

DIAGRAMA DE DISPERSIÓN

1.- INTRODUCCIÓN

Este documento describe el proceso completo a seguir para analizar la existencia de una relación lógica entre dos variables.

Describe la construcción de los Diagramas de Dispersión a partir de la recogida de datos acerca de dichas variables y el análisis posterior necesario para confirmar la correlación que puede mostrar dicho diagrama, ya que ésta no implica la existencia de una relación lógica.

2.- OBJETIVO Y ALCANCE

Definir las reglas básicas a seguir para la construcción e interpretación de los Diagramas de Dispersión, resaltando las situaciones en que pueden, o deben, ser utilizados.

Es de aplicación a todos aquellos estudios en los que es necesario analizar relaciones entre fenómenos o efectos y relaciones de causalidad.

Su utilización será beneficiosa para el desarrollo de los proyectos abordados por los Equipos y Grupos de Mejora y por todos aquellos individuos u organismos que estén implicados en la mejora de la calidad.

Además, se recomienda su uso como herramienta de trabajo dentro de las actividades habituales de gestión.

3.- RESPONSABILIDADES

a) Grupo de trabajo o persona responsable del estudio:

- Recoger los datos.
- Seguir las reglas que se señalan en el procedimiento para la construcción del Diagrama de Dispersión y para su correcta interpretación.

b) Dirección de Ingeniería de Calidad

- Asesorar, a quien así lo solicite, en las bases para la construcción y utilización de los Diagramas de Dispersión.

4.- DEFINICIONES / CONCEPTOS

4.1.- CORRELACIÓN

Definición

Se entiende por correlación el grado de relación existente entre dos variables.

Concepto

Cuando entre dos variables existe una correlación total, se cumple que a cada valor de una, le corresponde un único valor de la otra (función matemática).

Es frecuente que dos variables estén relacionadas de forma que a cada valor de una de ellas le correspondan varios valores de la otra.

En este caso es interesante investigar el grado de correlación existente entre ambas.

4.2.- DIAGRAMA DE DISPERSIÓN

Definición

Representación gráfica del grado de relación entre dos variables cuantitativas.

Características principales

A continuación se comentan una serie de características que ayudan a comprender la naturaleza de la herramienta.

Impacto visual

Un Diagrama de Dispersión muestra la posibilidad de la existencia de correlación entre dos variables de un vistazo.

Comunicación

Simplifica el análisis de situaciones numéricas complejas.

Guía en la investigación

El análisis de datos mediante esta herramienta proporciona mayor información que el simple análisis matemático de correlación, sugiriendo posibilidades y alternativas de estudio, basadas en la necesidad de conjugar datos y procesos en su utilización.

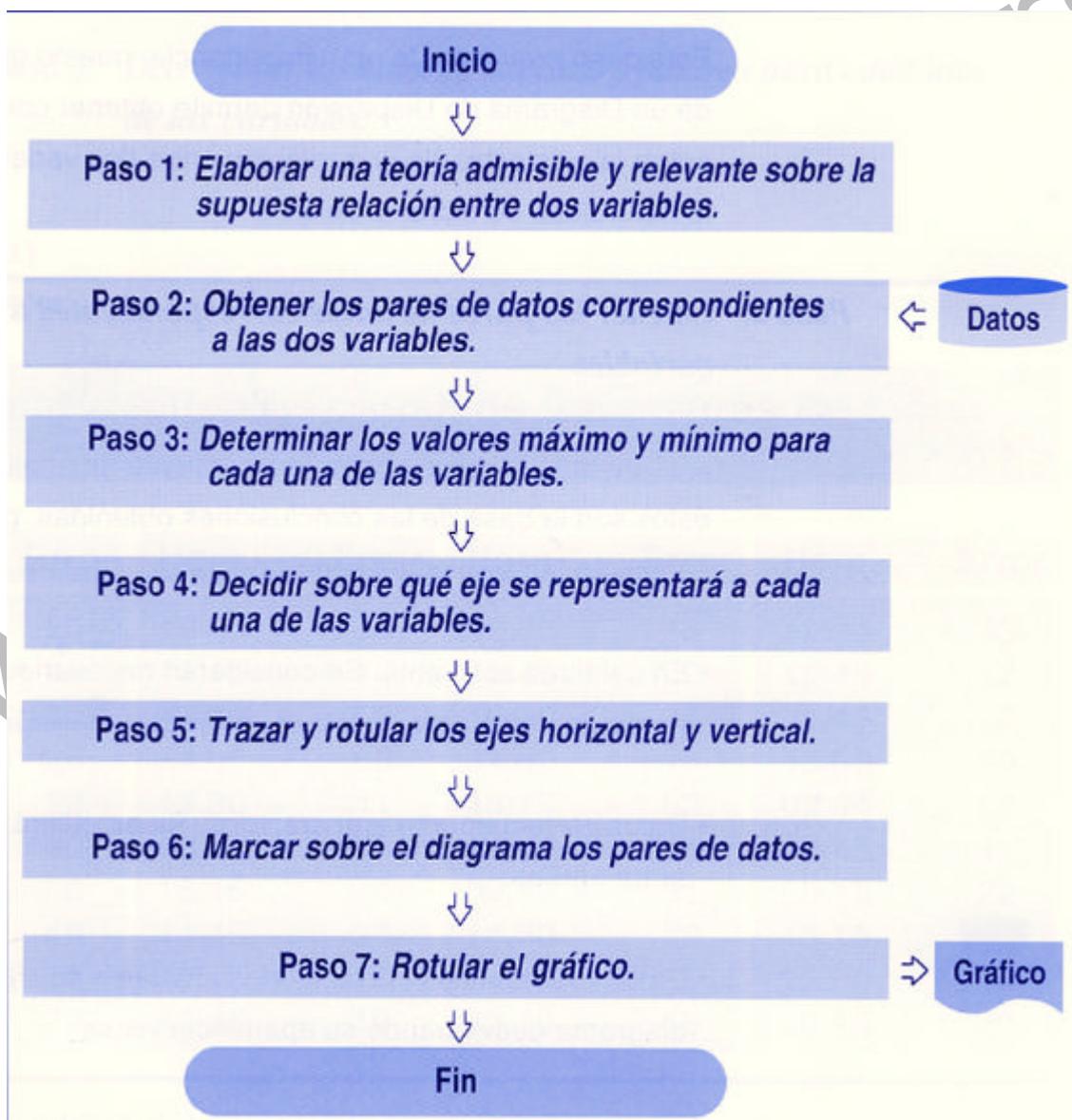
4.3.- ESTRATIFICACIÓN

Definición

Separar un conjunto de datos en diferentes grupos o categorías, de forma que los datos pertenecientes a cada grupo comparten características comunes que definen la categoría.

5.- PROCESO

5.1.- DIAGRAMA DE FLUJO



5.2.- CONSTRUCCIÓN

5.2.1.- Pasos previos a la construcción de un Diagrama de Dispersión

Paso 1: Elaborar una teoría admisible y relevante sobre la supuesta relación entre dos variables

Este paso previo es de gran importancia, puesto que el análisis de un Diagrama de Dispersión permite obtener conclusiones sobre la existencia de una relación entre dos variables, no sobre la naturaleza de dicha relación.

Paso 2: Obtener los pares de datos correspondientes a las dos variables

Al igual que en cualquier otra herramienta de análisis de datos, éstos son la base de las conclusiones obtenidas, por tanto cumplirán las siguientes condiciones:

- *En cantidad suficiente:* Se consideran necesarios al menos 40 pares de datos para construir un Diagrama de Dispersión.
- *Datos correctamente emparejados:* Se estudiará la relación entre ambos.
- *Datos exactos:* Las inexactitudes afectan a su situación en el diagrama desvirtuando su apariencia visual.
- *Datos representativos:* Asegúrese de que cubren todas las condiciones operativas del proceso.
- *Información completa:* Anotar las condiciones en que han sido obtenidos los datos.

5.2.2.- Pasos en la construcción de un Diagrama de Dispersión

Paso 3: Determinar los valores máximo y mínimo para cada una de las variables

Ejemplo: Tabla de los datos recogidos

Teoría: La fatiga es causa de los errores de tecleo							
Número de errores de tecleo según la hora del día							
Hora	Error	Hora	Error	Hora	Error	Hora	Error
11.00	25	13.30	38	09.30	15	12.45	33
14.15	45	12.15	14	15.45	72	10.45	17
10.00	7	16.30	56	10.30	35	11.45	8
13.45	26	14.45	60	11.30	18	15.30	30
09.15	22	12.30	11	16.15	63	09.45	22
16.00	50	13.30	55	15.30	62	12.45	41
12.30	60	15.15	40	09.45	31	09.15	22
13.45	19	14.45	25	14.30	32	15.15	80
14.30	78	10.45	10	14.45	56	12.00	30
13.00	22	13.45	19	12.15	45	10.15	22

Paso 4: Decidir sobre qué eje representará a cada una de las variables

Si se está estudiando una posible relación causa-efecto, el eje horizontal representará la supuesta causa.

Paso 5: Trazar y rotular los ejes horizontal y vertical

La construcción de los ejes afecta al aspecto y a la consiguiente interpretación del diagrama.

a) Los ejes han de ser aproximadamente de la misma longitud, determinando un área cuadrada.

b) La numeración de los ejes ha de ir desde un valor ligeramente menor que el valor mínimo de cada variable hasta un valor ligeramente superior al valor máximo de las mismas. Esto permite que los puntos abarquen toda el área de registro de los datos.

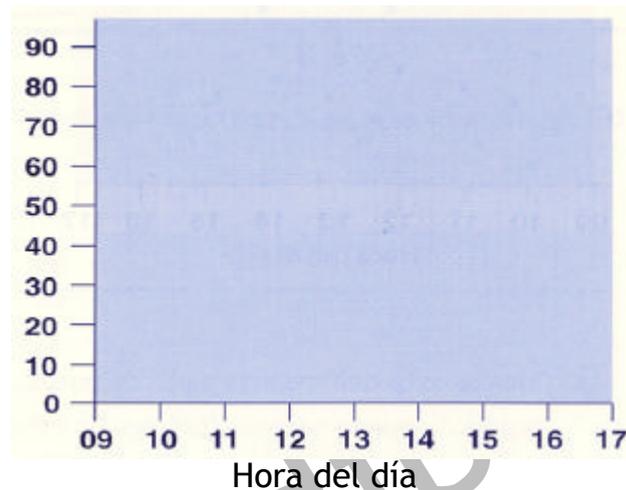
c) Numerar los ejes a intervalos iguales y con incrementos de la variable constantes.

d) Los valores crecientes han de ir de abajo a arriba y de izquierda a derecha en los ejes vertical y horizontal respectivamente.

e) Rotular cada eje con la descripción de la variable correspondiente y con su unidad de medida.

Ejemplo

Número de errores



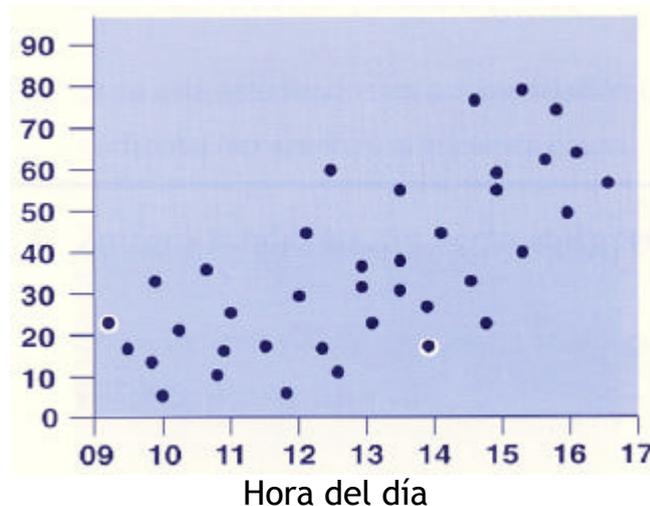
Paso 6: Marcar sobre el diagrama los pares de datos

a) Para cada par de datos localizar la intersección de las lecturas de los ejes correspondientes y señalarlo con un punto o símbolo.

Si algún punto coincide con otro ya existente, se traza un círculo concéntrico a este último.

Ejemplo

Número de errores



b) Cuando coinciden muchos pares de puntos, el Diagrama de Dispersión puede hacerse confuso. En este caso es recomendable utilizar una "Tabla de Correlación" para representar la correlación.

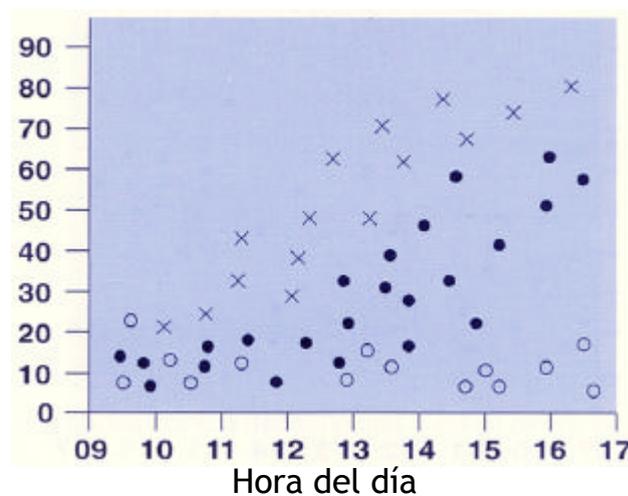
Ejemplo

Tabla de correlación							
Número de errores página	7						////
	6				/	###	///
	5			/	////	////	///
	4			///	///	////	///
	3	//	////	////	###	//	
	2	///	////	//			/
	1	### /	//				
		10.00	11.00	12.00	13.00	14.00	15.00
Hora del día							

c) En el caso en que se construye un Diagrama de Dispersión estratificado separando los pares de datos, por ejemplo, según el turno de trabajo, lote de materia prima, etc.), deben escogerse símbolos que pongan de manifiesto los diferentes grupos de puntos de forma clara.

Estratificación: Número de errores de tecleo según la hora del día y empleado

Número de errores



- José
- Javier
- x Juan

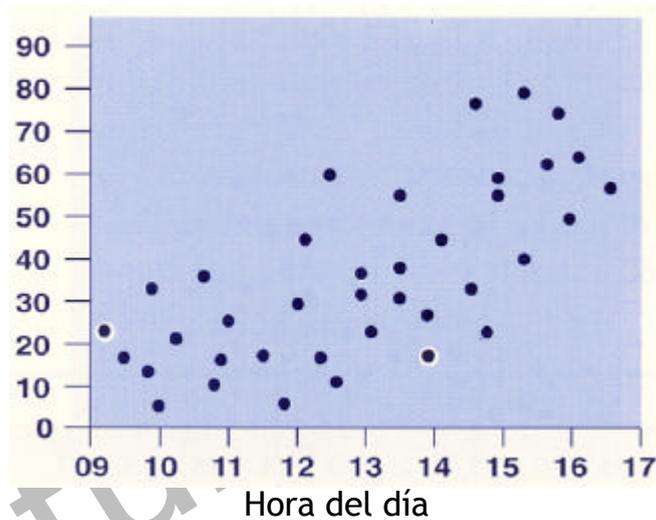
Paso 7: Rotular el gráfico

Se rotula el título del gráfico y toda aquella información necesaria para su correcta comprensión.

En general, es conveniente incluir una descripción adicional del objeto de las medidas y de las condiciones en que se han realizado, ya que esta información puede ayudar en la interpretación del diagrama.

Ejemplo: Número de errores de tecleo según la hora del día

Número de errores



5.3.- INTERPRETACIÓN

5.3.1.- Posibles tipos de relaciones entre variables

El Diagrama de Dispersión se puede utilizar para estudiar:

- Relaciones causa-efecto.

Este es el caso más común en su utilización para la mejora de la calidad. Se utiliza el diagrama a partir de la medición del efecto observado y de su posible causa.

Ejemplo: Comprobar la relación entre el número de errores y la hora en que se cometen.

- Relaciones entre dos efectos.

Sirve para contrastar la teoría de que ambos provienen de una causa común desconocida o difícil de medir.

Ejemplo: Analizar la relación entre el número de quejas que llegan y el aumento/disminución de las ventas, suponiendo que los dos dependen del nivel de satisfacción del cliente.

- Posibilidad de utilizar un efecto como sustituto de otro.

Se puede utilizar para controlar efectos difíciles o costosos de medir, a través de otros con medición más simple.

Ejemplo: Estudiar la relación existente entre reducción de costes y satisfacción del cliente para utilizar el parámetro de más fácil medición en la evaluación de las actividades de planificación.

- Relaciones entre dos posibles causas.

Sirve para actuar sobre efectos de forma más simple o adecuada y para analizar procesos complejos.

Ejemplo: Analizar la relación entre el porcentaje idóneo de contenido en potasio de un fertilizante y la cantidad media de lluvia recogida en la zona de cultivo, puesto que ambos elementos influyen en la calidad del vino y el régimen de lluvias no puede ser modificado.

5.3.2.- Proceso de interpretación

El Diagrama de Dispersión expresa el grado de relación entre dos variables, y dicha relación no necesariamente significa que una de ellas es la causa de la otra.

Ejemplo:

Suponiendo, por ejemplo, que haya sido identificada una fuerte correlación entre el precio de los periódicos y el precio de un vuelo aéreo de primera clase Madrid-Barcelona, permanecen todavía sin contestar las siguientes preguntas:

- ¿Qué relación existe entre los dos?
- ¿Influyen los precios de los periódicos distribuidos en el avión sobre el precio del vuelo?
- ¿Influye el costo de transporte por avión sobre el precio de los periódicos?
- ¿Dependen los dos de una tercera variable, por ejemplo el nivel de inflación?

El análisis de un Diagrama de Dispersión es un proceso de cuatro pasos:

Primero: Elaborar una teoría admisible y relevante sobre la supuesta relación entre dos variables.

Segundo: Recoger datos y construir el Diagrama.

Tercero: Identificar y clasificar la pauta de correlación.

Cuarto: Discutir la teoría original y considerar otras explicaciones.

La construcción y clasificación del Diagrama de Dispersión es la parte central del proceso. No es ni el principio ni el final.

5.3.3.- Pautas típicas de correlación

Correlación Fuerte

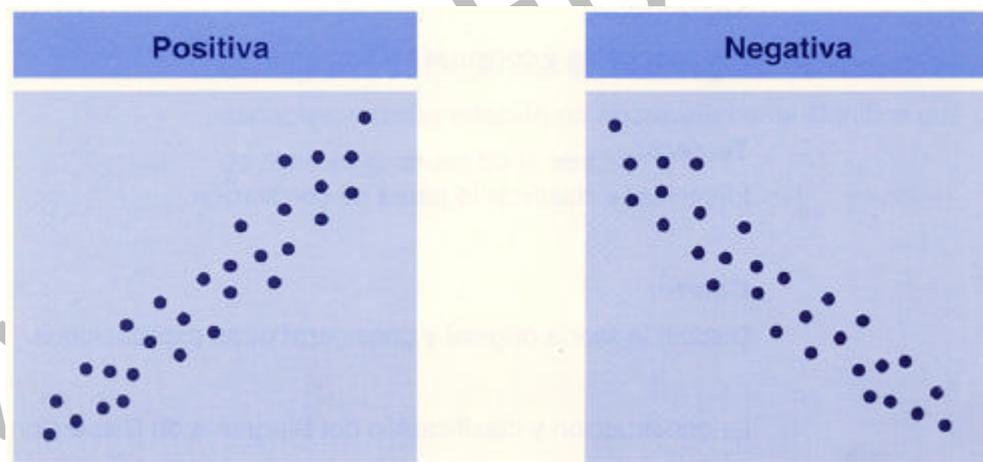
Los puntos se agrupan claramente alrededor de una línea imaginaria que pasa por el centro de la masa de los mismos. Estos casos sugieren que el control de una de las variables lleva al control de la otra.

Los datos parecen confirmar la teoría estudiada, pero hay que analizar la existencia de otras posibles explicaciones admisibles y relevantes para dicha relación.

Correlación Fuerte, Positiva: El valor de la variable "Y" (eje vertical) aumenta claramente con el valor de la variable "X" (eje horizontal).

Correlación Fuerte, Negativa: El valor de "Y" disminuye claramente cuando "X" aumenta.

Correlación fuerte



Correlación Débil

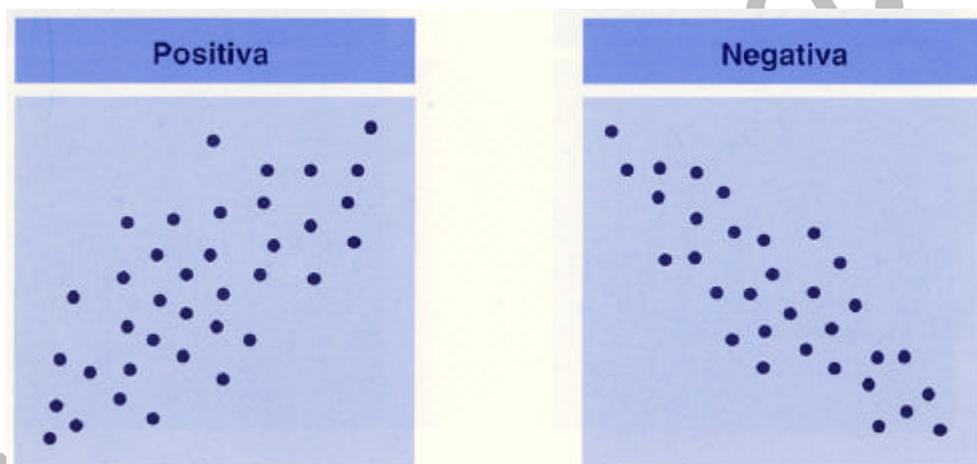
Los puntos no están suficientemente agrupados, como para asegurar que existe la relación. El control de una de las variables no necesariamente nos llevará al control de la otra.

Si lo que se busca es determinar las causas de un problema, se deben buscar otras variables con una relación mayor o más relevante sobre el efecto.

Correlación Débil, Positiva: El valor de la variable "Y" (eje vertical) tiende a aumentar cuando aumenta el valor de la variable "X" (eje horizontal)

Correlación Débil, Negativa: El valor de "Y" tiende a disminuir cuando aumenta el valor de "X".

Correlación débil



Correlación compleja

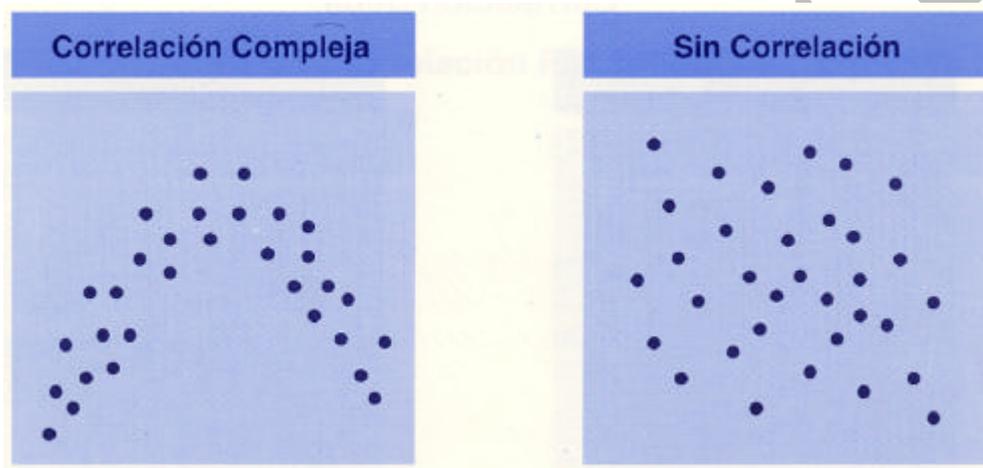
El valor de la variable "Y" parece estar relacionado con el de la variable "X", pero esta relación no es simple o lineal.

En este caso se estudia la relación más profundamente (¿Hay alguna ley no lineal que explique esta relación ?. ¿ Es esta relación el resultado de componer varias relaciones ?).

Sin correlación

Para cualquier valor de la variable "X", "Y" puede tener cualquier valor. No aparece ninguna relación especial entre ambas variables.

En este caso, nuestra teoría no es correcta y se deben buscar otros tipos de relaciones.



5.3.4.- Posibles problemas y deficiencias de interpretación

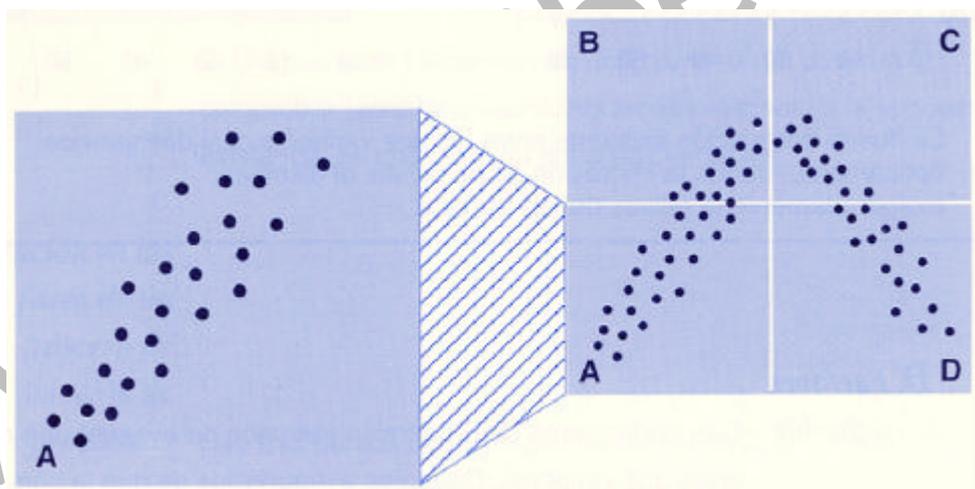
a) Correlación sin soporte lógico

Los Diagramas de Dispersión muestran solamente relaciones, no prueban relaciones causales. Ha de haber una explicación lógica y admisible para establecer una relación causa-efecto..

b) Recorrido de los datos

En el análisis del Diagrama se limitará su interpretación al recorrido de las observaciones. Generalizar las pautas de correlación para valores fuera de los límites del Grafico puede llevar a conclusiones completamente erróneas. Si se necesita conocer la relación entre dos variables para un rango de valores determinado, hay que obtener datos alrededor de ese rango.

Ejemplo



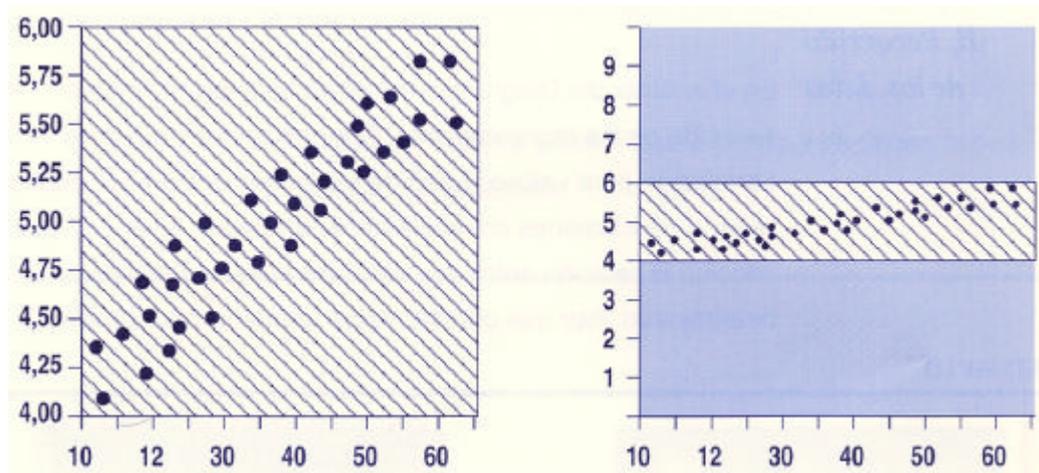
El examen de los datos relativos al campo "A" (Correlación positiva) no nos permite extraer conclusiones sobre el comportamiento de las variables para otros valores, por ejemplo en el campo "D" (Correlación negativa)

c) Efecto de la escala

Las escalas de los dos ejes influyen notablemente sobre la interpretación del Diagrama de Dispersión.

Escalas deficientes en alguno de los ejes puede enmascarar una relación o hacer ver relaciones inexistentes.

Ejemplo



La fuerte correlación existente entre las dos variables casi desaparece óptimamente, dando la impresión de no existir, al disminuir exageradamente la escala del eje vertical.

d) Factores de confusión

Con el Diagrama de Dispersión tratamos de estudiar una relación entre dos variables. Debemos asegurarnos de que la correlación que observamos no sea debida a una variable distinta de la que estamos registrando.

La forma ideal para tratar los factores de confusión es antes de la recogida de datos. Identificar los posibles factores de confusión y disponer la toma de datos de forma que se mantengan razonablemente constantes.

Los factores de confusión se medirán en la toma de datos, y si no ha sido posible mantenerlos constantes, se construirá un Diagrama de Dispersión estratificado según las condiciones de dichos factores.

e) Problemas con los datos

Si los datos son deficientes, la interpretación del Diagrama de Dispersión tiene, por fuerza, que ser deficiente.

5.4.- UTILIZACIÓN

El Diagrama de Dispersión es una herramienta útil para comprobar (aceptar o rechazar) teorías respecto a la supuesta existencia de una relación entre dos variables.

Utilización en las fases de un proceso de solución de problemas

Hay tres puntos de dicho proceso en los que el Diagrama de Dispersión puede ser una herramienta útil:

- Durante la fase de diagnóstico, para ensayar teorías sobre las causas e identificar las causas raíz.
- Durante la fase de corrección, en el diseño de soluciones.
- Para el diseño de un sistema de control que mantenga los resultados de una acción de mejora de la calidad.

6.- ANEXOS

Ejemplo 1 Errores en las facturas

Situación

Un equipo se encargó de analizar las causas de los frecuentes errores en las facturas.

El número de datos a rellenar variaba según el tipo de factura. Un miembro del equipo propuso concentrarse en simplificar las facturas más complicadas, seguramente causa de la mayoría de los errores.

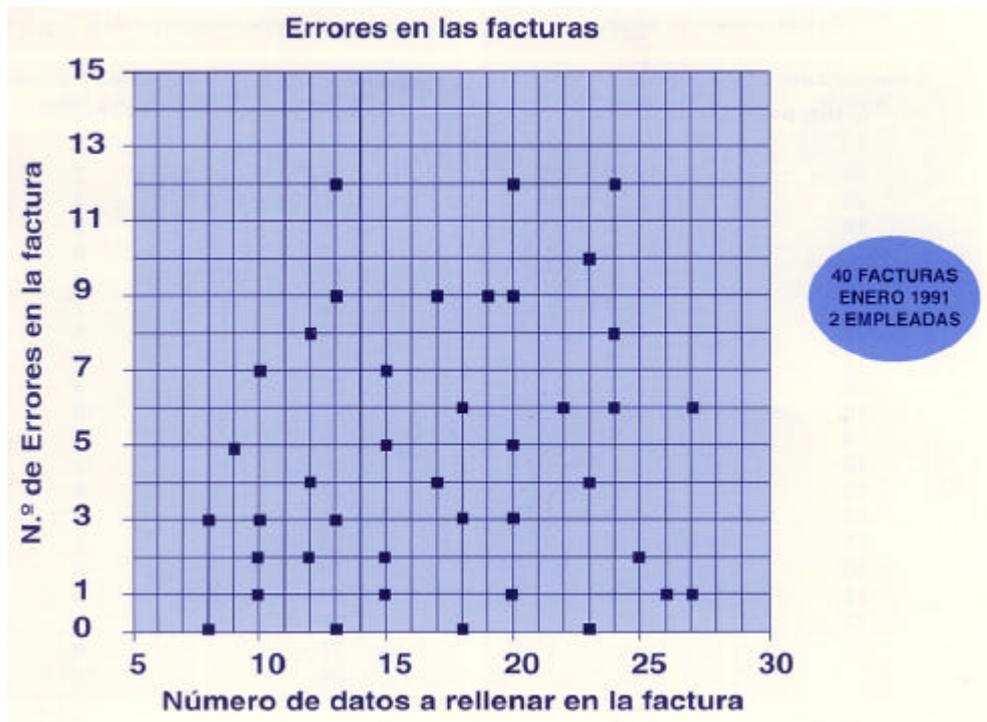
El equipo decidió investigar en primer lugar la teoría, aparentemente obvia, según la cual el número de errores en una factura dependía de la cantidad de datos a incluir en la misma.

El equipo recogió los datos relativos a los últimos meses y los representó en un Diagrama de Dispersión.

Tabla de los datos recogidos

Número de datos en la factura	Número de errores en la factura	Número de datos en la factura	Número de errores en la factura	Número de datos en la factura	Número de errores en la factura
8	3	15	7	15	1
15	2	26	1	22	6
15	5	20	5	24	6
12	4	10	2	19	9
10	7	10	3	12	8
25	2	9	5	27	6
20	3	13	3	24	12
17	9	20	9	10	1
18	3	23	10	12	2
23	4	20	12	17	4
27	1	13	12		
8	0	24	8		
18	6	13	9		
20	1	13	0		
18	0	23	0		

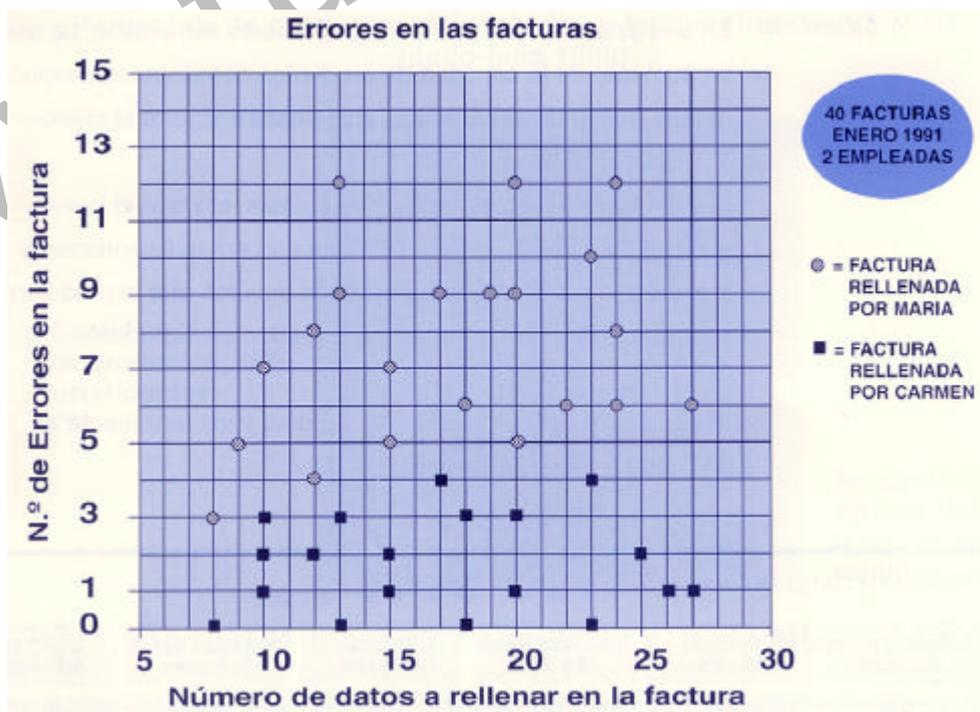
Diagrama de Dispersión



El Diagrama de Dispersión no parecía confirmar la teoría de una relación entre el número de datos a incluir en la factura y la cantidad de errores en la misma.

Una estratificación de los datos por empleada (una tenía mucha más experiencia que la otra) mostró que efectivamente no existía la correlación buscada pero si una clara diferencia en el número de errores entre las dos.

Facturas rellenas por Carmen		Facturas rellenas por María	
Número de datos en la factura	Número de errores en la factura	Número de datos en la factura	Número de errores en la factura
15	2	8	3
25	2	15	5
20	3	12	4
18	3	10	7
23	4	17	9
27	1	18	6
8	0	15	7
20	1	20	5
18	0	9	5
26	1	20	9
10	2	23	10
10	3	20	12
13	3	13	12
13	0	24	8
23	0	13	9
15	1	22	6
10	1	24	6
12	2	19	9
17	4	12	8
		27	6
		24	12



El proporcionar más formación a la empleada con menor experiencia llevó a una sensible reducción de los errores en las facturas.

Ejemplo 2 Temperatura en almacén

Situación

En una empresa que producía un producto alimenticio, se detectó un aumento de la cantidad de productos deteriorados después de una noche de almacenaje, antes del transporte al cliente.

Una de las teorías sobre posibles causas, era que el nuevo sistema de climatización del almacén no era suficientemente preciso y la temperatura superaba la máxima que el producto soportaba (temperatura máxima 5 grados centígrados).

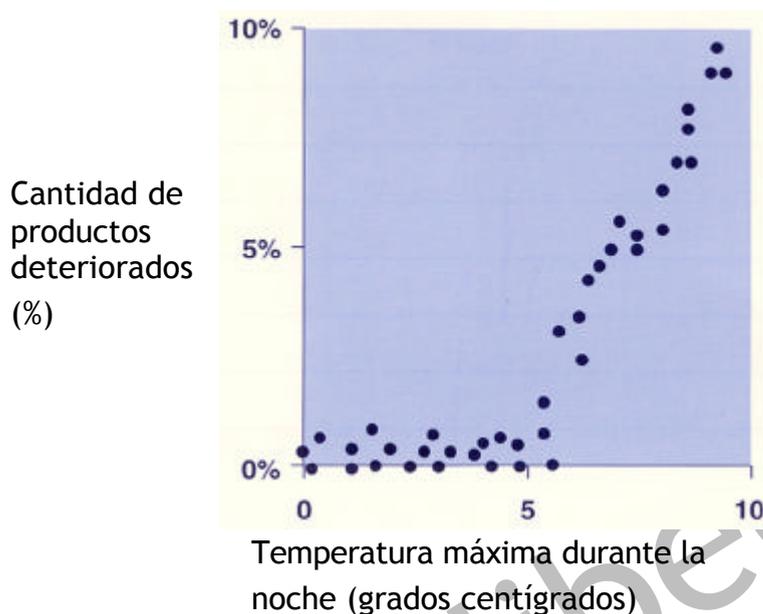
El equipo recogió datos durante 40 días y los representó en un Diagrama de Dispersión.

Tabla de los datos recogidos

Temperatura máxima en la noche	% Productos deteriorados	Temperatura máxima en la noche	% Productos deteriorados	Temperatura máxima en la noche	% Productos deteriorados
0,0	0,2	4,1	0,0	7,0	5,3
0,2	0,0	4,3	0,6	7,4	6,5
0,3	0,5	4,8	0,5	7,7	5,5
1,0	0,0	4,9	0,0	8,0	7,0
1,0	0,3	5,1	1,2	8,2	7,8
1,5	0,0	5,2	0,8	8,4	8,3
1,5	0,6	5,4	0,0	8,5	7,1
2,0	0,4	5,5	3,0	9,0	9,0
2,3	0,0	5,9	3,1	9,2	9,0
2,6	0,4	6,0	2,0	9,2	9,6
3,0	0,0	6,0	4,3		
3,0	0,7	6,1	4,7		
3,2	0,3	6,2	5,0		
3,9	0,3	6,5	5,8		
4,0	0,5	7,0	5,0		

Diagrama de dispersion:

Porcentaje de productos deteriorados y temperatura en almacén



Como se puede ver, la temperatura máxima de 5 grados centígrados se superó en 21 de 40 noches.

El Diagrama muestra una fuerte correlación entre la temperatura máxima de la noche y la cantidad de productos deteriorados, que se consiguió bajar casi a cero con otro nuevo sistema de refrigeración que garantizaba constantemente una temperatura menor de 5 grados centígrados.

7.- UTILIZACIÓN DE LA HERRAMIENTA

<i>La herramienta es muy útil para:</i>	<i>La herramienta es útil para:</i>
<ul style="list-style-type: none">- Determinación de causas- Diseño de soluciones y controles	<ul style="list-style-type: none">-Priorización de Causas

8.- RELACIÓN CON OTRAS HERRAMIENTAS

<i>La herramienta está fuertemente relacionada con:</i>	<i>La herramienta está débilmente relacionada con:</i>
<ul style="list-style-type: none">- Diagrama Matricial	<ul style="list-style-type: none">- Hojas de Comprobación y Recogida de Datos- Diseño de Experimentos