FACULTAD DE CIENCIAS EMPRESARIALES
ASIGNATURA: Econometría de Negocios

CORPORACIÓN UNIVERSITARIA REMINGTON
DIRECCIÓN PEDAGÓGICA

2011

CRÉDITOS

El módulo de estudio de la asignatura Econometría de Negocios es propiedad de la Corporación Universitaria Remington. Las imágenes fueron tomadas de diferentes fuentes que se relacionan en los derechos de autor y las citas en la bibliografía. El contenido del módulo está protegido por las leyes de derechos de autor que rigen al país.

Este material tiene fines educativos y no puede usarse con propósitos económicos o comerciales.

AUTOR

Carlos Guillermo Londoño Herrera
DIPLOMADO en Diseño Curricular y Herramientas significativas de Autoaprendizaje. Segundo semestre del 2008.
Docente de Estadistica y Matematicas Centro de atención de tutoría virtual para el aprendizaje de la estadística en la Corporación Universitaria REMINGTON durante el año 2011
Carlos.londono@remington.edu.co
Crow43@gmail.com

Nota: el autor certificó (de manera verbal o escrita) No haber incurrido en fraude científico, plagio o vicios de autoría; en caso contrario eximió de toda responsabilidad a la Corporación Universitaria Remington, y se declaró como el único responsable.

RESPONSABLES

FACULTAD DE CIENCIAS EMPRESARIALES
Director Dr. Gonzalo Jiménez Jaramillo

Director Pedagógico
Octavio Toro Chica
dirpedagogica.director@remington.edu.co

Coordinadora de Medios y Mediaciones
Angélica Ricaurte Avendaño
mediaciones.coordinador01@remington.edu.co

GRUPO DE APOYO

Personal de la Unidad de Medios y Mediaciones
EDICIÓN Y MONTAJE
Primera versión. Febrero de 2011.

TABLA DE CONTENIDO

1.1. Introducción 4

2. MAPA DE LA ASIGNATURA 5

3. REGRESIÓN LINEAL SIMPLE 6

3.1. Prueba inicial 6

3.2. Conceptos de Econometría 7

3.3. Regresión Lineal Simple 8

3.4. Análisis de Varianza 25

4. CONSTRUCCIÓN DE MODELOS ECONOMÉTRICOS 32

4.1. Prueba inicial 32

4.2. Regresión Lineal Múltiple 33

4.2.1. Modelos de Ecuaciones Lineales de Regresión Múltiple 33

5. PROBLEMÁTICA DE LOS MODELOS DE ECONOMETRÍA 49

5.1. Multicolinealidad 49

5.1.1. Hesteroscedasticidad 49

5.2. Autocorrelación 50

6. GLOSARIO 51

7. BIBLIOGRAFÍA 52

7.1. Fuentes Digitales o Electrónicas 53

1.1. Introducción

La econometría se ocupa de obtener, a partir del análisis estadístico y matemático (mas no de la teoría económica, como si se usa en las ciencias naturales, elementos de la física) de los valores reales de variables económicas, los valores que tendrían los parámetros de los modelos en los que esas variables económicas aparecieran, así como de comprobar el grado de validez de esos modelos, y ver en qué medida estos modelos pueden usarse para explicar la economía de un agente económico (como una empresa o un consumidor), o la de un agregado de agentes económicos, como podría ser un sector del mercado, o una zona de un país, o todo un país, o cualquier otra zona económica; su evolución en el tiempo (por ejemplo, decir si ha habido o no cambio estructural), poder predecir futuros valores de la variables, y sugerir medidas de política económica conforme a objetivos deseados (por ejemplo, para poder aplicar técnicas de optimización matemática para racionalizar el uso de recursos dentro de una empresa, o bien para decidir qué valores debería adoptar la política fiscal de un gobierno para conseguir ciertos niveles de recaudación impositiva).

2. MAPA DE LA ASIGNATURA

ECONOMETRÍA DE NEGOCIOS

PROPÓSITO GENERAL DEL MÓDULO

Busca establecer relaciones entre variable independiente y variable dependiente de los elementos de la Estadística de regresión lineal simple y la múltiple por medio de la fundamentación conceptual y procedimental para desarrollar la capacidad de analizar e interpretar los resultados hacia la toma de decisiones.

OBJETIVO GENERAL

Identificar los diferentes conceptos, herramientas, métodos, técnicas y procedimientos de los modelos econométricos para la solución de problemas de la vida cotidiana, tanto laborales como académicos.

OBJETIVOS ESPECÍFICOS

- Formular de manera sistemática un modelo de regresión lineal simple para la solución problemas de la vida cotidiana.
- Analizar la influencia de la relación entre dos o más variables independientes con respecto a una variable dependiente por medio de cálculos y gráficas.
- Conocer los diferentes problemas de la regresión lineal que se involucran en la econometría.

UNIDAD 1

Tiene habilidad para la construcción de modelos de econometría.

UNIDAD 2

Aplica con destreza los conceptos de la econometría.

UNIDAD 3

Maneja, con conocimiento, la problemática de la econometría.

3. REGRESIÓN LINEAL SIMPLE

OBJETIVO GENERAL

Formular de manera sistemática un modelo de regresión lineal simple para la solución problemas de la vida cotidiana.

OBJETIVOS ESPECÍFICOS

- Interpretar los diferentes diagramas de los modelos econométricos.
- Calcular los distintos parámetros de regresión lineal simple.
- Resolver algunos ejercicios de la problemática económica de la vida real.

3.1. Prueba inicial

A continuación encontrarás una serie de enunciados con cinco respuestas, de las cuales una sola es verdadera. Marque con una **X** la que usted considere correcta.

Dadas las siguientes definiciones, el estudiante estará en capacidad de responder a que concepto corresponde:

1. Analizar a que concepto corresponde la siguiente definición: __________ es una característica, cualidad o atributo o propiedad de un sujeto o unidad de una observación.

a. Variable b. Característica c. Escala de medición d. Parámetro

2. Analizar a que concepto corresponde la siguiente definición: _________ se refiere al fenómeno que se intenta explicar el objeto de estudio de la investigación.

a. Variable b. Variable dependiente c. Variable Independiente d. Parámetro

3. Analizar a que concepto corresponde la siguiente definición: _________ son todos aquellos elementos que explican un fenómeno.

a. Variable b. Variable dependiente c. Variable Independiente d. Parámetro

4. Analizar a que concepto corresponde la siguiente definición: _________ Es una función matemática, es una cantidad en la cual el operador puede asignarle un valor arbitrario.

a. Variable b. Variable dependiente c. Variable Independiente d. Parámetro

5. Analizar a que concepto corresponde la siguiente definición: _________ de una variable estadística es la suma de todos sus posibles valores, ponderada por las frecuencias de los mismos.

a. Media b. Parámetro c. Media Aritmética d. Media Ponderada

3.2. Conceptos de Econometría

Definición de econometría

La econometría es la aplicabilidad de la estadística avanzada a los datos económicos, para obtener resultados numéricos.

Características de la econometría

Los modelos econométricos son útiles para:

1. Para el análisis estructural y la comprensión de cómo funciona la economía.
2. Predecir los valores futuros de las diferentes variables económicas.
3. La simulación, con fines de planificación, de distintas posibilidades de las distintas variables exógenas.
4. La simulación con fines de controlar valores óptimos de variables instrumentales de las diferentes políticas económicas y de la empresa.

3.3. Regresión Lineal Simple

1. **Regresión Simple**

Muestra la relación entre dos variables, una de ellas es independiente y la otra una variable dependiente.

2. **Diagrama de dispersión**

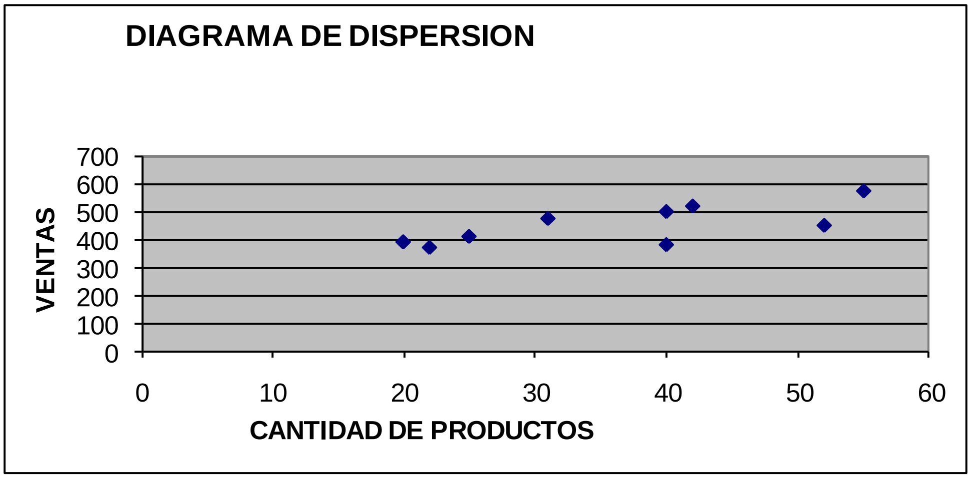
Es una gráfica de pares de datos X e Y en un espacio dimensional.

Ejemplo

Una tienda de ordenadores llevó a cabo un estudio para determinar la relación entre la cantidad de productos comprados por sus clientes mensuales y las ventas. Se obtuvieron los datos siguientes:

Cantidad de productos	**VENTAS**
40	380
25	410
20	390
22	370
31	475
52	450
40	500
20	390
55	575
42	520

GRÁFICA DE DIAGRAMA DE DISPERSIÓN



3. Coeficiente de correlación

Es un valor entre -1 y 1 que indica la fuerza de la relación lineal entre dos variables cuantitativas.

FÓRMULA

$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

TABLA DE DATOS

N	CANTIDAD DE PRODUCTOS (X)	VENTAS (Y)	XY	X2	Y2
1	40	380	15200	1600	144400
2	25	410	10250	625	168100
3	20	390	7800	400	152100
4	22	370	8140	484	136900
5	31	475	14725	961	225625
6	52	450	23400	2704	202500
7	40	500	20000	1600	250000
8	20	390	7800	400	152100
9	55	575	31625	3025	330625
10	42	520	21840	1764	270400
TOTAL	347	4460	160780	13563	2032750

r= 0,74

El coeficiente de correlación para la muestra de 10 datos puntuales es r=0,74. Esto indica que hay una relación lineal positiva bastante fuerte entre la cantidad de productos de computadores y las ventas de la tienda de ordenadores.

4. Prueba de hipótesis en el análisis de correlación

La preocupación especifica en el análisis de correlación es si se puede concluir, con arreglo a la evidencia muestral, que existe una relación lineal entre las dos variables continuas de la población.

La hipótesis nula que se quiere probar establece que no existe correlación en la población, es decir, p=0.

Ho= p=0
H1=p≠0 (Prueba de dos colas)

Ho= p=0
H1=p>0 (Prueba de una colas)

Ho= p=0
H1=p<0 (Prueba de una colas)

La distribución muestral adecuada para esta prueba es la distribución t con n-2 grados de libertad. Se pierden dos grados de libertad porque se estiman dos parámetros poblacionales (μx , μy) usando los estadísticos muestrales (X MEDIA e Y MEDIA). El valor del error estándar estimado de r se calcula mediante la ecuación.

$$Sr = \sqrt{\frac{1-r^2}{n-2}}$$

Sr= Error estándar del coeficiente de correlación
r= Coeficiente de correlación
n= Número de observaciones pareadas

La ecuación da el estadístico de la prueba adecuado:

$$t = \frac{(r-p)}{Sr}$$

r= Coeficiente de correlación muestral
p= Coeficiente de correlación poblacional hipotético
Sr= Error estándar del coeficiente de correlación

EJEMPLO (RETORNANDO AL EJEMPLO)

Una tienda de ordenadores llevó a cabo un estudio para determinar la relación entre la cantidad de productos comprados por los clientes mensualmente y las ventas. Existen diferencias significativas de correlación entre las dos variables.

Ho= p=0
H1=p≠0 (Prueba de dos colas)

El valor del error estándar estimado de r se calcula mediante la ecuación.

$$Sr = \sqrt{\frac{1-r^2}{n-2}}$$

$$Sr = \sqrt{\frac{1-(0{,}74)^2}{10-2}}$$ = 0,24

La ecuación da el estadístico de la prueba adecuado:

$$t = \frac{(r-p)}{Sr}$$

$$t = \frac{(0{,}74-0)}{0{,}24}$$ = 3,1

Se compara con la tabla de t de la tabla de t de student, con un nivel de significancia de 0,025 y con n-2=8, obteniendo un valor de t= 2,306

3,1>2,306

Se rechaza la hipótesis nula, y se concluye que no existe correlación entre la relación entre la cantidad de productos comprados por los clientes mensualmente y las ventas.

5. Ecuaciones lineales

Cuando se examina la correlación de dos variables; por lo general se hace con el propósito de usar una para pronosticar la otra.

$Y=\beta 0+\beta 1X+E$

Dónde**:**

$\beta 0$= Ordenada del origen (Intercepción en Y)
$\beta 1$= Pendiente de la recta.
E= error aleatorio

6. Método de mínimo cuadrados

RECTA DE REGRESIÓN MUESTRAL

Es la línea recta que mejor se ajusta a un conjunto de puntos X eY.

La ecuación es:

$Y=\beta 0+\beta 1X$

Dónde:

$\beta 0$= Ordenada del origen (Intercepción en Y)
$\beta 1$= Pendiente de la recta.
Y= Valor pronosticado de la variable pendiente
X= Variable independiente

Para determinar la ecuación para la línea recta que minimiza la suma de los cuadrados de las distancias verticales entre los puntos y la recta:

Para determinar la pendiente:

$$\beta 1=\frac{n\sum XY-(\sum X)(\sum Y)}{n\sum X^{2}-(\sum X)^{2}}$$

Para determinar la ordenada al origen y de la población:

$$\beta 0 = Y(MEDIA) - \beta 1 X(MEDIA)$$

EJEMPLO (RETORNANDO AL EJEMPLO)

Una tienda de ordenadores llevó a cabo un estudio para determinar la relación entre la cantidad de productos comprados por los clientes mensualmente y las ventas. Por medio del método de los mínimos cuadrados encontrar la línea recta.

Para determinar la pendiente:

$$\beta 1 = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

$$\beta 1 = \frac{10(160780) - (347)(4460)}{10(13563) - (347)^2} = 3{,}95$$

Para determinar la ordenada al origen y de la población:

$$\beta 0 = Y(MEDIA) - \beta 1 X(MEDIA)$$

$$\beta 0 = 446 - 3{,}95(34{,}7) = 308{,}80$$

La ecuación es:

$Y = \beta 0 + \beta 1X$

$Y = 308{,}80 + 3{,}95X$

7. Residuales

Un residual es la diferencia entre el valor y el valor ŷ pronosticado por la ecuación de regresión muestral.

$E = Y - \hat{y}$

E= Residual

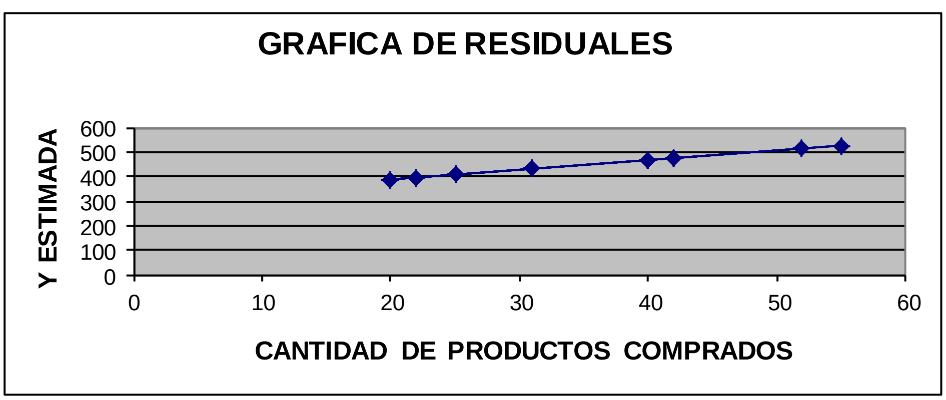
Y= valor real de y
Ŷ= Valor estimado de la variable dependiente al usar la ecuación de regresión muestral.

EJEMPLO (RETORNANDO AL EJEMPLO)

La gráfica presenta un diagrama de dispersión para los datos de la tienda de ordenadores llevó a cabo un estudio para determinar la relación la cantidad de productos comprados por los clientes mensualmente y las ventas.

$\hat{y}$ =308,80+3,95X

N	CANTIDAD DE PRODUCTOS (X)	VENTAS (Y)	ŷ
1	40	380	466,8
2	25	410	407,55
3	20	390	387,8
4	22	370	395,7
5	31	475	431,25
6	52	450	514,2
7	40	500	466,8
8	20	390	387,8
9	55	575	526,05
10	42	520	474,7
TOTAL	**347**	**4460**	



8. error estándar de la estimación

La desviación estándar de un conjunto sencillo de datos se usa para medir la variabilidad o la dispersión de los datos, alrededor de la media. El error estándar de la estimación se usa para medir la variabilidad o la dispersión de los valores pronosticados por la ecuación de regresión y los valores de y reales. Esto se puede observar en la fórmula del error estándar de la estimación:

$$Syx = \sqrt{\frac{\sum (y - \mathrm{y})^2}{\mathrm{n - 2}}}$$

EJEMPLO (RETORNANDO AL EJEMPLO)

Una tienda de ordenadores llevó a cabo un estudio para determinar la relación entre la cantidad de productos comprados por los clientes mensualmente y las ventas. Con los datos de la tabla, se pide encontrar el error estándar de la estimación.

N	**CANTIDAD DE PRODUCTOS (X)**	**VENTAS (Y)**	**ŷ**	**(y-ŷ)**	**(y-ŷ)2**
1	40	380	466,8	-86,8	7534,24
2	25	410	407,55	2,45	6,0025
3	20	390	387,8	2,2	4,84
4	22	370	395,7	-25,7	660,49
5	31	475	431,25	43,75	1914,0625
6	52	450	514,2	-64,2	4121,64
7	40	500	466,8	33,2	1102,24
8	20	390	387,8	2,2	4,84
9	55	575	526,05	48,95	2396,1025
10	42	520	474,7	45,3	2052,09
TOTAL	**347**	**4460**			**19796,55**

$$Syx = \sqrt{\frac{\sum (y - \mathrm{y})^2}{\mathrm{n - 2}}}$$

$$Syx = \sqrt{\frac{19796{,}55}{10 - 2}} = 49{,}75$$

9. Intervalo de Predicción

Una estimación puntual no proporciona información sobre la distancia a la que se encuentra del parámetro poblacional. Para determinar esta información, se desarrolla una predicción o **intervalo de confianza.** De hecho, los analistas pueden elegir entre dos tipos de intervalos: Intervalo de predicción para un valor especifico de y. **Intervalo de confianza** para el valor esperado de y para un valor dado de x.

$\hat{Y} \pm t\ S\hat{y}X$

Dónde:

$\hat{Y}$= Valor estimado de la variable dependiente al usar la ecuación de regresión muestral.
t= valor de la distribución t basado en n-2 grados de libertad para un nivel de predicción dado.
$S\hat{y}X$= Error estándar de la estimación del pronóstico.

El error estándar estimado del pronóstico $S\hat{y}X$ es una estimación de la desviación estándar de la distribución muestral para el estimador y:

$$\mathrm{Sy\,X} = \mathrm{Sy\,x}\sqrt{1+\frac{1}{\mathrm{n}}} + \frac{(Xp - X(MEDIA)^2)}{\sum(Xi - X(Media)^2)}$$

Dónde:

$S\hat{y}X$= Error estándar de la estimación del pronóstico.
SyX= Error estándar de la estimación.
Xp= El valor dado de x.
X (Media)= La media de x
$\sum$(x-x (media)^2= La suma de cuadrados total para la variable x.

EJEMPLO (RETORNANDO AL EJEMPLO)

Con respecto a la tienda de ordenadores se desea estimar las ventas del próximo mes en el caso de que la cantidad de productos comprados, por los clientes mensualmente, se incrementarán en 55. Usando los datos y la ecuación de regresión calculada, desarrollar una estimación puntual para dichas ventas:

$$\text{Sy X} = \text{Syx}\sqrt{1 + \frac{1}{\text{n}} + \frac{(Xp - X(MEDIA)^2)}{\sum (Xi - X(Media)^2)}}$$

N	CANTIDAD DE PRODUCTOS (X)	X(MEDIA)	(CANTIDAD DE PRODUCTOS (X)-X(MEDIA))	(CANTIDAD DE PRODUCTOS (X)-X(MEDIA))2
1	40	34,7	5,3	28,09
2	25	34,7	-9,7	94,09
3	20	34,7	-14,7	216,09
4	22	34,7	-12,7	161,29
5	31	34,7	-3,7	13,69
6	52	34,7	17,3	299,29
7	40	34,7	5,3	28,09
8	20	34,7	-14,7	216,09
9	55	34,7	20,3	412,09
10	42	34,7	7,3	53,29
TOTAL	**347**	**34,7**	**312,3**	**1522,1**

$$\text{Sy X} = 49{,}75\sqrt{1 + \frac{1}{10} + \frac{(55 - 34{,}7)^2)}{1522{,}1}}$$

$$\text{Sy X} = 49{,}75\sqrt{1 + 0{,}1} + \frac{(412{,}09)}{1522{,}1}$$

$$\text{Sy X} = 49{,}75\sqrt{1 + 0{,}1} + 0{,}27 \;=\; 49{,}75\sqrt{1{,}37}$$

$$\text{Sy X} = 58{,}23$$

El intervalo de predicción:

Ŷ ± t SŷX
526,05 ± 2,306(58,23)
526,05 ± 134,28

El intervalo de predicción del 95% para las ventas de la próxima semana en el que caso de que los gastos de publicidad se incrementarán 55, estará dentro del intervalo de 391,77 y 660,33, donde se encuentra el valor de **y** estimado.

10. Intervalo de confianza

Se utiliza para estimar el valor medio de y para un valor específico de X.

Ŷ ± t SûX

Dónde:

Ŷ= Valor estimado de la variable dependiente al usar la ecuación de regresión muestral.
t= valor de la distribución t basado en n-2 grados de libertad para un nivel de predicción dado.
SûX= Error estándar de la estimación del pronóstico con respecto a la media.

El error estándar estimado del pronóstico SŷX es una estimación de la desviación estándar de la distribución muestral para el estimador y:

$$\text{SûX} = \text{Syx}\sqrt{\frac{1}{\text{n}} + \frac{(Xp - X(MEDIA)^2)}{\sum(Xi - X(Media)^2)}}$$

Dónde:

SûX= Error estándar de la estimación del pronóstico.
SyX= Error estándar de la estimación.
Xp= El valor dado de x.
X(Media)= La media de x
∑(x-x(media)²= La suma de cuadrados total para la variable x.

EJEMPLO (RETORNANDO AL EJEMPLO)

Con respecto a la tienda de ordenadores se desea obtener un intervalo del 95% de confianza para las ventas media semanal, cuando se incremente la cantidad de productos comprados por los clientes mensualmente.

$$\text{SûX} = \text{Syx}\sqrt{\frac{1}{\text{n}} + \frac{(Xp - X(MEDIA)^2)}{\sum(Xi - X(Media)^2)}}$$

N	CANTIDAD DE PRODUCTOS (X)	X(MEDIA)	(CANTIDAD DE PRODUCTOS (X)-X(MEDIA))	(CANTIDAD DE PRODUCTOS (X)-X(MEDIA))2
1	40	34,7	5,3	28,09
2	25	34,7	-9,7	94,09
3	20	34,7	-14,7	216,09
4	22	34,7	-12,7	161,29
5	31	34,7	-3,7	13,69
6	52	34,7	17,3	299,29
7	40	34,7	5,3	28,09
8	20	34,7	-14,7	216,09
9	55	34,7	20,3	412,09
10	42	34,7	7,3	53,29
TOTAL	**347**	**34,7**	**312,3**	**1522,1**

$$\text{Sû}X = 49{,}75\sqrt{\frac{1}{10} + \frac{(55-34{,}7)^2)}{1522{,}1}}$$

$$\text{Sû}X = 49{,}75\sqrt{0{,}1} + \frac{(412{,}09)}{1522{,}1}$$

$$\text{Sû}X = 49{,}75\sqrt{0{,}1} + 0{,}27 \;=\; 49{,}75\sqrt{0{,}37}$$

$$\text{Sû}X = 30{,}26$$

El intervalo de confianza:

Ŷ ± t SûX
526,05 ± 2,306(30,26)
526,05 ± 69,78

El intervalo de confianza del 95% para las venta media de la próxima semana en el caso de que la cantidad de productos comprados por los clientes mensualmente se incrementarán 55, estará dentro del intervalo de 456,27 y 595,83, donde se encuentra el valor de y estimado.

11. Coeficiente de determinación simple, r^2

Mide el porcentaje de variabilidad en Y que puede ser explicado por la variable X.

Suma de cuadrados de total

La cantidad de desviación total en la variable dependiente se llama suma de cuadrados del total.

$$SCT = \sum (yi - y(media))^2$$

Suma del cuadrado del error

La recta de regresión de mínimos cuadrados minimiza la suma de cuadrados del error. La SCE mide la variabilidad de los valores Y de la muestra alrededor de Y.

$$SCE = \sum (yi - y(pronosticado))^2$$

Suma de cuadrado de la regresión

La cantidad de la desviación en la variable dependiente explicada por la ecuación de regresión.

$$SCR = SCT - SCE$$

La ecuación de r^2, el porcentaje de variabilidad de la variable dependiente, Y, que puede explicarse por la variable predoctora, X, se puede definir ahora como:

$$r2 = 1 - \frac{\sum (y - y(pronosticado))^2}{\sum (y - y(media))^2}$$

0 $\quad r2 = 1 - \dfrac{SCE}{SCT}$

El coeficiente que está después del signo menos representa el porcentaje de la variabilidad de Y que todavía, no se puede explicar en la ecuación de regresión.

EJEMPLO (RETORNANDO AL EJEMPLO)

Una tienda de ordenadores llevó a cabo un estudio para determinar la relación entre la cantidad de productos comprados por los clientes mensualmente y las ventas. Se pide determinar el coeficiente de determinación.

CANTIDAD DE PRODUCTOS (X)	VENTAS (Y)	ŷ	(y-ŷ)	(y-ŷ)2	y(media)	(y-y(media))	(y-y(media))2
40	380	466,8	-86,8	7534,24	446	-66	4356
25	410	407,55	2,45	6,0025	446	-36	1296
20	390	387,8	2,2	4,84	446	-56	3136
22	370	395,7	-25,7	660,49	446	-76	5776
31	475	431,25	43,75	1914,0625	446	29	841
52	450	514,2	-64,2	4121,64	446	4	16
40	500	466,8	33,2	1102,24	446	54	2916
20	390	387,8	2,2	4,84	446	-56	3136
55	575	526,05	48,95	2396,1025	446	129	16641
42	520	474,7	45,3	2052,09	446	74	5476
347	**4460**			**19796,55**			**43590**

Suma de cuadrados de total

$$SCT = \sum (yi - y(media))^2$$

SCT= 43590

Suma del cuadrado del error

La recta de regresión de mínimos cuadrados minimiza la suma de cuadrados del error. La SCE mide la variabilidad de los valores Y de la muestra alrededor de Y.

$$SCE = \sum (yi - y(pronosticado))^2$$

SCE=19796,55

Suma de cuadrado de la regresión

La cantidad de la desviación en la variable dependiente explicada por la ecuación de regresión.

$$SCR = SCT - SCE$$

SCR=43590-19796,55 =23793,45

La ecuación de r^2, el porcentaje de variabilidad de la variable dependiente, Y, que puede explicarse por la variable predoctora, X, se puede definir ahora como:

$$r2 = 1 - \frac{\sum (y - y(\text{pronosticado}))^2}{\sum (y - y(\text{media}))^2}$$

0 $$r2 = 1 - \frac{SCE}{SCT}$$

$$r2 = 1 - \frac{19796{,}55}{43590}$$

$$r2 = 0{,}55$$

El porcentaje de variabilidad en las ventas que puede ser explicado por la variable la cantidad de productos comprados por los clientes mensualmente es del 55%.

12. Coeficiente de correlación

Mide la fuerza de la relación de variabilidad en Y que puede ser explicado por la variable X.

$$r = \sqrt{1 - \frac{\sum (y - y(\text{pronosticado}))^2}{\sum (y - y(\text{media}))^2}}$$

$$r = \sqrt{1 - \frac{SCE}{SCT}}$$

EJEMPLO (RETORNANDO AL EJEMPLO)

Una tienda de ordenadores llevó a cabo un estudio para determinar la relación entre la cantidad de productos comprados por los clientes mensualmente y las ventas. Se pide determinar el coeficiente de determinación.

$$r = \sqrt{1 - \frac{\sum (y - y(\text{pronosticado}))^2}{\sum (y - y(\text{media}))^2}}$$

$$r = \sqrt{1 - \frac{SCE}{SCT}}$$

$$r = \sqrt{1 - \frac{19796{,}55}{43590}}$$

$$r = \sqrt{0{,}55}$$

r= 0,74

Mide la fuerza de la relación de las ventas con respecto a la cantidad de productos comprados por los clientes mensualmente semanal es del 74%.

PRUEBA DE HIPÓTESIS EN EL ANÁLISIS DE REGRESIÓN

13. Prueba de hipótesis para la pendiente

Otro estadístico importante es el valor t, que se usa para probar la hipótesis nula que la pendiente de la ecuación de regresión para la población es 0. Si una ecuación de regresión tiene pendiente 0, un cambio en X no afecta Y. En otras palabras, X e Y no tienen correlación poblacional.

Ho=β1=0
H1=β1$\neq$0

$$SB = \frac{Syx}{\sqrt{\sum (x - x(media))^2}}$$

Dónde:

Sb= Error estándar del coeficiente de regresión
Syx= Error estándar de la estimación

La ecuación del estadístico adecuado es:

$$t = \frac{b1 - \beta 1}{Sb}$$

EJEMPLO (RETORNANDO AL EJEMPLO)

Una tienda de ordenadores llevó a cabo un estudio para determinar la relación entre la cantidad de productos comprados por los clientes mensualmente y las ventas. Se pide determinar la hipótesis para la pendiente.

Ho=β1=0
H1=β1≠0

$$SB = \frac{Syx}{\sqrt{\sum (x - x(media))^2}}$$

$$SB = \frac{49{,}75}{1522{,}1}$$ =0,03

La ecuación del estadístico adecuado es:

$$t = \frac{b1 - \beta 1}{Sb}$$

Y=β0+β1X

Y=308,80+3,95X

$$t = \frac{3{,}95 - 0}{0{,}03}$$

t=120,85

El valor de t calculado (120,85) es mayor t de la tabla (2,306).Por tanto, se rechaza la hipótesis nula. Se concluye que la recta de regresión poblacional no tiene pendiente 0. Existe alguna relación lineal entre X e Y en la población.

3.4. Análisis de Varianza

Otro dato estadístico importante en el análisis de regresión es el estadístico F, que se usa para probar la hipótesis nula de que la ecuación de regresión muestral, no explica un porcentaje significativo de la varianza de la variable Y. Las hipótesis nula y alternativa son:

$Ho=p^2=0$
$H1=p^2\neq0$

El estadístico de prueba para la hipótesis nula establecida se obtiene de la distribución F si la hipótesis nula es cierta.

TABLA DE ANÁLISIS DE VARIANZA (ANOVA)

FUENTE DE VARIACIÓN	G.L.	SUMA DE CUADRADOS	ESTIMACIÓN DE VARIANZA
REGRESIÓN	k-1	SCR	SCR/(k-1)
ERROR RESIDUAL	n-k	SCE	SCE/(n-k)
TOTAL	n-1	SCT	

El estadístico de la prueba F:

$$f = \frac{\mathrm{SCR}/(k-1)}{\mathrm{SCE}/(n-k)}$$

EJEMPLO (RETORNANDO AL EJEMPLO)

Una tienda de ordenadores llevó a cabo un estudio para determinar la relación entre la cantidad de productos comprados por los clientes mensualmente y las ventas. Se pide determinar la hipótesis para la pendiente.

Ho=p²=0
H1=p²≠0

El estadístico de prueba para la hipótesis nula establecida se obtiene de la distribución F si la hipótesis nula es cierta.

TABLA DE ANÁLISIS DE VARIANZA (ANOVA)

FUENTE DE VARIACIÓN	G.L.	SUMA DE CUADRADOS	ESTIMACIÓN DE VARIANZA
REGRESIÓN	1	23793,45	23793,45
ERROR RESIDUAL	8	19796,55	2474,57
TOTAL	9	43590	

El estadístico de la prueba F:

$$f = \frac{\text{SCR}/(\text{k} - 1)}{\text{SCE}/(\text{n} - \text{k})}$$

$$f = \frac{23793{,}45}{2474{,}57}$$

$$f = 9{,}6$$

F DE LA TABLA PARA F(0.025, 1, 8)=7,57086

Por tanto, se rechaza la hipótesis nula. Con muy poca probabilidad de error. Se concluye que la ecuación de regresión explica un porcentaje significativo de la varianza de las ventas.

Ejercicios

1. Un hipermercado ha decidido ampliar el negocio. Decide estudiar de forma exhaustiva el número de cajas registradoras que va a instalar, para evitar grandes colas. Para ello, se obtuvieron los siguientes datos procedentes de otros establecimientos similares acerca del número de cajas registradoras y del tiempo medio de espera.

N	NÚMERO DE CAJAS REGISTRADORAS	TIEMPO PROMEDIO DE ESPERA
1	10	30
2	12	25
3	13	32
4	14	34
5	15	35
6	16	28
7	18	30
8	20	32
9	12	24
10	14	36

Bajo el supuesto de que el tiempo de espera medio depende linealmente del número de cajas registradoras se pretende saber, e Interpretar:

1. Realizar el diagrama de dispersión.
2. Realizar el coeficiente de correlación.
3. Realizar la prueba de hipótesis en el análisis de correlación. Con t del 0.95
4. Realizar la prueba de hipótesis en el análisis de correlación. Con t del 0.90
5. Encontrar la ecuación de la línea recta de regresión muestral con el método mínimos cuadrados
6. Encontrar la línea recta cuando x vale 10, 12, 24.
7. Encontrar los residuales.
8. Realizar el grafico de y pronostica con respecto a x.
9. Encontrar la línea recta de regresión con el valor mínimo y el valor máximo
10. Realizar el grafico de los residuales.

3. Un investigador cree que la inteligencia de los niños, medida a través del coeficiente intelectual (CI en puntos), depende del número de hermanos. Toma una muestra aleatoria de 15 niños y ajusta una regresión lineal simple. Los resultados aparecen en la salida adjunta.

N	CL	HERMANOS
1	112	0
2	114	1
3	110	2
4	113	3
5	114	2
6	115	4
7	110	2
8	112	1
9	117	2
10	111	3
11	118	2
12	120	4
13	122	5
14	121	4
15	124	5

Bajo el supuesto de que el tiempo de espera medio depende linealmente del número de cajas registradoras se pretende saber, e Interpretar:

1. Realizar el diagrama de dispersión.
2. Realizar el coeficiente de correlación.
3. Realizar la prueba de hipótesis en el análisis de correlación. Con t del 0.95
4. Realizar la prueba de hipótesis en el análisis de correlación. Con t del 0.90
5. Encontrar la ecuación de la línea recta de regresión muestral con el método mínimos cuadrados
6. Encontrar la línea recta cuando x vale 1, 5, 2.
7. Encontrar los residuales.
8. Realizar el grafico de y pronostica con respecto a x.
9. Encontrar la línea recta de regresión con el valor mínimo y el valor máximo
10. Realizar el grafico de los residuales.

4. La entrada a cine en los teatros de cine Colombia en los centros comerciales, indica una relación de cantidad de personas que ingresan y el valor en el precio a pagar (miles de pesos) . Con la siguiente tabla responder las preguntas:

N	CANTIDAD	PRECIO
1	12	96000
2	14	112000
3	16	121000
4	18	132000
5	19	135000
6	15	118000
7	14	110000
8	12	90000
9	16	119000
10	15	117500
11	18	130000
12	17	127600

Se pide encontrar:

1. Realizar el diagrama de dispersión.
2. Realizar el coeficiente de correlación.
3. Realizar la prueba de hipótesis en el análisis de correlación. Con t del 0.95
4. Realizar la prueba de hipótesis en el análisis de correlación. Con t del 0.90
5. Encontrar la ecuación de la línea recta de regresión muestral con el método mínimos cuadrados
6. Encontrar la línea recta cuando x vale 18, 15, 12.
7. Encontrar los residuales.
8. Realizar el grafico de y pronostica con respecto a x.
9. Encontrar la línea recta de regresión con el valor mínimo y el valor máximo
10. Realizar el grafico de los residuales.

5. Dada la difícil situación por la que atraviesa actualmente la empresa PALMA CARIBE en la que hemos empezado a trabajar, se propone la reducción de determinados gastos. Para ello se estudia la relación que existe entre dos variables como son: los gastos en publicidad (variable X) y los beneficios (variable Y). De ambas variables disponemos de los siguientes datos:

AÑO	GASTOS EN PUBLICIDAD	UTILIDADES
1985	60	32
1986	65	35
1987	78	37
1988	79	38
1989	82	42
1990	86	44
1991	88	46
1992	92	56
1993	98	58
1994	99	60

Se pide:

1. ¿Se puede considerar que ambas variables guardan algún tipo de relación? ¿Cuál sería la variable dependiente y cuál la independiente?

2. Realizando un gráfico adecuado. ¿Se puede suponer que la relación que las liga es de tipo lineal?

3. Construye las dos rectas de regresión mínimo cuadrática asociada con las variables.

4. Si la empresa para el próximo año realizará un esfuerzo para poder invertir 12.550.000 pesos en publicidad. ¿Cuáles resultarían ser sus beneficios? ¿Con qué fiabilidad realizaría usted la predicción?

5. ¿Cuáles resultarían ser sus beneficios si la predicción se efectúa considerando tan solo como variable explicativa el tiempo? ¿Cuál sería la fiabilidad de esta otra predicción? Comente los resultados.

Prueba Final

1. ¿Cómo define usted el concepto de regresión simple?

2. ¿Cómo define usted el concepto de diagrama de dispersión y coeficiente de regresión?

3. ¿Qué es un análisis de regresión simple?

4. Con una aplicación en su empresa calcule los parámetros de regresión simple y defina sus características.

ACTIVIDAD

El estudiante debe realizar un proyecto aplicando la regresión simple y analizar si es viable o no teniendo en cuenta el análisis de cálculos y gráficas.

4. CONSTRUCCIÓN DE MODELOS ECONOMÉTRICOS

OBJETIVO GENERAL

Analizar la influencia de la relación entre dos o más variables independientes con respecto a una variable dependiente por medio de cálculos y gráficas.

OBJETIVOS ESPECÍFICOS

- Interpretar los diagramas estadísticos de regresión múltiple.
- Calcular los parámetros de regresión múltiple.
- Analizar la influencia de cada una de las variables independientes con respecto a la variable dependiente.
- Efectuar análisis de residuales.

4.1. Prueba inicial

A continuación encontrará una serie de enunciados con cinco respuestas, de las cuales una sola es verdadera. Marque con una **X** la que usted considere correcta.

1. El profesor de la materia de estadística desea conocer el promedio de las notas finales de los 10 alumnos de la clase. Las notas de los alumnos son 3.2, 3.1, 2.4, 4, 3.5, 3, 3.5, 3.8, 4.2, 4. ¿Cuál es el promedio de notas de los alumnos de la clase?
   a. a. 3.5 b.3 c.4 d.4.2 e.3.8

2. Los miembros de una cooperativa de viviendas tienen las siguientes edades: 21, 23, 24, 56, 35, 38, 41, 45, 34 ,35.
   a. 3.6 b.3.5 c. 3.4 d.3.7 e.4

3. En el primer parcial de estadística las notas de los estudiantes fueron 2, 2.5, 3, 4, 5, 5, 4, 5, 3.5, 3.8, 4.2. Cuál es la nota promedio?
   a. 3 b.3.1 c.3.3 d.3.4 e.3.5

4. Los datos representan la edad de los miembros de un grupo de niños de una institución educativa de buenos aires, para el cual seleccionaron 6 niños(4, 1, 11, 13, 2, 7). Se pide calcular la desviación estándar.
    a. 4.9 b.4.8 c.5 d.4.3 e.4.7

5. Un fabricante de componentes electrónicos está interesado en determinar el promedio de vida de un tipo de batería. Toma una muestra de 9 baterías y obtiene que duran un total de 121, 118, 123, 112, 165, 136, 145, 151, 128 horas.
    a.132 b.133 c.134 c.135 d.136

4.2. Regresión Lineal Múltiple

Los modelos de análisis de regresión múltiple son similares a los modelos regresión lineal simple, excepto es que se tienen más de una variable independiente y pueden llegar a tener un mayor grado de complejidad en el análisis del modelo y en la relación de la línea recta ajustada a la regresión.

4.2.1. Modelos de Ecuaciones Lineales de Regresión Múltiple

Los modelos de análisis de regresión múltiple son similares a los modelos regresión lineal simple, excepto es que se tienen más de una variable independiente y pueden llegar a tener un mayor grado de complejidad en el análisis del modelo y en la relación de la línea recta ajustada a la regresión.

1. **Modelo de regresión múltiple es el siguiente:**

$E(Y) = \beta 0 + \beta 1X1 + \beta 2X2$

2. **Supuestos del modelo.**

Los supuestos del modelo de análisis de regresión son los siguientes:

1. La media de e ES 0. Esto implica que la media de y equivale al componente determinístico del modelo, esto es,

$E(y)= \beta 0 + \beta 1X1 + \beta 2X2+\ldots\ldots.+ \beta kXk$

2. Para todos los valores de las variables independientes X1,X2, ……,Xk, la varianza de E es constante.
3. La distribución de probabilidad de E es normal
4. Los errores aleatorios son independientes.

3. Las ecuaciones de mínimos cuadrados y su resolución.

Para encontrar un modelo de regresión múltiple se utiliza el procedimiento de álgebra lineal por medio de operaciones con matrices, teniendo en cuenta el siguiente modelo:

E (y)= β0 + β1X1 + β2X2+…….+ βkXk+ε

Las matrices de datos Y y X, la matriz de β y la matriz del error

$$Y = \begin{bmatrix} y1 \\ y2 \\ . \\ . \\ . \\ yn \\ \\ \\ \end{bmatrix} \quad X = \begin{bmatrix} 1 & X11 & X12 & X13 \\ 1 & X21 & X22 & X23 \\ 1 & X31 & X32 & X33 \\ 1 & Xn1 & Xn2 & Xn3 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta 0 \\ \beta 1 \\ \beta 2 \\ \beta 3 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon 1 \\ \varepsilon 2 \\ \varepsilon 3 \\ \varepsilon 4 \end{bmatrix}$$

Ecuación de matrices de mínimos cuadrados

(X` X) β = X` Y

Ejemplo

Se realizó un estudio en almacenes éxito del poblado sobre la congestión en la entrada de vehículos almacén, para determinar el comportamiento de la entrada se tomaron las variables del número de vehículos(X) y el tiempo de congestionamiento. Considere el modelo de lineal y observe si existe o no la relación entre las variables.

Número de vehículos	Tiempo de congestionamiento
1	O
2	O
3	1
4	1
5	2
6	2
7	3
8	2
9	3
10	4

La estimación de mínimos cuadrados.

Solución

$$Y = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 2 \\ 2 \\ 3 \\ 2 \\ 3 \\ 4 \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \\ 1 & 10 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} \mathbf{B0} \\ \mathbf{B1} \end{pmatrix}$$

Entonces:

$$X`X= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \\ 1 & 10 \end{pmatrix} = \begin{pmatrix} 10 & 55 \\ 55 & 385 \end{pmatrix}$$

$$X`Y= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 2 \\ 2 \\ 3 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 18 \\ 133 \end{pmatrix}$$

Por último, determinar la inversa de la matriz.

$$(X`X)^{-1} = \begin{pmatrix} 0,47 & -0,07 \\ 0,07 & 0,01 \end{pmatrix}$$

Y la solución de la ecuación de mínimos cuadrados es entonces:

$$\mathbf{B = (X`X)^{-1} X`Y} = \begin{pmatrix} 0,47 & -0,07 \\ -0,07 & 0,01 \end{pmatrix} \begin{pmatrix} 18 \\ 133 \end{pmatrix} = \begin{pmatrix} 0,9 \\ 0,1 \end{pmatrix}$$

Por tanto, $\beta 0 = -0,9$ y $\beta 1 = 0,1$ y la ecuación de predicción es

$E(y) = -0,9 + 0,1X$

4. Propiedades de los estimadores de mínimos cuadrados.

Las propiedades de la distribución de muestreo de βi (i=0, 1, 2, 3, 4,….., k) es normal, o se comporta según la distribución de probabilidad Normal con:

1. $E(\beta i) = \beta i$
2. $V(\beta i) = Cii\, \sigma^2$
3. $\sigma \beta = \sigma \sqrt{Cii}$

5. Estimación de δ^2, la varianza de E.

Las varianzas de los estimadores de todos los parámetros β y de y estimada dependen del valor de σ^2, la varianza del error aleatorio ε que se encuentra en el modelo lineal. Puesto que σ^2 casi nunca se conoce por adelantado, se debe utilizar las observaciones de la muestra para encontrar su respectivo valor.

$$S^2 = \frac{SSE}{n - \text{Numero de parámetros de } \beta \text{ en el modelo}}$$

SSE = Y` Y - β` X` Y

Continuando con el ejemplo anterior, se debe de calcular la estimación de la varianza del modelo.

SSE = Y` Y - β` X` Y

SSE = 48 - (-2,9)= 48 + 2,9 = 50,9

$$S^2 = \frac{SSE}{n - \text{Numero de parámetros de } \beta \text{ en el modelo}}$$

$$S^2 = \frac{50,9}{10 - 2}$$

S^2= 6,4

La variabilidad con respecto a la media de la muestra del tiempo de congestionamiento es del 6,4.

6. Intervalos de confianza y pruebas de hipótesis para βo, βi, …βk.

El intervalo de confianza de (1 - α) 100% para βi

El intervalo de confianza para la muestra de un modelo de regresión múltiple.

βi ± t α/2 (Error estimado estándar de βi), o sea,

$\beta i \pm t\, \alpha/2\, s \sqrt{Cii}$

Donde t α/2 se basa en el número de grados de libertad a s.

Ejemplo

Retomando el ejemplo anterior, se pide determinar el intervalo de confianza del 95% para $\beta 1=0,1$ y de la matriz inversa c11=0,01

$\beta i \pm t\, \alpha/2\; s \sqrt{Cii}$

$\beta 1 \pm t\, \alpha/2\; s \sqrt{C11}$

$0,1 \pm 2,306\,(2,5)\sqrt{0,01}$

$0,1 \pm 0,6$

El intervalo de confianza del 95% para el tiempo de congestionamiento con respecto a la media oscila entre -0,5 y 0,7.

Prueba de Hipótesis de un coeficiente de parámetro individual en el modelo de regresión múltiple

Prueba de una cola

Ho: $\beta i = 0$
H1: $\beta i > 0$ o $\beta i < 0$

Prueba de dos colas

Ho: $\beta i = 0$
H1: $\beta i \neq 0$

Estadístico de prueba:

$$t = \frac{\beta i}{s\beta i} = \frac{\beta i}{s\sqrt{Cii}}$$

Région de rechazo:

$T > t\alpha$ o $T < -t\alpha$

Donde

n= Número de observaciones
k= Número de variables independientes en el modelo

y t $\alpha/2$ se basa en (n- (k+1)) gl

Ejemplo

Retomando el ejemplo anterior, Se pide calcular el valor estadístico para probar que $\beta 1 = 0$

Prueba de dos colas

Ho: $\beta i = 0$
H1: $\beta i \neq 0$

Estadístico de prueba:

$$t = \frac{\beta i}{s\beta i} = \frac{\beta i}{s\sqrt{Cii}}$$

$$t = \frac{0,1}{2,306\ \sqrt{0,01}}$$

t= 0,4

Región de rechazo:

T > tα o T < - tα

0,4 < 2,306

Se puede concluir, que con un nivel de confianza del 95% existen diferencias significativas entre los dos tipos entre la relación del número de vehículos y el tiempo en el congestionamiento.

6. Coeficiente de determinación múltiple.

Mide el porcentaje de variabilidad de la variable dependiente (Y) que puede ser explicado por una de las variables independientes (X).

$$R^2 = 1 - \frac{SSE}{SSyy}$$

Donde SSE= $\Sigma (Y - Y\ (MEDIA))^2$, SSyy = $\Sigma(Y - Y\ (PRONOSTICADA))^2$

El coeficiente que está después del signo menos representa el porcentaje de la variabilidad de Y que todavía, no se puede explicar en la ecuación de regresión múltiple.

Ejemplo

Retomando el ejemplo anterior, se pide encontrar el coeficiente de determinación.

$$R^2 = 1 - \frac{SSE}{SSyy}$$

$$R^2 = 1 - \frac{38{,}88}{55{,}85}$$

$r^2 = 1- 0, 70$

$r^2 = 0{,}30$

El porcentaje de variabilidad en el tiempo de congestionamiento que puede ser explicado por la variable del número de vehículos es del 30%.

8. Un intervalo de confianza para E (y).

Se utiliza para estimar el valor medio de y para un valor específico de X, según el nivel de confiabilidad del estudio.

$£ \pm t\, \alpha/2\ s \sqrt{}a`(x`x)^{-1} a$

Donde

E (y)= $\beta0 + \beta1X1 + \beta2X2 + \ldots\ldots. + \beta kXk$

£=Y (PRONOSTICADA)= $\beta0 + \beta1X1 + \beta2X2 + \ldots\ldots. + \beta kXk$

$$a= \begin{pmatrix} 1 \\ x1 \\ x2 \\ x3 \\ . \\ . \\ . \\ xk \end{pmatrix}$$

Si y $`(x`x)^{-1}$ se obtienen del análisis de mínimos cuadrados, y t $\alpha/2$ se basa en el número de grados de libertad asociados a s, es decir, (n-(k+1)).

Ejemplo

Refiérase al ejercicio anterior tomando sus datos sobre el tiempo de congestión y el número de vehículos sometido a la fuerza comprensiva de x. Calcule el intervalo de confianza de 95% para la comprensión de la media E(y) cuando toma el valor de x = 3.

$$a= \begin{pmatrix} 1 \\ 3 \end{pmatrix} \qquad (X`X)^{-1}= \begin{pmatrix} 0,47 & -0,07 \\ 0,07 & 0,01 \end{pmatrix}$$

$$a`(x`x)^{-1} a = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \begin{pmatrix} 0,47 & -0,70 \\ -0,70 & 0,01 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$a`(x`x)^{-1}a = 3,64$

$£ \pm t\,\alpha/2\ s\ \sqrt{a`(x`x)^{-1}a}$

$-0,6 \pm 2,306\ (2,5)\ \sqrt{3,64}$

$-0,6 \pm 11$

El intervalo de confianza del 95% para el tiempo de congestionamiento en el que caso de que el número de vehículos es de 3, estará dentro del intervalo de -11,6 y 10,4, donde se encuentra el valor de y estimado.

9. Un intervalo de predicción para un valor futuro de Y.

Una estimación puntual no proporciona información sobre la distancia a la que se encuentra del parámetro muestral poblacional. Por tanto el Intervalo de predicción para un valor especifico de y.

$Y\ (\text{pronosticado}) \pm t\,\alpha/2\ s\ \sqrt{1+a`(x`x)^{-1}a}$

Donde

$Y\ (\text{pronosticado})) = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots\ldots. + \beta kXk$

S y $`(x`x)^{-1}$ se obtienen del análisis de mínimos cuadrados.

$$a = \begin{pmatrix} 1 \\ x1 \\ x2 \\ x3 \\ . \\ . \\ . \\ xk \end{pmatrix}$$

Contiene los valores numéricos de x1, x2, x3,……,xk y t $\alpha/2$ se basa en el número de grados de libertad asociados a s, es decir, (n-(k+1)).

Ejemplo

Refiérase al ejercicio anterior tomando sus datos sobre el tiempo de congestión y el número de vehículos sometido a la fuerza comprensiva de x. Calcule el intervalo de predicción de 95% para la comprensión de la media E (y) cuando toma el valor de x = 3.

Y (pronosticado) $\pm$ t $\alpha/2$ s $\sqrt{1 + a`(x`x)^{-1} a}$

-0,6 $\pm$ 2,306 (2,5) $\sqrt{4{,}64}$

-0,6 $\pm$ 2,306 (2,5) $\sqrt{4{,}64}$

-0,6 $\pm$ 12, 4

El intervalo de predicción del 95% para el tiempo de congestión en el que caso de que el número de vehículos fuera de 3, estará dentro del intervalo de -13 y 11,8 donde se encuentra el valor de y estimado.

Ejercicio

1. La entrada a cine en los teatros de cine Colombia en los centros comerciales, indica una relación de cantidad de personas que ingresan y el valor en el precio a pagar (miles de pesos) . Con la siguiente tabla responder las preguntas:

N	CANTIDAD	PRECIO
1	10	85500
2	12	102600
3	18	126700
4	16	130800
5	17	132300
6	14	108000
7	16	112500
8	10	96000
9	16	129000
10	14	127500
11	19	135000
12	17	126500

Se pide encontrar:

1. Modelos lineales generales.
2. Las ecuaciones de mínimos cuadrados y su resolución.
3. Estimación de δ^2, la varianza de E.
4. Intervalos de confianza
5. pruebas de hipótesis para βo, βi, …βk.
6. Coeficiente de determinación múltiple.

7. Un intervalo de confianza para E (y).
8. Un intervalo de predicción para un valor futuro de Y.

2. La secretaria de transito del Área Metropolitana realizo un estudio sobre un conjunto de conductores para analizar la edad (Y) y el número de accidentes que han sufrido (X). A partir de la misma, se obtuvieron los siguientes resultados:

N	EDAD	NUMERO DE ACCIDENTES
1	21	2
2	22	3
3	23	1
4	24	2
5	25	3
6	26	2
7	28	4
8	30	5
9	32	6
10	44	7
11	45	8
12	47	6

Se pide encontrar:

1. Modelos lineales generales.
2. Las ecuaciones de mínimos cuadrados y su resolución.
3. Estimación de δ^2, la varianza de E.
4. Intervalos de confianza
5. pruebas de hipótesis para βo, βi, …βk.
6. Coeficiente de determinación múltiple.
7. Un intervalo de confianza para E(y).
8. Un intervalo de predicción para un valor futuro de Y.
9. Análisis de residuales.

3. Se supone que se puede establecer la relación lineal entre las exportaciones de un país de Suramérica y la producción interna de dicho país. En el caso de Argentina, tenemos los datos anuales (expresados en miles de millones de pesetas) para tales variables correspondientes al 1968 a 1977 en la siguiente tabla:

N	**Producción**	**Exportaciones**
1968	45.567	7.890
1969	46.559	8.245
1970	47558	8.656
1971	48.904	8.756
1972	49.678	9.245
1973	50.565	9.458
1974	51.456	9.656
1975	55.689	9.586
1976	56.789	9.656
1977	60.755	10.897

Se pide encontrar:

1. Modelos lineales generales.
2. Las ecuaciones de mínimos cuadrados y su resolución.
3. Estimación de δ^2, la varianza de E.
4. Intervalos de confianza
5. pruebas de hipótesis para βo, βi, …βk.
6. Coeficiente de determinación múltiple.
7. Un intervalo de confianza para E(y).
8. Un intervalo de predicción para un valor futuro de Y.

10. Análisis de residuales.

4. La empresa COLOMBINA. ha trabajado hasta ahora con la hipótesis de que las ventas de un período dependen linealmente de los gastos de publicidad efectuados en el período anterior. En este momento, le solicitan a usted la realización de un análisis que ponga de manifiesto si la hipótesis, hasta ahora mantenida, se puede seguir sosteniendo en función de los datos que le suministran.

AÑO	GASTOS	VENTAS
1980	22	123
1981	23	124
1982	26	125
1983	28	132
1984	29	133
1985	30	134
1986	32	136
1987	34	138
1988	36	140
1989	38	142

Se pide encontrar:

1. Modelos lineales generales.
2. Las ecuaciones de mínimos cuadrados y su resolución.
3. Estimación de δ^2, la varianza de E.
4. Intervalos de confianza
5. pruebas de hipótesis para βo, βi, …βk.
6. Coeficiente de determinación múltiple.
7. Un intervalo de confianza para E(y).
8. Un intervalo de predicción para un valor futuro de Y.
9. Análisis de residuales.

Prueba Final

1. ¿Cómo define usted el concepto de regresión múltiple?

2. ¿Cuál es el procedimiento del método de mínimos cuadrados?

3. ¿Qué es un análisis de regresión múltiple?

4. Con una aplicación en su empresa calcule los parámetros de regresión múltiple y defina sus características.

ACTIVIDAD

El estudiante debe realizar un proyecto aplicando la regresión múltiple y analizar si es viable o no teniendo en cuenta el análisis de cálculos y gráficas.

5. PROBLEMÁTICA DE LOS MODELOS DE ECONOMETRÍA

OBJETIVO GENERAL

Conocer los diferentes problemas de la regresión lineal que se involucran en la econometría.

OBJETIVOS ESPECÍFICOS

- Identificar en qué consisten los problemas de regresión lineal aplicada a la econometría.
- Determinar posibles soluciones a estos problemas.

5.1. Multicolinealidad

El primer supuesto con respecto al modelo de regresión lineal, es que no se debe tener un alto grado de correlación entre las diferentes variables independientes y la dependiente, traería serios problemas, puede decir:

- En primer caso se tendría que los estimadores por mínimos cuadrados se encuentran en forma lineal, insegados y óptimos pero el problema se presentaría con la varianza y covarianza que pueden ser demasiado grandes.

- En algún caso se puede llegar a que el estimador de t no sea una prueba significativa.

- También se puede llegar a presentar que el coeficiente de determinación sea demasiado alto, lo que llevaría que el efecto de cada variable no el comportamiento adecuado.

5.1.1. Hesteroscedasticidad

El problema de Heteroscedasticidad se presenta cuando no se cumple con la hipótesis de varianza constante para el término de la perturbación. Cuando se presenta este caso, en no todos los términos de la diagonal principal la matriz de la varianza y covarianza serán iguales o idénticas.

Puede decir, que el estimador de mínimos cuadrados de β es insesgado y consistente, es decir, la varianza no será mínima, por lo que si se utiliza el estimador de mínimos cuadrados en lugar del eficiente para encontrar intervalos de confianza se estará perdiendo precisión ya se tendrá intervalos demasiado grandes de lo que proporciona el estimador de eficiencia.

5.2. Autocorrelación

En el momento en que se trabaja con un modelo de regresión múltiple, donde se tiene que la primera variable vale 1 y es acompañada por el termino independiente, es decir, que la matriz de varianzas y covarianzas en términos de perturbación, en la hipótesis se refiere al hecho de que la varianza será una matriz diagonal y su alrededor serán ceros. Cuando no se presenta esto se denomina auto correlación o correlación serial.

Prueba Final

1. ¿Cuáles son los pasos en la construcción de modelos?
2. ¿Cuáles son los problemas de regresión múltiple?
3. ¿Cuál es el procedimiento de comparación de las pendientes de dos o más líneas y comparación de dos o más curvas de respuestas?
4. Con una aplicación en su empresa realice la construcción de un modelo de regresión múltiple.

ACTIVIDAD

El estudiante debe realizar un proyecto aplicando la construcción de un modelo de regresión múltiple y analizar si es viable o no teniendo en cuenta el análisis de cálculos y gráficas.

6. GLOSARIO

Regresión lineal simple: Relación entre una variable independiente y una variable dependiente.

Regresión lineal Múltiple: Relación entre varias variables independientes y una variable dependiente.

Heteroscedasticidad: Este se presenta cuando no se cumple con la hipótesis de varianza constante para el término de la perturbación.

Multicolineallidad: El primero es que no se debe tener un alto grado de correlación lineal entre los diferentes tipos de variables independientes y la dependiente.

Autocorrelación: En el momento en que se trabaja con un modelo de regresión múltiple, donde se tiene que la primera variable vale 1 y es acompañada por el termino independiente

7. BIBLIOGRAFÍA

Fuentes bibliográficas

Greene, W. (1998) - “Análisis Econométrico” 3era. Ed., Prentice Hall Iberia SRL, páginas1075.

Johnston, J. y Dinardo, J. (2001) - “Métodos de Econometría”, 1era. Ed., Vinces Vives, páginas590.

Maddala, G. (1996) - “Introducción a la Econometría”, 2da. Ed., Prentice Hall Hispanoamericana, S.A., páginas715.

Novales, A. (1993)- “Econometría”, 2da. Ed., McGraw Hill, 1993. Páginas676.

7.1. Fuentes Digitales o Electrónicas

¿Qué es econometría?
www.icesi.edu.co/~jcalonso/bas/Seccion1.pdf

ColombiaLink.com - La Econometría - Economía y Finanzas
www.colombialink.com/01_INDEX/index_finanzas/13_econometria.html - 31k -

Programa Universidad Virtual
www.virtual.unal.edu.co/cursos/economicas/2001078/index.html

Econometría - Wikipedia, la enciclopedia libre
es.wikipedia.org/wiki/Economía_cuantitativa - 45k

Modelos econométricos y series temporales
books.google.com.co/books?isbn=8429126112.

Econometría: Modelos deterministas y estocásticos
books.google.com.co/books?isbn=8480040491.

Nelson Álvarez Vásquez. Aplicaciones de econometria
books.google.com.co/books?isbn=8480043784.