

Práctica 1.

Introducción a las Hojas de Cálculo y Estadística Descriptiva (1ª Parte).

Objetivos.

En esta práctica trataremos de familiarizarnos con un programa de hojas de cálculo (EXCEL) mediante la creación de una tabla de frecuencias de una variable estadísticas discreta.

Índice.

Objetivos.	1
Índice.	1
1. Introducción a una hoja de cálculo.	2
1.1 Abrimos el programa.	2
1.2. Conceptos básicos de Excel.	5
2. Creación de una tabla de frecuencias de una variable estadística discreta.	5

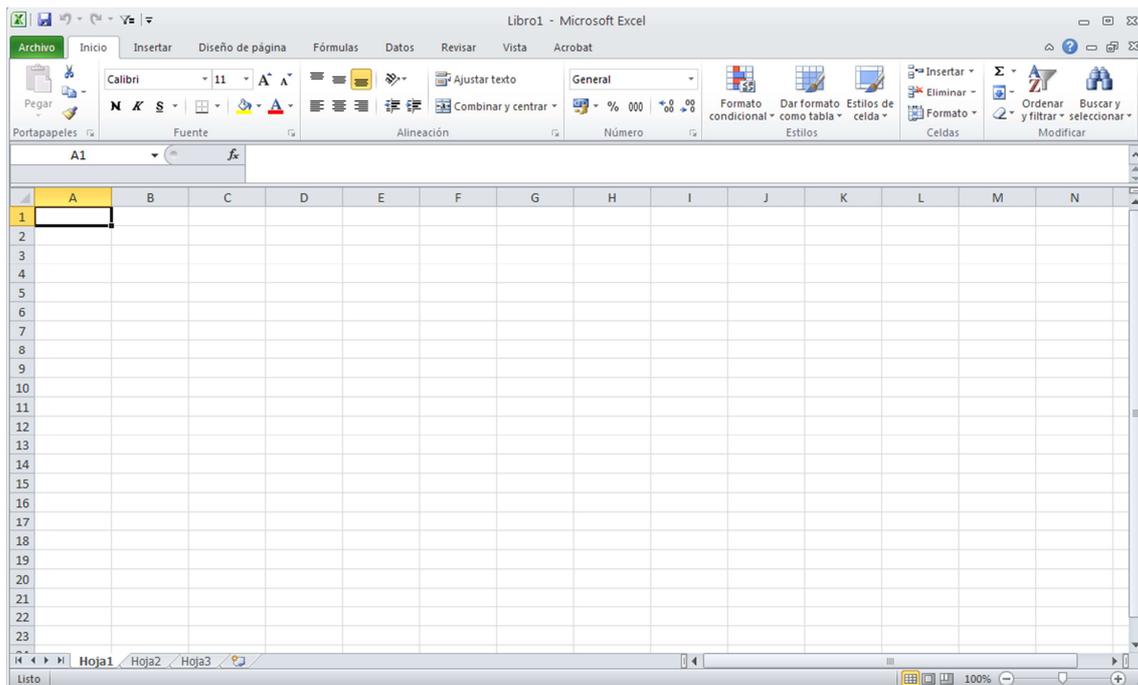
1. Introducción a una hoja de cálculo.

Utilizar un programa de Hojas de Cálculo para Estadística en lugar de un paquete propiamente estadístico (como por ejemplo SPSS) tiene unas ventajas y unos inconvenientes. La ventaja principal (que nosotros encontramos) es que aprenderemos a usar una herramienta que nos servirá para otras disciplinas (no sólo para Estadística). Quizás el principal inconveniente es que al ser un programa no propiamente estadístico puede ser muy engorroso realizar determinadas tareas.

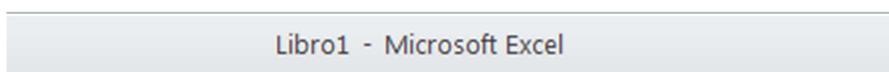
1.1 Abrimos el programa.

Si en el escritorio tenemos un acceso directo a Excel lo podemos hacer desde ahí, o bien, desde el botón de inicio → Todos los programas → Microsoft Office → Microsoft Excel 2010 (u otro en una ubicación similar).

La pantalla inicial nos muestra algo así como lo siguiente:



La barra de título.



Contiene nombre del documento sobre el que se está trabajando en ese momento.

La barra de acceso rápido.



La barra de acceso rápido tiene las operaciones más habituales que podemos hacer con Excel.

Es personalizable desde .

La banda de opciones.



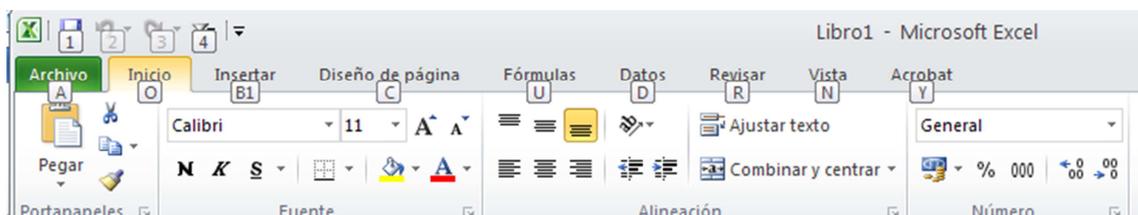
La Banda de opciones contiene todas las opciones del programa agrupadas en pestañas. Al hacer clic en **Insertar**, por ejemplo, veremos las operaciones relacionadas con la inserción de los diferentes elementos que se pueden crear en Excel.

Todas las operaciones se pueden hacer a partir de estos menús, pero las más habituales podríamos añadirlas a la barra de acceso rápido como hemos visto en el punto anterior.

En algunos momentos algunas opciones no estarán disponibles, las reconocerás porque tienen un color atenuado.

Las pestañas que forman la banda pueden ir cambiando según el momento en que te encuentres cuando trabajes con Excel. Está diseñada para mostrar solamente aquellas opciones que te serán útiles en cada pantalla. Haciendo clic sobre  que aparece en algunos bloques de la cinta nos presenta más opciones.

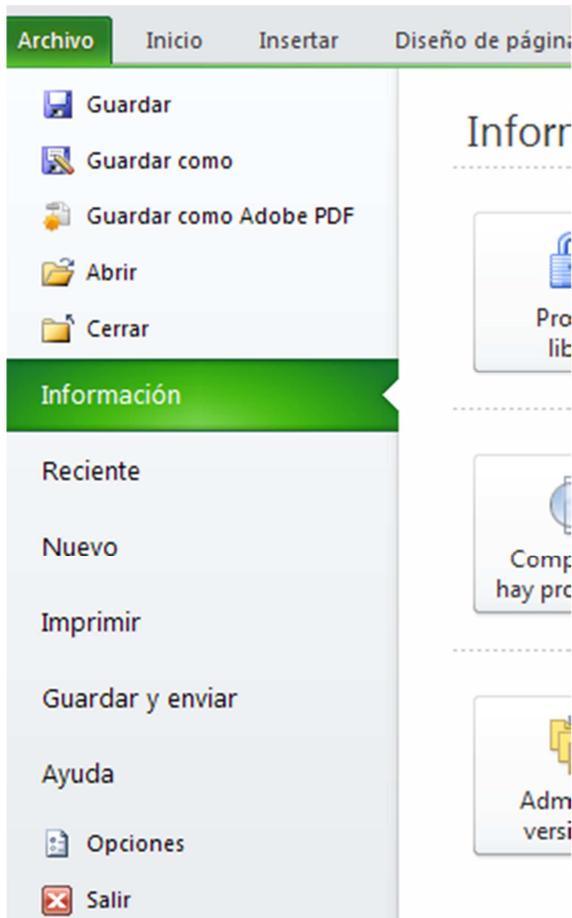
Pulsando la tecla **ALT** entraremos en el modo de acceso por teclado. De esta forma aparecerán pequeños recuadros junto a las pestañas y opciones indicando la tecla (o conjunto de teclas) que deberás pulsar para acceder a esa opción sin la necesidad del ratón.



Para salir del modo de acceso por teclado vuelve a pulsar la tecla **ALT**.

El Botón Archivo.

Haciendo clic en el botón  que se encuentra en la parte superior izquierda de la pantalla podrás desplegar un menú desde donde podrás ver las acciones que puedes realizar sobre el documento, incluyendo Guardar, Imprimir o crear uno Nuevo.



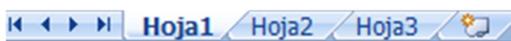
A este menú también puedes acceder desde el modo de acceso por teclado tal y como vimos para la Banda de opciones.

La barra de fórmulas.



Nos muestra el contenido de la celda activa, es decir, la casilla donde estamos situados. Cuando vayamos a modificar el contenido de la celda, dicha barra variará ligeramente, pero esto lo estudiaremos más adelante.

La barra de etiquetas.



Nos permite movernos entre las distintas hojas del libro de trabajo.

Las barras de desplazamiento.



Nos permite desplazarnos por el documento de una forma más rápida y sencilla.

1.2. Conceptos básicos de Excel.

Excel (al igual que otros programas de hojas de cálculo) trabaja con ficheros llamados *libros* que a su vez están formados por varias *hojas*. Por defecto un libro se crea con tres hojas (hoja1, hoja2 y hoja3). Un libro tiene que tener al menos una hoja. Cada hoja está formada por una matriz de *celdas* en la que almacenamos la información. Con las teclas del cursor podemos desplazarnos por las distintas celdas. En cada una de las celdas de la hoja, es posible introducir textos, números o fórmulas.

En una Hoja de Cálculo, los distintos *Tipos de datos* que podemos introducir son:

- **Valores constantes**, es decir, un dato que se introduce directamente en una celda. Puede ser un número, una fecha u hora, o un texto.
- **Fórmulas**, es decir, una secuencia formada por: valores constantes, referencias a otras celdas, nombres, funciones, u operadores. Es una técnica básica para el análisis de datos. Se pueden realizar diversas operaciones con los datos de las hojas de cálculo como +, -, x, /, Sen, Cos, etc... En una fórmula se pueden mezclar constantes, nombres, referencias a otras celdas, operadores y funciones. La fórmula se escribe en la barra de fórmulas y debe empezar por el signo =.

Cuando introducimos una fórmula en una celda puede ocurrir que se produzca un error. Dependiendo del tipo de error puede que Excel nos avise o no.

2. Creación de una tabla de frecuencias de una variable estadística discreta.

En esta sección vamos a empezar a introducir datos constantes y empezaremos a utilizar fórmulas sencillas. Para realizar esto vamos a plantear un ejercicio que iremos resolviendo poco a poco.

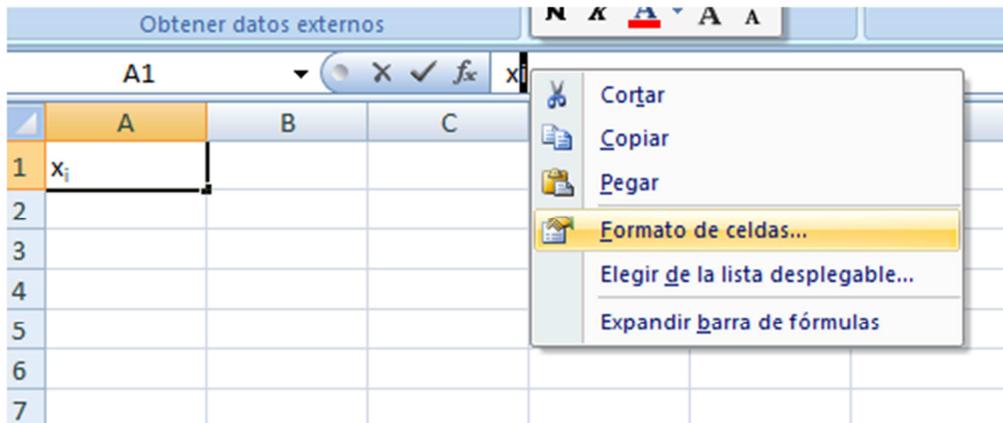
Ejercicio 1.

El número de convocatorias agotadas por un alumno hasta aprobar la asignatura de Estadística el año pasado fue el siguiente: con una convocatoria 30 alumnos, con dos convocatorias 12 alumnos, con tres 5, con cuatro 2, con cinco 0 y con seis 1. Se pide:

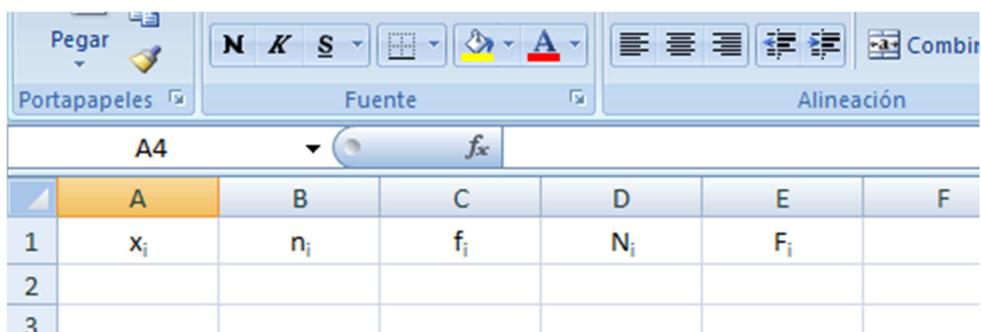
- ¿De qué tipo de variable estadística estamos hablando?
- Realiza una tabla de frecuencias en Excel.
- Representarla gráficamente mediante un diagrama de barras.
- Calcular la media, mediana y moda.
- Calcular la varianza (añadir columnas a la tabla si lo crees conveniente).
- Calcular el coeficiente de asimetría y el de apuntamiento de Fisher.

Obviamente la solución del apartado a) es que se trata de una variable discreta. El conjunto de posibles resultados que puede tomar esta variable es discreto, solo puede tomar los valores 1, 2, 3, 4, 5 y 6.

Pues procedamos a realizar una tabla de frecuencias utilizando Excel. Vamos a utilizar la primera fila para poner los encabezados de la tabla de frecuencias. En la primera columna pondremos los distintos valores x_i de la variable que en este caso corresponde al número de convocatorias agotadas. Para poder poner subíndices seleccionamos el texto que queremos poner con subíndices y con el botón izquierdo del ratón seleccionamos *Formato de celdas...* y posteriormente *subíndice*.



Y podemos crear la cabecera de otras cuatro columnas que corresponden a las frecuencias que hemos estudiado (frecuencia absoluta, frecuencia relativa, frecuencia absoluta acumulada y frecuencia relativa acumulada). Para una mejor presentación de la tabla podemos centrar todos los encabezados de las columnas pinchando en . Debemos tener un resultado como este:



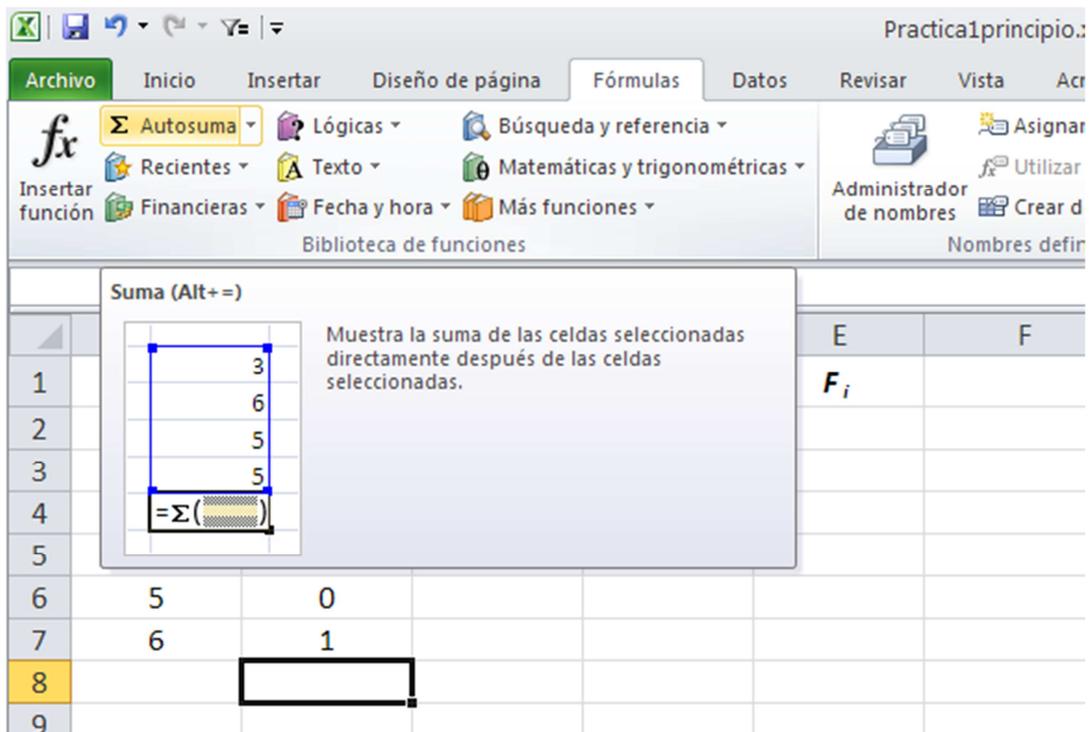
Vamos ahora a completar las dos primeras columnas con datos constantes. El Excel nos permite autocompletar una serie para que no tengamos que introducir todos los datos. Si después de poner los datos 1 y 2 seleccionamos dichas celdas (A2 y A3), pinchamos en la esquina inferior derecha hasta que aparece un cruz y, finalmente arrastramos hasta la fila 7, el programa nos completa las celdas restantes con los números siguientes, tal y como se ilustra en la figura.

	A
1	x_i
2	1
3	2
4	
5	
6	
7	
8	

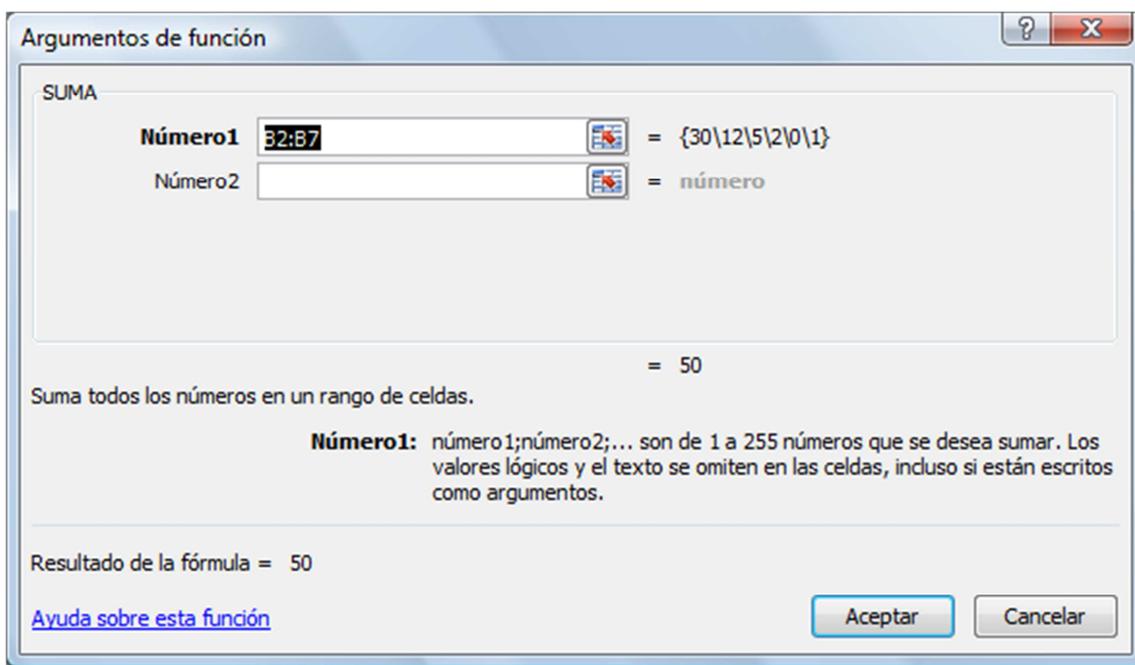
El resto de los datos no nos queda mas remedio que introducirlos a mano. El resultado debe ser el siguiente:

	A	B	C	D	E
1	x_i	n_i	f_i	N_i	F_i
2	1	30			
3	2	12			
4	3	5			
5	4	2			
6	5	0			
7	6	1			
8					
9					

Vamos a introducir nuestra primera fórmula. El número de observaciones total es la suma de todos los n_i . Pues vamos a poner esa cantidad en la celda B8. Mediante la función **SUMA()** excel nos devuelve la suma de todos los argumentos que pongamos, los argumentos deben separarse por “;” (punto y coma). Sin embargo esta función también admite como argumento toda una región (o varias). Para indicar una región en Excel lo hacemos indicandoc la referencia de la celda más arriba y a la izquierda y las celda más abajo y a la derecha separadas por “:” (dos puntos). La fórmula que debemos introducir en la casilla B8 sería = **SUMA(B2:B7)**. Sin embargo en lugar de introducir todo esto es suficiente situarnos en la celda B8 y pinchar en el icono de autosuma de la cinta de fórmulas **Σ Autosuma**. El programa interpreta que queremos sumar los valores numéricos que hay encima de la casilla B8.



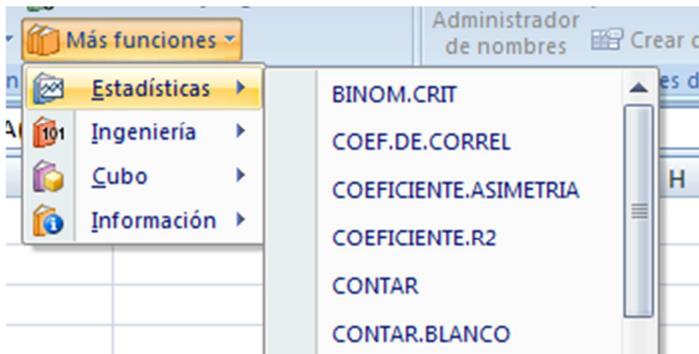
También pudimos hacerlo eligiendo la función suma desde las *funciones matemáticas y trigonométricas*. Se nos abre un formulario donde podemos poner los argumentos de cada función (en este caso de la función SUMA).



Pinchando en  podemos seleccionar la celda o región de celdas que queremos sumar. Por defecto nos aparece la región que queremos sumar es decir B2:B7. Desde este menú podemos ver de antemano cuánto va a dar el resultado (50) y solicitar ayuda sobre esta función.

Desde esta cinta de fórmulas podemos acceder a todas las funciones de que dispone el programa. Están divididas por categorías (*Usadas recientemente, Financieras, Lógicas, Texto,*

Fecha y hora, Búsqueda y referencia, Matemáticas y trigonométricas y Más funciones). Precisamente en Más funciones podemos elegir las funciones estadísticas que utilizaremos en esta y otras prácticas.



Bueno pues ahora que tenemos el número total de observaciones (N) podemos calcular la frecuencia relativa mediante la fórmula $f_i = n_i / N$. Así la casilla C2 debe ser igual a B2/B8, es decir debemos teclear = B2/B8 en la casilla C2 y el resto de forma análogo.

Fijense que cuando la celda C2 tecleamos =B2/B8 aparecen las celdas marcadas con colores. Esto nos ayudará a no equivocarnos con la fórmula.

	A	B	C	D
1	x_i	n_i	f_i	N_i
2	1	30	=B2/B8	
3	2	12		
4	3	5		
5	4	2		
6	5	0		
7	6	1		
8		50		
9				

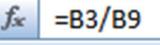
Seguidamente nos aparecerá el resultado en la celda C2 . Podemos cambiar el formato. Por ejemplo vamos a poner el resultado con dos cifras decimales, lo podemos hacer desde la cinta de inicio pinchando en para aumentar una cifra decimal. Otra posibilidad de hacer lo mismo es pinchando en el botón derecho del ratón y marcando y en la pestaña número elegimos número y marcamos 2 cifras decimales (). Desde este formulario podemos hacer cualquier cambio respecto al formato de la celda (o grupo de celdas en el caso de haber seleccionado varias).

¿Qué ocurre si copiamos y pegamos una fórmula? Hagámoslo. Vamos a copiar el contenido de

la celda C2 en la celda C3. Esto lo podemos hacer de varias formas: 1) Seleccionando la celda C2 pinchamos en  y luego seleccionamos la celda C3 y pinchamos en , 2) pinchamos al botón derecho sobre la celda y seleccionamos  Copiar y luego pinchamos al botón derecho sobre la celda C3 y seleccionamos  Pegar, 3) con la celda C2 seleccionada pulsamos la combinación de teclas Ctrl + C y con la celda C3 seleccionada pulsamos Ctrl + V, y 4) Teniendo seleccionada la casilla C2 arrastrarla hasta la casilla C3.

Bueno, haciendo cualquiera de estas opciones nos aparecerá lo siguiente.

n_i	f_i	N_i
30	0,60	
12	#DIV/0!	
5		
2		

Se ha producido un error de división por cero. Lo que ocurre es que el programa al copiar toda la fórmula de forma relativa, es decir, como nos hemos situado en una fila más abajo (en la casilla C3 en lugar de C2) ha añadido una fila a la fórmula que teníamos en la casilla C2, con esto la referencia que hacíamos a B2 ha pasado a ser B3 y la referencia que teníamos a B8 ha pasado a ser B9 . En realidad el primer valor si queremos que sea B3 pero el segundo debería seguir siendo B8. Esto lo podemos solucionar diciéndole a Excel que la casilla B8 se tiene que tomar de forma absoluta cuando copiamos una fórmula, una forma de hacerlo es marcando con el signo \$ la fila y la columna de la referencia B8 es decir \$B\$8 (en realidad nos bastaría con marcar como absoluta la fila es decir B\$8).

¡Ahora sí! Si copiamos la celda C2 a la celda C3 obtenemos la fórmula $=B3/\$B\8 que nos da el resultado de 0,24.

Copiamos esta fórmula al resto de casillas de la columna C. Para ello bastaría con seleccionar la casilla C3 y arrastrarla hasta la casilla C7 obteniendo el siguiente resultado:

	A	B	C	D	E
1	x_i	n_i	f_i	N_i	F_i
2	1	30	0,60		
3	2	12	0,24		
4	3	5	0,10		
5	4	2	0,04		
6	5	0	0,00		
7	6	1	0,02		
8		50			
9					

Se deja como ejercicio que ustedes completen las celdas de la frecuencia absoluta acumulada.

El resultado después de completar el resto de frecuencias debe ser:

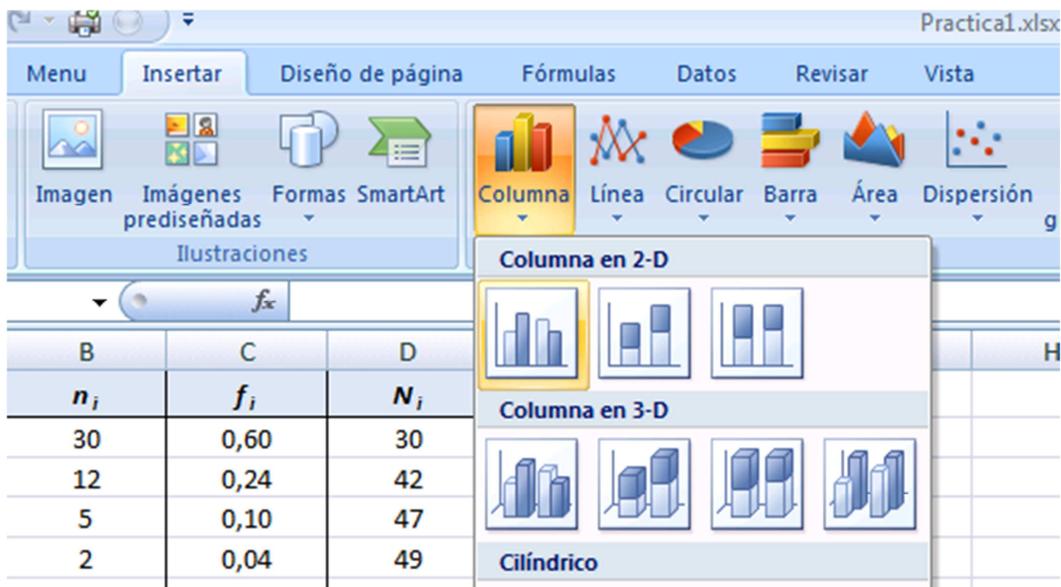
C16		f _x				
	A	B	C	D	E	
1	x _i	n _i	f _i	N _i	F _i	
2	1	30	0,60	30	0,60	
3	2	12	0,24	42	0,84	
4	3	5	0,10	47	0,94	
5	4	2	0,04	49	0,98	
6	5	0	0,00	49	0,98	
7	6	1	0,02	50	1,00	
8		50	1			

Podemos darle otro formato para que el resultado final sea más atractivo. Digamos hasta obtener algo como esto:

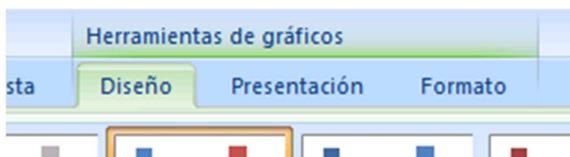
C11		f _x				
	A	B	C	D	E	
1	x _i	n _i	f _i	N _i	F _i	
2	1	30	0,60	30	0,60	
3	2	12	0,24	42	0,84	
4	3	5	0,10	47	0,94	
5	4	2	0,04	49	0,98	
6	5	0	0,00	49	0,98	
7	6	1	0,02	50	1,00	
8		50	1			

Construcción de un digrama de barras. Vamos a hacerlo de las frecuencias absolutas (pero de forma análoga lo podríamos hacer de la frecuencias relativas, absolutas acumuladas o relativas acumuladas).

Desde la cinta Insertar, seleccionamos [Columna](#) → [Columna Agrupada](#) o cualquiera de los gráficos de columnas.

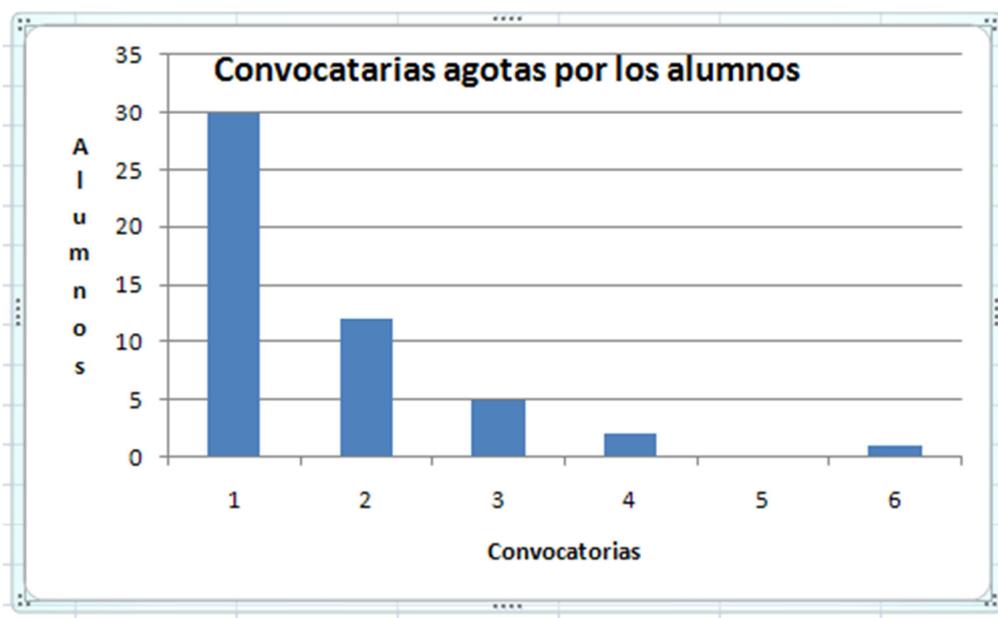


Una vez seleccionado el gráfico se habilitan las [Herramientas de gráficos](#) que la forman tres pestañas ([Diseño](#), [Presentación](#) y [Formato](#)) que nos dan la posibilidad de configurar y cambiar el gráfico.



Desde [Diseño](#), picamos en [Seleccionar datos](#). Y agregamos los datos, podemos elegir como nombre de la serie la celda [B1](#) (n_i) y valores de la serie la celdas [B2:B7](#). Automáticamente el Excel interpreta para el eje X los valores de las celdas [A2:A7](#) (x_i = número de convocatorias).

Desde estas herramientas podemos modificar el gráfico para poner un título personalizado o personalizar los ejes de coordenadas. Al final podemos llegar a este resultado:



Procedamos a resolver el apartado **d)** del ejercicio. Bueno, la moda es el valor que más se repite. Simplemente echando una ojeada al diagrama de barras nos daremos cuenta que ese valor es 1 (con 30 observaciones). La mediana es el valor que separa la mitad de las observaciones menores de la mitad mayores, en nuestro caso como el número de observaciones es 50 valdría la que ocupa la posición 25 y 26 (o el punto medio de estas). Pero como con 1 convocatoria hay más de la mitad el dato 25º y 26º son iguales a 1 y esa es la mediana.

Para la media ($\frac{1}{N} \sum_{i=1}^k x_i n_i$) vamos a añadir nuevas columnas en la tabla que nos facilitan el cálculo. Hasta obtener el siguiente resultado:

Portapapeles						
Fuente						
Alineación						
F12		=F8/B8				
	A	B	C	D	E	F
1	x_i	n_i	f_i	N_i	F_i	$x_i n_i$
2	1	30	0,60	30	0,60	30
3	2	12	0,24	42	0,84	24
4	3	5	0,10	47	0,94	15
5	4	2	0,04	49	0,98	8
6	5	0	0,00	49	0,98	0
7	6	1	0,02	50	1,00	6
8	SUMA	50	1			83
9						
10					moda	1
11					mediana	1
12					media	1,66
13						

Para el apartado **e)** tenemos que calcular la varianza, $\sigma^2 = \frac{1}{N} (\sum_{i=1}^k x_i^2 n_i) - \bar{x}^2$. Así podemos crear una nueva columna con $x_i^2 n_i$ y realizar las operaciones mediante la fórmula

=G8/B8-F12*F12

. También podríamos utilizar la fórmula $\sigma^2 = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^2 n_i$.

El apartado **f)** se puede resolver de forma análoga a los dos anteriores creando las columnas adecuadas y teniendo en cuenta que $A_F = \frac{\frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^3 n_i}{\sigma^3}$.

Práctica 2.

Introducción a las Hojas de Cálculo y Estadística Descriptiva (2ª Parte).

Objetivos.

En esta práctica continuaremos utilizando el programa Excel para realizar estadística Descriptiva. Ahora haremos más hincapié en variables estadísticas continuas.

Índice.

Objetivos.....	1
Índice.....	1
1 Creación de una tabla dinámica.....	2
1.1 Creación de una Tabla.....	2
1.2 Creación de una Tabla Dinámica.....	3
2 Agrupación de datos de una variable estadística continua.....	6
2.1 Agrupar datos y construir una tabla de frecuencias.....	6
2.2 Histograma.....	10
2.3 Cálculo de la media, varianza, desviación típica y coeficiente de variación de Pearson. 12	
2.4 Cálculo de la Moda con los datos agrupados.....	12
2.5 Cálculo de un percentil con los datos agrupados.....	13
2.6 Observaciones finales.....	14
3 Cálculo de medidas estadísticas de un conjunto de datos.....	14
4 Módulo de Análisis de datos de Excel.....	15
4.1 Estadística Descriptiva.....	16
4.2 Histograma.....	18

1 Creación de una tabla dinámica.

Excel tiene una herramienta que permite realizar tablas de forma automática. Mediante las [Tablas Dinámicas](#) podemos realizar una tabla (de frecuencias u otra) para un conjunto de datos. Para ver esto vamos a utilizar el fichero  [EjemploTablaDinamica.xls](#). Es fichero contiene una hoja denominada **Datos**, que contiene la **edad**, **sexo** y **nota** de 50 alumnos.

Ya hemos dicho antes que el Excel no es un programa propiamente estadístico, por lo generalmente permite cosas que un programa estadístico no permitiría. Por ejemplo, los datos de una variable estadística siempre son del mismo tipo. Es decir en este ejemplo la variable **edad** es numérica (en este caso sin valores decimales), la variable **sexo** es cualitativa (puede tomar los valores **mujer** o **hombre**) y la variable **nota** es numérica (con un decimal). En el Excel nosotros podríamos poner texto en un dato de **edad** y no pasaría nada mientras que en un programa propiamente estadístico esto no lo podríamos hacer. Todos los datos de una misma variable tienen que ser del mismo tipo. De igual forma aquí podemos colocar el conjunto de datos donde nosotros queramos (sin tener que empezar por la segunda fila). Aún así vamos a tratar de ser algo rigurosos con el tratamiento de datos y trataremos de colocar en una hoja los datos correspondientes a un conjunto de individuos (población estadística), utilizaremos la primera fila para poner el nombre de las variables, los datos de cada columna deben ser del mismo tipo y no utilizaremos esa hoja para introducir fórmulas u otros cálculos (salvo que por razones pedagógicas queramos tener en la misma pantalla los datos y los cálculos).

	A	B	C	D
1	Alumno	Edad	Sexo	Nota
2	1	18	Hombre	2,9
3	2	20	Hombre	7,4
4	3	18	Hombre	4,3
5	4	18	Mujer	8,1
6	5	19	Mujer	4.1

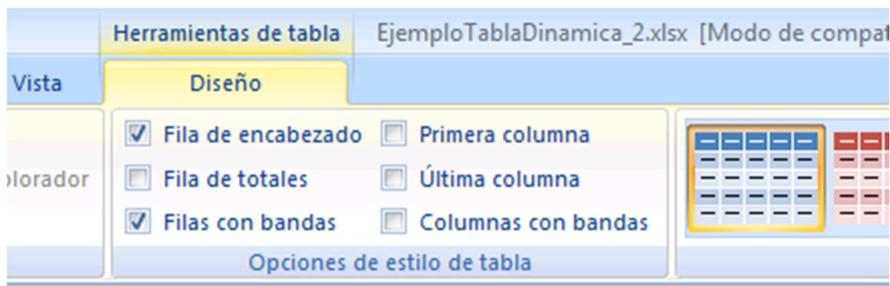
1.1 Creación de una Tabla.

Otra cosa que podemos hacer con los datos es modificar su presentación (así como agregar filas o columnas resumen). Si seleccionamos todos los datos y vamos a [Insertar](#) → [Tabla](#) podemos hacer estos cambios.

	A	B	C	D
1	Alumno	Edad	Sexo	Nota
2	1	20	Mujer	4,7
3	2	18	Hombre	6,9
4	3	19	Hombre	6,4
5	4	18	Mujer	4,6
6	5	18	Mujer	8.5

Una vez seleccionado la tabla se habilitan las [Herramientas de tabla](#) que la forma una pestaña ([Diseño](#)). Las herramientas de tabla nos permiten mecanizar muchas de las operaciones que

necesitamos hacer con los datos de una tabla.



Mediante una tabla las operaciones que hagamos con en una fila automáticamente se actualiza en todas ellas.

1.2 Creación de una Tabla Dinámica.

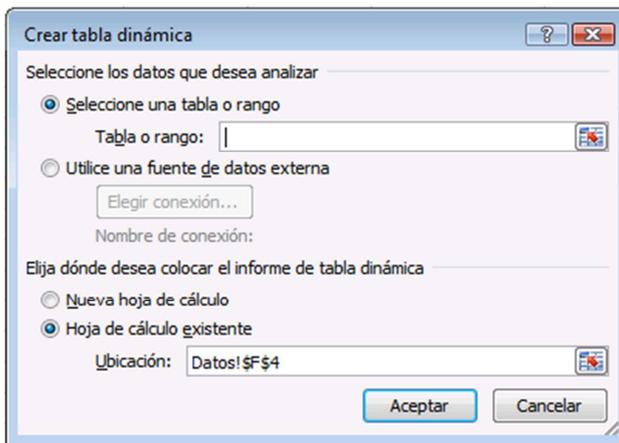
Una [tabla dinámica](#) nos permite agrupar los datos de forma automática. Para ello vamos a realizar el siguiente ejercicio:

Ejercicio 1.

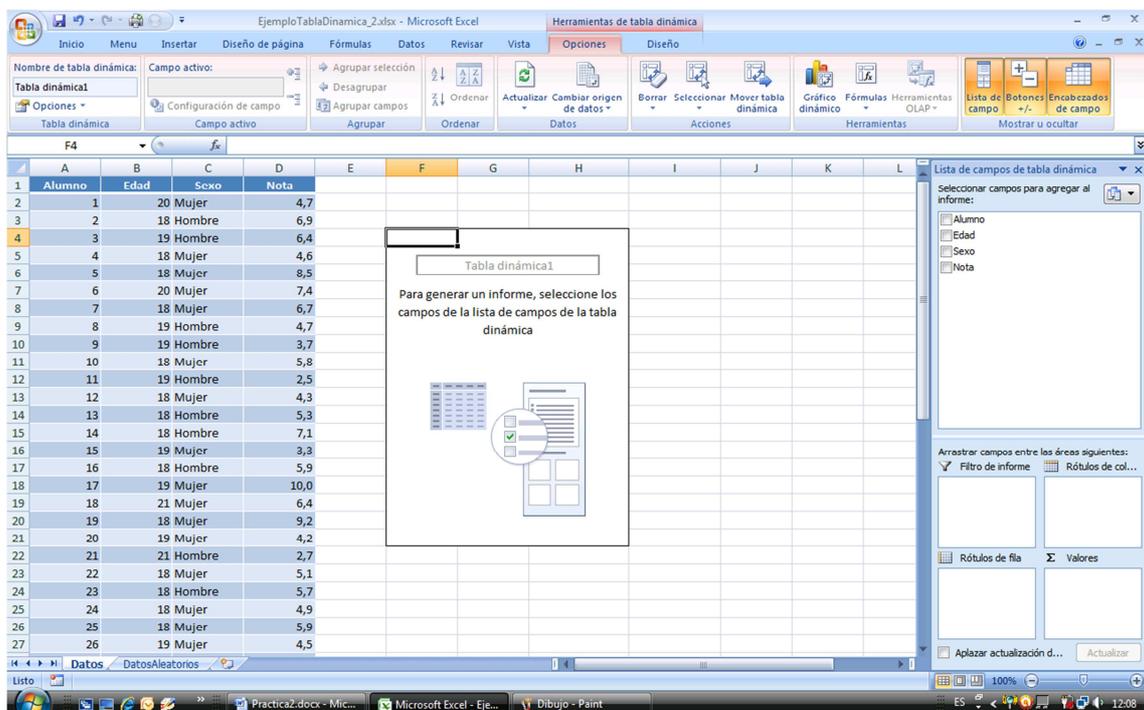
Utiliza el fichero [EjemploTablaDinamica.xlsx](#) para realizar mediante una Tabla Dinámica lo siguiente:

- Una tabla de frecuencias de la variable sexo. Muestra también la nota media por sexo.
- Una tabla de frecuencias de la variable edad. Muestra la nota media por edad.
- Una tabla bidimensional de la variables sexo y edad.
- Modifica algunos valores de la tabla de datos, ¿se modifican también en la tabla dinámica?

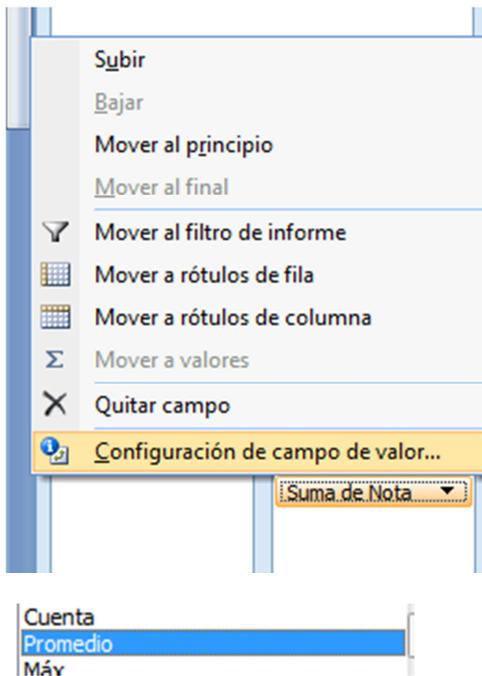
A partir de los datos anteriores vamos a realizar el siguiente ejercicio. Vamos a insertar una tabla dinámica para realizar el apartado [a\)](#), para ello nos situamos en una celda vacía y vamos a la pestaña [Insertar ->Tabla dinámica](#). Se nos presenta el siguiente formulario:



Vamos a seleccionar todas las celdas de la tabla actual y, aunque sería más lógico realizar la tabla en una **nueva hoja**, para poder visualizar la tabla resumen (tabla dinámica) y los datos a la vez lo vamos a hacer en la **hoja** de cálculo ya existente.



Al realizar esto se presenta dos nuevas pestañas con **herramientas de tabla dinámica** (**Opciones** y **Diseño**) y un formulario a la derecha en el que aparecen las distintas variables que podemos arrastrar a los cuadrados de abajo. Vamos a arrastrar **sexo** a **Rótulos de fila** y la misma la vamos a poner en **Valores**. Observen como automáticamente aparece el recuento de observaciones de los dos posibles valores que toma esta variable (carácter). También la variable **Nota** la arrastramos a **Valores**. En este caso pulsando en la flecha hacia abajo que aparece en la variable **Nota** y seleccionando **Configuración de campo de valor** podemos cambiar la función de la variable **Nota** a representar y así que represente la media (**Promedio**) en lugar de la **suma**.



El resultado final debe ser una cosa como la siguiente:

Valores		
Rótulos de fila	Cuenta de Sexo	Promedio de Nota
Hombre	21	5,14
Mujer	29	5,63
Total general	50	5,42

De forma muy similar se pueden realizar el resto de apartados de este ejercicio. Para el apartado **b)** tenemos:

Valores		
Rótulos de fila	Cuenta de Sexo	Promedio de Nota
18	25	5,77
19	15	5,01
20	5	5,88
21	4	3,93
22	1	6,60
Total general	50	5,42

Para el apartado **c)**:

Cuenta de Alumno		Rótulos de columna		
Rótulos de fila	Hombre	Mujer	Total general	
18	8	17	25	
19	9	6	15	
20	1	4	5	
21	2	2	4	
22	1		1	
Total general	21	29	50	

Observen que para el apartado d), para que se actualice la tabla dinámica es necesario pinchar



en  de la pestaña .

2 Agrupación de datos de una variable estadística continua.

Para ver esto vamos a utilizar el fichero **Cubiertas.xlsx** que contiene los cien datos del ejercicio 8 de la hojas de problemas del Tema 2 e intentaremos resolver dicho ejercicio utilizando Excel.

Ejercicio 2.

En el fichero **Cubiertas.xlsx** se han introducido los datos de talleres correspondientes a los miles de de kilómetros, recorridos por vehículos con un modelo particular de neumático, antes de un pinchazo o un reventón.

Se pide:

- Construir la tabla de frecuencias usando la fórmula de Sturges para determinar el número de intervalos. Esta expresión establece que, dado un conjunto de N datos ($N < 500$), el número ideal k de intervalos de clase puede aproximarse por $k = [1 + \log N / \log 2]$. En este ejemplo, $k = 8$.
- Dibujar el histograma de frecuencias relativas acumuladas y sin acumular.
- Calcular e interpretar los resultados para la media aritmética, la mediana y la moda.
- Calcular e interpretar los resultados para la desviación típica y el coeficiente de variación de Pearson.
- Analiza la asimetría y la curtosis de la distribución.
- Si el fabricante de este modelo de neumáticos quiere proponer un kilometraje para realizar el cambio de cubiertas, ¿qué valor propondría para que sólo 3 de cada 10 vehículos puedan tener un pinchazo o reventón antes de ese kilometraje?

2.1 Agrupar datos y construir una tabla de frecuencias.

Procedamos a agrupar los datos en intervalos de clase. Para ello vamos antes a calcular algunas medidas de los datos mediante las fórmulas oportunas:

Mínimo
Máximo
N
Nº de clases
Amplitud

Para el mínimo de clases utilizaremos la función `MIN()`.

A	B	C	D
Km			
52.452			
50.432	Mínimo	=MIN(A:A)	
37.748	Máximo		
51.831	N		
73.808	Nº de clases		

La referencia `A:A` (como no contiene el número de fila) se refiere a todas las celdas de la columna A. Así por ejemplo `A:C` se refiere a todas las celdas de las columnas de la A a la C.

Para obtener el máximo utilizaremos la función `MAX()` y para el número de datos N vamos a utilizar `CONTAR()`. Para el número de clases vamos a utilizar la función que nos da el ejercicio $k = \lceil 1 + \log N / \log 2 \rceil$ (pero con las funciones del Excel). Mediante la expresión `=CEILING(1+LN(D5)/LN(2))` me devuelve el número **7,64385619**. Ahora sólo faltaría redondear al entero por arriba. La función `ENTERO()` nos redondea al entero por abajo, pero también podemos utilizar la función `REDONDEAR.MAS()` que nos redondea por arriba según el número de cifras decimales especificadas (si ponemos 0 nos redondeará al entero por arriba). Así con `=REDONDEAR.MAS(1+LN(D5)/LN(2);0)` obtenemos lo que queríamos.

Vamos a crear las clases en función de estos datos. Por defecto vamos a poner un número grande de clases aunque no vayamos a utilizar sino 8. Vamos a poner 20 (por ejemplo).

Clase	Extremo Inf.	Extremo Sup.
1	10.504,0	20.511,6
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		

Y para los siguientes extremos vamos a coger el anterior más la amplitud de clase, siempre y cuando estemos entre las 8 primeras clases. Así mediante la fórmula

=SI(\$F3<=\$D\$6;G2+\$D\$7;"")	
SI(prueba_lógica; [valor_si_verdadero]; [valor_si_falso])	
Clase	Extremo Inf.

conseguimos lo que queremos.

Fíjense que las casillas D6 y D7 las hemos tomado de forma absoluta (con \$) y la columna de la casilla F3 también. De esta forma conseguimos que copiando la fórmula obtengamos los extremos de los intervalos. Sólo falta copiar la fórmula a los extremos y sorprendentemente a partir de la 8ª clase no nos aparecen datos (esto lo hemos conseguido gracias a la función SI()).

Procedamos a calcular ahora la marca de clase (x_i) y las distintas frecuencias (n_i , f_i , N_i y F_i). Para ello utilizaremos el mismo truco de la función SI para que aparezcan los datos sólo en las k primeras clases.

El cálculo de la marca de clase es muy sencillo es el valor medio de los extremos de la clase. Utilizamos la siguiente fórmula para poderla copiar =SI(\$F2<=\$D\$6;(G2+H2)/2;"").

Para las frecuencias vamos a proceder primero calculando la frecuencia absoluta acumulada N_i . Para ello vamos a utilizar la función CONTAR.SI(). Esta función nos cuenta el número de datos que verifican una determinada condición, la condición que queremos que verifique es que los datos sean menores que el extremo superior de la clase. Lo escribimos de la siguiente forma CONTAR.SI(A:A;"<"&H2). Estamos contando los datos de la columna A que verifican que son menores (tenemos que ponerlo entre comillas) que el contenido de la celda H2, esto hay que indicarlo en Excel con el símbolo &. Sin embargo para el último intervalo (la clase 8 en este caso) lo debemos coger cerrado por arriba y a partir de de la clase 8 no deben aparecer datos.

Esto lo podemos hacer mediante dos funciones SI() anidadas. Quedando al final la siguiente fórmula:

`=SI($F2<$D$6;CONTAR.SI(A:A;"<"&H2);SI($F2=$D$6;CONTAR.SI(A:A;"<="&H2);""))`

Al copiar tenemos lo que buscábamos:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Km					Clase	Extremo Inf.	Extremo Sup.	x_i	n_i	f_i	N_i	F_i
2	52.452					1	10.504,0	20.511,6	15.507,8			1	
3	50.432	Mínimo	10.504			2	20.511,6	30.519,3	25.515,4			4	
4	37.748	Máximo	90.565			3	30.519,3	40.526,9	35.523,1			15	
5	51.831	N	100			4	40.526,9	50.534,5	45.530,7			40	
6	73.808	Nº de clases	8			5	50.534,5	60.542,1	55.538,3			58	
7	61.065	Amplitud	10007,625			6	60.542,1	70.549,8	65.545,9			82	
8	35.807					7	70.549,8	80.557,4	75.553,6			93	
9	57.277					8	80.557,4	90.565,0	85.561,2			100	
10	48.698					9							
11	65.854					10							
12	75.850					11							
13	36.949					12							
14	75.548					13							
15	69.010					14							
16	61.477					15							
17	65.585					16							
18	44.411					17							
19	41.886					18							
20	34.754					19							
21	59.888					20							

Una vez calculadas las frecuencias N_i , el resto es un juego de niños. Recordad que nos interesa que en una clase posterior a k (8 en este caso) no nos muestre datos. Para el cálculo de $n_i = N_i$ y el resto $n_i = N_i - N_{i-1}$.



Las frecuencias relativas (f_i y F_i) son todavía más sencillas de calcular. Al final debemos tener lo siguiente:

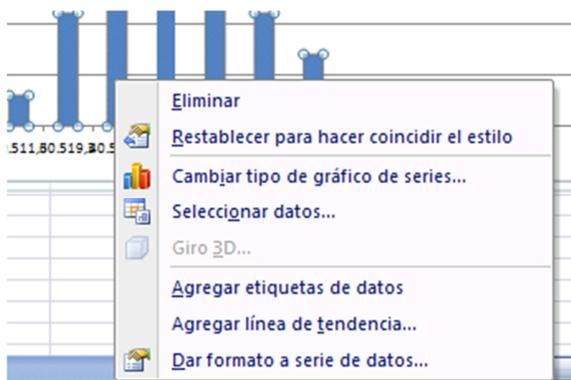
	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1				Clase	Extremo Inf.	Extremo Sup.	x_i	n_i	f_i	N_i	F_i				
2				1	10.504,0	20.511,6	15.507,8	1	0,01	1	0,01				
3	Mínimo	10.504		2	20.511,6	30.519,3	25.515,4	3	0,03	4	0,04				
4	Máximo	90.565		3	30.519,3	40.526,9	35.523,1	11	0,11	15	0,15				
5	N	100		4	40.526,9	50.534,5	45.530,7	25	0,25	40	0,4				
6	Nº de clases	8		5	50.534,5	60.542,1	55.538,3	18	0,18	58	0,58				
7	Amplitud	10007,625		6	60.542,1	70.549,8	65.545,9	24	0,24	82	0,82				
8				7	70.549,8	80.557,4	75.553,6	11	0,11	93	0,93				
9				8	80.557,4	90.565,0	85.561,2	7	0,07	100	1				
10				9											
11				10											
12				11											
13				12											
14				13											
15				14											
16				15											
17				16											
18				17											
19				18											
20				19											
21				20											
22															

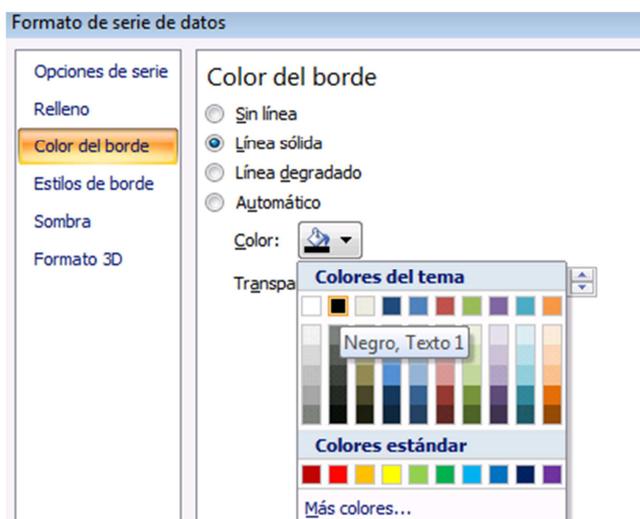
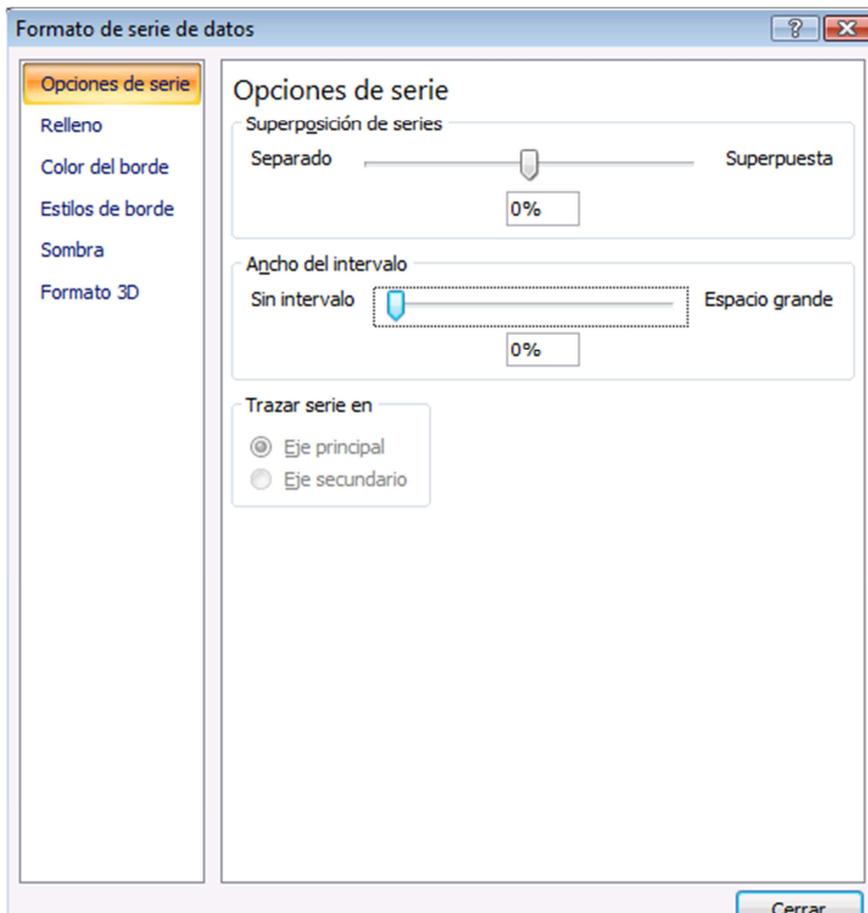
Ya tenemos la tabla de frecuencias.

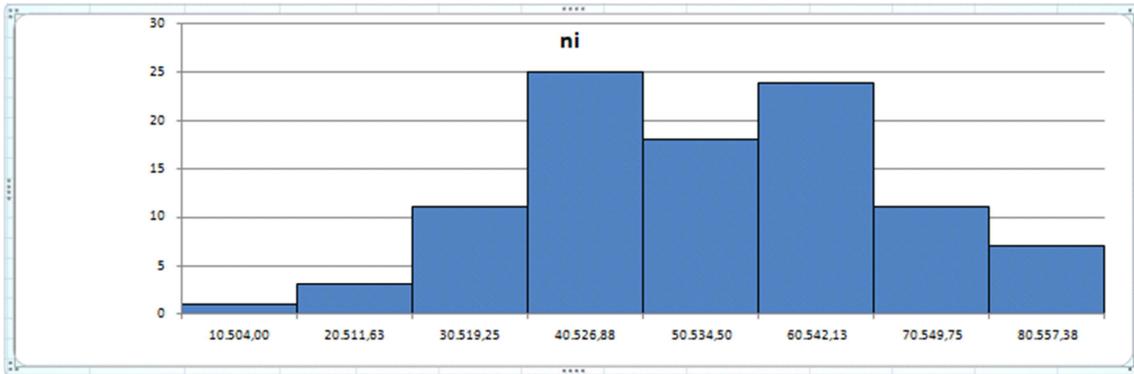
2.2 Histograma.

Para realizar una gráfica del histograma procedemos de forma similar a la práctica 1. En **Insertar** → **Gráficos de columnas** seleccionamos uno, señalamos los datos.

Pulsando al botón derecho del ratón sobre las columnas del gráfico podemos modificar el espaciado entre las columnas (en un histograma no debe haber espacios entre cada rectángulo) y darle un borde sólido a cada columna (rectángulo).







2.3 Cálculo de la media, varianza, desviación típica y coeficiente de variación de Pearson.

Para calcular la media y varianza podemos proceder como en la práctica anterior que teníamos datos de una variable discreta. En este caso he calculado una columna $x_i * n_i$ y otra $x_i^2 * n_i$.

Hemos calculado la media, varianza, desviación típica y coeficiente de variación de Pearson utilizando los datos agrupados y marcas de clase.

Clase	Extremo Inf.	Extremo Sup.	x_i	n_i	f_i	N_i	F_i	$x_i * n_i$	$x_i^2 * n_i$
1	10.504,0	20.511,6	15.507,8	1	0,01	1	0,01	15507,8125	240492248,5
2	20.511,6	30.519,3	25.515,4	3	0,03	4	0,04	76546,3125	1953112652
3	30.519,3	40.526,9	35.523,1	11	0,11	15	0,15	390753,688	13880767663
4	40.526,9	50.534,5	45.530,7	25	0,25	40	0,4	1138267,19	51826087606
5	50.534,5	60.542,1	55.538,3	18	0,18	58	0,58	999689,625	55521074796
6	60.542,1	70.549,8	65.545,9	24	0,24	82	0,82	1573102,5	1,0311E+11
7	70.549,8	80.557,4	75.553,6	11	0,11	93	0,93	831089,188	62791748871
8	80.557,4	90.565,0	85.561,2	7	0,07	100	1	598928,313	51245017645
SUMA				100	1			5.623.884,6	340.568.779.627,8

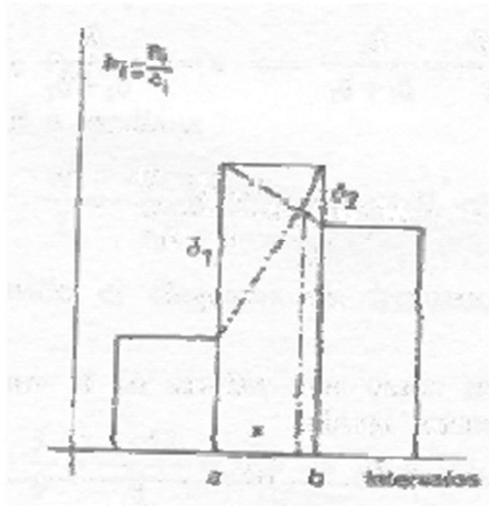
Estadística	Valor
Media agrupada	56.238,8
Varianza agrupada	242.879.968,7
Desviación típica A	15.584,6
CV Agrupado	0,2771

2.4 Cálculo de la Moda con los datos agrupados.

Para ello vamos a suponer que sólo tenemos un único intervalo modal, es decir, un intervalo que en la representación gráfica del histograma alcanza la máxima altura. Podemos utilizar la función **MAX()** para ver cuál es la frecuencia mayor. La función **COINCIDIR()** nos dice que valor coincide con un número dado en un rango de celdas, luego mediante la expresión **=COINCIDIR(MAX(J2:J21);J2:J21;0)** nos dirá en que clase se alcanza el máximo.

Si nos fijamos en cómo calcular la moda cuando tenemos los datos agrupados:

$$M_d = a + c \frac{\delta_1}{\delta_1 + \delta_2}; \text{ con } c = b - a$$



Podemos utilizar la función `INDICE()` para determinar el valor que ocupa justo antes (y justo después) de la moda. De esta forma podemos calcular a , δ_1 y δ_2 . El parámetro a lo calculamos mediante `=INDICE(G2:G21;D13)`, el siguiente será `=INDICE(J2:J9;D13)-INDICE(J2:J9;D13-1)` y el último `=INDICE(J2:J21;D13)-INDICE(J2:J21;D13+1)`. Finalmente tendremos el valor de la moda.

Intervalo modal	4
a	40526,875
delta1	14
delta2	7
moda	47198,6

2.5 Cálculo de un percentil con los datos agrupados.

Vamos a calcular el percentil 50 que es exactamente la mediana. Primero observamos en que intervalo cae el valor que deja exactamente el 50% de las observaciones por debajo. Esto lo podemos hacer con la función `COINCIDIR()`, si como segundo parámetro colocamos 1, nos dará justo la posición del último valor que sea menor o igual que él, luego el primer valor mayor será el siguiente. En lugar de hacerlo con una proporción de datos podemos hacerlo directamente con la frecuencia relativa acumulada, si buscamos el percentil 50, equivale a una frecuencia relativa acumulada de 0,5. Y si buscamos el percentil x equivale a una frecuencia relativa acumulada de $x/100$. Ese es el valor que tenemos que buscar con la función `COINCIDIR()`. Nos quedaría lo siguiente:

Percentil	50,0
Intervalo del Perce	=COINCIDIR(D19/100;M2:M21;1)+1

Una vez que sabemos en qué clase cae, para calcular el percentil tendremos que sumar al extremo inferior la parte proporcional de la frecuencia que cae en dicha clase. Eso se consigue mediante la fórmula:

`=INDICE(G2:G21;D20)+D7*(D19/100-INDICE(M2:M21;D20-1))/INDICE(K2:K21;D20)`

Mediante `INDICE(G2:G21;D20)` obtenemos el extremo inferior del intervalo, `D7` es la amplitud de la clase (todas las clases tienen la misma amplitud), `D19/100` es la frecuencia relativa equivalente al percentil que buscamos, `INDICE(M2:M21;D20-1)` es la frecuencia relativa acumulada que ya tenemos de las clases anteriores e `INDICE(K2:K21;D20)` es la frecuencia relativa del intervalo en el que cae el percentil.

2.6 Observaciones finales.

Observe que todo es configurable, podemos cambiar manualmente el extremo inferior y superior de donde queremos que empiecen a contar las clases, el número de clases la amplitud de los intervalos, etc. Si añadimos más datos el programa hará todos los cálculos necesarios para sacarnos la tabla de frecuencias adecuada. Por ejemplo si copiamos los últimos 50 datos a partir del último tenemos 150 datos en el que todas las cuantas se hacen de forma automática. Se generan una nueva clase ya que en este caso $k = 9$.

Sin embargo hay cosas que con el Excel no conseguimos el acabado que deseáramos, como por ejemplo que los valores del histograma nos los pone al centro del rectángulo y no en los extremos como sería deseable.

3 Cálculo de medidas estadísticas de un conjunto de datos.

Vamos a utilizar el conjunto de datos del fichero **Cubiertas.xlsx** para calcular diferentes medidas estadísticas.

Ejercicio 2.

En el fichero Cubiertas.xlsx se han introducido los datos de talleres correspondientes a los miles de kilómetros, recorridos por vehículos con un modelo particular de neumático, antes de un pinchazo o un reventón.

Se pide:

Calcular la media, varianza, cuasi-varianza, desviación típica, cuasi-desviación típica, coeficiente de variación de Pearson y coeficiente de asimetría.

Con el Excel podemos calcular medidas estadísticas de todos los datos mediante las funciones estadísticas. Vamos a hacerlo desde **Formulas** → **Más Funciones** → **Estadísticas** vamos a seleccionar la función `PROMEDIO()` que calcula la media aritmética de un conjunto de datos. En este caso podemos poner lo siguiente: `=PROMEDIO(A:A)`

Para la Varianza existen varias funciones: `VAR.S()`(o `VAR()`) calcula la cuasi-varianza (también

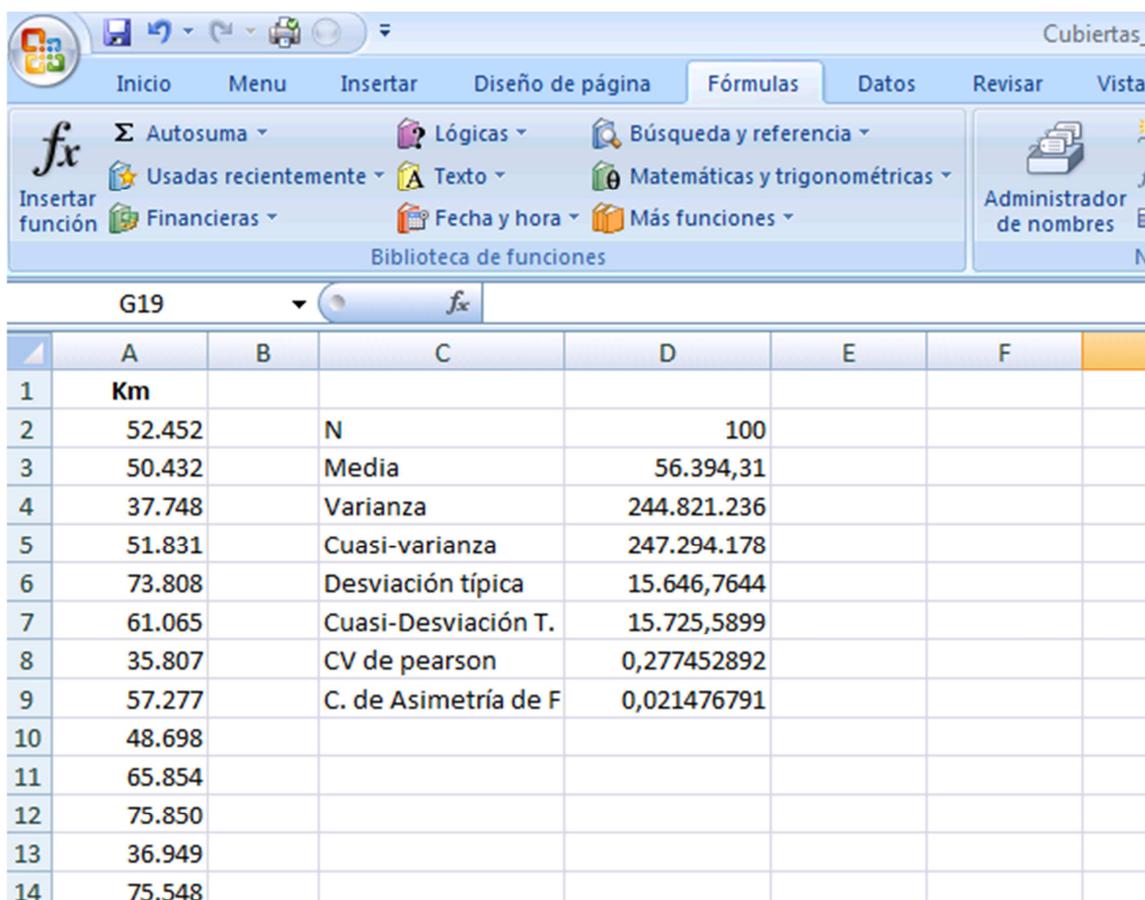
conocida como varianza muestral), $\text{VAR.P}()$ (o $\text{VARP}()$) calcula la varianza (también conocida como varianza poblacional) y también existen $\text{VARA}()$ y $\text{VARPA}()$ que tienen en cuenta valores de texto y lógicos (a nosotros no nos interesan).

De forma análoga existen las funciones $\text{DESVEST}()$, $\text{DESVESTP}()$, $\text{DESVESTA}()$ y $\text{DESVESTPA}()$ que son las raíces cuadradas de las anteriores. Es decir para calcular la cuasi-desviación típica y desviación típica.

El coeficiente de variación de Pearson no nos lo da el Excel de forma directa, pero sabemos que es la desviación típica entre la media.

Existe la función $\text{COEFICIENTE.ASIMETRIA}()$ para calcular al coeficiente de asimetría de Fisher. Aunque no es exactamente la hemos visto en teoría.

Así, obtenemos:



The screenshot shows the Excel interface with the 'Fórmulas' ribbon selected. The spreadsheet contains the following data:

	A	B	C	D	E	F
1	Km					
2	52.452	N		100		
3	50.432	Media		56.394,31		
4	37.748	Varianza		244.821.236		
5	51.831	Cuasi-varianza		247.294.178		
6	73.808	Desviación típica		15.646,7644		
7	61.065	Cuasi-Desviación T.		15.725,5899		
8	35.807	CV de pearson		0,277452892		
9	57.277	C. de Asimetría de F		0,021476791		
10	48.698					
11	65.854					
12	75.850					
13	36.949					
14	75.548					

4 Módulo de Análisis de datos de Excel.

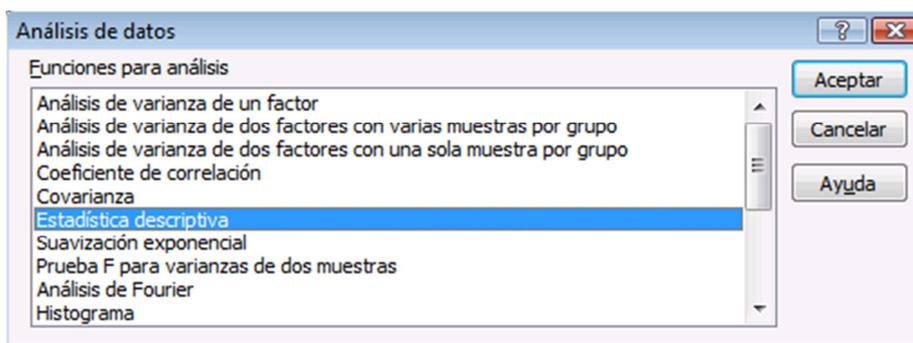
Excel también nos permite obtener distintas medidas estadísticas (y otros análisis estadísticos) mediante un complemento (que por defecto no está instalado) que es el [Módulo de Análisis de Datos](#). Para instalarlo tenemos que ir a [Archivo-> Opciones -> Complementos](#), seleccionar

Herramientas para análisis y pinchar en el botón [ir...](#). De esta forma se instalará un nuevo conjunto de iconos en la cinta de **Datos**

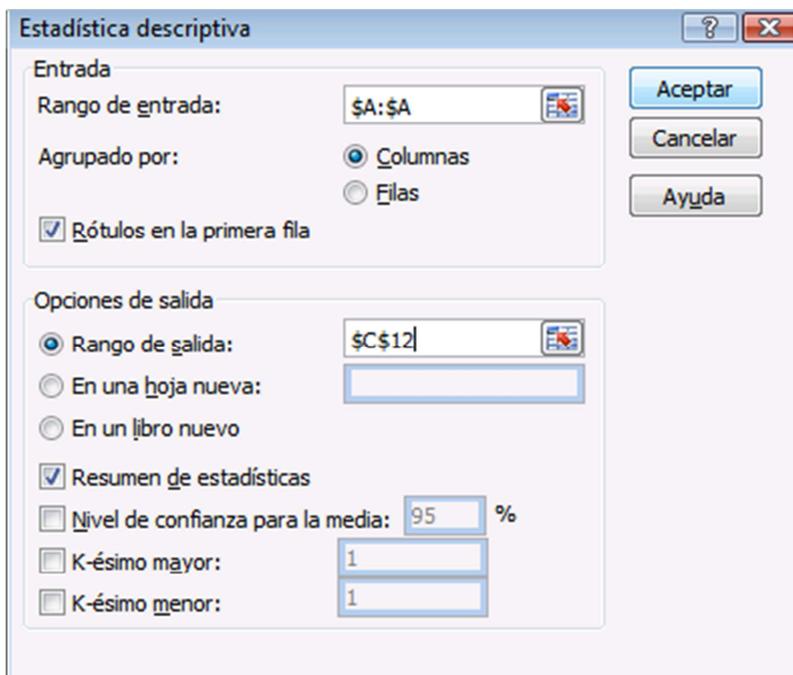


4.1 Estadística Descriptiva.

Desde ella podemos obtener las distintas medidas de **Estadística Descriptiva** en sólo unos pasos:



Nos sale el formulario:



Y después de rellenarlo como muestra en la figura obtenemos:

<i>Km</i>	
Media	56394,31
Error típico	1572,55899
Mediana	56072
Moda	#N/A
Desviación estándar	15725,5899
Varianza de la mues	247294177,8
Curtosis	-0,251824027
Coefficiente de asim	0,021476791
Rango	80061
Mínimo	10504
Máximo	90565
Suma	5639431
Cuenta	100

Hay que tener en cuenta que si hacemos una actualización en los datos, dicha actualización no se ve reflejada en estas medidas estadísticas a no ser que se vuelva a ejecutar:

Por ejemplo si cambiamos el primer dato al valor 40000 observamos que la media que calculamos mediante la fórmula sí cambia mientras que la que hicimos por este complemento no.

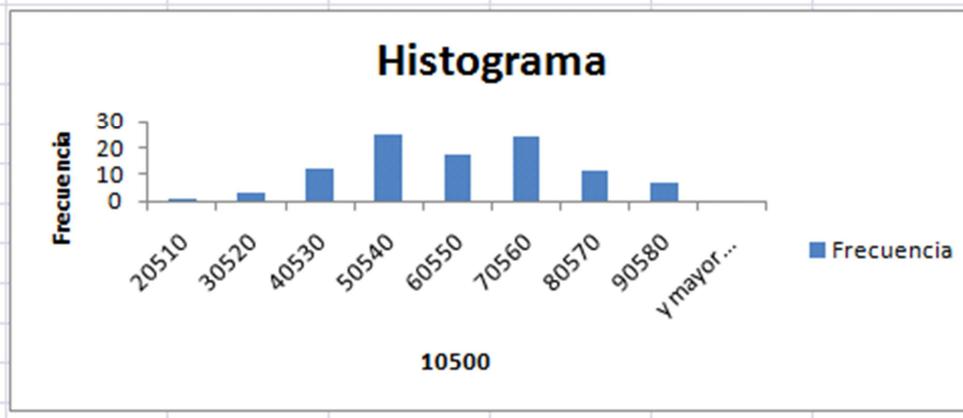
	A	B	C	D
1	Km			
2	40.000	N		100
3	50.432	Media		56.269,79
4	37.748	Varianza		247.338.047
5	51.831	Cuasi-varianza		249.836.411
6	73.808	Desviación típica		15.726,9847
7	61.065	Cuasi-Desviación T.		15.806,2143
8	35.807	CV de pearson		0,279492507
9	57.277	C. de Asimetría de F		0,033925335
10	48.698			
11	65.854			
12	75.850	<i>Km</i>		
13	36.949			
14	75.548	Media		56394,31
15	69.010	Error típico		1572,55899
16	61.477	Mediana		56072
17	65.585	Moda		#N/A
18	44.411	Desviación estándar		15725,5899
19	41.886	Varianza de la mues		247294177,8
20	34.754	Curtosis		-0,251824027
21	59.888	Coefficiente de asim		0,021476791
22	59.449	Rango		80061
23	67.632	Mínimo		10504
24	89.116	Máximo		90565
25	69.483	Suma		5639431
26	63.692	Cuenta		100

4.2 Histograma.

De forma análoga podemos ir a [Datos](#) → [Análisis de datos](#) → [Histograma](#) para crear una tabla de frecuencias de una forma más mecanizada (dándole previamente los extremos del intervalo) y no también nos permite crear un histograma.

Obteniendo los siguientes resultados:

Extremos		
10500	10500	Frecuencia
20510	20510	1
30520	30520	3
40530	40530	12
50540	50540	25
60550	60550	17
70560	70560	24
80570	80570	11
90580	90580	7
	y mayor...	0



Para que el histograma estuviera bien representado los rectángulos tienen que estar unidos. Pero eso ya sabemos cómo hacerlo.

Práctica 3.

Regresión lineal.

Objetivos.

En esta práctica continuaremos utilizando el programa Excel para realizar regresión lineal. Por ejemplo obtener medidas de variables bidimensionales como la covarianza y el coeficiente de correlación lineal y el gráfico de dispersión.

Índice.

Objetivos.....	1
Índice.....	1
1 Tablas de doble entrada.....	2
2 Diagrama de dispersión de una variable estadística bidimensional.....	2
3 Medidas estadísticas de una variable estadística bidimensional.....	5
4 Regresión lineal con el Módulo de Análisis de datos de Excel.....	6

1 Tablas de doble entrada.

Además de poder construirnos una tabla de doble entrada en Excel utilizando fórmulas hemos visto como a partir de un conjunto de datos realizar tablas dinámicas. Esta misma técnica la podemos utilizar para realizar una tabla de doble entrada.

En la pasada práctica, a partir del fichero [EjemploTablaDinamica.xlsx](#), construimos una tabla bidimensional de las variables **edad** y **sexo**.

Cuenta de Edad	Etiquetas de fila	Hombre	Mujer	Total general
18		8	17	25
19		9	6	15
20		1	4	5
21		2	2	4
22		1		1
Total general		21	29	50

2 Diagrama de dispersión de una variable estadística bidimensional.

Veamos cómo construir con el Excel un diagrama de dispersión de una variable bidimensional. Así podemos decir si encontramos algún tipo de regresión entre el par de variables.

Para ello resolvamos el siguiente ejercicio.

Ejercicio 1.

En una multinacional, dedicada a la fabricación de ordenadores, se están probando varios prototipos de arquitectura multinúcleo. Tras tomar los tiempos de respuesta (Y), en millonésimas de segundo, ante un código de prueba, el equipo de desarrollo sospecha que a medida que aumenta el número de procesadores (X) se reducen los tiempos de cómputo:

X	2	3	4	4	5	5	6	7	7	9	9	10	11	11	12
Y	11	12	10	13	11	9	10	7	12	8	7	3	6	5	5

Asesora al equipo de desarrollo confirmándoles que:

- la sospecha es cierta, y
- prediciendo el tiempo de cómputo para un prototipo con 13 núcleos.

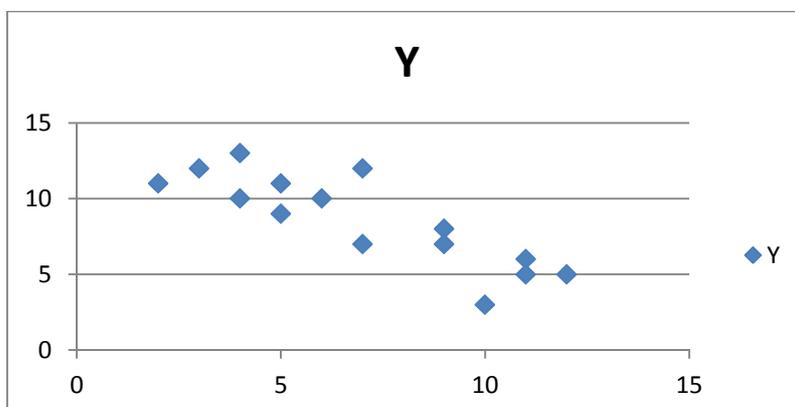
Vamos a introducir estos datos en Excel, lo podemos hacer tecleando cada dato, o copiando los datos del fichero y seleccionar pegado especial y marcar la casilla de transponer (para que

nos cambie filas por columnas y viceversa). Recuerda que siempre que estemos trabajando con un conjunto de datos estadístico conviene utilizar una única hoja para poner los datos y utilizar la primera fila para poner el nombre de las variables de forma que colocamos por filas cada una de las observaciones realizadas a cada elemento de la población. Además, podemos darle formato de tabla para tener una mejor visualización de los datos. Después de hacer esto nos debe quedar una cosa como la siguiente:

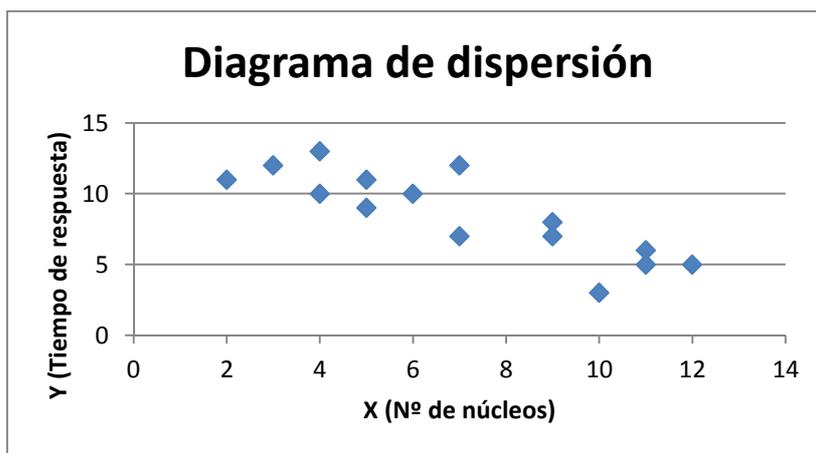
	A	B	C
1	X	Y	
2	2	11	
3	3	12	
4	4	10	
5	4	13	
6	5	11	
7	5	9	
8	6	10	
9	7	7	
10	7	12	
11	9	8	
12	9	7	
13	10	3	
14	11	6	
15	11	5	
16	12	5	
17			

Representemos estos datos mediante un diagrama de dispersión. Para ello vamos a seleccionar toda la tabla y desde **Insertar -> Dispersión** seleccionamos el primero de los gráficos.

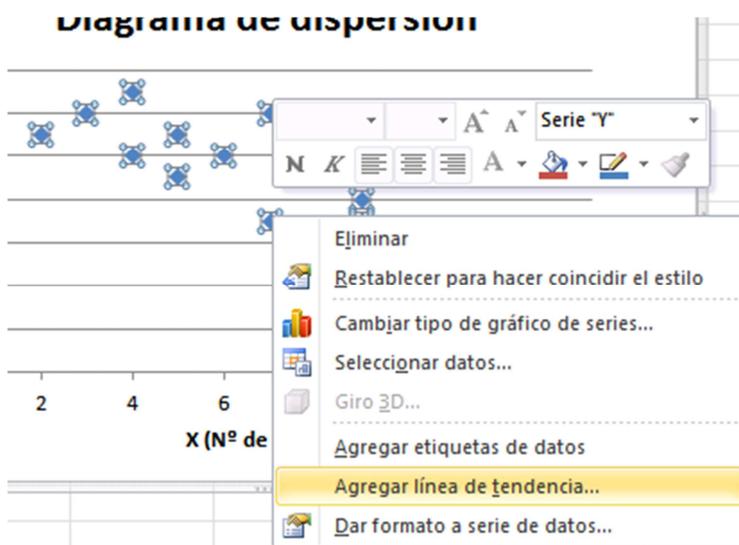
Si se fijan ya tenemos el gráfico de dispersión que queríamos, aunque lo correcto sería darle nombres más adecuados a los ejes.



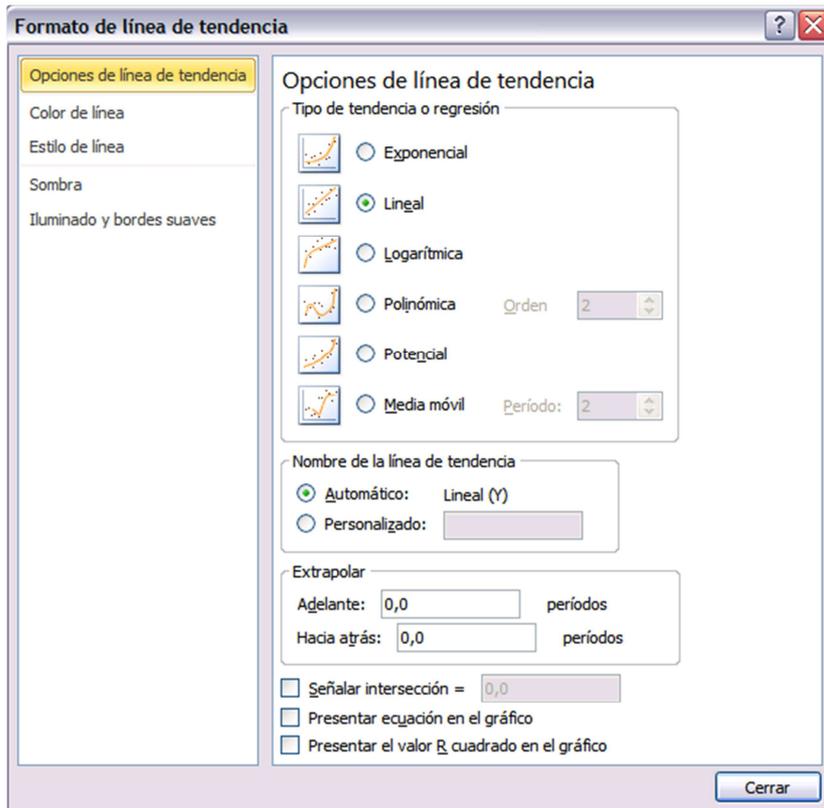
Podríamos cambiar el título y algunas otras características para que el gráfico quedara así:



También podemos agregar la recta de regresión (así como otros tipos de curvas de regresión). Para ello, nos situamos encima de uno de los datos, pinchamos en el botón derecho del ratón y seleccionamos [Agregar línea de tendencia](#):



Nos aparece el siguiente formulario:



Si dejamos seleccionada la **Lineal** nos muestra la recta de regresión lineal por mínimos cuadrados de Y sobre X de este conjunto de datos.

3 Medidas estadísticas de una variable estadística bidimensional.

Efectivamente según la gráfica parece que hay cierta regresión lineal inversa. Para basarnos en datos menos subjetivos que la gráfica podemos obtener el coeficiente de correlación lineal, como otras medidas estadísticas de una variable bidimensional.

Las funciones:

COEF.DE.CORREL(): Muestra el coeficiente de correlación lineal.

COEFICIENTE.R2(): Muestra el coeficiente de correlación lineal al cuadrado.

COVARIANCE.P(): Covarianza de las dos variables.

COVARIANZA.M(): Es un tipo especial de covarianza (que no han visto en este curso).

ERROR.TIPICO.XY(): Devuelve el error estándar del valor de "y" previsto para cada "x" de la regresión.

ESTIMACION.LINEAL(): Devuelve los parámetros de una tendencia lineal. Lo devuelve en formato de matriz. Así, si queremos ver todos los parámetros de la regresión lineal es

necesario, una vez introducida la fórmula en la primera celda, que seleccionemos un número de celdas para devolver los parámetros (en el caso de regresión lineal simple son dos), pulsemos la tecla **F2** y luego y pulsemos la combinación de teclas **Ctrl + Mayús + Enter**.

PRONOSTICO(): Devuelve el pronóstico y para un valor dado x a través de la recta de regresión.

Procedamos a calcular las siguientes medidas de regresión con los datos que tenemos. Obtenemos los siguientes resultados. También se han calculado la varianza de X e Y para comprobar que efectivamente el coeficiente de correlación r se obtiene también a partir de su fórmula. Una cosa parecida se ha hecho para calcular la pendiente de la recta α .

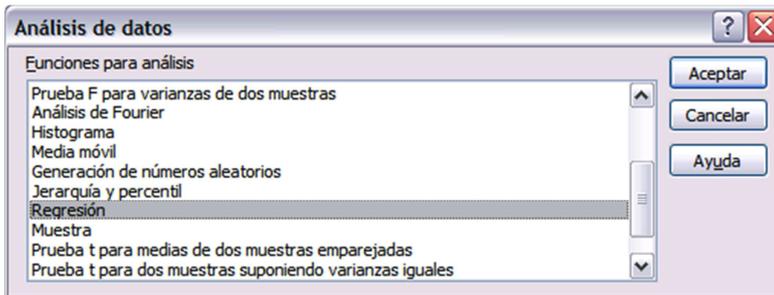
X	Y			
2	11		Coeficiente de correlación r	-0,84690626
3	12		R cuadrado	0,71725021
4	10		Covarianza	-7,6
4	13		Estimación Lineal α y β	-0,8028169 14,2197183
5	11			
5	9			
6	10		Pronóstico para x=	13 3,78309859
7	7			
7	12		Comprobación de las fórmulas del coeficiente de correlación y la pendiente de la recta:	
9	8		Varianza de V	9,46666667
9	7		Varianza de Y	8,50666667
10	3			
11	6		Coeficiente de correlación r	-0,84690626
11	5		Estimación Lineal α	-0,8028169
12	5			

4 Regresión lineal con el Módulo de Análisis de datos de Excel.

Excel también nos permite obtener distintas medidas estadísticas (y otros análisis estadísticos) mediante un complemento (que por defecto no está instalado) que es el **Módulo de Análisis de Datos**. Para instalarlo tenemos que ir a **Archivo-> Opciones -> Complementos**, seleccionar **Herramientas para análisis** y pinchar en el botón **ir...**. De esta forma se instalará un nuevo conjunto de iconos en la cinta de **Datos**



Desde **Análisis de datos** podemos obtener las distintas medidas de **Regresión lineal** en sólo unos pasos:



Nos sale el formulario:



Y después de rellenarlo obtenemos las siguientes medidas:

Resumen								
<i>Estadísticas de la regresión</i>								
Coefficiente de correlación múltiple	0,84690626							
Coefficiente de determinación R ²	0,71725021							
R ² ajustado	0,69550023							
Error típico	1,66592255							
Observaciones	15							
ANÁLISIS DE VARIANZA								
	<i>Grados de libertad</i>	<i>de cuadrado de los cua</i>	<i>F</i>	<i>Valor crítico de F</i>				
Regresión	1	91,5211268	91,5211268	32,9770456	6,79325E-05			
Residuos	13	36,0788732	2,77529794					
Total	14	127,6						
	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>nferior 95,0%</i>	<i>uperior 95,0%</i>
Intercepción	14,2197183	1,06896824	13,3022833	6,017E-09	11,91035283	16,5290838	11,9103528	16,5290838
Variable X 1	-0,8028169	0,13980111	-5,74256437	6,7933E-05	-1,104838848	-0,50079495	-1,10483885	-0,50079495

Algunas de las cuales ya las obtuvimos utilizando las fórmulas.

Hay que tener en cuenta que si hacemos una actualización en los datos, dicha actualización no

se ve reflejada en estas medidas estadísticas a no ser que se vuelva a ejecutar.

Práctica 4.

VARIABLES ALEATORIAS.

Objetivos.

En esta práctica utilizaremos las funciones que tienen el Excel para calcular probabilidades de variables aleatorias, calcular valores críticos de variables aleatorias y generar valores aleatorios.

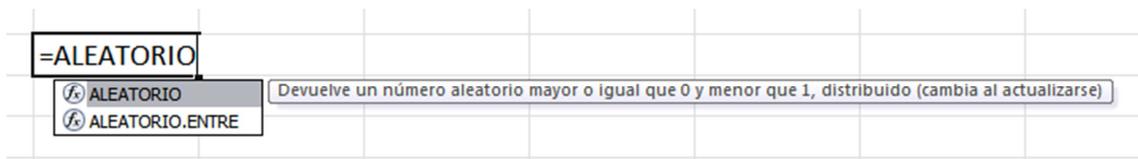
Índice.

Objetivos.....	1
Índice.....	1
1 Generación de valores aleatorios de una distribución uniforme.....	2
1.1 Valores aleatorios de una uniforme continua.....	2
1.2 Valores aleatorios de una uniforme discreta.....	3
2 Cálculo de probabilidades de variables aleatorias.....	4
2.1 Variables aleatorias discretas.....	4
2.2 Variables aleatorias continuas.....	7
2.3 Función de densidad de una variable aleatoria continua.....	9
3 Inversa de la función de distribución.....	11
3.1 Valores críticos.....	11
3.2 Generación de valores aleatorios de una distribución.....	12

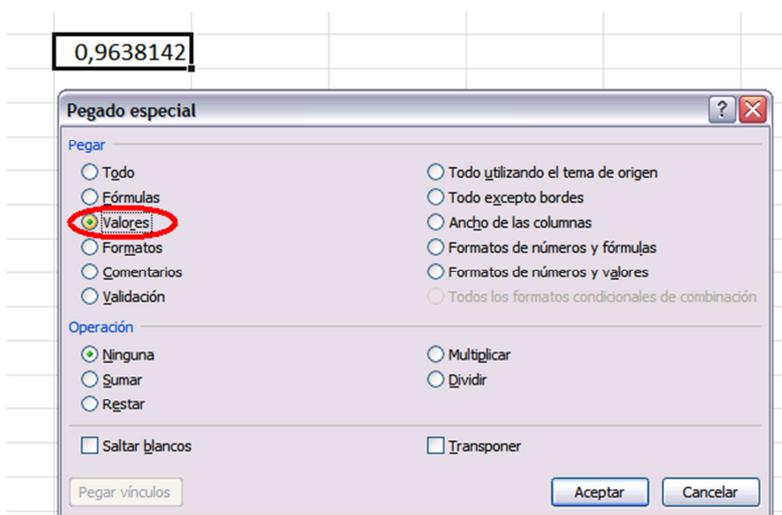
1 Generación de valores aleatorios de una distribución uniforme.

1.1 Valores aleatorios de una uniforme continua.

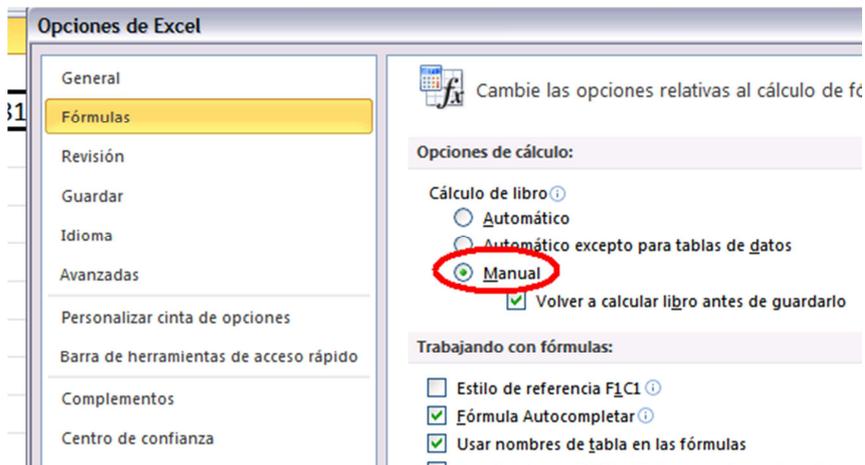
El Excel dispone de una función **ALEATORIO()** que nos genera un número aleatorio entre 0 y 1, es decir, nos genera valores de una distribución Uniforme Continua de parámetros 0 y 1 ($X \sim U(0,1)$). Veámoslo. A partir de un libro nuevo de Excel vamos a poner en una casilla esta función.



Observen que cada vez que actualizamos algo en el documento se actualiza este valor. Si queremos fijar este valor como definitivo podemos copiar la casilla sobre sí misma como valor.



Otra cosa que podemos hacer si queremos controlar cuando forzar el cambio es modificando las opciones de actualización desde **Archivo** → **Opciones** → **Fórmulas** y seleccionando la opción **Manual**.



Ahora, si queremos actualizarlo lo podemos hacer pulsando la tecla **F9** (para actualizar todos los libros abiertos) o mediante la **MAYÚSCULAS + F9** (actualizamos la hoja actual).

Establecer la actualización de las fórmulas manual también tiene mucho sentido cuando tenemos muchas fórmulas y datos ya que si no se relentecería mucho el trabajo con el programa.

Si queremos simular una Uniforme Continua en un intervalo general (a,b) lo podemos hacer mediante la expresión $a + (b-a)*\text{ALEATORIO}()$. Posteriormente veremos cómo podemos simular otras distribuciones continuas.

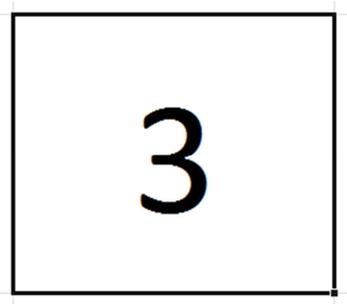
1.2 Valores aleatorios de una uniforme discreta.

Ejercicio 1.

Simula utilizando el Excel el lanzamiento de un dado.

Para simular una variable uniforme discreta podemos hacerlo redondeando (apropiadamente) lo descrito para una variable continua o utilizando directamente la función $\text{ALEATORIO.ENTRE}()$.

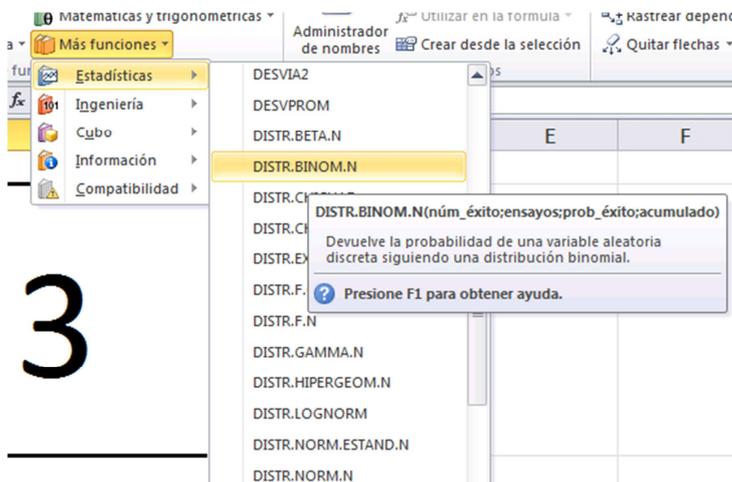
Mediante la expresión $\text{ALEATORIO.ENTRE}(1;6)$ obtenemos un número aleatorio (entero) entre 1 y 6. Podemos cambiar el formato del Excel para obtener una presentación más bonita.



2 Cálculo de probabilidades de variables aleatorias.

2.1 Variables aleatorias discretas.

Existen una enorme cantidad de funciones en Excel que nos permiten calcular probabilidades de variables aleatorias tanto discretas como continuas. Algunas de estas funciones para distribuciones discretas son: **DISTR.BINOM.N()**, **DISTR.HIPERGEOM.N()** y **POISSON.DIST()**.



En todas ellas hay que pasarle primero el valor que queremos comparar (es decir k), los parámetros de la distribución (por ejemplo en la binomial n y p) y si queremos la probabilidad acumulada o no (es decir si queremos la $P(X \leq k)$ ó $P(X = k)$).

Ejercicio 2.

Una caja con 12 artículos tiene 4 defectuosos. Si se toma una muestra de 3, en un caso con reemplazamiento y en otro sin reemplazamiento, ¿cuál será la probabilidad de no incluir artículos defectuosos en la muestra?

En el primer caso (la extracción se realiza con reemplazamiento) se puede calcular dicha probabilidad teniendo en cuenta que la variable $X =$ "Nº de artículos defectuosos en las tres extracciones (con reemplazamiento)" se distribuye según una binomial de parámetros $n = 3$ y $p = \frac{4}{12} = \frac{1}{3}$. Para esta distribución tenemos que calcular la $P(X = k)$ para $k = 0$.

Pero vamos a resolver el ejercicio de forma que si cambiamos n , p y/o k nos de cuál es el resultado de $P(X = k)$.

Dist. Binomial	
Parámetros	
n	3
p	0.3333
k	0
P(X=k)	0.2963

A la función `DISTR.BINOM.N(B6;B4;B5;FALSO)` le pasamos los siguientes 4 argumentos: el valor k para el que queremos calcular la probabilidad (en nuestro caso está en la celda B6), el número de ensayos n (que está en B4), la probabilidad de éxito en un ensayo p (B5) y en el último parámetro (acumulado) pondremos `FALSO` si queremos calcular $P(X = k)$ y `VERDADERO` si queremos calcular $P(X \leq k)$ (en este caso ponemos `FALSO`).

En el caso de que la elección de los 3 artículos se realiza sin reemplazamiento lo podemos resolver mediante una distribución Hipergeométrica donde la población total $N = 12$, los artículos que poseen cierta característica (son defectuosos) $N_1 = 4$ y el número de elementos que escogemos es $n = 3$. Al igual que antes tenemos que calcular $P(X = k)$ para $k = 0$.

Dist. Hipergeométrica	
Parámetros	
N	12
N1	4
n	3
k	0
P(X=k)	0.2545

Ejercicio 3.

Una centralita telefónica recibe 300 llamadas de media cada hora. No puede establecer más de 12 conexiones por minuto. Se pide:

- a) La probabilidad de que quede rebasada en un minuto dado.

En este caso se trata de una distribución de Poisson. El número medio de llamadas por minuto es $\lambda = \frac{300}{60} = 5$. Queremos calcular $P(X > k)$ para $k = 12$. Lo podemos calcular como $P(X > k) = 1 - P(X \leq k)$.

Es más podemos crearnos una hoja Excel que nos permita calcular cualquier tipo de

probabilidad con estas variables aleatorias discretas. De forma que al final tenemos algo así.

	A	B	C	D	E	F	G	H
1	Dist. Binomial			Dist. Hipergeométrica			Dist. Poisson	
2								
3	Parámetros			Parámetros			Parámetros	
4	n	3		N	12		lambda	5
5	p	0.3333		N1	4		k	12
6	k	0		n	3			
7				k	0			
8								
9	P(X=k)	0.2963		P(X=k)	0.2545		P(X=k)	0.0034
10								
11	P(X<=k)	0.2963		P(X<=k)	0.2545		P(X<=k)	0.9980
12								
13	P(X>k)	0.7037		P(X>k)	0.7455		P(X>k)	0.0020

Ejercicio 4.

Una caja contiene 100 bombillas, de las que 4 son defectuosas. Sea X el nº de bombillas defectuosas encontradas en una muestra de 9. Se pide:

- Hallar $P(X=2)$ y $P(X\leq 2)$.
- Aproximar las probabilidades anteriores por una binomial.
- Aproximar las probabilidades anteriores por una Poisson.

Vamos a realizar el problema todo lo parametrizado que podamos, es decir, de forma que podamos cambiar en Excel todos los valores. Podemos presentar los resultados de la siguiente forma:

	A	B	C
1	Parámetros		
2	Tamaño de la población N	100	
3	Unidades defectuosas Np	4	
4	Tamaño de la muestra n	9	
5	k	2	
6			
7		No acumulada	Acumulada
8		P(X=k)	P(X≤k)
9	Probabilidad de la Hipergeométrica		
10	Probabilidad de la Binomial		
11	Probabilidad de la Poisson		

Con fondo azul hemos colocado los parámetros, de esta forma utilizamos el criterio de que los valores que tienen fondo azul pueden ser cambiados mientras que el resto están formado por fórmulas (o texto). Veamos que fórmula debemos poner en cada una de las celdas de la tabla que falta por rellenar.

En la celda B9 iría la fórmula `DISTR.HIPERGEOM.N(B5;B4;B3;B2;FALSO)` y en la casilla C9 lo mismo pero con el último apartado a `VERDADERO`.

La distribución Hipergeométrica se puede aproximar por una distribución Binomial cuando N (el tamaño de la población) es grande con respecto a n (el tamaño de la muestra). En ese caso

una Hipergeométrica $H(N, n, p)$ la podemos aproximar por una Binomial, $B(n, p)$. Para el parámetro p de la distribución binomial debemos poner el cociente entre el número de unidades con esa determinada característica (defectuosas) entre el número total de unidades. Podemos poner esta operación intermedia en la casilla B6. El Luego en la casilla B10 colocaremos la fórmula `DISTR.BINOM.N(B5;B4;B6;FALSO)`. Para obtener la probabilidad acumulada (celda C10) basta con poner el último parámetro a `VERDADERO`.

Finalmente, la binomial se puede aproximar por la Poisson cuando n es grande y p muy pequeño, donde el parámetro $\lambda=n \cdot p$. Así en la B11 podemos poner `POISSON.DIST(B5;B4*B6;FALSO)` y para el valor acumulado `POISSON.DIST(B5;B4*B6;VERDADERO)`.

Al final de esto nos debe quedar algo similar a lo siguiente.

	A	B	C
1	Parámetros		
2	Tamaño de la población N	100	
3	Unidades defectuosas (n_A ó $N \cdot p$)	4	
4	Tamaño de la muestra n	9	
5	k	2	
6	p	0,04	
7		No acumulada	Acumulada
8		P(X=k)	P(X≤k)
9	Probabilidad de la Hipergeométrica	0,037595394	0,99801848
10	Probabilidad de la Binomial	0,043283375	0,99551762
11	Probabilidad de la Poisson	0,045209426	0,99404923

Si jugamos un poco con los parámetros veremos cuanto mejor son las aproximaciones.

Excel no saca directamente probabilidades de otras distribuciones discretas como la binomial negativa, aunque conociendo su función de probabilidad nos es fácil calcular probabilidades de esta y otras distribuciones discretas. Concretamente para ésta nos podemos ayudar de la función `COMBINAT()`.

2.2 Variables aleatorias continuas.

De forma análoga a las discretas podemos calcular probabilidades de las variables aleatorias continuas. Para ello podemos utilizar las siguientes funciones:

FUNCIÓN	DESCRIPCIÓN
<code>DISTR.BETA</code>	Devuelve la función de distribución beta acumulativa
<code>DISTR.CHICUAD</code>	Devuelve la probabilidad de una variable aleatoria continua siguiendo una distribución chi cuadrado
<code>DISTR.CHICUAD.CD</code>	Devuelve la probabilidad de una variable aleatoria continua siguiendo una distribución chi cuadrado considerando la Cola Derecha. Es decir $P(X \geq x)$
<code>DISTR.EXP.N</code>	Devuelve la distribución exponencial
<code>DISTR.F</code>	Devuelve la distribución de probabilidad F

DISTR.F.CD	Devuelve la distribución de probabilidad F considerando la Cola Derecha. Es decir $P(F \geq x)$
DISTR.GAMMA	Devuelve la distribución gamma
DISTR.NORM	Devuelve la distribución normal acumulativa
DISTR.NORM.ESTAND	Devuelve la distribución normal estándar acumulativa
DISTR.T	Devuelve los puntos porcentuales (probabilidad) de la distribución t de Student
DISTR.T.2C	Devuelve los puntos porcentuales (probabilidad) de la distribución t de Student Considerando las 2 Colas. Es decir $P(T \geq x)$
DISTR.T.CD	Devuelve la distribución de t de Student considerando la Cola Derecha. Es decir $P(T \geq x)$
DISTR.WEIBULL	Devuelve la distribución de Weibull

Ejercicio 5.

Utiliza Excel para obtener las tablas de probabilidad de la $N(0,1)$.

La función `DISTR.NORM.ESTAND()` devuelve las probabilidades de una distribución Normal Estándar (es decir $N(0,1)$). Por defecto Excel utiliza probabilidades hacia abajo (no hacia arriba como las tablas que tenemos) pero es muy fácil obtenerlas a partir de las anteriores (sería uno menos las probabilidades hacia abajo).

En primer lugar vamos a realizar los encabezados de la tabla. Quedaría algo así:

	A	B	C	D	E	F	G	H	I	J	K
1		Areas bajo la curva normal, $N(0; 1)$									
2	z_α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
3	0										
4	0,1										
5	0,2										
6	0,3										
7	0,4										
8	0,5										
9	0,6										
10	0,7										
11	0,8										
12	0,9										
13	1										
14	1,1										
15	1,2										
16	1,3										
17	1,4										
18	1,5										
19	1,6										
20	1,7										
21	1,8										
22	1,9										

A continuación colocaremos las fórmulas adecuadas en la casilla B3 (es decir cuando $z_\alpha = 0,00$) para poder copiarla al resto de la tabla. Para ello necesitamos obtener el valor de z_α sumando la cabecera de la fila y la columna. Si nos fijamos bien para las filas necesitamos dejar como referencia absoluta la columna A y para las columnas necesitamos dejar como referencia

absoluta la fila 2. Para obtener el valor z_α tenemos que hacer referencia a $\$A3+B\2 .

Finalmente falta poner adecuadamente la fórmula de la probabilidad total:

$1-DISTR.NORM.ESTAND.N(\$A3+B\$2;VERDADERO)$.

Y copiamos la fórmula en toda la tabla.

	A	B	C	D	E	F	G	H	I	J	K
1			Áreas bajo la curva normal, N(0; 1)								
2	z_α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
3	0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
4	0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
5	0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
6	0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
7	0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
8	0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
9	0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
10	0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
11	0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
12	0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
13	1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
14	1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
15	1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
16	1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
17	1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
18	1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
19	1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
20	1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
21	1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294

2.3 Función de densidad de una variable aleatoria continua.

En las variables aleatorias continuas si ponemos como falso el parámetro acumulado lo que nos devuelve es la función de densidad. Esto no nos vale para calcular probabilidades pero sí nos valdría para representar gráficamente la función de densidad.

Ejercicio 6.

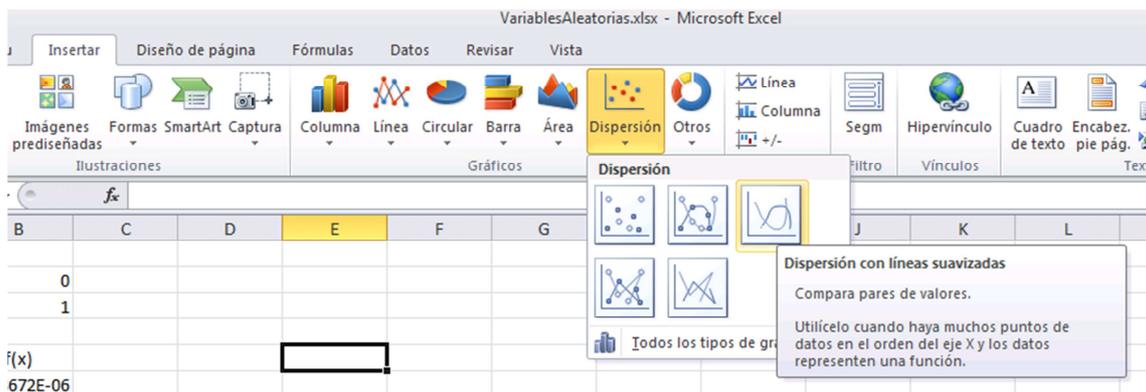
Representa gráficamente la función de densidad de una variable aleatoria $N(\mu, \sigma)$, donde μ y σ son dos parámetros modificables.

Para ello vamos a realizar una tabla con distintos valores de x y de $f(x)$ (la función de densidad).

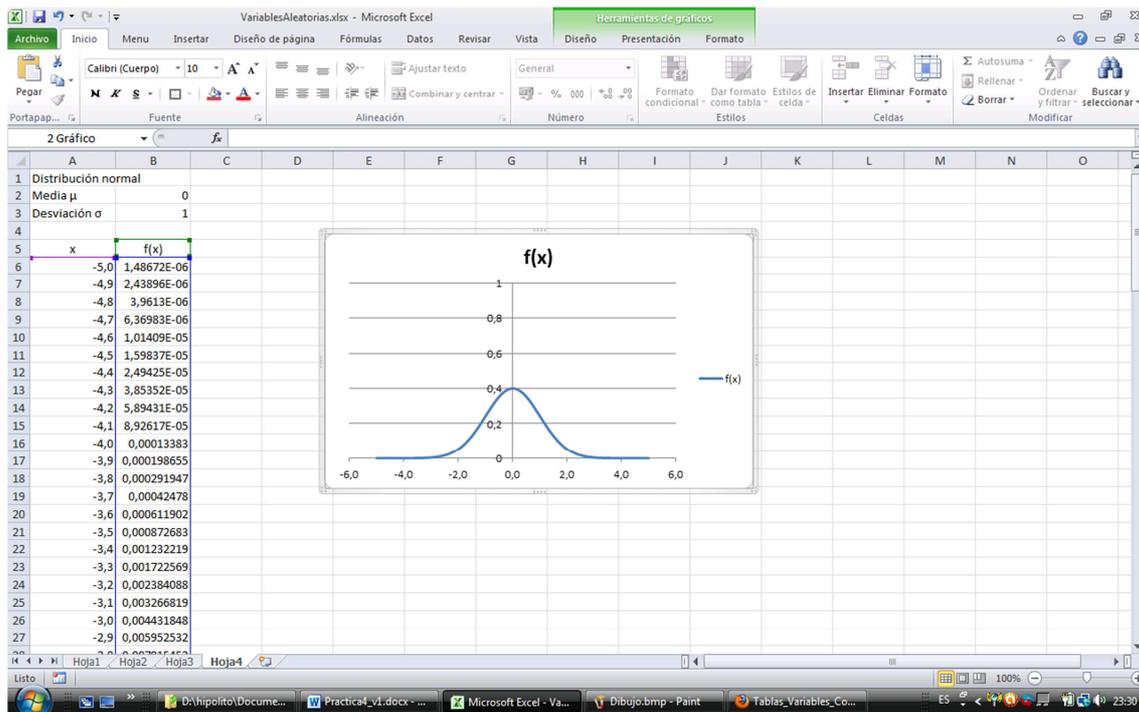
	A	B	C
1	Distribución normal		
2	Media μ		0
3	Desviación σ		1
4			
5	x	f(x)	
6	-5,0	1,48672E-06	
7	-4,9	2,43896E-06	
8	-4,8	3,9613E-06	
9	-4,7	6,36983E-06	
10	-4,6	1,01409E-05	
11	-4,5	1,59837E-05	
12	-4,4	2,49425E-05	
13	-4,3	3,85352E-05	
14	-4,2	5,89431E-05	
15	-4,1	8,92617E-05	
16	-4,0	0,00013383	

Donde $f(x)$ (en la celda B6) lo pondremos con la fórmula `DISTR.NORM.N(A6;B2;B3;FALSO)` (en el resto de la columna B basta con copiar la fórmula).

A continuación para representarla gráficamente vamos a utilizar el diagrama de dispersión donde vamos a unir cada dos puntos consecutivos de la tabla.



Configurando adecuadamente el gráfico llegamos a algo así:



Modifica los parámetros μ y σ para que veas que ocurre con la función de densidad. Por ejemplo prueba a representar una $N(1,1)$, $N(-2,1)$, $N(0,2)$ y $N(0,0'5)$.

3 Inversa de la función de distribución.

3.1 Valores críticos.

Excel también nos permite realizar la operación inversa al cálculo de probabilidades de variables aleatorias. Es decir, si dada una variable aleatoria X y un valor concreto x podemos calcular $\alpha = P(X \leq x)$, también podemos dada la variable aleatoria X y la probabilidad α calcular el valor x que verifica $\alpha = P(X \leq x)$. Para ello disponemos de las funciones inversas a la función de distribución.

FUNCIÓN	DESCRIPCIÓN
INV.BINOM	Devuelve el menor valor cuya distribución binomial acumulativa es menor o igual a un valor de criterio
INV.CHICUAD	Devuelve la función inversa de la probabilidad de una variable aleatoria continua siguiendo una distribución chi cuadrado de una sola cola
INV.CHICUAD.CD	Devuelve la función inversa de la probabilidad de una variable aleatoria continua siguiendo una distribución chi cuadrado de una sola cola
INV.F	Devuelve la función inversa de la distribución de probabilidad F
INV.F.CD	Devuelve la función inversa de la distribución de probabilidad F
INV.GAMMA	Devuelve la función inversa de la distribución gamma acumulativa
INV.NORM	Devuelve la función inversa de la distribución normal acumulativa
INV.NORM.ESTAND	Devuelve la función inversa de la distribución normal estándar acumulativa
INV.T	Devuelve el valor t de la distribución t de Student en función de la probabilidad y los grados de libertad
INV.T.2C	Devuelve la función inversa de la distribución de t de Student

Para las distribuciones discretas devuelve el menor valor cuya probabilidad acumulada es igual o mayor que α .

Ejercicio 7.

Calcula:

- El valor que deja un área por debajo de 0,27 de: $N(1,2)$, Chi-Cuadrado(7) y una t_7 .
- Calcula los valores críticos $z_{0,05}$, $z_{0,025}$, $z_{0,01}$, $t_{5;0,05}$, $\chi^2_{4;0,95}$.

Para el apartado a) basta con utilizar las funciones inversas de la función de distribución y para el apartado b) sería uno menos los valores que devuelven las funciones de distribución inversas.

Ejercicio 8.

Utiliza Excel para obtener la tabla de la Chi-Cuadrado.

3.2 Generación de valores aleatorios de una distribución.

Teniendo en cuenta como se utiliza la inversa de la función de distribución podemos utilizar para generar valores aleatorios de una distribución concreta. Por ejemplo vamos a simular 200 valores de una distribución normal de media 10 desviación 5. Bastaría con utilizar la fórmula $INV.NORM(ALEATORIO());20;5)$.

	A	B
1	N(20;5)	
2	23,66099766	
3	13,06315058	
4	16,37183363	
5	18,28286129	
6	20,32179779	
7	29,58794142	
8	31,54085565	
9	30,54490055	
10	12,30843297	
11	10,8717277	
12	18,25763282	
13	22,04335676	
14	26,25760126	

Para el caso de variables discretas también se procede exactamente igual.

Práctica 5.

Intervalos de Confianza y Contrastes de Hipótesis.

Objetivos.

Generaremos muestras aleatorias y a partir de ellas calcularemos los intervalos de confianza asociados. A continuación, identificamos y contamos cuántos intervalos contienen al verdadero valor del parámetro. Utilizaremos también la función lógica SI y los formatos condicionales.

Índice.

Objetivos.....	1
Índice.....	1
1 Intervalo de confianza para una normal.....	2
1.1 Generación de valores aleatorios de una distribución normal.....	2
1.2 Intervalo de confianza para una normal con varianza desconocida.....	2
2 Módulo de Análisis de datos de Excel.....	4
2.1 Intervalos de confianza y contrastes de hipótesis.....	5

1 Intervalo de confianza para una normal.

1.1 Generación de valores aleatorios de una distribución normal.

Vamos a generar 100 muestras de tamaño n (por ejemplo para $n = 5$) de una distribución normal de parámetros μ y σ de forma que cada muestra equivale a una fila. Ya hemos visto cómo hacerlo utilizando las funciones `ALEATORIO()` e `INV.NORM()`. Para que todo ello quede completamente parametrizado vamos a incluir en la parte alta de la hoja los valores de la media μ , la desviación típica σ , el nivel de significación α y el tamaño muestral n . Algo similar a lo que aparece a continuación:

	A	B
1	mu	0
2	sigma	1
3	alfa	0,05
4	n	5
5		

A partir de la fila 7 podemos generar una matriz de valores aleatorios de una $N(\mu, \sigma)$ mediante una fórmula como la siguiente `=INV.NORM(ALEATORIO();B1;B2)`. Algo similar a lo que se aprecia en el siguiente dibujo.

5					
6	Valores simulados de una Normal(mu,sigma)				
7	9,55445786	8,53839596	11,1342021	12,4354166	4,64207939
8	10,7780268	7,873833	11,3693556	14,2838898	11,9433994
9	7,07585688	11,7268879	13,1511158	12,017344	12,0913192
10	9,71252683	9,36621758	12,9815631	9,38865304	9,18253838
11	7,17917282	3,76968189	11,7499069	9,35721398	9,20502154
12	9,62383004	10,5989045	8,16466842	6,94075148	8,26617423
13	12,5302805	7,71128655	9,78420263	9,90686405	14,4094667
14	10,3318758	16,5102254	7,63433894	9,38386342	10,9708724
15	10,2640032	12,3178905	11,8911586	11,6538977	10,409984
16	11,9998811	11,7393067	10,8782874	8,1621653	7,38496343
17	9,99351671	6,51288855	8,22156357	10,4412342	10,2136216

Para evitar que estos valores se actualicen cada vez que hagas una modificación en el fichero puedes seleccionar una actualización de las fórmulas manual. Desde [Archivo](#) → [Opciones](#) → [Fórmulas](#) y seleccionando la opción [Manual](#).

1.2 Intervalo de confianza para una normal con varianza desconocida.

Nos interesa calcular un intervalo de confianza para la media μ (con varianza σ^2 desconocida)

$$\left(\bar{X} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} \right)$$

Vamos a calcular dicho valor crítico $t_{n-1, \alpha/2}$ utilizando la inversa de la distribución t de Student, concretamente mediante la fórmula `=INV.T(1-B3/2;B4-1)`.

	A	B	C	D	E	F	G
1	mu	10		t_n-1;alfa/2	=INV.T(1-B3/2;B4-1)		
2	sigma	2			INV.T(probabilidad; grados_de_libertad)		
3	alfa	0,05					
4	n	5					

En las columnas de la derecha podemos colocar las medias muestrales \bar{X} las desviaciones típicas muestrales S y los extremos inferiores y superiores de cada intervalo de confianza para la media poblacional. Nos debe quedar una cosa similar a esto.

6	Valores simulados de una Normal(mu,sigma)					media	S	L	U
7	13,5394904	10,0775426	7,60307859	13,1963791	8,73904454	10,631107	2,65021719	7,34042777	13,9217863
8	11,4005898	10,3227563	9,29171137	10,252496	8,77260419	10,0080315	1,01703348	8,74521768	11,2708454
9	10,7514692	9,92584861	5,37378718	11,4171612	11,9326072	9,88017469	2,6288579	6,61601648	13,1443329
10	9,17456228	10,5049338	12,4760509	14,0704184	10,760595	11,3973121	1,90075899	9,03720808	13,7574161
11	10,8445927	7,83286598	11,9107651	9,79658088	10,923564	10,2616737	1,55022691	8,33681278	12,1865347
12	10,1105761	9,60072682	13,9419838	6,49585594	6,50339084	9,33050669	3,08119606	5,50469646	13,1563169
13	6,64305768	10,1943269	10,4625053	10,2598061	11,8129198	9,87452315	1,92329087	7,48644212	12,2626042
14	9,4842747	10,4107468	11,1830751	11,3727413	4,15684439	9,32153646	2,98159666	5,61939523	13,0236777
15	9,43268646	8,02004691	5,05764265	8,59977015	11,0183915	8,42570753	2,19515011	5,70006867	11,1513464
16	8,28889639	11,1969565	8,26442332	10,3605557	11,172845	10,1567354	2,41296983	7,12338767	13,1900831

¿Contiene el verdadero valor? Vamos a comprobar si el verdadero valor del parámetro se encuentra o no en el intervalo calculado. Para ello haremos uso de la función lógica `SI()` y la función `Y()`.

`SI(prueba_lógica; valor_si_verdadero; valor_si_falso)`

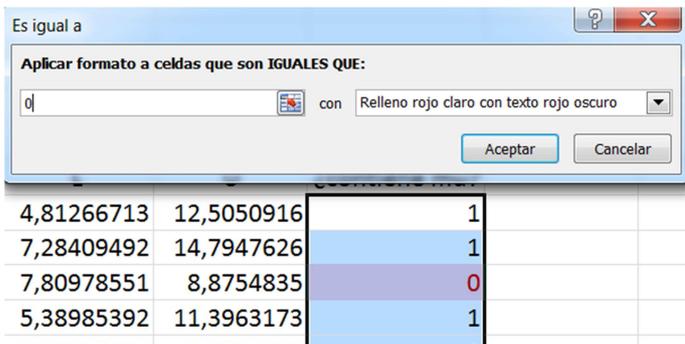
`Y(prueba_lógica1; prueba_lógica2;...)`

Por ejemplo, si las columnas **J** y **K** contienen los límites inferior y superior de un intervalo para la media, la siguiente función devolvería un 1 si μ está contenido en el intervalo asociado con la muestra de la fila 7, y un 0 en otro caso:

`=SI(Y(J7<mu;K7>mu); 1; 0)`

Con una función similar a la anterior comprueba para cada muestra y cada tipo de intervalo si el verdadero valor está contenido o no en el intervalo. Guarda el resultado en una columna al lado de la del límite superior de cada tipo de intervalo.

Podemos destacar cuándo el verdadero valor del parámetro no cae dentro del intervalo mediante **Inicio** → **Formato condicional** → **Resaltar reglas de celdas** → **Igual a ...** y ponemos cero y seleccionamos una regla que destaque dicha celda.



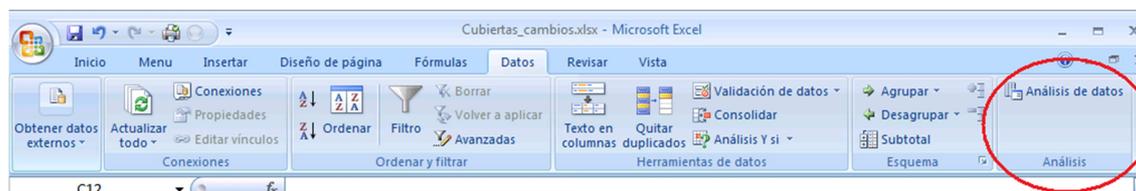
Cuenta el número de intervalos que contienen el parámetro. ¿Es algo similar a $100(1 - \alpha)$? Pulsa F9 para recalcularlo todo de nuevo y observa los diferentes valores que te dan.

Finalmente nuestra hoja de cálculo será similar a la siguiente.

	B	C	D	E	F	G	H	I	J	K	L	M
1	10		t _{n-1;alfa/2}	2,7764								
2	2											
3	0,05											
4	5										Total	
5												94
6	valores simulados de una Normal(mu,sigma)				media	S	L	U	¿contiene mu?			
7	8,83544447	7,77703625	11,9613863	10,7729621	8,65887934	3,09762723	4,81266713	12,5050916	1			
8	10,8260587	14,2153544	6,45178491	10,4219553	11,0394288	3,02443646	7,28409492	14,7947626	1			
9	8,64620562	7,7994988	7,98537307	8,49329235	8,3426345	0,42914105	7,80978551	8,8754835	0			
10	5,27767268	6,83661417	9,65541559	11,4882451	8,39308562	2,41871529	5,38985392	11,3963173	1			
11	13,993535	10,6466312	7,984684	11,5871676	11,3241073	2,23504559	8,54893167	14,099283	1			
12	7,32178859	11,7091238	5,64591054	7,15394058	8,77344574	2,90590277	5,16529089	12,3816006	1			
13	8,82372503	9,58028163	9,53501089	10,6947091	9,55540067	0,70785838	8,6764784	10,4343229	1			
14	6,98399248	13,3842384	9,43369243	10,4507737	10,0303927	2,29558274	7,18005025	12,8807352	1			
15	9,96370215	9,28471251	8,52433782	10,789698	9,36564877	1,0379813	8,07682475	10,6544728	1			
16	9,00532948	11,6235062	7,15007428	11,9510355	10,7339334	2,66373193	7,42647331	14,0413934	1			
17	12,1191912	10,5895456	9,06755425	8,1000562	10,4557691	1,8102160	8,10629261	12,7141426	1			

2 Módulo de Análisis de datos de Excel.

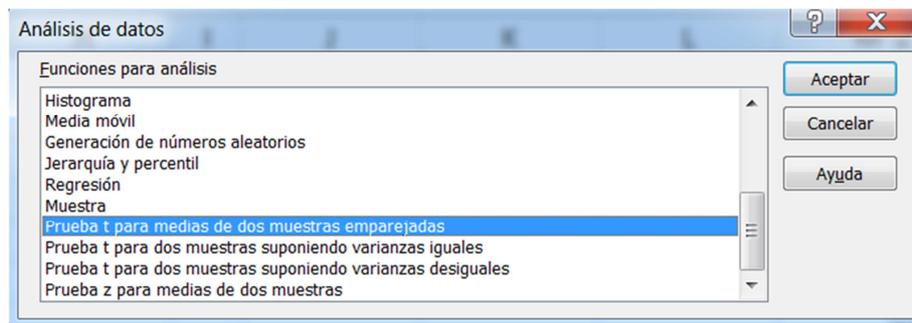
Excel también nos permite obtener contrastes de hipótesis (y otros análisis estadísticos) mediante un complemento (que por defecto no está instalado) que es el [Módulo de Análisis de Datos](#). Para instalarlo (si no lo hemos hecho ya) tenemos que ir a [Archivo-> Opciones -> Complementos](#), seleccionar Herramientas para análisis y pinchar en el botón [ir...](#) De esta forma se instalará un nuevo conjunto de iconos en la cinta de [Datos](#)



2.1 Intervalos de confianza y contrastes de hipótesis.

Desde este módulo también podemos obtener los siguientes intervalos de confianza y contrastes de hipótesis para:

- Diferencia media $\mu_D = \mu_X - \mu_Y$ de datos apareados (X_i, Y_i) $i = 1, \dots, n$. Se supone que $D = X - Y \sim N(\mu_D, \sigma_D)$ con σ_D desconocida.
- Para la diferencia de medias $\mu_1 - \mu_2$ de dos distribuciones normales independientes con varianzas desconocidas e iguales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$).
- Para la diferencia de medias $\mu_1 - \mu_2$ de dos distribuciones normales independientes con varianzas desconocidas y distintas ($\sigma_1^2 \neq \sigma_2^2$).
- Para el cociente de varianzas $\beta = \sigma_1^2 / \sigma_2^2$ de dos distribuciones normales independientes con medias μ_1 y μ_2 desconocidas.



Hay que tener mucho cuidado con esto, porque debemos saber siempre qué está haciendo Excel en cada prueba.

Práctica 6.

Programación Lineal.

Objetivos.

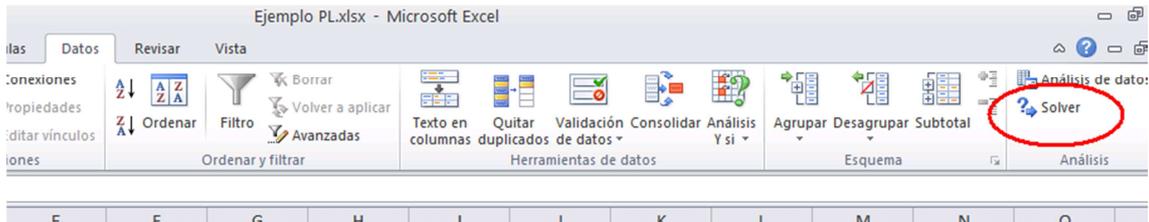
En esta práctica utilizaremos el complemento de Excel Solver para resolver problemas de Programación Lineal.

Índice.

Objetivos.....	1
Índice.....	1
1 Instalación del complemento Solver.....	2
2 ¿Cómo utilizar el complemento Solver de Excel?.....	2
3 Resolución de un problema práctico.....	5

1 Instalación del complemento Solver.

Este complemento se activa de forma similar al complemento de análisis de datos que en su momento instalamos para obtener medidas estadísticas y de regresión lineal. Procedemos de la siguiente manera. Tenemos que ir a **Archivo-> Opciones -> Complementos**, seleccionar **Herramientas para análisis** y pinchar en el botón **ir...** Y de la ventana que nos aparece seleccionamos **Solver**. De esta forma se instalará un nuevo icono en la cinta de **Datos**.



2 ¿Cómo utilizar el complemento Solver de Excel?

Veámoslo con un sencillo ejemplo:

$$\text{Máx } 10x_1 + 16x_2$$

s.a.:

$$2x_1 + 2x_2 \leq 8$$

$$1x_1 + 3x_2 \leq 6$$

$$x_1 \geq 0, x_2 \geq 0$$

Se colocan los parámetros a una planilla de cálculo.

	A	B	C	D
1	x1	x2	V(P)	
2				
3	10	16	0	
4				
5	2	2	0	8
6	1	3	0	6
7				
8				

Las celdas marcadas en amarillo corresponden a las **Celdas Cambiantes** o variables de decisión del modelo. La Celda **C3** corresponde al Valor de la Función Objetivo que está dada por: $A2 \cdot A3 + B2 \cdot B3$. Las Celdas **C5** y **C6** almacenan el valor o lado izquierdo de las restricciones 1 y 2, quedando definidas como $A2 \cdot A5 + B2 \cdot B5$ (o $\text{SUMAPRODUCTO}(A2:B2;A5:B5)$) y $A2 \cdot A6 + B2 \cdot B6$ (o $\text{SUMAPRODUCTO}(A2:B2;A6:B6)$), respectivamente.

Se inicia la aplicación **Solver** y se cargan los datos de la plantilla.

Parámetros de Solver

Establecer objetivo:

Para: Máx. Min Valor de:

Cambiando las celdas de variables:

Sujeto a las restricciones:

Convertir variables sin restricciones en no negativas

Método de resolución:

Método de resolución
 Seleccione el motor GRG Nonlinear para problemas de Solver no lineales suavizados. Seleccione el motor LP Simplex para problemas de Solver lineales, y seleccione el motor Evolutionary para problemas de Solver no suavizados.

En **Establecer objetivo** seleccionamos la celda **C3**. Marcamos como **Maximizar** la función objetivo y en **Cambiando las celdas de variables** colocamos las celdas de **A2**. Para establecer las restricciones pinchamos en **agregar** y vamos poniendo la celda donde está la fórmula de la primera restricción (es decir **C5**) y la celda con el lado derecho de la restricción (es decir **D5**) podemos cambiar el tipo de restricción (\leq , $=$ ó \geq).

Agregar restricción

Referencia de celda:

Restricción:

Y procedemos igual para todas las restricciones del problema.

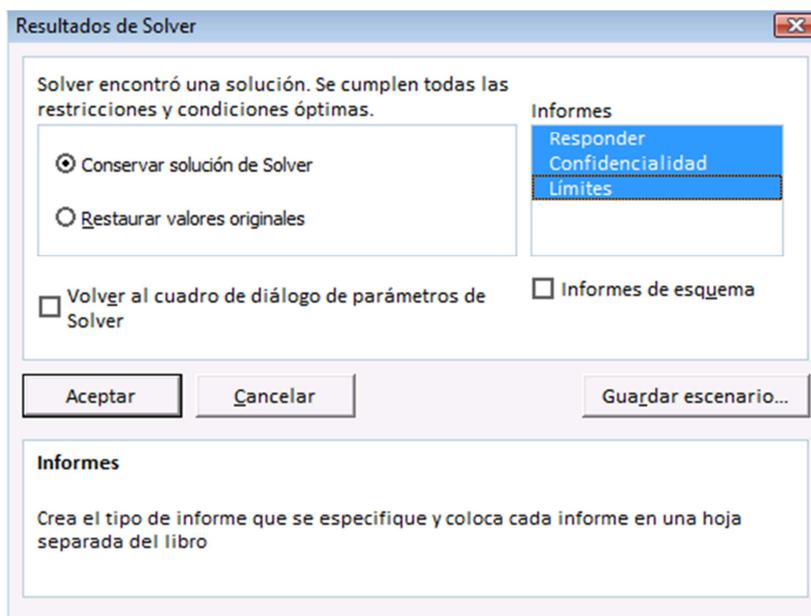
Finalmente la restricciones de no negatividad de las variables las podemos fijar seleccionado la opción de **Convertir variables sin restricciones en no negativas** y seleccionamos como Método de resolución el Simplex LP (ya que se trata de un problema de Programación Lineal).

Una vez colocados todos los parámetros pinchamos en [Resolver](#).

Si el modelo admite solución se obtienen los resultados.

	A	B	C	D
1	x1	x2	V(P)	
2	3	1		
3	10	16	46	
4				
5	2	2	8	8
6	1	3	6	6
7				

Se recomienda seleccionar los Informes ([Responder](#), [Confidencialidad](#) y [Límites](#)) que sugiere [Solver](#) para una mayor comprensión del modelo resuelto.



Estos informes son guardados en tres hojas diferentes donde nos dan detalles de la resolución del problema.



Los resultados son desplegados en las celdas cambiantes y se verifica el cumplimiento de las restricciones del problema. La Solución Óptima es $x_1=2$, $x_2=2$ con Valor Óptimo $V(P)=52$. Adicionalmente, ambas restricciones se encuentran activas, es decir, se cumplen en igualdad.

Al seleccionar los [Informes de Confidencialidad](#) ([Análisis de Sensibilidad](#)) se obtiene información relevante sobre el modelo propuesto.

	A	B	C	D	E	F	G	H
1	Microsoft Excel 14.0 Informe de confidencialidad							
2	Hoja de cálculo: [Ejemplo PL.xlsx]Hoja1							
3	Informe creado: 11/05/2011 22:17:46							
4								
5								
6	Celdas de variables							
7			Final	Reducido	Objetivo	Permisible	Permisible	
8	Celda	Nombre	Valor	Coste	Coefficiente	Aumentar	Reducir	
9	\$A\$2	x1	3	0	10	6	4,666666667	
10	\$B\$2	x2	1	0	16	14	6	
11								
12	Restricciones							
13			Final	Sombra	Restricción	Permisible	Permisible	
14	Celda	Nombre	Valor	Precio	Lado derecho	Aumentar	Reducir	
15	\$C\$5	V(P)	8	3,5	8	4	4	
16	\$C\$6	V(P)	6	3	6	6	2	
17								

Respecto a las **Celdas de variables** se incluye un intervalo de variación para los coeficientes en la función objetivo que mantiene la actual Solución Óptima. Por ejemplo el coeficiente que acompaña a x_1 en la función objetivo (actualmente igual a 10) puede variar en el siguiente intervalo garantizando la actual Solución Óptima: $[10 - 4'67, 10 + 6] = [5'33, 16]$. De la misma forma para el coeficiente que acompaña a x_2 en la función objetivo (actualmente igual a 16) puede variar en el intervalo $[10, 20]$.

En cuanto a las **Restricciones**, el precio sombra de la restricción 1 es 3'5, el cual es válido siempre y cuando la variación en el lado derecho se encuentre en el intervalo $[8 - 4, 8 + 4] = [6, 12]$. De la misma forma, el precio sombra para la restricción 2 es 3, válido en el intervalo de variación del lado derecho entre $[4, 12]$.

3 Resolución de un problema práctico.

Ejercicio 1.

Se desea obtener la mezcla de petróleo a partir de crudos de distinta procedencia, cada uno de los cuales tienen distintas características. En la tabla adjunta se detallan los distintos crudos (cuatro en total) y sus características más importantes: el tanto por ciento de azufre, la densidad y el precio por Tm.

Origen	% de Azufre	Densidad	Precio (€)
Kuwait	0,45	0,91	350
Arabia	0,40	0,95	310
Noruega	0,38	0,89	390
Venezuela	0,41	0,92	340

Se exige a la mezcla que tenga unas características concretas, que se traducen en un porcentaje del 0.40 % de contenido de azufre y una densidad igual a 0.91. Se desea que el precio de la mezcla sea mínimo.

Los elementos fundamentales de este problema, y que caracterizan cualquier problema de programación matemática, son los siguientes:

- **Variable de decisión:**

Permite caracterizar matemáticamente la decisión a adoptar. En general es un vector de \mathbb{R}^n , En el ejemplo debe identificar la mezcla y puede ser el porcentaje o la proporción de cada uno de los crudos en la mezcla. Supondremos que es la proporción (x_i). Se introducen las variables (x_1, x_2, x_3, x_4) $\in \mathbb{R}^4$ que representan la proporción con que intervendrán en la mezcla los crudos procedentes de Kuwait, Arabia, Noruega y Venezuela respectivamente.

- **Conjunto de restricciones:**

La variable de decisión debe verificar una serie de restricciones de forma que una decisión válida debe pertenecer a un subconjunto de \mathbb{R}^n . En los problemas de programación lineal las restricciones se identificarán por un conjunto de ecuaciones lineales en la variable de decisión y se supondrá que todas las componentes de la variable de decisión son mayores o iguales que 0. En el ejemplo hay tres restricciones, aparte de la restricción natural de que cualquier proporción no debe ser negativa $x_i \geq 0$.

- La suma de proporciones debe ser igual a 1.

$$x_1 + x_2 + x_3 + x_4 = 1$$

- El contenido (en %) de Azufre de la mezcla debe ser 0.4.

$$0.45 x_1 + 0.40 x_2 + 0.38 x_3 + 0.41 x_4 = 0.40$$

- La densidad de la mezcla debe ser 0.91

$$0.91 x_1 + 0.95 x_2 + 0.89 x_3 + 0.92 x_4 = 0.91$$

- **Función objetivo:**

Existe una función que evalúa todas las decisiones válidas y el problema es elegir aquella decisión que minimiza dicha función. En los problemas de P.L. esta función también depende linealmente de la variable de decisión.

En el ejemplo, cada mezcla válida tiene asociado un coste de forma que la función objetivo es el precio de cada Tm. de mezcla:

$$350 x_1 + 310 x_2 + 390 x_3 + 340 x_4$$

Una vez planteado adecuadamente el problema colocamos los parámetros en el Excel:

	A	B	C	D	E	F	G	H
1		x1	x2	x3	x4			
2		Kuwait	Arabia	Noruega	Venezuela			
3						tipo de restr	Lado derecho	valores
4	F. Objetivo	350	310	390	340			0
5								
6	Restricción 1	1	1	1	1	=	1	0
7	Restricción 2	0,45	0,40	0,38	0,41	=	0,40	0
8	Restricción 3	0,91	0,95	0,89	0,92	=	0,91	0

Le indicamos al Solver dónde está cada cosa:

Parámetros de Solver

Establecer objetivo:

Para: Máx. Mín Valor de:

Cambiando las celdas de variables:

Sujeto a las restricciones:

\$H\$6 = \$G\$6
 \$H\$7 = \$G\$7
 \$H\$8 = \$G\$8

Convertir variables sin restricciones en no negativas

Método de resolución:

Método de resolución
 Seleccione el motor GRG Nonlinear para problemas de Solver no lineales suavizados. Seleccione el motor LP Simplex para problemas de Solver lineales, y seleccione el motor Evolutionary para problemas de Solver no suavizados.

Ayuda Resolver Cerrar

Y lo resolvemos:

	A	B	C	D	E	F	G	H
1		x1	x2	x3	x4			
2		Kuwait	Arabia	Noruega	Venezuela			
3		1,85962E-15	0	0,333333333	0,666666667	tipo de restric.	Lado derecho	valores
4	F. Objetivo	350	310	390	340			356,666667
5								
6	Restricción 1	1	1	1	1	=	1	1
7	Restricción 2	0,45	0,40	0,38	0,41	=	0,40	0,4
8	Restricción 3	0,91	0,95	0,89	0,92	=	0,91	0,91

La mezcla óptima verificando las restricciones anteriores se obtiene mezclando los crudos procedentes de Noruega y Venezuela en las proporciones 1/3 y 2/3 respectivamente y el precio de esta mezcla es de 356'67 €/Tm. La variable de decisión toma los valores $(x_1, x_2, x_3, x_4) = (0, 0, 1/3, 2/3)$.