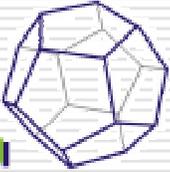


Aplicaciones de Estadística Descriptiva



Contenidos de la presentación

- ❑ Funciones estadísticas en Excel.
Gráficos.
 - ❑ El módulo de análisis de datos y las tablas dinámicas de Excel.
-



Información

- Intentaremos es obtener un resumen de la información que nos proporcionan los datos de que disponemos.
 - Para ello, primero construiremos tablas de frecuencias y observaremos los gráficos adecuados a las variables.
-



Frecuencias

- Partimos de un documento en el que están los datos organizados en una tabla y sin recontar.
 - Con Excel podemos calcular las frecuencias absolutas mediante dos funciones diferentes.
 - Las frecuencias relativas y acumuladas debemos calcularlas nosotros a través de fórmulas.
-



Función `CONTAR.SI`

- Es la función que debemos utilizar para calcular frecuencias de atributos
 - Recibe como argumentos:
 - un rango de datos (el que queremos recontar)
 - El valor del que queremos conocer la frecuencia.
 - Devuelve la frecuencia absoluta del valor indicado.
-



Función FRECUENCIA

- Esta función sólo es válida para variables numéricas y es la recomendada para agrupar en intervalos.
 - Recibe como argumentos:
 - Rango con los datos.
 - Los valores de los extremos finales de los intervalos.
 - Devuelve una matriz que tiene un elemento más que extremos hayamos indicado.
-



Representaciones

- Una vez calculadas las frecuencias, pasamos a representar los datos.
 - Un gráfico adecuado puede proporcionar mucha información sobre los datos y como se "distribuyen".
 - Son especialmente representativos en el caso de que tratemos con atributos pues serán nuestra principal referencia.
-

Gráficos

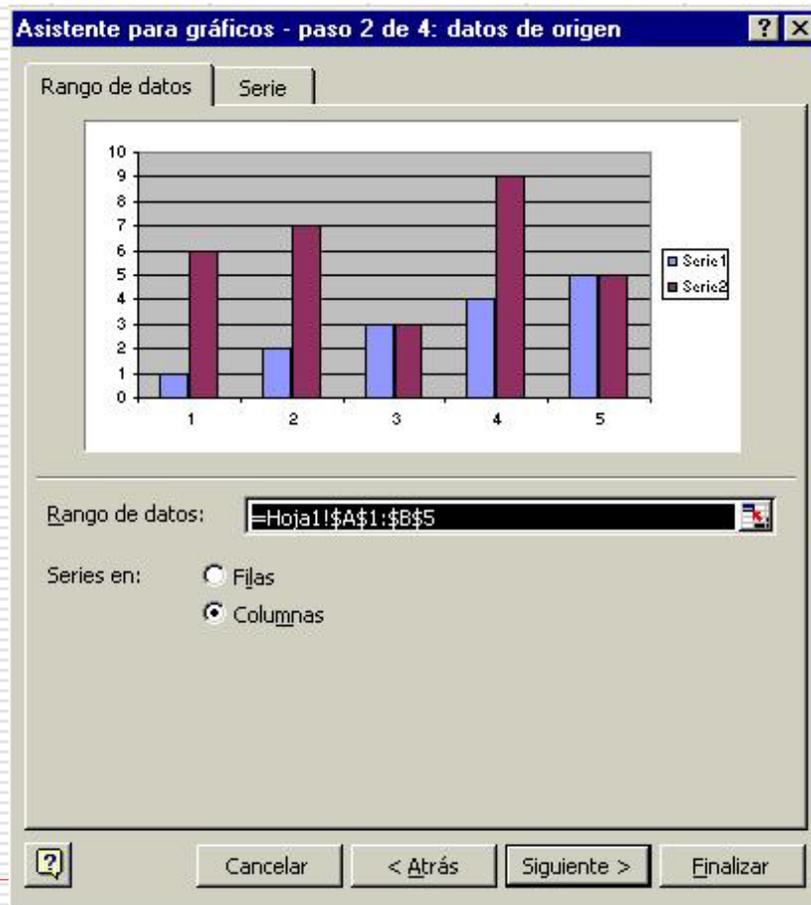
- Excel tiene bastantes posibilidades a la hora de hacer gráficas.
 - Posee un asistente que nos ayuda a la hora de construir nuestro gráfico.
 - Vamos a ir viendo los pasos que debemos seguir con dicho asistente.
 - El asistente aparecerá al pulsar el botón 
-

Gráficos (paso 1)



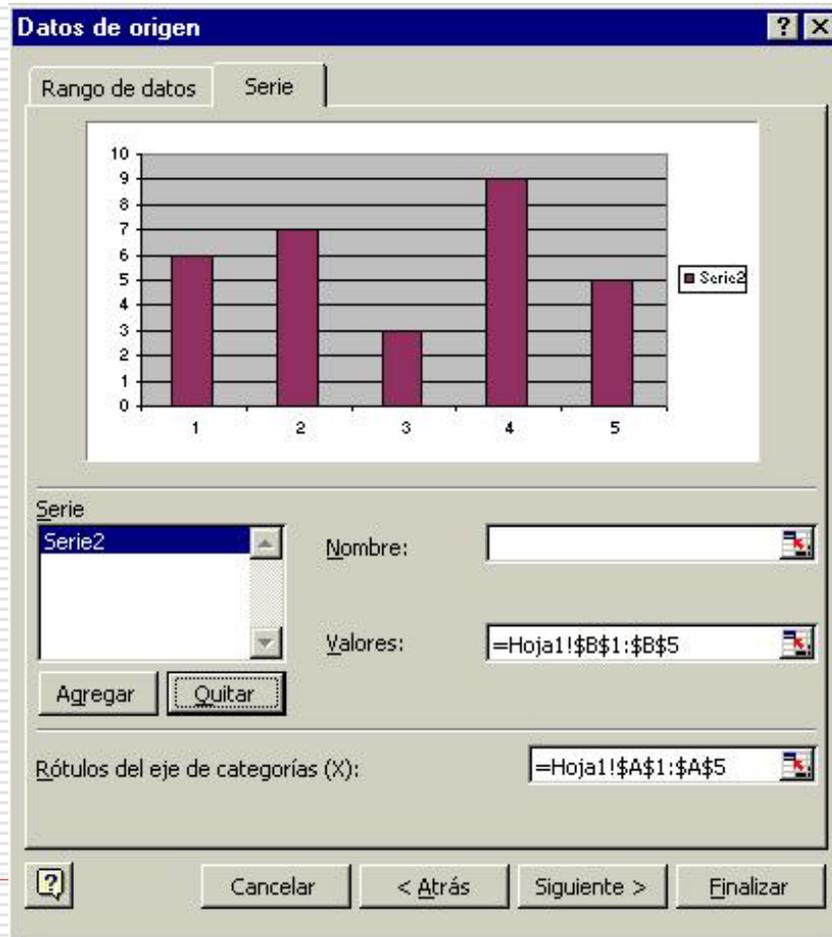
En el primer paso debemos elegir el tipo de gráfico más conveniente: barras, columnas, circular, XY (dispersión)...

Gráficos (paso 2)



Lo siguiente es elegir los datos de origen: primero si los datos están en filas o en columnas y dónde están

Gráficos (paso 2-cont)



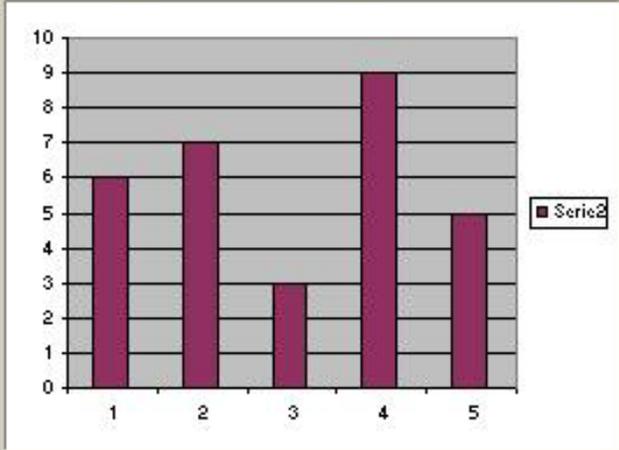
Ahora debemos elegir las series que queremos representar en este gráfico y dónde se encuentran los rótulos.

Gráficos (paso 3)

Asistente para gráficos - paso 3 de 4: opciones de gráfico

Eje principal

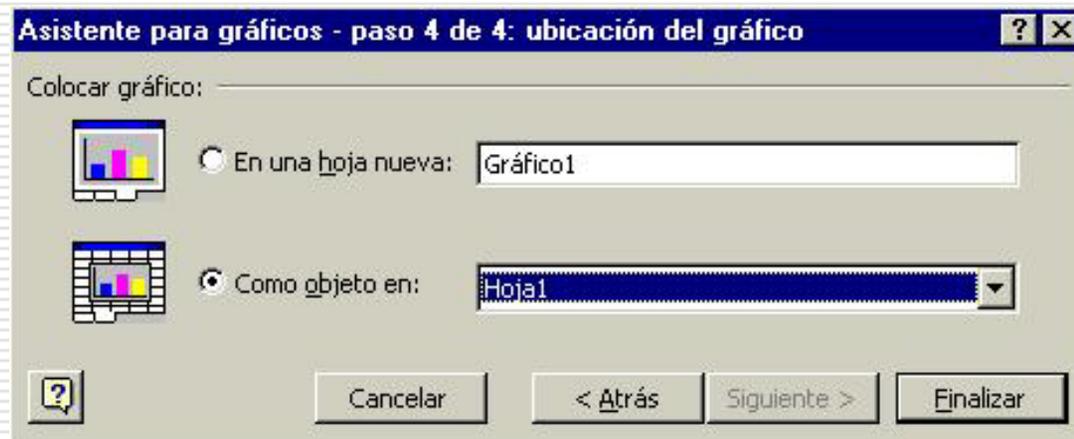
- Eje de categorías (X)
 - Automático
 - Categoría
 - Escala de tiempo
- Eje de valores (Y)



Ahora debemos elegir el aspecto de ejes, leyenda, rótulos, títulos...

Gráficos (paso 4)

Finalmente debemos elegir la ubicación del gráfico:





Gráficos (cambios)

Muchos de los elementos que hemos ido eligiendo en los pasos anteriores pueden ser cambiados utilizando la barra de herramientas gráfico:





Medidas de posición

- Excel nos permite el cálculo de las medidas de posición más importantes.
 - Las encontramos en el apartado estadística de la lista de funciones.
 - Dicha lista aparece pulsando el botón 
 - Calcularemos a partir de ahora medias, varianzas, percentiles, moda, mediana...
-



Función PROMEDIO

- La función PROMEDIO es la que utilizaremos para el cálculo de la media aritmética.
 - Esta función recibe un rango de celdas, en el que están los datos de los que queremos conocer la media
 - La función PROMEDIO **NO** permite trabajar con frecuencias, sólo podemos dar la lista de datos.
-



Función MEDIANA

- La función `MEDIANA` es la que utilizaremos para el cálculo de la mediana.
 - Recibe como argumento un rango de datos (no tabla de frecuencias).
 - En series con un número impar de valores, da el valor central, en series pares, la media de los valores centrales.
 - Elimina los valores extremos.
-



Función MODA

- La función MODA es la que utilizaremos para el cálculo de la moda de unos datos.
 - Recibe un rango de datos (no frecuencias) y devuelve el valor más frecuente.
 - En caso de que no hubiese valores repetidos o que fueran atributos, devuelve #N/A, si aparecen dos igual de frecuentes, devuelve el primero que aparece.
-



Otras funciones

- Otras medidas de posición central, aunque menos habituales son la media geométrica y la media armónica. Ambas podemos calcularlas.
 - La media geométrica se calcula mediante la función `MEDIA.GEOM`
 - La media armónica se calcula mediante la función `MEDIA.ARMO`.
-



Función PERCENTIL

- La función PERCENTIL es la que utilizaremos para el cálculo de cualquier percentil de los datos.
 - Recibe como argumentos el rango en el que están los datos y k , el valor del percentil, que debe estar entre 0 y 1 (para el percentil 30, indicaremos 0,3).
 - En caso de que el percentil se encuentre entre dos observaciones, Excel determina su valor por interpolación.
-



Función CUARTIL

- A pesar de que los cuartiles son un caso particular de los percentiles, Excel posee también una función para su cálculo.
 - Esta función recibe como argumentos el rango de datos y el número del cuartil que queremos calcular (1, 2, 3 o 4).
 - El cuartil 1 corresponde al percentil 25, el 2 a la mediana, el 3 al percentil 75 y el 4 al percentil 100.
-



Dispersión

- Veremos ahora las medidas del error que cometeríamos al representar el conjunto de datos a través de una medida de tendencia central.
 - Veremos el rango, rango intercuartílico, la varianza y desviación típica, que son las medidas más habituales.
-



Cálculo del rango

- Excel no tiene una función para calcular el rango, aunque es fácil calcularlo haciendo uso de dos de las funciones que posee.
 - Dichas funciones son la función `MAX` y la función `MIN`.
 - Ambas reciben como argumentos un rango de datos y devuelven, respectivamente el mayor y el menor valor.
 - El rango puede calcularse como la diferencia entre ambas.
-



Cálculo del rango intercuartílico

- Como en el caso del rango, Excel no tiene una función específica para el rango intercuartílico.
 - Lo que haremos es calcular los cuartiles primero y tercero y hallar la diferencia entre ambos.
-



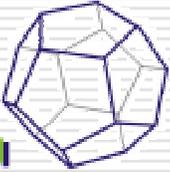
Función `VARP`

- La función `VARP` es la que nos permite calcular la varianza de todo un conjunto de datos.
 - Recibe como argumento el conjunto de datos del que queremos conocer su varianza.
 - OJO: La función `VAR` nos calcula la varianza de una muestra de los datos que tenemos, no de todos.
-



Función DESVESTP

- La función `DESVESTP` será la que utilizaremos para calcular la desviación típica de un conjunto de datos.
 - Recibe como argumentos los datos de los que queremos conocer la desviación típica.
 - OJO: la función `DESVEST` calcula la desviación estándar de una muestra de los datos que indiquemos.
-



Cálculo del coeficiente de variación

- Para el coeficiente de variación, Excel tampoco tiene una función específica.
 - Podemos calcularlo haciendo uso de la definición, ya que es la desviación típica dividida entre la media (en valor absoluto).
-



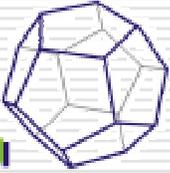
Otras funciones

- Excel nos permite el cálculo de otras medidas. Veamos algunas.
 - El coeficiente de asimetría lo podemos calcular mediante la función `COEFICIENTE.ASIMETRIA`
 - El coeficiente de curtosis se puede calcular mediante la función `CURTOSIS`
-



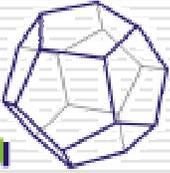
Cálculos para dos variables

- Excel tiene numerosas funciones que pueden ser utilizadas para el análisis descriptivo de dos variables.
 - Es una herramienta importante para el cálculo de rectas de regresión, covarianzas, coeficiente de correlación...
 - Como en el caso de una variable, comenzamos por resumir la información mediante tablas de frecuencias y gráficos.
-



Excel. Herramientas básicas

- Excel es una herramienta potente para realizar cálculos estadísticos.
 - Posee un conjunto de funciones, englobadas dentro de la categoría “Estadísticas” que permiten calcular la mayoría de medidas estadísticas comunes.
 - Asimismo posee un asistente para la realización de gráficos que permite hacer los gráficos más frecuentes.
-



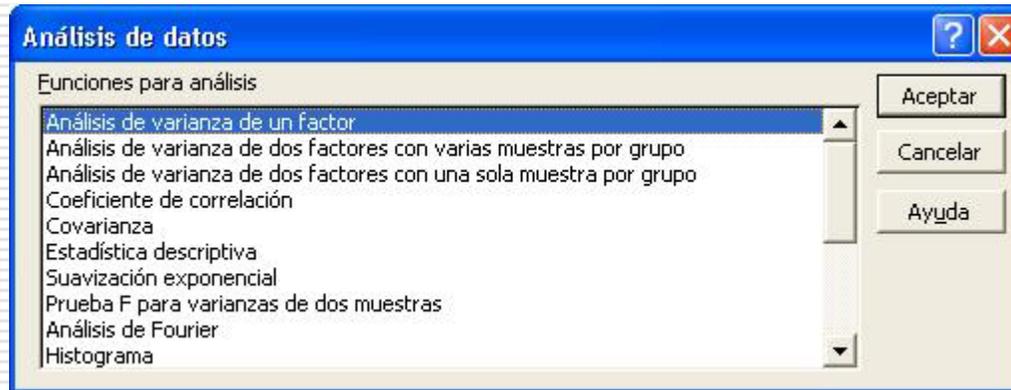
Excel. Módulo Análisis de Datos (1)

- El módulo Análisis de datos es un complemento de Excel que habitualmente no se instala por defecto.
 - Para instalarlo, es necesario seleccionar la opción “Herramientas para el análisis” dentro del menú Herramientas->complementos.
 - Es posible que necesitemos el cd de instalación para poder disponer de este módulo.
-



Excel. Módulo Análisis de Datos (2)

- Al elegir la opción "Análisis de Datos" nos aparece la siguiente ventana:



- Veremos las opciones "Estadística descriptiva", "Histograma" y "Regresión"
-



Excel. Módulo Análisis de Datos (3)

- La opción estadística descriptiva genera un informe estadístico de una sola variable para los datos del rango de entrada, y proporciona información acerca de la tendencia central y dispersión de los datos.
 - Dicho resumen presenta medidas como: media, mediana, moda, error típico, curtosis, desviación típica, varianza, coeficiente de asimetría, etc.
-



Excel. Módulo Análisis de Datos (4)

- El cuadro de diálogo que nos aparece es:

Estadística descriptiva

Entrada

Rango de entrada:

Agrupado por: Columnas Filas

Rótulos en la primera fila

Opciones de salida

Rango de salida:

En una hoja nueva:

En un libro nuevo

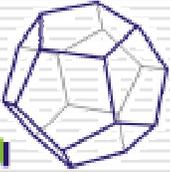
Resumen de estadísticas

Nivel de confianza para la media: %

K-ésimo mayor:

K-ésimo menor:

Aceptar
Cancelar
Ayuda

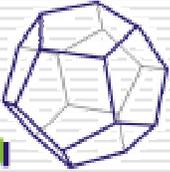


Excel. Módulo Análisis de Datos (5)

- Las opciones a detallar son:

- o Rango de entrada: celdas en la que se encuentran los datos. Asimismo debemos marcar si la primera celda es un el título.

- o Opciones de salida: lugar en el que nos va a aparecer el resumen de estadísticas. También debemos marcar qué queremos que nos calcule. Además del resumen, podemos pedir que nos calcule el elemento que ocupa el k menor lugar o el k mayor, así como el nivel de confianza



Excel. Módulo Análisis de Datos (6)

- La siguiente opción que veremos dentro del módulo análisis de datos es la opción histograma.
 - Excel tiene el problema de que dentro del menú gráficos no permite hacer un histograma de los datos de manera correcta. Esta opción viene a cubrir dicha carencia.
 - Podemos además no indicar la posible agrupación en clases de manera que se nos propone una clasificación que se puede utilizar como orientación.
-



Excel. Módulo Análisis de Datos (7)

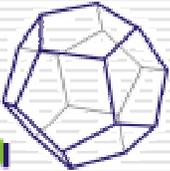
- El cuadro de diálogo que nos aparece al elegir la opción "histograma" es el siguiente:





Excel. Módulo Análisis de Datos (8)

- Como en el caso anterior, debemos indicar el lugar donde se encuentran los datos de los que queremos obtener el histograma, así como la salida que queremos obtener.
 - En este caso, podemos además indicar las clases en que hemos dividido los datos, en caso de tenerlas, o dejar este apartado en blanco para que Excel proponga una clasificación.
 - Por último, debemos indicar las salidas que queremos que Excel nos ofrezca: histograma ordenado, los porcentajes acumulados por clases y el gráfico (histograma) son las opciones.
-



Excel. Módulo Análisis de Datos (9)

- El módulo de regresión realiza un análisis de regresión lineal utilizando el método de los "mínimos cuadrados" para ajustar una línea a una serie de observaciones.
 - Presenta de manera conjunta varias de las opciones que ya se ofrecen entre las funciones estadísticas, y además proporciona un análisis de los residuos y la representación gráfica de los resultados obtenidos.
-



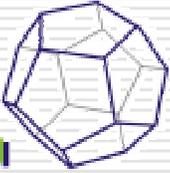
Excel. Módulo Análisis de Datos (10)

- El cuadro de diálogo que nos aparece es:

The screenshot shows the 'Regresión' dialog box in Excel. It has a blue title bar with a question mark and a close button. The dialog is divided into several sections:

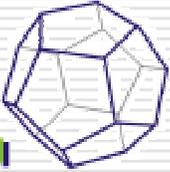
- Entrada:**
 - Rango Y de entrada: [Empty text box]
 - Rango X de entrada: [Empty text box]
 - Rótulos
 - Constante igual a cero
 - Nivel de confianza: 95 %
- Opciones de salida:**
 - Rango de salida: [Empty text box]
 - En una hoja nueva: [Empty text box]
 - En un libro nuevo
- Residuales:**
 - Residuos
 - Residuos estándares
 - Gráfico de residuales
 - Curva de regresión ajustada
- Probabilidad normal:**
 - Gráfico de probabilidad normal

On the right side, there are three buttons: 'Aceptar', 'Cancelar', and 'Ayuda'.



Excel. Módulo Análisis de Datos (11)

- En las opciones rango de entrada y opciones de salida volvemos a proceder como en los casos anteriores.
 - Adicionalmente, en este caso podemos marcar la casilla constante igual a cero, si queremos forzar la recta de regresión a pasar por el cero, y podemos indicar otro nivel del confianza diferente al predeterminado (95%)
 - En el apartado residuales se puede elegir que nos proporcione los residuos y los residuos estándares así como su gráfico, además de poder elegir que nos represente la recta de regresión.
 - Además podemos representar un gráfico de probabilidad normal
-



Excel. Módulo Análisis de Datos (12)

- El módulo análisis de datos tiene también otras opciones interesantes que facilitan ciertos cálculos que pueden resultar muy pesados. Entre ellas, podemos destacar las opciones "correlación" y "covarianza" que mejoran las funciones correspondientes que podemos encontrar de manera habitual en Excel.
-



Tablas dinámicas (1)

- Las tablas dinámicas son una herramienta de Excel que nos permite hacer múltiples operaciones para cruzar variables.
 - Las encontramos dentro del menú “datos” en la opción “Informe de tablas y gráficos dinámicos”.
-

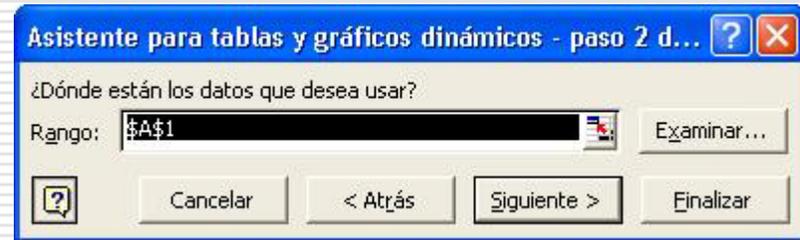
Tablas dinámicas (2)

- Cuando elegimos la correspondiente opción obtenemos el siguiente cuadro de diálogo

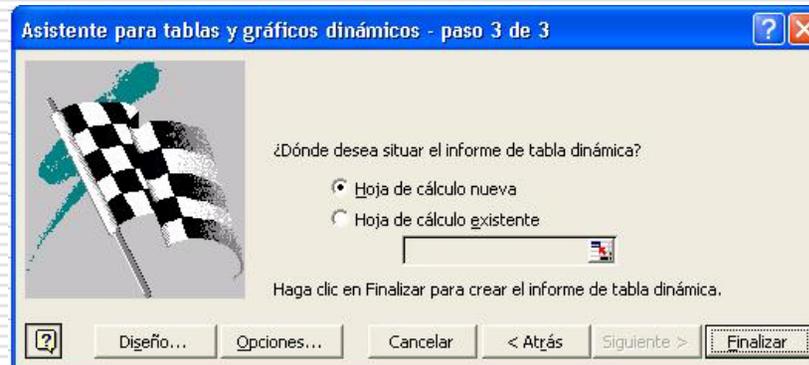


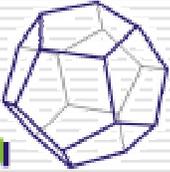
Tablas dinámicas (3)

- Pulsamos siguiente y aparece:



- Elegimos el rango de datos y pulsamos siguiente y vemos:





Tablas dinámicas (4)

- Si presionamos Finalizar ya nos aparecen los campos que podemos incluir en la tabla y los valores que queremos que sean calculados
 - Vemos en detalle cómo hacerlo a través de un ejemplo.
-



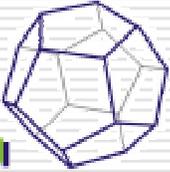
Gráficos

- En el caso bidimensional, los gráficos toman una relevancia especial.
 - El diagrama de dispersión (XY en Excel) nos permite obtener muchas conclusiones.
 - Podemos ver si la forma nos indica algún tipo de dependencia, si existe relación inversa o directa, si se aproxima a una recta, etc. Posteriormente pasamos a comprobar estas conclusiones mediante las correspondientes medidas.
-



Función `COVAR`

- La función `COVAR` es la que nos permite el cálculo de la covarianza de dos variables.
 - Recibe como argumentos dos rangos de datos.
 - Es claro que dichos rangos deben tener el mismo número de observaciones.
-



Función COEF.DE.CORREL

- Esta es la función que utilizaremos para el cálculo del coeficiente de correlación.
 - Recibe como argumentos dos rangos de datos.
 - Como en el caso anterior, los dos rangos deben tener el mismo número de datos.
-



Aproximaciones

- Una vez que tenemos calculados la covarianza y el coeficiente de correlación lineal, podemos calcular la recta de regresión para las dos variables.
-



Rectas de regresión (I)

- Podemos calcular rectas de regresión en Excel a través de la función `ESTIMACION.LINEAL`
- Veremos paso a paso qué es lo que debemos hacer para obtener dicha recta y otros estadísticos relacionados.
- Suponemos que vamos a calcular la recta de regresión entre Y y X con la forma

$$Y = mX + b$$



Rectas de regresión (II)

ESTIMACION.LI... = =ESTIMACION.LINEAL()

ESTIMACION.LINEAL

Conocido_y	<input type="text"/>		= referencia
Conocido_x	<input type="text"/>		= referencia
Constante	<input type="text"/>		= valor_lógico
Estadística	<input type="text"/>		= valor_lógico

=

Devuelve una matriz con la línea recta que mejor describe los datos, calculada usando el método de los mínimos cuadrados.

Conocido_y es el conjunto de valores de Y conocidos en la relación $y = mx + b$.

Resultado de la fórmula =

Éste es el cuadro de diálogo que nos aparece. Veamos los detalles.



Rectas de regresión (III)

- En **conocido_y** debemos introducir los valores de la variable independiente (obligatorio)
- En **conocido_x** debemos introducir los valores de la variable dependiente.
- En **constante** debemos introducir verdadero (o nada) si queremos que se calcule b o falso si tenemos la certeza de que la constante es 0.
- En **estadística** debemos introducir verdadero si queremos calcular estadísticos adicionales, o falso en caso contrario



Rectas de regresión (IV)

- Los estadísticos adicionales se nos muestran en la forma:

m	b
$Se(m)$	$Se(b)$
R^2	$Se(y)$
F	df
$Ss(reg)$	$Ss(res)$

Donde:



Rectas de regresión (V)

- **m** es la pendiente de la recta de regresión. Se puede calcular también con la función `PENDIENTE`
- **b** es el punto de intersección con el eje Y. Se puede calcular también con la función `INTERSECCION.eje`
- **se(m)** valor del error típico de la pendiente
- **se(b)** valor del error típico de la intersección
- **R²** coeficiente de determinación de la recta. Se puede calcular también con `COEFICIENTE.R2`
- **se(y)** error típico de la estimación. Se puede calcular también con `ERROR.TIPICO.XY`



Rectas de regresión (VI)

- **F** es el estadístico F o valor F observado. Se puede calcular también con `DIST.F`
 - **df** son los grados de libertad para la prueba F
 - **se(reg)** es la suma de los cuadrados de la regresión.
 - **se(resid)** es la suma de los cuadrados de los residuos.
-