

CAPITULO 1

ANÁLISIS DE REGRESIÓN

El análisis de regresión es uno de los métodos más utilizados para hacer estimaciones, y se emplea cuando existe relación entre dos o más variables.

El análisis de regresión, se basa en las observaciones que se realizan para cada una de las variables, donde se da a conocer una ecuación matemática que indica la relación de estas variables; cuando se tiene una ecuación está se utiliza para poder estimar valores futuros que puede tener una variable. En cualquier tipo de problema se deben tener muy bien estudiadas e identificadas las variables que existen en dicho problema, a estas variables se les conoce como variable dependiente (Y), y variable independiente (X). (1)

Cuando es usada la regresión permite encontrar una ecuación y trazar una línea que se acerque lo más posible a los puntos dados. La regresión se emplea en la estimación de una relación que puede existir en una población.

Los datos obtenidos para realizar el análisis de regresión, provienen de observaciones de las variables relacionadas.

REGRESIÓN LINEAL

¿QUÉ ES REGRESIÓN LINEAL?

Para tener un conocimiento más exacto de lo que es regresión lineal, primeramente se dará a conocer la definición de regresión y después de linealidad. Regresión es el proceso general de predecir una variable a partir de otra con medios estadísticos, usando datos anteriores.(1) Linealidad es la línea ajustada a un conjunto de puntos de datos para estimar la relación entre dos variables. (1) Por lo que *regresión lineal es el desarrollo de una ecuación*

(1) Estadística para Administradores. Richard I. Levin

de estimación que describe la relación entre dos variables. La regresión lineal tiene una finalidad muy importante que es la de estimar los valores de una variable con base a los conocidos de la otra.(1)

Cuando es utilizado este modelo para predicciones, es importante que solo se considere lo relevante de la variable independiente. Los valores de la variable dependiente se predicen cuando es conocida la variable independiente.

Algunas de las características de la ecuación lineal son:

- 1) La pendiente de la recta
- 2) La localización de la recta en algún punto.

Con este análisis no se puede establecer si una variable puede influir en los valores de la otra.

*** SUPOSICIONES DE LA REGRESIÓN LINEAL**

1.-Normalidad.- Todos los valores obtenidos para Y, deben estar distribuidos normalmente en cada valor de X.

2.-Homoscedasticidad.- Los valores de los puntos alrededor de la línea de regresión, son constantes para todos los valores de X.

3.-Independencia del error.- Es necesario que el valor se independice de cada valor de Y. (2)

4.- Linealidad.- La relación existente entre las variables es lineal.

*** RIESGOS DE LA REGRESIÓN LINEAL**

El proceso de análisis estadístico no puede establecer o determinar relaciones causales. El conocimiento de las causas debe venir del entendimiento de la física o naturaleza del sistema. Se hipotetizan causas para obtener soporte con respecto a la verdad de la hipótesis con la experimentación y el proceso de experimentación. No se puede probar causa y efecto al mostrar su correlación. (1)

(1) Estadística para Administradores. Richard I. Levin (2) Estadística Básica en la Admón. Berenson Levine

DIAGRAMA DE DISPERSIÓN

El diagrama de dispersión es la gráfica donde se encuentran todos los puntos de las observaciones, tanto de la variable dependiente como la variable independiente.(2)

El diagrama de dispersión observa que:

1) Debe existir relación lineal entre las dos variables (X,Y), dentro de cierto rango de variación.

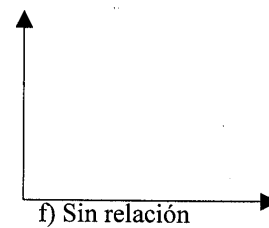
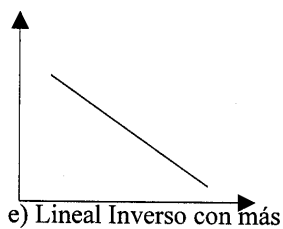
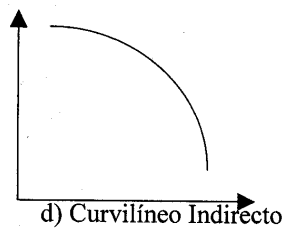
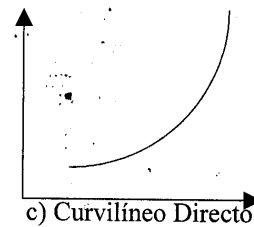
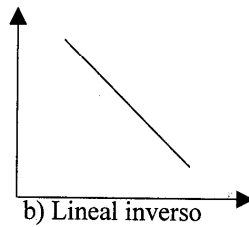
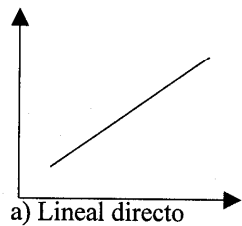
2) No es el mismo valor para las dos variables.

El diagrama de dispersión puede revelarnos dos tipos de información.

1) Relación de las variables.

2) Tipo de línea o ecuación de estimación.

Possible relación entre X y Y en los Diagramas de Dispersión (3)



(1) Probabilidad y Estadística Aplicada a la Ingeniería. Montgomery (2) Estadística para Administración y Economía. Williams J. Stevenson (3) Estadística para Administradores. Richard I. L

MÉTODO DE MÍNIMOS CUADRADOS

El método de mínimos cuadrados permite ajustar los datos observados a la línea recta, este ajuste se obtendrá minimizando el error entre los puntos estimados (\hat{Y}) y los observados (Y).

Se le llama método de mínimos cuadrados porque en la línea de estimación se busca minimizar la suma de cuadrados de los errores. Con este método debe calcularse la pendiente de la línea de estimación (β_1) y la intersección en Y (β_0). Al conocer estos dos factores conocemos la ecuación lineal, en donde al darle el valor de X , obtendremos el valor futuro de la variable dependiente. (1)

Para tener un valor del intervalo aproximado de las predicciones futuras es necesario conocer acerca del error estándar que a continuación se dará a conocer.

ERROR ESTÁNDAR DE ESTIMACIÓN

Se usa con el propósito de medir la confiabilidad de la ecuación lineal de estimación. Cual mayor es el error de la estimación más grandes serán las dispersiones de los puntos alrededor de la línea de regresión. Cuando se hace un ajuste de los puntos a la recta y, estos quedan alrededor de ellas, o sea, con error igual a cero, podría decirse que el estimador es perfecto. (2)

El error estándar se define como la medida de confianza de la ecuación de estimación, que indica la variabilidad de los números o datos observados alrededor de la línea de regresión. (2)

ANÁLISIS DE CORRELACIÓN

El Análisis de Correlación es la herramienta estadística de que nos valemos para describir el grado de relación que hay entre las dos variables; Este análisis se usa junto con el análisis de regresión para medir la eficacia de la línea de regresión, ya que revela la variación de la variable dependiente (Y). También puede usarse para medir el grado de asociación que existe entre dos variables. (2)

(1) Economía Administrativa. Marilú Hurt McCarty (2) Estadística para Administradores. Richard I. Levin

Su interpretación esta dado por el rango de valores determinado de 0 a 1; donde:

$$0 \leq R^2 \leq 1$$

El esquema se representa como sigue.

0.00	—————▶	Ninguna correlación
0.25	—————▶	Correlación débil
0.50	—————▶	Correlación Moderada
0.75	—————▶	Correlación Intensa
1.00	—————▶	Correlación Perfecta

Según algunos estadísticos existen 2 medidas para describir la correlación entre las variables: El coeficiente de determinación y el coeficiente de correlación.(1)

* COEFICIENTE DE DETERMINACIÓN

El Coeficiente de Determinación es la manera primaria de medir el grado o fuerza de la relación que existe entre dos variables X y Y.(1)

Dicho coeficiente se obtiene de la relación de dos tipos de variables. La variación de los valores de Y en el conjunto de datos alrededor de:

- 1) La línea de regresión ajustada
- 2) Su propia media.

El coeficiente de determinación se encuentra entre los extremos de 0 y 1. Pero sin olvidar que un r^2 cercano a 1 indica una fuerte correlación entre X y Y, mientras que un r^2 cercano a 0 significa que esas dos variables tienen poca correlación.(1)

* COEFICIENTE DE CORRELACIÓN

El coeficiente de correlación es la segunda medida con que puede describirse la eficacia con que una variable es explicada por otra. Este coeficiente se denota con la letra r y el la raíz cuadrada del coeficiente de determinación.

Cuando la pendiente de la ecuación de estimación es positiva, r es la raíz cuadrada positiva, pero si b es negativa, r es la raíz cuadrada negativa. El signo de r indica la dirección de la relación entre las dos variables X y Y. (1)

(1) Estadística para Administradores. Richard I. Levin

PRUEBA DE HIPÓTESIS E INTERVALO DE CONFIANZA

* Inferencia con respecto a β_1 :

El análisis para este estimador se hace desde dos puntos de vista:

- 1) Prueba de hipótesis
- 2) Intervalo de Confianza

En donde en ambos casos es necesario utilizar una t de student con (n-2) grados de libertad.(1)

* Inferencia con respecto a β_0 :

Al igual que el estimador β_1 , β_0 se calcula con una t de student con (n-2) grados de libertad.(1)

PRUEBA DE HIPÓTESIS

La prueba de hipótesis comúnmente utilizada es:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Donde la regla de decisión es:

$$\text{Aceptar sí: } t_{(n-2)} < t_{\alpha/2 (n-2)}$$

$$\text{Rechazar sí: } t_{(n-2)} > t_{\alpha/2 (n-2)}$$

INTERVALO DE CONFIANZA

El Intervalo de Confianza usado con más frecuencia es el de 95%, el cual indica que solo tenemos un 5% de error en dicho intervalo. Entre más grande sea el intervalo de confianza, el error será más pequeño, lo que indica que debemos tener mas precisión en nuestros resultados.(1)

¿DÓNDE SE APLICA?... ALGUNOS EJEMPLOS

El análisis de regresión lineal puede ser utilizado en varios tipos de problemas. Cuando se realiza la ecuación lineal, se emplean situaciones en las que dos variables miden aproximadamente lo mismo.

(1) Probabilidad y Estadística aplicada a la Ingeniería. Montgomery

Ejemplo:

La resistencia y dureza del metal, se puede estimar fácilmente su resistencia. Si al probar la resistencia se destituye metal, y no sucede lo mismo en la prueba de la dureza, una persona interesada en estimar la resistencia, obviamente podría interesarse en preferir basarse en prueba de dureza para estimar la resistencia.

Otra forma para aplicar las ecuaciones de regresión lineal, es para dar a conocer los valores de una variable en términos de la otra. Se puede intuir una relación de causa y efecto entre las dos variables.

Ejemplo.

Un economista puede intentar explicar los cambios en la demanda de automóviles usados en términos del nivel de desempleo.

Un agricultor puede creer que la cantidad de fertilizante que utilizó influyo en la cosecha lograda.

La velocidad de un automóvil podría ser un factor para determinar la distancia de frenado.

Un tercer uso de la regresión es para predecir valores futuros de la variable.

Ejemplo:

A menudo, se llevan a cabo pruebas de selección para posibles empleados o estudiantes, para predecir la potencia de tener éxito, tanto en la escuela como en un empleo. Supuestamente existe una relación matemática entre la calificación y el potencial del futuro.

Los ejemplos anteriormente mencionados se limitan a las ecuaciones lineales. Estas ecuaciones lineales son importantes porque aproximan muchas relaciones del mundo real y además porque es relativamente fácil trabajar con ellas e interpretarlas.(1)

REGRESIÓN MÚLTIPLE

¿QUÉ ES LA REGRESIÓN MÚLTIPLE?

Cuando se requiere utilizar más de una variable independiente para calcular la

(1) Estadística Elemental. Freud Simon

variable dependiente y tener una mejor exactitud de la estimación, es recomendable utilizar la técnica de regresión múltiple. En donde, *Regresión Múltiple se define como el procedimiento estadístico en virtud del cual algunas variables se usan para predecir a otra variables.*(1)

En teoría el análisis de regresión múltiple es solo una extensión del análisis de regresión lineal. Esto se refiere al desarrollo de una ecuación que se utiliza para predecir valores de Y, respecto a los valores dados de las variables independientes.

El objetivo del análisis de regresión múltiple consiste en dar a conocer aquellas variables que son de utilidad en la predicción del valor de una variable dependiente. Cuando existe una variable que no resulta de ayuda en la predicción del análisis, esta variable puede ser eliminada del modelo de regresión múltiple y así resultaría un modelo más fácil de utilizar.

La principal ventaja del análisis, es que permite utilizar una parte mayor de la información de que disponemos para estimar la variable dependiente.(1)

La regresión múltiple y el análisis de correlación son un proceso de tres pasos:

- I.- Definir la ecuación de regresión múltiple.
- II.- Examinar el error estándar de estimación.
- III.- Aplicar el análisis de correlación múltiple para ver que tan eficaz es la ecuación de regresión con los datos observados.(1)

Al aplicar estos tres pasos nos daremos cuenta si el modelo es adecuado y podemos empezar a predecir los valores de la variable dependiente (Y). Para dichas predicciones se utiliza la ecuación de regresión.(2)

ERROR ESTÁNDAR DE ESTIMACIÓN

Cuando se determina la relación de las variables se necesita alguna medida de

(1) Estadística para Administradores. Richard I. Levin (2) Estadística Elemental. Freud, Simon

dispersión alrededor del plano de regresión múltiple. Al igual que en la regresión lineal, en el análisis de regresión múltiple *la estimación es más precisa a medida que disminuye el grado de dispersión alrededor de la regresión.*

Se puede usar el error estándar de estimación y la distribución t para crear un intervalo de confianza aproximado alrededor de nuestro valor estimado de \hat{Y} .⁽¹⁾

COEFICIENTE DE CORRELACIÓN

Se usa con el fin de conocer que tan eficaz es el modelo usado para representar a sus datos, se puede calcular el coeficiente e interpretarse sus resultados de manera similar al del análisis de regresión lineal.

Este se define como la fracción de la variación de la variable dependiente que se explica por la regresión. Este coeficiente mide el grado en que la regresión múltiple se ajusta a los datos.⁽¹⁾

PRUEBA DE HIPÓTESIS E INTERVALO DE CONFIANZA

*** Inferencia sobre la pendiente β_1 :**

Al igual que en el análisis anterior, este estimador se hace desde dos puntos de vista:

- 1.- Prueba de hipótesis
- 2.- Intervalo de Confianza

En donde la diferencia al análisis anterior es que la distribución t es con $n-k-1$ grados de libertad.⁽¹⁾

INTERVALO DE CONFIANZA

Para realizar el intervalo de confianza es necesario conocer el margen de error con que se va a trabajar, como se menciona en el análisis de regresión lineal este margen es del 95%, que indica un límite de error del 5%.

*** PRUEBA DE HIPÓTESIS**

La prueba de hipótesis comúnmente utilizada es:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Donde la regla de decisión es:

$$\text{Aceptar sí : límite Superior} = \beta_1 + t (S_{b_1})$$

$$\text{Rechazar sí : límite Inferior} = \beta_1 - t (S_{b_1})$$

⁽¹⁾ Estadística para Administradores. Richard I. Levin

¿ DÓNDE SE APLICA?... ALGUNOS EJEMPLOS

Las técnicas de mínimos cuadrados utilizadas en el análisis de regresión múltiple al igual que en el análisis de regresión lineal, son para determinar las ecuaciones de regresión aún desde el punto de vista práctico, ya que es del hecho de que un problema sencillo requiere cálculos bastante complicados.

Ejemplo.

El estado general de la carrocería de un automóvil sería un factor importante, aunque puede ser difícil cuantificarlo. Una forma de resolver este problema podría ser obtener una escala en la que a los automóviles se les asigne un número de dígitos de 0 a 1 (ejemplo 0.65), dando la calificación de 1 a aquel que estuviera en mejores condiciones y 0 para el que fuera solo chatarra.

El uso de regresión múltiple en cualquier tipo de problema, se puede pronosticar una variable en términos de la otra con mucha precisión. Los resultados que arroje este análisis deben ser mejor cuando se tiene información adicional.

Ejemplo:

Podemos hacer pronósticos del desempeño de tres maestros contratados recientemente, si consideramos no solo su educación, sino sus años de experiencia y su personalidad.

Así mismo podríamos hacer mejores pronósticos del éxito de un libro de texto si consideramos no solo la calidad del trabajo sino la demanda potencial y la competencia.

En resumen, el análisis de regresión múltiple se utiliza en situaciones donde se tiene que más de un factor (variables de regresión) afectan en el resultado observado (Variable de respuesta).

Frecuentemente los modelos de regresión múltiple se emplean como funciones de aproximación, cuando se desconoce la verdadera relación entre Y y X_1, X_2, \dots, X_k . Dentro de los rangos de valores de las variables independientes, el uso de modelos de regresión múltiple constituye una buena aproximación.(1)

(1) Estadística para Administración y Economía. Williams J. Stevenson

Ejemplos de situaciones en las que se utiliza regresión múltiple (1)

Variable Dependiente (Y)	Variabes Independientes (X)
Cosecha	Cantidad de fertilizante, lluvia, tipo de suelo.
Salario anual	Años en la compañía, escolaridad.
Dureza de acero	Tiempo recorrido, cantidad de carbono, índice de enfriamiento.
Resistencia del concreto a la compresión	Composición, tiempo de curado, índice de enfriamiento.
Distancia de frenado de un automóvil	Velocidad, Coeficiente de rozamiento de la carretera y llantas, precio.
Volumen de ventas	Gasto de publicidad, precio.
Demanda de la carne de pollo	Precio de la carne de res, de cerdo, de pollo.

(1) Estadística para Administración y Economía. Williams J. Stevenson