ANÁLISIS DE REGRESIÓN SIMPLE Y CORRELACIÓN

Introducción a la unidad

El uso de la regresión lineal simple es muy utilizado para observar el tipo de relación que existe entre dos variables y poder llevar a cabo la toma de decisiones correspondiente dependiendo de la relación entre dichas variables, así por ejemplo, pudiera darse el caso en el que después de aplicar la regresión lineal no exista relación entre las variables involucradas y en consecuencia la decisión podría ser buscar cuál es la variable independiente que tiene influencia sobre la dependiente y volver a realizar el estudio completo; pero si fuera el caso en el cual si existiera una relación positiva entre las variables involucradas, la obtención del coeficiente de correlación nos daría más información sobre el porcentaje de relación existente y pudiendo determinar si es necesario la inclusión de otra variable independiente en el problema mismo, para lo cual el análisis de regresión ya sería del tipo múltiple.

Objetivo particular de la unidad

Analizar los conceptos fundamentales de regresión simple y correlación, su desarrollo y aplicación dentro del ámbito empresarial.

Lo que sé

Elige la respuesta correcta a las siguientes preguntas, una vez que concluyas, obtendrás de manera automática tu calificación.

1. Es una condición para determinar la ecuación de una recta:
   - a) conocer la pendiente de la ordenada al origen
   - b) conocer la pendiente y la ordenada al origen de la recta misma
   - c) conocer dos ordenadas al origen de la recta misma

2. La pendiente de una recta nos indica:
   - a) si la recta pasa por el origen
   - b) si la recta se encuentra en un cuadrante en particular
   - c) la inclinación de la recta

3. En la ecuación de una recta, la ordenada al origen nos indica:
   - a) el punto donde la recta intersecta al eje “x”
   - b) un punto fuera del plano
   - c) el punto donde la recta intersecta al eje “y”

4. Cuando se dice que la relación entre dos variables es de tipo lineal, sabemos que la grafica de su relación es
   - a) una línea recta
   - b) una parábola
   - c) una circunferencia

5. De las siguientes ecuaciones, cuál representa una línea recta:

- a) $x^2 + y^2 = 1$
- b) $y = mx + b$
- c) $y = mx^2 + b$

Temas de la unidad VI

1. Modelo lineal simple
2. Método de mínimos cuadrados
3. Inferencias relativas a la pendiente de la recta de regresión
4. Predicción de un valor particular de "y" para un valor dado de "x"
5. Coeficiente de correlación y coeficiente de determinación
6. Inferencias relativas al coeficiente de correlación

Resumen de la unidad

En la presente unidad se explicó el uso del análisis de regresión como medio para determinar como se relacionan una variable dependiente (y) con una variable independiente (x); en la regresión lineal la recta de regresión: $\hat{y}_i = b_0 + b_1 X_i$ describe como se relaciona el valor esperado de (y) con el valor correspondiente de (x). Esta recta se obtiene utilizando datos de una muestra y el método de mínimos cuadrados.

El coeficiente de determinación se interpreta como una medida de la bondad del ajuste para la recta de regresión y, en concreto nos indica la proporción de la variación de la variable dependiente que es explicada por la recta de regresión obtenida. El coeficiente de correlación lo interpretamos como una medida descriptiva de la intensidad de la relación entre ambas variables.

Durante el desarrollo de la unidad se explico de forma detallada un caso práctico para entender desde la obtención de la recta de regresión y los coeficientes de determinación y correlación, hasta la interpretación de estos mismos conceptos.

Terminamos la unidad con un caso práctico que integra el conocimiento adquirido durante la misma y que además pone a prueba la pericia del alumno en el uso de la calculadora.

Tema 1. Modelo lineal simple

Objetivo del tema

Reconocer al resultado de la relación bivariada es un modelo lineal simple.

Desarrollo

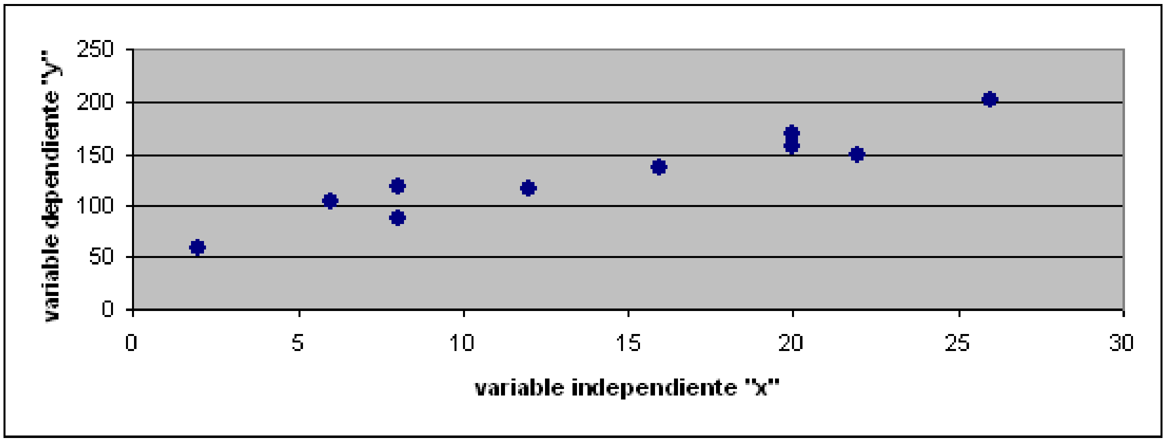
El **análisis de regresión lineal o bivariada**[1] es un procedimiento estadístico que sirve para estudiar la relación entre dos variables cuando una se considera como variable dependiente y la otra como variable independiente. Por ejemplo, podría ser de interés analizar la relación entre las ventas (variable dependiente) y la publicidad (variable independiente). Si el investigador estima la relación entre los gastos publicitarios y las ventas mediante el análisis de regresión, podrá predecir las ventas para diferentes niveles publicitarios. Cuando se emplean dos o más variables independientes en el problema (tales como la publicidad y el precio del producto) para pronosticar la variable dependiente de interés, se aplica el **análisis de regresión múltiple**.

Naturaleza de la relación[2]

Para estudiar la naturaleza de la relación entre la **variable dependiente y la independiente** se construye un **diagrama de dispersión**. La variable dependiente "**y**" se grafica en el eje vertical y la variable independiente "**x**" en el eje horizontal. Al examinar el diagrama de dispersión se ve si la relación entre las dos variables, en caso de que exista, es lineal o curva. Si la relación parece lineal o está cerca de ella, puede aplicarse la regresión lineal. Cuando se observa una relación no lineal en el diagrama de dispersión se emplean técnicas de regresión no lineal para la adaptación a una curva, en cuyo caso se utilizan técnicas que se encuentran más allá del alcance de este análisis.

[1] Carl McDaniel, y Roger Gates, *Investigación de mercados contemporánea*, p. 558.
[2] Op. Cit

Diagrama de dispersión que muestra la relación entre la variable dependiente e independiente



ACTIVIDAD 1

Discute en un foro las ventajas y posibles desventajas de la regresión lineal.

Pulsa el botón **Colocar un nuevo tema de discusión aquí**; pon en el apartado **Asunto** el título de tu aportación, redacta tu comentario en el área de texto y haz clic en **Enviar al foro.**

ACTIVIDAD 2

En periódicos como “el economista” o “el financiero” observa algún problema social donde consideres que se puede utilizar el análisis de regresión lineal simple y define las variables dependiente e independiente que estarían involucradas en el problema.

Envía al profesor tanto el problema como tus propuestas para que éste las evalúe

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar**, localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Elige la respuesta correcta a las siguientes preguntas, una vez que concluyas, obtendrás de manera automática tu calificación.

1. El análisis de regresión lineal o divariada

- ❍ a) es un procedimiento estadístico que sirve para estudiar la relación entre dos variables cuando una se considera como variable dependiente y la otra como variable independiente.
- ❍ b) es un procedimiento estadístico que sirve para estudiar la relación entre una variable dependiente con dos o más variables independientes.
- ❍ c) es un procedimiento estadístico que sirve para estudiar la relación entre dos variables dependientes con una variable independiente.

2. Cuando se emplean dos o más variables independientes en el análisis de regresión (tales como la publicidad y el precio del producto) para pronosticar la variable dependiente de interés, se aplica:

- ❍ a) el análisis de regresión múltiple.
- ❍ b) el análisis de regresión lineal.
- ❍ c) el análisis de regresión bivariado.

3. En el análisis de regresión lineal, un diagrama de dispersión sirve para:

- ❍ a) estudiar una muestra que es completamente homogénea.
- ❍ b) estudiar la naturaleza
- ❍ c) estudiar la naturaleza de la relación entre la variable dependiente y la independiente.

Tema 2. Método de mínimos cuadrados

Objetivos del tema

Reconocer el método de mínimos cuadrados como el proceso que define a la recta de regresión.

Desarrollo

Cualquier método estadístico que busque establecer una ecuación que permita estimar el valor desconocido de una variable, a partir del valor conocido de una o más variables, se denomina **análisis de regresión**.

El método de **mínimos cuadrados**, es un procedimiento para encontrar la ecuación de regresión que se origina al estudiar la relación estocástica que existe entre dos variables. Fue Karl Friedrich Gauss (1777-1855) quien propuso el método de los mínimos cuadrados y fue el primero en demostrar que la ecuación estimada de regresión minimiza la suma de cuadrados de errores.

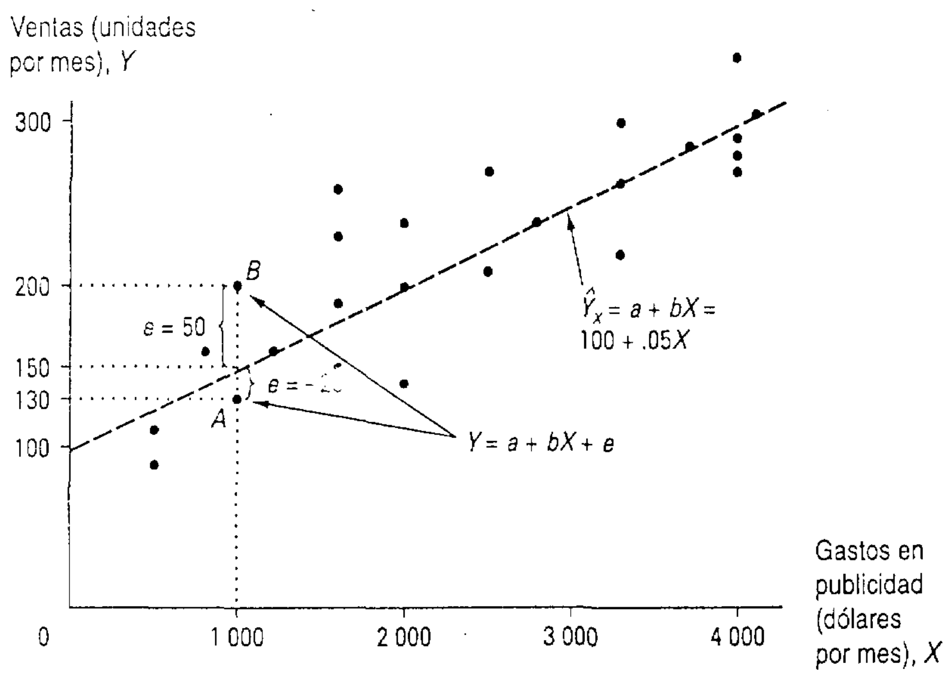
En el análisis de regresión [3], una variable cuyo valor se suponga conocido y que se utilice para explicar o predecir el valor de otra variable de interés se llama **variable independiente** y se simboliza por "X". Por el contrario, una variable cuyo valor se suponga desconocido y que se explique o prediga con ayuda de otra se llama **variable dependiente** y se simboliza por "Y".

Una **relación estocástica** [4] entre dos variables cualesquiera, *x* y *y*, es imprecisa en el sentido de que muchos valores posibles de "y" se pueden asociar con cualquier valor de "*x*". Sin embargo, un resumen gráfico de la relación estocástica entre la variable independiente "*x*" y la variable dependiente "*y*" estará dado por

[3] Heinz Kohler, *Estadística para negocios y economía,* pp. 528-529.
[4] Heinz Kohler, *Estadística para negocios y economía,* p. 530.

una línea de regresión, misma que reduce al mínimo los errores cometidos cuando la ecuación de esa línea se utilice para estimar y a partir de *x*.

Grafica que muestra la relación existente entre los gastos de publicidad y las ventas.



De esta gráfica podemos ver claramente que las ventas dadas en unidades por mes (variable dependiente) en este caso, si guardan relación con los gastos en publicidad y, que dicha relación puede ser denotada por la "recta de regresión"

De este análisis de relación estocástica que se da entre dos variables, surgen las ecuaciones que nos provee el método de mínimos cuadrados, que a saber son:

Ecuación de la recta de regresión: $\hat{y}_i = b_0 + b_1 X_i$

En la que:

x_i = es un valor dado de la variable independiente para el cual se quiere estimar el valor correspondiente de la variable dependiente

b_0 = ordenada al origen de la línea estimada de regresión,

b_1 = pendiente de la línea estimada de regresión,

$\hat{Y}_i$ = valor estimado de la variable dependiente, para el i-ésimo valor de la variable independiente

Resulta claro que para poder determinar la recta de regresión, es necesario que antes sean calculados los valores correspondientes a la pendiente de la recta y a la ordenada al origen.

La pendiente de la recta de regresión se calcula mediante la siguiente formula:

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}}$$

y la ordenada al origen se calcula mediante la formula:

$$b_0 = \overline{Y} - b_1 \overline{X}$$

Antes de continuar, es necesario advertir que el análisis de regresión no se puede interpretar como un procedimiento para establecer una relación de causa a efecto entre variables. Sólo puede indicar cómo o hasta qué grado las variables están asociadas entre sí. Cualquier conclusión acerca de causa y efecto se debe basar en el juicio del o los individuos con más conocimientos sobre la aplicación. Por ejemplo, un estadista puede llegar a determinar que la relación entre las ventas y el presupuesto asignado a mercadotecnia es positiva y que se tiene un coeficiente de correlación de 0.96, lo cual prácticamente nos indica que es recomendable incrementar el presupuesto al departamento de mercadotecnia para obtener mejores ingresos dentro de la compañía, sin embargo el director de operaciones puede llegar a determinar que debido a condiciones internas del país en el que se encuentre la empresa, o bien la aparición de una nueva ley que regule los medios utilizados por el mencionado departamento de mercadotecnia, pueden llegar a frenar o incluso generar conflictos dentro de la empresa si incrementamos el presupuesto al departamento correspondiente.

ACTIVIDAD 1

Explica los pasos a seguir utilizando el método de mínimos cuadrados.

Para enviar tu actividad, pulsa **Editar mi envío** y se mostrará un editor de texto en el que deberás redactar tu información. Cuando termines, guarda tu tarea haciendo clic en **Guardar cambios**.

ACTIVIDAD 2

Explica las formulas utilizadas para la pendiente de la recta en los libros de Lind y de Berenson de la bibliografía son equivalentes a las aquí presentadas.

Para enviar tu actividad, pulsa **Editar mi envío** y se mostrará un editor de texto en el que deberás redactar tu información. Cuando termines, guarda tu tarea haciendo clic en **Guardar cambios**.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Elige la respuesta correcta a las siguientes preguntas, una vez que concluyas, obtendrás de manera automática tu calificación.

1. ¿Es un procedimiento para encontrar la ecuación de regresión que se origina al estudiar la relación estocástica que existe entre dos variables?

- a) La suavización exponencial
- b) Una serie de tiempo
- c) El método de mínimos cuadrados

2. Es una variable cuyo valor se supone conocido y que se utiliza para explicar o predecir el valor de otra variable de interés

- a) variable independiente
- b) variable dependiente
- c) variable estocastica

3. Es una variable cuyo valor se supone desconocido y que se explica o predice con ayuda de otra variable.

- a) variable independiente
- b) variable dependiente
- c) variable estocastica

4. ¿Es el símbolo que se utiliza comúnmente para la variable independiente?

- a) “X”.
- b) “Y”.
- c) “Z”

5. Propuso el método de los mínimos cuadrados y fue el primero en demostrar que la ecuación estimada de regresión minimiza la suma de cuadrados de errores.

- a) Francis Galton
- b) Karl Friedrich Gauss (1777-1855)
- c) Kruskal-Wallis

6. Una relación estocástica entre dos variables cualesquiera, *x* y *y*, es imprecisa en el sentido de que muchos valores posibles de "y" se pueden asociar con cualquier valor de "*x*". Sin embargo, un resumen gráfico de la relación estocástica entre la variable independiente "*x*" y la variable dependiente "*y*" estará dado por

- a) Una circunferencia, porque circunscribe todos los valores de variable independiente "x"
- b) Una parábola, porque acota todos los valores posibles de la variable independiente "x"
- c) Una línea de regresión, misma que reduce al mínimo los errores cometidos cuando la ecuación de esa línea se utilice para estimar y a partir de *x*.

7. La fórmula para determinar la pendiente de la recta de regresión es:

- a) $\hat{y}_i = b_0 + b_1 X_i$
- b) $b_1 = \dfrac{\sum_{i=1}^{n} X_i Y_i - \dfrac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \dfrac{(\sum_{i=1}^{n} X_i)^2}{n}}$
- c) $b_0 = \overline{Y} - b_1 \overline{X}$

8. La fórmula para determinar la ordenada al origen de la recta de regresión es:

- a) $\hat{y}_i = b_0 + b_1 X_i$
- b) $b_1 = \dfrac{\sum_{i=1}^{n} X_i Y_i - \dfrac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \dfrac{(\sum_{i=1}^{n} X_i)^2}{n}}$
- c) $b_0 = \overline{Y} - b_1 \overline{X}$

9. La fórmula que caracteriza la recta de regresión es:

- a) $\bar{x} = \dfrac{\sum_{i=1}^{i=n} x_i}{n}$
- b) $\hat{y}_i = b_0 + b_1 X^2_i$
- c) $\hat{y}_i = b_0 + b_1 X_i$

Tema 3. Inferencias relativas a la pendiente de la recta de regresión

Objetivos del tema

Analizar el impacto de las diferentes pendientes que puede tener una recta de regresión.

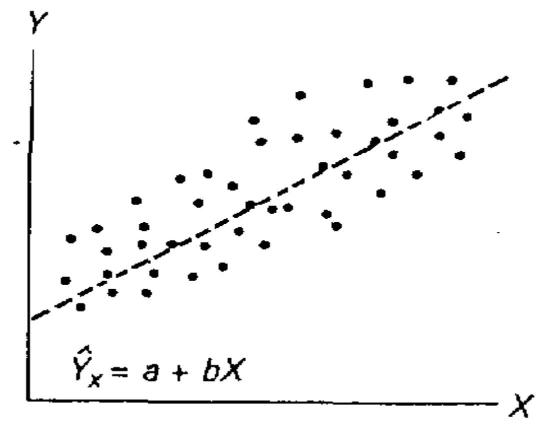
Desarrollo

Las inferencias acerca de la pendiente de la recta de regresión son importantes dado que la relación entre las dos variables en cuestión depende de ella precisamente, es decir, si la pendiente de la recta de regresión es positiva, entonces la naturaleza de la relación entre ambas variables será positiva, y la pendiente de la recta es negativa, entonces la relación entre las variables será negativa también, con lo cual podemos iniciar la toma de decisiones dependiendo del contexto del problema mismo. Como se menciono anteriormente la ecuación de la recta de regresión: $\hat{y}_i = b_0 + b_1 X_i$

b_0 representa la ordenada al origen de la línea estimada de regresión, y
b_1 es la pendiente de la línea estimada de regresión.

Donde $\mathbf{b_0}$ es en sí, el punto donde la recta corta al eje de las "x" y $\mathbf{b_1}$ nos da el grado de inclinación de la recta, de tal forma que cuando la pendiente de la recta es positiva, se dice que la relación que existe entre las dos variables dependiente e independiente es de naturaleza positiva, es decir, que posee una grafica como la indicada a continuación:

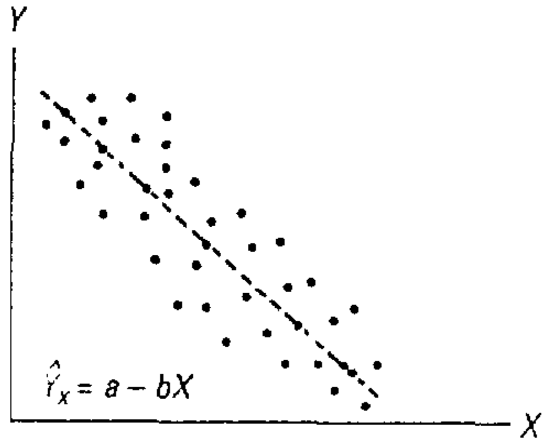
Relación positiva entre dos variables en regresión lineal



En este tipo de relación, los incrementos en los valores de la variable independiente traen como consecuencia un incremento en los valores correspondientes de la variable dependiente y la grafica tiene como podemos apreciar una forma ascendente.

Pero cuando la pendiente de la recta de regresión es negativa, es decir, que dicha ecuación tuviera la forma $\hat{y}_i = b_0 - b_1 X_i$ entonces la relación existente entre las variables es de tipo negativa, lo cual quiere decir, que a incrementos en los valores de la variable independiente, la variable dependiente responde con decrementos; la grafica resultante tendría la forma siguiente:

Relación negativa entre dos variables en regresión lineal.



En esta grafica podemos observar que la tendencia de la recta de regresión es descendente, lo cual implica como ya habíamos mencionado, que la relación entre ambas variables es negativa.

ACTIVIDAD 1

Haz un cuadro comparativo con los libros de Anderson y Black de lo que significa que la pendiente de la recta de regresión tenga un valor igual a cero.

	Significado de pendiente igual a cero
Anderson	
Black	

Descarga el siguiente cuadro para completarlo, una vez que lo tengas listo presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.

ACTIVIDAD 2

Si en un análisis de regresión lineal se dice que las variables no tienen relación alguna, ¿Qué valor tendrá la pendiente?

Pulsa el botón **Colocar un nuevo tema de discusión aquí;** pon en el apartado **Asunto** el título de tu aportación, redacta tu comentario en el área de texto y haz clic en **Enviar al foro**.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Elige la respuesta correcta a las siguientes preguntas, una vez que concluyas, obtendrás de manera automática tu calificación.

1. ¿Por qué son importantes las inferencias acerca de la pendiente de la recta de regresión?
   - a) porque de ella depende la relación entre las variables en cuestión
   - b) porque matemáticamente es obligatorio calcularla.
   - c) porque nos indica el punto donde la recta de regresión intersecta al eje de las “y”.

2. ¿Cuándo la pendiente de la recta de regresión es positiva, la relación entre las variables es?
   - a) negativa
   - b) cero
   - c) positiva

3. ¿Cuándo la pendiente de de la recta de regresión es negativa, la relación entre las variables es?
   - a) cero
   - b) negativa
   - c) positiva

4. ¿Es el símbolo comúnmente utilizado para denotar ala pendiente de la recta de regresión?:
   - a) b_0
   - b) b_1
   - c) b_2

Tema 4. Predicción de un valor particular de “y” para un valor dado de “x”

Objetivos del tema

Predecir un valor de de la variable dependiente “y” para un valor particular de la variable independiente “x” utilizando el análisis de regresión lineal.

Desarrollo

Para predecir un valor particular de “y” para un valor dado de “x” utilizaremos el análisis de regresión simple y para ilustrarlo consideremos el siguiente ejemplo: Domin’s Pizza es una cadena de restaurantes dedicado exclusivamente a la distribución de pizzas. El director general cree que los lugares donde sus establecimientos han tenido más éxito están cercanos a establecimientos de educación superior y para sustentar su creencia ha solicitado un estudio de las ventas de sus restaurantes asociadas al tamaño de la población estudiantil de los centros educativos correspondientes.

Los administradores creen que las ventas mensuales en esos restaurantes (representadas por “y”), se relacionan en forma positiva con la población estudiantil (representada por “x”). Esto es, que los restaurantes cercanos a centros escolares con gran población tienden a generar más ventas que los que están cerca de centros con población pequeña.

Para ilustrarlo, supongamos que en el caso de Domin’s Pizza se reunieron datos de una muestra de 10 restaurantes ubicados cerca de centros educativos.

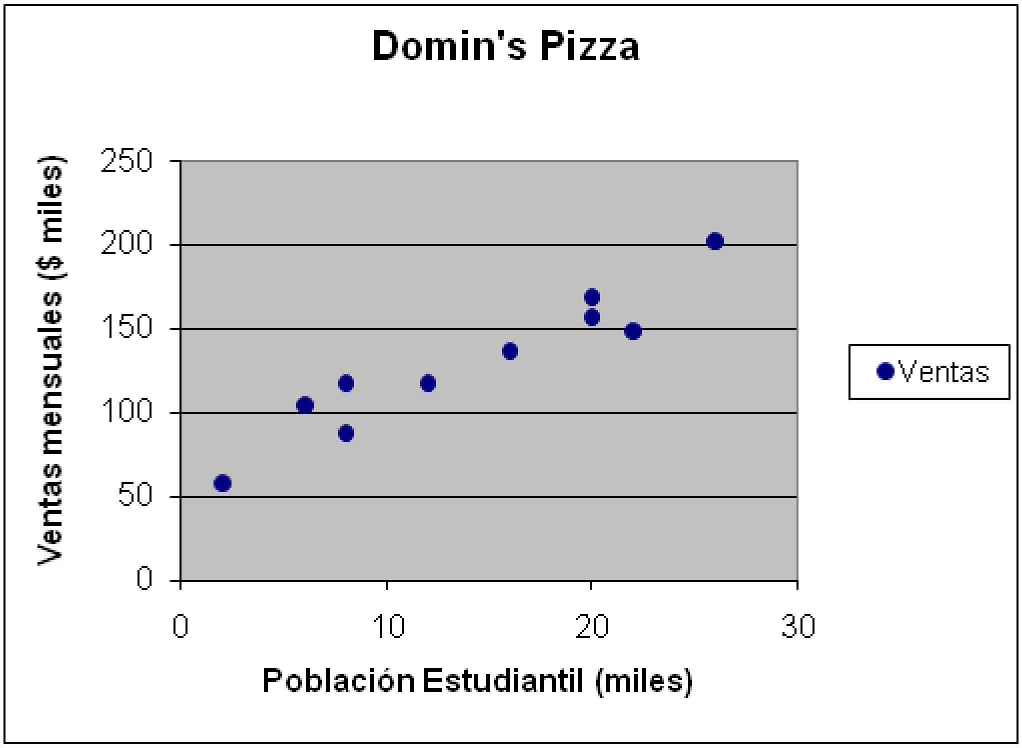
Para el iésimo restaurante de la muestra, x_i es el tamaño de la población estudiantil, en miles, y y_i son las ventas mensuales, en miles de pesos. Los valores de x_i y y_i para los 10 restaurantes de la muestra se resume en la siguiente tabla:

Restaurante i	Población de estudiantes x_i (miles)	Ventas mensuales y_i ($ miles)
1	2	58
2	6	105
3	8	88
4	8	118
5	12	117
6	16	137
7	20	157
8	20	169
9	22	149
10	26	202

Datos de población estudiantil y ventas trimestrales para 10 restaurantes de Domin's Pizza.

Lo primero que hay que hacer para resolver un problema de regresión lineal, es definir nuestras variables involucradas (variable dependiente y variable independiente); en este caso particular los administradores de Domin`s Pizza ya lo han hecho, definiendo las ventas mensuales de los restaurantes como la variable dependiente y la población estudiantil dada en miles como la variable independiente, debido a que claro esta se supone que las ventas de los restaurantes dependen de la población estudiantil de los centros educativos.

La siguiente gráfica corresponde al diagrama de dispersión de los datos de la tabla anterior:



¿Cuáles son las conclusiones que podemos sacar de la gráfica anterior?

- Parece que las ventas son mayores en los centros con más población de estudiantes.
- Para esos datos, la relación entre el tamaño de la población de estudiantes y las ventas parece poderse aproximar con una línea recta.
- Parece haber una relación lineal positiva entre “x” y “y”.

En consecuencia, elegimos el modelo de **regresión lineal simple** para esta opción; nuestra siguiente tarea será emplear los datos de la muestra de la tabla para determinar los valores de b_0 y b_1 en la ecuación de regresión lineal simple.

Por lo tanto, para el iésimo restaurante, la ecuación de regresión es: $\hat{y}_i = b_0 + b_1 X_i$

En la que:

x_i = tamaño de la población estudiantil (miles) para el iésimo restaurante

b_0 = ordenada al origen de la línea estimada de regresión.

b_1 = pendiente de la línea estimada de regresión.

$\hat{Y}_i$ = valor estimado de las ventas mensuales, en miles, para el pésimo restaurante.

Para desarrollar nuestros cálculos de manera más sencilla, complementamos la tabla de inicio del problema, misa que quedaría de la siguiente forma:

RESTAURANTE	POBLACIÓN DE ESTUDIANTES	VENTAS TRIMESTRALES			
I	x_i (miles)	y_i ($ miles)	X^2	Y^2	XY
1	2	58	4	3364	116
2	6	105	36	11025	630
3	8	88	64	7744	704
4	8	118	64	13924	944
5	12	117	144	13689	1404
6	16	137	256	18769	2192
7	20	157	400	24649	3140
8	20	169	400	28561	3380
9	22	149	484	22201	3278
10	26	202	676	40804	5252
TOTALES	140	1300	2528	184730	21040

Por lo tanto, calculando la pendiente de la recta tenemos que la formula es:

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}}$$

donde al sustituir los datos obtenidos de la tabla anterior:

$$b_1 = \frac{(21040) - \frac{(140)(1300)}{10}}{(2528) - \frac{(140)^2}{10}}$$

y al realizar los cálculos pertinentes tenemos que:

$$b_1 = 5$$

calculando ahora la ordenada al origen tenemos que la formula es:

$$b_0 = \overline{Y} - b_1 \overline{X}$$

en este caso podemos ver que nos falta el valor de la media en “Y” y el valor de la media de “X”, mismas que tienen un valor de:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

donde al sustituir datos tenemos que:

$$\overline{X} = \frac{140}{10}$$

es decir:

$$\overline{X} = 14$$

y para la media de “Y” tenemos que la formula es:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

donde la sustituir datos vemos que:

$$\overline{Y} = \frac{1300}{10}$$

es decir:

$$\overline{Y} = 130$$

por lo tanto, sabiendo que:

$$b_1 = 5$$

$$\overline{X} = 14$$

$$\overline{Y} = 130$$

podemos sustituir estos valores en la formula:

$$b_0 = \overline{Y} - b_1 \overline{X}$$

obteniendo:

$$b_0 = 130 - (5)(14)$$

finalmente realizando las operaciones indicadas:

$$b_0 = 60$$

por lo tanto, la ecuación de la recta de regresión sería:

$$\hat{y}_i = b_0 + b_1 X_i$$

donde al sustituir valores:

$$\hat{y}_i = 60 + 5X_i$$

Esta recta de regresión nos sirve para predecir cuales serían las ventas mensuales en función del tamaño de la población estudiantil. Por ejemplo, si se planea construir un nuevo centro en el cual la población estudiantil es de aproximadamente 30 mil, entonces el nivel de ventas estimado sería igual a

$$\hat{y}_i = 60 + 5(30)$$

$$\hat{y}_i = 210$$

Es decir, las ventas estimadas para ese restaurante serían de $210,000.00 mensuales (recuerde que las ventas están dadas en miles).

Resulta claro que la predicción de las ventas en función de la cantidad de personas es importante porque me ayuda a tomar decisiones que van desde el importe de renta que puedo pagar por el local donde vaya a ubicar la Pizzería, hasta programar mi presupuesto de insumos, la cantidad de personal del equipo de reparto, etc.

ACTIVIDAD 1

1. Considere que el departamento de recursos humanos de la empresa en la que laboramos está interesada en saber si el monto del salario guarda alguna relación directa con la el ahorro voluntario que los empleados sindicalizados de la empresa presentan. Para ello la empresa ha tomado una muestra aleatoria de 10 empleados quedando los datos obtenidos en la siguiente tabla.

Sueldo del empleado	8000	7000	6500	9200	6000	12000	10300	8700	7500	6250
Ahorro del empleado	4000	2000	3200	4500	1200	1000	2500	1500	1700	2250

¿Para este problema la recta de regresión considerando el sueldo del empleado como variable independiente es?:

- a) Y = -2607.25 + 0.58484 Xi
- b) Y = -3500.50 + 0.51831 Xi
- c) Y = 2687.23 – 0.03711 Xi
- d) Y = 5000 + 0.12581 Xi

2. El análisis de regresión es útil para estimar costos de producción. La siguiente tabla muestra de volúmenes de producción y costos asociados con una operación de manufactura.

Volumen de producción (unidades)	Costo total ($)
400	4000

450	**5000**
550	**5400**
600	**5900**
700	**6400**
750	**7000**

Considerando el costo como variable dependiente, ¿para este problema la recta de regresión es?

- a) Y = 1246.66 + 7.6 Xi
- b) Y = 1256.66 + 7.6 Xi
- c) Y = 1246.66 + 8.6 Xi
- d) Y = 1246.6 + 17.6 Xi

Pulsa el botón **Comenzar** para contestar las preguntas, una vez concluyas pulsa el botón **Enviar todo y terminar**

Autoevaluación

Selecciona si las siguientes aseveraciones son verdaderas (V) o falsas (F). Una vez que concluyas, obtendrás tu calificación de manera automática.

	Verdadera	Falsa
1. $b_1 = \dfrac{\sum_{i=1}^{n} X_i Y_i - \dfrac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \dfrac{(\sum_{i=1}^{n} X_i)^2}{n}}$ es la fórmula para calcular la pendiente de la recta de regresión.	( )	( )
2. $b_0 = \bar{Y} - b_1 \hat{X}$ es la fórmula para calcular la ordenada al origen de la recta de regresión.	( )	( )
3. $\hat{y}_i = b_1 + b_0 X_i$ es la fórmula de la recta de regresión.	( )	( )
4. La ecuación de la recta de regresión sirve para pronosticar el valor que tomaría la variable dependiente cuando se conoce un valor de la variable independiente.	( )	( )
5. Definir incorrectamente las variables dependiente e independiente no afecta el resultado de la regresión	( )	( )

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Tema 5. Coeficiente de correlación y coeficiente de determinación

Objetivos del tema

Calcular los coeficientes de determinación y de correlación.

Desarrollo

El coeficiente de determinación se utiliza para evaluar la bondad de ajuste para la ecuación de regresión y se define como:

$$r^2 = \frac{\text{Suma de Cuadrados de la regresión}}{\text{Suma de cuadrados Totales}} = \frac{SSR}{SST}$$

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{Y} - \bar{Y})^2}{\sum_{i=1}^{n}(Y - \bar{Y}_i)^2}$$

La ecuación anterior se puede interpretar como el porcentaje de variación de la variable Y que se puede explicar con el modelo de regresión; en el ejemplo de las pizzas, $r^2=0.9027$ así que el 90.27% de la variación de las ventas se puede explicar por el modelo de regresión, lo cual hace que sea un buen modelo.

Coeficiente de correlación

Cuando es necesario resumir aún más los datos (de una gráfica por ejemplo) se utiliza un solo número, que de alguna forma mide la fuerza de asociación entre dos variables como son el ingreso real y el nivel de educación escolar en nuestro caso. El análisis de correlación nos ayuda a obtener dicho número que se conoce como: **coeficiente de correlación**. Los valores de coeficiente de correlación siempre están entre –1 y +1 un valor de +1 indica que las dos variables tienen una relación lineal positiva perfecta. Esto es, todos los puntos de datos están en una línea recta con pendiente positiva. Un valor de –1 indica que las variables tienen una relación lineal negativa perfecta, y que todos los puntos de datos están en una recta con pendiente negativa. Los valores del coeficiente de correlación cercanos a cero indican que las variables no tienen relación línea [5].

A continuación presentamos la ecuación para calcular el coeficiente de correlación de la muestra. Si ya se ha hecho un análisis de regresión y se ha calculado el coeficiente de determinación, entonces, el coeficiente de correlación se puede calcular como sigue:

$$r = (signo\ de\ b_1)\sqrt{r^2}$$

donde b1 es la pendiente de la ecuación de regresión.

De esta formula, resulta claro que el signo del coeficiente de correlación es positivo si la ecuación de regresión tiene pendiente positiva (b1 > 0), y negativo si la ecuación de regresión tiene pendiente negativa (b1 < 0).

En nuestro ejemplo de las pizzas tendríamos que:

$$r = (signo\ de\ b_1)\sqrt{r^2}$$

$$r = +\sqrt{0.9027}$$

$$r = +0.9501$$

[5] Anderson, Sweeney & Williams, 1999. *Estadística para administración y economía*, p.p. 555.

ACTIVIDAD 1

Elabora un cuadro comparativo de los rangos que pueden tomar el coeficiente de correlación y el coeficiente de determinación.

	Rango del coeficiente
Coeficiente de correlación	
Coeficiente de determinación	

Descarga el siguiente cuadro para completarlo, una vez que lo tengas listo presiona el botón **Examinar.** Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.

ACTIVIDAD 2

Discute en un foro Coeficiente de correlación y coeficiente de determinación, cuál es la diferencia entre el coeficiente de determinación y el coeficiente de correlación.

Pulsa el botón **Colocar un nuevo tema de discusión aquí;** pon en el apartado **Asunto** el título de tu aportación, redacta tu comentario en el área de texto y haz clic en **Enviar al foro.**

Autoevaluación

Algunas empresas han optado por pagarles a sus directores y principales ejecutivos de acuerdo con las ganancias obtenidas por la empresa. La siguiente tabla es una lista de datos corporativos sobre el cambio porcentual en el rendimiento de las acciones durante un periodo de un año, y el cambio porcentual en la paga a los directores y principales ejecutivos, inmediatamente después de un año.

Elige la respuesta correcta a las siguientes preguntas, una vez que concluyas, obtendrás de manera automática tu calificación.

Empresa	Cambio bianual en el Rendimiento (%)	Cambio en el pago Al ejecutivo (%)
Walt mart	201.3	18
Bodega Aurrera	146.5	28
Grupo Soriana	76.7	10
Comercial Mexicana	158.2	28
Home mart	-34.9	15
Price club	73.2	-9
Sams club	-7.9	-20

1. Para este problema, considerando el cambio en el pago al ejecutivo como variable dependiente: la recta de regresión es:

- a) Y = -1.019795909 - 0.125817275 Xi
- b) Y = 1.019795909 + 0.125817275 Xi
- c) Y = -1.019795909 + 0.125817275 Xi

2. Para este mismo problema, el coeficiente de determinación es:

- a) $r^2 =$ -0.339859862
- b) $r^2 =$ 0.339859862
- c) $r^2 =$ 1.339859862

3. Para el problema en cuestión, el coeficiente de correlación es:

- a) r = -0.58297501
- b) r = 1.58297501
- c) r = 0.58297501

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Tema 6. Inferencias relativas al coeficiente de correlación

Objetivos del tema

Determinar la intensidad de la relación entre las variables involucradas y obtenga inferencias apropiadas sobre la población de interés.

Desarrollo

Cuando se tiene una relación lineal entre dos variables, el coeficiente de determinación y el coeficiente de correlación permiten tener medidas de la intensidad de una relación. El coeficiente de determinación da una medida entre 0 y 1, mientras que el coeficiente de correlación da una medida entre –1 y +1 aunque el coeficiente de correlación se restringe a una relación lineal entre dos variables, el coeficiente de determinación se puede emplear en relaciones no lineales y en relaciones que tengan dos o más variables independientes. En ese sentido, el coeficiente de determinación tiene una aplicabilidad más amplia.

Obtener una ecuación que describa la relación existente entre dos variables en un análisis de regresión lineal es muy importante, descubrir que esa relación es positiva, negativa o inexistente también lo es, pero tener un indicador que nos diga que tan intensa es la relación si que la hay entre las variables en cuestión, no deja de ser importante, además de complementar y sustentar las decisiones que se deriven del análisis de regresión.

Como podemos apreciar en el desarrollo del tema, el análisis de regresión es una herramienta matemática muy importante que nos ayuda a la mejor toma de decisiones en un ámbito como el actual lleno de cambios y de una competencia muy cerrada donde una respuesta rápida a los cambios presentados por el medio empresarial, por el mercado, etc. Puede representar la aparición de nueva competencia o bien la extinción de las empresas.

ACTIVIDAD 1

Explica las implicaciones del signo y valor del coeficiente de determinación del problema resuelto en la autoevaluación.

Para enviar tu actividad, pulsa **Editar mi envío** y se mostrará un editor de texto en el que deberás redactar tu información. Cuando termines, guarda tu tarea haciendo clic en **Guardar cambios.**

ACTIVIDAD 2

Explica las implicaciones del signo y valor del coeficiente de correlación del problema resuelto en la autoevaluación.

Para enviar tu actividad, pulsa **Editar mi envío** y se mostrará un editor de texto en el que deberás redactar tu información. Cuando termines, guarda tu tarea haciendo clic en **Guardar cambios.**

Autoevaluación

Un economista del Departamento del Distrito Federal está preparando un estudio sobre el comportamiento del consumidor. Los datos que obtuvo los plasmo en la siguiente tabla.

Consumidor	1	2	3	4	5	6	7	8	9	10	11	12
Ingreso	24.3	12.5	31.2	28	35.1	10.5	23.2	10	8.5	15.9	14.7	15
Consumo	16.2	8.5	15	17	24.2	11.2	15	7.1	3.5	11.5	10.7	9.2

Elige la respuesta correcta a las siguientes preguntas, una vez que concluyas, obtendrás de manera automática tu calificación.

1. Considerando el consumo como variable dependiente el coeficiente de determinación es:

- a) $r^2 =$ 0.844740208
- b) $r^2 =$ -0.844740208
- c) $r^2 =$ 1.844740208

2. Para el problema anterior, el coeficiente de correlación es:

- a) r =1.919097496
- b) r =-0.919097496
- a) r = 0.919097496

LO QUE APRENDÍ DE LA UNIDAD

Una tienda departamental, está considerando otorgar tarjetas de crédito a sus clientes, para lo cual realiza un estudio con el fin de observar el comportamiento de sus gastos en función de su salario. Los datos obtenidos en una muestra aleatoria de tamaño 11 se encuentran en la siguiente tabla.

Sueldo del cliente	18.0	15.0	19.0	9.2	8.6	12.0	10.7	14.3	17.8	16.0	15.0
Gastos del cliente	14.8	10.4	15.7	7.1	5.3	8.0	8.5	10.2	13.0	14.0	11.3

Nota: tanto el sueldo como los gastos del cliente son mensuales y están dados en miles de pesos.

Haga usted un análisis de regresión, defina las variables involucradas y determine:

a) la pendiente de la recta de regresión (0.60253)

b) la ordenada al origen de la recta de regresión (1.3223)

c) la recta de regresión lineal resultante. (Y = 1.3223 + 0.60253 Xi)

d) el coeficiente de determinación (0.37923)

e) el coeficiente de correlación (0.615817)

f) el pronóstico de gasto para un cliente que gana $21,000.00 ($11,330.76)

En conclusión, para este problema, entre más ganan los empleados, más gastan.

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar,** localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

Glosario de la unidad

Variable dependiente

Es la variable que se predice o se explica. Se representa matemáticamente por "y".

Variable independiente

Es la variable que sirve para predecir o explicar. Se representa matemáticamente por "x".

Regresión lineal simple

Análisis de regresión donde intervienen una variable independiente y una variable dependiente; en ella, la relación entre las variables se aproxima mediante una recta.

Recta de regresión

Estimación hecha a partir de datos de una muestra aplicando el método de mínimos cuadrados para la regresión lineal simple, la ecuación de regresión estimada es: $\hat{y}_i = b_0 + b_1 X_i$

Método de mínimos cuadrados

Procedimiento que se usa para determinar la recta de regresión. Su objeto es minimizar $\sum\left(y_i - \hat{y}_i\right)^2$

Diagrama de dispersión

Gráfica de datos de dos variables en la que la variable independiente está en el eje horizontal y la variable dependiente en el eje vertical.

Coeficiente de determinación

Medida de la bondad del ajuste de la recta de regresión. Se interpreta como la parte de la variación de la variable dependiente “y” que explica la recta de regresión.

Residual i-ésimo

Diferencia entre el valor observado de la variable dependiente y el valor predicho usando la recta de regresión; para la i-ésima observación, el residual es: $y_i - \hat{y}_i$

Coeficiente de correlación

Medida de la intensidad de la relación lineal entre dos variables.

Análisis de residuales

Análisis que se aplica para determinar si los supuestos acerca del modelo de regresión parecen válidos. También se usa para determinar observaciones extraordinarias o influyentes.

Observación influyente

Observación que tiene una fuerte influencia sobre el efecto de los resultados de la regresión.

Puntos de gran influencia.

Observaciones con valores extremos de la variable independiente.

MESOGRAFÍA

Bibliografía básica