



# Estadística: Grado en Información y Documentación

## Práctica 1: INTRODUCCIÓN AL USO DE EXCEL 2003

**OBJETIVO:** Los estudiantes deberán conocer el funcionamiento de la Hoja de Cálculo EXCEL y utilizarla para realizar estudios estadísticos, tanto de una como de dos variables, como los siguientes:

1. Obtener representaciones gráficas según el tipo de variable.
2. Hacer los cálculos necesarios para obtener los distintos estadísticos (media, mediana, moda, cuantiles, varianzas, etc.). A partir de unos datos (o de una distribución de frecuencias) hacemos los cálculos necesarios para desarrollar una fórmula y paso a paso calcular hasta obtener el resultado final. (Sumas, productos, divisiones, potencias, raíces cuadradas, etc.)
3. Obtener fácilmente estadísticos como promedios, varianzas, cuantiles, covarianzas, regresión, etc., aplicando unas funciones previamente programadas en la hoja de cálculo. (Insertar + Función o pinchar el botón  $f_x$  de la barra de herramientas). (Ver listado de funciones y argumentos que requieren)
4. Utilizar en el menú Herramientas la opción Análisis de Datos para hacer análisis estadísticos. Permite describir un conjunto de datos de una o varias variables. Los datos de entrada son con frecuencia unitaria (si un valor se repite varias veces aparece repetido ese número de veces). También permite aplicar técnicas estadísticas más sofisticadas como regresión, contrastes de hipótesis, estimación por intervalos, etc.

## INTRODUCCIÓN EN LA HOJA DE CÁLCULO EXCEL

Excel es un programa muy difundido y ampliamente utilizado. Dado que se trata de un programa del paquete Microsoft Office. Las acciones genéricas son muy similares a cualquier otro programa del mismo paquete, y ¿quien no maneja Microsoft Word hoy en día?. Un clic de ratón para seleccionar, un doble clic para accionar, el uso de teclado (flechas para desplazarse, enter para accionar, etc.) o la utilización de la Ayuda del programa son recursos que guardan similitud con cualquier otro software del Entorno de Windows.

La hoja de cálculo, a pesar de su manejo sencillo, constituye una herramienta muy potente de cálculo estadístico. Como su nombre indica, la hoja de cálculo nos permite calcular, hacer operaciones matemáticas, lógicas, estadísticas, financieras, etc.

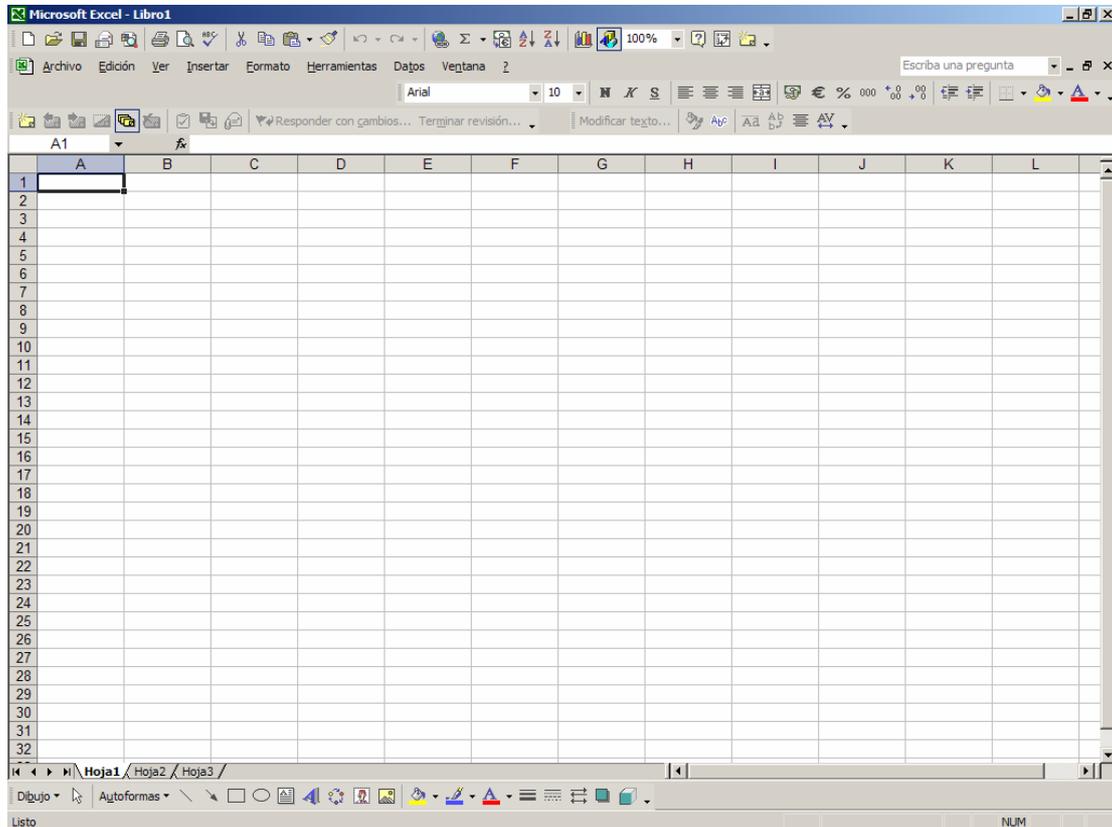
Excel es una hoja de cálculo que facilita los cálculos y la obtención de indicadores y estadísticos para estudiar el comportamiento de una o varias variables. Una hoja de cálculo es una matriz de filas y columnas, donde cada combinación de fila y columna se denomina celda; por ejemplo, las celdas A1, B5, C15, etc.

Un conjunto de celdas se denomina rango, y en Excel se expresa como A1:C5. Cada celda puede contener un número, una fórmula, texto, una fecha, etc.

Con la hoja de cálculo EXCEL se pueden obtener fácilmente estadísticos y medidas que resuman y caractericen una variable, una distribución de frecuencias, o una distribución bidimensional de frecuencias.

El tipo de fichero de Excel se llama Libro (análogo a lo que es el documento en Word) y su extensión habitual será .xls (de igual forma que es .doc para Word). Cada libro puede contener un número ilimitado de hojas de cálculo.

La primera pantalla que aparece en Excel es la siguiente:



Hay dos formas de acceder a la ayuda. Una es en el Menú con signo de interrogación "?". Otra es en el cuadro de texto situado en la esquina superior derecha, donde aparece "Escriba una pregunta". Escribiendo alguna pregunta o palabra clave, podemos acceder a la ayuda relacionada con dichas palabras.

Un primer recorrido por la hoja de cálculo nos permite conocer las:

- opciones de menú
- barra de herramientas
- barra de estado

## Manejo de Excel

### Introducir datos

Para introducir un dato numérico o textual en cada celda, se hace clic en la celda y se escribe.

### Introducir fórmulas

Algunas de las operaciones matemáticas que puede hacer son:

Suma = ..+..  
Multiplicación = ..\*..  
División = .. /..  
Elevar al cuadrado = ..^2  
Elevar a potencia n = ..^n  
Raíz cuadrada=raiz(..)  
Valor absoluto =abs(..)  
Raíz n-esima=...^(1/n)  
Logaritmo neperiano =LN(..)  
Etc.

Para introducir una fórmula en una celda, se hace clic en ella, se escribe el signo de igualdad (=) y, a continuación, la fórmula.

La mecánica es similar al procedimiento que utilizamos para resolver problemas de estadística utilizando una calculadora. Es importante disponer de manera clara y ordenada los cálculos. Esto facilita su posterior uso e incluso la revisión de los procedimientos si se detectan errores.

Como ya se ha indicado, una celda puede contener un número o una fórmula.

Para crear una fórmula empezamos tecleando en una celda, por ejemplo la C1, el símbolo "=", y la operación que deseamos realizar. Puede ser un simple cálculo matemático: =3+5, y nos da un resultado 8. Pero puede ser una fórmula =A1+B1. Si en la celda A1 hay un 3 y en la B1 un 5, aparecerá el mismo resultado 8, sin embargo ya no es un número, sino una fórmula. Esto tiene dos implicaciones importantes:

- 1.- Si cambian los valores de las celdas A1 y B1, cambia el resultado en C1 obteniendo la suma de los nuevos números.
- 2.- Cuando copiamos la celda C1 a otra celda, no copiamos el valor, sino la fórmula.

En algunas ocasiones nos interesa que la referencia de una celda no se modifique cuando se copie la fórmula que la contiene. Podemos indicarle en la fórmula que mantenga fija la fila, la columna o ambas. Para ello deberemos anteponer el símbolo \$ antes de la letra (columna), o antes del número (fila) o antes de las dos (fila y columna, es decir celda) de la identificación de la celda.

Este tipo de referencias que se mantienen constantes al copiar la fórmula se denominan referencias absolutas a celdas o celdas absolutas. Ejemplo : \$C\$26 es una referencia absoluta, mientras C26 es una referencia relativa.

Sumar un rango de valores (con la función =SUMA(..), o el botón autosuma  $\Sigma$ )

Con ello se pueden calcular estadísticos que utilizaremos en los siguientes temas.

#### **Seleccionar:**

- a) Una celda: se hace clic en ella.
- b) Una fila: se hace clic en el número de cabecera.
- c) Varias filas seguidas: se hace clic en la primera y se arrastra (o Shift + flecha descendente) hasta la última (también se puede hacer al revés).
- d) Una columna: se hace clic en la letra de la cabecera.
- e) Varias columnas: se hace clic en la primera y se arrastra hasta la última (también se puede hacer al revés).
- f) Un rango: se hace clic en la celda de uno de los vértices y se arrastra hasta la celda de la esquina opuesta.

#### **Insertar hojas:**

Para insertar una hoja nueva, se elige en la barra de menús **Insertar /Hoja de cálculo**

#### **Mover hojas**

Para mover una hoja, se selecciona en la parte inferior, en la barra de hojas del libro, y se arrastra al lugar correspondiente.

**EJERCICIO 1:** Se ha preguntado a 40 personas sobre el número de libros leídos en el último mes y se han obtenido los siguientes resultados:

1, 1, 2, 1, 1, 2, 1, 4, 2, 2, 1, 2, 1, 1, 0, 3, 1, 2, 1, 1  
2, 3, 2, 3, 0, 1, 1, 1, 2, 4, 1, 2, 2, 3, 3, 0, 2, 3, 0, 0

1. Abre la Hoja de Cálculo Excel.
2. Introduce todos los valores en el rango A4:H8

The screenshot shows a Microsoft Excel window titled 'Libro1.xls'. The spreadsheet contains the following data:

	A	B	C	D	E	F	G	H	I	
1										
2			LIBROS LEÍDOS EN EL ÚLTIMO MES							
3										
4	1	1	2	1	1	2	1	4		
5	2	2	1	2	1	1	0	3		
6	1	2	1	1	2	3	2	3		
7	0	1	1	1	2	4	1	2		
8	2	3	3	0	2	3	0	0		
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

The formula bar at the bottom shows 'H18' and an equals sign. The status bar at the bottom right shows 'NUM'.

### CALCULOS EN EXCEL APLICANDO FUNCIONES.

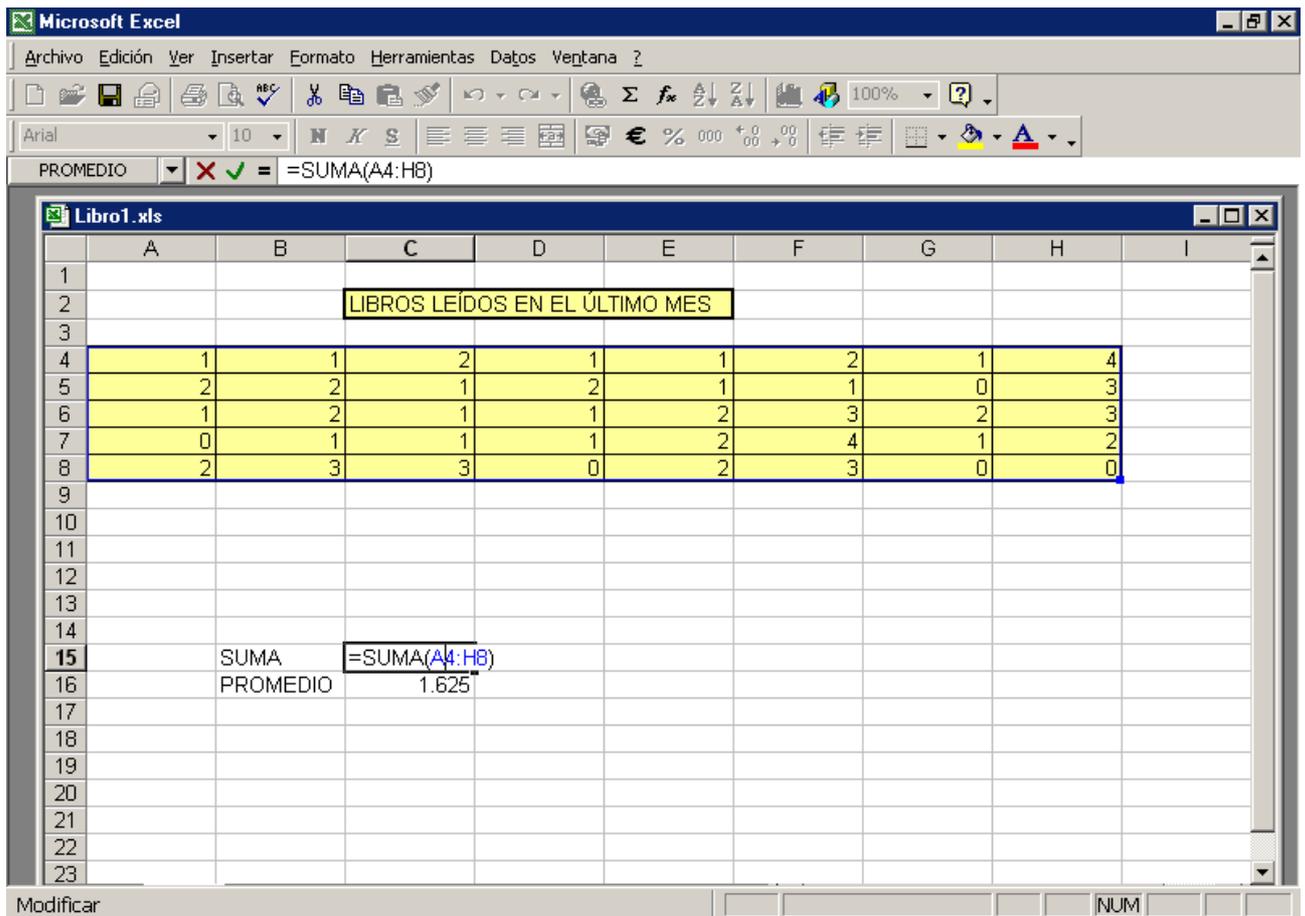
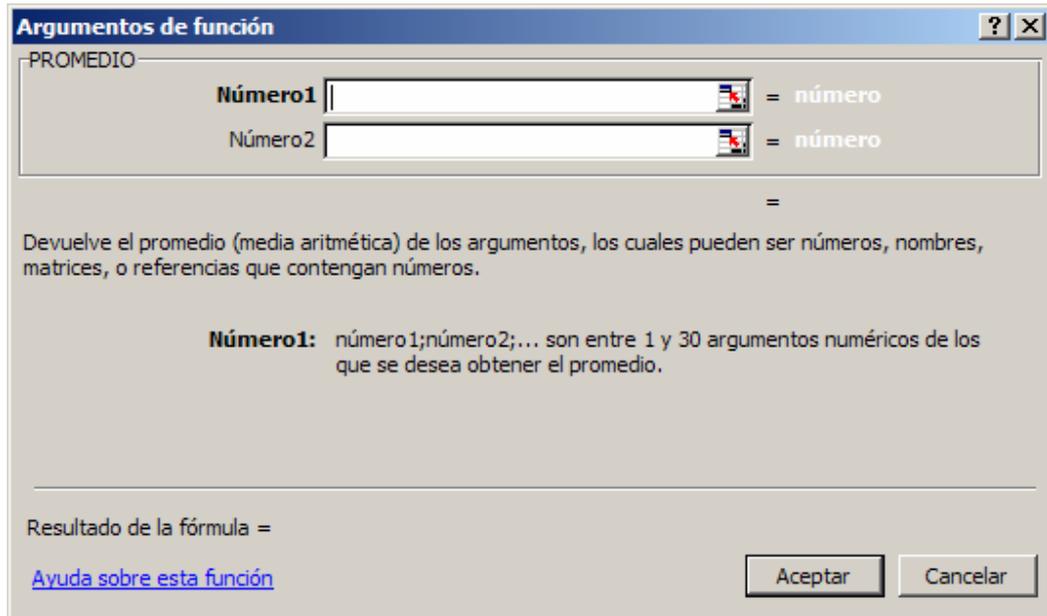
En el menú **Insertar**, seleccionar la opción **Función**, después la categoría de **Estadísticas** y dentro de ellas la que interese.

Una vez introducidos los datos, utilizaremos las funciones que Excel nos proporciona. Por ejemplo, la función `=SUMA(Rango)` o `=PROMEDIO(Rango)` calculan respectivamente la suma y el promedio de una muestra.

La función promedio permite calcular la media aritmética. Aparece en la parte inferior del cuadro de diálogo una breve descripción de la función, y permite solicitar más información sobre ella en Ayuda sobre esta función.

En este caso los argumentos de la función, lo que habría que rellenar en el cuadro de diálogo es el conjunto de celdas en los que se encuentran los valores de la variable.

En Número1 habrá que indicarle el rango de celdas, por ejemplo A4:H8, o pinchando en el recuadro que hay a la derecha, marcar con el ratón dichas celdas.



Usando las funciones podremos calcular: máximo y mínimo valor, recuento, promedios, cuantiles, mediana, moda, media armónica, geométrica, media acotada, varianza y desviación típica, medidas de forma, etc. También podemos calcular algunos estadísticos bidimensionales: covarianza, correlación, estimación lineal, estimación logarítmica, tendencia, pronóstico, etc. que estudiaremos en los siguientes temas de la asignatura.

## **FUNCIONES ESTADÍSTICAS Y SU CORRESPONDENCIA CON LOS ESTADÍSTICOS QUE SE EXPLICARÁN EN TEORÍA.**

### **COEF.DE.CORREL.**

Dar rango de valores de las variables X e Y.

### **COEFICIENTE.ASIMETRIA**

Rango de valores de la variable.

### **COEFICIENTE.R2**

Dar rango de valores de las variables X e Y.

### **CONTAR**

Rango de valores de la variable. Cuenta el número de valores.

### **CONTAR.SI**

Especificar rango de valores de la variable y criterio. Por ejemplo número de veces que en una variable se repite el 3. El criterio puede ser un número o una celda de la hoja de cálculo.

### **COVAR**

Dar rango de valores de las variables X e Y.

### **CUARTIL**

Hay que introducir el Rango de valores de la variable, y el número del cuartil que se busca: 1,2,3.

### **CURTOSIS**

Rango de valores de la variable.

### **DESVESTP**

Rango de valores de la variable.

### **DESVIA2**

(Media de desviaciones respecto a la media al cuadrado) Rango de valores de la variable.

### **DESVPROM**

(Media de valores absolutos de desviaciones respecto a la media al cuadrado). Rango de valores de la variable.

### **ESTIMACION.LINEAL**

(Recta de regresión).

Conocido\_y es el rango de valores de la variable dependiente.

Conocido\_x es el rango de valores de la variable independiente.

Constante, es un valor lógico, de forma que si se pone verdadero, calculará la recta  $y=a+bx$ , si se pone falso calculará  $y=bx$ .

Estadística, también es un valor lógico, lo dejaremos en falso.

La peculiaridad de esta función es que la salida es un conjunto de celdas, una matriz (Valores de la ordenada en origen y la pendiente).

Para visualizar toda la salida habrá que: Marcar con el ratón dos celdas, pulsar la tecla de funciones F2 y pulsar a continuación y al mismo tiempo las teclas

Control+Shift+Intro. El primer valor, corresponde a la pendiente y el segundo a la ordenada en el origen.

### **FRECUENCIA**

(Distribución de frecuencias). El argumento Datos, es el rango de valores de la variable. El argumento Grupos, son los extremos superiores del intervalo. El resultado de la función también es una matriz, y por tanto habrá que usar la opción de marcar el rango de celdas que tiene la matriz, pulsar la tecla de funciones F2 y pulsar a continuación y al mismo tiempo las teclas Control+Shift+Intro.

### **MAX**

(Valor máximo) Rango de valores de la variable.

### **MEDIA.ACOTADA**

(Media de valores centrales) El argumento Matriz, es el rango de valores de la variable. El argumento porcentaje es el % de valores que se excluyen. Si se pone 0,1 calcula la media aritmética del 90% de valores centrales. (quitando el 5% más altos y el 5% más bajos).

### **MEDIA.ARMO**

(Media armónica) Rango de valores de la variable.

### **MEDIA.GEOM**

(Media geométrica) Rango de valores de la variable.

### **MEDIANA**

Rango de valores de la variable.

### **MIN**

(Mínimo valor) Rango de valores de la variable.

### **MODA**

Rango de valores de la variable.

### **NORMALIZACION**

(Calcula valores tipificados). Habrá que darle como argumentos el valor de la variable y su correspondiente media y desviación típica.

### **PERCENTIL**

Los argumentos son: Matriz, que es el rango de valores de la variable, y K que es el percentil que se busca. Si se desea el percentil 10, K valdrá 0,1. Si se desea el percentil 90, K vale 0,9.

### **PROMEDIO**

(Media aritmética) Rango de valores de la variable.

### **PRONOSTICO**

(Predicción a partir de la recta de regresión). Suponiendo que se conoce el valor de X, qué valor tomará Y según la recta de regresión. Los argumentos son: X que es el valor que toma la variable independiente, Conocido\_X, que es el rango de valores de la variable X, y Conocido\_Y, que es el rango de valores de la variable Y.

### **VARP**

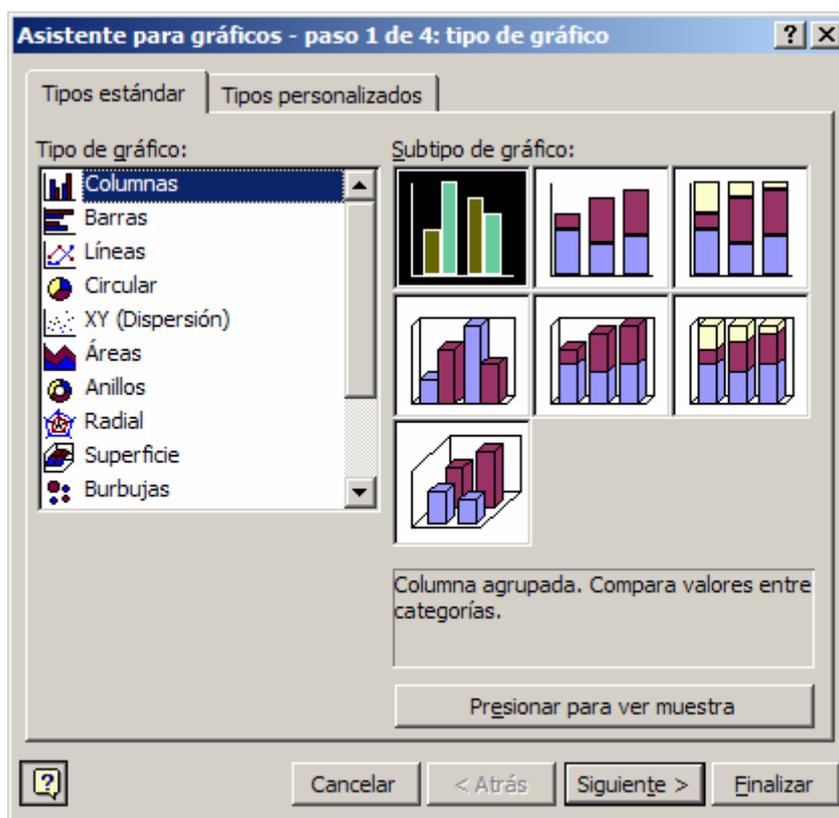
(Varianza) Rango de valores de la variable.

## REPRESENTACIONES GRÁFICAS

La hoja de cálculo EXCEL dispone de un asistente para gráficos que, mediante un sencillo cuadro de diálogo, permite elegir fácilmente la gráfica más adecuada a los datos que queremos representar. El asistente puede ser activado, indistintamente, a través de un icono en la barra de herramientas o mediante la opción Gráfico del apartado Insertar en el menú principal. Después tenemos que seguir los pasos que nos sugiere el asistente para gráficos:

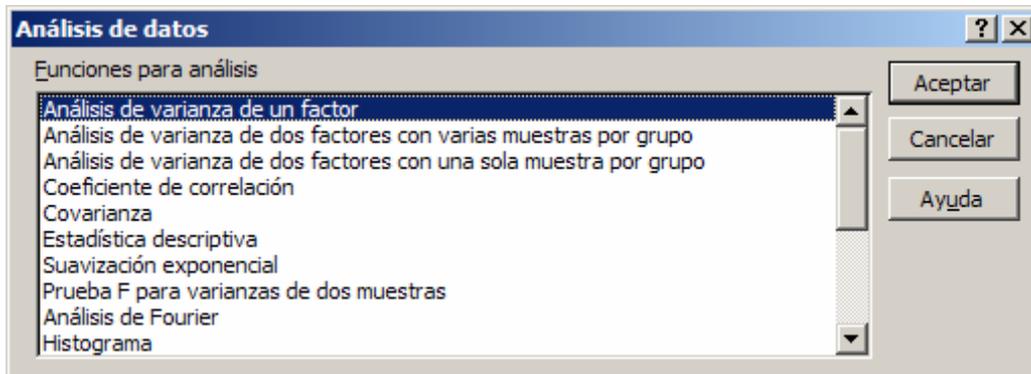
- Elegir tipo de gráfico
- Definir rango de datos
- Etiquetar ejes y gráfico, leyenda, etc.
- Decidir dónde se ha de representar el gráfico (en la hoja, en hoja nueva, etc.)
- Para cada variable, según sus propiedades de medida, elegimos uno o varios gráficos según los aspectos o detalles que queramos resaltar.
- Variables cualitativas: Barras, circular, columnas, anillos, ...
- Variables cuantitativas: Histograma (que aparece en Herramientas Analizar Datos)
- 2 variables cuantitativas: Dispersión XY
- 2 variables cualitativas: columnas en espacio tridimensional ( $x_i, y_j, n_{ij}$ )
- Listado de gráficos tipo estándar
- Listado de gráficos personalizados.

Una vez obtenido el gráfico, luego se puede editar y cambiar los aspectos de formato que nos interese: (rótulos, colores, rejillas, escalas, etc.). Incluso se le pueden añadir rótulos de textos, flechas o figuras, etc.

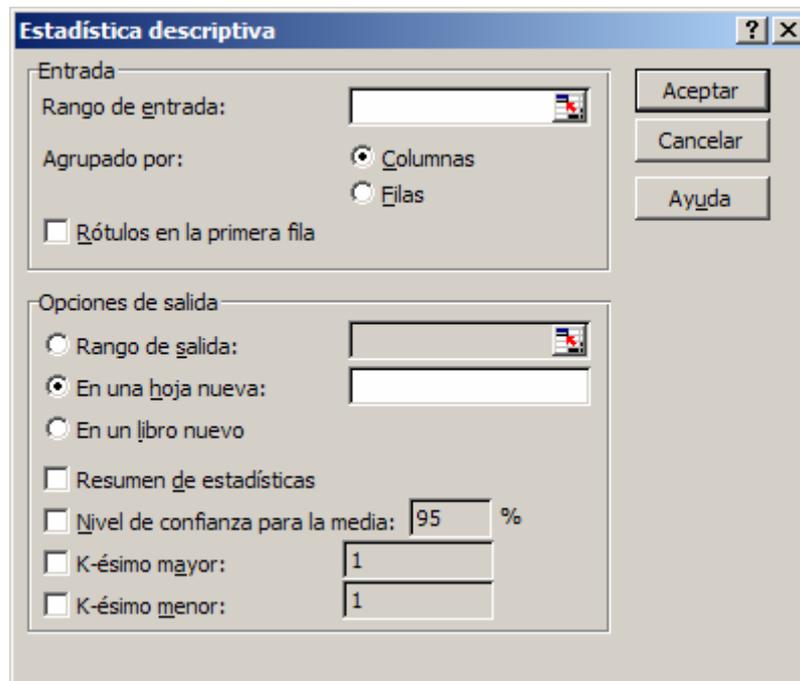


## USO DE LA HERRAMIENTA ANÁLISIS DE DATOS.

Es un complemento que no suele instalarse por defecto. Si en el menú Herramientas no aparece; habrá que instalarlo. El proceso sería: Herramienta + Complementos + Herramientas para Análisis. (Es posible que pida el disco de instalación del Office).

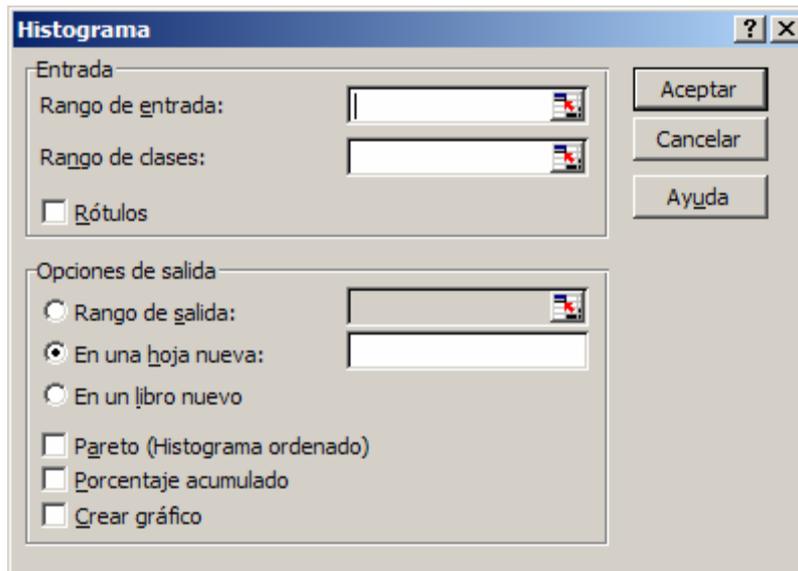


El procedimiento para realizar dichos análisis es rellenar los cuadros de diálogo.



Habrá que señalar el rango de entrada, indicarle si está en filas o columnas y si la primera celda corresponde al rótulo de la variable o no.

Habrá que señalar también el rango de salida, y seleccionar la opción resumen de estadísticas en la siguiente ventana:



Habrá que señalar el rango de entrada, indicarle si está en filas o columnas y si la primera celda corresponde al rótulo de la variable o no. Habrá que señalar también el rango de salida, y seleccionar la opción resumen de estadísticas.

En este caso habrá que señalar el rango de entrada que contiene los valores de la variable (se trata de distribuciones de frecuencias unitarias).

El rango de clases que es un conjunto de celdas donde se indican los extremos superiores del intervalo.

Pareto, ordena las clases o intervalos por la frecuencia de los mismos.

Porcentaje acumulado calcula las frecuencias acumuladas.

Todo ello se puede representar gráficamente marcando la opción crear gráfico.

La siguiente ventana:

The image shows a dialog box titled "Regresión" (Regression) with a blue header bar containing a help icon and a close button. The dialog is organized into several sections:

- Entrada (Input):** Contains two text boxes for "Rango Y de entrada" and "Rango X de entrada". Below these are three checkboxes: "Rótulos", "Constante igual a cero", and "Nivel de confianza" (set to 95%).
- Opciones de salida (Output options):** Contains three radio buttons: "Rango de salida", "En una hoja nueva" (which is selected), and "En un libro nuevo".
- Residuales (Residuals):** Contains four checkboxes: "Residuos", "Residuos estándares", "Gráfico de residuales", and "Curva de regresión ajustada".
- Probabilidad normal (Normal probability):** Contains one checkbox: "Gráfico de probabilidad normal".

On the right side of the dialog, there are three buttons: "Aceptar", "Cancelar", and "Ayuda".

Permite ajustar una recta de regresión a una distribución bidimensional de frecuencias.

Rango Y: Variable dependiente.

Rango X: Variable independiente.

Rótulos: Si se utilizan las primeras celdas para el nombre de la variable.

Constante igual a cero: Si se fuerza a la recta a pasar por el origen de coordenadas.

Habrá que dar rango de salida, y seleccionar las opciones de residuos, gráfico de residuales y curva de regresión ajustada.