



GOBIERNO
DE ESPAÑA

MINISTERIO
DE EDUCACIÓN

SECRETARÍA DE ESTADO
DE EDUCACIÓN Y
FORMACIÓN PROFESIONAL
DIRECCIÓN GENERAL DE
FORMACIÓN PROFESIONAL

INSTITUTO DE
TECNOLOGÍAS EDUCATIVAS

HOJA DE CÁLCULO EN LA ENSEÑANZA DE LAS CIENCIAS

SESIÓN 6: TRATAMIENTO DE
DATOS ESTADÍSTICOS

TRATAMIENTO DE DATOS ESTADÍSTICOS

CONTENIDOS

INTRODUCCIÓN

Las cuestiones estadísticas tienen un uso muy natural en las Hojas de Cálculo, pues su instrumento principal son las tablas y gráficos, que también son esenciales en Estadística.

Por otra parte, la abundancia de cálculos repetitivos hace que cada vez más se impartan las clases de esta materia con la ayuda de calculadoras gráficas o del ordenador, y en concreto de una Hoja de Cálculo, pues los paquetes estadísticos, o son muy caros, o tienen una gestión de entrada de datos muy complicada para los alumnos.

CÁLCULOS ESTADÍSTICOS

Abre el modelo `CUANTITA.ODS`, que es un ejemplo sencillo del uso de técnicas informáticas en los cálculos estadísticos. Si te pide *Habilitar macros*, responde afirmativamente.

Observa que en la primera hoja se incluyen tres niveles de trabajo:

Recogida de datos cuantitativos en una zona preparada para ello:

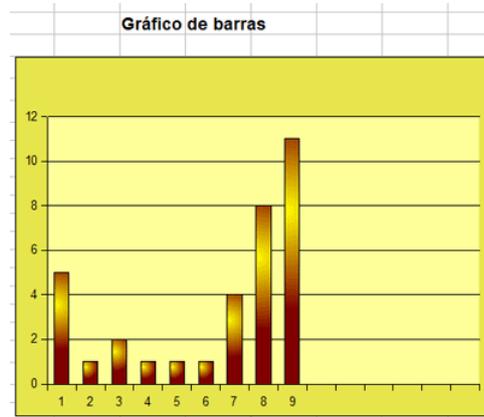
Datos cuantitativos aislados					
Puedes escribir un máximo de 60 datos numéricos, elegidos entre una gama de hasta 15 valores, dentro del rectángulo de color, en cualquier orden.					
Para borrar datos usa este botón:					
					Borrar datos
7		6			
7		5			
8	1	4			
7	3	3			
8	2	8			
9		8			
7	8	8		9	1
8	8		9	9	1
9	9	1	9	9	1
	9	9	9		

Obtención de frecuencias a partir de esos datos:

Escribe los valores cuyas frecuencias deseas encontrar en la columna Valores y en el resto de la columna escribe un espacio en blanco o borra

Valores	Frec. absolutas	Frec. Relativas
1	5	0,15
2	1	0,03
3	2	0,06
4	1	0,03
5	1	0,03
6	1	0,03
7	4	0,12
8	8	0,24
9	11	0,32
Total	34	1

Creación de un gráfico de barras sobre las frecuencias:



Con esto hemos querido representar las tres primeras operaciones estadísticas por orden de ejecución: recogida de datos, tabulación y presentación gráfica.

En la segunda hoja figuran todas las medidas más usuales a nivel elemental:

- De tendencia central: media, mediana y moda.
- De orden: cuarteles.
- De dispersión: varianza, desviación típica, etc.
- De asimetría y curtosis.

Estadísticos			
De orden		Paramétricos	
Cuartil 1	4,25	Media	6,44
Mediana	8	Varianza	8,6
Cuartil 3	9	Desviación tip.	2,93
Máximo	9	Coef. Variación	45,53%
Mínimo	1	Asimetría	-0,95
Rango	8	Curtosis	-0,69
Semiintervalo			
Intercuartílico	2,38		
Moda	9		

Reservamos la tercera para que practiques con cálculos estadísticos y compruebes algunos de los que ya figuran en la Hoja segunda. No se pretende que analices todo el contenido del modelo, sino mostrarte las posibilidades de

la Hoja de Cálculo en estos temas. Si no recuerdas qué era la varianza o la asimetría, no las tengas en cuenta.

RECOGIDA DE DATOS

Con este modelo aprenderás muchas funciones que te pueden interesar. Sigue la explicación teniendo abierto el archivo *cuantita.ods* mientras la lees.

Comenzamos investigando la columna de frecuencias absolutas. Hemos usado en la tabla de frecuencias algunas técnicas que conviene aprender para poder usar y diseñar modelos estadísticos.

Funciones lógicas y de información

Lee con atención las fórmulas contenidas en la columna de frecuencias, una de ellas es:

=SI (ESNÚMERO(B28);CONTAR.SI (\$B\$7:\$G\$16;B28);" ")

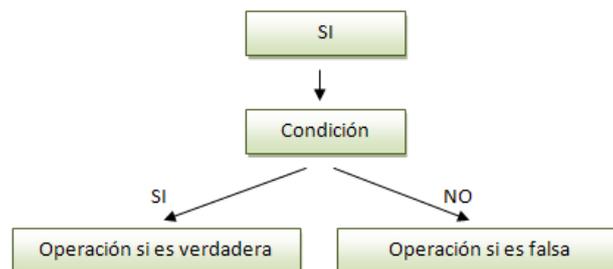
que requiere una explicación progresiva y pausada en los siguientes párrafos.

Función SI

Es la función lógica de tipo condicional, cuyo formato es

=SI(CONDICIÓN;OPERACIÓN SI ES VERDADERA;OPERACIÓN SI ES FALSA)

=SI(Condición; Operación si es verdadera; Operación si es falsa)



Se escribe el signo = para indicar que se ingresa una fórmula. Escribes SI y abres paréntesis. En él expresas una condición (D13=F13, A24>0, LOG(A23)<1, etc.) y a continuación del primer signo ";" . Luego escribes qué operación deseamos efectuar si **se cumple** esa condición, y detrás del segundo ";" lo que se debe calcular **si es falsa**.

Por ejemplo, la fórmula **SI (D12<100 ; D12*2 ; D12/2)** devuelve el doble del contenido de la celda D12 si éste es menor que 100 y en caso contrario devuelve la mitad de ese número. Estúdialo bien.

En este caso de las frecuencias, si la celda de la izquierda **es un número**, calculamos la frecuencia, y si no, escribimos un espacio en blanco " ", que es la última operación que se puede leer en la fórmula.

Consulta en el OpenOffice.org la lista de **FUNCIONES LÓGICAS**, algunas de las cuales como **Y O NO**, iremos usando en este curso. Usa para ello el botón **fx**:



Función ESNÚMERO

Es un ejemplo de **FUNCIÓN DE INFORMACIÓN**, que nos indica si la celda de valores B28 contiene un número, o bien le hemos escrito un espacio en blanco o bien hemos borrado el contenido. De esta forma sólo se calcula la frecuencia si esa celda contiene información numérica. Hemos escrito **ESNÚMERO(B28)** que se puede interpretar como **"si la celda B28 contiene un número"**.

Consulta en la lista de funciones alguna más de Información: **ESBLANCO**, **ESTEXTO**, **TIPO**, etc.

Por tanto, en la fórmula que estamos estudiando, ya tendríamos explicado la función y la condición:

=SI (ESNÚMERO(B28);CONTAR.SI (\$B\$7:\$G\$16;B28);" ")

"Si la celda B28 contiene un número..."

Función CONTAR.SI

La función **CONTAR.SI** es una variante de **CONTAR** que es muy usada en Estadística. Observa cómo se ha usado en la fórmula que estamos analizando:

CONTAR.SI (\$B\$7:\$G\$16;B28)

Esta estructura se explica por sí misma. Se cuentan todos los números pertenecientes al rango **\$B\$7:\$G\$16** (que es el rectángulo amarillo de recogida de datos) coincidentes con el que figura en la celda B28, es decir, en términos estadísticos, su **frecuencia**.

Estas funciones son algo difíciles de entender y manejar al principio, pero constituyen instrumentos imprescindibles cuando se desea construir modelos de cierta complejidad.

Si recorres la columna de frecuencias observarás que esta fórmula, con los cambios de referencia oportunos, figura en todas las celdas.

Por tanto, ya tenemos, además de la condición, la operación que hay que efectuar **si es verdadera**.

=SI (ESNÚMERO(B28);CONTAR.SI (\$B\$7:\$G\$16;B28);" ")

Por último, la operación segunda, que es la que se efectúa si la operación es falsa, en nuestra fórmula sería simplemente " "), es decir, que se escribe un espacio en blanco. Con eso tenemos la fórmula completa. Repasa todo si no lo has entendido, que la función SI es muy importante.

Función CONTAR

Observa la columna de **frecuencias relativas**. Recuerda que las frecuencia relativa de un dato se calcula dividiendo la absoluta entre el total de frecuencias. Para ello el modelo usa la fórmula

=SI (ESNÚMERO(B28);D28/CONTAR(\$B\$7:\$G\$16);" ")

Como la anterior, comienza con la pregunta de si B28 es un número. En caso afirmativo divide su frecuencia D8 entre el total de frecuencias de la zona de datos. Para eso cuenta los números que hay en ella, con la función **CONTAR** que actúa sobre el rango \$B\$11:\$G\$20, como casi todas las de esta hoja.

La función CONTAR es utilísima en todos los trabajos de Ciencias con las hojas de cálculo. Procura aprender bien su estructura, que es muy sencilla.

La construcción del gráfico no requiere explicación, pues ya se estudió en la sesión 3.

Función SUMA

El total de las frecuencias se ha calculado con la función **SUMA**, cuya sintaxis es:

SUMA (Conjunto de celdas que se desea sumar)

Dentro del paréntesis de la función SUMA se puede escribir:

- una referencia de rango =SUMA(F24:F38)
- de varios rangos: =SUMA(B5:B7;D5:D7)
- bien celdas aisladas: =SUMA(B5;D7)

Recuerda que también con el botón de autosuma



se consigue el mismo efecto, pero controlas menos los rangos o números que se suman.

FUNCIONES ESTADÍSTICAS

Dentro del archivo `CUANTITA.ODS`, en la Hoja 2 *Cálculos* figuran las medidas o estadísticos básicos de una distribución cuantitativa discreta, es decir aquella que está construida sobre datos numéricos que no están agrupados.

Nos detendremos tan sólo en las básicas.

Media

En la Hoja de Cálculo la **media** se halla con la función **PROMEDIO** que puede actuar de varias formas:

- Sobre números aislados: **=PROMEDIO(3;4;5;7) = 4,75**
- Sobre celdas aisladas: **=PROMEDIO(C7;C8;C9)**
- Sobre un rango de celdas:
=PROMEDIO(Recogida.\$B\$11:Recogida.\$G\$20)

Observa el último caso, que es el se usa en la hoja que estamos observando. Es el mismo rango desde B11 hasta G20 que hemos visto repetido en muchas fórmulas, pero en este caso precedido de la palabra *Recogida* y de un punto. Esto es así porque nos referimos a unas celdas que están **en una hoja distinta de la actual**, que es la hoja *Cálculos*.

Para referirnos a una celda situada en otra hoja debemos escribir su referencia precedida del nombre de dicha hoja y un punto separador.

Si seleccionas la celda con el ratón, el mismo programa escribe el nombre de la hoja por ti.

Cuartiles

Los cuartiles nos informan de los puntos en los que se encuentran los límites del 25% (primer cuartil) de los datos de menor valor, el 50% (segundo o *mediana*) y el 75% (tercer cuartil) del total de la distribución de datos. En **OpenOffice.org Calc** se usa la función **CUARTIL**, incluyendo como parámetro el número de cuartil deseado. Si te interesa este instrumento estadístico, consulta las celdas correspondientes en el modelo para ver cómo se sitúa el número de cuartil detrás del rango de referencia.

Rango

Seguimos con la Hoja 2 *Cálculos* del modelo *cuantita.ods*. Estudia en ella las fórmulas de *Máximo*, *Mínimo* y *Rango*. No requieren explicación adicional. Consulta **FUNCIONES ESTADÍSTICAS** y la ayuda de **OpenOffice.org**. Recuerda que sólo se trata de tener una visión global de los cálculos estadísticos.

Desviación típica

Para este cálculo se usa la función **DESVESTP** ("Desviación típica de la población", en la que la suma de desviaciones al cuadrado se divide entre **n**), para distinguirla de la *muestral* (en la que dividimos entre **n-1**), que se obtendría con la función **DESVEST**. Tiene el mismo formato que el **PROMEDIO** con las mismas variantes. Recuerda que esta es una medida de la dispersión o variabilidad de los datos.

Reproducción de los cálculos

Como práctica sobre el modelo intenta reproducir, por otros procedimientos, los cálculos de la *media* y la *desviación típica* en la Hoja 3. No dejes de hacerlo porque este aprendizaje es importante para ti.

En primer lugar cambia su nombre de Hoja 3, por el de *Práctica*. Para lograrlo, pulsa con el botón derecho del ratón sobre la pequeña pestaña inferior de la **HOJA** en la que figura su nombre *Hoja 3*. En el menú contextual que se abre elige **Cambiar nombre...** A partir de la versión 3.1 puedes efectuar la misma operación con doble clic.

En la Hoja ya hay una tabla construida, que es copia de la de frecuencias de la Hoja de Recogida. Es una copia dinámica (o conjunto de celdas vinculadas). Si cambias datos en la primera hoja, se ven reflejados en esta otra de forma automática. La razón es que contienen fórmulas con el signo = para que se recalculen automáticamente. Lee alguna para entenderlo. No escribas, por tanto, nada en ella. Deja que copie los datos de la Hoja 1.

Para calcular la media en una tabla con frecuencias usaremos la fórmula:

$$\bar{x} = \frac{\sum x}{N}$$

La forma más cómoda de calcularla es creando otra columna junto a la F en la que figuren los productos $x \cdot f$ para después sumarlos todos en columna. Organiza el trabajo de esta forma:

Con **Insertar Nombres...** asigna el nombre de **X** a toda la columna X (Selecciona previamente toda la columna de color naranja excluyendo el rótulo "X").

Por el mismo procedimiento, asigna el nombre **F** a la columna amarilla de frecuencias, omitiendo el rótulo.

A la derecha **de la primera frecuencia**, escribe la fórmula **=X*F** y arrástrala a toda la columna.

Calcula su suma en la parte inferior con la función Σ (autosuma) e igualmente suma toda la columna de F. Llamaremos a estas sumas, para entendernos, SUMAXF Y SUMAX, respectivamente.

Deberá quedar así:

Práctica		
X	F	X*F
1	5	5
2	1	2
3	2	6
4	1	4
5	1	5
6	1	6
7	4	28
8	8	64
9	12	108
		0
		0
		0
		0
		0
		0
Totales	35	228
	Media	6,51

De esta forma tenemos preparado el cálculo de la media. En una celda cualquiera inferior a la tabla inserta el cociente **SUMAXF/SUMAX** (No lo escribas así, sino con referencias de celdas. En la imagen equivale a dividir la celda del 228 entre la del 35) y comprueba si el resultado coincide con la media de la hoja de *Cálculos*. (Si no coincide puede deberse a que no figuren todos los valores posibles en la tabla de frecuencias).

Selecciona cualquier celda libre en la hoja de *Prácticas*, abre el ASISTENTE DE



y busca, entre las funciones de matriz la de SUMA.PRODUCTO, que te permite multiplicar ordenadamente los elementos de dos matrices (no confundir con el producto de matrices) y después sumarlos.

En nuestro caso deseamos calcular $\sum \mathbf{X} * \mathbf{F} / \mathbf{S F}$ sin la ayuda de la *autosuma*. Por tanto, como argumento de la función SUMA.PRODUCTO selecciona por separado las dos columnas **X** y **F** respectivamente. Observa la imagen:



Una vez construida la función SUMA.PRODUCTO, divide su resultado entre **SUMA** (columna de las frecuencias).

Quedará de esta forma:

$$=SUMA.PRODUCTO(B6:B20;C6:C20)/SUMA(C6:C20)$$

para obtener la media por tercera vez. Comprueba el resultado.

ESTADÍSTICA BIDIMENSIONAL

Aunque sea brevemente, describiremos a continuación todas las funciones que contiene una hoja de cálculo para el estudio conjunto de dos variables X e Y. Es un caso tan frecuente en las Ciencias, que resulta muy útil disponer de todas las funciones importantes implementadas de origen en el programa.

Para describir todas las funciones disponibles, puedes abrir la hoja de cálculo **BIDIMENSIONAL.ODS**. Si te pide *Habilitar macros*, responde afirmativamente.

Columna de la variable X	Columna para la Y	I
34	411,6	
35,5	428	
40	431,3	
47	441,2	
48,2	437,4	
48,2	439,6	
48,4	443,1	
49	443,4	
50	441,2	
51,5	447,6	
52	445,7	
52,2	450,2	
53	453,4	
53,2	442,1	
53,2	443,8	

Como ves, dispone de una zona de entrada de datos (los datos X e Y de la izquierda), a cuyas columnas se les han definido los nombres de X e Y respectivamente.

Puedes borrar con Supr o con el botón que se ha creado para esta operación. También puedes rellenarla desde otro archivo con Copiar y Pegar.

A su derecha dispones de una serie de cálculos que enumeraremos a continuación, y, por último, el gráfico de tipo XY con su recta de ajuste lineal incorporada.

Cálculos

Recorreremos todas las funciones importantes que se usan en estos cálculos:

Media: Ya la conocemos, es la función **PROMEDIO**. Como aquí comparamos dos variables, usaremos las fórmulas **PROMEDIO(X)** y **PROMEDIO(Y)**. Ve recorriendo en el modelo las celdas correspondientes.

Desviación típica: También la conocemos, se trata de **DESVTP(X)** y **DESVTP(Y)**.

Regresión: A continuación se han incluido los dos coeficientes de regresión, a fin de ajustar los datos a la recta de regresión $y = ax + b$ (ajuste por mínimos cuadrados) y poder efectuar así los pronósticos. Dispones, como ves, de dos: **PENDIENTE(Y;X)**, que te devuelve la inclinación **a** de la recta de ajuste, e **INTERSECCIÓN.EJE(Y;X)**, que nos calcula el coeficiente **b** de la recta.

También disponemos de la función **PRONÓSTICO**, que nos devuelve el pronóstico para Y, correspondiente a un valor concreto de X, tomado en la recta de regresión. Puedes calcular los pronósticos para una X determinada escribiéndola en la celda rotulada con la frase *Escribe un valor de X*.

Bondad del ajuste: El grado de ajuste de la recta de regresión respecto a los datos se mide en este modelo mediante otras dos funciones de OpenOffice Calc:

Coefficiente de correlación: Número entre -1 y 1 que mide el grado de paralelismo entre las dos columnas de datos. Su función es **COEF.DE.CORREL(Y;X)**

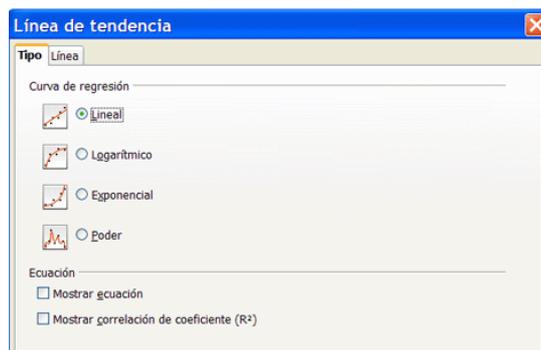
Error de estimación: Evalúa el error esperado en cada pronóstico individual. Se usa la función **ERROR.TÍPICO.XY(Y;X)**.

Gráfico

El gráfico adecuado para esta situación es el de tipo **XY**. Observa que se le ha añadido la recta de regresión. Si deseas incorporar esa recta a tus gráficos del tipo XY, basta que sigas estas operaciones:

Haz doble clic sobre el gráfico

Señala la línea de datos y usa el botón derecho para abrir el menú contextual. Elige **Insertar una línea de tendencia**.



Una vez obtenida la ventana de opciones, eliges la línea de tendencia de tipo **Regresión Lineal**. Es conveniente que actives **Mostrar ecuación** y **Mostrar correlación de coeficiente (R^2)** si dominas estos temas.

Con esto hemos presentado de forma sucinta todas las ayudas que la Hoja de Cálculo puede prestar en un análisis de dos conjuntos relacionados de datos. Nos detenemos aquí, para no alargar demasiado esta sesión.

PRÁCTICA

Errores en medidas repetidas

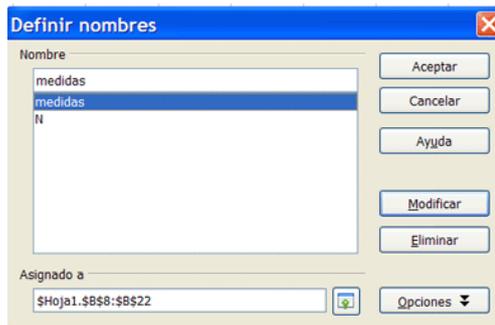
Como práctica de Estadística reproducirás los cálculos posteriores a una serie de medidas repetidas de una misma magnitud, procedimiento muy usado para estimar la medida verdadera y, en lo posible, su incertidumbre. Construirás un modelo que admita los resultados de las medidas y efectúe estimaciones sobre ellas.

Existen muchos métodos aproximados para estimar el error en una serie de medidas. Aunque la estimación de la medida verdadera se suele efectuar a partir de la **media**, en el cálculo de la incertidumbre de una medida se pueden usar diversos estadísticos, como la desviación media, la desviación estándar, el error típico, etc. Nosotros usaremos este último, que es el que mejor clasifica, a posteriori, las distintas medidas como fiables o desechables.

Comienza la práctica diseñando todos los textos y la estructura de las tablas de datos. Puede ser el de la imagen que sigue, pero cámbialo si tienes otras preferencias, o si usas ya otro esquema en trabajos de laboratorio. También puedes sustituir el error típico por otra medida que sueles usar. No escribas los símbolos O, X, y XX, que eso vendrá después. Por ahora escribe solo algunos datos, preferiblemente los mismos que figuran en la imagen de abajo.

Errores en medidas directas repetidas					
Escribe aquí los resultados de tus medidas (hasta 15)		Escribe la precisión del instrumento de medida (error instrumental)		0,05	
1	2,38	O			
2	2,45	O	El valor estimado de tu medida es	2,4030	
3	2,33	O			
4	2,27	X	El error típico equivale a	0,0292	
5	2,4	O			
6	2,35	O			
7	2,61	XX	Tomamos como error el máximo entre el típico y el instrumental	0,0500	
8	2,47	O			
9	2,39	O			
10	2,38	O	Según estos resultados, indica cuántos decimales se usarán	2	
11					
12					
13					
14			Medida estimada	2,4	
15			Error estimado	0,05	
Total medidas	10		Expresión final	2,4	± 0,05

En la imagen se ve que se podrán admitir hasta 15 medidas, cantidad que quizás sea un poco excesiva. La zona en la que escribirás esas medidas es la de color amarillo. Selecciónala completa, y dale el nombre de **medidas**. Recuerda que basta con que pidas **Insertar > Nombres > Definir** y en el marco correspondiente escribes **medidas** y pulsas **Aceptar**.



A partir de aquí te limitarás a rellenar fórmulas y más datos.

Total medidas: Basta que uses la fórmula **=CONTAR(medidas)**. En el ejemplo son 11. Puedes situar el total debajo de la tabla.

Dale a ese total el nombre de **N**, por el procedimiento explicado en el párrafo anterior.

Error instrumental: Ese dato lo escribirás directamente, según la precisión que tenga el instrumento de medida. En la imagen es 0,05

Valor estimado: Hemos indicado que se suele usar la media de los datos, luego escribe la fórmula **=PROMEDIO(medidas)**. En el ejemplo da el resultado de 2,3927

Error típico: Usaremos el error del muestreo, que equivale a la desviación estándar de la muestra (conjunto de medidas) dividida entre la raíz cuadrada de su número, luego la fórmula adecuada será **=DESVEST(medidas)/RAÍZ(N)**. Si has usado los datos del ejemplo obtendrás 0,0283

Error máximo: Se suele elegir como error definitivo el máximo entre el error que nos da la Estadística y la precisión del instrumento. Por tanto, la fórmula adecuada será **=MÁX(I6;I11)** (suponiendo que fueran esas las celdas. Tú deberás usar, en lugar de I6 e I11, las referencias de las celdas en las que tengas los dos errores).

Decimales: A la vista del error, deberás elegir cuántas cifras significativas se usarán. En este modelo sólo se ajustan los decimales. Si el error es mayor que 1, no funcionará y deberás escribir tú lo que sigue.

En la imagen figuran dos decimales, para adaptarse a la primera cifra significativa del error de 0,05.

Medida estimada y error estimado: Ambas celdas usan la función REDONDEAR. Para escribir la media redondeada escribirás esta fórmula

=REDONDEAR("Celda en la que esté la media";"Celda en la que estén los decimales")

Sustituye la celda de la media por la referencia que tenga en tu modelo, y haz lo mismo con la celda de los decimales.

Para redondear el error debes utilizar también REDONDEAR.

Te dejamos como ejercicio la expresión final. Deberás usar el signo = para copiar los dos últimos datos en otra celda.

Para el \pm usa la orden **Insertar > Símbolos...**

Distinción entre medidas aceptables y rechazables (opcional): Si la Estadística no es lo tuyo, olvida lo que sigue. Si te apetece intentarlo, lo puedes hacer así:

Si la distribución de errores siguiera la distribución normal (aproximadamente es así. Gauss descubrió su campana estudiando errores), el porcentaje de las medidas que se separa más de dos veces el error, sólo sería de un 4%, y de las que se separan más de 3, prácticamente es cero.

Podríamos aprovechar esto para clasificar las medidas en:

Aceptables: Si se desvían menos que el error multiplicado por 2. Las rotulamos con O.

Sospechosas: Si su desviación está entre 2 y 3. Las marcamos con X.

Rechazables: El resto de medidas, a las que marcaremos con XX. Estas medidas se deberían borrar y volver a calcular todo. En una hoja de cálculo, con su capacidad de recalcular, esta operación es instantánea.

Esta clasificación es convencional, y sólo nos servirá para destacar las medidas poco fiables. No tienes que seguirla al pie de la letra.

Para conseguir que a la derecha de cada medida aparezca el O, la X o la XX, podrás usar esta fórmula:

```
=SI(ESNÚMERO(B8);SI (ABS(B8-I$9)<2*I$14;"O";SI (ABS(B8-I$9)<3*I$14;"X";"XX"));" ")
```

Adáptala a tu modelo. En ella B8 es cualquier medida, por ejemplo la primera. I\$9 la media (protegida por \$ para poder arrastrar fórmulas hacia abajo), y I\$14 el error típico.

Consiste en tres funciones SI anidadas. En la primera nos preguntamos si la celda B8 es un número, porque puede estar vacía. En caso afirmativo nos preguntamos si su distancia a la media es menor que dos errores típicos, con lo que la calificamos con un "O". Por último, si está a más de 3 errores le asignamos "XX" y si cae entre 2 y 3, la "X".

EJERCICIOS

EJERCICIO 1

Comparación de dos grupos (experimental y de control)

En un colegio intentan cambiar un método de enseñanza en un área. Antes de proponerlo, para actuar con cierto fundamento, se eligen unos cursos del mismo nivel y se dividen en dos grupos. Durante un trimestre se les imparten unos conocimientos totalmente nuevos a ambos grupos, pero a uno de ellos con los métodos tradicionales (grupo de control) y al otro con el nuevo método (grupo experimental). Al final se les somete a una prueba de conocimientos evaluada de 0 a 10. Imagina que los resultados fueran los contenidos en la siguiente tabla. A partir de esos datos debes responder (o resolver), mediante un modelo de Hoja de Cálculo, a estas cuestiones:

- *¿Es superior el rendimiento del grupo experimental? (Usa la media)*
- *¿Qué grupo está más disperso? Debes usar el coeficiente de variación, que es el cociente entre la desviación típica y la media.*
- *Elige un gráfico adecuado para esta situación y confecciónalo de forma que se vean claramente las diferencias.*

Debes responder en la misma hoja en la que sitúes la tabla.

Grupo de control			Grupo experimental		
2	4	5	4	2	3
4	5	4	5	6	3
3	4	3	5	8	5
5	4	3	6	8	4
6	6	2	5	7	6
6	8	3	7	6	5
7	7	2	6	5	5
5	9	4	8	7	7
8	6	6	7	6	8
6	5	5	6	4	
5	4	4	7	3	
5	3	7	5	3	
4	4	8	8	5	
6	4	6	9	4	
5	3	5	2	7	

EJERCICIO 2

Confección de una tabla de medidas típicas

Cada docente tiene su forma de evaluar y su nivel de exigencia. Por ello, una misma calificación en una evaluación puede tener significados muy distintos según el grupo al que se pertenezca. Para situar una medida dentro de su grupo se han ideado varios procedimientos, entre ellos las **medidas típicas Z** y el **rango percentil** (o percentil inverso).

Te proponemos el siguiente ejercicio:

(a) Busca en Internet o en manuales de Estadística cómo se calcula la puntuación típica Z de una medida individual en un conjunto. Haz lo mismo con el concepto de rango percentil o percentil inverso.

(b) Escribe en columna en una hoja de cálculo unas 20 o 30 calificaciones numéricas de tu alumnado (o inventadas). Calcula su media y desviación típica.

(c) En una columna paralela a ella convierte las calificaciones en puntuaciones Z mediante la fórmula adecuada.

(d) En otra columna encuentra el rango percentil de cada calificación. Puedes usar la función RANGO.PERCENTIL, en la que tendrás que concretar el conjunto de notas y la calificación individual. Es costumbre expresarlo como porcentaje.

(e) Comenta los resultados en la misma hoja.

Tendrás una dificultad con las referencias absolutas y relativas. Cuida este aspecto.

En la imagen puedes ver un trabajo similar:

	Nota	Z	Rango-perc.
	5,0	-0,3	42%
	9,6	1,63	100%
	3,7	-0,85	29%
	3,6	-0,89	21%
	6,3	0,24	50%
	3,5	-0,93	17%
	8,4	1,13	83%
	5,8	0,03	46%
	9,1	1,42	96%
	6,8	0,45	58%
	8,5	1,17	92%
	4,4	-0,55	33%
	8,4	1,13	83%
	7,6	0,79	75%
	2,0	-1,56	0%
	8,3	1,08	79%
	4,5	-0,51	38%
	6,7	0,41	54%
	7,0	0,54	63%
	2,0	-1,56	0%
	2,0	-1,56	0%
	2,3	-1,44	13%
	3,6	-0,89	21%
	7,0	0,54	63%
	7,0	0,54	63%
Media	5,72		
Desv. Tip.	2,38		

EJERCICIO 3

Análisis de regresión

En este ejercicio analizarás una serie de datos bidimensionales. Partiremos de una situación imaginaria, que da lugar a este tipo de datos:

En una fábrica de plásticos sospechan que la densidad del producto, normalmente de 1,5 Kg. por litro, ha variado en las últimas semanas. También se duda de la uniformidad de dicha densidad en los distintos momentos de producción. Para verificarlo, recogen recortes de plástico sobrantes en las diversas máquinas y miden el volumen de cada sumergiéndolos en agua. Después los pesan todos por separado, pero dentro siempre de la misma caja, obteniendo los siguientes resultados, con el peso de la caja incluido:

Núm. muestra	Volumen en cm ³	Peso con caja en g.
1	34	411,6
2	35,5	428
3	40	431,3
4	47	441,2
5	48,2	437,4
6	48,2	439,6
7	48,4	443,1
8	49	443,4
9	50	441,2
10	51,5	447,6
11	52	445,7
12	52,2	450,2
13	53	453,4
14	53,2	442,1
15	53,2	443,8

En teoría, la fórmula de tipo funcional que relaciona los datos es:

$$\text{Peso del producto} = \text{Peso de la caja} + \text{Volumen} * \text{Densidad}$$

Por tanto, podemos admitir un modelo de regresión lineal para este problema.

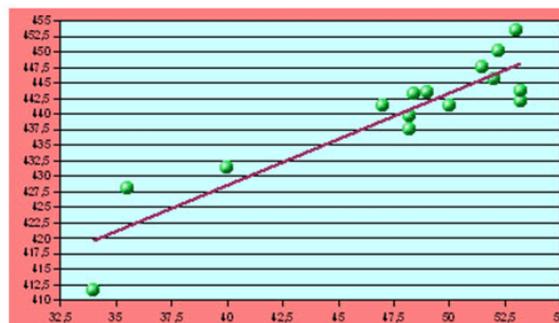
Abre el modelo `BIDIMENSIONAL.ODS`, traslada, con **Copiar** y **Pegar**, los datos de volumen y peso con caja de este documento a la zona de datos de la hoja. A la vista de los resultados, responde a estas cuestiones:

(1) Los técnicos han calculado que una desviación de 0,2 en la densidad del producto respecto a 1,5 se considera significativa y en ese caso se ha de revisar el proceso. Calcula la densidad media de los datos (¿qué resultado del modelo corresponde a la densidad?) y comenta, en la misma hoja si el proceso ha de revisarse o no por este motivo.

(2) También han decidido que si la correlación no alcanza, al menos, un valor de 0,95, se considerará que no se consigue una uniformidad mínima en la densidad. ¿Sería este caso?

(3) ¿Cuánto pesa la caja y qué inconvenientes presenta este dato?

(4) Confecciona un gráfico *de puntos* de tipo XY para esta situación y le añades posteriormente una línea de regresión lineal (puedes repasar el procedimiento en el apartado de los Contenidos *Estadística Bidimensional*). Debe quedar parecido a éste:



COMPLEMENTOS

HACER ESTADÍSTICAS

El uso de las técnicas estadísticas con la ayuda de una Hoja de Cálculo abarca tanta materia que se llenaría un curso completo. Por ello es imposible incluir en esta sesión todos los tipos de estudios estadísticos que se pueden emprender con este instrumento. No obstante, dado su interés, se ha adjuntado un

documento ([DATOS.PDF](#)) en la subcarpeta Documentos de esta sesión en el que se incluyen algunas ideas básicas para poder organizar algún trabajo estadístico con el alumnado.

En este documento se repasan los tipos de datos que existen, las operaciones que permiten y la forma de recogerlos.

¿QUÉ TIPO DE GRÁFICO ELIJO?

Es muy normal que en el momento de presentar datos sólo se usen los tipos de barras o sectores (tarta), sin darse cuenta de la riqueza informativa que se puede perder por no elegir el gráfico más adecuado a cada situación. Si has consultado en el apartado anterior el archivo *datos.pdf*, puedes leer unos consejos muy sencillos y prácticos sobre la relación entre el tipo de datos y el gráfico más adaptado a ellos en el archivo [DECIGRAF.PDF](#), situado en la misma subcarpeta Documentos.

¿QUÉ TENDENCIA SIGUEN MIS DATOS?

Cuando se estudian dos series paralelas de datos XY nos puede interesar qué tipo de tendencia siguen entre sí: lineal, exponencial, logarítmica, etc. También es posible que conozcamos previamente por la teoría el tipo de función al que se deben acercar. En una dilatación esperaremos una tendencia lineal, pero en una caída vertical puede ser de segundo grado, o en un crecimiento de una colonia de bacterias tal vez sea exponencial.

Nos interesaría disponer de un instrumento que nos ayudara a ajustar unos datos a una tendencia concreta, o bien elegir cuál es la que nos minimiza los errores cuadráticos.

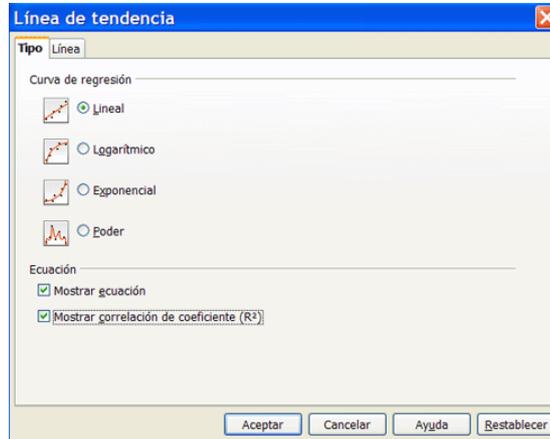
OpenOffice.org Calc nos ofrece el poder ajustar nuestros gráficos XY a distintas tendencias, y ofrece además la fórmula más adecuada y el valor de R^2 para evaluar la bondad de ese ajuste.

Puedes seguir la explicación con este ejemplo sencillo:

X	1	2	3	4	5	6	7	8	9	10	11
Y	3	4	5	7	9	12	15	17	21	23	25

Copia esta tabla y pégala en una hoja de cálculo. Si quieres los datos en columna, una vez pegados en la hoja, vuelve a copiarlos y usa **Pegado Especial y Transponer**. Si no, trabaja por filas y así practicas.

Crea un gráfico de dispersión. Haz doble clic sobre el gráfico y señala la línea que forman los datos y cuando aparezca la nota de **Serie (o punto) de datos...** abre el menú contextual con el botón derecho del ratón y elige **Insertar Línea de tendencia**. Abre la pestaña **Tipo**. Hay cuatro opciones: Lineal, Exponencial, Logarítmica o Potencial.



Para elegir el mejor ajuste activa las casillas de **Mostrar ecuación y Mostrar correlación de coeficiente (R^2)**. Prueba ahora con cada tendencia una por una y toma nota del valor de R^2 . Entre una y otra deberás suprimir la anterior. La tendencia que tenga el valor más alto será la más adecuada. En el caso del ejemplo es la lineal, con ecuación $Y=2,36X-1,36$ y un ajuste del 98%.

Este es un buen instrumento para ajustar tus datos experimentales.

SUGERENCIAS DE USO DIDÁCTICO

Como se indicó al principio de la sesión, el uso de una Hoja de Cálculo ha beneficiado bastante a la enseñanza y uso de la Estadística. Al eliminar la reiteración de cálculos y favorecer el lenguaje de tablas y gráficos, ha convertido a esta disciplina en un auxiliar importante en todas las asignaturas, dejando de ser un conjunto de capítulos molestos que eran sacrificados cuando faltaba tiempo para impartir todo un programa.

Podemos destacar algunos aspectos en los que este beneficio es más notable:

Organización de los cálculos estadísticos

Hemos experimentado desde hace algunos años la enseñanza de los conceptos y cálculos de tipo estadístico con la ayuda de calculadoras gráficas y Hoja de Cálculo. En el documento [CUANTITATIVOS.PDF](#) se puede apreciar el tipo de trabajo que se ha desarrollado, normalmente en las siguientes fases:

- Explicación teórica simultánea con la escritura y organización de las tablas y cálculos en las pantallas.
- Reparto de tareas de cálculo por equipos de alumnos y alumnas, debiendo construir modelos o seguir secuencias definidas de cálculo. Como producto de este trabajo deben obtener cálculos (medias, coeficientes de regresión, etc.) y gráficos.
- Estudio de algunos experimentos, cuestiones teóricas o cálculos rutinarios mediante el modelo que ha sido creado.
- Repaso posterior y ampliación teórica de lo aprendido, ya sin ayuda de instrumentos informáticos.

Hemos observado los siguientes beneficios:

Incremento de la atención en el alumnado: al tener que organizar los cálculos en pantalla simultáneamente a la recepción de la explicación no pueden distraerse, pues se desconectarían automáticamente del proceso general de la clase.

Atención más efectiva a la diversidad: en cursos con muchos niveles distintos de conocimientos en Estadística este proceso ayuda a respetar los distintos ritmos de aprendizaje y a diversificar el tipo de ejercicios propuestos.

Posibilidad de realización de experimentos sencillos: el cálculo automatizado permite reservar más tiempo a pequeños experimentos o recogida de datos, así como a los comentarios y conclusiones.

Esquemas de cálculos estadísticos

Puedes consultar las posibilidades en las [SUGERENCIAS DE USO](#).

Recogida de datos

En los talleres de Matemáticas, asignaturas experimentales e Informática se puede usar la Hoja de Cálculo para recoger y estudiar datos procedentes de experimentos, encuestas o simulaciones. En el documento [FRECUENC.PDF](#) se puede ver una guía para un experimento sencillo de comparación de frecuencias y probabilidades.

En la asignatura de Informática se puede organizar un ciclo de trabajo muy interesante, que abarque las siguientes fases:

Confección de modelos de simulación: trataremos de ellos en la sesión 9 de este curso. No es una tarea difícil simular una tirada de dados o una distribución de vocales, tanto en Hoja de Cálculo como en un lenguaje de programación.

Recogida de datos: los modelos citados producirán unos datos que se podrán recoger en otros modelos confeccionados para este fin (Ver **RECOGIDA DE DATOS**). En ellos se efectuarán todos los cálculos interesantes y se confeccionarán los gráficos.

Análisis y comunicación de resultados: sobre lo obtenido con la Hoja de Cálculo se puede confeccionar un Informe, tal como vimos en la sesión 4 del curso.

Este ciclo permite fijar el objetivo último en la confección del informe, que reflejará con bastante exactitud, mediante las tablas, comentarios y gráficos insertados, la calidad del trabajo.

Este tipo de trabajo, realizado en asignaturas de Técnicas de Información e Informática, interesa mucho al alumnado y le inclina más favorablemente a la Estadística, cuyo estudio está incluido en la mayoría de los programas de estas asignaturas y provoca el rechazo en alumnos que sólo desean "jugar" con los ordenadores.

Webquest y Cazas del tesoro

Las técnicas estadísticas son muy útiles en la metodología de las Webquest, pues permiten analizar y presentar mejor toda la información que se recoja en Internet sobre el tema. Por ejemplo, si organizamos un trabajo con el título de "España húmeda y España seca", deberemos apoyar esos dos adjetivos con búsquedas de estadísticas de lluvias, y las hojas de cálculo nos pueden servir para calcular medias y compararlas.

CONTENIDO

Tratamiento de datos estadísticos	2
Contenidos	2
Introducción	2
Cálculos estadísticos.....	2
Recogida de datos	4
Funciones estadísticas.....	7
Fórmulas matriciales	10
Estadística bidimensional	11
Práctica	14
Ejercicios.....	17
Ejercicio núm. 1	17
Ejercicio núm. 2	18
Ejercicio núm. 3	19
Complementos	20
Hacer estadísticas.....	20
¿Qué tipo de gráfico elijo?	21
¿Qué tendencia siguen mis datos?.....	21
Sugerencias de uso didáctico	22