



ESCUELA SUPERIOR POLITÈCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

Ingeniería en Estadística e Informática

**“DISEÑO DE UN MODELO DE PUNTUACIÓN
CREDITICIA PARA EMPRESAS COMERCIALES”**

PROYECTO INTEGRADOR

Previo a la obtención del Título de:

INGENIERO EN ESTADISTICA E INFORMÀTICA

Presentado por:

Danny Leonel Hernández Cano

Guayaquil – Ecuador

2015

AGRADECIMIENTO

A dios, por sus bendiciones espirituales;

A mis padres, por sus consejos, enseñanzas e impulso para terminar este objetivo. A mi esposa por el cariño brindado en el trayecto de mi vida. A mi gran amigo Christian Galarza por su apoyo con su experiencia en la profesión;

A nuestra profesora, amiga Ing. Paola Reyes por la paciencia, dedicación y apoyo brindado en la culminación de este proyecto.

DEDICATORIA

Por tu ejemplo, por tu apoyo, por tu dedicación, por tu amor y por todo lo que eres en mi vida este trabajo va dedicado a mi Padre.

A las mujeres de mi vida; mi madre, mi esposa y mis hijas, que sea un ejemplo para mis pequeñas y que de seguro me superen profesionalmente.

DECLARACIÓN EXPRESA

"La responsabilidad del contenido de este Trabajo final de graduación, nos corresponde exclusivamente; y el patrimonio intelectual de la misma a la Escuela Superior Politécnica del Litoral".

(Reglamento de Graduación de la ESPOL)

Danny Leonel Hernández Cano

RESUMEN

Este presente trabajo aborda el tema del riesgo de otorgar un crédito directo a clientes nuevos en las empresas comerciales, basado en los datos obtenidos en una empresa de RETAIL en la ciudad de Guayaquil, el trabajo es realizado con la finalidad de determinar un modelo estadístico que prediga si un cliente es moroso o no, además identificar los perfiles de los clientes considerados como “Bueno” y “Malos” en términos crediticios.

En el primer capítulo denominado “Descripción del Problema”, se redacta el problema actual de la empresa, su respectiva justificación que mediante un modelo estadístico mitigue el problema planteado por empresa. Luego se presenta los objetivos por la cual se está realizando dicho proyecto.

En el segundo capítulo denominado “Metodología”, se describe los conceptos relaciones con la cartera de crédito de la empresa, además se describe algunos avances relacionados al tema del proyecto. Luego se redacta los métodos que utilizar para realizar el trabajo.

En el tercer capítulo denominado “Análisis y Modelos”, se estudia los análisis descriptivos e inferencial de las variables más relevante de acuerdo al objetivo como también el modelo estadístico, regresión logística.

Como última sección se tiene las conclusiones y recomendaciones más relevantes.

Contenido

CAPÍTULO I.....	1
1. Descripción del Problema	1
1.1 Introducción	1
1.2 Justificación del Problema.....	2
1.3 Objetivos	3
CAPÍTULO 2.....	4
2. Metodología	4
2.1 Marco Referencial.....	4
2.2 Marco Conceptual.....	4
2.3 Marco Teórico	8
2.4 Población	13
2.5 Análisis Estadístico Univariado	13
2.6. Análisis Estadístico Inferencial.....	14
2.7. Regresión Logística	15
2.7.1 Estimación de parámetros en un modelo de Regresión Logística	18
2.7.2 Evaluación de los Modelos de la Regresión Logística	19
2.7.2.1 El estadístico de Wald.	19
2.7.2.2 El estadístico G de la razón de verosimilitud.....	20
2.8 Recolección de Información	21
CAPÍTULO 3.....	22
3. Análisis y Modelos.....	22
3.1. Análisis Estadístico Univariado	22
3.1.1 Distribución Por Edad.....	23
3.1.2 Distribución Por Género	24
3.1.3 Distribución Por Región Domicilio.....	25
3.1.4 Distribución Por Cargas Familiares.....	26
3.1.5 Distribución Por Clasificación del Buro.....	27
3.1.6 Distribución Por Cupo Otorgado.....	28

3.1.7 Distribución Por Cliente Malo vs Cliente Bueno	29
3.2. Análisis Estadístico Inferencial	30
3.2.1 Tabla de Contingencia: Clasificación del Buro vs Tarjeta	30
3.2.2 Tabla de Contingencia: Canal Apertura vs Compro	32
3.2.3 Tabla de Contingencia: GrupoHijos vs VariableRespuesta	34
3.2.4 Tabla de Contingencia: GrupoEdad vs VariableRespuesta	36
3.2.5 Tabla de Contingencia: Clasificación del Buro vs VariableRespuesta.....	38
3.2.6 Tabla de Contingencia: Canal vs VariableRespuesta	40
3.4. Regresión Logística	42
3.4.1 Modelo Estadístico	42
3.4.2 Coeficientes Estimados y Prueba de Wald	44
3.4.3 Resultados de los Razón de Propensión (Odd Ratio)	48
CONCLUSIONES Y RECOMENDACIONES	vii
Conclusiones	vii
Recomendaciones.....	ix
BIBLIOGRAFÍA	x
Anexo: Código de Programación en R	xii

CAPÍTULO I

1. DESCRIPCIÓN DEL PROBLEMA

1.1 Introducción

Medir el nivel de riesgo de un cliente en términos crediticios, presenta una gran incertidumbre, ofrecer un crédito directo a un cliente nuevo genera un riesgo en la cartera de crédito. En la empresa de RETAIL, la tasa de morosidad y las provisiones que se genera por el préstamo del crédito directo al cliente, son tan importantes como las ventas del producto principal.

La tasa de morosidad es un indicador que refleja la relación que existe entre la cartera que se encuentra en mora sobre el total de la cartera de créditos. Este indicador se presentan a los clientes que no cancelan su primera cuota vencida y alcanzan una edad de mora mayor a 30 días, actualmente la tasa de morosidad en la empresa de RETAIL supera el 16% del saldo de la cartera de crédito, lo cual representa un incremento del 4% del indicador al cierre del año 2014. Por lo que se estima que los clientes no están pagando su totalidad de saldo vencido.

La cartera de crédito crece al mismo ritmo que el mercado crediticio, en la actualidad la empresa de RETAIL ha crecido el 25% en saldo de cartera con relación al cierre del año 2014. La nómina del departamento de cobranzas

se mantiene con la misma cantidad de gestores, por lo que no se alcanzan a gestionar gran parte de la cartera, de tal manera que la gestión no está siendo efectiva.

La solución más eficiente no sería aumentar la nómina de cobranzas, sino disminuir el nivel de riesgo de los clientes nuevos. Para lograrlo, se debe cuantificar el comportamiento del cliente identificando los perfiles y motivos que genera el incremento de la morosidad.

1.2 Justificación del Problema

Es primordial que el nivel de riesgo sea mínimo sin disminuir el volumen de ingreso de clientes, es decir más exigentes en la aceptación al crédito directo, obteniendo mejores perfiles para el pago a tiempo de sus cuotas vencidas y optimizando los procesos de gestión en el departamento de cobranzas.

El riesgo de la cartera de crédito en la empresa de RETAIL se mitiga estableciendo un modelo estadístico, que nos permita identificar y tomar medidas acorde al nivel de riesgo en morosidad de los clientes.

1.3 Objetivos

- ✓ Identificar los perfiles sociodemográficos de los clientes que se caen en mora.

- ✓ Diseñar e implementar un modelo estadístico que permita disminuir el nivel de riesgo de la cartera de crédito de la empresa de RETAIL.

- ✓ Conglomerar la cartera de crédito de acuerdo a los perfiles de los clientes.

CAPÍTULO 2

2. Metodología

2.1 Marco Referencial

2.1.1 Antecedentes

La empresa RETAIL es una compañía que proporciona crédito directo a los clientes a nivel nacional, con una cartera de crédito de más de 500 mil clientes, se enfrenta al mundo crediticio con la incertidumbre y el riesgo que debe correr el aceptar el crédito a un cliente.

En los últimos meses, la cartera de crédito de RETAIL ha crecido considerablemente con respecto al año pasado por lo que la provisión y morosidad se ha elevado.

2.2 Marco Conceptual

La cartera de crédito de la empresa de RETAIL se encuentra distribuida de la siguiente manera:

Tabla 2.01: Distribución de la cartera de crédito – Empresa RETAIL
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Edad	Rango	Proceso de Cobro	Evento
Vigente	0 días	Cobranza Preventiva	Contacto telefónico y Clientes si puede comprar
30 días	1 a 30 días vencido	Cobranza administrativa (interna)	Contacto telefónico y Clientes no puede comprar
60 días	31 a 60 días vencido	Cobranza administrativa (interna)	Contacto telefónico y Clientes no puede comprar
90 días	61 a 90 días vencido	Cobranza administrativa (interna)	Contacto telefónico y Clientes no puede comprar
120-180 días	91 a 180 días vencido	Cobranza Pre-Jurídica	Lo gestionan las casas de cobro externas
210 en adelante	Desde 181 días de vencido	Cobranza Judicial	Se gestiona con agentes externos (Abogados), si la cuenta permanece 1 año y no se ha recuperado, es dado de baja.

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

Los clientes que genera mayor mora dentro de la cartera de crédito están entre las edad 30 y 90 días, por lo que el modelo tratará de mejorar la estrategia de cobro de la cartera interna, de tal manera que se minimice el riesgo y las funciones de la gestión del departamento de Cobranzas.

Es importante conocer algunos conceptos básicos de cobranzas, a continuación se detalla lo siguiente:

Índice de Morosidad: El índice de morosidad es la cartera de crédito vencida como proporción de la cartera total. Es uno de los indicadores más utilizados como medida de riesgo

$$\text{Índice de Morosidad} = \frac{\text{Cartera Vencida}}{\text{Cartera Total}}$$

Para clasificar los vencimientos se requieren una edad de mora más de 30 días.

Índice de Provisiones: Es un porcentaje de la cartera vencida que guarda la empresa en caso de que el cliente no pague su deuda, este porcentaje de provisión se va incrementando a medida que crece la edad de mora, es decir un cliente con una edad de mora mayor es menos probable que pague su deuda por lo que la empresa provisiona o asume gran parte de su deuda¹.

Cuando se presentan estafas en la cartera de crédito se provisionara el total de la deuda, con la autorización de Gerencia. El porcentaje de provisión a aplicarse será:

¹ **RUIZ H. (2013)** "Diseño de un modelo matemático para optimizar la gestión de un sistema de cobranza en una empresa comercial", ESPOL – GUAYAQUIL – ECUADOR.

Tabla 2.02: Porcentaje de Provisión para cada Edad de Mora
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Edad de Morosidad	% Provisión
01-30 días (Vigentes)	2%
30 días	4%
60 días	10%
90 días	20%
120 días	30%
150 días	40%
180 días	60%
más de 180 días	100%

Elaborado Por: *D. Hernández*, Facultad de Ciencias Naturales y Matemáticas (FCNM).

Políticas del Departamento:

- ✓ La gestión de recuperación cartera de crédito estará a cargo del departamento de cobranzas.
- ✓ Las cuentas con edad de mora mayor a 120 días serán cerradas de forma automática y no podrá utilizar su tarjeta hasta cancelar la totalidad de su deuda, previa una solicitud de reapertura de cuenta.
- ✓ Las cuentas con edad de mora mayor a 210 días, serán transferidas a cobranza judicial, para que se realice la gestión de cobranzas correspondiente por medio de agentes de cobro externos (abogados o casas de cobro).
- ✓ Las cuentas que presenten los siguientes motivos serán trasferidas directamente a cobranza judicial.

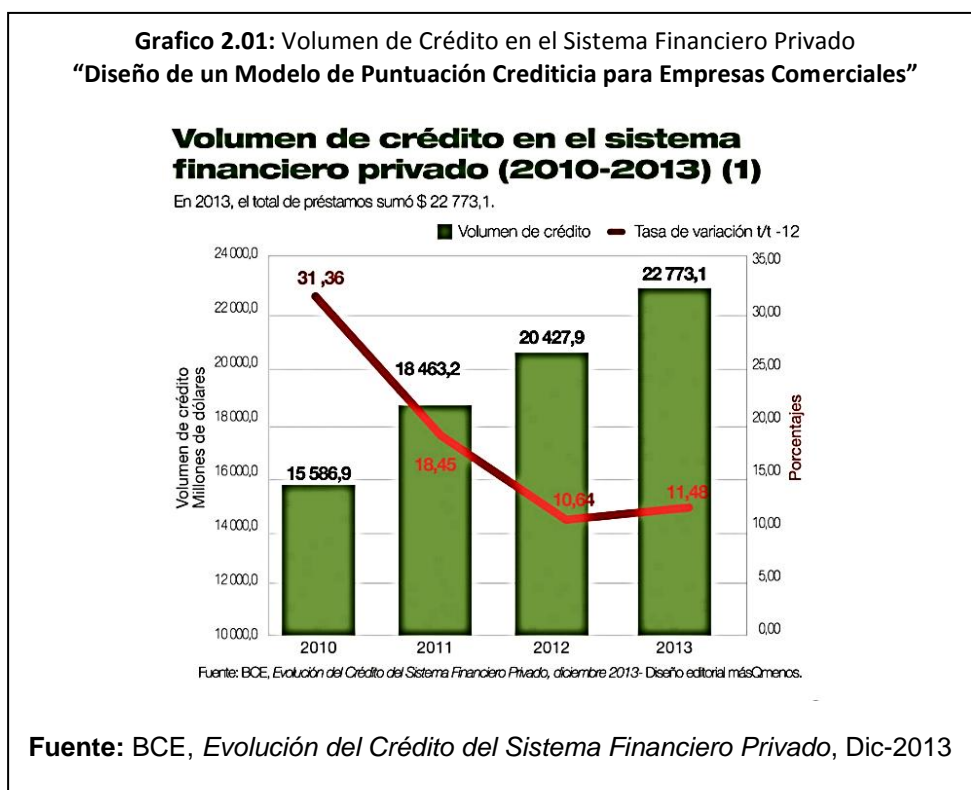
- ✓ Clientes desaparecidos y nadie se quiere hacer cargo de la deuda
- ✓ Cuentas que tienen indicios de estafas
- ✓ No se podrá divulgar información sobre cliente alguno para fines personales.
- ✓ Cuando el valor vencido de la cuenta es menor a \$ 4 dólares, no se reflejará la cuenta como vencida, por lo tanto no estará dentro de las cuentas asignadas para la gestión.

2.3 Marco Teórico

La otorgación de un crédito directo tiene dos puntos claves: si no hay confianza con el cliente no existe el crédito y si se otorga el crédito existe riesgo en él. Vale recalcar que la recuperación del crédito proviene de la viabilidad en la empresa y de la solvencia moral y económica del cliente. Las garantías no constituyen sino un refuerzo adicional para la aprobación del crédito pero en ningún momento sustituye la capacidad de pago del cliente. Es preferible declinar una operación o un modelo estadístico que basarla exclusivamente en garantías.

El riesgo crediticio se da cuando no se recupera todo o parte del dinero prestado junto con los intereses generados. Este riesgo que en los últimos años se ha vuelto incertidumbre para las empresas ya que el mercado crediticio ha crecido debido a la importancia en las operaciones de compra – venta, el volumen de crédito aumentó en 2010 31,36%; en 2011, 18,45% y

en 2012, 10,64%. El año pasado la cifra creció 11,48% y sumó \$ 22 773,13 millones². Según el analista económico Walter Spurrier³ explicó que de los datos se infiere que “el segmento del crédito de los bancos del sector público tiende a crecer más aceleradamente que el crédito de la banca privada... (y) en los últimos 10 años ha crecido muchísimo el sector de cooperativas comparado con el de bancos”.



² BANCO CENTRAL DEL ECUADOR (BCE) (2013)

<http://www.bce.fin.ec/index.php/component/k2/item/755>

³ WALTER SPURRIER Analista Económico graduado de un Masterado de Ciencias Políticas en la Universidad de California en los Ángeles (UCLA). Editorialista de Diario El Comercio y El Universo. Actualmente es Director de Análisis Semanal, Presidente de Grupo Spurrier.

Este riesgo crediticio puede mitigarse mediante técnicas estadísticas. Los estudios relacionados con el riesgo crediticio se han desarrollado desde mucho tiempo atrás, donde los métodos estaban más enfocados en las características de las carteras de crédito de cada empresa que de los perfiles de los clientes entrantes.

En el 2007 se calculó las probabilidades de impago para establecer el mínimo monto a cobrar y las provisiones de cartera, a través de las metodologías de redes neuronales, el objetivo del análisis es encontrar algunas relaciones para grupos determinados de la población, de acuerdo con sus características particulares. Las Redes Neuronales son un conjunto de algoritmos matemáticos que encuentran las relaciones no lineales entre conjuntos de datos; suelen ser utilizadas como herramientas para la predicción de tendencias y como clasificadoras de conjuntos de datos⁴.

La técnica de Matrices de Probabilidades de Transición fue desarrollada en 1997 con el objetivo de estimar la probabilidad de incumplimiento de pago del cliente, tomando como referencia la cartera de crédito en una institución financiera de Colombia. Se define la pérdida esperada como la probabilidad de incumplimiento de un cliente cuando este alcance una altura de mora n , en la cual la institución asume la pérdida del capital en un tiempo dado⁵.

⁴ **PEREZ F, FERNANDEZ H (2007)** "LAS REDES NEURONALES Y LA EVALUACIÓN DEL RIESGO DE CRÉDITO" - Revista Ingenierías Universidad de Medellín, volumen 6, No. 10, pp. 77-91 - ISSN 1692-3324 - Enero-junio de 2007/166p.

⁵ **TAMARA A, ARISTIZABAL R, VELASQUEZ E (2012)** "Matrices de transición en el análisis del riesgo crediticio como elemento fundamental en el cálculo de la pérdida esperada en una institución financiera colombiana"

Otra de las técnicas utilizadas para medir el riesgo crediticio es la minería de datos⁶, está contiene dos etapas de procesamiento; en la primera etapa, la agrupación de los clientes nuevos, en donde los grupos son homogéneos y se eliminan los grupos aisladas y grupos inconsistentes que son re etiquetados. En la segunda etapa, se construye el modelo de clasificación en el que se obtiene como resultado más de tres grupos, a diferencias de otros modelos de clasificación en donde solo dan como resultado dos grupos, “bueno” o “malo”.

La importancia de esta técnica, es que no solo predice si el cliente tiene buen crédito o no, si no que proporciona más de dos opciones, y para efecto de gestiones estos resultados identificaría mejor el riesgo de crédito.

Una de las técnicas que con mayor frecuencia se usa en los últimos tiempos para predecir el riesgo en la cartera de crédito es Arboles de Decisión, para segmentar la población o ver los perfiles de clientes moroso en un periodo de tiempo, y Regresión Logística para observar que tan propenso son los clientes al pago puntual, o en otros casos se utiliza la regresión para dar un score (calificación) a un grupo de clientes con ciertas características.

⁶ **CHEN W (2012)** “Evaluación del riesgo de crédito mediante la técnica de minería de datos híbrido” - Systems Engineering Procedia 3 (197 – 200).

Este es el caso de un estudio que desarrolló una metodología para elaborar un modelo que logre predecir el comportamiento de impago en función de la información sociodemográfica del cliente. Un scorecard clasifica a la población objetivo dentro de dos o más grupos usando diversas técnicas estadísticas⁷. Es común encontrar, por un lado, propuestas que emplean métodos econométricos de variables dependientes limitadas, como lo son los modelos logit, probit y logísticos; y por otro lado, los métodos de clasificación estadística.

Es así como existe en la literatura una variedad de estudios referente al análisis del riesgo crediticio para lo cual se han utilizado diferentes metodologías para la búsqueda de la probabilidad de incumplimiento. Por lo tanto, el modelo a utilizar dependerá de la información con que cuente la empresa, además de tener en cuenta el tipo de producto que se ofrece.

La metodología que se aplicará para mitigar la problemática de morosidad dentro de la empresa RETAIL será el diseño de un modelo estadístico que logre predecir el comportamiento de impago de un cliente en función de su información sociodemográfica, como también los comportamientos ciertos grupos de clientes con características homogéneas.

⁷ **ESPIN-GARCIA O, RODRIGUEZ-CABALLERO C (2013)** "Metodología para un Scoring de clientes sin referencias crediticias". Cuadernos de Economía, 32(59).

2.4 Población

La empresa RETAIL proporciona el crédito directo a Nivel nacional en el Ecuador a más de 500 mil clientes vigentes, es decir clientes que tienen su cuenta activa. Para nuestro efecto de proyecto, la población objetivo a analizar serán los clientes aprobados en el periodo del 2013 que tengan su cuenta activa. Se tomará este periodo con el fin de analizar el comportamiento de su pago en 12 meses, es decir las cuentas aprobadas en DIC-2013 se las medirá hasta DIC-2014.

Aproximadamente en la población objetivo hay 89 mil clientes aprobados, de las cuales cerca del 82% tienen su cuenta vigente.

2.5 Análisis Estadístico Univariado

En esta sección se realizará el análisis estadístico Univariado de las variables más relevante de acuerdo al objetivo, en las cuales se describirá, analizará y representará los datos de la muestra utilizando métodos estadísticos y gráficos que resuman la información contenida en ellas.

Entre las principales medidas descriptivas que serán analizadas tenemos, medidas de tendencia central como media y mediana, medidas de dispersión como desviación estándar y varianza, medidas de posición como cuartiles, mínimo y máximo y medidas de forma como sesgo o asimetría y curtosis o puntiagudez. Además para ilustrar la información se realizarán gráficos como histogramas de frecuencias y diagramas de caja

El objetivo del análisis es poder observar la distribución de la población objetivo a nivel sociodemográfico.

2.6. Análisis Estadístico Inferencial

En esta sección el análisis es aplicado con métodos estadísticos inferenciales, en la cual nos ayudará a determinar las propiedades de la población objetivo, su distribución de algunas variables y la relación que exista entre dos o más variables aplicando tablas de contingencia.

El objetivo de utilizar el método de las tablas de contingencia es identificar la relación de dependencia entre variables cualitativas, el contraste estadístico estará basado en el estadístico χ^2 (Chi-cuadrado), cuyo cálculo nos permitirá afirmar con un nivel de confianza estadístico determinado si los niveles de la variable cualitativa influyen en los niveles de la otra variable nominal analizada. La variable cualitativa a analizar será TARJETA (si tiene o no tarjeta de crédito externa) vs la variable CLASIFBR (Clasificación del Buró) ya que se estima algún tipo de relación entre las dos variables y el segundo caso analizar están entre las variables, CANAL (Canal de apertura de la cuenta) vs COMPRO (Si se utilizó o no el cupo del crédito directo), debido a que se quiere observar en cuanto es la relación de dependencia ya que ambas variables influyen en el indicador de morosidad.

Se estima que los clientes con mayor número de hijos tienen mayor probabilidad de caer en mora, es por esto que se plantea el método de contraste de hipótesis, en donde la población objetivo es agrupada en dos partes, el grupo 1, la población de clientes que tiene menor o igual que 1 hijo vs el grupo 2, la población de clientes que tienen mayor de 1 hijo. El contraste de hipótesis será; la probabilidad de caer en mora para el grupo 1 es igual que para el grupo 2. El segundo contraste de hipótesis es identificar en cuanto difiere estos grupos. El método para las pruebas de hipótesis será el Kolmogorov-Smirnov, ya que asumiremos normalidad en la variable. EL software estadístico a utilizar será SPSS.

2.7. Regresión Logística

La regresión logística es un modelo no lineal mediante el cual se puede determinar la relación entre una variable de respuesta Y que es binaria y una o más variables de explicación X_1, X_2, \dots, X_k , que son variables continuas.

La variable de respuesta Y es: si el cliente es “bueno” o “malo” en termino de mora, el cual tomará valores 1 si es “bueno” (si el cliente no cae en mora en un periodo de 12 meses después del mes de aprobación del crédito directo) y 0 si es “malo” (si el cliente cae en mora en un periodo de 12 meses después del mes de apertura del crédito directo).

Las variables de explicación X_1, X_2, \dots, X_k , serán las consideradas en el análisis multivariado y las que resulten del perfil de clientes medidos en el análisis de árboles de decisión.

La función de densidad de la variable aleatoria X que tiene Distribución Logística con parámetro θ , es:

$$f(x) = \frac{e^{-(x-\theta)}}{(1 + e^{-(x-\theta)})^2}; \text{ con soporte } S = \mathbb{R}; \theta \in \mathbb{R} \quad (4.01)$$

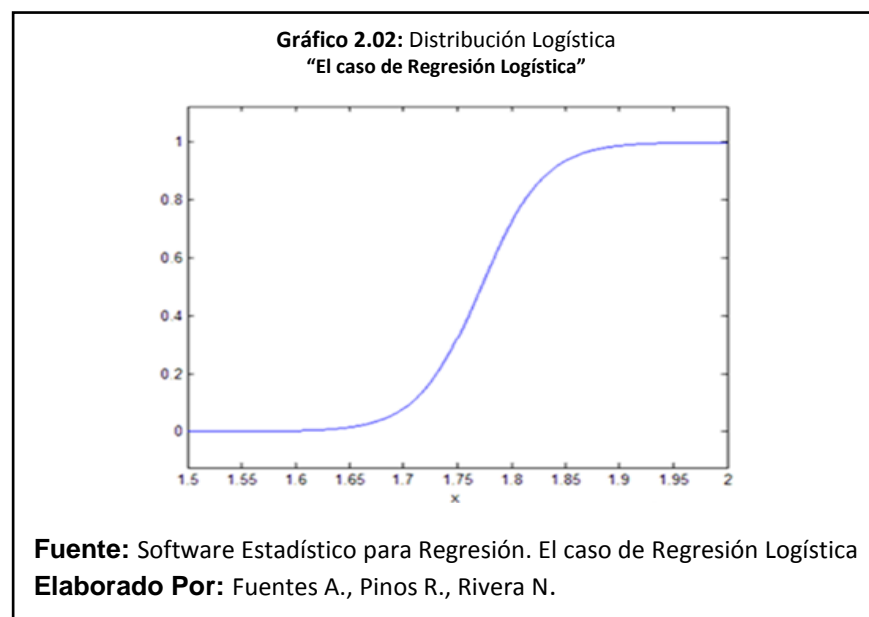
Para el caso cuando θ es cero se lo llama Distribución Logística, la cual es:

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}; S = \mathbb{R} \quad (4.02)$$

Su Distribución Acumulada $F(x) = P(X \leq x)$ es;

$$P(X \leq x) = F(x) = \frac{1}{1 + e^{-x}}, x \in \mathbb{R} \quad (4.03)$$

La representación gráfica de $f(x)$ se presenta en el Gráfico 2.03



Como pueden apreciar f es una curva que se extiende sobre R y cuyo dominio en el intervalo real que va desde cero hasta uno; la curva presentada es monótona creciente.

Este método estadístico es útil para modelar la probabilidad de un evento como función de otros factores, en este caso observaremos las probabilidades de éxito dado su comportamiento sociodemográfico, en otras palabras la probabilidad de un cliente sea “bueno” dado sus variables de relación.

El objetivo de aplicar este modelo es poder observar estadísticamente si el cliente paga su deuda a tiempo dado su comportamiento historial crediticio con variables sociodemográficas.

Observar que tan propensos son los clientes al pago, predecir si el cliente caerá en mora dado que su comportamiento en 12 meses. Comprobar que los perfiles de clientes “malos” y “buenos” se ajustan al modelo logístico.

Existen algunos software estadísticos en las que se aplica el modelo logístico tales como: STATA y SPSS. El software que aplicaremos este modelo será R-Studio, un software libre de lenguaje de programación para análisis estadístico. Esta herramienta es la más óptima para este tipo de análisis debido a que permite que los usuarios lo extiendan definiendo sus propias funciones, obteniendo los resultados y gráficos de forma dinámica.

2.7.1 Estimación de parámetros en un modelo de Regresión

Logística⁸

En Regresión Logística la estimación de los coeficientes del modelo y de sus errores estándar se recurre al método de Máxima Verosimilitud, es decir, estimaciones que hagan máxima, la probabilidad de obtener Y proporcionados por los datos de la muestra. Para el cálculo de estimaciones máximo-verosímiles en Regresión Logística, ya que no se obtienen expresiones explícitas para los valores de “los betas” incluidos en el modelo y por tanto debe recurrirse a métodos iterativos, como lo hemos enunciado, usaremos el método de Newton–Raphson

Para aplicar el método de Máxima Verosimilitud en Regresión Logística se trabaja con que cada observación Y_i de la muestra sigue la distribución de Bernoulli, suponiendo independencia de las n observaciones, donde la densidad de probabilidades conjuntas, dado β , de y_1, y_2, \dots, y_n está dada por:

$$p(y_1, y_2, \dots, y_n | \beta) = \prod_{i=1}^n [\pi_i^{y_i}] [1 - \pi_i]^{1-y_i} \quad (4.04)$$

Entonces la función de verosimilitud está dada por:

$$L(\beta | y_1, y_2, \dots, y_n) = \sum_{i=1}^n y_i \ln \pi_i + \sum_{i=1}^n (1 - y_i) \ln(1 - \pi_i) \quad (4.05)$$

⁸ **A. Fuentes, N. Rivera, R. Pinos (2011)** “Software Estadístico para Regresión. El caso de Regresión Logística y Regresión Poisson” ESPOL-GUAYAQUIL.

2.7.2 Evaluación de los Modelos de la Regresión Logística

El siguiente paso será comprobar la significación estadística de cada uno de los coeficientes de regresión en el modelo. Para ello podemos emplear dos métodos, el del Estadístico de Wald y el del Estadístico G de Verosimilitud:

2.7.2.1 El estadístico de Wald.

Se utiliza el denominado estadístico W de Wald que se define como:

$$W = \frac{\hat{\beta}_j^2}{\text{var}(\hat{\beta}_j^2)}$$

Que tiene aproximadamente una distribución χ^2 con $(p - 1)$ grados de libertad.

Para el caso multivariado, W se lo expresa como la expresión matricial:

$$W = \hat{\beta}'[\Sigma(\hat{\beta})]^{-1}\hat{\beta} = \hat{\beta}'(X'VX)\hat{\beta} \quad (4.15)$$

Se hace el siguiente contraste de hipótesis:

$$H_0: \beta_j = 0, \quad j = 1, 2, \dots, p$$

Vs.

$$H_1: \beta_j \neq 0, \quad j = 1, 2, \dots, p$$

Como regla general rechazamos H_0 a favor de H_1 si el nivel de significancia de la muestra (valor p) es menor que 0.05, definiendo la Región Crítica como: con $(1 - \alpha)100\%$ de confianza se rechaza H_0 a favor de H_1 si $W > \chi^2_{1-\alpha}(p - 1)$.

2.7.2.2 El estadístico G de la razón de verosimilitud.

Otra opción para verificar estadísticamente el valor de los parámetros β_1, β_2, \dots es utilizar el denominado estadístico G de la Razón de Verosimilitud, cuya definición se bosqueja a continuación:

Se trata de comparar el modelo que resulta de eliminar de forma aislada cada una de las covariables frente al modelo completo. En este caso cada estadístico G sigue una distribución χ^2 con 1 grado de libertad (no se supone normalidad). La ausencia de significación implica que el modelo sin la covariable eliminada no desmejora respecto al modelo completo (es decir, da igual su presencia o su ausencia), por lo que según la estrategia de obtención del modelo más reducido (principio de parsimonia), dicha covariable debe ser eliminada del modelo ya que no es significativa en el mismo. Esta prueba no supone distribución alguna, por lo que es la más recomendada.

Es más una método de “prueba y error”, que compara diferentes modelos donde se sustituyen las variables que se emplean, por lo que en si no tiene un contraste de hipótesis.

2.8 Recolección de Información

La recolección de información será por medio de las solicitudes de crédito, esto reposa en las bases de datos de la empresa. Es muy probable encontrar error en la información, proveniente de las solicitudes de crédito, ejemplo de ello serían ingresos exagerados, número de cargas familiares, menores de edad entre otros, por lo que se recomienda considerar estos casos como anormales para evitar que traigan complicaciones en los análisis posteriores.

Este tipo de casos fueron excluidos para el análisis, con el fin de tener una base depurada y un análisis más real para la solución a la problemática.

CAPÍTULO 3

3. Análisis y Modelos

3.1. Análisis Estadístico Univariado

A continuación se presentan el análisis descriptivo para las variables más relevantes dentro del análisis: En el periodo del 2013, se han aprobado un total de 89,677 cuentas, siendo los meses SEPT y DIC los de mayor aporte. El promedio de cuentas aprobadas es de 7,473 al año con una desviación estándar de 1,375 cuentas. El mes de menor aporte es Mayo con 5,431 cuentas aprobadas.

Tabla 3.01: Número de Aprobaciones por Meses de la Empresa RETAIL
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Aprobaciones de Crédito Directo en el 2013

Periodo	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
01-ene	8.064	9	9	9
02-feb	6.070	6,8	6,8	15,8
03-mar	7.672	8,6	8,6	24,3
04-abr	5.798	6,5	6,5	30,8
05-may	5.431	6,1	6,1	36,8
06-jun	7.482	8,3	8,3	45,2
07-jun	6.360	7,1	7,1	52,3
08-ago	7.790	8,7	8,7	61
09-sep	9.296	10,4	10,4	71,3
10-oct	7.458	8,3	8,3	79,6
11-nov	8.311	9,3	9,3	88,9
12-dic	9.945	11,1	11,1	100
Total	89.677	100	100	

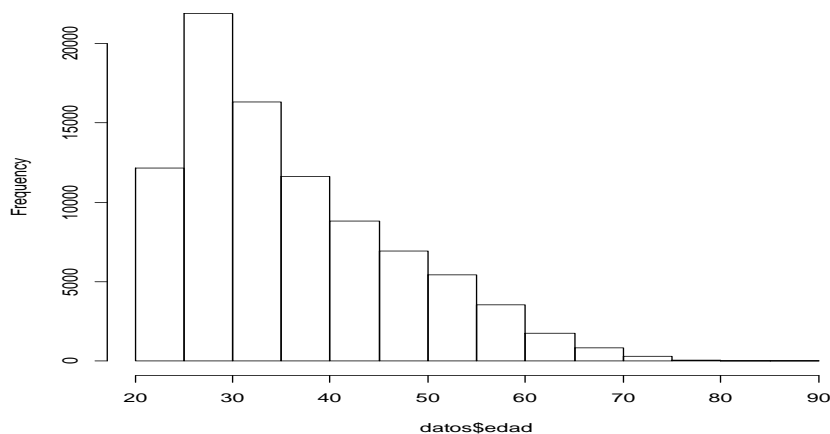
Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

3.1.1 Distribución Por Edad

Cuadro 3.01: Distribución de la cartera de crédito por Edad de Nacimiento
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Estadísticos descriptivos								
	N	Mínimo	Máximo	Media	Desv. típ.	Varianza	Curtosis	
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Error típico
EDAD_NAC	89677	20	86	36,50	11,007	121,150	,180	,016

Histogram of datos\$edad

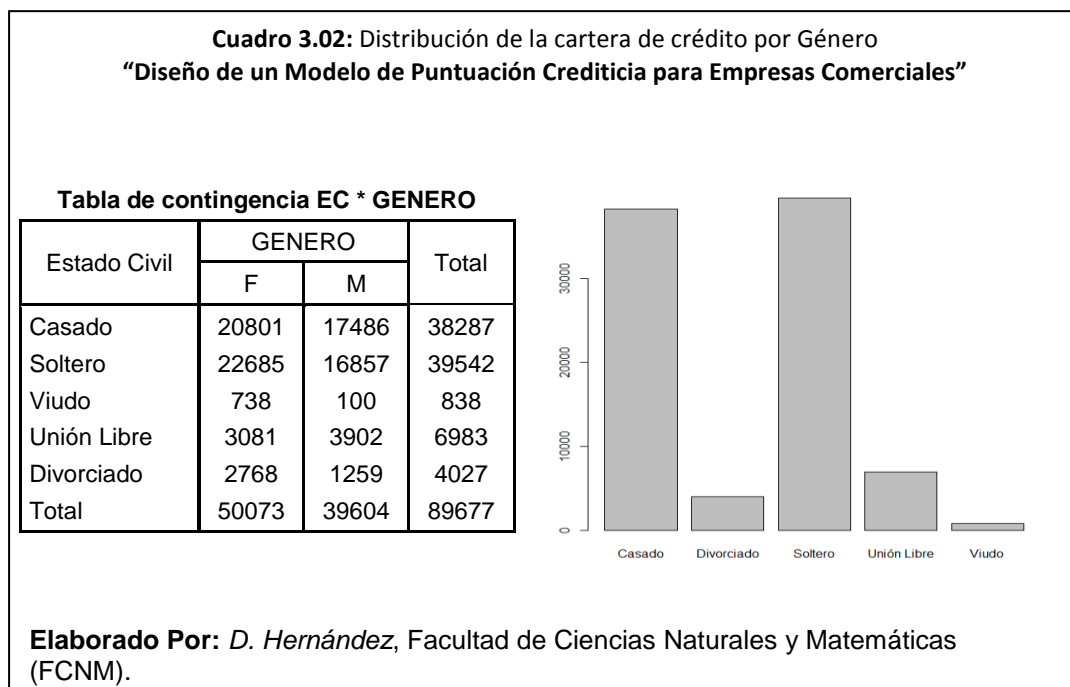


Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

El promedio de edad de los clientes aprobados en el 2013 es de 36.5 años con una desviación estándar de 11 años. La edad mínima es de 20 años y la máxima de 86 años, se estima que estos clientes de edad elevada son cuentas reaperTURADAS⁹. El error típico es 0.016. El 25% de los clientes están concentrados menores a 28 años, el tercer cuartil es menor o igual a 43 años.

⁹ **CUENTAS REAPER TURADAS** - son cuentas cerradas en su momento por algún motivo que luego de un tiempo considerado, 3 meses, puede aplicar a reabrir su cuenta, siempre y cuando no esté en mora y cumpla los requisitos.

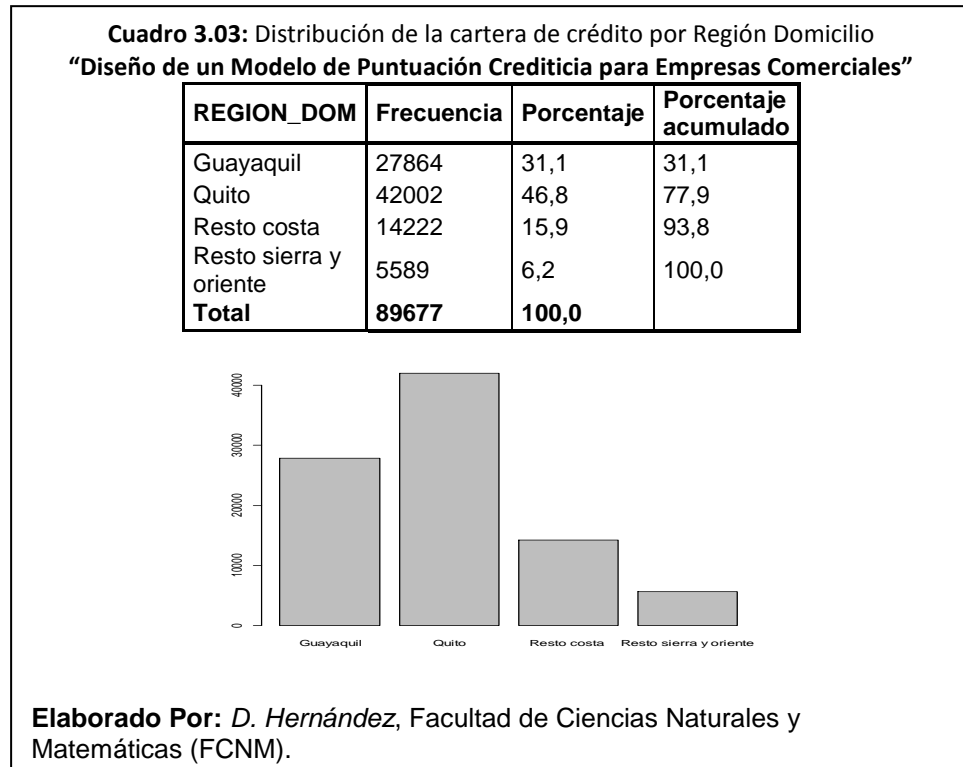
3.1.2 Distribución Por Género



El 56% de clientes que entraron en el 2013 son Mujeres (50,073 cuentas), de los cuales el 25% son solteras y 23% son casadas. Del 44% clientes hombres, el 19% son solteros y el mismo porcentaje son casados. El 44% de los clientes son Solteros, el 43% son Casados.

Uno de los perfiles de la empresa de RETAIL para dar crédito directo es a clientes Mujeres, por el producto que ofrece, de las cuales estas cifras se ajustan de cierta manera a los clientes entrantes en el periodo de medición.

3.1.3 Distribución Por Región Domicilio

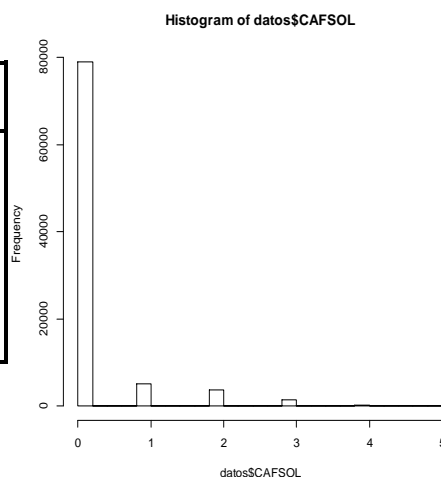


Como inicialmente se describió, la empresa RETAIL proporciona el crédito directo a nivel nacional, donde el 46.8% de los clientes que fueron aprobados en el periodo del 2013 son Quito, seguidos de Guayaquil con el 31.1%, la mayoría de los clientes de la cartera de crédito son de estas dos ciudades, el resto de los clientes pertenecen a provincias tales como: 15.9% al Restos Costa y 6.2% al Resto Sierra y Oriente. El 17% de los clientes de Guayaquil son Femeninos, el 20% de Quito son Masculinos. En promedio en Guayaquil se aprueban 2,322 clientes mensuales, en Quito el promedio es de 3,500 clientes mensuales.

3.1.4 Distribución Por Cargas Familiares

Cuadro 3.04: Distribución de la cartera de crédito por Carga Familiares
 “Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

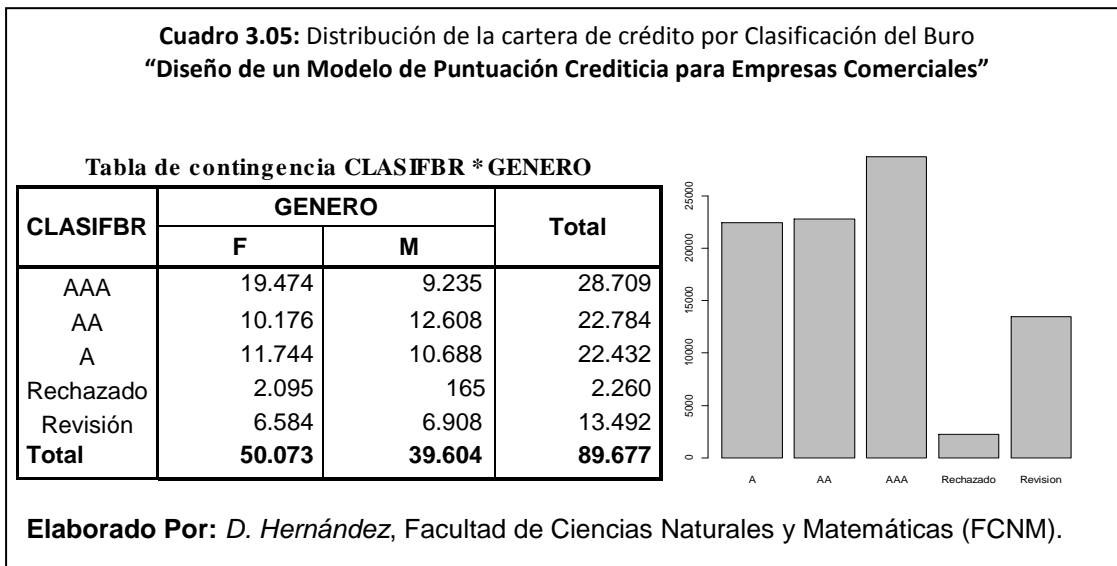
CAFSOL	Frecuencia	Porcentaje	Porcentaje acumulado
0	78.932	88	88
1	5.215	6	94
2	3.787	4	98
3	1.411	2	100
4	272	0	100
5	60	0	100
Total	89.677	100	



Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

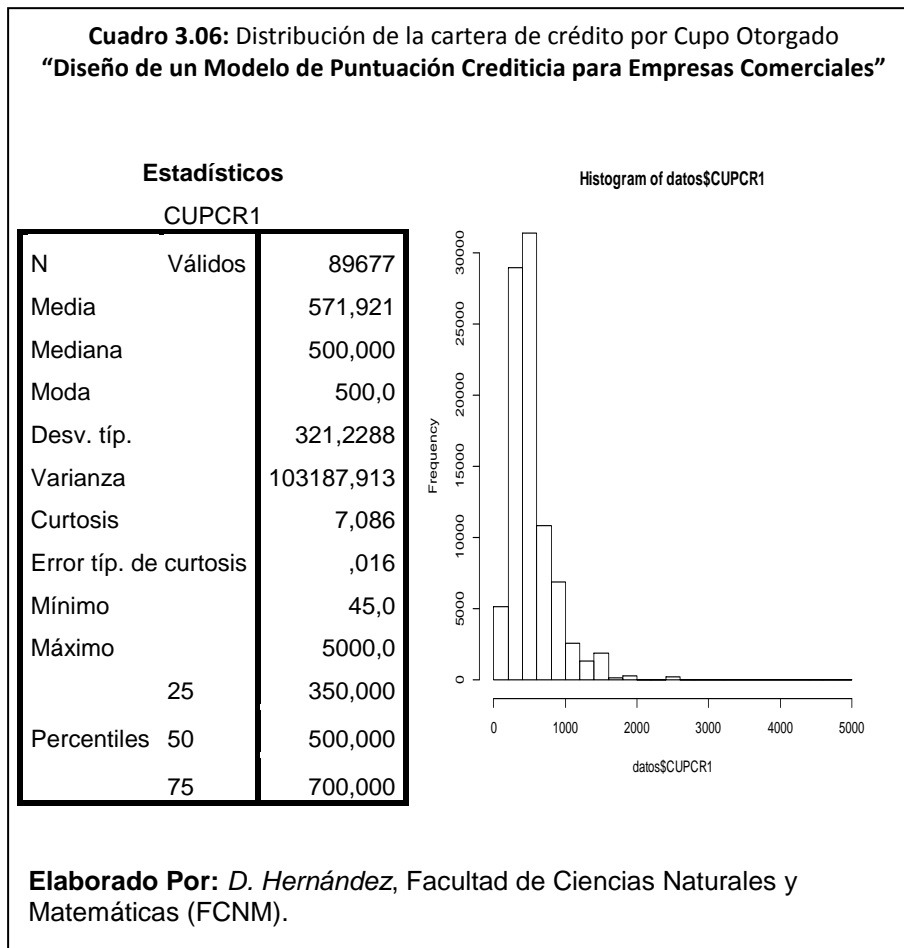
Esta variable, cargas familiares o números de hijos, se estima que tenga una relación fuerte con la morosidad, por ejemplo la probabilidad de los clientes con mayor número de cargas familiares es más alta de caer en mora. El 88% de clientes entrantes en el 2013 no tiene cargas familiares, el promedio de números de hijos es 0.6 equivalente a 1 hijo por cliente. Del grupo mayoritario que no tienen cargas familiares, el 56% son Mujeres y el 13% se caen en mora en promedio. De los clientes que tienen 5 cargas, el 67% son Hombres y el 23% tienen vencidas sus cuentas, es decir este grupo de clientes con estas características tienen mayor probabilidad de caer en mora.

3.1.5 Distribución Por Clasificación del Buro



Esta variable es proporcionada por el Buró de crédito, EQUIFAX, que brinda la información crediticia del cliente en términos de una calificación, siendo los AAA de mejor comportamiento crediticio y los Rechazados los peores. De los clientes que se aprobaron en el 2013, el 32% son cuentas AAA con un mora del 7% en promedio, el 25% son clientes AA con mora del 12% y clientes A con mora del 18%. El 3% de los clientes aprobados tienen una calificación en el Buró de Rechazados, este grupo de clientes fueron ingresados por excepción según la consideración del analista, teniendo una mora del 24% en promedio.

3.1.6 Distribución Por Cupo Otorgado



Los cupos de créditos son los valores que se les asignan o prestan la empresa a los clientes por el crédito directo, esto va medido de acuerdo al sueldo actual y otras variables sociodemográficas, el cupo promedio asignado a los clientes aprobados en el 2013 es de \$500 con una desviación del \$321. El cupo máximo asignado es de \$5000 y el mínimo de \$45 (crédito inmediato).

3.1.7 Distribución Por Cliente Malo vs Cliente Bueno

Cuadro 3.07: Distribución de la cartera de crédito por Cliente “Malo” vs Cliente “Bueno”
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”



Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

3.2. Análisis Estadístico Inferencial

3.2.1 Tabla de Contingencia: Clasificación del Buro vs Tarjeta

La relación que existe entre las variables Tarjeta, si tiene o no alguna tarjeta de crédito externa, y la variable de la Clasificación del Buró está dada por las tablas de contingencia que se presenta a continuación:

Tabla 3.02: Tabla de Contingencia: Clasificación del Buro vs Tiene Tarjeta Externa
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Tabla de contingencia CLASIFBR * TARJETA

			TARJETA		Total
			NO	SI	
A	Recuento	13569	8863	22432	
	% del total	15,1%	9,9%	25,0%	
AA	Recuento	10277	12507	22784	
	% del total	11,5%	13,9%	25,4%	
CLASIFBR AAA	Recuento	11102	17607	28709	
	% del total	12,4%	19,6%	32,0%	
Rechazado	Recuento	1911	349	2260	
	% del total	2,1%	0,4%	2,5%	
Revisión	Recuento	12744	748	13492	
	% del total	14,2%	0,8%	15,0%	
Total	Recuento	49603	40074	89677	
	% del total	55,3%	44,7%	100,0%	

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

Del 55% de los clientes que no presenta tarjeta de crédito, el 15% son clientes A, 11.5% AA y el 12.4% clientes AAA. Y de los clientes que si tienen tarjeta, el 19% son clientes AAA y el 13% son clientes AA.

Tabla 3.03: Prueba de Chi Cuadrado: Clasificación del Buro vs Tiene Tarjeta
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	13565,541 ^a	4	,000
Razón de verosimilitudes	15796,674	4	,000
N de casos válidos	89677		

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1009,93.

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

La prueba de chi-cuadrado nos muestra estadísticamente que existe relación de dependencia entre las dos variables, es decir la calificación del Buro depende si el cliente tiene o no tarjeta, con esa información podemos tomar ambas variables para el diseño de modelo de puntuación crediticia.

Otras de las variables que se puede considerar en el modelo estadístico son las variables, si utilizó o no el cupo y el canal de apertura de la cuenta, para aquello se realizan la prueba de tabla de contingencia para ver la relación entre ellas.

3.2.2 Tabla de Contingencia: Canal Apertura vs Compro

Tabla 3.04: Tabla de Contingencia: Canal vs Compro
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Tabla de contingencia CANAL * Compro

			Compro		Total
			NO	SI	
Atención cliente	Recuento		345	2126	2471
	% del total		0,4%	2,4%	2,8%
Empresa Ext	Recuento		8503	8964	17467
	% del total		9,5%	10,0%	19,5%
Internet	Recuento		127	1432	1559
	% del total		0,1%	1,6%	1,7%
CANAL Piso venta	Recuento		1997	12873	14870
	% del total		2,2%	14,4%	16,6%
Prom Cred Inmd	Recuento		2494	10146	12640
	% del total		2,8%	11,3%	14,1%
Prom Ext	Recuento		10849	13662	24511
	% del total		12,1%	15,2%	27,3%
Prom Tienda	Recuento		3254	12905	16159
	% del total		3,6%	14,4%	18,0%
Total	Recuento		27569	62108	89677
	% del total		30,7%	69,3%	100,0%

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

El 69.3% de los clientes ha utilizado el crédito directo, de los cuales el 15% se lo hizo por medio del canal Prom. Externa, el 14.4% son de los canales Piso de Venta y Prom. Tienda, y el 11% son del canal Prom. Cred. Inmd. Del

30% de los clientes aprobados que aún no consumen su crédito directo, el 12% son de Prom. Externas y el 9.5% de Empresas Externas. Está claro que los canales en donde los clientes más consumen su crédito directo son: Atención al Cliente, Piso de venta, Prom tienda.

Tabla 3.05: Prueba de Chi-Cuadrado: Canal vs Compro
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	9111,161 ^a	6	,000
Razón de verosimilitudes	9409,215	6	,000
N de casos válidos	89677		

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 479,28.

Elaborado Por: *D. Hernández*, Facultad de Ciencias Naturales y Matemáticas (FCNM).

Se evidencia estadísticamente en la prueba de chi-cuadrado de Pearson que existe relación entre las variables, es decir la utilización del cupo del crédito directo dependerá porque canal sea aprobado la cuenta.

3.2.3 Tabla de Contingencia: GrupoHijos vs VariableRespuesta

Tabla 3.06: Tabla de Contingencia: GrupoHijos vs Variable Y
 “Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Tabla de contingencia GrupoHijos * Yobjetivo

		Yobjetivo		Total	
		Cliente Malo	Cliente Bueno		
GrupoHijos	1,00	Recuento	44785	39362	84147
		% del total	49,9%	43,9%	93,8%
2,00		Recuento	2757	2773	5530
		% del total	3,1%	3,1%	6,2%
Total		Recuento	47542	42135	89677
		% del total	53,0%	47,0%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	23,616 ^a	1	,000		
Corrección por continuidad ^b	23,481	1	,000		
Razón de verosimilitudes	23,575	1	,000		
Estadístico exacto de Fisher				,000	,000
Asociación lineal por lineal	23,616	1	,000		
N de casos válidos	89677				

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2598,29.

b. Calculado sólo para una tabla de 2x2.

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

Se estima que a mayor números de hijos, mayor probabilidad que el cliente caiga morosidad, para evidenciar este caso se realiza la prueba de hipótesis, en donde se segmenta la población objetivo en dos grupos, n_1 = clientes aprobados con 0 y 1 hijo vs n_2 = clientes aprobados con 2 hijos en adelante. El contraste de hipótesis es:

$$H_0: p_1 = p_2;$$

vs.

$$H_1: p_1 \neq p_2;$$

En donde p es la probabilidad que el cliente se encuentre en mora. Los resultados que muestra el software SPSS son los siguientes.

El Promedio de hijos para $n_1 = 0.06$ y para $n_2 = 2.39$ el estadístico de la prueba de hipótesis no asume varianza iguales, es decir se rechaza la hipótesis nula, las probabilidades de que el cliente se encuentre en mora son diferente para ambos grupos. Realizando la tabla de contingencia para observar que en cuanto difiere la morosidad de cada grupo, los resultados son los siguientes: la probabilidad que un cliente caiga en mora con característica máximo 1 hijo es de 0.12 y la probabilidad que un cliente este en mora con característica mínimo 2 hijos es de 0.01, lo cual se rechaza la hipótesis inicial es decir si el cliente se encuentra en mora no dependerá del número de cargas familiares, las probabilidades no evidencia discriminación entre los grupos.

3.2.4 Tabla de Contingencia: GrupoEdad vs VariableRespuesta

Tabla 3.07: Tabla de Contingencia: GrupoEdad vs Variable Y
 “Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Tabla de contingencia GrupoEdad * Yobjetivo

		Yobjetivo		Total	
		Cliente Malo	Cliente Bueno		
GrupoEdad	1,00	Recuento	23361	24163	47524
		% del total	26,1%	26,9%	53,0%
	2,00	Recuento	24181	17972	42153
		% del total	27,0%	20,0%	47,0%
Total		Recuento	47542	42135	89677
		% del total	53,0%	47,0%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	604,286 ^a	1	,000		
Corrección por continuidad ^b	603,957	1	,000		
Razón de verosimilitudes	605,228	1	,000		
Estadístico exacto de Fisher				,000	,000
Asociación lineal por lineal	604,279	1	,000		
N de casos válidos	8967				
	7				

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 19805,71.

b. Calculado sólo para una tabla de 2x2.

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

Otra estimación se realizó con la partición de la población en la base a la edad de nacimiento, el objetivo es identificar que grupo de edad tiene mayor probabilidad de caer en mora, se dividió la población en dos grupos, $n_1 =$ clientes aprobados con edad menor a 35 años vs $n_2 =$ clientes aprobados con edad mayor o igual a 35 años. El contraste de hipótesis es:

$$H_0: p_1 = p_2;$$

vs.

$$H_1: p_1 \neq p_2;$$

En donde p es la probabilidad que el cliente caiga en mora. Los resultados que muestra el software SPSS son los siguientes.

El 47% de los clientes se encuentran en el grupo 2, de estos la mayoría tienen una mayor probabilidad de caer en mora, 27% clientes malos vs 20% clientes buenos. Del grupo de los clientes buenos, 47% de la población, el 27% son clientes con edad menores a 35 años, es decir los mejores clientes se encuentran en el grupo 1.

La prueba de chi-cuadrado muestra estadísticamente que las variables están altamente correlacionadas. En la tabla de contingencia se muestra que se rechaza la hipótesis nula, las probabilidades de caer en mora de cada grupo es diferente, 0.26 para el grupo 1 vs 0.27 para el grupo 2.

3.2.5 Tabla de Contingencia: Clasificación del Buro vs

Variable Respuesta

Tabla 3.08: Tabla de Contingencia: Clasificación del Buro vs Variable Y
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Tabla de contingencia CLASIFBR * Yobjetivo

		Yobjetivo		Total
		Cliente Malo	Cliente Bueno	
A	Recuento	11685	10747	22432
	% del total	13,0%	12,0%	25,0%
AA	Recuento	12523	10261	22784
	% del total	14,0%	11,4%	25,4%
CLASIFBR AAA	Recuento	16090	12619	28709
	% del total	17,9%	14,1%	32,0%
RE	Recuento	1059	1201	2260
	% del total	1,2%	1,3%	2,5%
RV	Recuento	6185	7307	13492
	% del total	6,9%	8,1%	15,0%
Total	Recuento	47542	42135	89677
	% del total	53,0%	47,0%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	461,343 ^a	4	,000
Razón de verosimilitudes	460,992	4	,000
N de casos válidos	89677		

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1061,87.

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

Se estima que la Calificación del Buro sea una variable fuertemente relacionada con la morosidad, por lo que se realiza el contraste hipótesis entre las dos variables, se agrupo la población de acuerdo a la calificación del buro, n_1 = clientes aprobados con calificación A, n_2 = clientes aprobados con calificación AA, n_3 = clientes aprobados con calificación AAA, n_4 = clientes aprobados con calificación Rechazado y n_5 = clientes aprobados con calificación Revisión. El contraste de hipótesis es:

$$H_0: p_1 = p_2 = p_3 = p_4 = p_5;$$

vs.

$$H_1: p_1 \neq p_2 \neq p_3 \neq p_4 \neq p_5;$$

En donde p es la probabilidad que el cliente caiga en mora. Los resultados que muestra el software SPSS son los siguientes.

El 47% de los clientes de la población se considera como clientes buenos de acuerdo a su comportamiento en 1 año, de los cuales el 14% tienen calificación AAA, 12% A y 11% AA, los mejores clientes son los que vienen con mejores calificaciones por parte del Buro. Del 53% de los clientes considerados como malos el 17% AAA y 14% AA, es decir que a pesar que el cliente se comporta bien en el Buro, dentro de la compañía tiene un comportamiento malo, una probabilidad alta de caer en mora. La prueba de chi-cuadrado muestra estadísticamente que las variables están altamente correlacionadas y que las probabilidades de caer en mora de cada muestra no son iguales.

3.2.6 Tabla de Contingencia: Canal vs Variable Respuesta

Tabla 3.09: Tabla de Contingencia: Canal vs Variable Y
 “Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Tabla de contingencia CANAL * Yobjetivo

		Yobjetivo		Total
		Cliente Malo	Cliente Bueno	
ATC	Recuento	1037	1434	2471
	% del total	1,2%	1,6%	2,8%
EXT	Recuento	26534	15444	41978
	% del total	29,6%	17,2%	46,8%
PCI	Recuento	5994	6646	12640
	% del total	6,7%	7,4%	14,1%
PT	Recuento	7588	8571	16159
	% del total	8,5%	9,6%	18,0%
PV	Recuento	5748	9122	14870
	% del total	6,4%	10,2%	16,6%
WEB	Recuento	641	918	1559
	% del total	0,7%	1,0%	1,7%
Total	Recuento	47542	42135	89677
	% del total	53,0%	47,0%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	3588,850 ^a	5	,000
Razón de verosimilitudes	3617,974	5	,000
N de casos válidos	89677		

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 732,50.

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

La variable Canal es un medio por donde se apertura la cuenta del cliente, es decir es un filtro en donde se puede detectar que Canal trae más cliente malos, por lo que se realiza la tabla de contingencia entre las dos variables, n_1 = clientes aprobados en el canal ATC, n_2 = clientes aprobados en el canal EXT, n_3 = clientes aprobados en el canal PCI, n_4 = clientes aprobados en el canal PT y n_5 = clientes aprobados en el canal WEB. El contraste de hipótesis es

$$H_0: p_1 = p_2 = p_3 = p_4 = p_5;$$

vs.

$$H_1: p_1 \neq p_2 \neq p_3 \neq p_4 \neq p_5;$$

En donde p es la probabilidad que el cliente caiga en mora. Los resultados que muestra el software SPSS son los siguientes.

De la población aprobada, el 47% se lo realiza por el Canal Externo (EXT), dentro de ellos cerca del 30% son clientes considerados como malos, es decir por este canal se están aprobando clientes con alta probabilidad de caer en mora, el 18% se aprueban clientes por el canal de Promoción Tienda (PT) de las cuales el 8.5% son clientes considerados como malos, y el 16% se aprueban por el canal Piso de Venta (PV) de las cuales el 6.4% son malos. La prueba de chi-cuadrado muestra estadísticamente que las variables están altamente correlacionadas y que las probabilidades de caer en mora de cada muestra no son iguales

3.4. Regresión Logística

3.4.1 Modelo Estadístico

El modelo de Regresión Logística Binomial es un modelo estadístico lineal generalizado que predice la probabilidad de una variable dependiente binomial en términos de unas variables de explicación. La variable dependiente Y será el comportamiento del cliente en un periodo de 12 meses a partir del mes de su apertura del crédito directo, en el que se definirán como “Cliente Bueno” a los clientes que utilizaron el su cupo de crédito y en 12 meses no cayeron en mora, es decir su edad máxima de mora es de 30 días, este grupo tomaran valores de 1. Se consideran “Clientes Malos” aquellos que en 12 meses cayeron en mora, es decir que en el periodo tuvieron una edad mínima de 60 días, este grupo tomaran valores de 0. Esto se puede apreciar en la siguiente tabla. Los resultados que muestra el software R-Studio son los siguientes.

Tabla 3.10: Codificación de la variable de Respuesta Y del modelo Regresión Logística “Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Codificación de la variable dependiente Y	
Valor Original	Valor Interno
<i>Cliente Bueno</i>	1
<i>Cliente Malo</i>	0

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

Las variables independientes o de explicación para el modelo de Regresión Logística que predecir si el cliente es Bueno o Malo se presenta en la siguiente tabla.

Tabla 3.11: Variables Independientes del Modelo de Regresión Logística “Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Covariables	Variables de Explicación
Genero	Femenino
	Masculino
Edad Nacimiento	Numérico
Estado Civil	Casado
	Soltero
	Unión Libre
	Divorciado
	Viudo
Región Domicilio	Quito
	Guayaquil
	Resto Costa
	Resto Sierra
Cargas Familiares	Numérico
Sueldo	Numérico
Tiene Tarjeta	1-SI
	0-NO
Puntaje	Numérico
Cupo	Numérico
Clasificación del Buro	A
	AA
	AAA
	Rechazo
	Revisión
Canal	Atención al Cliente
	Prom. Externa
	Prom. Cred. Inmd
	Prom. Tienda
	Piso de Venta
	Web

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM)

3.4.2 Coeficientes Estimados y Prueba de Wald

Para estimar los parámetros del modelo se utiliza el método de Máxima Verosimilitud y como los parámetros del modelo no es posible obtenerlos de forma explícita, se utilizó el método numérico de Newton-Raphson.

Tabla 3.12: Aplicación del Modelo Regresión Logística para el comportamiento del Cliente
Estimación de los Betas
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Variables		$\hat{\beta}_i$	Std. Error	Estadístico de Wald	Valor P
	Constante	0.38348	0.06566	5.840	5.21e-09
Genero	Masculino	-0.32533	0.01621	-20.059	< 2e-16
Edad Nacimiento	Cantidad	-0.00941	0.00084	-11.201	< 2e-16
Estado Civil	Divorciado	-0.00865	0.03850	-0.225	0.822189
	Soltero	0.04993	0.01808	2.761	0.005767
	Unión Libre	0.02001	0.03084	0.649	0.516402
	Viudo	0.20941	0.08059	2.598	0.009372
Región Domicilio	Quito	-0.38369	0.01834	-20.915	< 2e-16
	Resto Costa	-1.04859	0.02515	-41.689	< 2e-16
	Resto Sierra	-1.39478	0.03719	-37.503	< 2e-16
Sueldo	Cantidad	-0.15039	0.00265	-56.695	< 2e-16
Tarjeta	1-SI	-0.42950	0.02062	-20.819	< 2e-16
Puntaje	Cantidad	-0.00248	0.00020	-12.232	< 2e-16
Clasificación del Buro	AA	-0.22476	0.02187	-10.276	< 2e-16
	AAA	-0.34860	0.02154	-16.181	< 2e-16
	Rechazo	0.11244	0.05023	2.238	0.025192
	Revisión	0.14127	0.02564	5.509	3.62e-08
Canal	Prom. Externa	-0.73596	0.04694	-15.678	< 2e-16
	Prom. Cred. Inmd	-0.10591	0.04958	-2.136	0.032683
	Prom. Tienda	-0.12111	0.04860	-2.492	0.012706
	Piso de Venta	0.18012	0.04920	3.661	0.000252
	Web	0.32002	0.07378	4.338	1.44e-05
Cupo	Cantidad	0.45272	0.00442	102.384	< 2e-16

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

Tabla 3.13: Análisis de los Signos de los Coeficientes Betas Estimados
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Variables		Signo	Explicación
Genero	Femenino	Negativo	Ser Hombre disminuye la probabilidad que el cliente NO caiga en mora
	Masculino		
Estado Civil	Divorciado	Negativo	Estar Divorciado disminuye la probabilidad que el cliente NO caiga en mora. Estar Soltero, Unión Libre y Viudo aumentan la probabilidad de NO caer en mora
	Soltero	Positivo	
	Unión Libre	Positivo	
	Viudo	Positivo	
Región Domicilio	Quito	Negativo	Vivir en la Región de Quito, Resto Costa y Resto Sierra disminuye la probabilidad que el cliente NO caiga en mora.
	Resto Costa	Negativo	
	Resto Sierra	Negativo	
Tarjeta	1-SI	Negativo	Tener tarjeta externa disminuye la probabilidad que el cliente NO caiga en mora.
	0-NO		
Clasificación del Buro	AA	Negativo	Los clientes con calificación AA y AAA disminuyen la probabilidad de NO caer en mora, mientras que los clientes Rechazados y Revisión por el Buro aumentan la probabilidad de NO caer en mora.
	AAA	Negativo	
	Rechazo	Positivo	
	Revisión	Positivo	
Canal	Prom. Externa	Negativo	Las cuentas que son aprobadas por los canales: Prom. Externa, Cred. Inmd y Tienda disminuyen la probabilidad de NO caer en mora, mientras que las cuentas aprobadas por los canales Piso Venta y Web aumentan la probabilidad de NO caer en mora.
	Prom. Cred. Inmd	Negativo	
	Prom. Tienda	Negativo	
	Piso de Venta	Positivo	
	Web	Positivo	

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

En la Tabla 3.12 Las variables de referencia para la estimación de los parámetros β , para los factores están dadas de la siguiente manera; Género (Femenino), Estado Civil (Soltero), Región Domicilio (Guayaquil), Tiene Tarjeta (NO), Buro (A).

El “valor p” para los factores “Género, Edad, Región Domicilio, Sueldo, Tiene Tarjeta Externa, Puntaje y Cupo” es menor a 0,05 lo que significa que estas variables independientes ó de explicación son altamente significantes. (Véase Tabla 3.12, Valor p).

El modelo es utilizado para determinar la probabilidad de que un cliente NO caiga es mora, es decir que el cliente sea “Bueno” en términos de comportamiento crediticio, es decir, la probabilidad de que $y = 1$.

$$p(y = 1|x) = \frac{e^y}{1 - e^y} = \frac{1}{1 - e^{-y}}$$

Donde, la función de regresión es la transformación $\text{logit}(p)$, es decir:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \dots + \beta_p\chi_p$$

Donde β_0 es la constante y β_i es el i-ésimo coeficiente de la variables predictoras i.

Los coeficientes β_i asociados a cada una de las variables explicativas x_i representan el cambio en Odd ratio para un incremento de una unidad en x_i .

Con los betas estimados en el Tabla 3.13 el modelo se lo plantearía de la siguiente manera.

$$\begin{aligned}
y = & 0.38348 - 0,32533 \text{ Género Masculino}_1 - 0,00941 \text{ Edad} \\
& - 0,0086531 \text{ Estado Civil Divorciado} \\
& + 0,0499376 \text{ Estado Civil Soltero} \\
& + 0,0200163 \text{ Estado Civil Union Libre} \\
& + 0,2094115 \text{ Estado Civil Viudo} - 0,3836933 \text{ Región Quito} \\
& - 10,485939 \text{ Región Resto Costa} \\
& - 13,947891 \text{ Región Resto Sierra} - 0,1503947 \text{ Sueldo} \\
& - 0,4295023 \text{ Si Tiene Tarjeta Externa} - 0,0024893 \text{ Puntaje} \\
& - 0,2247656 \text{ Buro AA} - 0,3486071 \text{ Buro AAA} \\
& + 0,1124418 \text{ Buro Rechazado} + 0,1412776 \text{ Buro Revisión} \\
& - 0,7359609 \text{ Canal Prom. Externo} \\
& - 0,1059182 \text{ Canal Prom. Cred Inmd} \\
& - 0,1211152 \text{ Canal Piso Tienda} \\
& + 0,1801298 \text{ Canal Piso Venta} + 0,3200248 \text{ Canal Web} \\
& + 0,452727 \text{ Cupo}
\end{aligned}$$

El modelo indica que las cuentas que son aprobadas por los canales Piso de Venta y Web son más propensos a NO caer en mora, Estar Solteros, Unión Libre o Viudo y estar en Revisión en Buro aumenta la probabilidad de NO caer en mora.

El signo de los coeficientes $\widehat{\beta}_i$ de la función logística estimada se explica en la Tabla 3.13.

3.4.3 Resultados de los Razón de Propensión (Odd Ratio)

Tabla 3.14: Aplicación del Modelo Regresión Logística para el comportamiento del Cliente
Odds Ratio
“Diseño de un Modelo de Puntuación Crediticia para Empresas Comerciales”

Variables		Odds Ratio	I.C 95% para Exp()	
			Inferior	Superior
	Constante	1,4674	1,2903	1,6691
Genero	Masculino	0,7223	0,6997	0,7456
Edad Nacimiento		0,9906	0,9890	0,9923
Estado Civil	Divorciado	0,9914	0,9192	1,0690
	Soltero	1,0512	1,0146	1,0891
	Unión Libre	1,0202	0,9604	1,0838
	Viudo	1,2330	1,0528	1,4440
Región Domicilio	Quito	0,6813	0,6573	0,7063
	Resto Costa	0,3504	0,3336	0,3681
	Resto Sierra	0,2479	0,2304	0,2666
Sueldo		0,8604	0,8559	0,8648
Tarjeta	1-SI	0,6508	0,6250	0,6777
Puntaje		0,9975	0,9971	0,9979
Clasificación del Buro	AA	0,7987	0,7652	0,8337
	AAA	0,7057	0,6765	0,7361
	Rechazo	1,1190	1,0141	1,2348
	Revisión	1,1517	1,0953	1,2111
Canal	Prom. Externa	0,4790	0,4369	0,5252
	Prom. Cred. Inmd	0,8995	0,8161	0,9912
	Prom. Tienda	0,8859	0,8053	0,9744
	Piso de Venta	1,1974	1,0872	1,3185
	Web	1,3772	1,1919	1,5916
Cupo		1,5726	1,5591	1,5863

Elaborado Por: D. Hernández, Facultad de Ciencias Naturales y Matemáticas (FCNM).

La “Odds” son la razón de probabilidad de ocurrencia de un evento de interés, están definidos como un cociente entre el número de eventos y el número de “no” eventos. Las siguientes Odds fueron calculadas usando la expresión;

$$odd\ ratio = \frac{\frac{p_i}{1-p_i}}{\frac{p_{i+1}}{1-p_{i+1}}} = \exp(\beta_i)$$

Para cada covariable usada en el estudio los resultados se muestran en la Tabla 3.14, donde se muestra la variable de explicación, el estimador del coeficiente correspondiente a la variable y sus respectivos intervalos de confianza.

Los valores de los Exp ($\widehat{\beta}_i$) de la razón de propensión para cada $\widehat{\beta}_i$, entonces:

1. Un hombre es menos propenso de NO caer en mora que una mujer.
2. Un cliente que viva en Quito o Provincias es menos propenso de NO caer en mora que un cliente que viva en Guayaquil.
3. Tener tarjeta externa es menos propenso de NO caer en mora que un cliente que no tiene tarjeta externa.
4. Un Cliente que este Soltero o Unión Libre es 1.0 veces más propenso de NO caer en mora que un cliente Casado. Si el cliente está Divorciado es menos propenso de NO caer en mora si el cliente está Casado.
5. Tener calificación del Buro Rechazado o Revisión es 1.1 veces más propenso de NO caer en mora que un cliente con calificación A. Si el

cliente es AA o AAA es menos propenso de NO caer en mora que un cliente A.

6. Un crédito directo que es aprobado por el canal Piso de Venta es 1.1 más propenso de NO caer en mora que un crédito aprobado por ATC, si el canal de apertura es la WEB es 1.3 veces más propenso de NO caer en mora que ATC. Si el cliente aprueba su crédito directo por los canales Prom. Externa, Crédito Inmediato o Tienda es menos propenso de NO caer en mora que si el cliente lo aprueba por el canal ATC.

CONCLUSIONES Y RECOMENDACIONES

Como parte final de esta investigación se presentan las conclusiones y recomendaciones obtenidas luego de desarrollar el estudio, las mismas que tienen el objeto de ser utilizadas para operaciones inmediatas o como base de estudios científicos futuros.

Conclusiones

Las conclusiones de este proyecto son las siguientes:

1. Las mujeres son las más preferidas para obtener un crédito directo en la empresa de RETAIL.
2. Los clientes solteros y casados son los mayoritarios en el crédito directo.
3. De los clientes aprobados el 47% son de Quito, de las cuales el 20% son Masculino y su promedio de crédito directo es de 3500 cuentas aprobadas al mes. El 31% son de la ciudad de Guayaquil, el 17% son Femeninos y su promedio de cuenta 2,322 cuentas aprobadas al mes.
4. El número promedio de hijos por cliente es 1.
5. De los clientes en el 2013, el 32% son cuentas AAA, el 25% son clientes AA, y el 3% de clientes son Rechazados por el Buro que son aprobados por excepción.

6. El cupo promedio otorgado a su crédito directo es de \$572 con una desviación estándar \$321. El cupo mínimo es de \$45 y el cupo máximo es de \$5000.
7. Tener una tarjeta externa dependerá mucho de la calificación del Buro.
8. El consumo del cupo otorgado en el crédito directo depende mucho del canal por medio donde lo aprueben. Los que se aprueban por el canal Promoción Externa tienen mayor probabilidad de utilizar más su cupo.
9. La tasa de morosidad no depende del número de hijos del cliente. Son variables independientes.
10. Tener mayor edad aumenta la probabilidad que el cliente caiga en mora.
11. Un cliente AAA tiene mayor probabilidad de ser un cliente Bueno.
12. El canal que más clientes malos aprueba es Promoción Externa.
13. Las mujeres son más propensas de caer en mora que un hombre.
14. Los clientes de Guayaquil son más morosos que los de Quito.
15. Tener tarjeta externa es más probable des ser moroso.

Recomendaciones

1. Se recomienda que la base esté depurada para los análisis estadísticos correspondientes.
2. Tomar en cuenta las conclusiones para evitar en el futuro un aumento en la tasa de morosidad.
3. Se recomienda a la empresa disminuir la tasa de morosidad sin disminuir el promedio de cuentas aprobadas en el año.
4. En los datos de entrada, se recomienda escoger una variable relevante para el objetivo de regresión logística, como por ejemplo; tiempo (años) en el último trabajo, utilización de cupo de las tarjetas externas, Proviene de una educación fiscal o particular, entre otros.
5. Los perfiles de los clientes “Buenos” y “Malos” son un resultado a priori para unos próximos análisis a lo referente.
6. Se recomienda hacer un análisis exhaustivo inferencial entre las variables más prioritarias al objetivo, ejemplo; un perfil más ajustado de los clientes morosos en un periodo de tiempo.

BIBLIOGRAFÍA

1. **RUIZ H. (2013)** “Diseño de un modelo matemático para optimizar la gestión de un sistema de cobranza en una empresa comercial”, ESPOL – GUAYAQUIL – ECUADOR,
2. **BANCO CENTRAL DEL ECUADOR (BCE) (2013)**
<http://www.bce.fin.ec/index.php/component/k2/item/755>
3. **WALTER SPURRIER** Analista Económico graduado de un Masterado de Ciencias Políticas en la Universidad de California en los Ángeles (UCLA). Editorialista de Diario El Comercio y El Universo. Actualmente es Director de Análisis Semanal, Presidente de Grupo Spurrier.
4. **PEREZ F, FERNANDEZ H (2007)** “LAS REDES NEURONALES Y LA EVALUACIÓN DEL RIESGO DE CRÉDITO” - Revista Ingenierías Universidad de Medellín, volumen 6, No. 10, pp. 77-91 - ISSN 1692-3324 - Enero-junio de 2007/166p.
5. **TAMARA A, ARISTIZABAL R, VELASQUEZ E (2012)** “Matrices de transición en el análisis del riesgo crediticio como elemento

fundamental en el cálculo de la pérdida esperada en una institución financiera colombiana”

6. **CHEN W (2012)** “Evaluación del riesgo de crédito mediante la técnica de minería de datos híbrido” - Systems Engineering Procedia 3 (197 – 200).

7. **ESPIN-GARCIA O, RODRIGUEZ-CABALLERO C (2013)** “Metodología para un Scoring de clientes sin referencias crediticias”. Cuadernos de Economía, 32(59).

8. **A. Fuentes, N. Rivera, R. Pinos (2011)** “Software Estadístico para Regresión. El caso de Regresión Logística y Regresión Poisson” ESPOL-GUAYAQUIL.

9. **CUENTAS REAPERTURADAS** - son cuentas cerradas en su momento por algún motivo que luego de un tiempo considerado, 3 meses, puede aplicar a reabrir su cuenta, siempre y cuando no esté en mora y cumpla los requisitos.

Anexo: Código de Programación en R

#Lectura de los datos

```
data=read.delim("D:/BaseCredito.txt")
```

#Estructura de los datos

```
str(data)
```

#Creación del Modelo Binomial de Regresión Logística

```
m1=glm(y ~ sex + edad + ec + region + sueldo + tarjeta + puntaje + buro + canal +  
cupo, data = BaseCredito, family = binomial)
```

#Resumen del Modelo

```
summary(m1)
```

#Pureba Chisq

```
drop1(m1,test="Chisq")
```

#Odds Ratio

```
exp(cbind(OR = coef(m1), confint(m1)))
```