Diagrama de dispersión: esquema visual que refleja la correlación entre dos variables, y si existe permite determinar el tipo de relación presentada.

Gráfica 6.21 Diagrama de dispersión.

Fuente: Mario Gutiérrez. *Nociones de calidad*. México: Limusa, 1997. Adaptación de María Inés Díaz.

QUÉ ES

Herramienta que permite determinar si existe correlación entre dos variables, y si existe, entonces establecer una relación cuantitativa entre ellas.

PROCEDIMIENTO

1. Recolectar pareja de datos.
2. Asignar un eje a cada variable.
3. Graficar los puntos respectivos.
4. Determinar la mediana para X y Y y graficar.
5. Identificar los 4 cuadrantes.
6. Identificar los puntos existentes en cada cuadrante.
7. Analizar el tipo de relación que existe entre las variables.

APLICACIÓN

Establecer relaciones entre causa y efecto.

Reconocer relaciones estáticas y dinámicas entre las variables.

Ejemplo

Una empresa de la industria pesquera cuenta con los siguientes registros del número de veces que salieron sus barcos a realizar pesca (número de viajes por año) y las toneladas por año obtenidas:

Año	**Número de salidas por año**	**Toneladas por embarcación por año**
2001	456	21,1
2002	536	23,8
2003	554	25,2
2004	675	26,1
2005	702	27,4
2006	730	28,9
2007	750	30,5
2008	918	38,4
2009	928	43,5
2010	897	44,6

En el análisis de regresión debe determinarse si existe algún tipo de relación entre las variables; para ello deben seguirse los pasos siguientes:

- La relación más simple entre variables es la de tipo lineal.
- La gráfica de una función lineal es una recta, y la presentación general de la función es como sigue:

$$y = a + bx$$

Donde b es la pendiente de la recta y a es su punto de intersección con el eje y (en un sistema de coordenadas o plano cartesiano), es decir, el valor de y cuando $x = 0$.

Esto significa que:

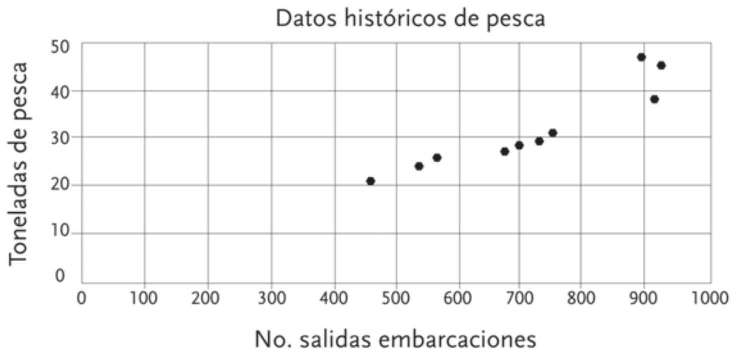
y = variable dependiente
x = variable independiente
a = variable exógena
b = relación de cambio de la variable dependiente con relación a un cambio de la variable independiente.

Lo primero que debe hacerse es determinar cuál es la variable independiente (x) y cuál es la variable dependiente (y) y elaborar una gráfica de puntos, para observar la tendencia de los datos y el tipo de relación existente entre las variables.

En este caso, el número de viajes realizados es la variable independiente (x) y las toneladas de pesca son la variable dependiente (y), relación que se explica muy fácilmente ya que sin salidas de los barcos no habría pesca alguna.

La gráfica que se obtiene es:

Gráfica 6.22 Diagrama de dispersión (caso).



La gráfica 6.22 muestra claramente la existencia de una relación entre las variables, aparentemente de carácter lineal.

Para estimar los valores de b y de a se aplica el método de los mínimos cuadrados, el cual emplea los siguientes parámetros:

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2} \qquad b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

La relación esperada entre variables será de la forma:

y (ton pesca) = b (pendiente) $\times$ x (No. de salidas de embarcaciones) + a

Con esto se pretende buscar una línea que ajuste los datos de manera óptima.

Otro elemento por considerar es la importancia de contar con datos fiables y que estén en un grado de actualización suficiente, ya que la metodología sirve para hacer pronósticos.

Para el caso de la pesca, los resultados obtenidos se reúnen en una matriz, mediante la cual se amplía y se obtienen los distintos valores necesarios para calcular los valores de b (pendiente) y de a (intersección con el eje).

Los datos obtenidos son:

Año	Número de salidas por año	Toneladas por embarcación por año		
	x	y	S X Y	S X∧2
2001	456,0	21,1	9.621,6	207.936,0
2002	536,0	23,8	12.756,8	287.296,0
2003	554,0	25,2	13.960,8	306.916,0
2004	675,0	26,1	17.617,5	455.625,0
2005	702,0	27,4	19.234,8	492.804,0
2006	730,0	28,9	21.097,0	532.900,0
2007	750,0	30,5	22.875,0	562.500,0
2008	918,0	38,4	35.251,2	842.724,0
2009	928,0	43,5	40.368,0	861.184,0
2010	897,0	44,6	40.006,2	804.609,0
	S X	S Y		
Total	7.146,0	309,5	232.788,9	5.354.494,0
	(S X)∧2			
	51.065.316,0			
n =	10			
b =	0,04686275			
a =	–2,53812126			

Es decir, la curva que mejor ajusta los datos es:

$y = 0{,}0469x - 2{,}5381$

Si la organización deseara saber cuántas toneladas de pesca puede esperar si realizara 1.000 salidas, el dato sería:

$y = 0{,}0469 \times (x) - 2{,}5381$
$y = 0{,}0469 \times (1.000) - 2{,}5381$
$y = 44{,}32$ ton por año