

ESTIMACIÓN Y PREDICCIÓN CON EL MODELO DE REGRESIÓN CÚBICA APLICADO A UN PROBLEMA DE SALUD

Diego Cardona¹, Javier González², Miller Rivera³, Edwin Cárdenas⁴

¹Director del Doctorado en Ciencias de la Dirección

²Director de la Maestría en Administración en Salud y de las Especializaciones de Gerencia en Salud

³Coordinador del Laboratorio de Modelamiento y Simulación, Escuela de Administración

Correo electrónico: miller.rivera@urosario.edu.co

Universidad del Rosario, Bogotá, Colombia

⁴Profesor de la Facultad de Ingeniería, Departamento de Ciencias Básicas

Corporación Universitaria CUN, Bogotá, Colombia

Recibido: 20 de enero del 2014. Aprobado: 10 de abril del 2014.

Cómo citar este artículo: D. Cardona, J. González, M. Rivera y E. Cárdenas, "Estimación y predicción con el modelo de regresión cúbica aplicado a un problema de salud". *Ingeniería Solidaria*, vol. 10, n.º 17, pp. 153-160, en.-dic., 2014. doi: <http://dx.doi.org/10.16925/in.v9i17.828>

Resumen. El artículo corresponde a un proyecto de investigación desarrollado en la Escuela de Administración de la Universidad del Rosario, dirigido a fortalecer la utilización de los métodos inferenciales de regresión lineal, no lineal y múltiple en la ejecución de procesos de toma de decisión, a través de la construcción de materiales didácticos para estudiantes, docentes e investigadores. Este artículo muestra las bondades del modelo de regresión polinómica de tercer orden y su aplicación en la administración y la ciencia, mediante el desarrollo de un caso real aplicado a la salud, en el que se estima el porcentaje de mujeres que consumen más de 20 cigarrillos diarios según la edad. Dentro del proyecto de investigación, iniciado el segundo semestre del 2012, se ha realizado la publicación de diferentes recursos didácticos entre los que se encuentran documentos de investigación como: "Una aproximación de la variable aleatoria a procesos de toma de decisión que implican condiciones de riesgo e incertidumbre", "Aplicación de colas de Poisson en procesos de 'toma de decisiones' en la gestión de servicios médicos" y guías de inferencia estadística de los métodos de regresión lineal y no lineal.

Palabras clave: inferencia estadística, regresión no lineal, modelo cúbico, estimación, predicción.

ESTIMATE AND PREDICTION WITH CUBIC REGRESSION MODEL APPLIED TO A HEALTH PROBLEM

Abstract. The article corresponds to a research project carried out at the School of Administration of the Universidad del Rosario, aimed at strengthening the use of inferential linear, nonlinear and multiple regression methods in decision-making processes by creating didactic materials aimed at students, teachers and researchers. This article shows the advantages of the third order polynomial regression model and its application in administration and science, through the development of a real case applied to health, in which the percentage of women who consume more than 20 cigarettes per day is estimated according to age. As part of the research project begun during the second half of 2012, diverse didactic guides have been published, including research documents such as: "An Approach using the Aleatorical Variable in Decision-Making Processes that Imply Conditions of Risk and Uncertainty" ("Una aproximación de la variable aleatoria a procesos de toma de decisión que implican condiciones de riesgo e incertidumbre"), "Application of Poisson Tails in 'Decision-making Processes' for Managing Medical Services" ("Aplicación de cola de Poisson en la gestión de servicios médicos") and statistical inference guides for linear and nonlinear regression methods.

Keywords: statistical inference, non-linear regression, cubic model, estimate, prediction.

ESTIMAÇÃO E PREDIÇÃO COM O MODELO DE REGRESSÃO CÚBICA APLICADO A UM PROBLEMA DE SAÚDE

Resumo. Este artigo corresponde a um projeto de pesquisa desenvolvido na Escola de Administração da Universidade do Rosario (Colômbia), dirigido a fortalecer a utilização dos métodos inferenciais de regressão linear, não linear e múltipla na execução de processos de tomada de decisão, por meio da construção de materiais didáticos dirigidos a estudantes, docentes e pesquisadores. Além disso, mostra os benefícios do modelo de regressão polinomial de terceira ordem e sua aplicação na administração e na ciência, mediante o desenvolvimento de um caso real aplicado à saúde, no qual se estima a porcentagem de mulheres que consomem mais de 20 cigarros diários segundo a idade. Dentro do projeto de pesquisa, iniciado no segundo semestre de 2012, realizou-se a publicação de diferentes recursos didáticos entre os quais se encontram documentos de pesquisa como: "Uma aproximação da variável aleatória a processos de tomada de decisão que implicam condições de risco e incerteza", "Aplicação do modelo de distribuição de Poisson em processos de 'tomada de decisões' em gestão de serviços médicos" e guias de inferência estatística dos métodos de regressão linear e não linear.

Palavras-chave: inferência estatística, regressão não linear, modelo cúbico, estimação, predição.



1. Introducción

Este artículo hace parte de una serie de documentos producto de una investigación docente, cuyo objetivo es fomentar y fortalecer el uso de técnicas de estadística inferencial que mejoren la utilización de los datos que recopilan los estudiantes de pregrado o de posgrado en sus trabajos e investigaciones aplicadas en diferentes campos del conocimiento.

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 [1].

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

En la terminología de la regresión [2], [3], la variable que se va a predecir se llama *dependiente*, a explicar, o *endógena*. La o las variables que se usan para predecir el valor de la variable dependiente se llaman *variables independientes, explicativas o exógenas*.

La relación entre las variables puede ser de dos tipos: la relación lineal y la relación no lineal. Se considera que la relación es lineal si al graficar los datos se observa una distribución de estos alrededor de una línea recta. La ecuación de la recta se determina a través del análisis de regresión por medio del modelo lineal. 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.

Uno de los modelos no lineales más utilizados es el modelo polinómico que procede del modelo lineal general [4], como se muestra en (1).

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p + \varepsilon \quad (1)$$

En la cual cada una de las variables independientes z_j (donde $j = 1, 2, \dots, p$) es una potencia de la variable independiente (x^k donde $k = 1, 2, \dots, p$).

Teniendo en cuenta lo anterior, se pueden proponer modelos de regresión cuadrática, cúbica o de orden superior, donde depende exclusivamente de x_1 , como se muestra en (2).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \dots + \beta_p x_1^p + \varepsilon \quad (2)$$

El interés particular de este artículo es mostrar situaciones en las que se aplica la regresión polinómica con $k \geq 3$. Dentro de las aplicaciones que se modelan con funciones polinómicas de orden mayor que dos se pueden mencionar fenómenos de la física y la ingeniería relativos a la medición de flujos o caudales; el llenado o vaciado de tanques o cualquier tipo de recipientes; y situaciones en las cuales el volumen varía en función del tiempo y de la geometría del recipiente [5].

En Economía, para determinar los valores en los cuales el ingreso promedio es máximo o igual al ingreso marginal, se deben hacer cálculos sobre funciones de ingreso cúbicas o de orden superior [6].

En Medicina, por medio de una función cúbica se puede determinar el número de semanas de gestación del feto en relación con su longitud [7].

Infortunadamente, en la práctica profesional, es común confiar en la intuición para juzgar cómo se relacionan dos variables. Debido a ello, se hacen pronósticos a tientas e incluso temerarios; sin embargo, si se tiene la posibilidad de tomar datos y utilizar un procedimiento estadístico de análisis para determinar cómo lo conocido se relaciona con el evento futuro, se podrían tomar decisiones acertadas para mejorar procesos y procedimientos que ocurren en casi todos los campos del conocimiento.

En este artículo, se describe el análisis de regresión donde intervienen una variable dependiente y una independiente, en una situación particular del campo de la salud, en la cual la relación entre ellas se aproxima por medio de una ecuación cúbica.

2. Materiales y métodos

Para hacer el análisis de regresión y determinar la relación entre las variables, se debe:

- Hacer una gráfica de los datos para observar qué clase de curva describen los puntos y por tanto tener una idea del modelo matemático a usar
- Encontrar la ecuación particular que mejor se ajuste a los datos y demostrar que la ecuación particular encontrada cumple con ciertos aspectos referentes a los méritos de esta para hacer pronósticos [8].

Con el fin de poner en práctica esta metodología, se hace uso de una situación real de las ciencias de la

salud. En una encuesta de salud realizada en el 2009 en España, se preguntó a las mujeres fumadoras (cerca de 4 millones) por el número de cigarrillos que fumaban diariamente [9]. La encuesta arrojó los datos que se muestran en la tabla 1.

Tabla 1. Porcentaje de mujeres en España que consumen cigarrillos diariamente, por edad y número de unidades

Mujeres	Porcentaje de consumidoras		
	Entre 1 y 9 unidades	Entre 10 y 19 unidades	20 o más unidades
De 16 a 24 años	43,7	34,2	22,1
De 25 a 34 años	41,3	39,5	19,2
De 35 a 44 años	32,9	39,6	27,5
De 45 a 54 años	29,5	34,0	36,5
De 55 a 64 años	23,7	32,7	43,6
De 65 a 74 años	31,4	32,3	36,3
De 75 y más años	59,3	34,1	6,6

* Se ajusta el último intervalo de la tabla 1 de 75 a 84 años, asumiendo que el porcentaje de mujeres mayores de 84 que sean fumadoras no es significativo

Fuente: [9]

Luego se procede a realizar la gráfica de dispersión para observar si existe una posible relación y de qué tipo es (véase figura 1).

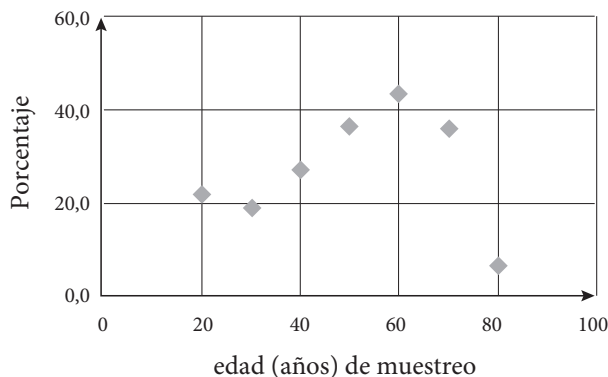


Figura 1. Porcentaje de mujeres españolas consumidoras de 20 o más cigarrillos (según los datos de la tabla 1)

Fuente: elaboración propia

La figura 1 de dispersión nos sugiere que existe una relación curvilínea entre las variables. Se utiliza primero el modelo de regresión cuadrática, dado por (3).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \varepsilon \quad (3)$$

Con ayuda de la hoja Excel® se hace el análisis de regresión, que arroja la información de la tabla 2.

Tabla 2. Análisis de regresión cuadrática

Estadísticas de la regresión		
Coefficiente de correlación múltiple	0,76448317	
Coefficiente de determinación R^2	0,584434517	
R^2 ajustado	0,376651775	
Error típico	9,946306565	
Observaciones	7	
ANÁLISIS DE VARIANZA		
	Grados de libertad	Suma de cuadrados
Regresión	2	556,5190857
Residuos	4	395,7160571
Total	6	952,2351429
Promedio de los cuadrados	F	Valor crítico de F
278,2595429	2,81272	0,172694671
98,92901429		

Fuente: elaboración propia

El coeficiente de determinación expresado como un porcentaje indica la variación de los valores de la variable independiente que se puede explicar con la ecuación de regresión [10]; la relación matemática se muestra en (4).

$$r^2 = \frac{SSR}{SST} \quad (4)$$

Donde, usualmente, el coeficiente de determinación aumenta siempre a medida que se agregan variables independientes (z_j) al modelo general de la ecuación (1). Por lo tanto, se prefiere ajustar r^2 para evitar una sobre estimación del impacto de agregar otra variable independiente.

El coeficiente de determinación ajustado es [11], como se muestra en (5).

$$SSR = \sum (\hat{y}_i - \bar{y})^2 \quad (5)$$

$$SSR = \sum (y_i - \bar{y})^2$$

Donde n es el número de observaciones y p es el grado del polinomio del modelo caracterizado por la ecuación(2).

Así, pues, al aumentar el grado del polinomio el ajuste de la curva a las observaciones será mejor pero el número de datos también debe ser mayor [12], [13].

De acuerdo con la tabla 2, el coeficiente de determinación R^2 y el de determinación ajustada son bajos; por lo tanto, no hay un buen ajuste de la curva con los puntos (figura 2). Además, el valor del estadístico de prueba F es menor que el valor observado en una tabla de distribución F para un nivel de significancia del 0,05 con dos grados de libertad en el numerador y cuatro en el denominador ($2,81 < 6,94$); por lo tanto, este modelo de regresión no es significativo.

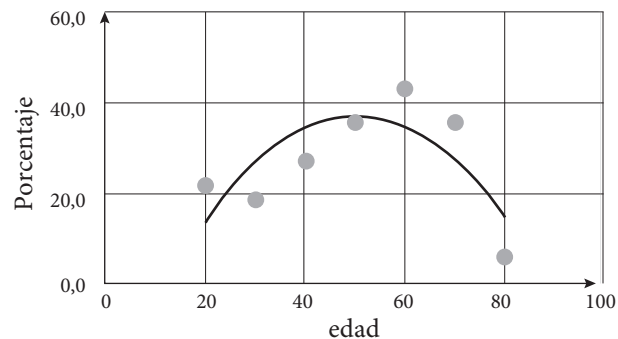


Figura 2. Función cuadrática estimada de ajuste
Fuente: elaboración propia

Dado que la regresión cuadrática no es significativa, se procede a hacer el análisis con el modelo cúbico. Al tomar sólo los cuatro primeros términos de la ecuación (2), se tiene el modelo de regresión cúbica presentado en (6).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3 + \epsilon \quad (6)$$

Este es el modelo de tercer orden con una variable predictora [11] y la ecuación estimada de regresión se aprecia en (7).

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + b_3 x^3 \quad (7)$$

Una vez más con ayuda de la hoja de cálculo Excel®, se desarrolla el análisis, y se obtiene los resultados mostrados en la tabla 3 y 4.

Los resultados de la regresión muestran que el ajuste de la ecuación cúbica con las observaciones es muy alto ($r^2 = 99\%$; tabla 3).

La relación encontrada es estadísticamente significativa ya que el estadístico de prueba F es mucho ma-

Tabla 3. Análisis de regresión del modelo cúbico

Estadísticas de la regresión				
Coefficiente de correlación múltiple				0,99883467
Coefficiente de determinación R^2				0,9976707
R^2 ajustado				0,9953414
Error típico				0,85985279
Observaciones				7
	Coefficientes	Error típico	Estadístico t	Probabilidad
Intercepción	94,198	5,6897	16,5558	0,0005
Variable X 1	-6,59225	0,4091	-16,115	0,00052
Variable X 2	0,17673	0,0088	20,0241	0,000272
Variable X 3	-0,00135	5,8E-05	-23,07	0,00018

Fuente: elaboración propia

Tabla 4. Prueba de significancia de la regresión cúbica

ANÁLISIS DE VARIANZA		
	Grados de libertad	Suma de cuadrados
Regresión	3	950,017102
Residuos	3	2,21804048
Total	6	952,235143
Promedio de los cuadrados	F	Valor crítico de F
316,672367	428,313691	0,00019071
0,73934683		

Fuente: elaboración propia

por que el valor observado en una tabla de distribución F con una significancia de 0,05 y tres grados de libertad en el numerador y en el denominador $F_{0,5} = 9,28$. Por ello, la probabilidad o valor crítico es casi cero (tabla 4).

Al remplazar los valores de los coeficientes de regresión (tabla 3) en (7) se tiene (8).

$$\hat{y} = 94,198 - 6,5922x + 0,176729x^2 - 0,001349x^3 \quad (8)$$

Al graficar la curva de (8) se observa un ajuste casi perfecto con los datos originales (figura 3).

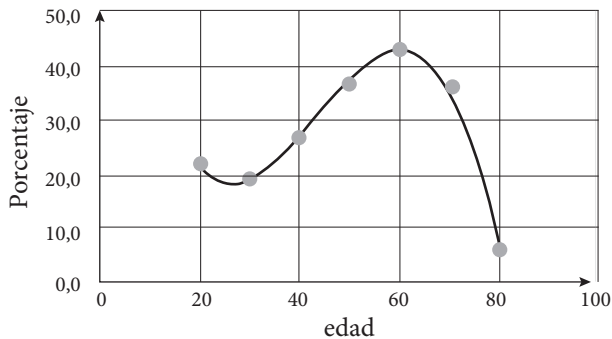


Figura 3. Gráfica de la ecuación de regresión cúbica
Fuente: elaboración propia

Sin embargo, no se puede pasar por alto la validación de los supuestos del modelo acerca del término de error: la homocedasticidad, $E(\epsilon) = 0$ y distribución normal de ϵ [14], [15], [16].

La relación cúbica encontrada cumple con los supuestos del modelo, dado que los residuos estandarizados (e_{zi}) se encuentran entre -2 y 2 desviaciones estándar (tabla 5 y figura 4) demostrando con ello una distribución normal del factor de error y no se evidencia aumento en la varianza conforme aumenta \hat{y} .

Tabla 5. Análisis de residuales de la relación cúbica

		Análisis de los residuales		
		Pronóstico para y	Residuos	Residuos estandarizados
x	y	\hat{y}	$e_i = y - \hat{y}$	e_{zi}
20	22,1	22,2469	-0,16690476	-0,27451
30	19,2	19,0445	0,19547619	0,3215
40	27,5	26,8931	0,63690476	1,0475
50	36,5	37,6943	-1,16428571	-1,91492
60	43,6	43,34977	0,2202381	0,36223
70	36,3	35,7612	0,52880952	0,86974
80	6,6	6,8302	-0,2502381	-0,041157

Fuente: elaboración propia

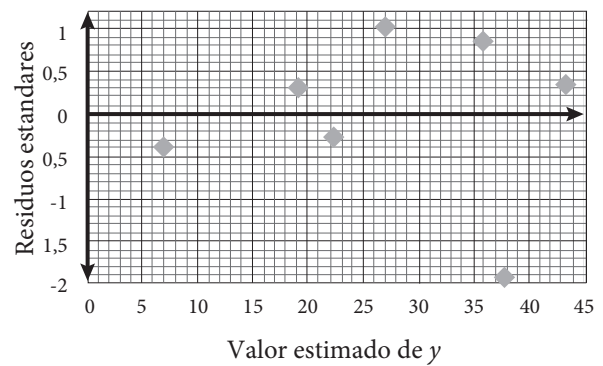


Figura 4. Gráfica de residuales de la relación cúbica
Fuente: elaboración propia

3. Resultados

El análisis de regresión con la ecuación cúbica obtenida a partir de los datos demuestra que existe una relación estadísticamente significativa entre las variables, y el ajuste que proporciona es mejor que lo obtenido con la ecuación cuadrática. Por tal motivo, esta expresión puede usarse para hacer estimaciones y predicciones de valores de la variable dependiente (porcentaje de mujeres que fuman 20 o más cigarrillos diarios) a partir de valores de la variable independiente (edad de las mujeres) que están dentro del rango de la muestra, pero diferentes a los observados, como se detalla a continuación.

3.1 Estimación de intervalo

Al hacer una estimación puntual de un valor de y 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 y simplemente remplazando valores de x en (8).

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

Supóngase que se desea estimar el porcentaje de mujeres fumadoras de 50 años de edad que fuman 20 o más cigarrillos diarios. Remplazando x por 50 en (8), se tiene:

$$\hat{y} = 94,198 - 6,5922(50) + 0,176729(50)^2 - 0,001349(50)^3$$

$$\hat{y} = 37,69428571$$

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:

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

Así, el estimado de la varianza de un valor individual está dado por (9):

$$s_{ind}^2 = s^2 + s_{\hat{y}_p}^2 \quad (9)$$

Por consiguiente, un estimado de la desviación estándar de un valor individual de \hat{y} está dado por (10).

$$s_{ind} = s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{S_{xx}}} \quad (10)$$

La ecuación general para un estimado del intervalo de predicción para un valor individual de y 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 $n-4$ grados de libertad.

Para determinar un estimado de intervalo de predicción del 95% para el porcentaje de mujeres fumadoras de 50 años de edad que fuman 20 o más cigarrillos diarios, se necesita el valor de t para $\alpha/2=0.025$ y $n-4=3$ grados de libertad. Así, con $\hat{y}_p = 37,6943$ $t_{0,025} = 3,182$ y $s_{ind} = 0,859853$, se tiene:

$$37,6943 \pm 3,182 \cdot 0,859853$$

$$37,6943 \pm 2,736$$

Entonces, con una confianza del 95% se puede decir que el porcentaje de mujeres fumadoras de 50 años de edad que fuman 20 o más cigarrillos diarios se encuentra entre 34,96% y 40,43%.

4. Discusión

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 cúbica 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.

La tabla 6 muestra los valores inferior y superior para el intervalo de confianza del 95% de los parámetros del modelo.

Tabla 6. Intervalos de confianza

	Coefficientes	Inferior 95%	Superior 95%
Intercepción	94,19786	76,0906525	112,305062
Variable X 1	-6,59225	-7,894107	-5,2904
Variable X 2	0,17673	0,148642	0,20481753
Variable X 3	-0,00135	-0,0015359	-0,00116353

Fuente: elaboración propia

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 [10]. 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 particular, para la situación que se ha analizado no se puede asegurar que el comportamiento de las

mujeres fumadoras mantenga la misma tendencia que describe la ecuación hallada para aquellas con edades superiores a 84 años y menores de 16 años, pues si se asume que el consumo de cigarrillo en niñas inicia desde los 14 años y se utiliza la ecuación para determinar el porcentaje, se obtiene un 32,8% que sería aproximadamente un 50% mayor que el consumo a la edad de 20 años. Por esta razón, solo se puede utilizar la ecuación cúbica encontrada para conocer el porcentaje de consumo de 20 o más cigarrillos diarios en mujeres de cualquier edad específica en el rango de 16 a 84 años.

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 [17].

5. 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 polinómico de tercer orden que se ajusta de manera adecuada a situaciones aplicadas a la salud como el porcentaje de mujeres españolas que consumen más de 20 cigarrillos diarios con relación a su edad. Este modelo también es muy utilizado en economía para la solución de problemas de optimización de precios e ingresos.

Esta herramienta de análisis estadístico proporciona 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 de tercer orden podría estimar el ingreso promedio máximo en un proceso de venta, así como optimizar el proceso de inventario en los almacenes. De igual forma, los funcionarios de la administración de la

salud pública en España, a partir del análisis de regresión hecho con respecto al consumo de cigarrillos observado en las mujeres, podrían establecer políticas que disminuyan estos indicadores.

Referencias

- [1] J. E. Freund y G. A. Simon, *Estadística elemental*, 8a ed., México: Prentice Hall, 1994.
- [2] J. L. Devore, *Probabilidad y estadística para ingeniería y ciencias*, 6a ed., México: Thomson Learning, 2005.
- [3] R. E. Walpole y R. H. Myers, *Probabilidad y estadística para ingenieros*, 6a ed., México: Prentice Hall, 1999.
- [4] H. Mendoza, J. Vargas, L. López y G. Bautista, “Métodos de regresión”, 2002. [En línea]. Disponible en: <http://www.virtual.unal.edu.co/cursos/ciencias/2007315/> [Último acceso: 2 octubre 2013].
- [5] F. De Mendiburu, “Modelos no lineales”, 2006. [En línea]. Disponible en: <http://tarwi.lamolina.edu.pe/~fmendiburu/index-filer/academic/Foresteria%20I/Teoria/Teoria%20modelos%20no%20lineales.pdf> [Último acceso: 12 septiembre 2013].
- [6] A. Sancho y G. Serrano, “Econometría de Económicas: apuntes para el tema 6”, 2006. [En línea]. Disponible en: <http://www.uv.es/~sancho/panel.pdf>
- [7] H. Mendoza y G. Bautista, “Bioestadística fundamental”, 2002. [En línea]. Disponible en: <http://www.virtual.unal.edu.co/cursos/ciencias/2001091/>
- [8] L. A. Muñoz R., “Comprobación de los supuestos del modelo de regresión lineal”, 2006. [En línea]. Disponible en: http://augusta.uao.edu.co/moodle/file.php/284/18_supuestos_de_la_regresion_lineal.pdf [Último acceso: 26 noviembre 2013].
- [9] Ministerio de Sanidad, Servicios Sociales e Igualdad, “Encuesta Europea de Salud en España”, 2009. [En línea]. Disponible en: https://www.msssi.gob.es/estadEstudios/estadisticas/EncuestaEuropea/Principales_Resultados_Informe.pdf [Último acceso: 20 octubre 2013].
- [10] R. I. Levin y D. S. Rubin, *Estadística para administración y economía*, México: Pearson Educación, 2004.
- [11] D. R. Anderson, D. J. Sweeney y T. A. Williams, *Estadística para administración y economía*, 7a ed., vol. II, México: Thomson, 2001.
- [12] A. Novales, *Econometría*, 2ª ed., Madrid: McGraw-Hill, 1993.
- [13] M. Evans y J. S. Rosenthal, *Probabilidad y estadística. La ciencia de la incertidumbre*, Barcelona: Reverté S.A., 2005.

- [14] L. Orellana, "Análisis de regresión", 2008. [En línea]. Disponible en: http://www.dm.uba.ar/materias/estadistica_Q/2011/1/clase%20regresion%20simple.pdf [Último acceso: 15 diciembre 2013].
- [15] C. M. Lopera, "Análisis de Residuales", 2002. [En línea]. Disponible en: [http://www.docentes.unal.edu.co/cmlopera/docs/Estad2/2_RLM/2.\(Complemento\)Análisis de Residuales y Otros en RLM.pdf](http://www.docentes.unal.edu.co/cmlopera/docs/Estad2/2_RLM/2.(Complemento)Análisis de Residuales y Otros en RLM.pdf) [Último acceso: 10 noviembre 2013].
- [16] P. Pacheco, "Verificación de supuestos", 2012. [En línea]. Disponible en: [http://www.virtual.unal.edu.co/cursos/ciencias/dis_exp/und_3/pdf/validaciondesupuestosunidad3b\[1\].pdf](http://www.virtual.unal.edu.co/cursos/ciencias/dis_exp/und_3/pdf/validaciondesupuestosunidad3b[1].pdf) [Último acceso: 12 diciembre 2013].
- [17] D. F. Cardona, J. L. González, M. Rivera y E. H. Cárdenas, *Módulo de regresión lineal*, Bogotá: Universidad del Rosario, 2013.