Estimación y predicción para la toma de decisiones: Métodos semilogarítmico y recíproco*

Diego Fernando Cardona Madariaga[1]
Javier Leonardo González Rodríguez[2]
Miller Rivera Lozano[3]
Edwin Hernán Cárdenas Vallejo[4]

Fecha de recepción:
Julio 26, 2013

Fecha de aceptación:
Agosto 28, 2013

Resumen

En este artículo se presentan las bondades de la estadística inferencial en lo referente al análisis de regresión no lineal.

Para tal fin, se recurre al estudio de los modelos semilogarítmico y el recíproco, mostrando su aplicación en la solución de problemas pertinentes a la administración, la ciencia y la ingeniería. Desarrollándose como ejemplo especifico, un caso real aplicado a la administración y la economía.

Palabras claves: Estadística inferencial, regresión no lineal, mmétodos semilogarítmico y recíproco

Decision Making Methods: Estimation and Trend Determination Using Reciprocal and Semi-Logarithmic Procedures

Abstract

In this article we present the benefits of inferential statistics in relation to non-linear regression analysis.

To this end, we resort to the study of semi-log and reciprocal models, showing its application in solving relevant problems to management, science and engineering. Developing as a specific example, a real case applied to the administration and the economy.

Keywords: Inferential statistic, non-linear regression analysis, reciprocal and semi-logarithmic procedures.

* Artículo derivado de la investigación de los autores por desarrollar referentes conceptuales, método lógicos y de aplicación de los modelos semilogarítmicos y recíproco en áreas administrativas e ingenierías.

1 Diego Fernando Cardona Madariaga. Matemático, Ingeniero Civil, Msc y PhD en Ciencias Administrativas; profesor titular de la Escuela de Administración de la Universidad del Rosario con funciones de director del Doctorado en Ciencias de la Dirección. Correo electrónico: diego.cardona@urosario.edu.co.

2 **Javier Leonardo González Rodríguez.** Médico, Especialista en Salud Pública y Ph. D. en Economía y Gestión de la Salud; profesor principal de la Escuela de Administración de la Universidad del Rosario con funciones de director de la Maestría en Administración de la Salud y Especializaciones de Gerencia en la Salud. Correo electrónico: javier.gonzalez@urosario.edu.co.

3 **Miller Rivera Lozano.** Ingeniero de Sistemas, Especialista en Auditoría de Sistemas e Ingeniería de Software, M. Sc. en Administración; con funciones de coordinador del Laboratorio de Modelamiento y Simulación en la Escuela de Administración de la Universidad del Rosario. Correo electrónico: miller.rivera@urosario.edu.co.

4 **Edwin Hernán Cárdenas Vallejo.** Ingeniero Electrónico, Especialista en Educación Matemática, aspirante a Magister en Educación, con funciones de docencia en la Secretaría de Educación Distrital de Bogotá y Corporación Unificada Nacional de Educación Superior. Correo electrónico: edwin_cardenas@cun.edu.co.

Introducción

El análisis de regresión es una herramienta estadística que permite, al investigador o al profesional, encontrar relaciones matemáticas entre las variables que intervienen en un proceso, experimento o situación que implique cambios en la información objeto de análisis.

Por medio de estas relaciones se hacen estimaciones y predicciones sobre el comportamiento futuro de esas variables y con las predicciones realizadas se procede a tomar decisiones administrativas, establecer políticas, o a realizar diseños con base en los modelos matemáticos.

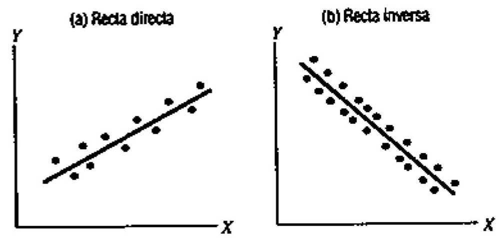
El término regresión fue utilizado por primera vez como un concepto estadístico en 1877 por sir Francis Galton, quien llevó a cabo un estudio que mostró que la estatura de los niños nacidos de padres altos tiende a retroceder o "regresar" hacia la estatura media de la población. Designó la palabra regresión como el nombre del proceso general de predecir una variable (la estatura de los niños) a partir de otra (la estatura del padre o de la madre). Posteriormente, los estadísticos acuñaron el término regresión múltiple para describir el proceso mediante el cual mediante la utilización de varias variables se puede predecir otra. (Devore, 2005).

En la terminología de la regresión, la variable que se va a predecir se llama dependiente, a explicar, o endógena, mientras que la o las variables que se usan para predecir el valor de la variable dependiente se identifican como variables independientes, explicativas o exógenas.

En general, existen cuatro posibles formas en que las variables se pueden relacionar:

Relación lineal directa e inversa.

Figura 1. Relación lineal.

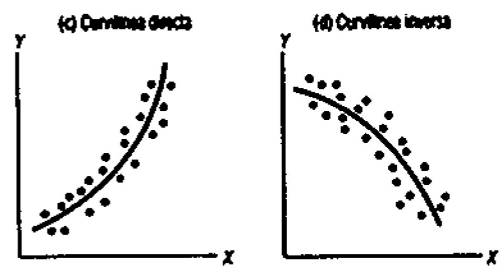


FUENTE: Levin & Rubin, 2004

Se considera que la relación es lineal si al graficar los datos se observa una distribución de los mismos alrededor de una línea recta (figura 1). Esta ecuación, se determina a través del análisis de regresión por medio del modelo lineal.

Relación no lineal directa e inversa

Figura 2. Relación no lineal.



FUENTE: Levin & Rubin, 2004

Si la relación es no lineal, los datos observados se distribuyen alrededor de una curva y la ecuación de esa curva se determina a partir de alguno de los modelos no lineales. En la práctica, es muy común encontrar que la mayoría de las aplicaciones en las que se establecen relaciones entre variables sea del tipo no lineal.

Los modelos de regresión no lineal se clasifican en:

1. Modelos curvilíneos que se ajustan al modelo lineal general.
2. Modelos intrínsecamente lineales.
3. Modelos no lineales específicos.

Los modelos no lineales que se ajustan al modelo lineal general, son aquellos que se pueden expresar como una combinación lineal entre las variables y los parámetros a estimar, se encuentran en este grupo: los modelos: cuadrático, cúbico, polinómicos de orden superior, semilogarítmico (linear-log) y el recíproco entre otros.

Los modelos intrínsecamente lineales son aquellos en los que los parámetros no son lineales pero se pueden hacer transformaciones algebraicas para linealizar la relación; siendo los más representativos: el modelo hiperbólico, el potencial y el exponencial.

Los modelos no lineales específicos son aquellos que no se pueden linealizar y se aplican otros métodos como el de estimación de máxima verosimilitud, el de Gauss-Newton, el de Newton-Raphson, entre otros.

En este artículo se describe el análisis de regresión donde intervienen una variable dependiente y una independiente, en una situación particular en administración y economía. El objetivo del análisis es encontrar que curva se ajusta mejor a los datos y decidir entre el modelo semilogarítmico (linear-log) y el recíproco.

El modelo lineal general

En una situación en la que se tiene una variable dependiente y, k y variables independientes ; $(x_1, x_2, \dots, x_k)$ se puede deducir una ecuación estimada de regresión que sea la mejor relación entre las variables por medio del modelo lineal general.

En este modelo la variable dependiente se puede expresar como la combinación lineal de las variables independientes y los parámetros.

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \cdots + \beta_p z_p + \epsilon \qquad (1)$$

En esta ecuación, cada una de las variables independientes z_j (donde $j = 1,2, \dots, p$) es función de las variables x_k (donde $k = 1,2, \dots, p$)

El caso más sencillo es cuando sólo se tienen datos de la variable x_1 y se desea estimar empleando una relación lineal directa. En esta situación, $z_1 = x_1$ y la ecuación (1) se transforma en:

$$y = \beta_0 + \beta_1 x_1 + \epsilon \qquad (2)$$

Dentro del léxico estadístico, este modelo se conoce como *modelo simple de primer orden con una variable predictora*. (Anderson, Sweeney, & Williams, 2001).

El modelo de regresión semilogaritmico o LINEAR-LOG

Al hacer $z_1 = \log x$ en la ecuación (1) se obtiene el modelo:

$$y = \beta_0 + \beta_1 \log x + \epsilon \qquad (3)$$

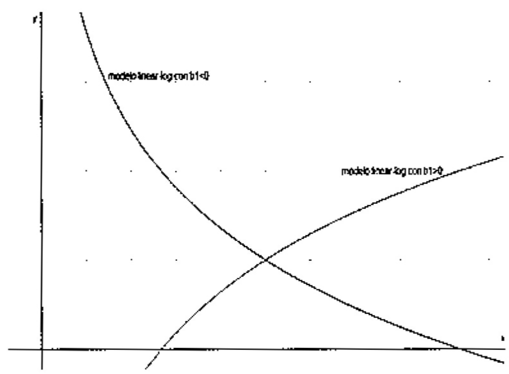
Por tanto, la ecuación estimada de ajuste de es:

$$\hat{y} = b_0 + b_1 \log x \qquad (4)$$

Este modelo se usa cuando se observan cambios porcentuales en x para cambios absolutos en y (Maul, 2010).

$$\Delta y = \beta_0 + \beta_1 \log(x + \Delta x) - [\beta_0 + \beta_1 \log x]$$
$$\Delta y = \beta_1 [\log(x + \Delta x) - \log x]$$
$$\Delta y \cong \beta_1 \left[\frac{\Delta x}{x}\right]$$

Figura 3. Modelo semilogarítmico linear–log



FUENTE: Elaboración propia

La figura (3), señala las posibles curvas que se pueden ajustar con este modelo. Nótese que para ambos casos $b_1 > 0$ y $b_1 < 0$ las curvas tienen asíntotas en x=0, es decir, que no cortan el eje y. además las curvas no son acotadas; es decir, que si x aumenta sin límite y también seguirá aumentando o decreciendo sin límite.

MODELO RECÍPROCO

Al hacer $z_1 = \frac{1}{x}$ en la ecuación (1) se obtiene el modelo:

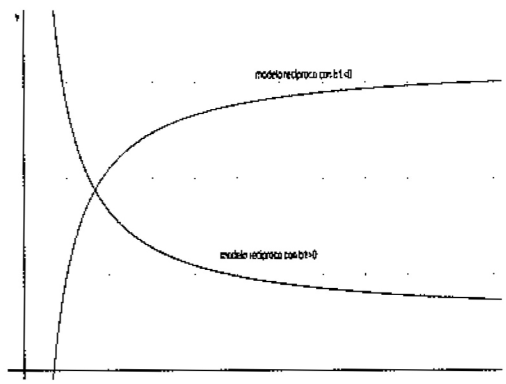
$$y = \beta_0 + \beta_1 \frac{1}{x} + \epsilon \qquad (5)$$

Por tanto, la ecuación estimada de ajuste de y es:

$$\hat{y} = b_0 + b_1 \frac{1}{x} \qquad (6)$$

Como se observa, este modelo es muy parecido al modelo semilogarítmico, que en la práctica es mucho más utilizado.

Figura 4. Modelo recíproco



FUENTE: Los autores

La semejanza de este modelo con el semilogarítmico radica en que tiene una asíntota vertical en x=0, pero su diferencia radica en que tiene además una asíntota horizontal en $y = b_0$. Esto quiere decir que la función recíproca es acotada, tiene un límite de crecimiento o de decrecimiento. (Figura 4).

El modelo recíproco se utiliza cuando el ajuste es mucho más significativo que el semilogarítmico y se trata de modelar algún tipo de relación de crecimiento o variación de tasas de crecimiento.

En economía existen varias situaciones que se modelan con la relación recíproca, entre ellas se resaltan: La curva de Philips y la curva de gasto de Engel la primera expresa la relación entre la tasa de cambio absoluta del salario monetario y la tasa de cambio porcentual del desempleo; mientras que la curva de gasto de Engel, muestra la existencia de un umbral o cota superior tal que por mucho que incremente el ingreso no se incrementará más el gasto de un determinado bien. (Sancho P., Serrano, & Pérez, 2006).

Caso de aplicación de los modelos de regresión

Con el fin de mostrar una aplicación de los modelos mencionados, se presenta la siguiente situación en el campo de la economía.

Tras 10 años de funcionamiento, una empresa del sector de las telecomunicaciones, quiere estudiar el beneficio obtenido en dicho periodo en función al número de clientes que utilizan sus servicios. Los datos quedan recogidos en la siguiente tabla:

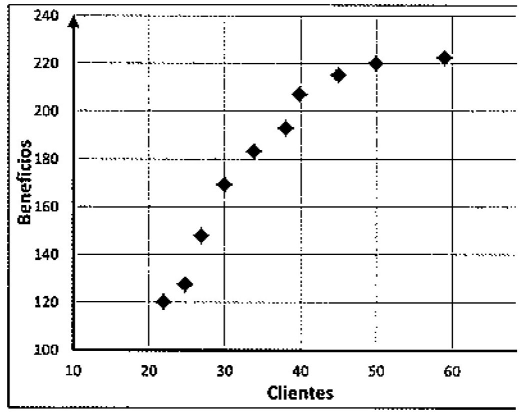
Tabla 1. Beneficios en relación al Nº de clientes de una empresa de telecomunicaciones.

Beneficios (millones U$)	Nº Clientes (miles)
120	22
127	25
148	27
169	30
183	34
293	38
207	40
215	45
220	50
222	59

FUENTE: Chica, 2011

Para encontrar el modelo que se ajusta mejor a los datos debe hacerse primero una gráfica de dispersión de los valores observados (Figura 5).

Figura 5. Gráfica de dispersión de los datos del ejemplo



FUENTE: Elaboración propia

Validándose entonces que los puntos describen una curva que puede ajustarse con el modelo semilogarítmico o el modelo recíproco.

Para hacer los respectivos análisis se recurrirá al uso de la herramienta de cálculo EXCEL® ingresando los datos como se presentan en la siguiente tabla.

Tabla 2. Datos de entrada para el análisis de regresión del ejemplo.

Nº Clientes (miles)			Beneficios (millones U$)
x	*Ln X*	*1/x*	*y*
22	3,0910434	0,04545455	120
25	3,21887681	0,04	127
27	3,29583787	0,03703704	148
30	3,40119842	0,03333333	169
34	3,5263616	0,02941176	183
38	3,63758727	0,02631579	193
40	3,68888058	0,025	207
45	3,80666365	0,02222222	215
50	3,9120242	0,02	220
59	4,07753869	0,01694915	222

FUENTE: Elaboración propia

Para el modelo semilogarítmico, los resultados son:

Tabla 3. Estadísticas de la regresión para el modelo semilogarítmico

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,96713461
Coeficiente de determinación R^2	0,93534935
R^2 ajustado	0,92726802
Error típico	10,2688088
Observaciones	10

FUENTE: Elaboración propia

Tabla 4. Prueba de significancia para el modelo semilogarítmico

ANÁLISIS DE VARIANZA		
	Grados de libertad	*Suma de cuadrados*
Regresión	1	12204,8125
Residuos	8	843,587481
Total	9	13048,4
Promedio de los cuadrados	*F*	*Valor crítico de F*
12204,8125	115,741998	4,9057E-06
105,448435		

FUENTE: Elaboración propia

Tabla 5. Parámetros estimados para el modelo semilogarítmico.

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*
Intercepción	-233,1353	38,5754803	-6,04361368	0,00030809
Variable X 1	115,979122	10,7803864	10,7583455	4,9057E-06

FUENTE: Elaboración propia

De acuerdo con la tabla (5), la ecuación estimada de regresión semilogarítmica es:

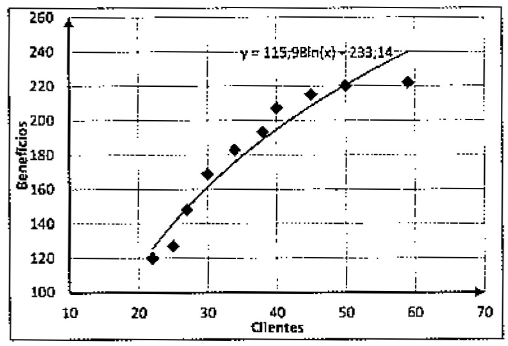
$$\hat{y} = -233{,}13 + 115{,}98 \ln x \qquad (7)$$

Se observa que el ajuste de este modelo es muy bueno porque el coeficiente de determinación (tabla 3), indica que la relación explica en un 93.54% los valores de a partir de los datos de.

Adicionalmente, la relación es estadísticamente significativa pues el valor crítico del estadístico F es cero. (Tabla 4).

Luego se gráfica la ecuación encontrada junto con las observaciones o datos registrados para comparar.

Figura 6. Gráfica de la ecuación del modelo semi logarítmico



FUENTE: Elaboración propia

Se observa que el ajuste de la curva semilogarítmica es bueno para todos los valores excepto el último donde la curva se empieza alejar de los valores observados y si la tendencia de los datos se mantiene, entonces esta ecuación no sirve para hacer estimaciones o proyecciones.

Para el modelo recíproco, los resultados son:

Tabla 6. Estadísticas de la regresión para el modelo recíproco

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,98675219
Coeficiente de determinación R^2	0,97367989
R^2 ajustado	0,97038987
Error típico	6,55205474
Observaciones	10

FUENTE: Elaboración propia

Tabla 7. Prueba de significancia para el modelo recíproco

ANÁLISIS DE VARIANZA		
	Grados de libertad	*Suma de cuadrados*
Regresión	1	12704,9646
Residuos	8	343,43537
Total	9	13048,4

Promedio de los cuadrados	*F*	*Valor crítico de F*
12704,9646	295,950056	1,3263E-07
42,9294213		

FUENTE: Elaboración propia

Tabla 8. Parámetros estimados para el modelo recíproco

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*
Intercepción	300,728056	7,29494352	41,2241789	1,3203E-10
Variable X 1	-4068,93318	236,521892	-17,203199	1,3263E-07

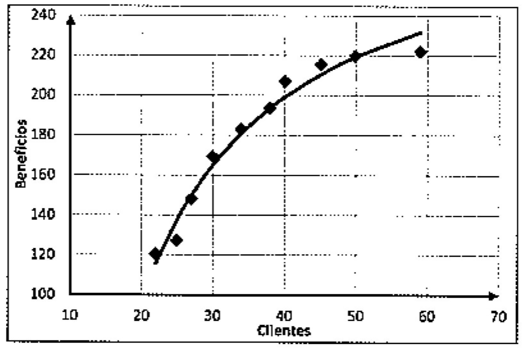
FUENTE: Elaboración propia

De acuerdo con la tabla (8) la ecuación estimada de regresión para el modelo recíproco es:

$$\hat{y} = 300{,}73 - 4068{,}93\frac{1}{x} \qquad (8)$$

Comprobándose así, que el ajuste de este modelo es mucho mejor porque el coeficiente de determinación (tabla 6), indica que la relación explica en un 97. 37% los valores de a partir de los datos de X. Adicionalmente, la relación es estadísticamente significativa pues el valor crítico del estadístico F es cero. (Tabla 7).

Figura 7. Gráfica de la ecuación del modelo recíproco.

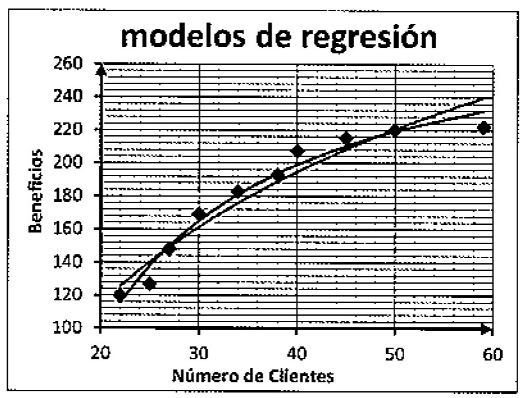


FUENTE: Elaboración propia

Al comparar la ecuación de regresión del modelo recíproco con los datos observados se nota un ajuste más apropiado para la relación entre beneficios y número de clientes.

Comparando gráficamente las dos ecuaciones se tiene:

Figura 8. Modelos de regresión semilogarítmico y recíproco para los datos del ejemplo



FUENTE: Elaboración propia

De acuerdo con los resultados obtenidos, el modelo más adecuado es el recíproco pues tiene un mejor ajuste (R^2=97.37%) y se observa en las curvas trazadas de la figura 8.

Estimación y predicción

Si el análisis de la ecuación de regresión obtenida con los datos demuestra que existe una relación estadísticamente significativa entre las variables, y si el ajuste que proporciona la ecuación es bueno, esa ecuación podría usarse para estimaciones y predicciones.

Estimación de intervalo

Al hacer una estimación puntual de y un valor de dado un valor de x, no se tiene idea alguna de la precisión asociada con el valor estimado. Por ello aunque la regresión tenga un gran ajuste y sea estadísticamente significativa, no se deben hacer estimaciones de valores de simplemente remplazando valores de en la ecuación.

El *estimado de intervalo de predicción* se usa cuando se desea un estimado de intervalo de valor individual de que corresponda a determinado valor de.

En la estimación y la inferencia, un error común es suponer que la línea de regresión, así el ajuste sea muy bueno (valor de r^2 muy alto), puede aplicarse en cualquier intervalo de valores. Aun cuando una relación se cumpla para el intervalo de puntos de la muestra, puede existir una relación completamente distinta para un intervalo diferente. Por ejemplo, la relación edad y talla puede ser lineal para cierto intervalo del crecimiento de los niños en su primera infancia pero en la adolescencia esa relación ya no es lineal.

Una ecuación de estimación es válida para el mismo rango dentro del cual se tomó la muestra inicialmente (Levin & Rubin, 2004). Sin embargo, si el investigador tiene la certeza de que el comportamiento entre las variables será el mismo en otros intervalos fuera del rango de la muestra, entonces puede usar la ecuación para hacer predicciones.

En la situación que se plantea como ejemplo para el análisis de regresión se supondrá que los beneficios crecen conforme aumenta el número de clientes con un límite superior; es decir, el modelo recíproco será válido para hacer predicciones fuera del rango de los datos observados.

Supóngase que se desea obtener el monto de los beneficios estimado para la empresa cuando sus clientes lleguen a ser 70.000.

Reemplazando X por 70 en la ecuación (8), se tiene:

$$\hat{y} = 300{,}73 - 4068{,}93\frac{1}{70} = 242{,}6$$

Para determinar un estimado del intervalo de predicción debemos determinar primero la varianza asociada al empleo de $\hat{y}$ como estimado de un valor individual de y. Esta varianza está formada por la suma de dos componentes:

1. La varianza de los valores individuales de respecto del promedio cuyo estimado es s^2
2. La varianza asociada al uso de para estimar E (y) cuyo estimado es $s_{\hat{y}_p}$.

Así, el estimado de la varianza de un valor individual es:

$$s_{ind}^2 = s^2 + s_{\hat{y}_p}$$

Por consiguiente, un estimado de la desviación estándar de un valor un individual de $\hat{y}$ es:

$$s_{ind} = s\sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{S_{xx}}}$$

La ecuación general para un estimado del intervalo de predicción para un valor individual de dado un valor particular de x es:

$$\hat{y} \pm t_{\alpha/2} \cdot s_{ind}$$

En donde el coeficiente de confianza es $1-\alpha$ y $t_{\alpha/2}$ se basa en una distribución t con n–2 grados de libertad.

Para determinar un estimado de intervalo de predicción del 95% para el valor de los beneficios esperados por la empresa cuando sus clientes sean 70.000, se necesita el valor de t para α/2=0.025 y n–2= 8 grados de libertad. Así, con $\hat{y}_p = 242{,}6$, $t_{0.025}=2{,}306$ y $s_{ind} = 6{,}9117$ se tiene:

$$242{,}6 \pm 2{,}306 \cdot 6{,}9117$$

$$242{,}6 \pm 15{,}938$$

Entonces, con una confianza del 95% se puede decir que el valor de los beneficios recibidos por la empresa cuando sus clientes sean 70.000 se encuentra entre 226,66 y 258,54 millones de dólares.

Estimación de los parámetros del modelo de regresión

Uno de los conceptos fundamentales sobre el que se ha basado este análisis, es que la ecuación de regresión cuadrática obtenida a partir de los datos de la muestra es un estimado de los parámetros del modelo para la población. Por lo tanto, es posible determinar intervalos de confianza para los coeficientes de la ecuación de regresión.

El análisis de regresión con EXCEL® muestra, por defecto, el intervalo de confianza del 95% para cada uno de los parámetros.

Tabla 9. Intervalo de confianza de los parámetros.

	Coeficientes	*Inferior 95%*	*Superior 95%*
Intercepción	300,728056	283,905887	317,550226
Variable X 1	-4068,938	-4614,353	-3523,5127

FUENTE: Los autores

Finalmente, es importante mencionar que se puede cometer otro error al utilizar el análisis de regresión, y es suponer que un cambio en una variable es "ocasionado" por un cambio en la otra variable. Los análisis de regresión y correlación no pueden, de ninguna manera, determinar la causa y el efecto. Si se dice, por ejemplo, que existe una relación entre el número de canas y de arrugas que van apareciendo en una persona, no se puede decir que una ocasiona la otra pues es muy posible que existan otras variables asociadas que sean la causa; en este caso la edad de la persona, por ejemplo.

La validez de una conclusión de tipo causa y efecto requiere de una justificación teórica, o del buen juicio por parte del analista. (Anderson, Sweeney, & Williams, 2001).

Conclusiones

El análisis de regresión es una herramienta matemática poderosa que permite determinar modelos sobre el comportamiento de las variables que intervienen en una situación en cualquier campo del conocimiento con el fin de hacer estimaciones y predicciones dentro de un intervalo de confianza deseado.

Dentro de estos modelos, que pueden ser lineales o no lineales, se encuentra el modelo de regresión semilogarítmico y el modelo de regresión recíproco que se ajusta de manera adecuada a situaciones aplicadas a la administración y la economía, como la relación entre el beneficio obtenido por una empresa de telecomunicaciones y el número de clientes.

Esta herramienta de análisis estadístico le da al profesional la posibilidad de hacer ajustes en los procesos, tomar decisiones o establecer políticas. Por ejemplo, si un profesional de la administración utiliza la regresión semilogarítmica puede encontrar la relación entre la variación de los ingresos o utilidades con respecto a variaciones porcentuales en el precio de un producto. Mientras que con la regresión recíproca puede determinar la cota superior o límite de crecimiento de las ganancias o utilidades en función de la producción.

Asimismo, el presidente o gerente de la empresa de telecomunicaciones puede usar los resultados del ejemplo para decidir si invierte en más publicidad y promociones para captar más clientes o si por el contrario al observar la tendencia del límite en las utilidades decide invertir en tecnología y mejoras en el servicio.

Bibliografía

Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2001). Estadística para administración y economía (7a ed., Vol. II). México: Thomson.

Chica, J. (2011). *Proyecto de Innovación Docente: Guía multimedia para la elaboración de un modelo econométrico.* Obtenido de Universidad de Granada: http://www.ugr.es/~jchica/Pagina2/GUIME/Otros%20ejercicios/Ejercicios%20Modelos%20Linealizables.pdf

De Mendiburu, F. (2006). *Modelos no lineales.* Obtenido de Universidad Nacional Agraria La Molina: http://tarwi.lamolina.edu.pe/~fmendiburu/index-filer/academic/Foresteria%20I/Teoria/Teoria%20modelos%20no%20lineales.pdf

Devore, J. L. (2005). *Probabilidad y estadística para ingeniería y ciencias* (6a ed.). México: Thomson Learning.

Evans, M., & Rosenthal, J. S. (2005). *Probabilidad y estadística. La ciencia de la incertidumbre.* Barcelona, España: Reverté S.A.

Freund, J. E., & Simon, G. A. (1994). *Estadística elemental* (8a ed.). México: Prentice Hall.

Levin, R. I., & Rubin, D. S. (2004). *Estadística para administración y economía.* México: Pearson Educación.

Maul, H. (2010). *Modelo de Regresión múltiple parte I.* Obtenido de Universidad Francisco Marroquin: http://www.youtube.com/watch?v=pMKQKJY0jXI&list=PLD307E757E5488563

Mendoza, H., & Bautista, G. (2002). *Bioestadística fundamental.* Obtenido de Universidad Nacional de Colombia: http://www.virtual.unal.edu.co/cursos/ciencias/2001091/

Mendoza, H., Vargas, J., López, L., & Bautista, G. (2002). *Métodos de regresión.* Obtenido de Universidad Nacional de Colombia: http://www.virtual.unal.edu.co/cursos/ciencias/2007315/

Mendoza, H., Vargas, J., López, L., & Bautista, G. (2002). *Métodos de regresión.* Obtenido de Universidad Nacional de Colombia: http://www.virtual.unal.edu.co/cursos/ciencias/2007315/

Miller, I. (2000). *Estadística matemática con aplicaciones.* (6a ed.). México: Pearson Educación.

Muñoz R., L. A. (2006). *Comprobación de los supuestos del modelo de regresión lineal.* Obtenido de Universidad Autónoma de Occidente: http://augusta.uao.edu.

co/moodle/file.php/284/18_supuestos_de_la_regresion_lineal.pdf

Pacheco, P. (2012). *Validación de supuestos.* Obtenido de Universidad Nacional de Colombia: http://www.virtual.unal.edu.co/cursos/ciencias/dis_exp/und_3/pdf/validaciondesupuestos-unidad 3b[1].pdf

Rivera, M., & Cárdenas V., E. H. (2013). *Módulo de regresión lineal.* Bogotá: Universidad del Rosario.

Sancho P., A., Serrano, G., & Pérez, P. (2006). *Econometría de Económicas.* Obtenido de Universidad de Valencia: http://www.uv.es/~sancho/panel.pdf

Walpole, R. E., & Myers, R. H. (1999). *Probabilidad y estadística para ingenieros* (6a ed.). México: Prentice Hall.