

UNIDAD DIDÁCTICA 2

USO DE EXCEL EN ESTADÍSTICA

1. FÓRMULAS MATRICIALES

Las fórmulas matriciales de Excel permiten ejecutar operaciones en rangos de celdas simultáneamente. Supongamos que tenemos una lista de números y queremos conocer la media de variación que experimentan los valores entre sí.

	A	B	C	D	E	F	G
1	17,24						
2	12,445	-4,795	←	=A2-A1			
3	18,639	6,194					
4	17,065	-1,574					
5	13,76	-3,305					
6	13,671	-0,089					
7	12,437	-1,234					
8	15,379	2,942					
9	14,876	-0,503					
10	17,923	3,047					
11							
12		0,076	←	=PROMEDIO(B2:B10)			
13							
14							
15		#¡VALOR!	←	=PROMEDIO(A2:A10-A1:A9)			
16							
17		0,076	←	{=PROMEDIO(A2:A10-A1:A9)}			
18							
19							

El procedimiento que se seguiría normalmente para abordar un problema de este tipo, consistiría en añadir una nueva columna en la que se efectuarían los cálculos de los pasos intermedios. En la celda **B2**, calculamos la diferencia entre **A2** y **A1**. La fórmula correspondiente se introduce en la celda **B10**. De esta forma, en la celda **B12** obtendríamos sencillamente el valor de la media.

Con una fórmula matricial, obtendríamos el mismo resultado utilizando una sola celda. La fórmula que se introduciría en la celda **B15** sería la siguiente:

=PROMEDIO(A2:A10-A1:A9)

Hasta aquí todo resulta lógico. Queremos calcular la media de las diferencias, y este es precisamente el valor que solicita la fórmula introducida; sin embargo, *Excel* nos devuelve un valor de error. Esto se debe a que debemos introducir la fórmula de forma que le indiquemos a *Excel* que ésta debe aplicarse a rangos de celdas y no a celdas individuales. Para ello, deberemos pulsar **Ctrl-Mayús-Intro**, al introducir la fórmula. *Excel* mostrará entonces la fórmula entre corchetes, tal y como puede apreciarse en la celda **B17**.

{=PROMEDIO(A2:A10-A1:A9)}

Las fórmulas matriciales permiten ganar espacio en la hoja de cálculo, pudiéndose eliminar con frecuencia columnas enteras de fórmulas. Al reducirse la complejidad visual, se facilita en gran medida la comprensión de los datos y se agilizan los cálculos efectuados en la hoja de trabajo.

2. DIRECCIONES E INDIRECCIONES

Las funciones **INDICE**, **INDIRECTO**, **DIRECCION** y **DESREF** se utilizan frecuentemente, proporcionándonos fórmulas que ofrecen una mayor flexibilidad de la que nos facilitan las referencias a celdas.

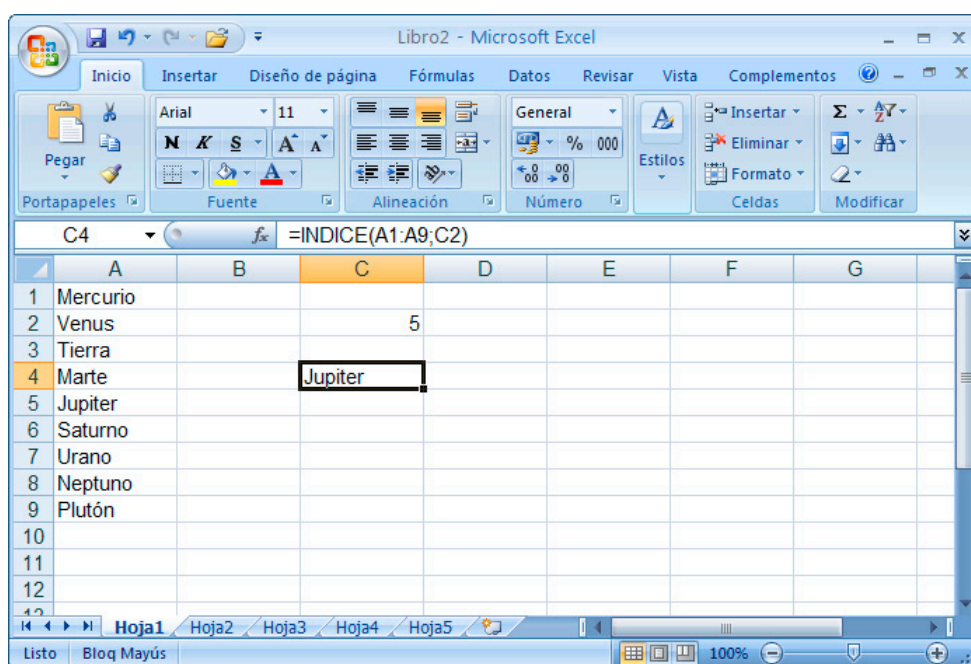
Todas las celdas de una hoja de cálculo tienen asignada una dirección única (por ejemplo, A1) y *Excel* nos permite hacer referencia a las mismas utilizando su dirección. Pero qué sucedería si desconocemos la dirección de la celda. Esto podría ocurrir, por ejemplo en los casos en que *Excel* haya actualizado la dirección de la información requerida para reflejar los cambios introducidos en la hoja de cálculo.

Supongamos que tenemos una lista de elementos en la que deseamos realizar una selección. La función **INDICE** se ocupará de hacerlo. Esta lista puede encontrarse en una columna o en una fila.

Tenemos una lista de nueve planetas. En la celda C4, la fórmula es:

=INDICE (A1:A9; C2)

El primer argumento de la fórmula corresponde al rango en el que están ubicados los nombres de los planetas. El número de la celda **C2**, indica a la fórmula el planeta que buscamos. En nuestro caso, se trata del planeta número 5. El planeta número 5, ubicado en la celda **A5**, es Jupiter. La fórmula de la celda **C4** nos devuelve el valor Jupiter en forma de texto.



La función **INDICE** resulta apropiada en la mayoría de los casos, pero puede ocurrir que no sepamos a veces en qué fila o columna se encuentra el valor requerido. Dicho valor podría encontrarse en cualquier lugar de la hoja de cálculo e incluso en una hoja de cálculo diferente.

Observe el siguiente ejemplo:

	A	B	C	D	E	F	G
1	Productos	TRI 1	TRI 2	TRI 3	TRI 4		
2	Ordenadores	309	212	216	201		
3	Periféricos	159	411	373	175		
4	Consumibles	262	385	8	215		
5	Almacenamiento	53	412	19	6		
6							
7							
8							
9							
10							
11			Columna	4			
12			Fila	3			
13							
14			Direccion	\$D\$3			
15							

En este caso, los datos ocupan varias columnas. El valor que contiene la celda **D3** es 373. **D3** es la celda de la cuarta columna y la tercera fila. Los números correspondientes a la fila y a la columna se muestran en las celdas **D12** y **D11**. La función **DIRECCION**, de la celda **D14** utiliza estos valores para generar la dirección.

=DIRECCION(D12;D11)

Siempre debe colocarse la fila antes que la columna. En caso de que necesitemos incluir el nombre de la hoja de cálculo en la dirección porque tengamos que hacer referencia a datos que se encuentran en otra hoja de cálculo, se deberá utilizar la siguiente fórmula:

=DIRECCION(D12;D11;;; "Libro3")

Esta fórmula permite incluir el nombre de la hoja y generarla en el valor que nos devuelve Excel.

	A	B	C	D	E	F	G
1	Productos	TRI 1	TRI 2	TRI 3	TRI 4		
2	Ordenadores	309	212	216	201		
3	Periféricos	159	411	373	175		
4	Consumibles	262	385	8	215		
5	Almacenamiento	53	412	19	6		
6							
7							
8							
9							
10							
11			Columna	4			
12			Fila	3			
13							
14			Direccion	\$D\$3			
15							
16			Con nombre				
17			de hoja	Libro3!\$D\$3			
18							
19							

La dirección por sí sola no resulta de gran ayuda, pero la función **INDIRECTO** nos devuelve el valor correspondiente a una dirección. En la celda **D20** la función **INDIRECTO** utiliza la dirección generada en la celda **D14** para recuperar el valor de la celda **D3**.

	A	B	C	D	E	F	G
1	Productos	TRI 1	TRI 2	TRI 3	TRI 4		
2	Ordenadores	309	212	216	201		
3	Periféricos	159	411	373	175		
4	Consumibles	262	385	8	215		
5	Almacenamiento	53	412	19	6		
6							
7							
8							
9							
10							
11			Columna	4			
12			Fila	3			
13							
14			Direccion	\$D\$3			
15							
16			Con nombre				
17			de hoja	Libro3!\$D\$3			
18							
19							
20			El valor	373			
21							

Así pues, la fórmula de la celda **D20** podría escribirse también de la forma siguiente:

=INDIRECTO (DIRECCION (D12; D11))

Aunque estas funciones pueden parecer a primera vista bastante complejas, resultan muy útiles cuando tenemos que hacer referencia a datos utilizando valores tanto de filas como de columnas. Puede que necesite un poco de práctica hasta llegar a dominar las funciones **DIRECCION** e **INDICE**; sin embargo, merece la pena familiarizarse con ellas ya que proporcionan una gran flexibilidad en los casos en que necesitamos hacer referencia a datos en *Excel*.

Excel, como es habitual, proporciona más de un método para realizar esta misma operación. La función **DESREF** nos permite operar del mismo modo que las funciones **INDIRECTO** y **DIRECCION**, utilizando un enfoque diferente.

DESREF(ref;filas;columnas;alto;ancho)

- **Ref:** es la referencia en la que se desea basar la desviación. Ref debe referirse a una celda o rango de celdas adyacentes; en caso contrario, DESREF devuelve el valor de error #¡VALOR!
- **Filas:** es el número de filas, hacia arriba o hacia abajo, al que se desea que haga referencia la celda superior izquierda. Si el argumento filas es 5, la celda superior izquierda de la referencia pasa a estar cinco filas más abajo que la referencia. Filas puede ser positivo (lo que significa que está por debajo de la referencia de inicio) o negativo (por encima).
- **Columnas:** es el número de columnas, hacia la derecha o izquierda, al que se desea que haga referencia la celda superior izquierda del resultado. Si el argumento columnas es 5, la celda superior izquierda de la referencia pasa a estar cinco columnas hacia la derecha de la referencia. Columnas puede ser positivo (lo que significa a la derecha de la referencia de inicio) o negativo (a la izquierda).
- **Alto:** es el alto, en número de filas, que se desea que tenga la referencia devuelta. El alto debe ser un número positivo.
- **Ancho:** es el ancho, en número de columnas, que se desea que tenga la referencia devuelta. El argumento ancho debe ser un número positivo.

En la siguiente imagen se muestra cómo la función **DESREF** resuelve la misma situación.

	A	B	C	D	E
1	Productos	TRI 1	TRI 2	TRI 3	TRI 4
2	Ordenadores	309	212	216	201
3	Periféricos	159	411	373	175
4	Consumibles	262	385	8	215
5	Almacenamiento	53	412	19	6
6					
7				373	
8					

3. PROMEDIOA

La función **PROMEDIO**, resulta normalmente apropiada para el cálculo de la media aritmética de un conjunto de valores, pero no lo es así en el caso de que una celda tenga un valor lógico o de texto. Por ejemplo, en un rango de 10 celdas hay una que contiene texto, la función **PROMEDIO** realizará la media dividiendo por 9 y obviarán la celda de texto. En cambio, la función **PROMEDIOA**, realizará una media dividiendo por 10, esto es así porque a la celda que contiene texto la interpreta con el valor 0.

La función **PROMEDIOA** se utiliza en casos de este tipo. **PROMEDIOA** opera de modo muy similar a **PROMEDIO**, con la excepción de que esta función asigna un valor 0 a las celdas no numéricas.

Cuando esté calculando el promedio de celdas, tenga en cuenta la diferencia existente entre las celdas vacías y las que contienen el valor cero. Las celdas vacías no se tienen en cuenta pero sí los valores cero.

4. BDPROMEDIO

Excel también permite calcular la media de los elementos seleccionados de una lista mediante la función **BDPROMEDIO**.

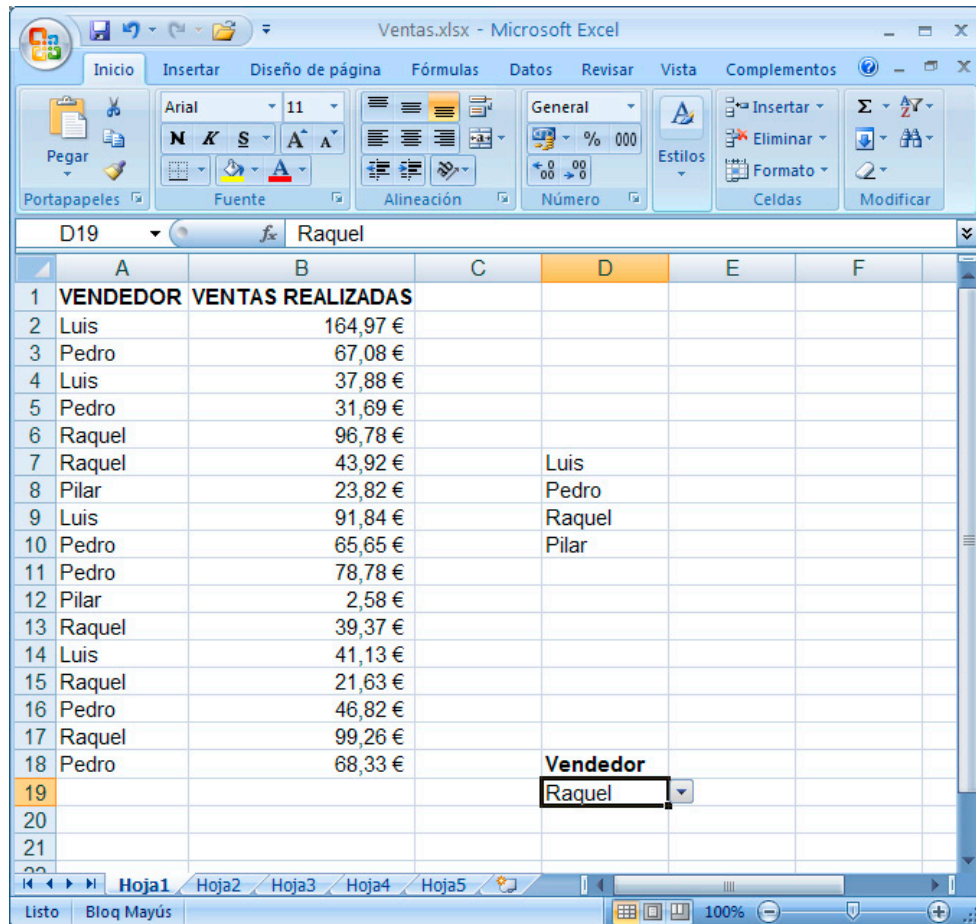
Su sintaxis es la siguiente:

BDPROMEDIO(base_de_datos;nombre_de_campo;criterios)

- **Base_de_datos:** es el rango de celdas que compone la lista o base de datos. Una base de datos es una lista de datos relacionados en la que las filas de información son registros y las columnas de datos, campos. La primera fila de la lista contiene los r tulos de cada columna.
- **Nombre_de_campo:** indica qu  columna se utiliza en la funci n. Escriba el r tulo de la columna entre comillas, como por ejemplo "Edad" o "Rendimiento", o un n mero (sin las comillas) que represente la posici n de la columna en la lista: 1 para la primera columna, 2 para la segunda y as  sucesivamente.
- **Criterios:** es el rango de celdas que contiene las condiciones especificadas. Puede utilizar cualquier rango en el argumento Criterios mientras  ste incluya por lo menos un r tulo de columna y al menos una celda debajo del r tulo de columna en la que se pueda especificar una condici n de columna.

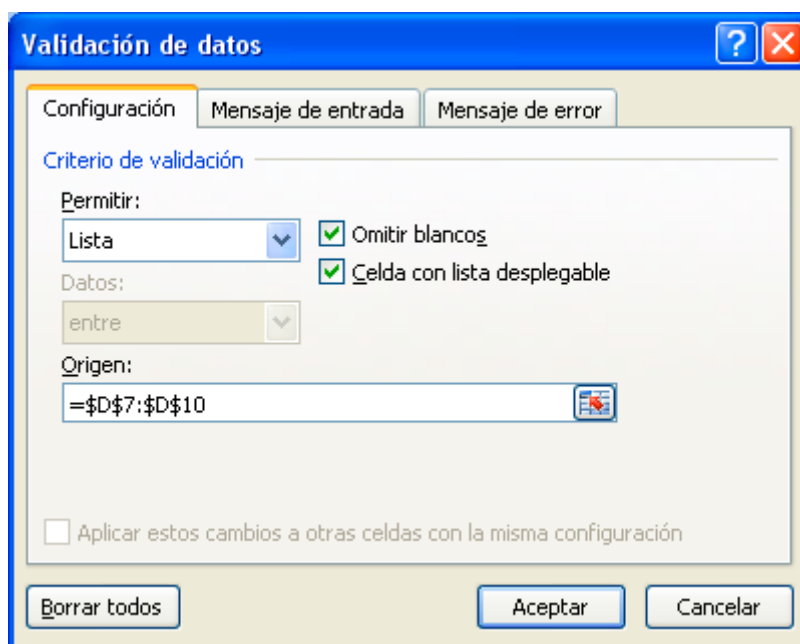
Veremos el funcionamiento de esta funci n con un ejemplo:

El rango A1:A18 contiene los porcentajes de ventas y el vendedor que las ha realizado. Esta Hoja de c lculo, permite al usuario calcular la media del vendedor seleccionado.



En primer lugar, debemos proporcionar un método para seleccionar un vendedor. Esto lo conseguimos introduciendo los nombres de los vendedores en el rango de celdas D7:D10 y añadiendo una lista desplegable.

Para crear la lista desplegable con los vendedores, nos situaremos en la ficha **Datos** y seleccionaremos la opción **Validación de datos**. Nos aparecerá un cuadro de diálogo como el que se muestra en la imagen.



En la sección **Permitir**, seleccionamos **Lista** y en el origen seleccionamos el rango donde se encuentran los vendedores. De esta forma, tendremos una celda en la que podremos seleccionar el vendedor que queramos para operar con él.

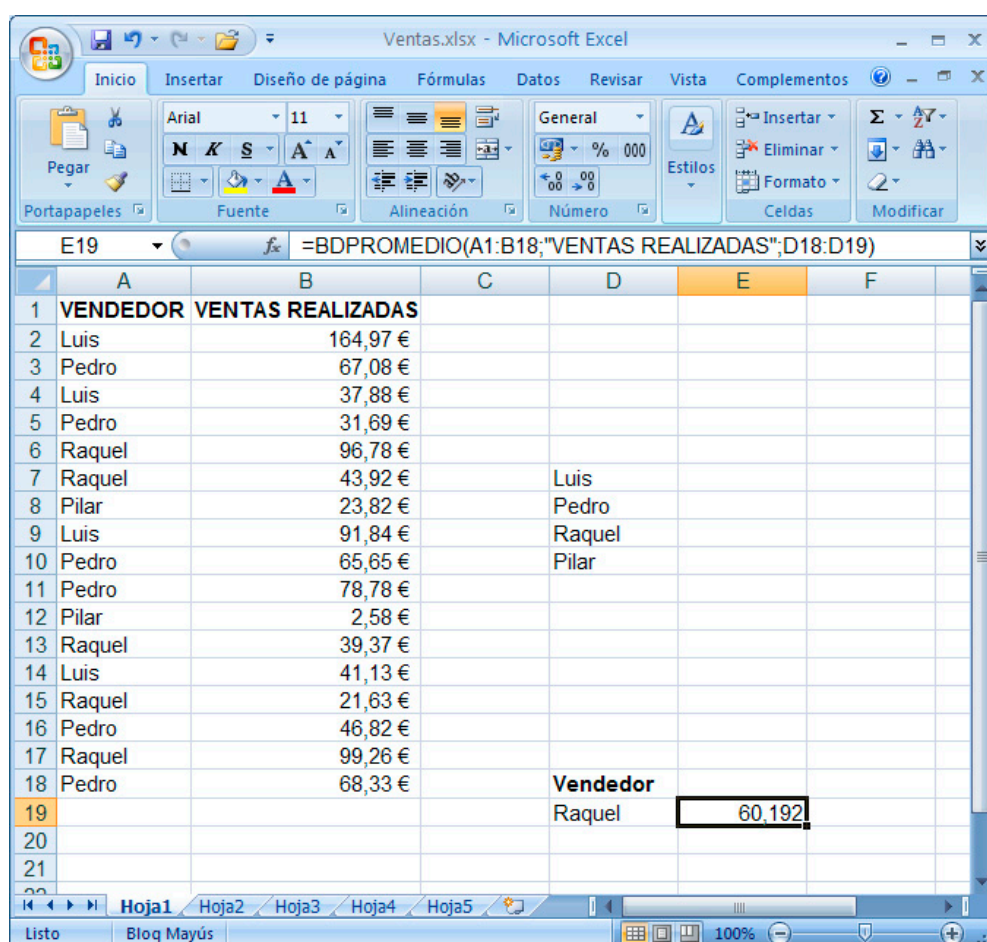
A continuación, podemos utilizar la función **BDPROMEDIO**. Por ejemplo, podemos introducirla en la celda **E19** y presenta la siguiente sintaxis:

= BDPROMEDIO(A1:B18;B1;D18:D19)

La fórmula incluye tres parámetros. El primero de ellos es el rango **A1:B18**, que contiene la totalidad de los datos, incluido el encabezado de columna. El segundo parámetro es **B1**. Éste es el encabezado de la columna que contiene los valores para los que queremos calcular la media. La fórmula podría haberse introducido como:

=BDPROMEDIO(A1:B18;"VENTAS REALIZADAS";D18:D19)

El tercer parámetro es el rango de criterios. Consta de un encabezado y de una muestra lógica de datos. El encabezado le indica a la función, la columna a la que debe aplicarse el criterio especificado. BDPROMEDIO calcula el promedio de las filas para las que la condición es verdadera.



Si seleccionamos a otro vendedor de la lista, la función BDPROMEDIO se recalculará automáticamente para ofrecernos el valor correcto. De este modo, tenemos una hoja interactiva sin necesidad de utilizar ninguna macro.

BDPROMEDIO es una función de base de datos. Las funciones de bases de datos permiten efectuar cálculos frecuentes en los datos de las hojas de cálculo. No obstante, es preciso recordar que estos datos deben configurarse correctamente para que *Excel* pueda operar con ellos. Deberá especificarse un encabezado y un rango de criterios. Esto no presenta el menor inconveniente si se ha diseñado una solución *Excel* a partir de cero; sin embargo, puede dar lugar a complicaciones cuando se añaden cálculos a una hoja existente.

Excel cuenta con un gran número de funciones y la mayoría de las veces existe más de un método para obtener el resultado deseado.

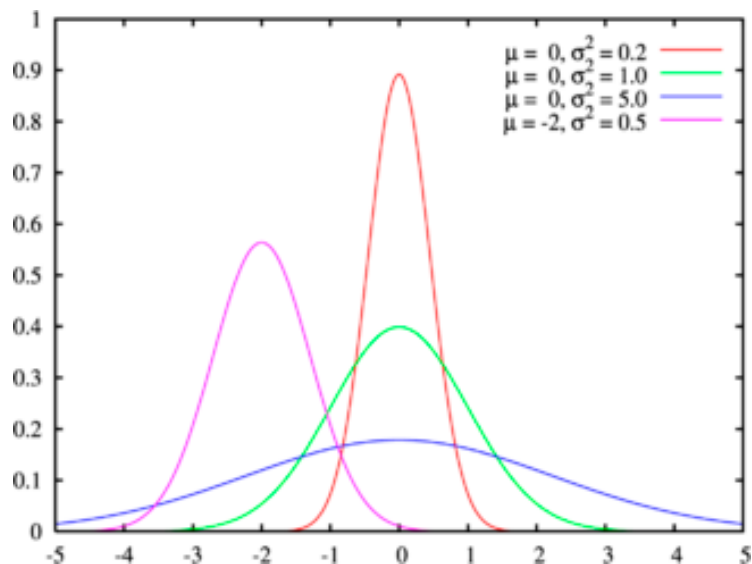
5. DISTRIBUCIONES

El promedio de llamadas diarias recibidas en una central telefónica es 12.534; sin embargo, no es frecuente que en un día se reciban exactamente 12.534 llamadas. Para poder gestionar este proceso, necesitamos conocer la medida de dispersión de los datos y el modo en que estos datos están distribuidos.

Excel nos proporciona funciones para operar con los tipos de distribuciones que se dan con más frecuencia en el mundo de la empresa.

DISTRIBUCIONES NORMALES

Una distribución normal presenta la forma de una curva de campana. Los valores en este tipo de distribución se sitúan simétricamente alrededor de la media. Las colas de distribución incluyen valores excepcionalmente altos o bajos. Ésta es la situación más común en negocios.



La medida más sencilla de la variación de una distribución normal es la varianza. Para calcular la varianza, se estima la desviación de la media de cada uno de los valores y se utiliza el cuadrado de la desviación de la media. Luego se suman estos cuadrados y se dividen entre el número de valores menos 1. En Excel, estas operaciones se realizan mediante la función **VAR**. La sintaxis es la siguiente:

VAR(número1;número2; ...)

Observe el siguiente ejemplo.

	L	M	N	O	P	Q	R	S
1			45,57	0,0025	←			=(N1-PROMEDIO(N\$1:N\$20))^2
2			45,74	0,0144				
3			46,28	0,4356				
4			45,56	0,0036				
5			45,96	0,1156				
6			45,72	0,01		0,33166316	←	=SUMA(O1:O20)/19
7			45,59	0,0009				
8			45,75	0,0169		0,33166316	←	=VAR(N1:N20)
9			45,53	0,0081				
10			45,06	0,3136				
11			44,98	0,4096				
12			44,74	0,7744				
13			44,71	0,8281				
14			45,16	0,2116				
15			45,5	0,0144				
16			45,25	0,1369				
17			45,5	0,0144				
18			46,48	0,7396				
19			46,45	0,6889				
20			46,87	1,5625				
21								

Queremos calcular la varianza de los 20 números de la columna N. En la celda O1, escribimos la siguiente fórmula:

=(N1- PROMEDIO (N\$1:N\$20))^2

Esta fórmula calcula la desviación de la media de los valores contenidos en la celda N1 y la media de los valores de las 20 celdas.

La celda O6 contiene la siguiente fórmula:

=SUMA(O1:O20)/19

La celda Q1 contiene el valor de la media y la celda Q2 el valor de la desviación estándar. La fórmula que utilizaremos es la siguiente:

=DISTR.NORM(46,75;Q1;Q2;VERDADERO)

El primer argumento, es el valor con el que queremos trabajar (46,75). A continuación, indicamos los valores de la media y de la desviación estándar. Finalmente, utilizamos el argumento acum. VERDADERO que nos devuelve la probabilidad acumulada de 46,75. Si especificáramos FALSO como argumento, la función devolvería el valor máximo de la curva de distribución para 46,75.

La función **DISTR.NORM**, devuelve el valor de 0,9751. Esto quiere decir que el 97,51% de los valores de distribución se encuentran por debajo de 46,75.

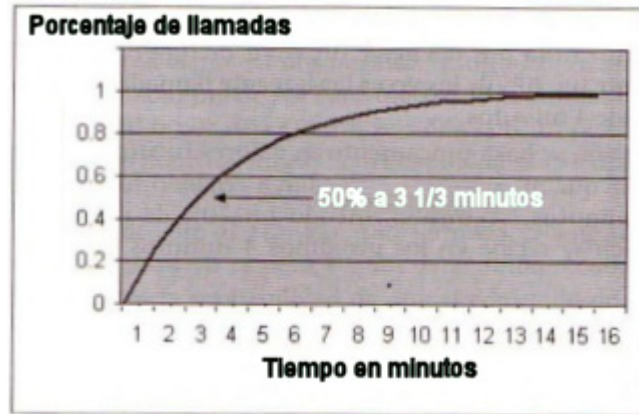
La función **DISTR.NORM.INV** opera en sentido inverso. Supongamos que queremos saber qué valores se encuentran por encima del 25% de la distribución. El valor que buscamos es el que se encuentra en el punto 75%. La siguiente fórmula nos permitirá obtener la respuesta:

=DISTR.NORM.INV(0,75;Q1;Q2)

El primer argumento es el porcentaje con el que queremos trabajar. A continuación, indicamos los valores de la media y de la desviación estándar. La función no contiene los argumentos VERDADERO/ FALSO. La fórmula devuelve un valor de 46. Los valores 46 y todos los valores superiores a éste se encuentran por encima del 25% de la distribución.

DISTRIBUCIONES EXPONENCIALES

Nuestra central de llamadas recibe tres llamadas cada diez minutos. El tiempo transcurrida entre llamadas, presenta la distribución que se muestra en la siguiente imagen.



La curva puede iniciarse a partir de cualquier unidad de tiempo. En el momento preciso en que ésta se origina, aún no se ha recibido ninguna llamada, por lo que existe un porcentaje cero de llamadas en el intervalo de tiempo cero. Un 50% de las veces la siguiente llamada, se recibirá transcurrido un intervalo de tiempo de 3 1/3 minutos. La siguiente llamada, se recibe casi siempre a los 15 minutos. Un tiempo de espera de 16 minutos es posible pero casi nunca sucede.

En este caso, el valor que nos interesa es el intervalo de tiempo transcurrido entre llamadas, de lo que obtenemos como resultado una distribución exponencial. Excel nos permite calcular probabilidades en esta distribución mediante la función **DISTR.EXP**. La sintaxis es la siguiente:

DISTR.EXP(x;lambd;a;acum)

Antes de poder utilizar esta función, debemos calcular el valor del parámetro **Lambda**. El parámetro Lambda, es el valor inverso del intervalo medio. Así pues, si nuestra central de llamadas recibe 3 llamadas cada 10 minutos, el intervalo medio entre llamadas será 3,33... minutos. El parámetro Lambda, se calcula dividiendo 1 entre 3,33... lo que equivale a 0,3. Hemos utilizado los minutos como unidad de tiempo para determinar el parámetro lambda. Por consiguiente, el resultado que obtendremos al utilizar la función **DISTR. EXP** se expresará en minutos.

Supongamos que queremos saber en qué porcentaje de tiempo se recibirá la siguiente llamada dentro de un intervalo de 4 minutos. Para ello, introduciremos la siguiente fórmula:

=DISTR.EXP(4;0,3;VERDADERO)

El primer argumento es 4, ya que éste es el intervalo en minutos con el que deseamos trabajar. A continuación, especificamos el parámetro lambda (0,3). El argumento **VERDADERO**, indica a la función que queremos obtener la probabilidad acumulada. La fórmula nos devuelve un valor de 0,698. Esto quiere decir que aproximadamente un 70% de las veces, la siguiente llamada se recibirá transcurrido un intervalo de 4 minutos.

Esta distribución se basa únicamente en valores futuros; carece de memoria. Esto quiere decir, que aun cuando no se haya recibido ninguna llamada en los últimos cinco minutos, seguirá existiendo un 70% de probabilidad de que la siguiente llamada se reciba en los próximos 4 minutos. El pasado no cuenta, sólo el futuro.

DISTRIBUCIÓN GAMMA

La distribución gamma es similar a la exponencial, salvo que es más general. Esta distribución permite calcular la probabilidad de múltiples eventos.

Obtendremos el mismo resultado que con DISTR.EXP utilizando esta fórmula:

=DISTR. GAMMA (4;1;3,3333; VERDADERO)

Existen dos diferencias respecto a los argumentos especificados en DISTR. EXP. En este caso el parámetro Lambda es 0,3 pero la función DISTR. GAMMA requiere el valor 1/Lambda. El argumento es 3,3333. El valor 1 es el número de coincidencias con el que queremos trabajar. En este caso, estamos interesados en la siguiente llamada, de manera que únicamente queremos calcular la probabilidad de una llamada. Si quisiéramos calcular la probabilidad de recibir tres llamadas en los cuatro minutos siguientes, tendríamos que introducir una fórmula como ésta:

=DISTR.GAMMA(4;3;3,3333;VERDADERO)

Esta fórmula devuelve un valor de 0,1205. Existe un 12% de probabilidad de recibir tres llamadas en los cuatro minutos siguientes.

La función DISTR.GAMMA. INV opera en sentido inverso. Si buscamos una probabilidad del 90% de que la siguiente llamada se reciba en un determinado número de minutos, la fórmula que tendremos que introducir es la siguiente:

=DISTR.GAMMA.INV(0,9;1;3,333)

El valor del primer argumento (0, 9) es la probabilidad que queremos evaluar, el número de coincidencias es 1, y 1 dividido entre el parámetro Lambda es 3,333. El resultado es 7,67. Esto quiere decir que existe un 90% de probabilidad de que la siguiente llamada se reciba en los 7,67 minutos siguientes.

DISTRIBUCIÓN BINOMIAL

Algunas cosas sólo pueden tener dos posibles resultados. Al lanzar una moneda al aire, el resultado sólo puede ser cara o cruz; una llamada reclamando un cobro puede tener éxito o no; una cuenta se salda con un resultado positivo o negativo. Para este tipo de situaciones, la probabilidad se calcula utilizando una distribución binomial. La sintaxis es la siguiente:

DISTR.BINOM(núm_éxito;ensayos;prob_éxito;acumulado)

Si la probabilidad de que al realizar una llamada reclamando un pago se obtenga un resultado positivo es de 0,1 y un determinado cobrador realiza 75 llamadas al día, ¿cuál es la probabilidad de que el resultado de las llamadas dé exactamente 10 resultados positivos?

Excel proporciona una función que permite contestar a esta pregunta. La fórmula es la siguiente:

DISTR.BINOM (10;75;0,1;FALSO)

El primer argumento (10) es el número de llamadas que vamos a evaluar. Se han realizado 75 llamadas y la probabilidad de éxito de cada llamada es 0,1. El argumento VERDADERO le indica a la función que queremos calcular la probabilidad.

La fórmula devuelve un valor de 0,873. Existe un 8,73% de probabilidad de obtener exactamente 10 pagos realizando 75 llamadas.