuencias relativas acumuladas. Dejamos este último paso como ejercicio para el lector. El resultado debe ser el de la figura:

Tut03-PracticaConCalc.ods - OpenOffice.org Calc

Archivo Editar Ver Insertar Formato Herramientas Datos Ventana Ayuda

Arial 10 N C S

S1 $\sum$ =

	A	B	C	D	E	F	G	H	I	J	K	L
1	var1	var2	var3				valor	frecuencia	frec. relativa	frec. acumulada	frec. rel. Acumulada	
2	A	54,717	4				0	9	0,0069230769	9	0,006923077	
3	E	52,676	8				1	25	0,0192307692	34	0,026153846	
4	A	7,278	4	máx var 3=	16		2	100	0,0769230769	134	0,103076923	
5	E	1,253	4	mín var 3=	0		3	188	0,1446153846	322	0,247692308	
6	C	24,436	5				4	244	0,1876923077	566	0,435384615	
7	B	82,398	5				5	246	0,1892307692	812	0,624615385	
8	F	94,411	3				6	186	0,1430769231	998	0,767692308	
9	E	17,865	6				7	131	0,1007692308	1129	0,868461538	
10	D	27,52	6				8	87	0,0669230769	1216	0,935384615	
11	F	14,274	2				9	43	0,0330769231	1259	0,968461538	
12	A	61,88	4				10	28	0,0215384615	1287	0,99	
13	A	22,722	4				11	6	0,0046153846	1293	0,994615385	
14	C	95,965	3				12	2	0,0015384615	1295	0,996153846	
15	B	39,324	3				13	2	0,0015384615	1297	0,997692308	
16	D	7,697	3				14	2	0,0015384615	1299	0,999230769	
17	C	90,413	2				15	0	0	1299	0,999230769	
18	C	27,803	6				16	1	0,0007692308	1300	1	
19	E	3,667	4						suma			
20	B	82,971	5						1			
21	D	12,873	2									
22	C	24,736	5									

Ejercicio 8.

1. Obtén esa tabla de frecuencias relativas acumuladas.
2. Usando esa tabla de frecuencias relativas acumuladas, calcula la mediana de var3.
3. Confirma la respuesta usando la función MEDIANA de Calc.

□

Cuando los datos se presentan agrupados por intervalos, los cálculos son similares, pero utilizando las marcas de clase. No entraremos en detalles.

6. Varianza y desviación típica.

Hemos recorrido ya una buena parte de nuestra introducción a la Estadística Descriptiva con la hoja de cálculo Calc. Para concluir este tutorial, vamos a aprender a calcular la varianza y desviación típica (poblacionales) con Calc. También aprenderemos a calcular la cuasivarianza y la cuasidesviación típica (muestrales). No te preocupes si todavía no entiendes por qué es necesaria esta diferencia entre poblacional y muestral, ya se aclarará más adelante. De momento, lo único importante es que sepas que hay dos tipos de objetos, y que te fijas en cuál estamos calculando en cada caso. Al principio del curso el protagonismo es para las cantidades poblacionales, pero poco a poco los muestrales irán ganando en importancia.

Sea $x = (x_1, \dots, x_n)$ un vector de datos (no agrupados). La fórmula de la varianza poblacional es:

$$Var(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}.$$

El valor $\bar{x}$ representa la media de x , así que este cálculo *presupone* que ya hemos calculado la media de x . Como en casos anteriores, podemos ver esta fórmula como una receta para realizar, en varios pasos, el cálculo correspondiente:

1. Como hemos dicho, debemos haber calculado la media $\bar{x}$.
2. Debemos restarle la media $\bar{x}$ a cada uno de los valores x_i .

$$(x_1 - \bar{x}), (x_2 - \bar{x}), \dots, (x_n - \bar{x}).$$

3. Después elevamos cada diferencia al cuadrado, para obtener los n valores

$$(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, \dots, (x_n - \bar{x})^2.$$

4. Sumamos los cuadrados y
5. Dividimos por n .

Como se ve, buena parte del trabajo consiste en aplicar la misma operación a toda una lista de valores, y eso hace que esta tarea sea fácil de abordar con una hoja de cálculo como Calc. Y, al igual que sucedía con la media, el cálculo se puede hacer de varias formas distintas, dependiendo del punto de partida y de nuestros deseos.

Para practicar con un ejemplo vamos a volver al mismo punto del comienzo de la Sección 4.1 (página 22), es decir, que volvemos a cargar el fichero

`Tut01-mediaAritmeticaConCalc.csv`,

que aparece adjunto en esa página. Cuidado: usamos el `csv`, no el fichero de tipo `ods` que obtuvimos al final de esa sección, y que vamos a usar en breve.

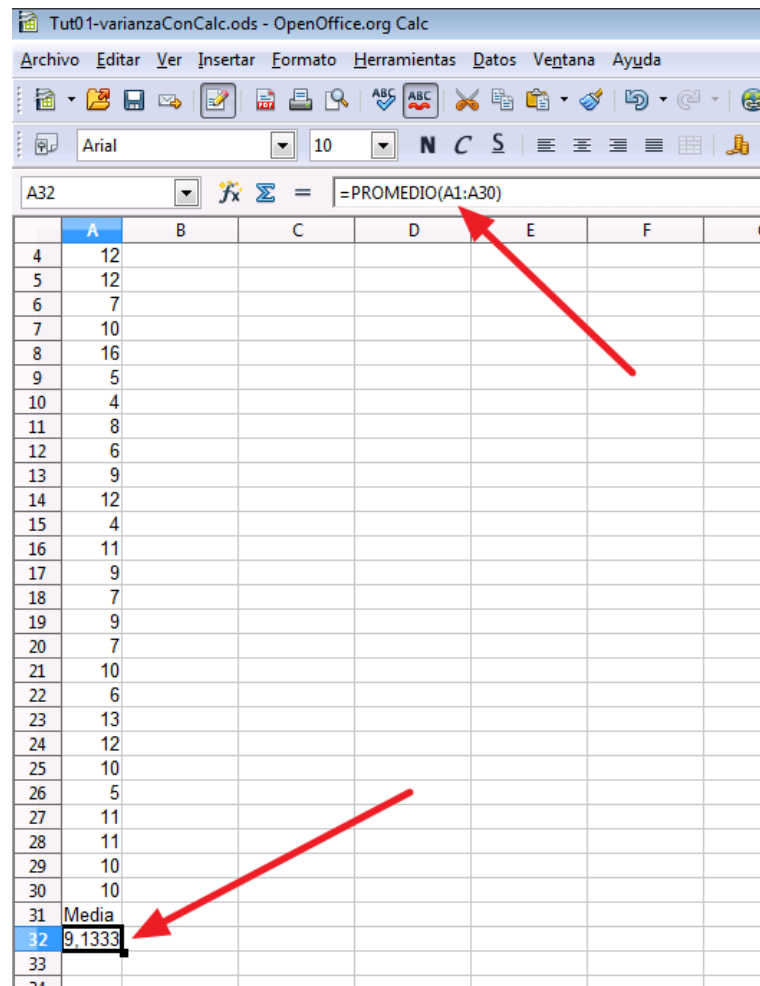
Una vez abierto el fichero, para el primer paso del método vamos a calcular la media otra vez, usando `PROMEDIO` para abreviar.

Ejercicio 9.

Abre el fichero `csv` y haz ese cálculo de la media. Coloca el resultado, con un rótulo, en las celdas `A31:A32`. □

¡No sigas, si no has hecho este ejercicio!

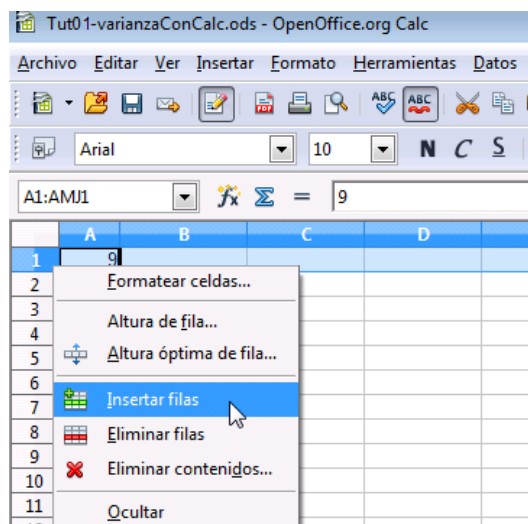
El resultado se muestra en la siguiente figura:



Obtendrás, como ya sabíamos, $\bar{x} = 9.133$. Guardamos el fichero en formato ods, con nombre, por ejemplo

Tut01-varianzaConCalc.ods.

Antes de seguir adelante, vamos a insistir en la necesidad de rotular y comentar bien nuestro trabajo. El problema es que, puesto que los datos empiezan en la celda A1 parece que nos hemos quedado sin sitio para incluir un rótulo. No hay problema, hagamos algo de sitio. Haz clic con el botón derecho del ratón en el número 1 al principio de la primera fila de la hoja de cálculo. En el menú que aparece haz clic en **Insertar filas**, y verás como aparece una fila vacía, y todo el contenido de la hoja se desplaza una posición hacia abajo.



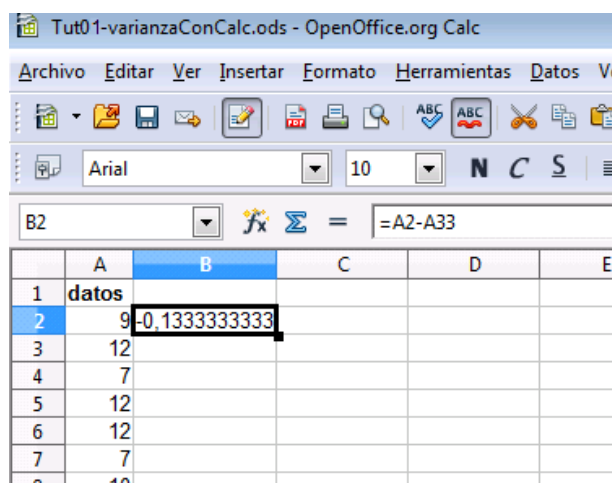
Y si necesitas insertar, por ejemplo, cinco filas vacías, empieza seleccionando cinco filas (si lo haces bien, verás sombreada completamente toda esas filas), antes de usar **Insertar filas**. Se insertarán tantas filas como filas tengas seleccionados. No te preocupes, cuando se inserta una fila, Calc actualiza todas las fórmulas automáticamente, así que, por ejemplo, el cálculo que hemos hecho de la media no se ve afectado. En concreto, puedes comprobar que ahora la media aparece en la celda **A33**, y se calcula con `=PROMEDIO(A2:A31)`, usando el nuevo rango que ocupan los datos. Escribamos en la celda **A1** el rótulo **datos**.

Referencias absolutas en la hoja de cálculo

Ahora vamos con el segundo paso, para calcular las diferencias $(x_i - \bar{x})$. Vamos a colocarlas en el rango **B2:B31**. Para calcular la primera, nos situamos en la celda **B2**, e introducimos la fórmula:

$$=A2-A33$$

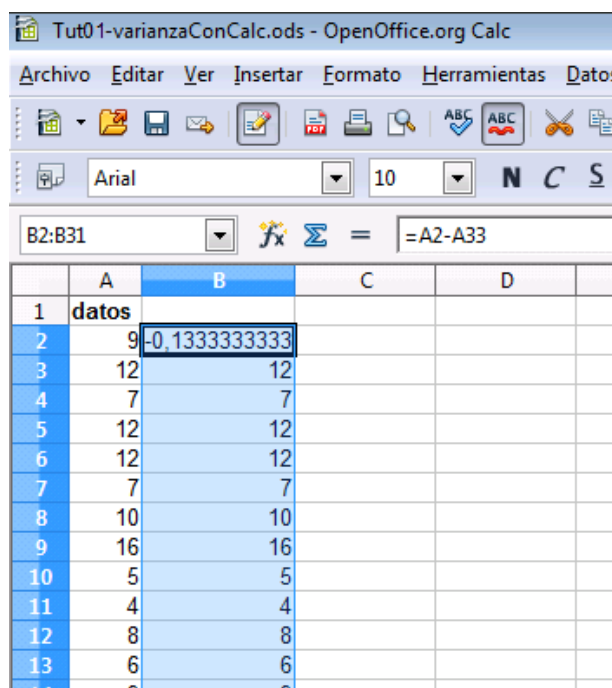
Recuerda que la media $\bar{x}$ está en **A33**. El resultado es $x_1 - \bar{x} = -0.1333$ (con cuatro cifras significativas):



The screenshot shows the OpenOffice Calc interface with the file 'Tut01-varianzaConCalc.ods'. The formula bar displays '=A2-A33' for cell B2. The spreadsheet shows the following data:

	A	B	C	D	E
1	datos				
2	9	-0.1333333333			
3	12				
4	7				
5	12				
6	12				
7	7				
8	10				

Ahora, como hemos hecho en casos anteriores, vamos a copiar esa fórmula a las restantes celdas del rango **B3:B31**. Recuerda que puedes pinchar en el pequeño cuadrado negro de la esquina inferior derecha de **B2** y arrastrar para cubrir todo el rango. Se obtiene (sólo se muestran las primeras filas):



The screenshot shows the same OpenOffice Calc interface, but now the range B3:B31 is selected, and the formula bar still shows '=A2-A33'. The spreadsheet shows the following data:

	A	B	C	D	E
1	datos				
2	9	-0.1333333333			
3	12	12			
4	7	7			
5	12	12			
6	12	12			
7	7	7			
8	10	10			
9	16	16			
10	5	5			
11	4	4			
12	8	8			
13	6	6			
14	10	10			

Una pequeña luz de alarma se debería encender en algún rincón de nuestra cabeza: si estamos restando 9.133, ¿cómo es que los resultados son números enteros, sin decimales? Ahora que algo

ha llamado nuestra atención, fíjate mejor: los resultados son los datos de partida, sin modificación alguna. ¡Y la razón por la que esto sucede es que estamos restando 0!

Enseguida vamos a ver lo que ha pasado, y a ponerle remedio. Pero queremos llamar la atención del lector sobre el hecho de que es necesario, imprescindible de hecho, preguntarse constantemente por las operaciones que hacemos, si son correctas, incluso tratando de anticipar el resultado para detectar posibles errores o problemas.

Veamos cuál ha sido el problema. Hagamos una pequeña comprobación. En la celda B3 esperamos tener el resultado de A3, menos la media 9.133, que está en A33. Es decir, que la fórmula para calcular B3 tiene que ser:

$$=A3-A33$$

Pero si te sitúas en B3, verás que la fórmula que aparece en esa celda es:

$$=A3-A34$$

De la misma forma, en B4 tenemos

$$=A4-A35$$

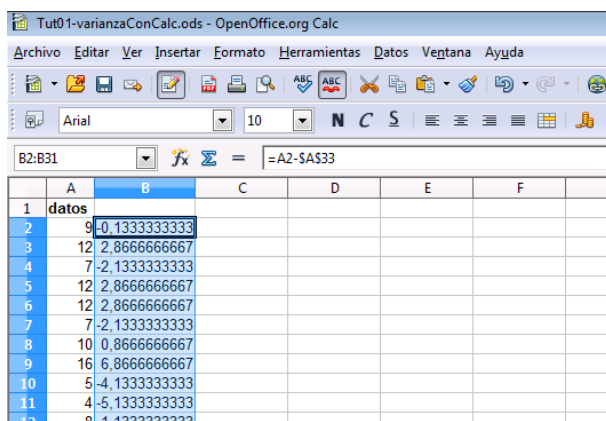
Ahora ya debería empezar a estar claro lo que ha pasado: hemos sido víctimas de la forma de actualizar las *referencias relativas* en una hoja de cálculo. Cuando copiamos la fórmula de B2 en la celda de abajo, todas las referencias a celdas que aparecen en la fórmula se sustituyen por referencias a las correspondientes celdas de abajo. Y la celda situada debajo de A33 es A34, que contiene nada, es decir un 0. Por eso hemos restado ceros.

De acuerdo, ese es el problema. ¿Y la solución? Calc normalmente utiliza *referencias relativas*, pero en ocasiones como esta necesitamos una forma de emplear *referencias absolutas*, que apunten a posiciones inmutables dentro de la hoja de cálculo. Afortunadamente, es muy sencillo hacer esto.

Empecemos por deshacer los últimos pasos (usando el icono  de la barra de herramientas de Calc, o con **Ctrl+Z**), hasta dejar la columna B libre. Y ahora, en B2 introducimos la fórmula:

$$=A2-\$A\$33$$

con dos símbolos \$ adicionales. Con cada uno esos símbolos le estamos diciendo a Calc que deje fijo el elemento correspondiente. Es decir \$A significa “no cambies la columna A”, y “\$33” significa “no cambies la fila 33”. Al introducir esa fórmula, el resultado en B3 es el mismo que antes (claro), pero cuando copiamos esa fórmula en el resto del rango B3:B31, las cosas cambian (a mejor):



	A	B	C	D	E	F	G
1	datos						
2	9	-0.1333333333					
3	12	2.866666667					
4	7	-2.133333333					
5	12	2.866666667					
6	12	2.866666667					
7	7	-2.133333333					
8	10	0.866666667					
9	16	6.866666667					
10	5	-4.133333333					
11	4	-5.133333333					
12	8	1.133333333					

Antes de avanzar, escribe en B1 un rótulo para esa columna, por ejemplo, **datos - media**.

De acuerdo, ya hemos superado este escollo, tenemos calculadas las diferencias $x_i - \bar{x}$ y podemos seguir con el tercer paso, elevando esas diferencias al cuadrado. ¿Cómo se hace esto en Calc? En la mayoría de los lenguajes informáticos, Calc incluido, la operación “elevar al cuadrado” se representa mediante la notación $\wedge 2$ (en algunos casos se utiliza $**2$). Así que en la celda C2 escribimos la fórmula:

$$=B2\wedge 2$$

y la copiamos en el rango C3:C31. No te olvides de rotular la columna C.

Tut01-varianzaConCalc.ods - OpenOffice.org Calc

Archivo Editar Ver Insertar Formato Herramientas Datos Ventana Ayuda

Arial 10

C1 $\sum$ = (datos-media)^2

	A	B	C	D	E
1	datos	datos - media	(datos-media)^2		
2	9	-0,1333333333	0,0177777778		
3	12	2,866666667	8,217777778		
4	7	-2,133333333	4,551111111		
5	12	2,866666667	8,217777778		
6	12	2,866666667	8,217777778		
7	7	-2,133333333	4,551111111		
8	10	0,866666667	0,751111111		
9	16	6,866666667	47,15111111		
10	5	-4,133333333	17,08444444		
11	4	-5,133333333	26,35111111		

Los últimos pasos son sencillos. Nos situamos en C33 (serviría cualquier celda libre, pero esta es conveniente) e introducimos la fórmula para sumar todos los valores del rango C2:C31. Es decir:

=SUMA(C2:C31)

Tut01-varianzaConCalc.ods - OpenOffice.org Calc

Archivo Editar Ver Insertar Formato Herramientas Datos Ventana Ayuda

Arial 10

C33 $\sum$ = =SUMA(C2:C31)

	A	B	C	D	E
10	5	-4,133333333	17,08444444		
11	4	-5,133333333	26,35111111		
12	8	-1,133333333	1,284444444		
13	6	-3,133333333	9,817777778		
14	9	-0,133333333	0,017777778		
15	12	2,866666667	8,217777778		
16	4	-5,133333333	26,35111111		
17	11	1,866666667	3,484444444		
18	9	-0,133333333	0,017777778		
19	7	-2,133333333	4,551111111		
20	9	-0,133333333	0,017777778		
21	7	-2,133333333	4,551111111		
22	10	0,866666667	0,751111111		
23	6	-3,133333333	9,817777778		
24	13	3,866666667	14,95111111		
25	12	2,866666667	8,217777778		
26	10	0,866666667	0,751111111		
27	5	-4,133333333	17,08444444		
28	11	1,866666667	3,484444444		
29	11	1,866666667	3,484444444		
30	10	0,866666667	0,751111111		
31	10	0,866666667	0,751111111		
32	Media		Suma		
33	9,1333		243,4666667		

El resultado, 243.5 (con cuatro cifras significativas) es el numerador de la varianza:

$$\sum_{i=1}^n (x_i - \bar{x})^2.$$

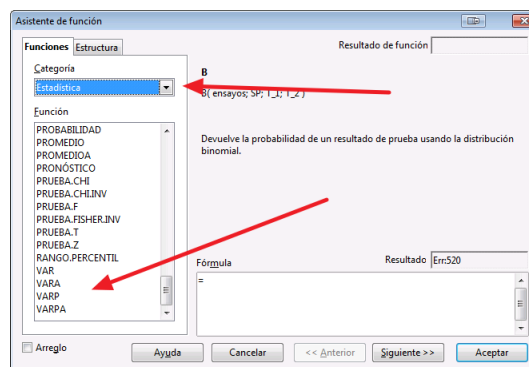
Para calcular la varianza sólo tenemos que dividir por 30, cosa que hacemos en la celda C35:

	A	B	C	D	E	F
10	5	-4,133333333	17,08444444			
11	4	-5,133333333	26,35111111			
12	8	-1,133333333	1,284444444			
13	6	-3,133333333	9,817777778			
14	9	-0,133333333	0,017777778			
15	12	2,866666667	8,217777778			
16	4	-5,133333333	26,35111111			
17	11	1,866666667	3,484444444			
18	9	-0,133333333	0,017777778			
19	7	-2,133333333	4,551111111			
20	9	-0,133333333	0,017777778			
21	7	-2,133333333	4,551111111			
22	10	0,866666667	0,751111111			
23	6	-3,133333333	9,817777778			
24	13	3,866666667	14,95111111			
25	12	2,866666667	8,217777778			
26	10	0,866666667	0,751111111			
27	5	-4,133333333	17,08444444			
28	11	1,866666667	3,484444444			
29	11	1,866666667	3,484444444			
30	10	0,866666667	0,751111111			
31	10	0,866666667	0,751111111			
32	Media		Suma			
33	9,1333		243,4666667			
34			Varianza			
35			8,115555556			
36						
37						

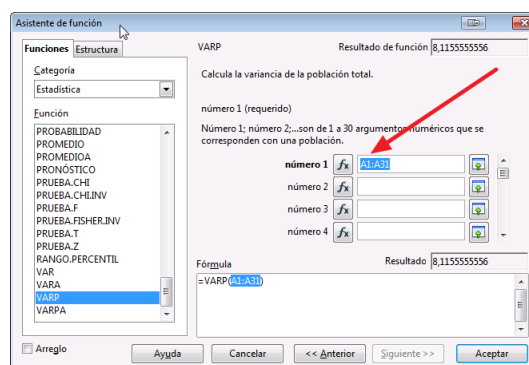
La varianza, por tanto, es 8.156 (con cuatro cifras significativas).

La varianza usando funciones propias de Calc

Hay otra forma de obtener ese resultado. Puesto que la varianza de un conjunto de datos es un valor que se calcula con mucha frecuencia, Calc incluye una función para obtenerla directamente. Nos situamos en C37 y usamos el menú **Insertar** → **Función**. En el cuadro de diálogo que aparece, en **Categoría** elegimos **Estadística** y entonces, en el cuadro **Función** navegamos hacia abajo, hasta el final de la lista de funciones, donde veremos varias funciones cuyo nombre empieza por **VAR**, como en esta figura (los nombres en inglés son los mismos):



La que nos interesa ahora es **VARP** (recuerda **VAR**ianza **PO**blacional). La que se llama **VAR** es la cuasivarianza muestral (con $n - 1$ en lugar de n en el denominador), que usaremos mucho más adelante en el curso. Las dos versiones con una *a* al final, las que se llaman **VARA** y **VARPA**, son variantes de estas, que sirven para los casos en los que el conjunto de datos contiene alguna omisión. Este problema de los *datos ausentes* (*missing data*, en inglés) es uno de los problemas prácticos más frecuentes en Estadística, y a veces influye enormemente en la manera adecuada de proceder. De momento, no obstante, sólo necesitamos **VARP**. La seleccionamos haciendo doble clic sobre su nombre, y en el siguiente paso la aplicamos al rango **A2:A31** así:



Obtenemos, en C37, el mismo resultado que ya teníamos en C35. Y el lector se volverá a preguntar, ¿y no podríamos habernos ahorrado todos estos pasos, entonces? La respuesta es que, en este caso, sí, podríamos haber usado directamente **VARP**. Pero cuando lo que tenemos, como punto de partida, no son los datos, sino una tabla de frecuencias, entonces es necesario seguir pasos parecidos a estos. Ese es, de hecho, casi todo el trabajo que nos falta por hacer en esta sección.

Desviación y cuasidesviación típica.

Pero, antes de ir a eso, un último paso. La *desviación típica* (poblacional) es la raíz cuadrada de la varianza (poblacional). La podemos obtener, por tanto, calculando la raíz cuadrada de C35 (o de C37). ¿Cómo se calcula una raíz cuadrada en Calc? Usando la función **RAIZ** (en inglés es **SQRT**, de *square root*). Dejamos como ejercicio para el lector comprobar por este método que la desviación típica es 2.849. Hay otra forma, usando la función **DESVESTP**, de desviación estándar poblacional. Hay cuatro funciones que empiezan por **DESVEST**, que son las raíces de las correspondientes **VAR** (sus análogas en inglés tienen nombres que empiezan por **STDEV**, de *standard deviation*). El siguiente ejercicio para el lector es comprobar que **DESVESTP** produce el mismo resultado que **RAIZ** aplicado a **VARP**.

Acuérdete de guardar tu trabajo de esta sección en un fichero, con formato **ods**, claro. Nosotros lo hemos llamado **Tut01-varianzaConCalc.ods**.

6.1. Varianza a partir de la tabla de frecuencias.

Volvamos a cargar el fichero **Tut01-mediaAritmeticaConCalc.ods**, que guardaste al final de la Sección 4.2, y que contiene una tabla de frecuencias para los 30 datos con los que hemos trabajado en aquella sección. Vamos a usar ahora esa tabla de frecuencias para calcular la varianza poblacional de los datos (que ya hemos calculado, y sabemos que vale aprox. 8.156). Cuando los datos se presentan así, en forma de una tabla de frecuencias, no existe una función de Calc, como **VARP**, que nos permita obtener la varianza a partir de la tabla. Así que lo que haremos en estos casos es usar la fórmula:

$$\frac{\sum_{i=1}^k f_i \cdot (x_i - \bar{x})^2}{\sum_{i=1}^k f_i} = \frac{f_1 \cdot (x_1 - \bar{x})^2 + \dots + f_n \cdot (x_n - \bar{x})^2}{f_1 + f_2 + \dots + f_n}.$$

La siguiente figura resume el trabajo que hay que hacer para conseguir el resultado. Dejamos al lector la tarea de reproducir estos valores. No hay nada nuevo aquí, sólo hay que poner juntas las piezas.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	9												
2	12												
3	7				4		Datos	Frecuencias	Dato*Frec	Dato-media	(Dato-Media)^2	Frec*(Dato-Media)^2	
4	12				16		5	2	10	-4,133333333	17,084444444	34,168888889	
5	12						6	2	12	-3,133333333	9,817777778	19,635555556	
6	7						7	4	28	-2,133333333	4,551111111	18,204444444	
7	10		9,133333333				8	1	8	-1,133333333	1,284444444	1,284444444	
8	16						9	4	36	-0,133333333	0,017777778	0,071111111	
9	5						10	5	50	0,866666667	0,751111111	3,755555556	
10	4		9,133333333				11	3	33	1,866666667	3,484444444	10,453333333	
11	8						12	5	60	2,866666667	8,217777778	41,088888889	
12	6						13	1	13	3,866666667	14,951111111	14,951111111	
13	9						14	0	0	4,866666667	23,684444444	0	
14	12						15	0	0	5,866666667	34,417777778	0	
15	4						16	1	16	6,866666667	47,151111111	47,151111111	
16	11						Suma:	30	Suma:	274		Suma	243,466666667
17	9												
18	7												
19	9												
20	7												
21	10								Media	9,133333333		Varianza	8,115555556
22	6												
23	13												

Hemos destacado en color rojo las novedades con respecto al cálculo de la media, que puedes usar como referencia para completar el primer apartado del siguiente ejercicio:

Ejercicio 10.

1. Repite las operaciones de la anterior figura, hasta completar el cálculo de la varianza.
2. Calcula la varianza y desviación típica de la variable `var3` del fichero

`Tut01-PracticaConCalc.csv`,

(adjunto en la página 1), aplicando la función `VARP` (y `DESVESTP`) directamente sobre los datos.

3. Vuelve a calcular esos mismos valores a partir de la tabla de frecuencias que obtuvimos en la Sección 1, y que debes tener guardada en el fichero

`Tut01-PracticaConCalc.ods`

Compara los resultados con los del apartado anterior.

□

7. Ejercicios adicionales y soluciones.

Ejercicios adicionales.

11. Queremos saber la edad media de los empleados de una empresa que tiene tres fábricas. La primera fábrica tiene 150 trabajadores, con una edad media de 32 años. La segunda tiene 241 trabajadores con una edad media de 39 años, y la tercera tiene 165 trabajadores, con una edad media de 37 años. ¿Cuál es la edad media del conjunto de trabajadores de la empresa? Una indicación. La media no es “la media de las medias”:

$$\frac{32 + 39 + 37}{3}.$$

Para calcular correctamente la media necesitas algo más parecido a lo que hacemos cuando nos dan una tabla de frecuencias.

12. Este ejercicio explora una idea similar a la del anterior.
 - a) Las calificaciones finales de un estudiante en cuatro asignaturas fueron 82, 86, 90 y 70. Si los respectivos créditos¹ otorgados a esos cursos son 3, 5, 3 y 1 determinar una calificación media apropiada. De nuevo, es posible que pienses que la calificación media es:

$$\frac{82 + 86 + 90 + 70}{4},$$

pero si te preguntamos por la *calificación media por crédito*, ¿crees que ese es el resultado correcto? ¿Qué tipo de media es la que estamos calculando aquí?

- b) Supongamos que he hecho un viaje a pie en tres etapas. En la primera etapa he recorrido 25 kilómetros en 6 horas. En la segunda he recorrido 21 kilómetros en 4 horas y en la tercera, 32 kilómetros en 7 horas. ¿Cuál ha sido mi velocidad media en ese viaje, medida en kilómetros por hora? ¿Ves la relación con el apartado anterior?
- c) El átomo de Cloro tiene dos isótopos, cuya masa es de 35u (unidades de masa atómica) y 37u, respectivamente. El 75.5% de los átomos de Cloro son isótopos del primer tipo, y el resto son isótopos del segundo tipo. ¿Cuál es la masa atómica media del Cloro?

Las medias que aparecen en este ejercicio son ejemplos de **medias ponderadas**, ver el enlace:

http://es.wikipedia.org/wiki/Media_ponderada.

13. En la página 19 hemos hablado de números pseudoaleatorios con `Calc`, y hemos presentado la función `ALEATORIO.ENTRE()`. Esta función genera números aleatorios enteros. Si queremos generar números reales no enteros (como los valores de una variable cuantitativa continua) podemos usar la función `ALEATORIO()`. El resultado de esta función es un número pseudoaleatorio entre 0 y 1. Vamos a hacer algunos experimentos con ella.

¹El *crédito* es una unidad de medida académica que, en esencia, mide el tiempo de formación del estudiante. Ver http://es.wikipedia.org/wiki/Crédito_académico

- a) Usa la función `ALEATORIO()` para generar 100 números aleatorios entre 0 y 1. Calcula su media. Calcula su desviación típica. Repite el proceso varias veces (recuerda `Ctrl + Mays + F9`) y observa los valores de la media y la desviación típica cada vez. ¿Qué observas?
 - b) Ahora calcula 100 números aleatorios entre 30 y 50, y repite los pasos anteriores, para observar la media y la desviación típica. Indicación: para generar estos números sólo necesitas multiplicar por 20 y sumar 30.
 - c) Más ambicioso: trata de construir una hoja Calc en la que, tras introducir dos números a y b (con $a < b$), Calc construya 100 números pseudoaleatorios del intervalo (a, b) y calcule su media y desviación típica.
14. Para percibir algunas de las dificultades que nos vamos a encontrar en el trabajo con los datos, nada como la práctica. El fichero:

[Tut01-DatosTemperaturaINE.csv](#)

contiene datos sobre temperaturas medias mensuales registradas en distintos observatorios meteorológicos españoles, desde enero de 2007 a diciembre de 2012. Los datos proceden del INE (Instituto Nacional de Estadística de España, ver <http://www.ine.es>). El fichero, además de la tabla de datos propiamente dicha, contiene varias filas iniciales y varias finales con información adicional sobre los datos. En cuanto a la tabla, daremos algunas indicaciones:

- La primera columna contiene el nombre del observatorio.
- La primera fila (la cabecera) contiene el código del mes correspondiente a la observación. Por ejemplo 2011M09 significa *Septiembre de 2011*.
- Las columnas se han separado con tabuladores. Los valores aparecen entrecomillados.
- Un valor como "." significa que ese dato no está disponible, es un dato ausente. EL problema de los datos ausentes (en inglés, *missing data*) va a ser uno de nuestros compañeros inseparables en el Análisis de Datos.

Sabiendo esto:

- a) Lee el fichero con Calc. Es muy posible que antes tengas que *esquilarlo* un poco.
- b) Calcula la media de las temperaturas en Guadalajara durante ese periodo de tiempo.
- c) Calcula la temperatura media durante el mes de Agosto en todos los observatorios.

Fin del Tutorial01. ¡Gracias por la atención!