

ANÁLISIS DE CORRELACIÓN EMPLEANDO EXCEL Y GRAPH

Cuando se estudian en forma conjunta dos características (variables estadísticas) de una población o muestra, se dice que estamos analizando una variable estadística bidimensional. La correlación es el grado de relación que existe entre ambas características, y la regresión es la forma de expresar matemáticamente dicha relación.

Dado dos variables, la correlación permite hacer estimaciones del valor de una de ellas conociendo el valor de la otra variable.

1) DIAGRAMA DE DISPERSIÓN

Los diagramas de dispersión son planos cartesianos en los que se marcan los puntos correspondientes a los pares ordenados (X,Y) de los valores de las variables.

2) CLASIFICACIÓN DE LA CORRELACIÓN

2.1) Según la relación entre variables

- **Correlación lineal:** Se representa mediante una línea recta.
- **Correlación no lineal:** Se representa con una línea curva.

2.2) Según el número de variables

- **Correlación simple:** La variable dependiente actúa sobre la variable independiente.
- **Correlación múltiple:** Cuando la variable dependiente actúa sobre varias variables independientes.
- **Correlación parcial:** Cuando la relación que existe entre una variable dependiente y una independiente es de tal forma que los demás factores permanezcan constantes.

2.3) Según el valor cuantitativo

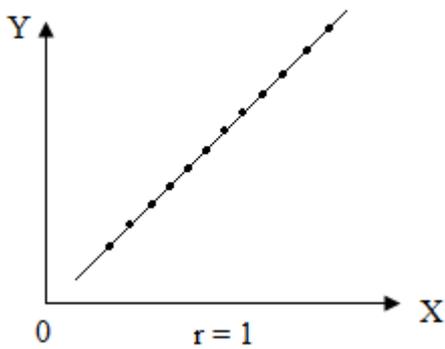
- **Correlación perfecta:** El valor del coeficiente de correlación es 1
- **Correlación imperfecta:** El coeficiente de correlación es menor a 1 sea en sentido positivo o negativo.
- **Correlación nula:** El coeficiente de correlación es 0. No existe correlación entre las variables. Ejemplo: Número de calzado de una persona y su cociente intelectual.

2.4) Según el signo

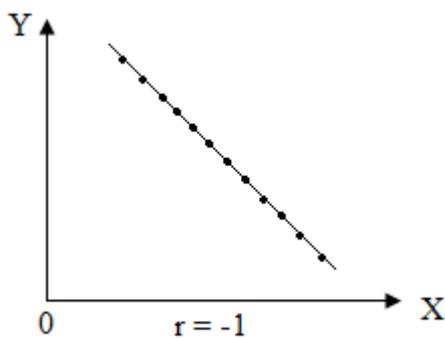
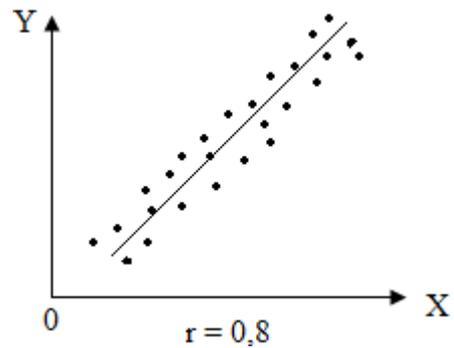
- **Correlación positiva.-** Dos variables tiene correlación positiva cuando al aumentar o disminuir el valor de una de ellas entonces el valor correspondiente a la otra aumentará o disminuirá respectivamente, es decir, cuando las dos variables aumentan en el mismo sentido. Ejemplo: Peso de una persona y su talla.
- **Correlación negativa.-** Dos variables tiene correlación negativa cuando al aumentar o disminuir el valor de una de ellas entonces el valor de la otra disminuirá o aumentará respectivamente, es decir, una variable aumenta y otra disminuye o viceversa. Ejemplo: Número de partidos ganados por un equipo en una temporada y su posición final en la tabla.

3) COEFICIENTES DE CORRELACIÓN

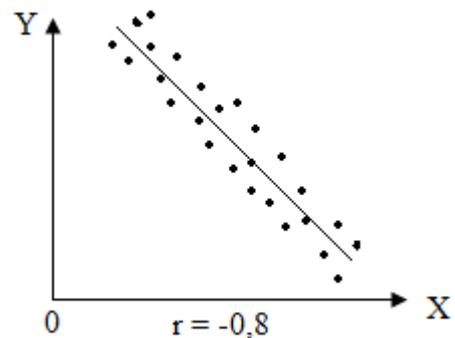
Los coeficientes de correlación son medidas que indican la situación relativa de los mismos sucesos respecto a las dos variables, es decir, son la expresión numérica que nos indica el grado de relación existente entre las 2 variables y en qué medida se relacionan. Son números que varían entre los límites +1 y -1. Su magnitud indica el grado de asociación entre las variables; el valor $r = 0$ indica que no existe relación entre las variables; los valores ± 1 son indicadores de una correlación perfecta positiva (al crecer o decrecer X, crece o decrece Y) o negativa (Al crecer o decrecer X, decrece o crece Y).



Correlación Positiva



Correlación Negativa



Para interpretar el coeficiente de correlación utilizamos la siguiente escala:

Valor	Significado
-1	Correlación negativa grande y perfecta
-0,9 a -0,99	Correlación negativa muy alta
-0,7 a -0,89	Correlación negativa alta
-0,4 a -0,69	Correlación negativa moderada
-0,2 a -0,39	Correlación negativa baja
-0,01 a -0,19	Correlación negativa muy baja
0	Correlación nula
0,01 a 0,19	Correlación positiva muy baja
0,2 a 0,39	Correlación positiva baja
0,4 a 0,69	Correlación positiva moderada
0,7 a 0,89	Correlación positiva alta
0,9 a 0,99	Correlación positiva muy alta
1	Correlación positiva grande y perfecta

3.1) COEFICIENTE DE CORRELACIÓN DE KARL PEARSON

Llamando también coeficiente de correlación producto-momento.

a) Para datos no agrupados se calcula aplicando la siguiente ecuación:

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

r = Coeficiente producto-momento de correlación lineal

$$x = X - \bar{X}; y = Y - \bar{Y}$$

Ejemplo ilustrativo:

Con los datos sobre las temperaturas en dos días diferentes en una ciudad, determinar el tipo de correlación que existe entre ellas mediante el coeficiente de PEARSON.

X	18	17	15	16	14	12	9	15	16	14	16	18	$\Sigma X = 180$
Y	13	15	14	13	9	10	8	13	12	13	10	8	$\Sigma Y = 138$

Solución:

Se calcula la media aritmética

$$\bar{x} = \frac{\sum x_i}{n}$$

Para X:

$$\bar{X}_X = \frac{180}{12} = 15$$

Para Y:

$$\bar{Y}_Y = \frac{138}{12} = 11,5$$

Se llena la siguiente tabla:

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	xy	y^2
18	13	3	1,5	9	4,5	2,25
17	15	2	3,5	4	7	12,25
15	14	0	2,5	0	0	6,25
16	13	1	1,5	1	1,5	2,25
14	9	-1	-2,5	1	2,5	6,25
12	10	-3	-1,5	9	4,5	2,25
9	8	-6	-3,5	36	21	12,25
15	13	0	1,5	0	0	2,25
16	12	1	0,5	1	0,5	0,25
14	13	-1	1,5	1	-1,5	2,25
16	10	1	-1,5	1	-1,5	2,25
18	8	3	-3,5	9	-10,5	12,25
180	138			72	28	63

Se aplica la fórmula:

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{28}{\sqrt{(72)(63)}} = 0,416$$

Existe una correlación moderada

En Excel se calcula de la siguiente manera:

a) Se inserta la función COEF.DE.CORREL y pulsar en Aceptar.

The image shows an Excel spreadsheet with the following data in columns A and B:

A	B
X	Y
18	13
17	15
15	14
16	13
14	9
12	10
9	8
15	13
16	12
14	13
16	10
18	8

The 'Insertar función' dialog box is open, showing the 'Estadísticas' category selected. The function 'COEF.DE.CORREL' is highlighted in the list. The description below the list reads: 'COEF.DE.CORREL(matriz1;matriz2) Devuelve el coeficiente de correlación de dos conjuntos de datos.'

b) En el cuadro de argumentos de la función, en el recuadro de la Matriz 1 seleccionar las celdas de X, y en el recuadro de la Matriz 2 seleccionar las celdas de Y.

The screenshot shows an Excel spreadsheet with columns A and B. Column A contains values: X, 18, 17, 15, 16, 14, 12, 9, 15, 16, 14, 16, 18. Column B contains values: Y, 13, 15, 14, 13, 9, 10, 8, 13, 12, 13, 10, 8. A dialog box titled 'Argumentos de función' is open, showing the function 'COEF.DE.CORREL'. The 'Matriz1' argument is 'A2:A13' and the 'Matriz2' argument is 'B2:B13'. The dialog box displays the result of the formula as 0,41573971. The formula bar at the bottom of the spreadsheet shows '=COEF.DE.CORREL(A2:A13;B2:B13)'. The spreadsheet also shows a formula in cell C15: '=COEF.DE.CORREL(A2:A13;B2:B13)'.

c) Pulsar en Aceptar.

	A	B	C	D	E
1	X	Y			
2	18	13			
3	17	15			
4	15	14			
5	16	13			
6	14	9			
7	12	10			
8	9	8			
9	15	13			
10	16	12			
11	14	13			
12	16	10			
13	18	8			
14					
15		0,41573971	=COEF.DE.CORREL(A2:A13;B2:B13)		

El diagrama de dispersión en Excel se realiza de la siguiente manera:

a) Seleccionar los datos e insertar diagrama de dispersión.

Correlación - Microsoft Excel

Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista Complementos

Tabla dinámica Tabla Imagen Imágenes prediseñadas Formas SmartArt Columna Línea Circular Barra Área Dispersión Otros gráficos Hipervínculo Cuadro Encabezado de texto pie de página

Tablas Ilustraciones Gráficos

A2 fx 18

	A	B	C	D	E	F	G	J	K
1	X	Y							
2	18	13							
3	17	15							
4	15	14							
5	16	13							
6	14	9							
7	12	10							
8	9	8							
9	15	13							
10	16	12							
11	14	13							
12	16	10							
13	18	8							

Dispersión

Dispersión sólo con marcadores

Compara pares de valores.

Utilícelo cuando los valores no están en el orden del eje X o cuando representan medidas separadas.

b) En diagrama dispersión, escoger el primero.

Cambiar tipo de gráfico Guardar como plantilla Tipo

Cambiar entre filas y columnas Seleccionar datos Datos

Diseños de gráfico

3 Gráfico fx

	A	B	C	D	E	F	G	H	I
1	X	Y							
2	18	13							
3	17	15							
4	15	14							
5	16	13							
6	14	9							
7	12	10							
8	9	8							
9	15	13							
10	16	12							
11	14	13							
12	16	10							
13	18	8							
14									
15									

16

14

12

10

8

6

4

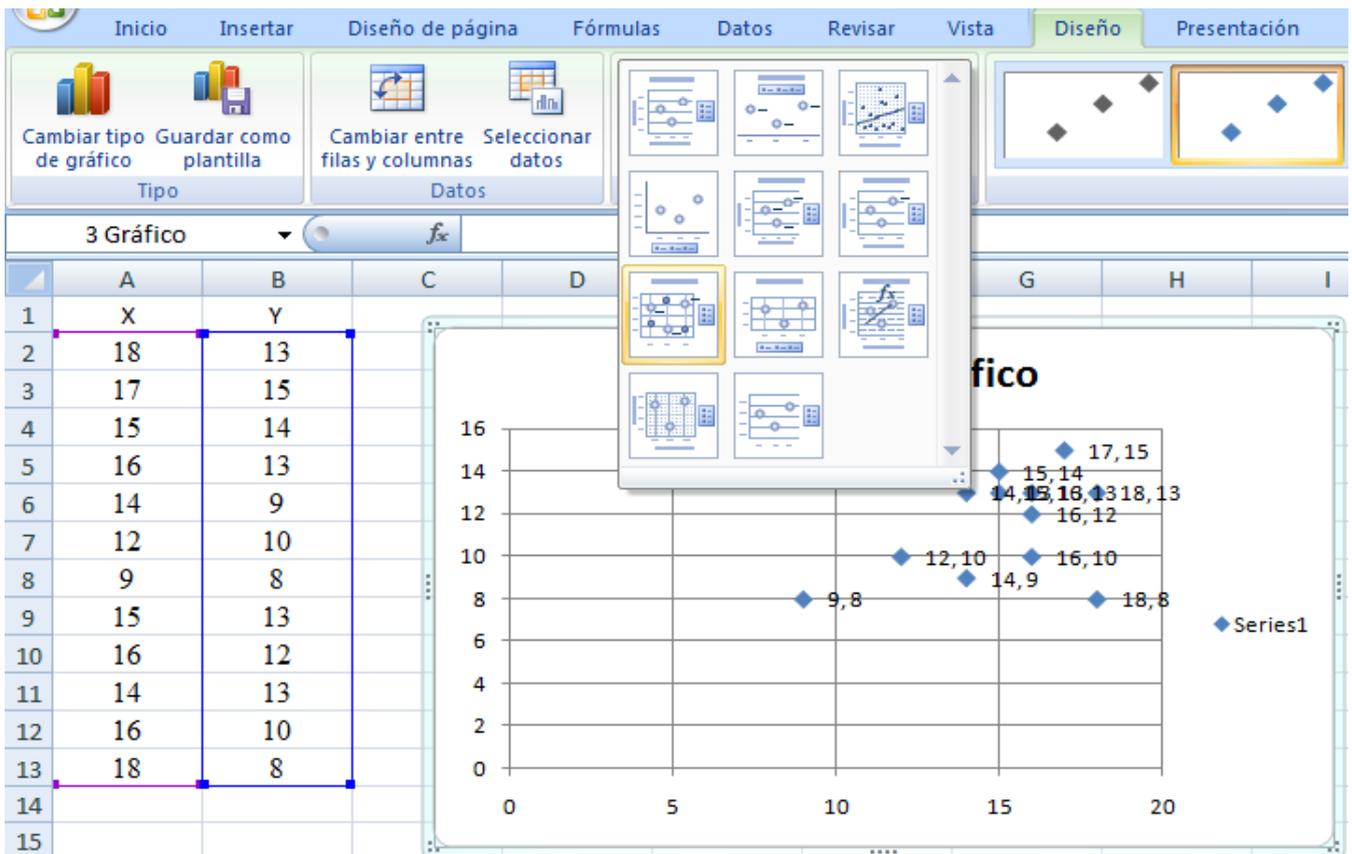
2

0

0 5 10 15 20

◆ Series1

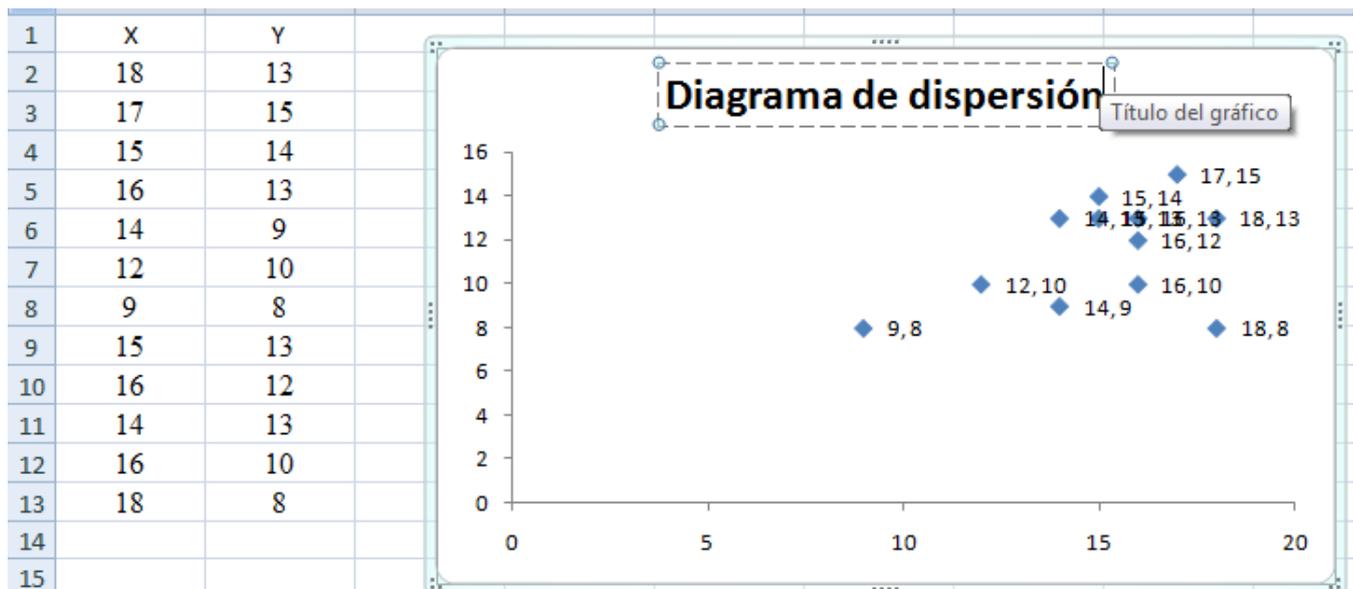
c) Para que ver las coordenadas escoger el diseño N° 7.



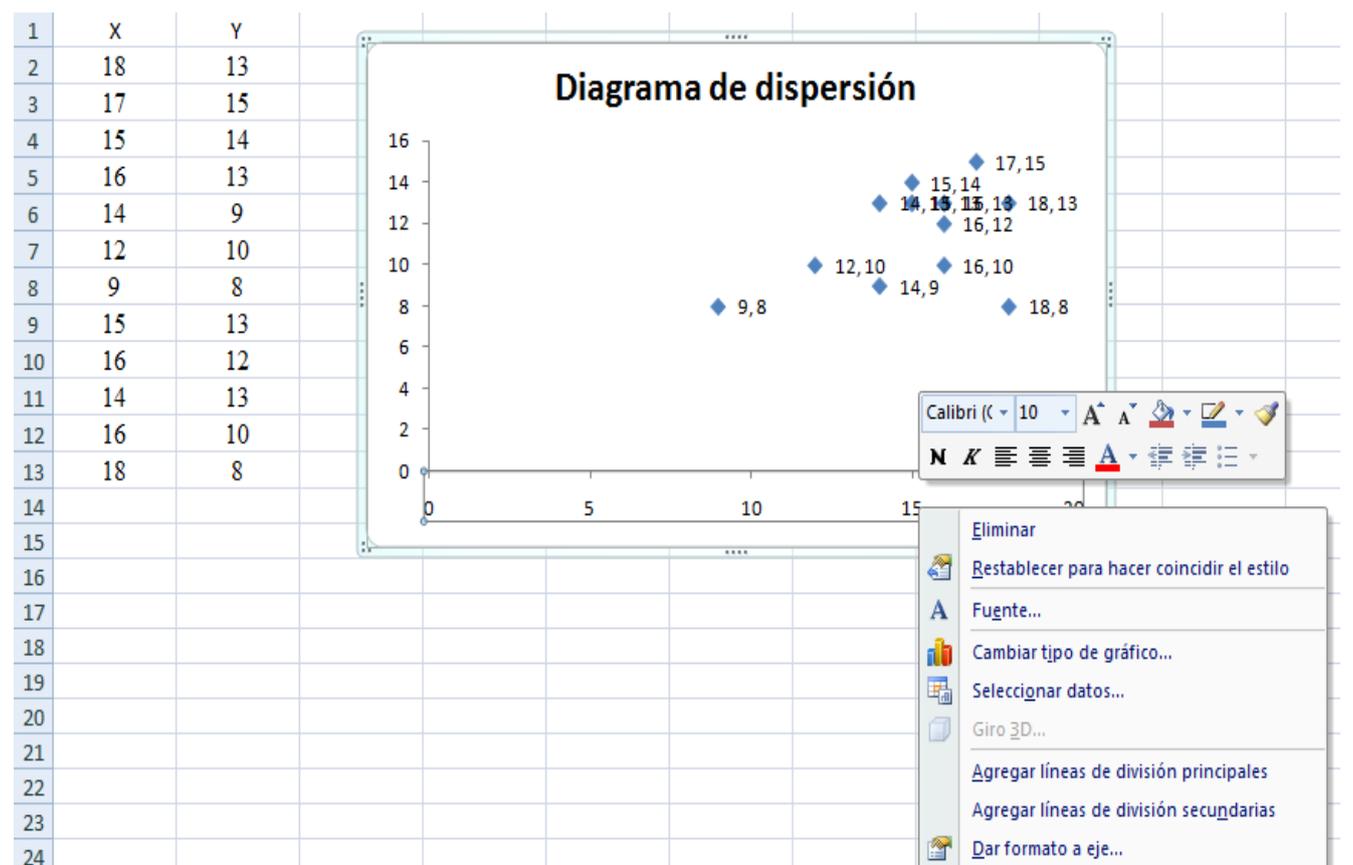
d) Borrar Serie 1, las líneas horizontales y verticales (haciendo clic y suprimir en cada objeto).



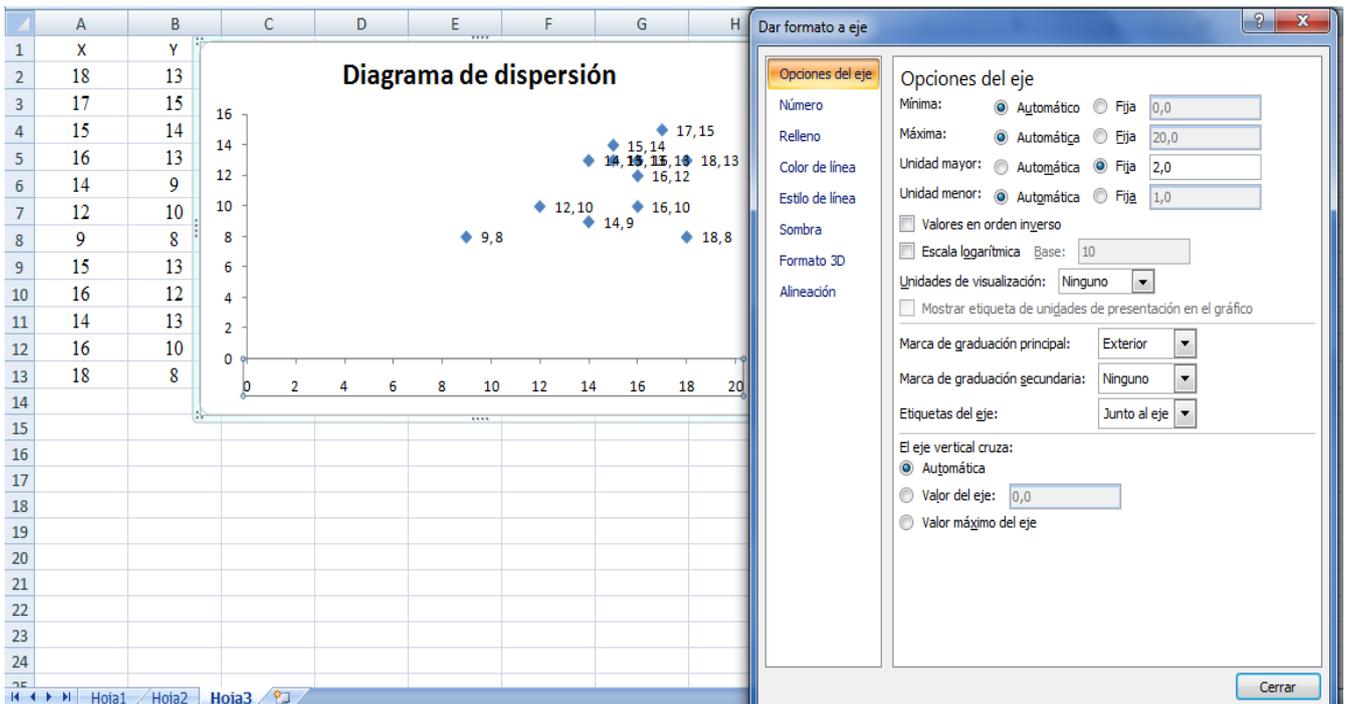
e) En título del gráfico escribir Diagrama de dispersión.



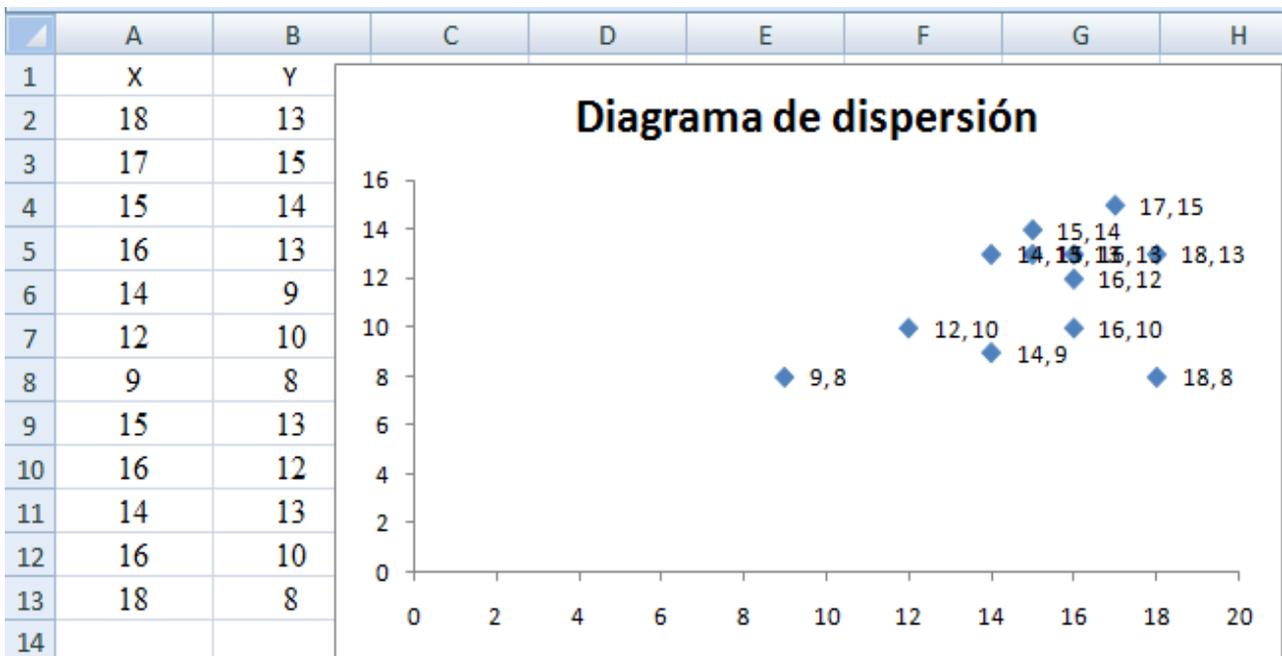
f) Clic en el eje x, y luego clic derecho para dar formato al eje.



g) Poner 2 en la casilla unidad mayor para ver los números de 2 en 2 en el eje x.

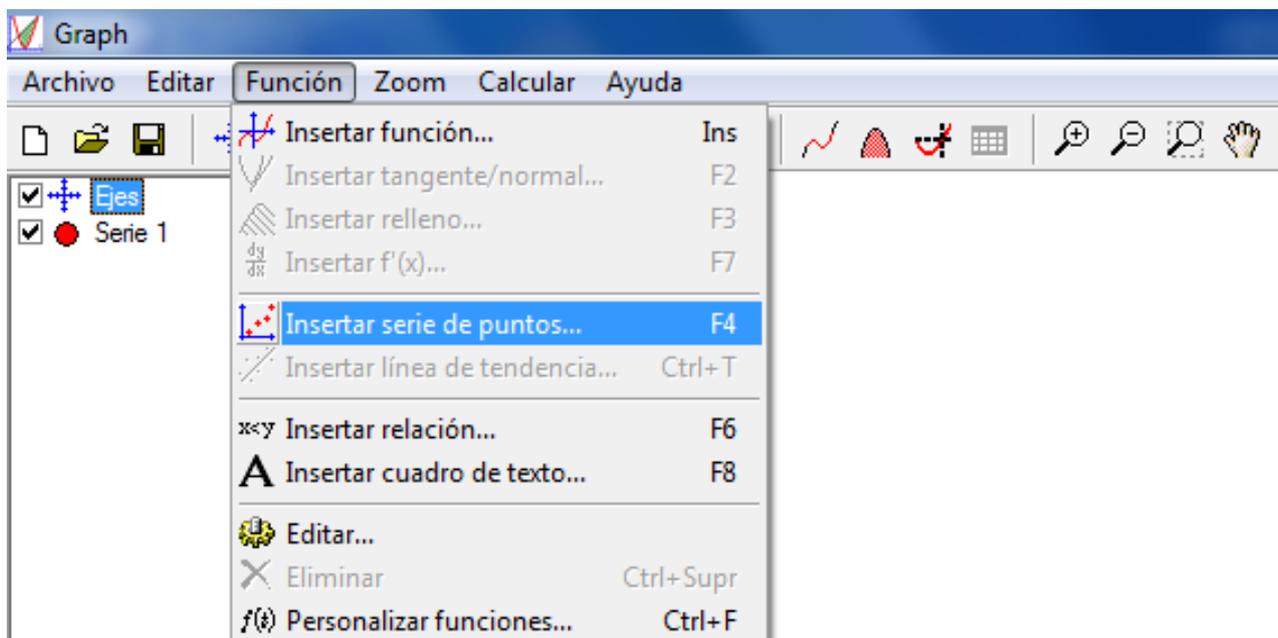


h) Clic en Cerrar para culminar la elaboración del diagrama de dispersión, aunque se le puede seguir haciendo más mejoras.

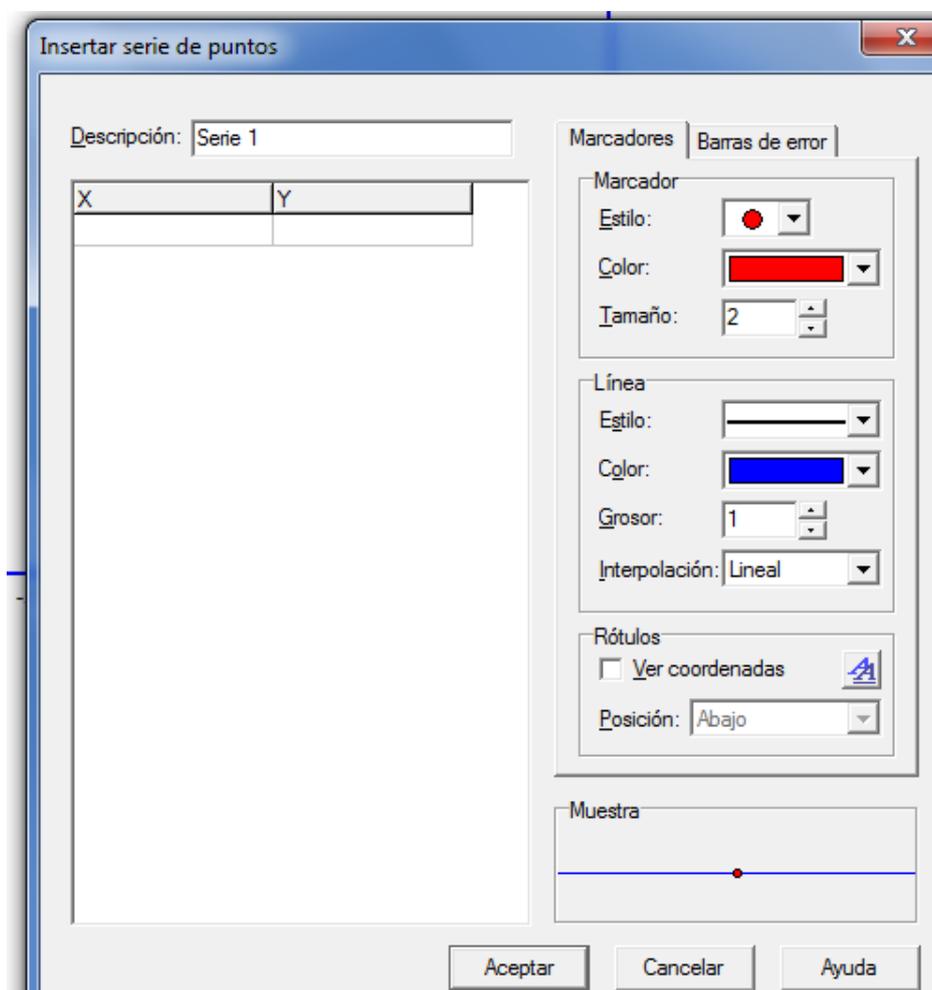


Para realizar el diagrama de dispersión en el programa Graph se procede de la siguiente manera:

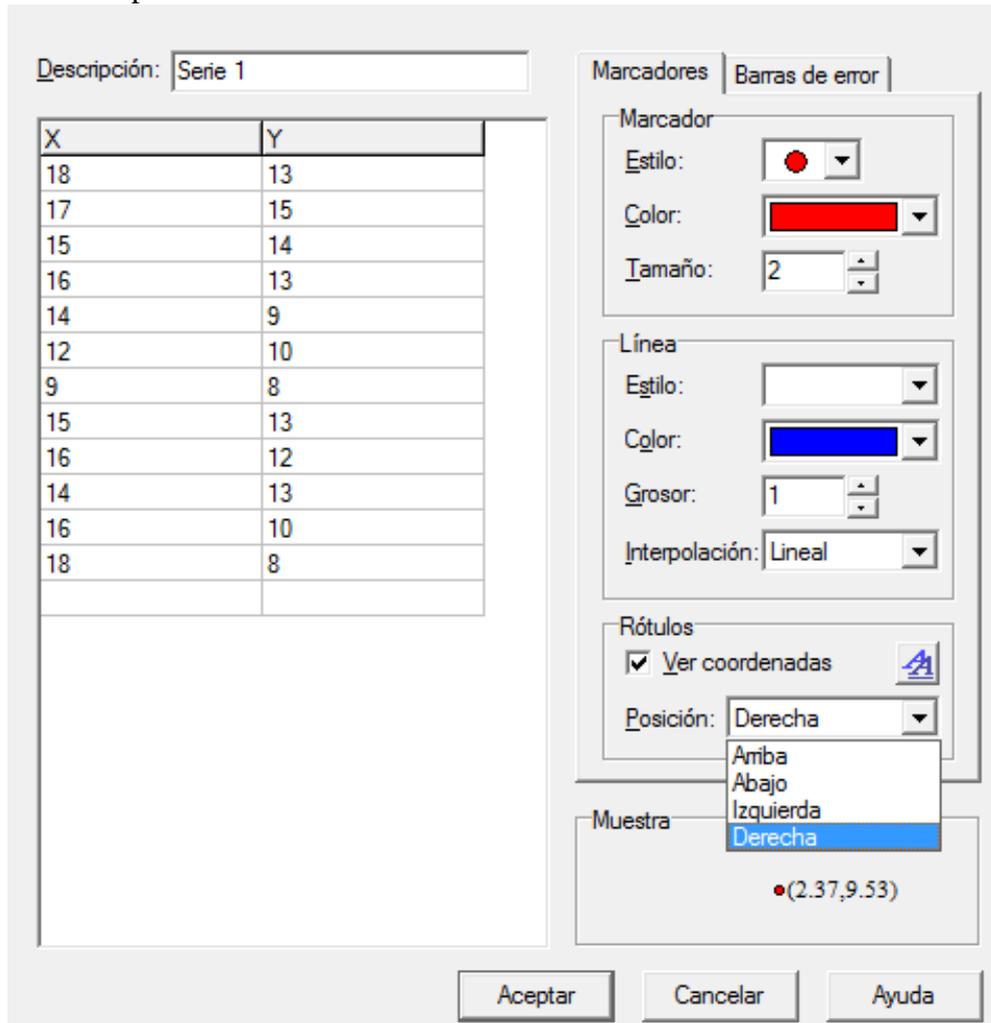
a) Clic en Función.



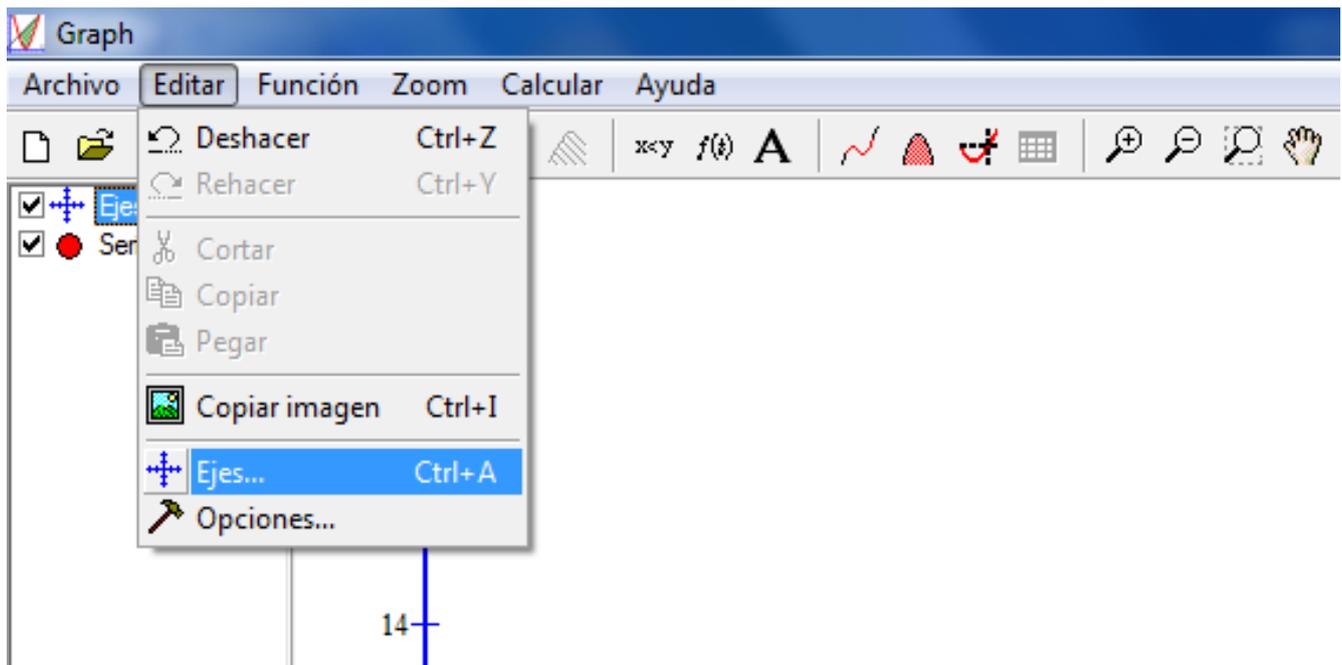
b) Clic en Insertar serie de puntos.



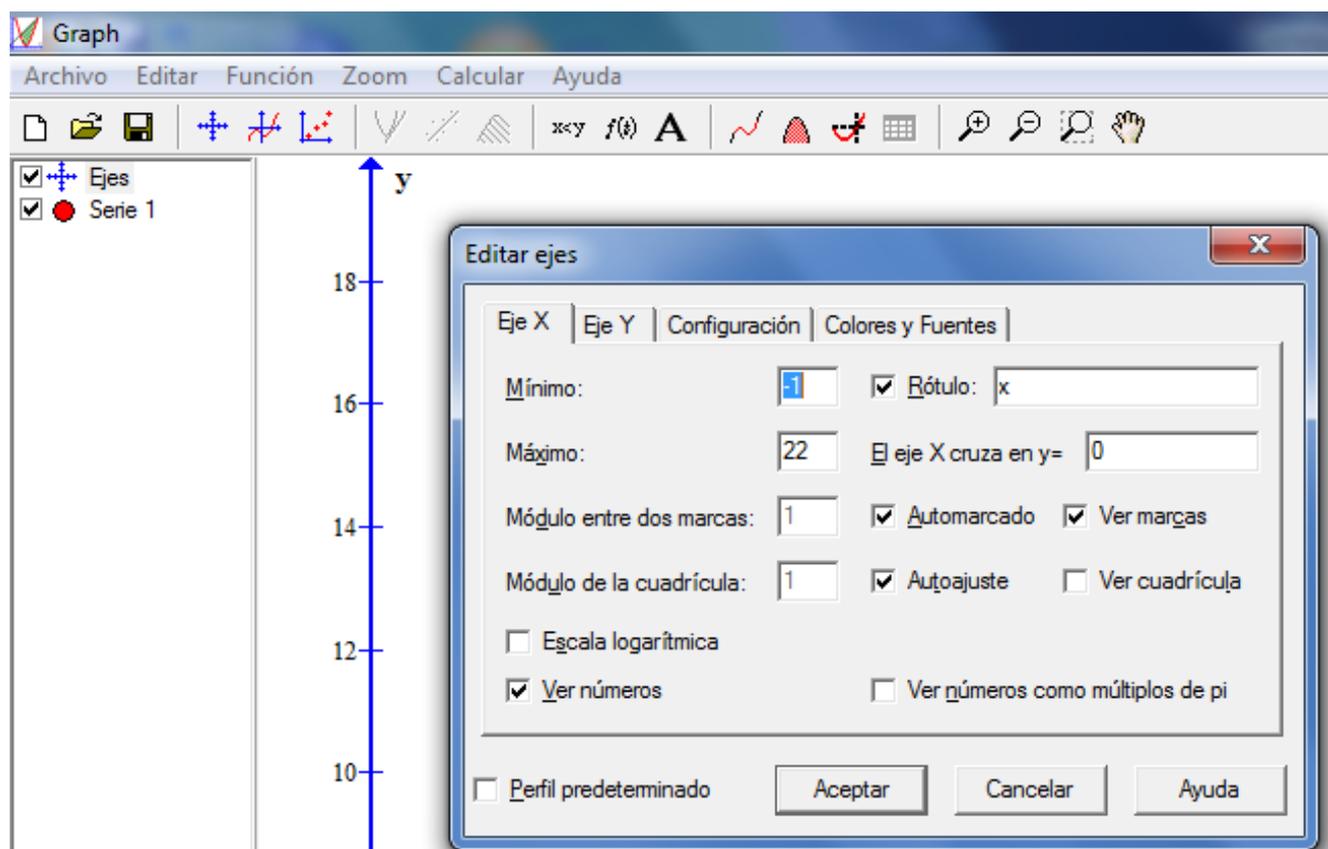
c) Escribir los puntos, y en estilo de línea, escoger sin línea. En rótulos poner en ver coordenadas a la derecha. Pulsar en Aceptar.



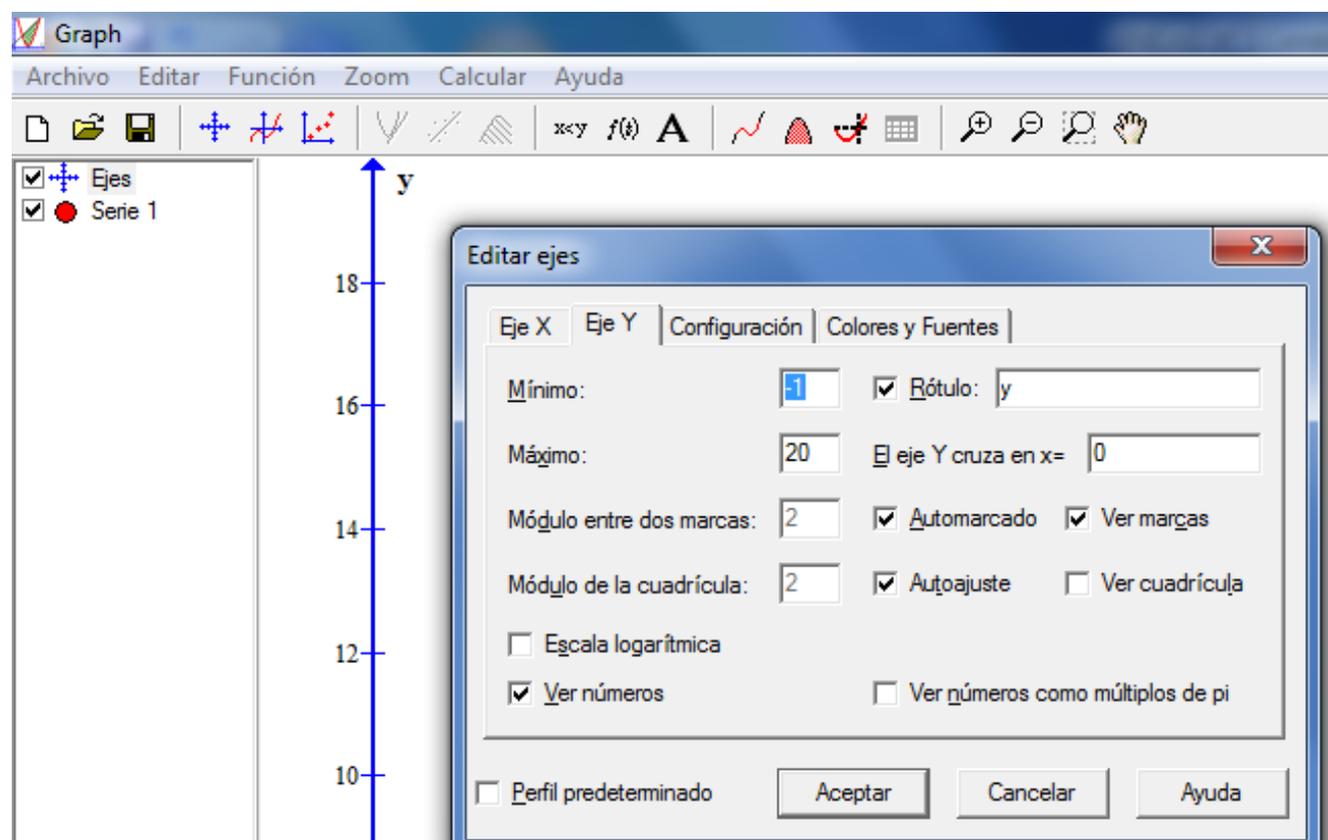
d) Para editar los ejes, hacer clic en Editar y luego en Ejes.



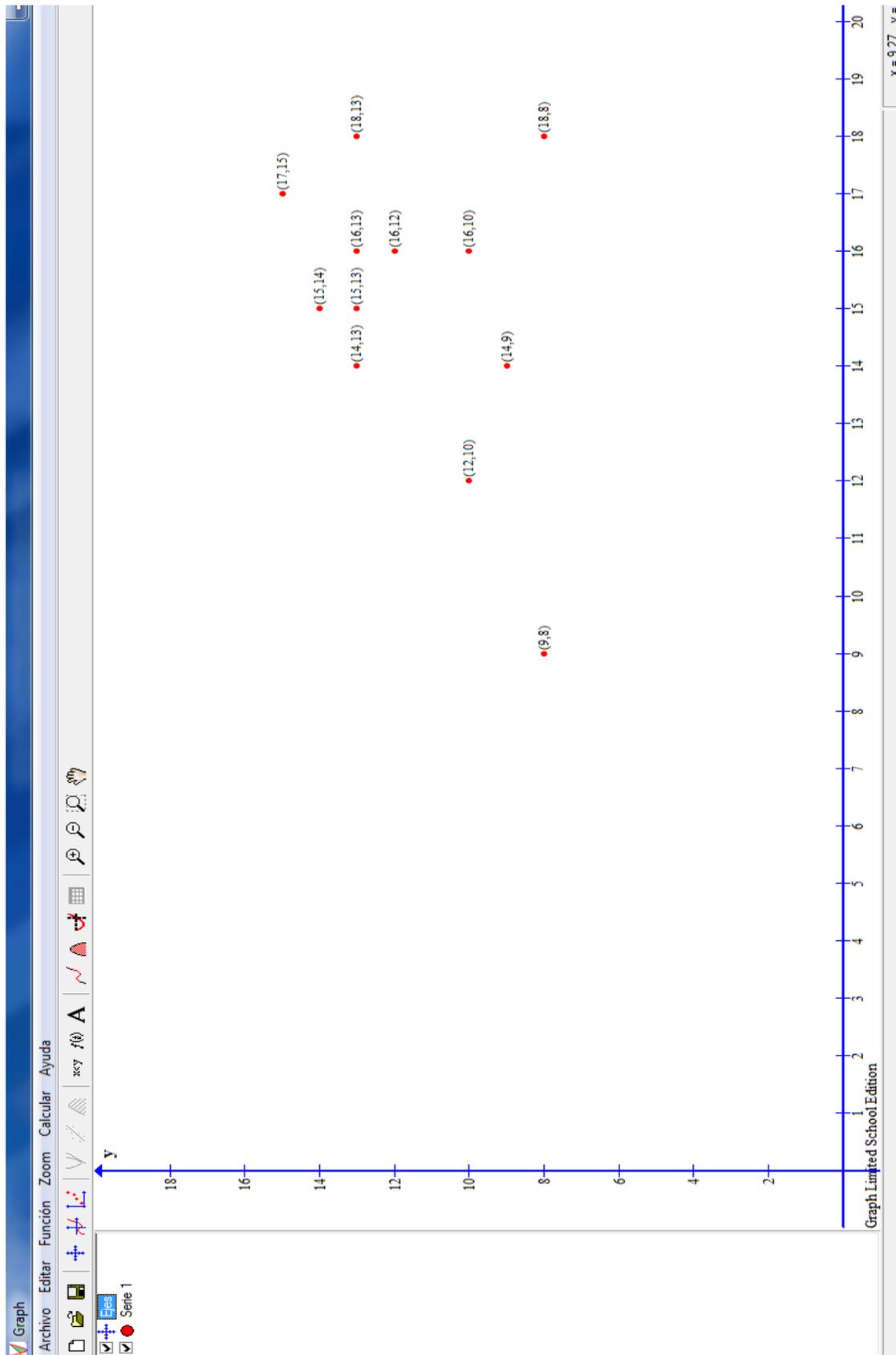
e) Llenar las casillas del Eje X de acuerdo a los datos del ejercicio.



f) Llenar las casillas del Eje Y de acuerdo a los datos del ejercicio.



g) Pulsar en Aceptar para dar por culminado la elaboración del diagrama de dispersión, el cual se presenta en la siguiente figura:



b) Para datos agrupados, el coeficiente de Correlación de Pearson se calcula aplicando la siguiente fórmula:

$$r = \frac{n \cdot \sum f \cdot dx \cdot dy - (\sum fx \cdot dx) (\sum fy \cdot dy)}{\sqrt{[n \cdot \sum fx \cdot dx^2 - (\sum fx \cdot dx)^2][n \cdot \sum fy \cdot dy^2 - (\sum fy \cdot dy)^2]}}$$

Donde:

n = número de datos.

f = frecuencia de celda.

fx = frecuencia de la variable X.

fy = frecuencia de la variable Y.

dx = valores codificados o cambiados para los intervalos de la variable X, procurando que al intervalo central le corresponda $dx = 0$, para que se hagan más fáciles los cálculos.

dy = valores codificados o cambiados para los intervalos de la variable X, procurando que al intervalo central le corresponda $dy = 0$, para que se hagan más fáciles los cálculos.

Ejemplo ilustrativo:

Con los siguientes datos sobre los Coeficientes Intellectuales (X) y de las calificaciones en una prueba de conocimiento (Y) de 50 estudiantes:

N° de estudiante	X	Y	N° de estudiante	X	Y
1	76	28	26	88	40
2	77	24	27	88	31
3	78	18	28	88	35
4	79	41	29	88	26
5	79	43	30	89	30
6	80	45	31	89	24
7	80	34	32	90	18
8	80	18	33	90	11
9	82	40	34	90	15
10	82	35	35	91	38
11	83	30	36	92	34
12	83	21	37	92	31
13	83	22	38	93	33
14	83	23	39	93	35
15	84	25	40	93	24
16	84	11	41	94	40
17	84	15	42	96	35
18	85	31	43	97	36
19	85	35	44	98	40
20	86	26	45	99	33
21	86	30	46	100	51
22	86	24	47	101	54
23	86	16	48	101	55
24	87	20	49	102	41
25	88	36	50	102	45

- 1) Elaborar una tabla de dos variables
- 2) Calcular el coeficiente de correlación

Solución:

1) En la *tabla de frecuencias de dos variables*, cada recuadro de esta tabla se llama una *celda* y corresponde a un par de intervalos, y el número indicado en cada celda se llama *frecuencia de celda*. Todos los totales indicados en la última fila y en la última columna se llaman *totales marginales o frecuencias marginales*, y corresponden, respectivamente, a las frecuencias de intervalo de las distribuciones de frecuencia separadas de la variable X y Y.

Para elaborar la tabla se recomienda:

- Agrupar las variables X y Y en un igual número de intervalos.
- Los intervalos de la variable X se ubican en la parte superior de manera horizontal (fila) y en orden ascendente.
- Los intervalos de la variable Y se ubican en la parte izquierda de manera vertical (columna) y en orden descendente.

Para elaborar los intervalos se procede a realizar los cálculos respectivos:

En la variable X:

Calculando el Rango se obtiene:

$$R = x_{máx} - x_{mín} = 102 - 76 = 26$$

Calculando el número de intervalos se obtiene:

$$n_i = 1 + 3,32 \cdot \log(n) = 1 + 3,32 \cdot \log 50 = 6$$

Calculando el ancho se obtiene:

$$i = \frac{R}{n_i} = \frac{26}{6} = 4,33$$

En la variable Y:

Calculando el Rango se obtiene:

$$R = y_{máx} - y_{mín} = 55 - 11 = 44$$

Calculando el número de intervalos se obtiene:

$$n_i = 1 + 3,32 \cdot \log(n) = 1 + 3,32 \cdot \log 50 = 6$$

Calculando el ancho se obtiene:

$$i = \frac{R}{n_i} = \frac{44}{6} = 7,33$$

Nota: Para la variable X se tomará un ancho de intervalo igual a 5 y para la variable Y un ancho de intervalo igual a 8 para obtener un número de intervalos igual a 6 para cada variable.

Contando las frecuencias de celda para cada par de intervalos de las variables X y Y se obtiene la siguiente tabla de frecuencias de dos variables:

		Coeficientes Intellectuales (X)					<i>f_y</i>	
		76-80	81-85	86-90	91-95	96-100		101-105
Calificaciones (Y)	51-58					1	2	3
	43-50	2					1	3
	35-42	1	3	3	3	3	1	14
	27-34	2	2	3	3	1		11
	19-26	1	4	5	1			11
	11-18	2	2	4				8
	<i>f_x</i>	8	11	15	7	5	4	50

Interpretación:

- El número 5 es la frecuencia de la celda correspondiente al par de intervalos 86-90 en Coeficiente Intellectual y 19-26 en Calificación obtenida en la prueba de conocimiento.
- El número 8 en la fila de *f_x* es el total marginal o frecuencia marginal del intervalo 76-80 en Coeficiente Intellectual.
- El número 14 en la columna de *f_y* es el total marginal o frecuencia marginal del intervalo 35-42 en Calificación obtenida en la prueba de conocimiento.
- El número 50 es total de frecuencias marginales y representa al número total de estudiantes.

2) Realizando los cálculos respectivos se obtiene la siguiente tabla:

		X									
		76-80	81-85	86-90	91-95	96-100	101-105	<i>f_y</i>	<i>f_y · d_y</i>	<i>f_y · d_y²</i>	<i>f · d_x · d_y</i>
Y	<i>d_x</i> <i>d_y</i>	-2	-1	0	1	2	3				
51-58	3					1 6	2 18	3	9	27	24
43-50	2	2 -8					1 6	3	6	12	-2
35-42	1	1 -2	3 -3	3 0	3 3	3 6	1 3	14	14	14	7
27-34	0	2 0	2 0	3 0	3 0	1 0		11	0	0	0
19-26	-1	1 2	4 4	5 0	1 -1			11	-11	11	5
11-18	-2	2 8	2 4	4 0				8	-16	32	12
<i>f_x</i>		8	11	15	7	5	4	50	2	96	46
<i>f_x · d_x</i>		-16	-11	0	7	10	12	2			
<i>f_x · d_x²</i>		32	11	0	7	20	36	106			
<i>f · d_x · d_y</i>		0	5	0	2	12	27	46			

Nota:

Los números de las esquinas de cada celda en la anterior tabla representan el producto $f \cdot dx \cdot dy$, así por ejemplo, para obtener el número -8 de los intervalos 76-80 en X y 43-50 en Y se obtiene multiplicando $2 \cdot (-2) \cdot (2) = -8$. Para obtener el número 6 de los intervalos 96-100 en X y 51-58 en Y se obtiene multiplicando $1 \cdot 2 \cdot 3 = 6$.

Los números de la última columna (24, -2, 7, 0, 5 y 12) se obtienen sumando los números de las esquinas en cada fila, así por ejemplo, para obtener el número 24 se suma $6 + 18 = 24$.

Los números de la última fila (0, 5, 0, 2, 12 y 27) se obtienen sumando los números de las esquinas en cada columna, así por ejemplo, para obtener el número 27 se suma $18 + 6 + 3 = 27$.

Para obtener el número 2 de la antepenúltima columna se obtiene sumando los resultados de $fy \cdot dy$, es decir, representa la $\sum fy \cdot dy$

Para obtener el número 2 de la antepenúltima fila se obtiene sumando los resultados de $fx \cdot dx$, es decir, representa la $\sum fx \cdot dx$

Para obtener el número 96 de la penúltima columna se obtiene sumando los resultados de $fy \cdot dy^2$, es decir, representa $\sum fy \cdot dy^2$

Para obtener el número 106 de la penúltima fila se obtiene sumando los resultados de $fx \cdot dx^2$, es decir, representa $\sum fx \cdot dx^2$

Para obtener último número de la última columna se obtiene sumando los resultados de la última columna ($46=24-2+7+0+5+12$), es decir, representa $\sum f \cdot dx \cdot dy$.

Para obtener último número de la última fila se obtiene sumando los resultados de la última fila ($46=0+5+0+2+12+27$), y tiene que ser igual al último número de la última columna como comprobación que los cálculos de la tabla han sido correctos.

Observando los datos en la tabla anterior se reemplaza los valores en la ecuación del Coeficiente de Correlación de Pearson para datos agrupados, obteniéndose:

$$r = \frac{n \cdot \sum f \cdot dx \cdot dy - (\sum fx \cdot dx) (\sum fy \cdot dy)}{\sqrt{[n \cdot \sum fx \cdot dx^2 - (\sum fx \cdot dx)^2][n \cdot \sum fy \cdot dy^2 - (\sum fy \cdot dy)^2]}}$$
$$r = \frac{50 \cdot 46 - (2)(2)}{\sqrt{[50 \cdot 106 - (2)^2][50 \cdot 96 - (2)^2]}} = \frac{2300 - 4}{\sqrt{[5300 - 4][4800 - 4]}} = \frac{2296}{\sqrt{[5296][4796]}}$$
$$r = \frac{2296}{\sqrt{25399616}} = \frac{2296}{5039,803} = 0,456$$

Existe una correlación positiva moderada

TAREA DE INTERAPRENDIZAJE

1) Elabore un organizador gráfico de los tipos de correlación.

2) Con los datos de la siguiente tabla sobre las temperaturas del día X y del día Y en determinadas horas en una ciudad

X	9	10	12	14	16	18	20	22	24	26	28	30
Y	12	14	15	16	17	20	22	23	26	28	31	32

2.1) Calcule el coeficiente de correlación de Pearson empleando la fórmula y mediante Excel.

0,99

2.2) Elabore el diagrama de dispersión de manera manual.

2.3) Elabore el diagrama de dispersión empleando Excel.

2.4) Elabore el diagrama de dispersión empleando el programa Graph.

3) Cree y resuelva un ejercicio similar al anterior.

4) Dada la siguiente tabla de frecuencias de dos variables, con los datos sobre las calificaciones obtenidos en un curso de 50 estudiantes en la asignatura de Matemática (X) y en la asignatura de Estadística (Y), determinar el tipo de correlación que existe entre ellas mediante el coeficiente de Pearson.

		X					fy
		1-2	3-4	5-6	7-8	9-10	
Y	9-10				7	8	15
	7-8				6		6
	5-6			3	4		7
	3-4	5	5	1			11
	1-2	7	4				11
	fx	12	9	4	17	8	50

Correlación positiva muy alta de 0,91

5) Dada la siguiente tabla de frecuencias de dos variables, con los datos sobre los pesos en kilogramos en dos barrios diferentes en una ciudad, determinar el tipo de correlación que existe entre ellas mediante el coeficiente de Pearson.

		X					fy	
		40-49	50-59	60-69	70-79	80-89		90-99
Y	90-99				3	3	4	10
	80-89			8	2	2	4	16
	70-79			5	10	8	1	24
	60-69	8	1	2	5	2		18
	50-59	3	10	6	2			21
	40-49	4	6	1				11
	fx	15	17	22	22	15	9	100

Correlación positiva moderada de 0,688

6) Dada la siguiente tabla de frecuencias de dos variables, con los datos sobre las calificaciones obtenidos en un curso de 100 estudiantes en la asignatura de Matemática (X) y en la asignatura de

Estadística (Y), determinar el tipo de correlación que existe entre ellas mediante el coeficiente de Pearson para datos agrupados.

N° de estudiante	X	Y									
1	40	60	26	57	73	51	71	86	76	84	83
2	41	50	27	58	78	52	72	88	77	84	84
3	42	55	28	60	79	53	72	89	78	85	86
4	43	59	29	61	60	54	72	70	79	86	88
5	44	40	30	62	61	55	73	71	80	86	89
6	45	42	31	63	62	56	74	72	81	86	70
7	45	49	32	64	63	57	74	73	82	87	78
8	45	60	33	64	64	58	74	74	83	87	79
9	45	62	34	65	65	59	75	75	84	88	78
10	48	66	35	65	66	60	76	76	85	88	77
11	49	69	36	66	67	61	76	77	86	88	79
12	50	50	37	66	69	62	77	78	87	88	78
13	50	52	38	66	50	63	77	79	88	89	78
14	56	54	39	66	52	64	78	60	89	89	60
15	56	56	40	67	55	65	78	67	90	89	69
16	56	59	41	68	56	66	78	65	91	90	90
17	56	59	42	68	57	67	78	68	92	91	96
18	56	40	43	68	59	68	79	69	93	92	97
19	57	45	44	69	40	69	79	50	94	93	99
20	57	47	45	69	45	70	79	59	95	94	80
21	57	48	46	69	47	71	80	90	96	95	81
22	57	49	47	69	49	72	81	94	97	96	82
23	57	80	48	70	90	73	82	96	98	97	83
24	57	70	49	70	99	74	82	99	99	98	89
25	57	72	50	70	80	75	83	80	100	99	70

	X	40-48						94-102	
Y	$\begin{matrix} dx \\ dy \end{matrix}$	-3	-2	-1	0	1	1	3	f_y
94-102	3								7
	2								
	1								
	0								
	-1	5							
	-2	3							
40-48	-3	2							9
	f_x	10							100

Correlación positiva moderada de 0,62

7) Crear y resolver un ejercicio similar al anterior sobre cualquier tema de su preferencia.

3.2) COEFICIENTE DE CORRELACIÓN POR RANGOS DE SPEARMAN

Este coeficiente se emplea cuando una o ambas escalas de medidas de las variables son ordinales, es decir, cuando una o ambas escalas de medida son posiciones. Ejemplo: Orden de llegada en una carrera y peso de los atletas.

Se calcula aplicando la siguiente ecuación:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

r_s = Coeficiente de correlación por rangos de Spearman

d = Diferencia entre los rangos (X menos Y)

n = Número de datos

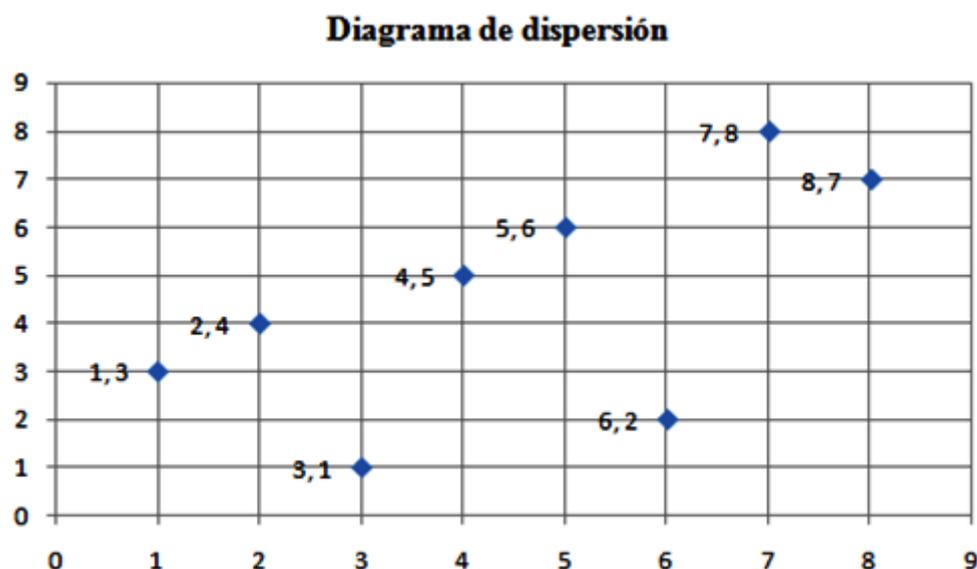
Nota: Los datos hay que traducirlos u ordenarlos en rangos. A los puntajes más elevados le asignamos el rango 1 al siguiente el rango 2 y así sucesivamente. Si se repiten dos puntajes o más se calculan las medias aritméticas.

Ejemplo ilustrativo N° 1: La siguiente tabla muestra el rango u orden obtenido en la primera evaluación (X) y el rango o puesto obtenido en la segunda evaluación (Y) de 8 estudiantes universitarios en la asignatura de Estadística. Realizar el diagrama de dispersión y calcular el coeficiente de correlación por rangos de Spearman.

Estudiante	X	Y
Dyanita	1	3
Elizabeth	2	4
Mario	3	1
Orlando	4	5
Mathías	5	6
Josué	6	2
Anita	7	8
Lucía	8	7

Solución:

El diagrama de dispersión hecho en Excel se muestra en la siguiente figura:



Para calcular el coeficiente de correlación por rangos de Spearman se llena la siguiente tabla:

Estudiante	X	Y	$d = X - Y$	$d^2 = (X - Y)^2$
Dyanita	1	3	-2	4
Elizabeth	2	4	-2	4
Mario	3	1	2	4
Orlando	4	5	-1	1
Mathías	5	6	-1	1
Josué	6	2	4	16
Anita	7	8	-1	1
Lucía	8	7	1	1
				$\Sigma d^2 = 32$

Se aplica la fórmula:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 32}{8(8^2 - 1)} = 1 - \frac{192}{504} = \frac{504 - 192}{504} = \frac{312}{504} = 0,619$$

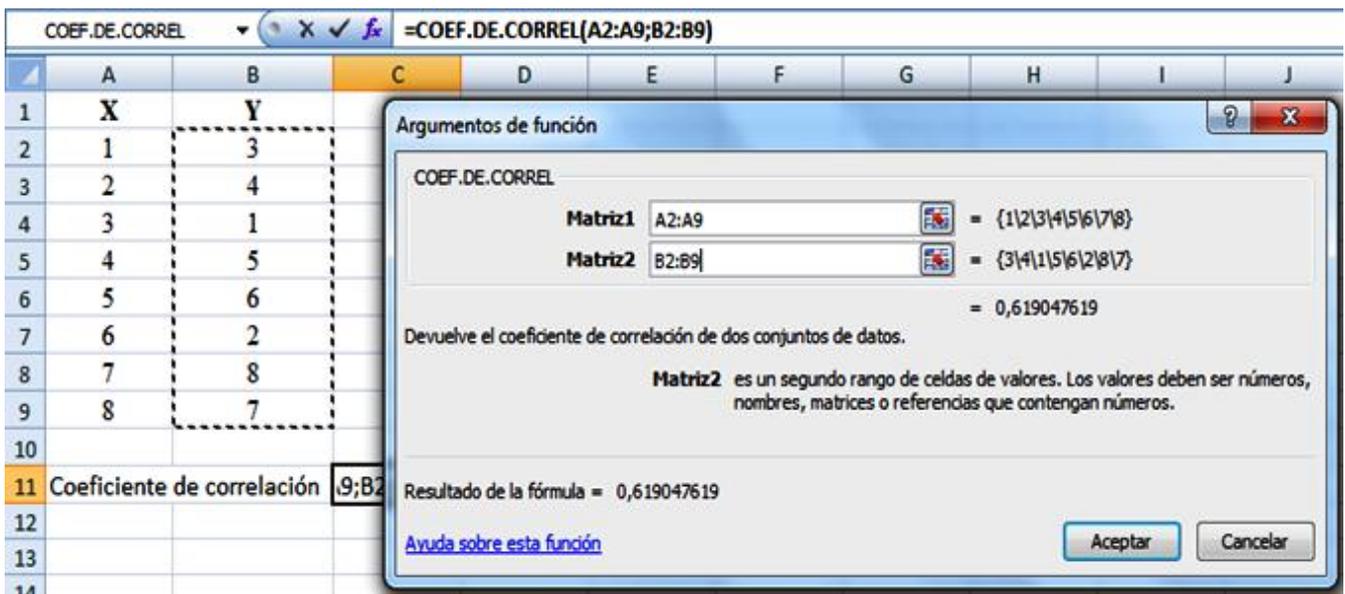
Por lo tanto existe una correlación positiva moderada entre la primera y segunda evaluación de los 8 estudiantes.

En Excel se calcula de la siguiente manera:

a) Se inserta la función COEF.DE.CORREL y pulsar en Aceptar.

The image shows an Excel spreadsheet with two columns, X and Y, containing the values 1 through 8. The formula bar shows the start of a formula: '='. Overlaid on the spreadsheet is the 'Insertar función' (Insert Function) dialog box. The 'Buscar una función:' (Search for a function) section has a search box and a category dropdown set to 'Estadísticas'. The 'Seleccionar una función:' (Select a function) list shows several functions, with 'COEF.DE.CORREL' highlighted. Below the list, the function name 'COEF.DE.CORREL(matriz1;matriz2)' is displayed, along with its description: 'Devuelve el coeficiente de correlación de dos conjuntos de datos.' (Returns the correlation coefficient of two sets of data.) At the bottom of the dialog are 'Ayuda sobre esta función' (Help on this function), 'Aceptar' (OK), and 'Cancelar' (Cancel) buttons.

b) En el cuadro de argumentos de la función, en el recuadro de la Matriz 1 seleccionar las celdas de X, y en el recuadro de la Matriz 2 seleccionar las celdas de Y.



c) Pulsar en Aceptar.

	A	B	C	D	E	F
1	X	Y				
2	1	3				
3	2	4				
4	3	1				
5	4	5				
6	5	6				
7	6	2				
8	7	8				
9	8	7				
10						
11	Coeficiente de correlación		0,6190476			=COEF.DE.CORREL(A2:A9;B2:B9)

Ejemplo ilustrativo N° 2

La siguiente tabla muestra las calificaciones de 8 estudiantes universitarios en las asignaturas de Matemática y Estadística. Calcular el coeficiente de correlación por rangos de Spearman y realizar el diagrama de dispersión.

N°	Estudiante	Matemática	Estadística
1	Dyana	10	8
2	Elizabeth	9	6
3	Mario	8	10
4	Orlando	7	9
5	Mathías	7	8
6	Josué	6	7
7	Anita	6	6
8	Lucía	4	9

Solución:

Para calcular el coeficiente de correlación por rangos de Spearman se procede a clasificar u ordenar los datos en rangos (X para Matemática y Y para Estadística) tomando en cuenta las siguientes observaciones:

En la asignatura de Matemática se observa:

- Dyana tiene la más alta calificación, ocupando el primer puesto, por lo que su rango es 1
- Elizabeth ocupa el segundo puesto, por lo que su rango es 2
- Mario se encuentra ubicado en el tercer lugar, por lo que su rango es 3
- Orlando y Mathías ocupan el cuarto y quinto puesto, por lo que su rango es la media aritmética de 4 y 5 que da por resultado 4,5
- Josué y Anita ocupan el sexto y séptimo lugar, por lo que su rango es la media aritmética de 6 y 7 que da por resultado 6,5
- Lucía se encuentra ubicada en el octavo lugar, por lo que su rango es 8

En la asignatura de Estadística se observa:

- Mario tiene la más alta calificación, ocupando el primer puesto, por lo que su rango es 1
- Orlando y Lucía ocupan el segundo y tercer puesto, por lo que su rango es la media aritmética de 2 y 3 que da por resultado 2,5
- Dyana y Mathías ocupan el cuarto y quinto puesto, por lo que su rango es la media aritmética de 4 y 5 que da por resultado 4,5
- Josué se encuentra ubicado en el sexto lugar, por lo que su rango es 6
- Elizabeth y Anita ocupan el séptimo y octavo lugar, por lo que su rango es la media aritmética de 7 y 8 que da por resultado 7,5

Los rangos X y Y se presentan en la siguiente tabla:

N°	Estudiante	Matemática	Estadística	X	Y
1	Dyana	10	8	1	4,5
2	Elizabeth	9	6	2	7,5
3	Mario	8	10	3	1
4	Orlando	7	9	4,5	2,5
5	Mathías	7	8	4,5	4,5
6	Josué	6	7	6,5	6
7	Anita	6	6	6,5	7,5
8	Lucía	4	9	8	2,5

Calculando d , d^2 y Σd^2 se obtiene los siguientes resultados:

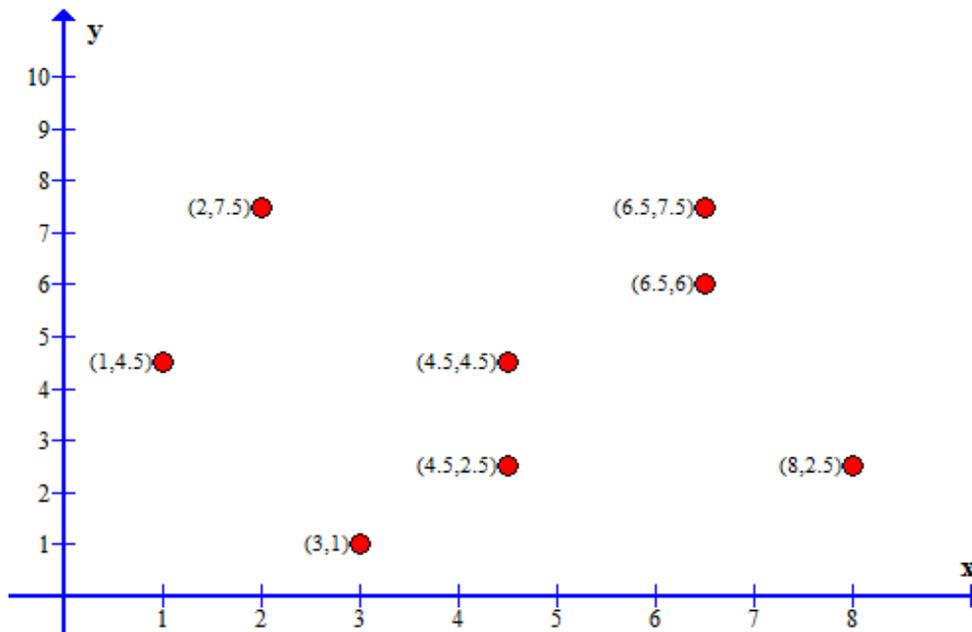
N°	Estudiante	Matemática	Estadística	X	Y	$d = X - Y$	$d^2 = (X - Y)^2$
1	Dyana	10	8	1	4,5	-3,5	12,25
2	Elizabeth	9	6	2	7,5	-5,5	30,25
3	Mario	8	10	3	1	2	4
4	Orlando	7	9	4,5	2,5	2	4
5	Mathías	7	8	4,5	4,5	0	0
6	Josué	6	7	6,5	6	0,5	0,25
7	Anita	6	6	6,5	7,5	-1	1
8	Lucía	4	9	8	2,5	5,5	30,25
							$\Sigma d^2 = 82$

Aplicando la fórmula se obtiene:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 82}{8(8^2 - 1)} = 1 - \frac{492}{504} = \frac{504 - 492}{504} = \frac{12}{504} = 0,024$$

Por lo tanto existe una correlación positiva muy baja

El diagrama de dispersión hecho en Graph se muestra en la siguiente figura:



TAREA DE INTERAPRENDIZAJE

- 1) Consulte sobre la biografía de Spearman y realice un organizador gráfico de la misma.
- 2) La siguiente tabla muestra el rango u orden obtenido en la primera evaluación (X) y el rango o puesto obtenido en la segunda evaluación (Y) de 8 estudiantes universitarios en la asignatura de Matemática.

X	Y
1	4
2	5
3	6
4	8
5	3
6	2
7	1
8	7

2.1) Realice el diagrama de dispersión en forma manual.

2.2) Realice el diagrama de dispersión empleando Excel.

2.3) Realice el diagrama de dispersión empleando el programa Graph.

2.4) Calcule el coeficiente de correlación por rangos de Spearman empleando la ecuación.

-0,19

2.5) Calcule el coeficiente de correlación empleando Excel.

-0,1905

3) Cree y resuelva un ejercicio similar al anterior.

4) La siguiente tabla muestra las calificaciones de 18 estudiantes universitarios en las asignaturas de Matemática y Estadística.

N°	Estudiante	Matemática	Estadística
1	Dyanita	10	3,5
2	Elizabeth	9	1
3	Mario	8	6
4	Orlando	8	8
5	Mathías	7	7
6	Benjamín	6	10
7	Segundo	6	4
8	Bertha	6	3,5
9	Alberto	5	1
10	Victoria	4	3
11	Sandra	4	9
12	Ximena	3	5
13	Darío	3	2,5
14	Santiago	2	0,7
15	José	1	2
16	Tomás	0,7	1,5
17	Paola	0,5	2,5
18	Kevin	0,5	0,5

4.1) Realice el diagrama de dispersión en forma manual.

4.2) Realice el diagrama de dispersión empleando Excel.

4.3) Realice el diagrama de dispersión empleando el programa Graph.

4.4) Calcule el coeficiente de correlación por rangos de Spearman empleando la ecuación.

0,49

4.5) Calcule el coeficiente de correlación empleando Excel.

0,49

5) Cree y resuelva un ejercicio similar al anterior.

4) COEFICIENTE DE DETERMINACIÓN

Revela qué porcentaje del cambio en Y se explica por un cambio en X. Se calcula elevando al cuadrado el coeficiente de correlación.

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

$$x = X - \bar{X}$$

$$y = Y - \bar{Y}$$

r = Coeficiente de correlación de Pearson

r^2 = Coeficiente de determinación

La ecuación del coeficiente producto-momento (Coeficiente de Pearson) $r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$ puede escribirse en la forma equivalente:

$$\text{Coeficiente de Pearson} = r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

De donde coeficiente de determinación = $r^2 = (\text{Coeficiente de Pearson})^2$

Ejemplo ilustrativo

Con los datos de la siguiente tabla sobre las temperaturas, calcular el coeficiente de determinación empleando la ecuación obtenida de la forma equivalente del coeficiente de Pearson.

X	18	17	15	16	14	12	9	15	16	14	16	18
Y	13	15	14	13	9	10	8	13	12	13	10	8

Solución:

Se calcula el coeficiente de Pearson llenando la siguiente tabla:

X	Y	XY	X²	Y²
18	13	234	324	169
17	15	255	289	225
15	14	210	225	196
16	13	208	256	169
14	9	126	196	81
12	10	120	144	100
9	8	72	81	64
15	13	195	225	169
16	12	192	256	144
14	13	182	196	169
16	10	160	256	100
18	8	144	324	64
$\Sigma X = 180$	$\Sigma Y = 138$	$\Sigma XY = 2098$	$\Sigma X^2 = 2772$	$\Sigma Y^2 = 1650$

Se aplica la ecuación para calcular el coeficiente de Pearson.

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} = \frac{12 \cdot 2098 - 180 \cdot 138}{\sqrt{[12 \cdot 2772 - (180)^2][12 \cdot 1650 - (138)^2]}}$$

$$r = \frac{25176 - 24840}{\sqrt{[33264 - 32400][19800 - 19044]}} = \frac{336}{\sqrt{[864][756]}} = \frac{336}{\sqrt{653184}} = \frac{336}{808,198} = 0,4157$$

Elevando al cuadrado coeficiente de Pearson queda calculado el coeficiente de determinación.

$$\text{Coeficiente de determinación} = r^2 = (0,4157)^2 = 0,1728$$

Esto establece que 17,28% del cambio en Y se explica mediante un cambio en X.

Nota:

El r^2 tiene significado sólo para las relaciones lineales. Dos variables pueden tener $r^2 = 0$ y sin embargo estar relacionadas en sentido curvilíneo. El valor de r^2 no se interpreta como si la variable Y fuera causado por un cambio de la variable X, ya que la correlación no significa causa.

En Excel se calcula elevando al cuadrado el coeficiente de correlación o insertando la función =COEFICIENTE.R2 como muestra la siguiente figura:

COEFICIENTE.R2						
	A	B	C	D	E	
1	X	Y				
2	18	13				
3	17	15				
4	15	14				
5	16	13				
6	14	9				
7	12	10				
8	9	8				
9	15	13				
10	16	12				
11	14	13				
12	16	10				
13	18	8				
14						
15	r	0,4157397	=COEF.DE.CORREL(A2:A13;B2:B13)			
16	r ²	0,1728395	=B15^2			
17	r ²	0,1728395	=COEFICIENTE.R2(B2:B13;A2:A13)			

TAREA DE INTERAPRENDIZAJE

1) La siguiente tabla muestra el dinero en miles de dólares gastado en publicidad por una empresa (X) para vender sus productos, y el número en miles de clientes (Y) que compran los productos de la empresa.

X	15	17	14	13	18	20	17	18	16	14	20	18
Y	30	34	28	26	32	40	34	36	32	25	40	36

1.1) Realice el diagrama de dispersión en forma manual.

1.2) Realice el diagrama de dispersión empleando Excel.

1.3) Realice el diagrama de dispersión empleando el programa Graph.

1.4) Calcule el coeficiente de Pearson empleando las dos fórmulas.

0,96015

1.5) Calcule el coeficiente de determinación empleando las dos fórmulas y mediante Excel.

0,9219

2) La siguiente tabla muestra el tiempo en minutos dedicado al estudio y la calificación sobre 10 obtenida.

X	140	150	130	120	170	190	180	160	200	110	100	90
Y	7	8	7	6	8	10	9	8	10	6	5	4

2.1) Realice el diagrama de dispersión en forma manual.

2.2) Realice el diagrama de dispersión empleando Excel.

2.3) Realice el diagrama de dispersión empleando el programa Graph.

2.4) Calcule el coeficiente de Pearson empleando las dos fórmulas.

0,9817

2.5) Calcule el coeficiente de determinación empleando las dos fórmulas y mediante Excel.

0,9638

3) Cree y resuelva un ejercicio similar a los anteriores.

REFERENCIAS BIBLIOGRÁFICAS

SUÁREZ, Mario, (2012), Interaprendizaje de Estadística Básica, Universidad Técnica del Norte
TAPIA, Fausto Ibarra, Ecuador.

SUÁREZ, Mario, (2011), Coeficiente de correlación de Pearson para datos agrupados en intervalos,
www.monografias.com/trabajos86/

SUÁREZ, Mario, (2011), Coeficiente de Correlación por Rangos de Spearman,
www.monografias.com/trabajos85/