

REGRESION LOGISTICA CON VARIABLE DICOTOMICA: UNA APLICACIÓN AL  
ESTUDIO DE CREDITO DE VEHICULO

JUAN FELIPE GUERRA

DIEGO ALEJANDRO BEDOYA SUAZA

Proyecto presentado para optar al título de ingeniería industrial

Asesor

Leonidas Millán, Ingeniero industrial, Msc.

UNIVERSIDAD DE SAN BUENAVENTURA SECCIONAL MEDELLÍN

FACULTAD DE INGENIERÍAS

INGENERÍA INDUSTRIAL

MEDELLIN

2014

## CONTENIDO

<b>1. JUSTIFICACIÓN .....</b>	<b>3</b>
<b>2. PLANTEAMIENTO DEL PROBLEMA .....</b>	<b>4</b>
<b>3. OBJETIVO GENERAL .....</b>	<b>5</b>
<b>4. OBJETIVOS ESPECÍFICOS .....</b>	<b>5</b>
<b>5. MARCO REFERENCIAL .....</b>	<b>6</b>
<b>6. DISEÑO METODOLÓGICO .....</b>	<b>10</b>
<b>6.1 Área de estudio.....</b>	<b>10</b>
<b>6.2 Materiales .....</b>	<b>10</b>
<b>6.2 Métodos .....</b>	<b>13</b>
<b>6.2.1 Obtención del crédito. ....</b>	<b>13</b>
<b>6.2.1 Determinación del tamaño muestral.....</b>	<b>14</b>
<b>6.2.2 Modelo estadístico.....</b>	<b>16</b>
<b>7. RESULTADOS Y ANALISIS .....</b>	<b>17</b>
<b>8. CONCLUSIONES .....</b>	<b>22</b>
<b>9. REFERENCIAS BIBLIOGRAFICAS .....</b>	<b>23</b>
<b>LISTA DE TABLAS.....</b>	<b>27</b>
<b>LISTA DE FIGURAS.....</b>	<b>29</b>

## 1. JUSTIFICACIÓN

Debido a los inconvenientes que presentan algunas entidades financieras con respecto a la capacidad y velocidad de respuesta a la atención del usuario para la solicitud de créditos, se encontró la necesidad de implementar nuevas estrategias para dar una respuesta más rápida y eficaz. Por tal motivo se ha recurrido a la regresión logística con el fin de proporcionar una posible solución a la problemática planteada.

Para realizar esta investigación es necesario determinar un modelo funcional que arroje la probabilidad según la información suministrada; realizando un gran número de pruebas para determinar si el modelo es viable, si es así, buscar automatizar el modelo con el fin de hacerlo más ágil. Otros alcances importantes que se obtendrá de la implementación de este sistema, es la reducción de costos en cuanto a mano de obra, calidad de servicio y fidelización de los clientes por una pronta respuesta.

## 2. PLANTEAMIENTO DEL PROBLEMA

En la industria bancaria se han implementado diversas técnicas de calificación de un crédito, con la finalidad de generar tablas de puntuación para dicho crédito, una de las técnicas más usadas ha sido el modelo de regresión logística ya que este otorga una cantidad de características (solidez y transparencia) que lo hacen imprescindible a la hora de la implementación.

Aunque algunas técnicas nuevas (por ejemplo, máquinas de vectores) se han aplicado a la puntuación de crédito y han mostrado mayor precisión de la predicción, tienen problemas con la interpretabilidad de los resultados. Por lo tanto, estas técnicas avanzadas no se han aplicado ampliamente en la práctica. Para mejorar la precisión de la predicción de la regresión logística, se propone regresión logística con coeficientes aleatorios. El modelo propuesto puede mejorar la precisión de la predicción de regresión logística sin sacrificar las características deseables. Se espera que el índice de calificación de crédito propuesto pueda contribuir en la práctica de gestión eficaz del riesgo de crédito. (Tarjeta de puntuación de créditos basados en la regresión logística con coeficientes al azar), (Dong, Lai y Yen, 2012)

### **3. OBJETIVO GENERAL**

Establecer un método aplicando regresión logística para encontrar una probabilidad según la información que se tenga de un cliente.

### **4. OBJETIVOS ESPECÍFICOS**

- Encontrar las variables adecuadas para hacer funcional el método de regresión
- Determinar un modelo funcional que arroje la probabilidad según la información suministrada.

## 5. MARCO REFERENCIAL

Mittlböck & Heinz (2001) Realizaron un estudio medidas  $r^2$  para Poisson y regresión logística cuando ambos modelos son aplicables; concluyen que la medida de la  $r^2$  de la regresión de Poisson cuantifica la previsibilidad de los tipos de eventos, pero no es adecuado para cuantificar la predictibilidad de los resultados de las observaciones individuales. Además, realizaron un estudio basado en una escala modificada para estimar la función de regresión logística multinomial en muestras pequeñas. El método se basa en una modificación adecuada de la función de puntuación que elimina el sesgo de primer orden. Los investigadores (Janssen, y otros, 2012) enfocaron sus estudios en el desarrollo y validación de modelos de predicción clínica: diferencias marginales entre regresión logística, estimación de máxima verosimilitud penalizada, y la programación genética. Ahora (Dong, Tung, Chen, Liao, & Wen, 2011), implementaron Modelo de regresión logística para predecir la probabilidad de fallo de un dique de deslizamiento de tierra, sobre la base de un conjunto de datos japonesa, en este estudio se utilizó el método de regresión logística y la técnica jackknife para identificar las variables geomorfológicas importantes, incluyendo el flujo pico (o área de captación), altura de la presa, la anchura y la longitud de la secuencia que afecta a la estabilidad de las presas de deslizamientos. El resultado de alta potencia predicción general, pone de manifiesto la robustez de los modelos de regresión logística propuestos. En consecuencia, la probabilidad de fallo de un dique de deslizamiento también se puede evaluar sobre la base de este enfoque. Por consiguiente, se sugiere que el modelo de predicción utilizando el flujo máximo como factor causal debe ser utilizado para evaluar la estabilidad de un dique de deslizamiento si el flujo máximo disponible junto con una estimación del impacto de una inundación, la probabilidad de fallo de la presa de deslizamiento de tierra predicha por el modelo de regresión logística propuesto podría ser útil para evaluar el riesgo asociado. Por su parte (Heijden, 2012), ejecuto un estudio sobre el apoyo a las decisiones para la selección de modelos de regresión logística, dado un conjunto de variables métricas y una variable dependiente binaria, selecciona el

subconjunto óptimo de variables que mejor pueden predecir esta variable dependiente, en este trabajo se documentó la heurística, se formalizó los algoritmos, y, finalmente, se presentó un sistema de soporte de decisiones interactiva para facilitar la selección de este tipo de modelo. Los estadistas (Pardo, Pardo, & Pardo, 2006), consideran que la inferencia basada en medidas de divergencia son muy generales bajo los supuestos de un modelo de regresión logística. Otra investigación (Heredia, Hernández, & Vilalta, 2012), aplicaron la regresión logística para predecir el Rendimiento Académico, en particular con temas difíciles que requieren la intervención oportuna para reducir la tasa de fracaso escolar, las estimaciones de probabilidad que se obtuvieron con este modelo se utilizaron como punto de partida para desarrollar un método que permite la predicción del estado del MPP para aquellos estudiantes que tomaron este tema en el curso académico 2010-2011. (Hobza, Pardo, & Vajda, 2008), presentan un estimador de la mediana de los parámetros de regresión logística. Se define como la clásica  $L_1$  –estimador, se aplica a los datos continuos  $Z_1, \dots, Z_n$  obtenidos por un suavizado estadístico de las observaciones binarias originales de regresión logística  $Y_1, \dots, Y_n$ . La consistencia y normalidad asintótica de este estimador se han demostrado. Ahora (Mazzocco & Hussain, 2012), implementaron la regresión logística para facilitar el diagnóstico de la demencia. El objetivo de este estudio es el de mejorar en el rendimiento de una aplicación reciente de redes de creencias bayesianas utilizando un enfoque alternativo basado en la regresión logística.

Por su parte (Steyerberg, Eijkemans, & Habbema, 1999), realizaron un estudio de simulación de sesgo en el análisis de regresión logística, las muestras fueron extraídas al azar que incluyeron 3, 5, 10, 20 o 40 eventos por variable (EPV). Selección por pasos hacia atrás se aplicó en los modelos con 8 ó 16 predictores de mortalidad a 30 días. Se encontró una sobreestimación considerable de los coeficientes de regresión de las covariables seleccionadas. Mientras que (Xie, Pendergast, & ClarkeWilliam, 2008), estudian la bondad del ajuste de prueba para los modelos de regresión logística con predictores continuos. Se propone un nuevo método para la bondad del ajuste de prueba que utiliza una estrategia de división muy general (clustering) en el espacio

covariable y, o bien una estadística de Pearson, o una puntuación estadística. En otro momento (Kim, 2009), plantea un modelo de regresión logística combinada con árbol de decisiones para hacer frente a una interacción significativa entre las variables explicativas. Los datos nacionales de pensiones se aplican a este modelo, y como una aplicación del modelo propuesto, las estrategias para la reducción de los riesgos financieros en la gestión y planificación de financiación de las pensiones están ilustrados. No obstante (Brazer, y otros, 1991), encuentran la utilidad de la regresión logística ordinal en la predicción de la neoplasia colonrrectal, se demostró en un grupo de 461 pacientes consecutivos sometidos a colonoscopia en una práctica comunitaria. El riesgo de neoplasia colonrrectal podría predecirse utilizando la edad del paciente, sexo, hematocrito, número de ensayo de sangre fecal oculta y la indicación de la colonoscopia. Mientras tanto (Wattimena, Kramadibrata, & Azizi, 2013) utilizaron la regresión logística para calcular la probabilidad de que un pilar particular de una geometría determinada (relación de anchura a altura) y una condición de estrés conocido (fuerza al cociente de tensión) fuera estable. Hout, Heijden, & Gilchrist, (2007), indican que el modelo de regresión logística univariante y multivariante se discute que las variables de respuesta están sujetos a la respuesta al azar (RR). RR es una técnica de entrevista que se puede utilizar cuando las preguntas sensibles tienen que ser hechas y los encuestados son reacios a contestar directamente. Las variables RR pueden ser descritos como mal clasificados variables categóricas donde las probabilidades condicionales de errores de clasificación son conocidos. Sen Roy & Guria (2008), estudiaron el diagnóstico de un modelo de regresión logística utilizando la técnica de supresión de observación. El modelo se ajustó utilizando el método de máxima verosimilitud y los cambios en las estimaciones y la desviación se observan después de eliminar una observación. Las expresiones se deriva de modo que no es necesario volver a ejecutar la regresión después de cada eliminación, así de manera considerable ahorro de tiempo de cálculo. (Sebastiani & Settimi, 1997), demuestran que un diseño de 2 puntos sugeridos por Ford et al. (J.Roy. estadistas. Soc. Ser B54 (1992), (569-583) para un modelo de regresión logística simple. También se proponen

diseños que dan alternativas eficaces a los diseños D-óptimos y requieren menos información sobre los parámetros. (Le Calvé & Saboya.Jacques, 2000), el estudio se ocupa de un aspecto particular de un proceso de fusión, por lo que sólo el rango de los registros recuperados está disponible. Sobre la base de esta información bastante limitada, en este documento se describe el fundamento teórico y rendimiento de la recuperación del enfoque de base de datos basado en la fusión de regresión logística. Otro estudio realizado por (Hirsch, 1987) propone un programa para la Texas Instruments TI-59, un soporte de impresión descrito que permite el cálculo de estimaciones puntuales e intervalos de confianza para las razones de posibilidades utilizando información incluida en la salida de la mayoría de los programas de regresión logística. El programa contempla un máximo de quince variables independientes continuas y / o dicotómica. Los investigadores (Broll, Glaser, & Kreienbrock, 2002), encontraron que el cálculo del tamaño de un estudio de muestra requerido es uno de los aspectos más importantes de la validez de un estudio epidemiológico. La regresión logística a menudo se utiliza en el modelado en epidemiología. Un método simplificado para calcular el tamaño de muestra para la regresión logística múltiple modelo fue propuesto por Hsieh et al. [Stat. Med. 17 (1998) 1623]. Los investigadores (Qian & Field, 2002), establecen una ley del logaritmo iterado para el estimador de máxima verosimilitud de los parámetros en un modelo de regresión logística con enlace canónico. Este resultado establece la fuerte consistencia de algunos criterios pertinentes de selección del modelo.

## 6. DISEÑO METODOLÓGICO

### 6.1 Área de estudio

El área de estudio abarca principalmente los clientes en potencia pertenecientes a los estratos 3 y 4 de la ciudad de Medellín que van a solicitar crédito para vehículo usado en la entidad Sufí Bancolombia; Sufí es una línea de negocio y marca del Grupo Bancolombia que se especializa en promover el crédito de consumo a través de canales indirectos, la cual tiene como objetivo identificar las necesidades de los clientes para llegarles a través de los canales, ofreciendo financiación como producto complementario. Dentro de los productos que maneja Sufí como Marca Bancolombia se encuentra la financiación de vehículos, financiación de motos, créditos de consumo, etc (Bancolombia S.A., 2013).

### 6.2 Materiales

Para poder determinar cuál es la variable que más influye en la obtención de un crédito de vehículo, se deben tener claro las variables que el banco analiza para la obtención del crédito de vehículo usado, dichas variables se mencionan en las siguientes tablas.

**Tabla 1.** Variables Crédito Vehículo (Línea de financiación e información personal).

<b>LINEA DE FINANCIACION</b>	
Valor solicitado	País
Plazo	Regional
Fecha de diligenciamiento	Oficina
<b>DATOS DEL VEHICULO A FINANCIAR</b>	
Marca	
Modelo	
Valor comercial	
Concesionario	
Vendedor	
Cedula vendedor	
<b>INFORMACION PERSONAL</b>	
Nombres	Nombres del conyugue
Nº Documento	Nº de documento
Sexo	Teléfono
Fecha de nacimiento	Lugar de trabajo
Lugar de nacimiento	
Estado civil	
Estudios Realizados	
Tipo de vivienda	
Estrato	
Departamento	
Ciudad	
Barrio	
Dirección	
Teléfonos	
Email	

**Tabla 2.** Datos laborales, Información Financiera, respaldo patrimonial y referencias.

<b>DATOS LABORALES</b>	
<b>EMPLEADO</b>	<b>INDEPENDIENTE</b>
Ocupación	Actividad Económica
Cargo	Tiempo en la Actividad
Tipo de contrato	
Nombre de la empresa	
Nit de la empresa	
tiempo en la empresa	
Dirección	
Ciudad	
Teléfono	
<b>INFORMACION FINANCIERA</b>	
<b>INGRESOS</b>	<b>EGRESOS</b>
Salario	Préstamo Vivienda
Honorarios	Préstamo Vehículo
Otros	Arriendo
Ingresos conyugue	Gastos familiares
	Tarjeta de crédito
	Otros prestamos
	Egresos conyugue
<b>RESPALDO PATRIMONIAL</b>	
<b>VEHICULOS</b>	<b>BIENES RAICES</b>
Número de vehículos	Numero de bienes
Marca y modelo	Tipo
Valor Comercial	Dirección
Total Activos	Ciudad
	Valor comercial
	Hipoteca
<b>REFERENCIAS</b>	
Comerciales	
Familiares	
Personales	

## **6.2 Métodos**

### **6.2.1 Obtención del crédito.**

Sufí como entidad financiera que otorga créditos para la adquisición de vehículos a personas de diferentes niveles socioeconómicos, siempre y cuando cumplan con una serie de requisitos establecidos por la entidad los cuales son informados a las personas que aspiran obtener dicho crédito, luego de que el solicitante envíe al asesor de créditos la respectiva documentación, este la remite a un comité encargado del análisis de las peticiones donde se realiza el estudio de la solicitud y verificación de que se cumplan con los requisitos exigidos inicialmente. A los 5 días hábiles siguientes a la fecha de solicitud, se da respuesta al cliente si fue aprobado o no el crédito, en caso de ser positivo se pasa a hacer todos los trámites, establecer modalidades de pago y tiempo a financiar, posteriormente la adquisición del vehículo. En la Figura 1, se encuentra el flujo-grama de cómo es la obtención del crédito.

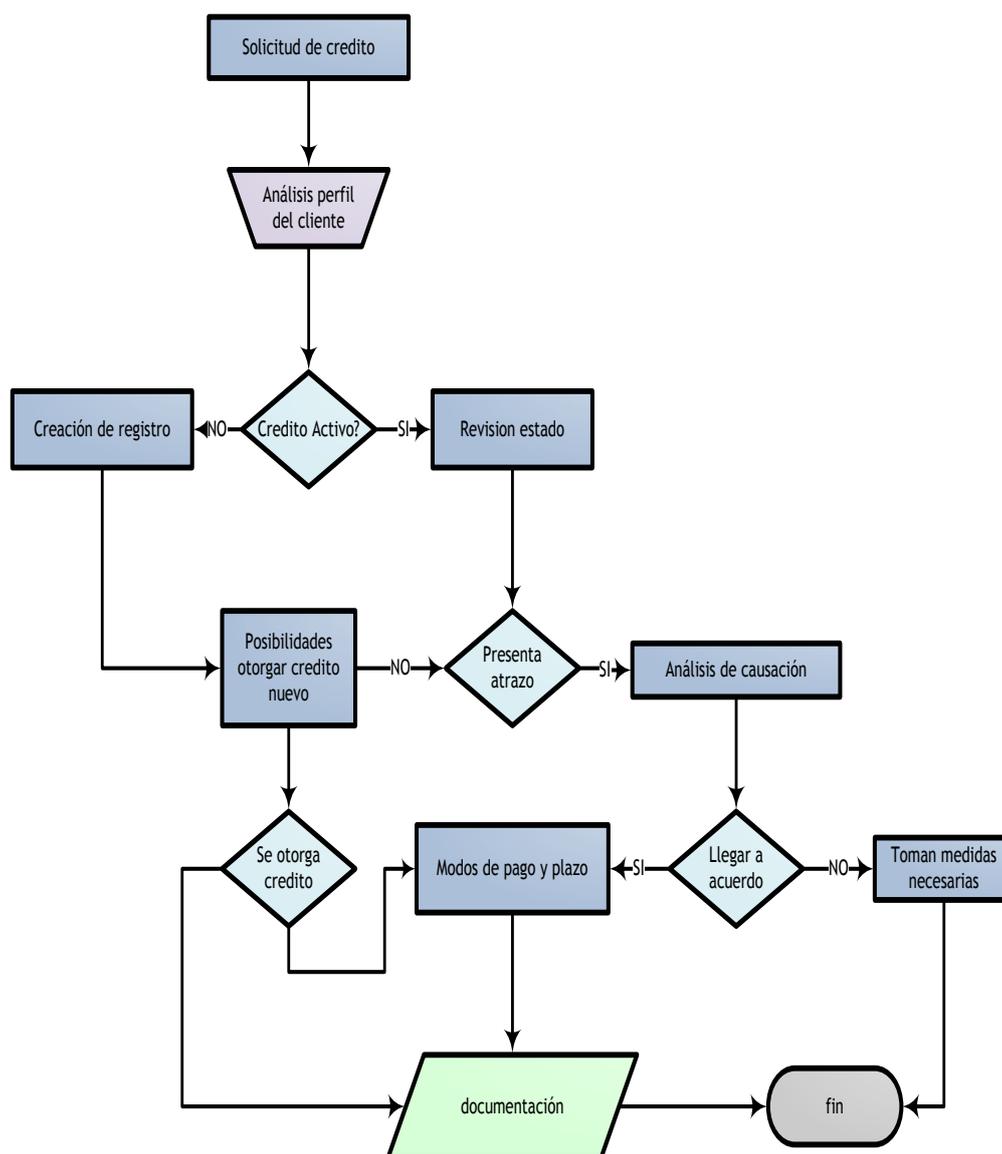


Figura 1. Flujo-grama de cómo obtener el crédito.

### 6.2.1 Determinación del tamaño muestral.

Mediante un estudio realizado por el Departamento Administrativo de Planeación (2011), el número de personas ubicadas en el estrato 3 fue de 689.029 y para el estrato 4 el análisis arrojó que había un número de 235.821; partiendo de estos resultados se buscaba establecer el tamaño de la muestra

para ser aplicada en la regresión logística, mediante la fórmula estadística (Feedback Networks Technologies, 2013):

$$n = \frac{K^2 * p * q * N}{(e^2 * (N - 1)) + K^2 * p * q} \quad [1]$$

Donde:

N: Es el tamaño de la población o universo (número total de posibles encuestados).

K: Es una constante que depende del nivel de confianza que se asigne. El nivel de confianza indica la probabilidad de que los resultados de la investigación sean ciertos: un 95,5% de confianza es lo mismo que decir que se puede equivocar con una probabilidad del 4,5%. Los valores k más utilizados y sus niveles de confianza se observan en la tabla

Tabla 3. Valores de k.

K	1.15	1.28	1.44	1.65	1.96	2	2.58
Nivel de confianza	75%	80%	85%	90%	95%	95.5%	99%

e: es el error muestral deseado. El error muestral es la diferencia que puede haber entre el resultado que se obtiene preguntando a una muestra de la población y el que se obtendría si preguntáramos al total de ella.

p: es la proporción de individuos que poseen en la población la característica de estudio. Este dato es generalmente desconocido y se suele suponer que  $p=q=0.5$  que es la opción más segura.

q: es la proporción de individuos que no poseen esa característica, es decir, es  $1-p$ .

n: es el tamaño de la muestra (número de encuestas que se van a realizar).

Para calcular el tamaño muestral, se empleo la formula número 1, la cual tiene como resultado:

$$\frac{1.96^2 * 0.5 * 0.5 * 924850}{(0.05^2 * (924850 - 1)) + 1.96^2 * 0.5 * 0.5} = 384$$

Para el análisis de los datos se tomara una muestra de 384 individuos de la base datos, para así determinar cuáles variables de las anteriormente descritas se relacionan en mayor medida con la probabilidad de que le acepten un préstamo de automóvil.

### **6.2.2 Modelo estadístico.**

Para el análisis de la información se empleara el software SPSS licencia amparada por la Universidad San Buenaventura Sede Medellín. Este software fue creado como el acrónimo de *Statistical Package for the Social Sciences* aunque también se ha referido como "Statistical Product and Service Solutions" (Pardo Merino & Ruíz Díaz, 2002).

Se escogió este programa para el análisis de la información, en primer lugar porque la Universidad tiene licencia de uso y en segundo lugar porque es uno de los programas estadísticos más conocidos teniendo en cuenta su capacidad para trabajar con grandes bases de datos y su sencilla interfaz para la mayoría de los análisis, además de ofrecer una completa gama de funciones (SPSS Regression Models, 2007).

El modelo que se empleará para el análisis de los datos, será un modelo de regresión logística, donde la variable respuesta es dicotómica (aceptar o no aceptar el crédito) versus las variables (Tabla 2) que están en el formato que llena el solicitante para presentar el crédito.

## 7. RESULTADOS Y ANALISIS

Demostración, El modelo de regresión logística parte de la hipótesis de que los datos siguen el siguiente modelo:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_k * x_k + u = x * b + u$$

Con el fin de simplificar la notación, definimos Z:

$$z = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_k * x_k$$

Por lo tanto, el modelo se puede presentar como:

$$\ln\left(\frac{p}{1-p}\right) = z + u$$

Donde p es la probabilidad de que ocurra el suceso de estudio operando algebraicamente sobre el modelo:

$$\ln\left(\frac{p}{1-p}\right) = z$$

$$\frac{p}{1-p} = e^z$$

$$p = (1-p) * e^z$$

$$p = e^z - p * e^z$$

$$p + p * e^z = e^z$$

$$p(1 + e^z) = e^z$$

$$p = \frac{e^z}{1 + e^z}$$

Como la función de distribución logística es:

$$F(x) = \frac{e^x}{1 + e^x}$$

Por lo tanto, podemos reescribir el modelo de forma mucho más compacta:

$$p = \frac{e^z}{1 + e^z} = F(z) = F(x * b)$$

De donde se deduce que el modelo de regresión logística es, en principio, un modelo de regresión no lineal, pero es lineal en escala logarítmica atendiendo su definición original:

$$\ln\left(\frac{p}{1-p}\right) = z$$

$$\ln(p) - \ln(1-p) = z$$

$$\ln(p) - \ln(1-p) = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_k * x_k$$

Es decir la diferencia de la probabilidad de que ocurra un suceso respecto de que no ocurra es lineal pero en escala logarítmica. Por tanto, el significado de los coeficientes, aunque guardando una cierta relación con un modelo de regresión lineal, va a ser algo más complejo de interpretar.

Ahora recordemos las dos formas más importantes de expresar el modelo de regresión logística.

$$\ln(p) - \ln(1-p) = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_k * x_k$$

$$\frac{p}{1-p} = e^{b_0} * e^{b_1 * x_1} * e^{b_2 * x_2} \dots e^{b_k * x_k}$$

La primera expresión se llama logit y a la segunda odds ratio o cociente de probabilidades.

El resultado muestra los resultados de ajustar un modelo de regresión logística para describir la relación entre RESPUESTA total y 11 variables independientes. La ecuación del modelo ajustado es

$$\text{RESPUESTA TOTAL} = \exp(\eta) / (1 + \exp(\eta))$$

Donde

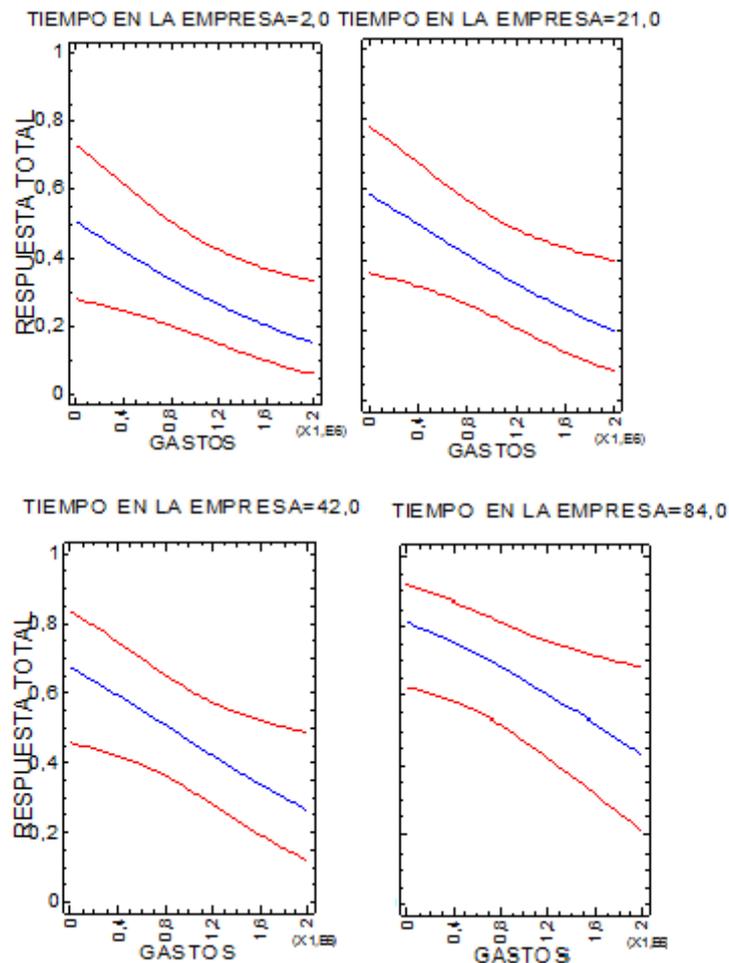
$$\begin{aligned} \eta = & -0,766716 - 0,00972632 * \text{NUMEROS DE BIENES RAICES} + 0,0535793 * \\ & \text{NUMERO DE VEHICULOS PROPIOS} + 0,0176234 * \text{TIEMPO EN LA} \\ & \text{EMPRESA} - 1,86765 \text{ E-8} * \text{VALOR SOLICITADO} + 1,07855 \text{ E-8} * \text{VALOR} \\ & \text{COMERCIAL DEL VEHICULO} + 3,80081 \text{ E-7} * \text{Honorarios} + 2,37068 \text{ E-7} * \\ & \text{OTROS INGRESOS} + 1,64362 \text{ E-8} * \text{Salarios} - 8,72851 \text{ E-7} * \text{GASTOS} + \\ & 0,108612 * \text{ESTRATO} = 3 + 0,0854257 * \text{MODELO} = 2,007 + 0,26904 * \\ & \text{MODELO} = 2,008 - 0,233502 * \text{MODELO} = 2,009 - 0,377319 * \text{MODELO} = \\ & 2,010 - 0,234226 * \text{MODELO 2011} = - = 0,00178792 * \text{MODELO 2012} \end{aligned}$$

Para el análisis de los datos es totalmente dispensable contar con los datos propuestos en las tablas 1 y 2, la cual hace referencia a las posesiones económicas, y seguridad laboral y socioeconómica que dan pie a que el crédito tenga mayor o menor probabilidad de que un crédito sea aprobado o no, en cuanto al análisis por medio de regresión logística. En la tabla 4, se puede percibir los estados de significancia de cada una de las variables del estudio tanto cualitativas como cuantitativas, evaluadas con un porcentaje de confianza del 95% en la cual se determina que los valores que están por debajo de un valor p de 0.05 son los valores que tienen más relevancia en cuanto al proceso de asignar el crédito, en este caso los gastos con un valor p de 0.0156 y el tiempo en la empresa con un valor p de 0.0001, son las dos variables que presentan diferencias significativas, las otras variables por presentar un valor p mayor a 0,05 (no son significativas) al momento de aprobación de un crédito.

Tabla 4. Pruebas de Razón de Verosimilitud

Factor	Chi-cuadrado	g.l	Valor-p
Números de bienes raíces	0,014589	1	0,9039
Número de vehículos propios	0,3135	1	0,5755
Tiempo en la empresa	15,1579	1	0,0001
Valor solicitado	2,65486	1	0,1032
Valor comercial del vehículo	1,52749	1	0,2165
Honorarios	3,06876	1	0,0798
Otros ingresos	3,33826	1	0,0677
Salarios	0,0396557	1	0,8422
Gastos	5,85022	1	0,0156
Estrato	0,246724	1	0,6194
Modelo	3,61161	6	0,7291

En la Figura 2, se observa el modelo de regresión logística ajustado. Esta figura muestra la proporción de gastos a razón del tiempo para un porcentaje de aprobación, en el cual a menor cantidad de gastos y mayor cantidad de tiempo en la empresa será más elevado la probabilidad de aceptación, es notable los porcentajes de variación según los meses propuestos en los que un empleado ha estado en servicio.



## Figura 2. Modelo de regresión logística ajustado

Una persona que lleva empleado 2 meses en una la empresa, según las condiciones básicas, da como resultado que tiene un porcentaje de probabilidad de aceptación del crédito de 50.9%; para una persona que lleva empleado 21 meses, la probabilidad de que le acepten el crédito es de 58.9%; en el caso de que una persona lleve vinculada con la empresa 42 meses, la probabilidad que le aprueben el crédito es del 66.8%. Y para un cliente que lleve 81 meses trabajando en su empresa, la aprobación del crédito tiene un porcentaje de 81.8%. Ver Figura 2.

Con los resultados propuestos se puede tener mayor certeza para aprobar un crédito; teniendo los datos básicos se puede pedir y otorgar de manera fácil, en el momento de tener los datos, se podrá decir en la asesoría al cliente, su crédito tiene la probabilidad X de ser aprobado, por tanto, al cliente se le dará más confianza para que busque la papelería necesaria para hacer el estudio final del crédito, en el departamento de análisis teniendo ya la información veraz y precisa del cliente, aplicando el mismo sistema por parte del analista, hará que la aprobación sea mas ágil ya que al revisar la información del cliente no tendrá que ser tan exhaustiva y así poderle dar una respuesta más pronta teniendo como repercusión una fidelización por el buen y eficiente servicio prestado por la compañía, adicional que mejorarían los tiempos por tanto se reducirían los costos ya que en si necesitarían menos personal para el análisis final del crédito a estudiar.

## 8. CONCLUSIONES

Las variables que influyen significativamente en la aprobación o no de un crédito son los gastos de las personas y el tiempo que lleva laborando en la empresa.

A medida que una persona aumenta los gastos la probabilidad que le apruebe un crédito disminuye.

Entre mayor sea el tiempo que una persona lleve laborando en una empresa, la probabilidad que le aprueben un crédito es mayor.

El conocimiento de los principales factores que inciden en la aprobación y rechazo de un crédito para vehículo es necesario ya que permite a las entidades financieras optimizar el tiempo de respuesta a sus clientes, y a su vez generar resultados que determinen el perfil de la persona que aspira al crédito y así otorgar una mayor confiabilidad y garantía paraa la entidad.

## 9. REFERENCIAS BIBLIOGRAFICAS

- Brazer, S., Pancotto, F., Long, T., Harrell, F., Lee, k., Tyor, M., & Pryor, D. (1991). El uso de regresión logística ordinal para estimar la probabilidad de neoplasia colorrectal. *Revista de Epidemiología Clínica*, 44(11), 1263-1270.
- Broll, S., Glaser, S., & Kreienbrock, L. (2002). Cálculo de límites de tamaño de muestra para la regresión logística. *Medicina Veterinaria Preventiva*, 54(2), 105-111.
- Dong, J., Tung, Y.-H., Chen, C.-C., Liao, J.-J., & Wen, Y. (2011). Modelo de regresión logística para predecir la probabilidad de fallo de un dique de deslizamiento de tierra. *Ingeniería Geológica*, 117(1-2), 52-61.
- Donga, G., Keung Lai, K., & Yen, J. (2012). Credit scorecard based on logistic regression with random coefficients. *Procedia Computer Science*, 1(1), 2463-2468.
- Heijden, H. (2012). Apoyo a las decisiones para la selección de modelos de regresión logística óptima. *Sistemas Expertos con aplicaciones*, 39(10), 8573-8583.
- Heredia Rico, J., Hernández Rodríguez, A., & Vilalta Alonso, J. (2012). Empleo de la regresión logística ordinal Para La predicción del Rendimiento Académico. *Investigacion Operacional*, 33(3), 252-267.
- Hirsch, R. P. (1987). El cálculo de odds ratios en la TI-59 de salida de regresión logística. *Los equipos de la Biología y la Medicina*, 17(2), 101-108.
- hobza, T., Pardo, L., & Vajda, I. (2008). Estimador robusto mediano de regresión logística. *Diario de Planificación Estadística e inferencia*, 138(12), 3822-3840.

- Hout, A., Heijden, G. P., & Gilchrist, R. (2007). El modelo de regresión logística con las variables de respuesta sujetas a la respuesta al azar. *Estadística Computacional y Análisis de Datos*, 51(12), 6060-6069.
- Janssen, K., Siccama, I., Vergouwe, Y., Koffijberg, H., Debray, Keijzer, M., Lunas, K. (2012). Desarrollo y validación de modelos de predicción clínica: diferencias marginales entre la regresión logística, la estimación de máxima verosimilitud penalizada, y la programación genética. *Revista de Epidemiología Clínica*, 65(4), 404-412.
- Kim, M. (2009). Modelo de dos etapas de regresión logística. *Sistemas Expertos con aplicaciones*, 36(3), 6727-6734.
- Le Calvé, A., & Saboya.Jacques. (2000). Base de datos fusionando estrategia basada en la regresión logística. *Procesamiento de la Información y Gestión*, 36(3), 341-359.
- Mazzocco, T., & Hussain, A. (2012). Nuevos modelos de regresión logística para facilitar el diagnóstico de la demencia. *Sistemas Expertos con aplicaciones*, 39(3), 3356-3361.
- Mittlböck, M., & Heinz, H. (2001). A note on R2 measures for Poisson and logistic regression models when both models are applicable. *Journal of Clinical Epidemiology*, 54(1), 99-103.
- Pardo, J. A., Pardo, L., & Pardo, M. d. (2006). Prueba de los modelos de regresión logística basado en  $\phi$  -divergencias medidas. *Diario de Planificación Estadística e inferencia*, 136(3), 982-1006.
- Qian, G., & Field, C. (2002). Ley de criterio reiterado modelo logaritmo y consistente en la selección de regresión logística. *Estadísticas y Probabilidad Cartas*, 56(1), 101-112.

- Sebastiani, P., & Settimi, R. (1997). Una nota sobre D-óptimos diseños para un modelo de regresión logística. *Diario de Planificación Estadística e inferencia*, 59(2), 359-368.
- Sen Roy, S., & Guria, S. (2008). Diagnóstico en modelos de regresión logística. *Revista de la Sociedad de Estadística de Corea*, 37(2), 89-94.
- Steyerberg, E., Eijkemans, M., & Habbema, D. (1999). Selección por pasos en pequeños conjuntos de datos : un estudio de simulación de sesgo en el análisis de regresión logística. *Revista de Epidemiología Clínica*, 52(10), 935-942.
- Wattimena, R., Kramadibrata, S., & Azizi, M. (2013). Desarrollo de carbón pilar estabilidad tabla mediante regresión logística. *Revista Internacional de Mecánica de Rocas y Ciencias de la Minería*, 52(1), 55-60.
- Xie, X.-J., Pendergast, J., & ClarkeWilliam. (2008). El aumento del poder: Un enfoque práctico para la bondad del ajuste de prueba para los modelos de regresión logística con predictores continuos. *Estadística Computacional y Análisis de Datos*, 52(5), 2703-2713.
- Bancolombia S.A. (2013). *Sufi: Una Marca Bancolombia*. Recuperado el 10 de 03 de 2013, de [http://www.sufi.com.co/cs/Satellite?c=Page&cid=1266348980798&pagenam e=Sufi%2FS\\_TemplateInterna](http://www.sufi.com.co/cs/Satellite?c=Page&cid=1266348980798&pagenam e=Sufi%2FS_TemplateInterna)
- Departamento Administrativo de Planeacion. (2011). Perfil socioeconómico estrato 4. Medio bajo. *Encuesta de Calidad de Vida* . Medellín.
- Feedback Networks Technologies. (2013). *Experiencia: Calcular la muestra correcta*. Recuperado el 19 de Marzo de 2013, de <http://www.feedbacknetworks.com/cas/experiencia/sol-preguntar-calcul ar.html>

- Glz Solis, O. (2013). Product Information: Departamento de Crédito.
- Pardo Merino, A., & Ruíz Díaz, M. Á. (2002). *Guía para el análisis de datos*. Madrid: McGraw-Hill.
- J.M Rojo Abuin. (2007). *Regresion con variable dependiente cualitativa, 6-8*.

## LISTA DE TABLAS

Tabla 1. Variables Crédito Vehículo.	11
Tabla 2. Datos laborales, Información Financiera, respaldo patrimonial y referencias.	12
Tabla 3. Valores de K	15
Tabla 4. Pruebas de Razón de Verosimilitud	20

## LISTA DE FIGURAS

Figura 1. Flujo-grama de cómo obtener el crédito.	14
Figura 2. Modelo de regresión logística ajustado	20