

# Introducción a la Estadística con Excel

En el siguiente guión vamos a introducir el software Excel 2007 y la manera de trabajar con Estadística Descriptiva.

## Cargar o importar datos

En Excel 2007 podemos meter los datos *a mano*, generarlos con el programa a través de algunas funciones (por ejemplo, "ALEATORIO()"), o cargarlos desde un archivo externo. Excel 2007 puede abrir archivos que estén en varios formatos, además de los suyos propios (xls,xlsx).

El conjunto de datos que se utiliza en este guión está recogido en el fichero "Países.xlsx".

Los datos se refieren a 91 países y describen 7 variables, de ellas 6 variables cuantitativas (tasa de natalidad, tasa de mortalidad, tasa de mortalidad infantil, expectativa de vida – hombres, Expectativa de vida – mujeres, PIB) y 1 variable cualitativa (zona).

Para abrir el fichero "Países.xlsx" hay que seguir los siguientes pasos:

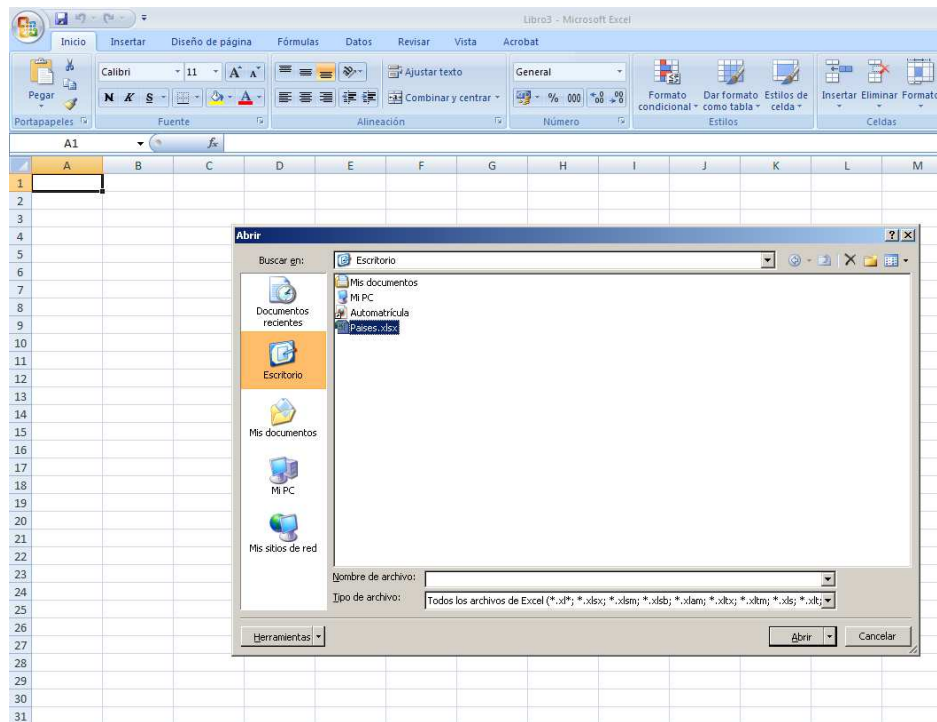
1. Pulsar el Botón de *Office*:



2. Seleccionar *Abrir*:



3. Seleccionar el fichero "Países.xlsx" en su ubicación y pulsar *Abrir*. Por ejemplo, en este caso:



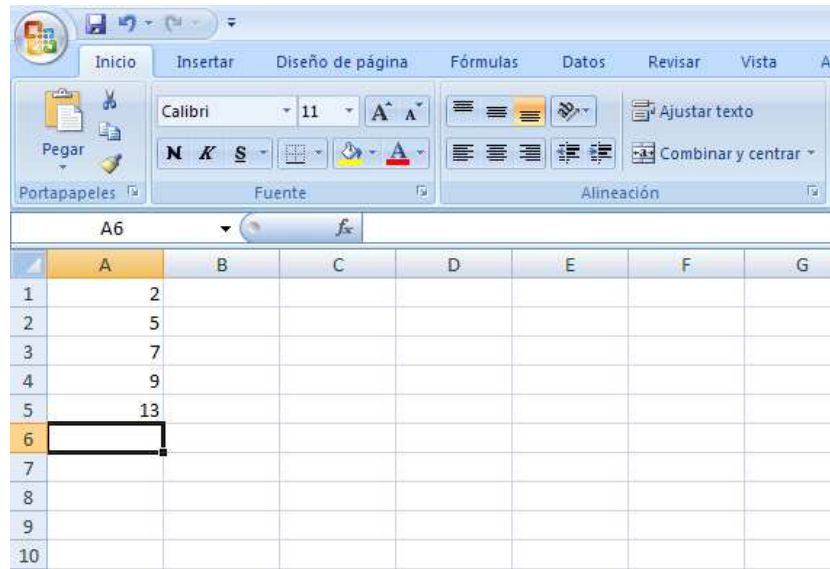
Una vez cargados los datos, Excel 2007 nos ofrecerá la siguiente vista:

Países.xlsx - Microsoft Excel								
G1 PIB								
	A	B	C	D	E	F	G	H
1	País	Tasa de natalidad	Tasa de mortalidad	Tasa de mortalidad infantil	Expectativa de vida - hombres	Expectativa de vida - mujeres	PIB	Zona
2	Algeria	35,5	8,3	74	61,6	63,3	2060	AFR
3	Angola	47,2	20,2	137	42,9	46,1	610	AFR
4	Botswana	48,5	11,6	67	52,3	59,7	2040	AFR
5	Congo	46,1	14,6	73	50,1	55,3	1010	AFR
6	Egypt	38,8	9,5	49,4	57,8	60,3	600	AFR
7	Ethiopia	48,6	20,7	137	42,4	45,6	120	AFR
8	Gabon	39,4	16,8	103	49,9	53,2	390	AFR
9	Gambia	47,4	21,4	143	41,4	44,6	260	AFR
10	Ghana	44,4	13,1	90	52,2	55,8	390	AFR
11	Kenya	47	11,3	72	56,5	60,5	370	AFR
12	Libya	44	9,4	82	59,1	62,6	5310	AFR
13	Malawi	48,3	25	130	38,1	41,2	200	AFR
14	Morocco	35,5	9,8	82	59,1	62,5	960	AFR
15	Mozambique	45	18,5	141	44,9	48,1	80	AFR
16	Namibia	44	12,1	135	55	57,5	1030	AFR
17	Nigeria	48,5	15,6	105	48,8	52,2	360	AFR
18	Sierra_Leone	48,2	23,4	154	39,4	42,6	240	AFR
19	Somalia	50,1	20,2	132	43,4	46,6	120	AFR
20	South_Africa	32,1	9,9	72	57,5	63,5	2530	AFR
21	Sudan	44,6	15,8	108	48,6	51	480	AFR
22	Swaziland	46,8	12,5	118	42,9	49,5	810	AFR
23	Tanzania	50,5	14	106	51,3	54,7	110	AFR
24	Tunisia	31,1	7,3	52	64,9	66,4	1440	AFR
25	Uganda	52,2	15,6	103	49,9	52,7	220	AFR
26	Zaire	45,6	14,2	83	50,3	53,7	220	AFR
27	Zambia	51,1	13,7	80	50,4	52,5	420	AFR
28	Zimbabwe	41,7	10,3	66	56,5	60,1	640	AFR
29	Argentina	20,7	8,4	25,7	65,5	72,7	2370	AME
30	Bolivia	46,6	18	111	51	55,4	630	AME
31	Brazil	28,6	7,9	63	62,3	67,6	2680	AME
32	Canada	14,5	7,3	7,2	73	79,8	20470	AME
33	Chile	23,4	5,8	17,1	68,1	75,1	1940	AME
34	Columbia	27,4	6,1	40	63,4	69,2	1260	AME
35	Ecuador	32,9	7,4	63	63,4	67,6	980	AME
36	Guyana	28,3	7,3	56	60,4	66,1	330	AME
37	Mexico	29	23,2	43	62,1	66	2490	AME
38	Paraguay	34,8	6,6	42	64,4	68,5	1110	AME
39	Peru	23,8	6,3	100,0	66,0	66,5	1160	AME

## Fórmulas programadas

Excel 2007 ofrece la posibilidad de utilizar algunas funciones ya programadas. Empezamos con un caso sencillo.

En una hoja Excel en blanco vamos a introducir los siguientes números (2, 5, 7, 9, 13) de la siguiente manera:

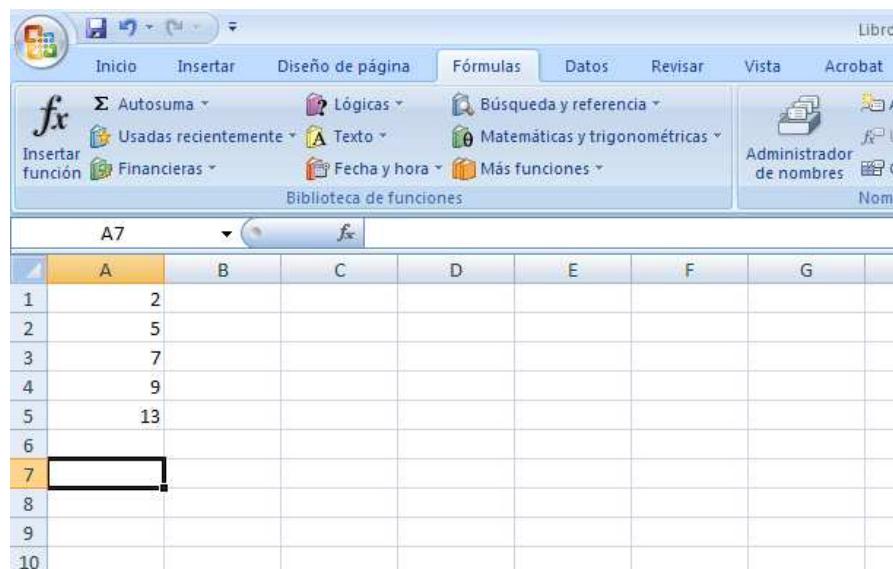


The screenshot shows the Excel 2007 interface with the 'Inicio' ribbon selected. The spreadsheet has the following data in column A:

	A	B	C	D	E	F	G
1	2						
2	5						
3	7						
4	9						
5	13						
6							
7							
8							
9							
10							

Supongamos que queremos calcular la media de los números anteriores y escribirla en la celda A7. Para ello podemos utilizar una de las fórmulas *programadas* en *Excel 2007*, siguiendo los siguientes pasos:

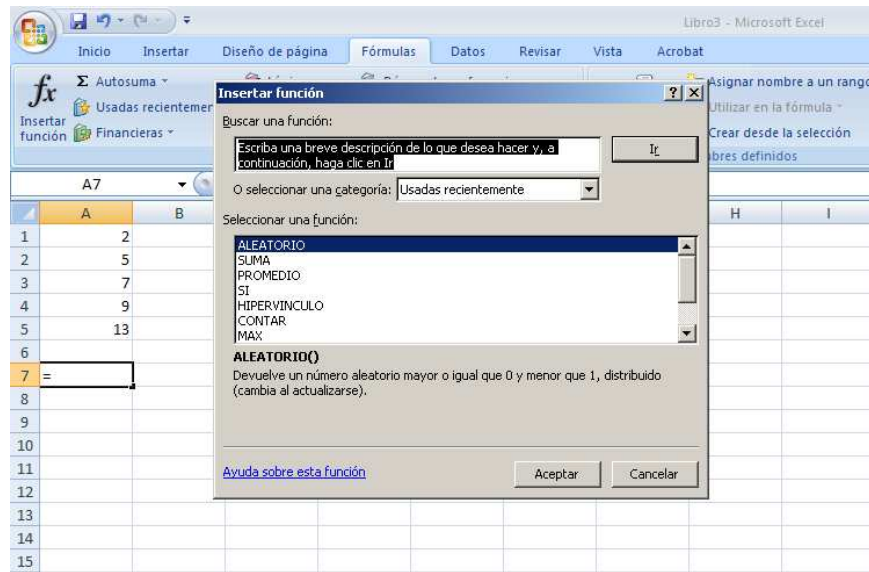
1. Seleccionar la celda A7 y *Fórmulas* en el Menú de arriba (se trata de la línea que contiene *Inicio*, *Insertar*, *Diseño de página*, *Fórmulas*, *Datos*, etc.):



The screenshot shows the Excel 2007 interface with the 'Fórmulas' ribbon selected. The spreadsheet is the same as in the previous image, but now cell A7 is selected. The 'Fórmulas' ribbon shows various function categories like Autosuma, Usadas recientemente, Financieras, Lógicas, Texto, Fecha y hora, Búsqueda y referencia, Matemáticas y trigonométricas, and Más funciones.

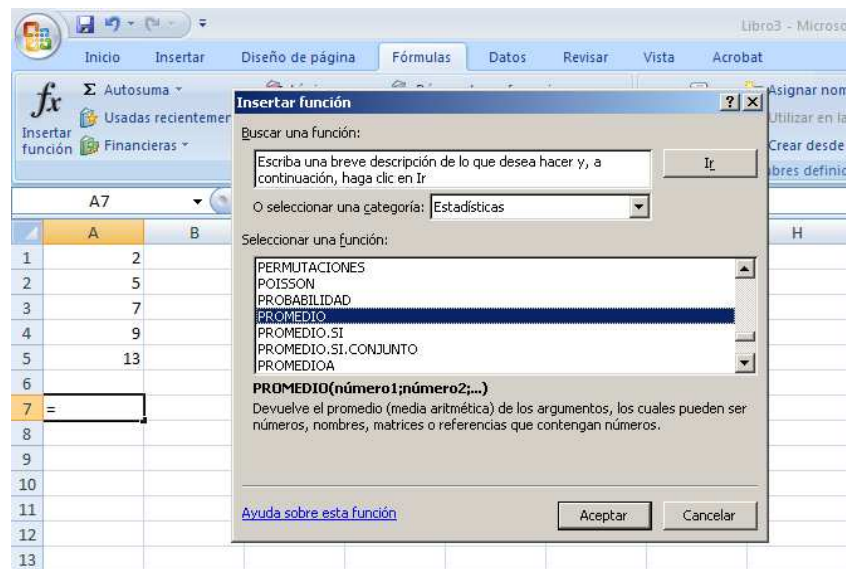
	A	B	C	D	E	F	G
1	2						
2	5						
3	7						
4	9						
5	13						
6							
7							
8							
9							
10							

2. Seleccionar *Insertar función*. Una vez seleccionado insertar función. Entonces, Excel 2007 muestra la siguiente vista:

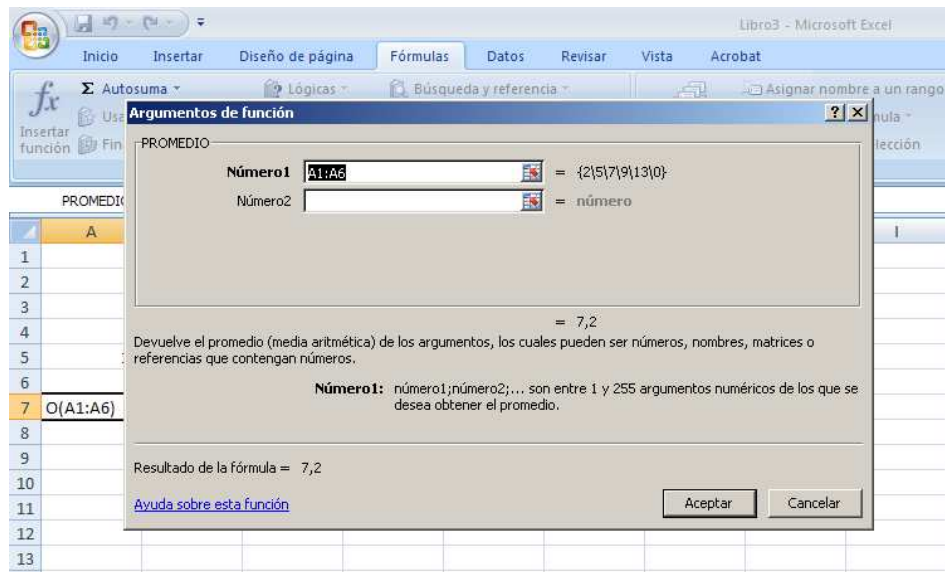


3. Buscar la función que permite calcular la media de un conjunto de números: "PROMEDIO()"

Para hacer eso hay dos maneras: la primera, intentar buscar la función que nos interesa entre las funciones *Usadas recientemente* (en este caso se trata de la tercera función); la segunda manera, y más general, buscarla en la categoría *Todas* o en su categoría, en este caso *Estadísticas*. En los dos casos, pulsar *Aceptar* una vez encontrada y seleccionada la función:

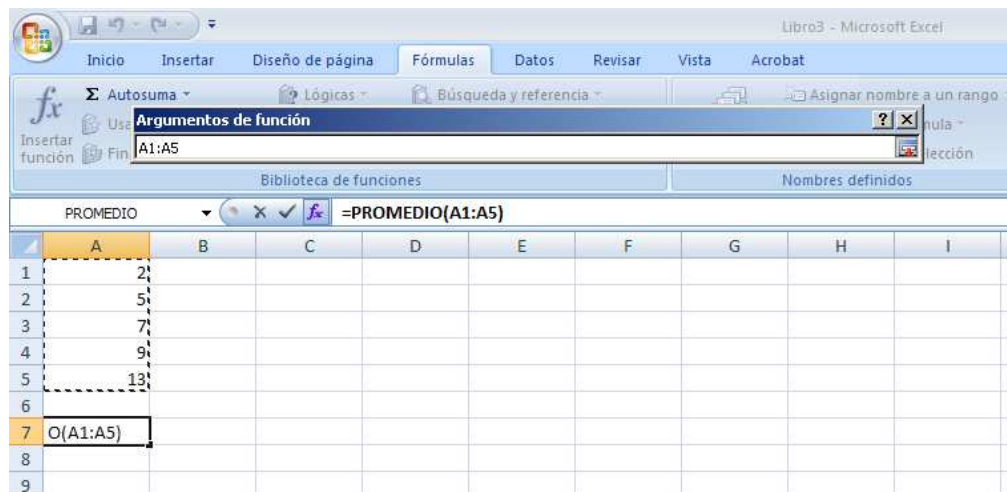


4. Una vez pulsado *Aceptar* Excel 2007 ofrecerá la siguiente vista:

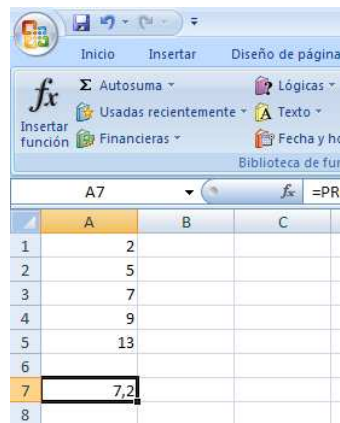


Podemos notar que Excel 2007 ha seleccionado automáticamente unas celdas (A1:A6) que puede traducirse en “desde A1 hasta A6”, pero nosotros queremos la media de los números contenidos en las celdas “desde A1 hasta A5”. Entonces, tendremos que introducir en el recuadro **Número 1** el texto “A1:A5”.

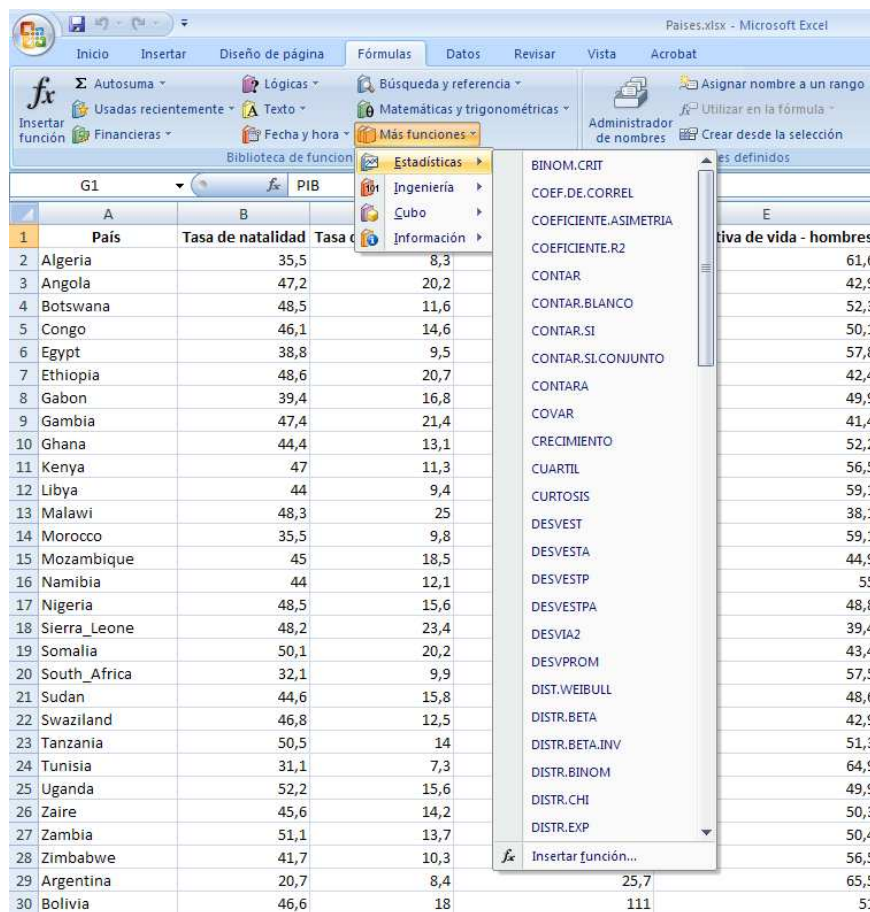
Otra manera de hacerlo, que muchas veces resulta mucho más cómoda, es la siguiente: Primero, minimizar la ventana *Argumentos de función* usando el botón del interior del recuadro **Número 1**. Segundo, seleccionar con el ratón las celdas desde A1 hasta A5, pulsar el botón *Enviar* del teclado y después pulsar *Aceptar* :



5. Excel 2007 nos ofrecerá la siguiente vista, donde 7,2 representa la media que se buscaba:



**Observación:** En Excel 2007 resultan más accesibles sus funciones programadas. En nuestro caso, desde la pestaña *Fórmulas* podemos acceder directamente a las funciones estadísticas:





## Cómo cargar el módulo para Estadística

En Excel 2007 existe la posibilidad de cargar un módulo que nos facilita el trabajo a la hora de hacer cálculos estadísticos avanzados. Como no es un módulo básico, no aparece en la pestaña *Datos* por defecto y hay que cargarlo expresamente.

Para hacer que el submenú *Análisis de datos* aparezca en la pestaña *Datos*, podemos seguir los pasos que nos proporciona la ayuda de *Excel 2007*:

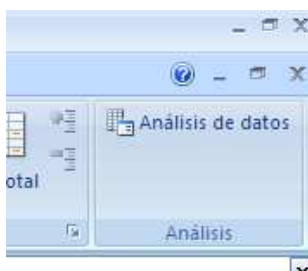
1. Pulsar el *Punto Interrogación* (arriba a la derecha) para abrir la *Ayuda* de Excel 2007:



Introducir *Herramientas para análisis* en el buscador de la *Ayuda* y pulsar *Buscar*. En nuestro caso lo que estamos buscando es el primer artículo, es decir Cargar Herramientas para análisis. Seleccionarlo entonces, y seguir las instrucciones.



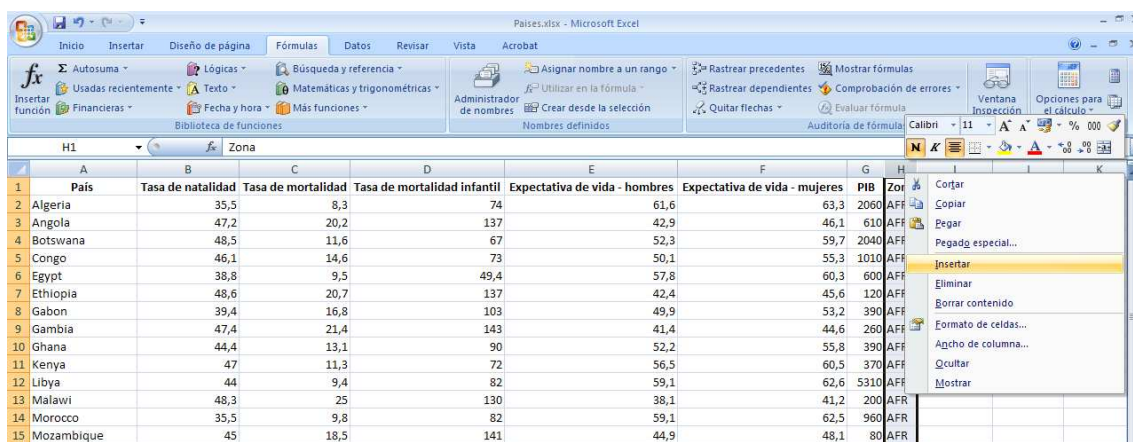
Una vez completado el proceso, en la pestaña *Datos* aparece una nueva herramienta (a la derecha): la herramienta para el *Análisis de datos*:



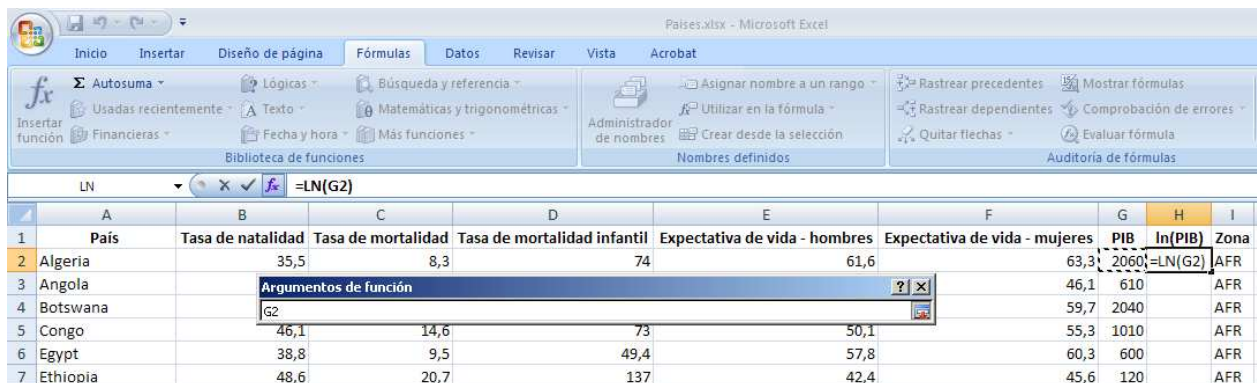
## Cómo transformar los datos

En Estadística, en algunos casos hace falta trabajar con datos transformados en lugar de los datos originales. Uno de los casos más comunes es el uso de la transformación logarítmica en datos relativos al PIB. La penúltima variable de los datos contenidos en el fichero "Países.xlsx" representa, de hecho, el PIB de los 91 países. Aplicamos, así, a esta variable la transformación logarítmica:

1. Se inserta una nueva columna a la derecha de la columna G. Para hacer eso, se pone el cursor encima de la letra de la columna siguiente (en este caso H) y se pulsa el botón izquierdo del ratón (con esta acción estamos seleccionando toda la columna H). Se pulsa el botón derecho del ratón y se selecciona *Insertar*:



2. Se introduce en la celda H1 un nombre para la nueva variable, por ejemplo **ln(PIB)**. Nos posicionamos en la celda H2 y repetimos los pasos que vimos antes para la introducción de una función de Excel 2007. Esta vez habrá que buscar la función LN(), es decir la función que calcula el logaritmo natural de un numero real positivo mayor que 0. Esta función está en la categoría *Matemáticas y trigonométricas*. El parámetro de la función será la celda G2, es decir, la celda en el que está almacenado el PIB de Argelia, que es el valor que queremos transformar:





3. Para repetir la misma operación con los otros países podemos utilizar una característica de Excel 2007: si copiamos y pegamos el contenido de la celda H2 en la celda H3, Excel no copiará el valor 7,63046, sino escribirá en la celda H3 lo siguiente: “=LN(G3)”. Ahora bien, si copiamos el contenido de la celda H2 en todas las celdas hasta H92, obtendremos la nueva variable **ln(PIB)**:

The screenshot shows the Microsoft Excel 2007 interface. The active sheet is 'Países.xlsx'. The formula bar shows '=LN(G2)'. The spreadsheet contains data for 36 countries, with columns A through I. Column H is highlighted, showing the formula being applied to the data in column G.

	A	B	C	D	E	F	G	H	I
56	Malaysia	31,6	5,6	24	67,5	71,6	2320	7,74932	ASIA
57	Mongolia	36,1	8,8	68	60	62,5	110	4,70048	ASIA
58	Nepal	39,6	14,8	128	50,9	48,1	170	5,1358	ASIA
59	Oman	45,6	7,8	40	62,2	65,8	5220	8,56025	ASIA
60	Pakistan	30,3	8,1	107,7	59	59,2	380	5,94017	ASIA
61	Philippines	33,2	7,7	45	62,5	66,1	730	6,59304	ASIA
62	Saudi_Arabia	42,1	7,6	71	61,7	65,2	7050	8,86078	ASIA
63	Singapore	17,8	5,2	7,5	68,7	74	11160	9,32009	ASIA
64	Sri_Lanka	21,3	6,2	19,4	67,8	71,7	470	6,15273	ASIA
65	Thailand	22,3	7,7	28	63,8	68,9	1420	7,25841	ASIA
66	United_Arab_Emir	22,8	3,8	26	68,6	72,9	19860	9,89646	ASIA
67	Albania	24,7	5,7	30,8	69,6	75,5	600	6,39693	EU
68	Austria	14,9	7,4	8	73,3	79,6	17000	9,74097	EU
69	Belgium	12	10,6	7,9	70	76,8	15540	9,65117	EU
70	Bulgaria	12,5	11,9	14,4	68,3	74,7	2250	7,71869	EU
71	Byelorussia	15,2	9,5	13,1	66,4	75,9	1880	7,53903	EU
72	Czechoslovakia	13,4	11,7	11,3	71,8	77,7	2980	7,99968	EU
73	Denmark	12,4	11,9	7,5	71,8	77,7	22080	10,0024	EU
74	Finland	13,2	10,1	5,8	70,7	78,7	26040	10,1674	EU
75	France	13,6	9,4	7,4	72,3	80,5	19490	9,87766	EU
76	Germany	11,4	11,2	7,4	71,8	78,4	22320	10,0132	EU
77	Greece	10,1	9,2	11	65,4	74	5990	8,69785	EU
78	Hungary	11,6	13,4	14,8	65,4	73,8	2780	7,93021	EU
79	Ireland	15,1	9,1	7,5	71	76,7	9550	9,1643	EU
80	Italy	9,7	9,1	8,8	72	78,6	16830	9,73092	EU
81	Netherlands	13,2	8,6	7,1	73,3	79,9	17320	9,75962	EU
82	Norway	14,3	10,7	7,8	67,2	75,7	23120	10,0485	EU
83	Poland	14,3	10,2	16	67,2	75,7	1690	7,43248	EU
84	Portugal	11,9	9,5	13,1	66,5	72,4	7600	8,9359	EU
85	Romania	13,6	10,7	26,9	66,5	72,4	1640	7,40245	EU
86	Spain	10,7	8,2	8,1	72,5	78,6	11020	9,30747	EU
87	Sweden	14,5	11,1	5,6	74,2	80	23660	10,0715	EU
88	Switzerland	12,5	9,5	7,1	73,9	80	34064	10,436	EU
89	Turkey	29,2	8,4	76	62,5	65,8	1630	7,39634	EU
90	U.K.	13,6	11,5	8,4	72,2	77,9	16100	9,68657	EU
91	Ukraine	13,4	11,6	13	66,4	74,8	1320	7,18539	EU
92	USSR	17,7	10	23	64,6	74	2242	7,71512	EU
93									

## ¿Qué más podemos hacer?

En esta última sección, mediante un ejemplo, vamos a mostrar como introducir fórmulas manualmente, y al mismo tiempo mostraremos la utilidad del símbolo “\$” (dólar) a la hora de copiar una fórmula con *Excel 2007*.

Los datos que vamos a utilizar son los siguientes:

	A	B	C	D	E	F	G
1	Estudiante	Parcial 1	Parcial 2	Examen final		Nota Final con Método 1	Nota Final con Método 2
2	Amanda	9	7	7,5			
3	Beatriz	7	8	8			
4	Carlos	5	7	6			
5	David	7,5	8,5	9			
6	Elena	9,5	9	10			
7							
8		Método 1	Método 2				
9	Peso del Parcial 1	0,3	0,25				
10	Peso del Parcial 2	0,2	0,15				
11	Peso del Examen final	0,5	0,6				
12							

Se trata de las notas de cinco estudiantes, (dos pruebas parciales y un examen final). Para calcular la nota final, la profesora ha decidido utilizar el método de la media ponderada, pero no está segura de los pesos que hay que asignar a cada prueba, y por este motivo ha planteado dos métodos, el 1 y el 2.

Para empezar, vamos a calcular en la celda F2 la nota final de Amanda con el método 1. Para hacer eso nos posicionamos en la celda F2 e introducimos la siguiente expresión:

`"=B9*B2+B10*C2+B11*D2"`

El resultado será igual a 7,85.

Ahora, copiando el contenido de la celda F2 en la celda F3, se obtiene la siguiente expresión:

`"=B10*B3+B11*C3+B12*D3"`

¿Qué ha pasado? Mientras todas las letras no cambian, todos los números aumentan de **una** unidad porque hemos copiado la celda F2 en **una** celda más abajo. En este caso, para calcular la nota final de Beatriz con el método 1 tenemos que bloquear las filas de las celdas que contienen los pesos del método 1. ¿Cómo? Utilizando el símbolo "\$" de la siguiente manera en la siguiente expresión que vamos a introducir en la celda F2:

`"=B$9*B2+B$10*C2+B$11*D2"`

Ahora, si volvemos a copiar el contenido de la celda F2 en la celda F3, se obtiene la siguiente, y correcta, expresión:

`"=B10*B3+B11*C3+B12*D3"`

Finalmente, podemos copiar el contenido de la celda F2 (o F3, los resultados no cambiarían) en F4, F5 y F6. Las notas finales con el método 1 serían las siguientes:

#### Nota Final con Método 1

7,85

7,7

5,9

8,45

9,65

Falta ahora calcular las notas finales con el método 2. Copiando el contenido de la celda F2 en la celda G2 obtenemos la siguiente expresión:

$$"=C\$9*C2+C\$10*D2+C\$11*E2"$$

¿Qué ha pasado esta vez? Mientras todos los números no cambian, todas las letras "aumentan" de **una** unidad porque hemos copiado la celda F2 en **una** celda más a la derecha. En este caso, para calcular la nota final de Amanda con el método 2 tenemos que bloquear las columnas de las celdas que contienen las notas. ¿Cómo? Utilizando el símbolo "\$" de la siguiente manera en la siguiente expresión que vamos a introducir en la celda F2:

$$"=B\$9*\$B2+B\$10*\$C2+B\$11*\$D2"$$

Ahora, si volvemos a copiar el contenido de la celda F2 en la celda G2, se obtiene la siguiente, y correcta, expresión:

$$"=C\$9*\$B2+C\$10*\$C2+C\$11*\$D2"$$

Finalmente, podemos copiar el contenido de la celda G2 en G3, G4, G5 y G6. Las notas finales con el método 2 serían las siguientes:

#### Nota Final con Método 2

7,8  
7,75  
5,9  
8,55  
9,725

**Observación:** hemos visto como bloquear una fila (ejemplo: A\$1) y como bloquear una columna (ejemplo: \$A1). Podemos también bloquear al mismo tiempo una fila y una columna (ejemplo \$A\$1): cuando bloqueamos una fila y una columna de hecho estamos bloqueando una celda.

#### Otra forma de resolver el ejercicio, mediante producto de matrices:

Primero, antes de poner el = en la barra de escritura, seleccionar todas las celdas reservadas para la matriz producto.

Con las celas marcadas, se introduce la función *MMULT* en la barra, y **NO** se pulsa directamente *Enter*. Se pulsa la combinación de teclas: **Ctrl+Shift+Enter**.

## Ejercicios para entrenamiento

1. Escribe una columna con 20 valores numéricos. Puedes simular los datos mediante una función de Excel, por ejemplo, **ALEATORIO()** o escribirlos directamente en la hoja de Excel.

**Nota:** Una vez generados los datos, para evitar que se generen nuevos valores cada vez que se ejecuta un nuevo comando, es mejor copiar los valores usando la opción de pegar **Valores** en el **pegado especial**.

2. Escribe una función para calcular la media. Compara el resultado con la función original de Excel.

3. Escribe otra función para calcular la varianza mediante el cálculo de la media de los cuadrados menos el cuadrado de la media. Compara el resultado con la función original de Excel.

Usar sobre los datos y comentar los resultados de las siguientes funciones:

**DESVEST** y **DESVESTA**. Ambas calculan la cuasi-desviación típica. Pero, la primera ignora casillas en blanco y texto dentro del rango seleccionado y la otra no.

**DESVESTP** y **DESVESTPA**. Calculan la desviación típica

**DESVIA2**. Calcula la suma de los cuadrados de las desviaciones a la media.

**DESVPROM**. Calcula el promedio de las *desviaciones absolutas* a la media.