

Relación entre variables

Todos los días, los administradores toman decisiones personales y profesionales basadas en predicciones de sucesos futuros. Para hacer estos pronósticos, se basan en la relación (intuitiva y calculada) entre lo que ya se sabe y lo que se debe estimar. Si los responsables de la toma de decisiones pueden determinar cómo lo conocido se relaciona con el evento futuro, pueden ayudar considerablemente al proceso de toma de decisiones. Ése es el objetivo de este capítulo: cómo determinar la relación entre variables.

Diferencia entre la ji-cuadrada y los temas de este capítulo

pruebas de ji-cuadrada de independencia para determinar si existía una relación estadística entre dos variables. La prueba ji-cuadrada nos dice si existe tal relación, pero no nos dice cuál es esa relación. Los análisis de regresión y correlación nos mostrarán cómo determinar tanto la naturaleza como la fuerza de una relación entre dos variables. De esta forma, aprenderemos a pronosticar, con cierta precisión, el valor de una variable desconocida basándonos en observaciones anteriores de ésta y otras variables.

Origen de los términos regresión y regresión múltiple

El término regresión fue utilizado por primera vez como un concepto estadístico en 1877 por sir Francis Galton, quien llevó a cabo un estudio que mostró que la estatura de los niños nacidos de padres altos tiende a retroceder o “regresar” hacia la estatura media de la población. Designó la palabra regresión como el nombre del proceso general de predecir una variable (la estatura de los niños) a partir de otra (la estatura del padre o de la madre). Más tarde, los estadísticos acuñaron el término regresión múltiple para describir el proceso mediante el cual se utilizan varias variables para predecir otra.

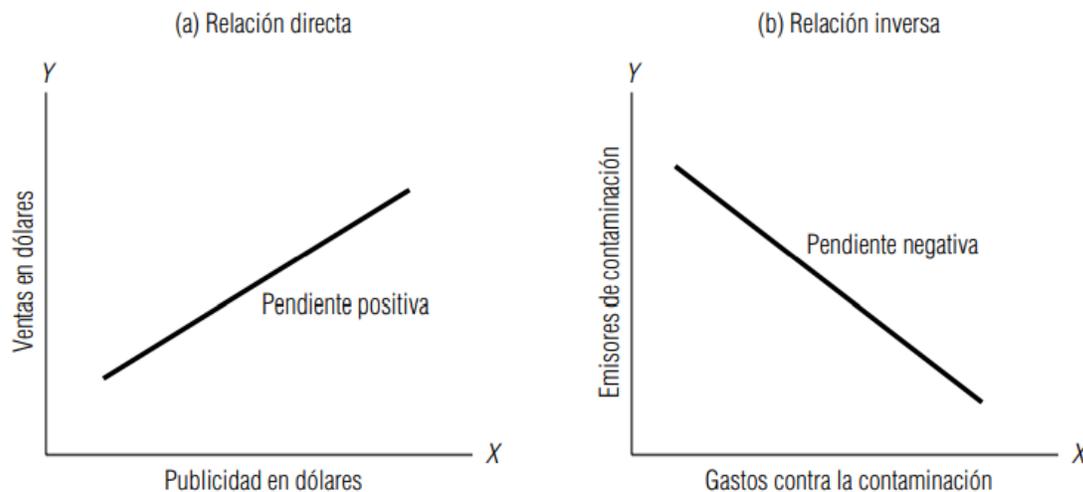
Desarrollo de una ecuación de estimación

En el análisis de regresión, desarrollaremos una ecuación de estimación, esto es, una fórmula matemática que relaciona las variables conocidas con la variable desconocida. Después de conocer el patrón de esta relación, podremos aplicar el análisis de correlación para determinar el grado en el que las variables se relacionan. El análisis de correlación, entonces, nos indica qué tan bien la ecuación de estimación describe realmente la relación.

Tipos de relaciones

Variables independientes y dependientes

Los análisis de regresión y de correlación se basan en la relación, o asociación, entre dos (o más) variables. La variable (o variables) conocida(s) se llaman variable(s) independiente(s); la que tratamos de predecir es la variable dependiente.



Relaciones directas e inversas entre la variable independiente X y la variable dependiente Y

Los científicos saben, por ejemplo, que existe una relación entre las ventas anuales de latas de aerosoles y la cantidad de fluorocarburos liberados a la atmósfera cada año. Si estudiáramos esta relación, “el número de latas de aerosol vendidas cada año” sería la variable independiente y “la

cantidad de fluoro carburos liberados anualmente” sería la variable dependiente.

Relación directa entre X y Y

Al incrementarse la variable independiente, la variable dependiente también lo hace. De manera similar, esperamos que las ventas de una compañía se incrementen al aumentar el presupuesto de publicidad. Podemos graficar una relación directa de este tipo colocando la variable independiente en el eje X y la variable dependiente.

Relación inversa entre X y Y

Las relaciones pueden ser inversas en vez de directas. En estos casos, la variable dependiente disminuye al aumentar la variable independiente. El gobierno supone que existe una asociación inversa entre un mayor gasto anual de una compañía en dispositivos anticontaminantes y menores emisiones contaminantes. La gráfica (b) de la figura 12-1 ilustra este tipo de relación, que se caracteriza por una pendiente negativa (la variable dependiente Y disminuye al aumentar la variable independiente X).

A menudo encontramos una relación causal entre variables, esto es, la variable independiente “causa” cambios en la variable dependiente. Éste es el caso en el ejemplo de la contaminación. Pero en muchos casos, otros factores ocasionan los cambios tanto en las variables dependientes como en las independientes. Podríamos predecir las ventas de aretes de diamantes observando las de Cadillac nuevos, pero no podríamos decir que una origina a la otra. Más bien, nos damos cuenta que otro factor, como el nivel de ingresos disponibles, es la causa de los niveles de ventas tanto de Cadillac como de aretes de diamantes.

Relaciones de asociación, no de causa y efecto

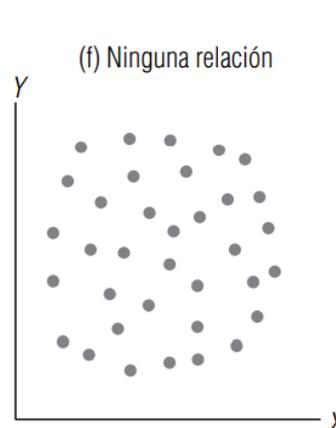
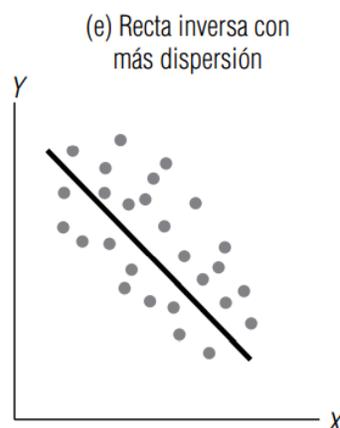
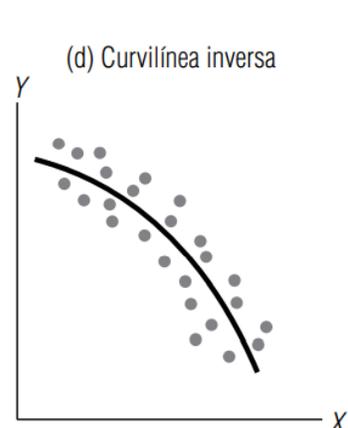
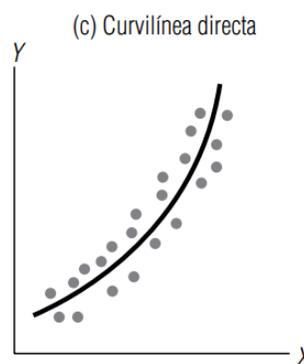
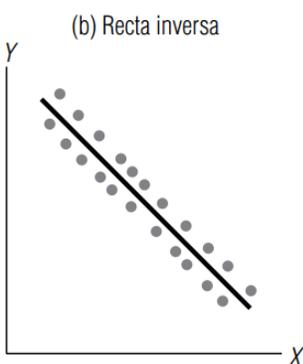
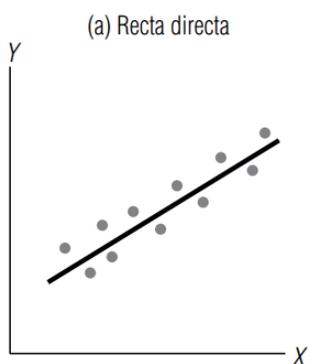
Por esta razón, es importante considerar que las relaciones encontradas por la regresión son relaciones de asociación, pero

no necesariamente de causa y efecto. A menos que tenga razones específicas para creer que los valores de la variable dependiente se originan por los valores de las variables independientes, no infiera causalidad en las relaciones encontradas por la regresión.

Diagramas de dispersión

El primer paso para determinar si existe una relación entre dos variables es examinar la gráfica de los datos observados (o conocidos). Esta gráfica, o dibujo, se llama diagrama de dispersión.

Estudiante	A	B	C	D	E	F	G	H
Calificaciones de examen de admisión (100 = máxima calificación posible)	74	69	85	63	82	60	79	91
Promedio general acumulado (4.0 = A)	2.6	2.2	3.4	2.3	3.1	2.1	3.2	3.8



Un instructor está interesado en saber cómo se relaciona el número de estudiantes ausentes con la temperatura media del día. Usó una muestra aleatoria de 10 días para el estudio. El siguiente número de estudiantes ausentes (AUS) y la temperatura media (TEMP) para cada día.

ausente	8	7	5	4	2	3	5	6	8	9
tiempo	10	20	25	30	40	45	50	55	59	60

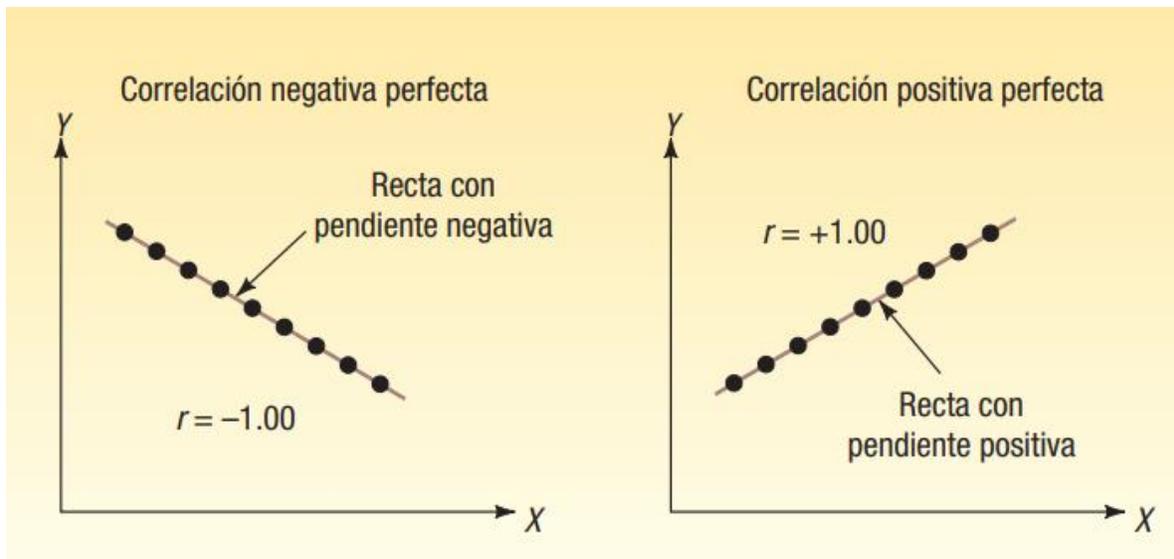
- Establezca la variable dependiente (Y) y la variable independiente (X).
- Dibuje un diagrama de dispersión para estos datos.
- ¿La relación entre las variables parece lineal o curvilínea?
- ¿Qué tipo de curva puede dibujar a través de los datos?
- ¿Cuál es la explicación lógica para la relación observada?

¿Qué es el análisis de correlación?

Grupo de técnicas para medir la asociación entre dos variables.

¿Hay alguna relación entre el número de horas que estudiaron los alumnos para un examen y la calificación que obtuvieron?

Coeficiente de correlación.



COEFICIENTE DE CORRELACIÓN Medida de la fuerza de la relación lineal entre dos variables.

CARACTERÍSTICAS DEL COEFICIENTE DE CORRELACIÓN

1. El coeficiente de correlación de la muestra se identifica con la letra minúscula r .
2. Muestra la dirección y fuerza de la relación lineal (recta) entre dos variables en escala de intervalo o en escala de razón.
3. Varía de -1 hasta $+1$, inclusive.
4. Un valor cercano a 0 indica que hay poca asociación entre las variables.
5. Un valor cercano a 1 indica una asociación directa o positiva entre las variables.

6. Un valor cercano a -1 indica una asociación inversa o negativa entre las variables.

¿Cómo se determina el coeficiente de correlación?

Traza el diagrama de dispersión que corresponda a los datos

Semanas de experiencia	7	9	6	14	8	12	10	4	2	11	1	3
Numero de rechazos	26	20	28	16	23	18	24	26	38	22	32	25

que se proporcionan a continuación. Al observar la nube de puntos resultante, indica el tipo de relación que existe entre las variables.

Los datos relacionan el numero de semana de experiencia en un trabajo de instalación de cables de componentes electrónicos en miniaturas, y el numero de componentes que se rechazaron por presentar algún tipo de defecto.

Mediante el análisis del diagrama de dispersión identificar el tipo de correlación que se presenta en cada uno de los siguientes casos.

El editor de un periódico esta convencido de que, cuando las prensas trabajan, el grado de ruido crea niveles no saludables de tensión y ansiedad, por lo que un psicólogo realizo una prueba durante la cual los prensistas se situaron en cuartos con niveles variable de ruido, y luego se midió el nivel de ansiedad y nerviosismo que presentaban. La siguiente tabla muestra los resultados obtenidos.

Ruido	4	3	1	2	6	7	2	3
Ansiedad	39	38	16	18	41	45	25	38

En pruebas diseñadas para medir el efecto de cierto aditivo en el tiempo de secado de pintura se obtuvieron los siguientes datos:

Concentración del aditivo (%)	4.0	4.2	4.4	4.6	4.8	5.0	5.2	5.4
Tiempo de secado	8.7	8.8	8.3	8.7	8.1	8.0	8.1	7.7

Los siguientes datos representan la temperatura diaria promedio, en grados Fahrenheit, durante un periodo de seis días y el numero correspondiente de luces de tráfico defectuosas.

Temperatura (°F)	72	78	75	74	78	76
Luces defectuosas	4	5	5	2	2	3