



Práctica 7: CORRELACIÓN

En esta práctica vamos a aprender a analizar la relación entre dos variables continuas. Construiremos diagramas de dispersión que nos permitirán visualizar esta relación y calcularemos medidas de relación entre variables continuas: covarianza y coeficiente de correlación de Pearson.

1. Introducción de datos: En esta práctica vamos a partir de los datos ya introducidos en nuestra base de datos (Hoja **Datos** de nuestro fichero Excel). Vamos a utilizar las variables Peso y Estatura.

	A	B	C	D	E	F	G	H	I	J	K
1	Sexo	Edad	Núm. Hermanos	Nota Selectividad	Horas Estudio	Horas Deporte	Vive	Trabaja	Libros	Peso	Estatura
2	1	49	4	8,5	7,5	5	5	1	38	98	1,76
3	2	18	1	18	4	4	4	1	4	60	1,85
4	2	33	2	5	11,5	1	2	1	13	50	1,7
5	2	44	3	7,34	17,5	0	2	1	15	54	1,51
6	1	18	4	5,825	12	0	3	1	28	97	1,75
7	2	20	2	7,8	4	1	1	2	3	51	1,55
8	2	18	3	7,8	14	6	5	1	4	58	1,6
9	2	19	1	6	4	2	1	1	8	52	1,6
10	2	18	3	5,8	4	3	1	1	4	50	1,64
11	2	19	2	6,4	1	5	1	2	2	61	1,7
12	1	18	2	8	13	14	2	2	15	-	-
13	2	18	2	8,29	12	0	1	1	10	52,5	1,85
14	1	18	1	6,4	12	6	1	1	16	70	1,85
15	2	18	1	6,16	8	2	1	1	3	58	1,6
16	1	18	2	8	0	2	1	1	8	62	1,75
17	2	22	4	5,06	2,5	3	1	2	5	49	1,5
18	2	35	7	6,69	0	2	5	1	50	58	1,7
19	2	20	1	5	-	0	1	1	6	75	1,8
20	1	36	3	-	7,5	3	1	1	6	76	1,85
21	2	19	2	6,9	6,5	0	5	1	30	66	1,58
22	2	27	4	6,03	0	0	1	1	3	50	1,6
23	2	18	2	5,9	4	2	1	1	3	60	1,59
24	2	36	3	-	-	-	-	-	-	-	-
25	2	18	2	6	6	2	1	1	10	60	1,63
26	2	20	2	6,24	2	1	1	1	2	67	1,58
27	2	21	1	5,32	-	0	3	2	4	53	1,64
28	1	23	1	5,4	2	2	1	1	2	-	1,83
29	1	18	2	6,7	7,5	5	1	1	11	70	1,9
30	1	18	1	8,14	8	1	1	1	10	65	1,67
31	1	21	4	6,5	10	10	1	2	20	72	1,75

Creamos una nueva hoja llamada **Peso-Estatura**. Copiamos las columnas correspondientes a las variables Peso y Estatura en las dos primeras columnas de esta nueva hoja.

Suprime las filas en las que falta algún dato. Para ello, sitúate en la cabecera de la fila, pincha con el botón derecho del ratón y selecciona Eliminar.

2. Diagrama de dispersión:

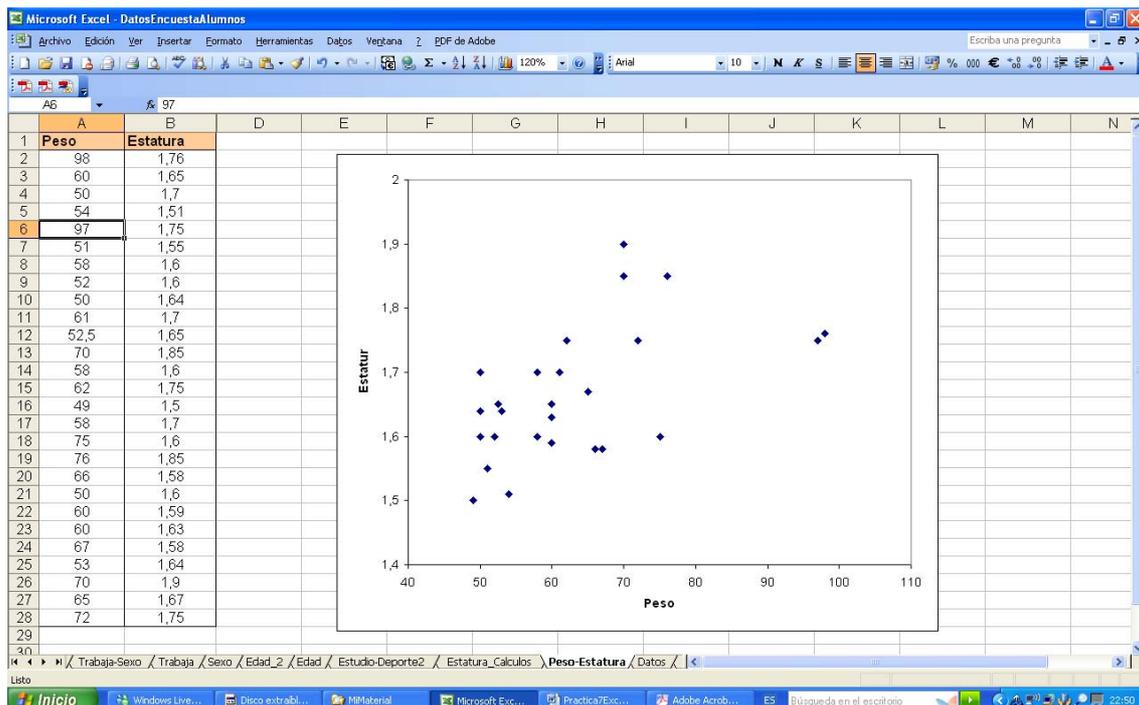
Selecciona con el ratón las dos columnas de valores de Peso y Estatura. Después presiona en el botón de Gráficos, elije XY (Dispersión) y presiona en Siguiente. Quita la leyenda y las líneas de división del gráfico, y en la ficha Títulos pon el nombre de las variables graficadas en cada eje. Presiona Siguiente y después Finalizar.

3. Modificación del formato del gráfico: Pincha con el botón derecho del ratón en el gráfico y selecciona Formato de área de trazado. En la ficha Área selecciona Ninguna.

Ahora pincha en el eje X con el botón derecho y selecciona Formato de ejes. En la ficha Escala pon Mínimo: 40. De la misma forma pero en el eje Y, pon Mínimo: 1,4. De esta forma podemos ver mejor los puntos graficados.

¿Existe relación entre el Peso y la Estatura de los alumnos?

Si existe, ¿Cómo es esta relación?

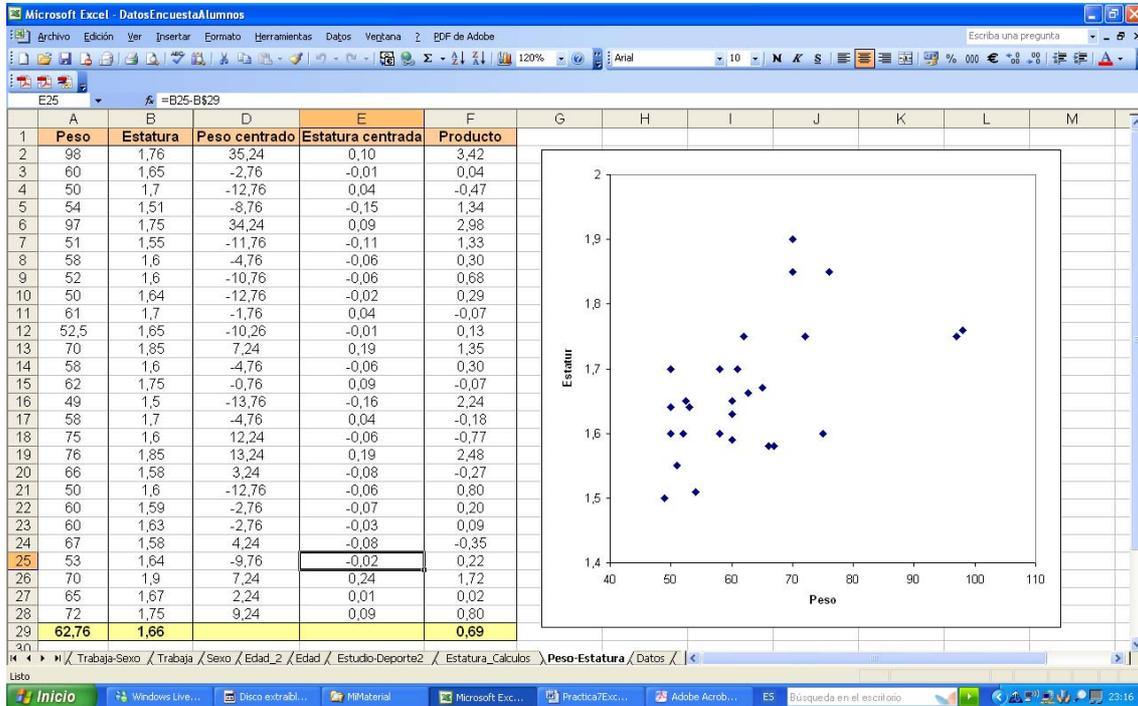


4. Covarianza: Calcula el Peso centrado o diferencias entre Peso y el peso medio en la columna D (para eso primero necesitas calcular el peso medio). Haz lo mismo para la Estatura en las columnas E (mueve el gráfico hacia la derecha). En la columna F, calcula el producto entre el Peso centrado y la Estatura centrada. Al final de esta columna, calcula la media. Este valor es la covarianza, cuya fórmula es:

$$S_{X,Y} = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})$$

¿Qué signo toma la covarianza?

¿Qué tipo de relación indica ese signo?

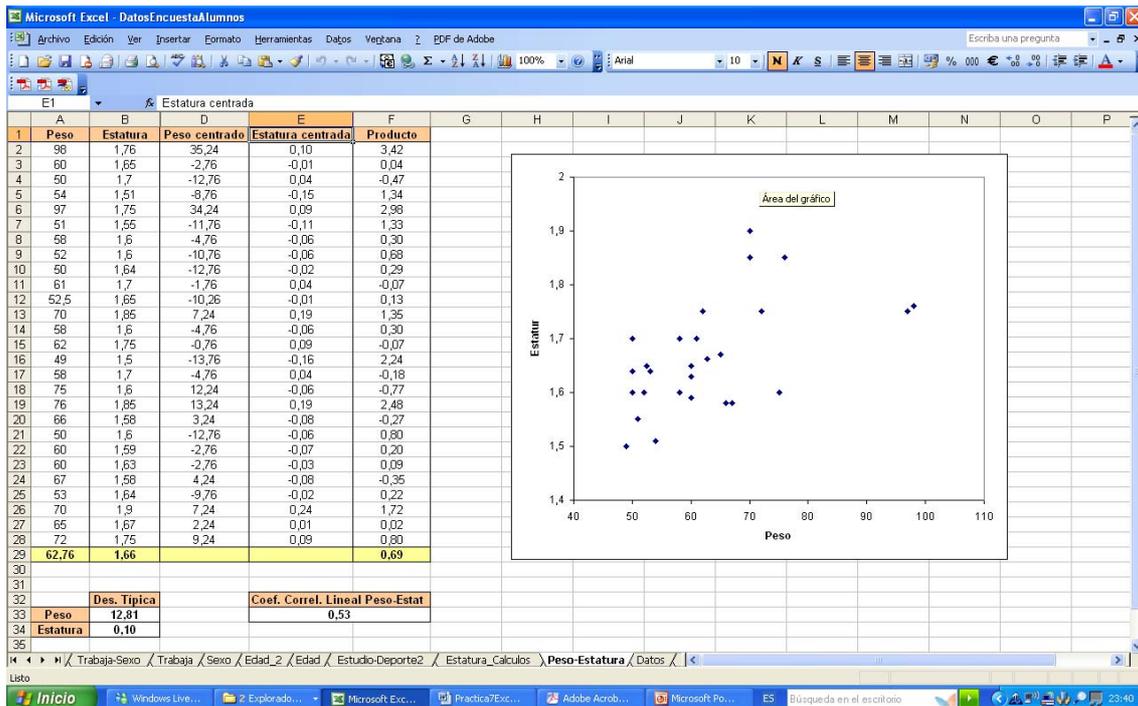


¿Qué tiene que ocurrir con el Peso y la Estatura de un estudiante para que su valor en la columna Producto sea positivo?

¿Y qué tiene que ocurrir para que la suma de los valores de la columna Producto sea un número positivo?

5. Coeficiente de correlación lineal: Calcula las desviaciones típicas del Peso y de la Estatura en las celdas B33 y B34 respectivamente. Después calcula el coeficiente de correlación lineal entre el Peso y la Estatura en la celda E33, el cual viene dado por la fórmula:

$$r_{x,y} = \frac{s_{x,y}}{s_x s_y}$$



¿Existe mucha o poca correlación lineal entre el Peso y la Estatura?

Ejercicio 1: RELACIÓN ENTRE HORAS DE ESTUDIO Y HORAS DE DEPORTE

Estudia la relación entre las Horas de estudio semanales y las Horas de deporte semanales de los alumnos haciendo el gráfico de dispersión y calculando la covarianza y el coeficiente de correlación lineal.