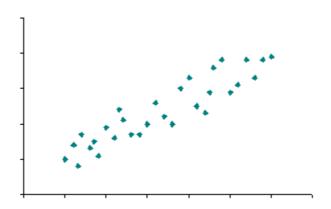


DIAGRAMA DE CORRELACIÓN-DISPERSIÓN

¿QUÉ ES EL DIAGRAMA DE CORRELACIÓN-DISPERSIÓN?

- Es una herramienta gráfica que permite demostrar la relación existente entre dos clases de datos y cuantificar la intensidad de dicha relación.
- Se utiliza para conocer si efectivamente existe una correlación entre dos magnitudes o parámetros de un problema y, en caso positivo, de qué tipo es la correlación.



Ejemplo de diagrama de correlación-dispersión

La secuencia a seguir para realizar un diagrama de dispersión es:

1. Recoger y ordenar los datos que se cree que tienen una posible correlación.

Los datos son recogidos en una tabla, indicando el número de muestras y los valores de las características que se quiere investigar. Es conveniente que el número de mediciones sea de al menos 30.

EJEMPLO:

Una empresa se plantea cambiar la composición de uno de sus productos utilizando un nuevo material. Antes de tomar una decisión, la empresa decide realizar un ensayo para estudiar la posible relación entre la utilización de dicho material y el número de defectos. Para ello analiza lotes con diferentes porcentajes del nuevo material y toma los siguientes datos:

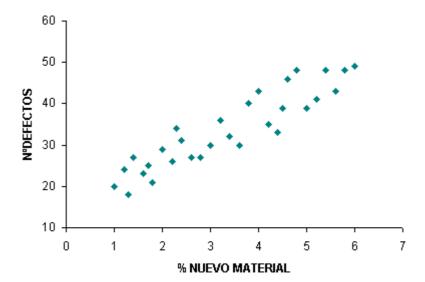
% Nuevo material	Nº defectos	% Nuevo material	Nº defectos
1	20	3,4	32
1,2	24	3,6	30
1,3	18	3,8	40
1,4	27	4	43
1,6	23	4,2	35
1,7	25	4,4	33
1,8	21	4,5	39
2	29	4,6	46
2,2	26	4,8	48
2,3	34	4,8 5	39
2,4	31	5,2 5,4	41
2,6	27	5,4	48
	27	5,6	43
2,8 3	30	5,8	48
32	36	1 6	40

2. Representar gráficamente los datos:

- Dibujar, en un diagrama cartesiano, los ejes vertical y horizontal de la misma longitud. Observar los valores máximo y mínimo de los grupos de datos, para escoger la escala de representación adecuada a los mismos y evitar así errores de interpretación.
- Representar en el diagrama mediante puntos cada par de datos, reflejando los valores de la clase de datos que se considera dependiente (causa) sobre el eje horizontal X y los valores de la clase de datos que se considera independiente (efecto) sobre el eje vertical Y.
- Si dos o más pares de datos caen en el mismo punto, dibujar círculos concéntricos alrededor del punto individualizado.

EJEMPLO:

La figura siguiente muestra la representación gráfica de los datos:



- 3. Una vez construido el diagrama se analiza la forma que tiene la nube de puntos obtenida, para así determinar las relaciones entre los dos tipos de datos. Este análisis puede efectuarse por técnicas estadísticas que permitan determinar si existe o no relación, y el grado de existencia en su caso. Las herramientas utilizadas son:
 - La recta de regresión, y
 - El coeficiente de correlación lineal.

¿CÓMO SE UTILIZA?

3.A. RECTA DE REGRESIÓN

La recta de regresión es la línea meior representa а conjunto de puntos. La función donde: que aproxima la recta es:

 $\hat{v} = a + bx$

donde:

ŷ: variable

dependiente (causa)

a : ordenada en el

origen.

b: pendiente de la recta de regresión

x: variable

independiente (efecto)

La pendiente se halla mediante la expresión:

$$b = \frac{\sum xy - n\overline{x}\overline{y}}{\sum x^2 - n\overline{x}^2}$$

x: valores de la variable

independiente

y: valores de la variable

dependiente

x: media de los valores

y: media de los valores

de y

número de observaciones o pares

de datos

La ordenada en el origen se calcula como:

 $a = \overline{y} - b\overline{x}$

EJEMPLO:

Calculamos la recta de regresión a partir de los siguientes datos:

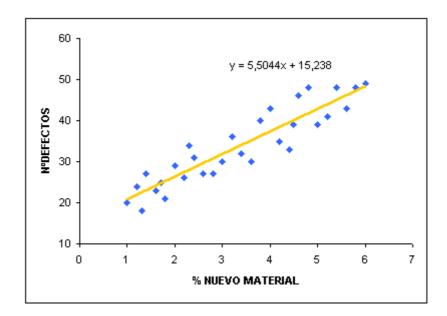
$$\begin{split} \Sigma x = &100,8 \quad \Sigma y = 1012 \quad \Sigma x^2 = 407,12 \quad \Sigma y^2 = 36600 \\ \Sigma x y = &3777 \quad x = \ 3,36 \quad y = \ 33,73 \end{split}$$

$$b = \frac{\Sigma x y - n \overline{x} \overline{y}}{\Sigma x^2 - n \overline{x}^2} = \frac{3777 - \left(30 \cancel{3},36 \cancel{3}3,73\right)}{407,12 - \left(30 \cancel{3},36^2\right)} = \frac{377}{68,432} = 5,5 \end{split}$$

$$a = \overline{y} - b\overline{x} = 33,73 - 5,5(3,36) = 15,25$$

$$y = 15,25 + 5,5 x$$

Representamos en el gráfico la recta de regresión:



Observamos que existe una clara correlación entre los datos.

3B. COEFICIENTE DE CORRELACIÓN

El coeficiente de correlación lineal r, viene determinado por la expresión:

$$r = \frac{\left(\sum xy - n\overline{xy}\right)}{\sqrt{\left(\sum x^2 - n\overline{x}^2\right)\left(\sum y^2 - n\overline{y}^2\right)}}$$

Toma valores comprendidos entre -1 y 1. Cuanto más próximo a 0 sea r menor será la relación entre los datos, y cuanto más próximo a 1 (en valor absoluto) mayor será dicha relación. Su signo indica si se da una relación positiva o negativa entre las variables x e y.

EJEMPLO

Calculamos el coeficiente de correlación:

$$r = \frac{\left(\sum xy - n\overline{x}\overline{y}\right)}{\sqrt{\left(\sum x^2 - n\overline{x}^2\right) \cdot \left(\sum y^2 - n\overline{y}^2\right)}} = 0,918$$

$$r = 0.918$$

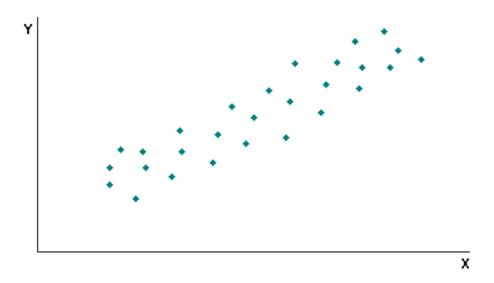
El valor obtenido es muy próximo a 1, lo que nos confirma que la correlación es fuertemente positiva.

CONSEJOS PARA ELABORAR Y USAR LOS DIAGRAMAS DE CORRELACIÓN-DISPERSIÓN

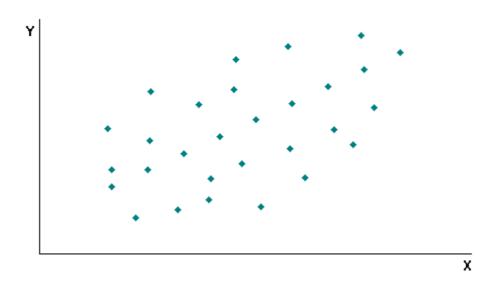
- Si existen puntos muy apartados en el diagrama se puede deber a errores en la medición o registro de los datos o a algún cambio en las condiciones del proceso. En cualquier caso, se deben estudiar las causas que los han originado y apartarlos del análisis.
- Es conveniente estratificar los datos, ya que esto nos puede permitir descubrir correlaciones donde aparentemente no las hay y desechar otras que lo parecen pero no lo son. Ejemplo: Supongamos que producimos ruedas con caucho procedente de dos países distintos y analizamos la correlación entre el diseño del dibujo y la resistencia al desgaste. Es posible que encontremos correlación si estratificamos los datos entre los dos orígenes del caucho.
- No hay que olvidar que, a veces, la casualidad hace aparecer correlaciones donde no las hay. Por ello, una vez que creamos haber encontrado una correlación hay que analizar los motivos de ésta antes de tomar otras medidas.

¿CÓMO SE INTERPRETA?

Correlación positiva: A un crecimiento de X (causa) corresponde un crecimiento de Y (efecto). Controlando la evolución de los valores de X, quedan controlados los valores de Y, Figura siguiente.



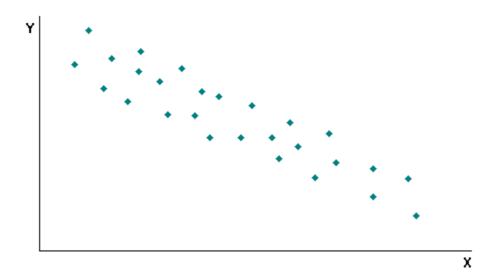
Correlación positiva débil: A un crecimiento de X se observa una tendencia a crecer de Y, pero se presume que existen otras causas de dependencia, Figura siguiente.



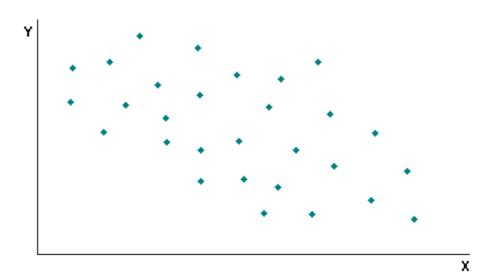
CALIDAD JOSÉ MANUEL DOMENECH ROLDÁN PROFESOR DE ENSEÑANZA SECUNDARIA

¿CÓMO SE INTERPRETA?

Correlación negativa: A un crecimiento de X se observa una tendencia a disminuir de Y, Figura siguiente.

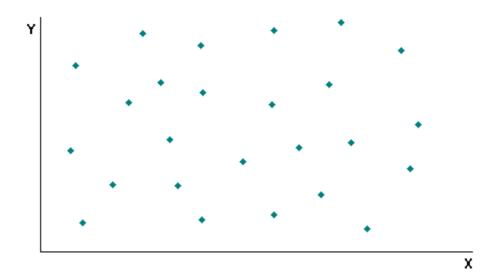


Correlación negativa débil: A un crecimiento de X se observa una tendencia a disminuir de Y, pero se presume que existen otras causas de dependencia, Figura siguiente.

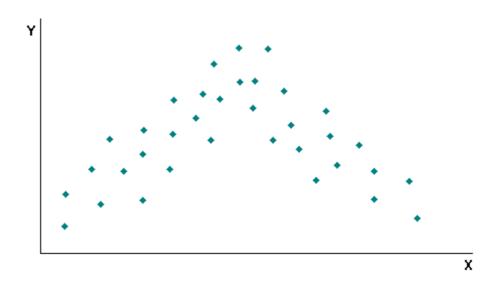


¿CÓMO SE INTERPRETA?

No existe una correlación evidente en la Figura siguiente.



Puede existir una Correlación Compleja en la Figura siguiente.





EJERCICIO PROPUESTO

La Dirección de una mina está preocupada por el alto porcentaje de indisponibilidad de sus máquinas cargadoras. Encarga al Jefe de Mantenimiento que analice si está influyendo la antigüedad de dichas máquinas en su porcentaje de indisponibilidad. Para ello, recoge la información de la fecha de compra y del porcentaje de indisponibilidad de cada máquina y la traslada a la siguiente tabla.

Máquina	Fecha de compra	% Indisponibilidad
C-0037	1994	29
C-0038	1994	39
C-0039	1995	24
C-0040	1995	32
C-0041	1995	43
C-0042	1996	20
C-0043	1996	41
C-0044	1996	30
C-0045	1997	20
C-0046	1997	25
C-0047	1998	12
C-0048	1998	19
C-0049	1999	10
C-0050	1999	30
C-0051	2000	9
C-0052	2000	14

Utilizar un diagrama de correlación para realizar este estudio.

----- 00000 -----