



**UNIVERSIDAD
DEL PACÍFICO**

**Escuela de
Postgrado**

**“APLICACIÓN DEL MODELO DEL CREDIT SCORING A
LAS TARJETAS DE CREDITO EN LIMA
METROPOLITANA: CASO DE UN BANCO PERUANO”**

**Trabajo de Investigación presentado
Para optar al Grado Académico de
Magíster en Finanzas**

**Presentado por
Víctor Antonio David Tineo
Edwin Chuquipul Ruiz**

Asesor: Oscar Antonio Basso Winffel
[0000-0003-1767-2105](tel:0000-0003-1767-2105)

Lima, marzo 2008

INTRODUCCION

La situación actual de nuestro país muestra que existe optimismo por parte del ciudadano común, así como por muchas instituciones financieras, esto lo vemos reflejado en el incremento porcentual que se viene dando en los créditos de consumo, hasta octubre del presente se ha producido un crecimiento anual del 30% y se espera que esta tendencia continúe hasta diciembre (campaña navideña). Con respecto a las tarjetas de crédito, se ha observado que la oferta de los plásticos ha venido incrementándose. Ya resulta común recibir constantemente llamadas de distintas instituciones financieras ofreciendo sus tarjetas, se ofrece “líneas aprobadas”, aprobaciones en 15 minutos, están las famosas “T x T”. La pregunta que debería hacerse es si estas instituciones cuentan con herramientas expertas que le permitan hacer estos ofrecimientos, o es simplemente el ímpetu comercial de querer ganar participación de mercado, a costa de incurrir en mayor riesgo Crediticio.

En el presente trabajo se busca enfocarlo desde el punto de vista de un banco del país, es decir, se busca elaborar una herramienta que sea útil en su proceso de evaluación de las solicitudes de las tarjetas de crédito.

El objetivo que se plantea tomando en consideración el contexto actual que vive el país en lo que respecta a los créditos de consumo y en particular a las tarjetas de crédito es: “Contribuir con una herramienta que pueda ayudar en la toma de decisión, no que tome la decisión, al momento de otorgar una tarjeta de crédito”.

Para ello, se ha dividido el presente trabajo en las siguientes partes: en una primera parte se busca esbozar cual es la situación actual de las tarjetas de crédito en el Perú. Se hace un análisis de la información estadística que se ha obtenido, en el anexo 1 se incorpora información sobre los créditos de

consumo. En una segunda parte se esboza el marco teórico que sirve como base para el desarrollo del modelo. Principalmente se centra en la metodología a aplicar para obtener el modelo, en el anexo 2 se dan más alcances acerca de teoría acerca del Scoring. En una tercera parte, se realiza la aplicación práctica a nuestra realidad a través del desarrollo de un modelo de Scoring personalizado con los datos obtenidos de un banco del país, se sigue todos los pasos requeridos para la construcción del modelo y se realiza las respectivas pruebas de validación y un Backtesting. El trabajo concluye con las respectivas conclusiones y recomendaciones.

Se ha considerado que la importancia del presente trabajo de investigación radica en la posibilidad de elaborar una herramienta que permita a las instituciones financieras realizar una mejor selección de sus clientes en lo que respecta a las tarjetas de crédito, se quiere dejar en claro que estos modelos no sustituyen al analista de riesgos, por el contrario, es una muy importante herramienta de apoyo para que esta junto a su expertise, pueda tomar una decisión más fundamentada, lo que debería tener un impacto positivo en la cartera atrasada, por consiguiente una mayor rentabilidad para el negocio.

Las limitaciones del presente trabajo son el acceso a los datos para correr el modelo y la no incorporación del modelo de capacidad de pago. En lo primero, dado el alto grado de confidencialidad que muestran las instituciones financieras con su base de datos, no se ha podido acceder a todo el universo de información con la que cuentan para tarjetas de crédito, esto definitivamente influye en la predictividad del modelo. Con respecto a lo segundo, el modulo de capacidad de pago involucra la estimación de monto disponible de cada persona luego de restar los gastos implícitos y explícitos. La dificultad de este modelo está en la estimación de los gastos implícitos, lo que significa realizar un estudio socioeconómico, que no es materia de este trabajo de investigación. Los analistas pueden contar con una hoja de Excel para estos cálculos.

CAPITULO I: LAS TARJETAS DE CRÉDITO EN EL PERU

Actualmente, las tarjetas de crédito, más que un símbolo de alta categoría económica y social, han pasado a convertirse en una prenda común. Las tarjetas de crédito permiten al usuario adquirir todo tipo de bienes de consumo masivo. Este negocio se basa, en que, al tratarse de un medio de pago, facilita al tarjeta habiente la adquisición de bienes y servicios, así como la disponibilidad de efectivo. De esta manera el cliente mantiene una línea de crédito aprobada que se encuentra disponible en momentos de falta de liquidez o en aquellos en los que por algún motivo se desee financiar una compra.

En el presente capítulo se revisan los principales indicadores relacionados a las tarjetas de crédito en el Perú, habiendo dado un valor agregado a partir del análisis de la información estadística obtenida.

No es posible hablar de tarjetas de crédito, sin mencionar a los créditos de consumo, en consecuencia, en el anexo 1 se ha recopilado información acerca de Los Créditos de Consumo en el Perú.

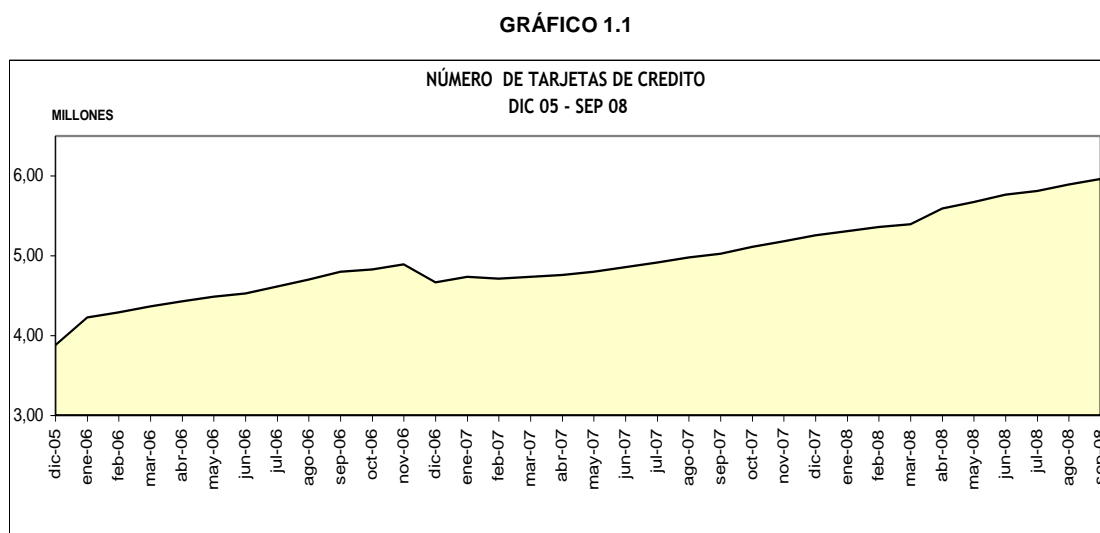
1.1 Tarjetas de Crédito¹

La Tarjeta de Crédito es un medio de pago personal e intransferible, que permite a su titular y, si fuera el caso, a los adicionales, adquirir bienes y/o servicios en los establecimientos y puntos de venta afiliados a la marca emisora, así como disponer de efectivo en las oficinas de los bancos y Cajeros Automáticos en el ámbito nacional o internacional.

1.2 Evolución del Número de Tarjetas de Crédito en el Perú

¹ Norma 18.10.017 Sobre Tarjetas de Crédito BBVA Banco Continental, Abril 2001

En el gráfico 1.1 se muestra como ha ido evolucionando el número de tarjetas de crédito



Fuente: Asbanc
Elaboración Propia

Ha habido un incremento importante en el número de tarjetas a partir del año 2008, a partir de abril 2008 se ha incrementado en 18% en promedio el número de plásticos en relación al mismo mes del año anterior. Esto guarda relación con el comportamiento dinámico que se ha observado en el mercado de tarjetas de crédito en el país, en donde la mayoría de las entidades de financieras han empezado a ofrecer tarjetas de crédito bajo la modalidad de TXT, con la simple fotocopia de otra tarjeta de crédito (como único sustento de ingreso), o en muchos casos sin presentar ningún documento solo con el llenado del formulario de solicitud.

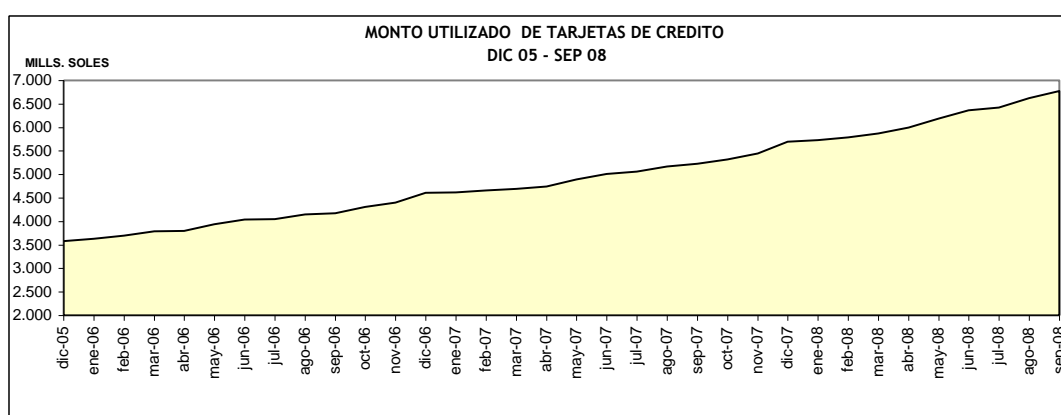
También se ha venido observando que los tiempos de respuesta de aprobación han venido disminuyendo (existen entidades bancarias que aprueban una solicitud en 15 minutos), lo cual resulta positivo en términos de eficiencia siempre y cuando no se sacrifique la calidad de riesgo. Estos son los motivos que consideramos que pueden explicar este incremento importante en tarjetas

de crédito a pesar de las medidas que los organismos reguladores han venido tomando para ir frenando los créditos de consumo.

1.3 Evolución del Monto Utilizado de Las Tarjetas de Crédito en el Perú

Otro aspecto importante a observar cuando se analizan tarjetas de crédito, es la evolución del monto utilizado de las tarjetas de crédito, los cuales se observan en el siguiente gráfico 1.2

GRÁFICO 1.2



Fuente: Asbanc
Elaboración Propia

Habiendo observado el crecimiento importante del número de tarjetas en el punto anterior, es lógico observar un crecimiento también en el monto utilizado de las tarjetas de crédito, a continuación mostramos un cuadro de crecimiento anual de los últimos 5 meses en el monto utilizado de las tarjetas de crédito.

CUADRO 1.1

Consumo de Tarjetas de Crédito
Crecimiento Anual (%)

may-08		26,42%
jun-08		27,00%
jul-08		27,10%
ago-08		28,29%
Set-08	↓	29,56%

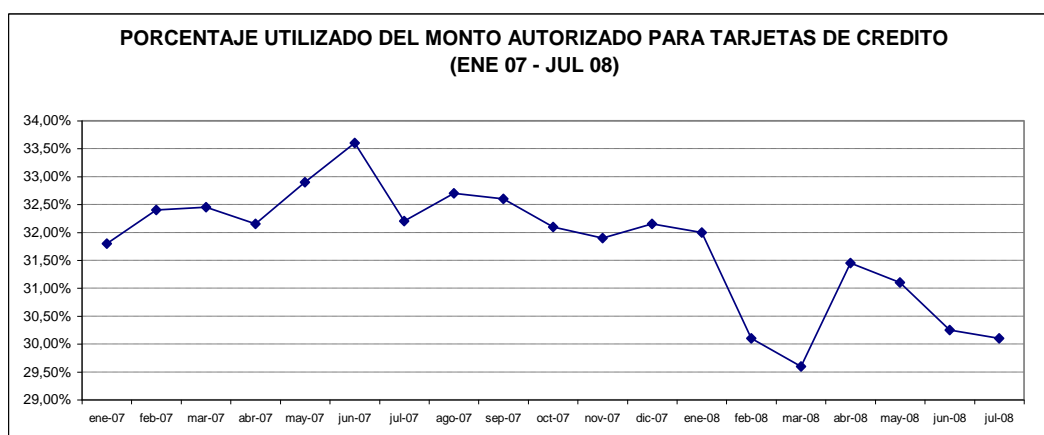
Fuente: Asbanc
Elaboración Propia

Observamos que el consumo de tarjetas de crédito va en claro aumento, mes a mes el porcentaje del consumo de tarjetas comparado con el año anterior está

en aumento. Esto puede resultar preocupante en la medida en que la gente está usando más sus tarjetas para cubrir sus déficit personales (gastos personales mayores a ingresos personales), mientras que si este crecimiento es reflejo de un mayor ingreso por parte de las personas (generalmente un mayor ingreso, involucra un mayor gasto que puede ser financiado con tarjetas de crédito), esto sería razonable.

Esta duda la podríamos ir despejando con el siguiente gráfico 1.3 en donde se observa el monto utilizado de las líneas de crédito respecto al total de líneas otorgadas.

GRAFICO 1.3



Fuente: Asbanc
Elaboración: Propia

En este gráfico podemos observar que el porcentaje utilizado de la línea autorizada a disminuido llegando a estar en 30%, con esto se puede concluir que el cada vez mayor incremento del monto utilizado de las tarjetas de crédito no responden a un problema de déficit de ingresos personales sino a que los bancos y financieras han ido otorgando nuevas tarjetas y aumentos de líneas a sus clientes.

Esto último estaría justificando el porque La S.B.S. en Agosto del presenta año publica una norma de sobre endeudamiento para Personas Naturales. Esta norma lo que busca es prevenir, es decir, ante un posible shock negativo en la

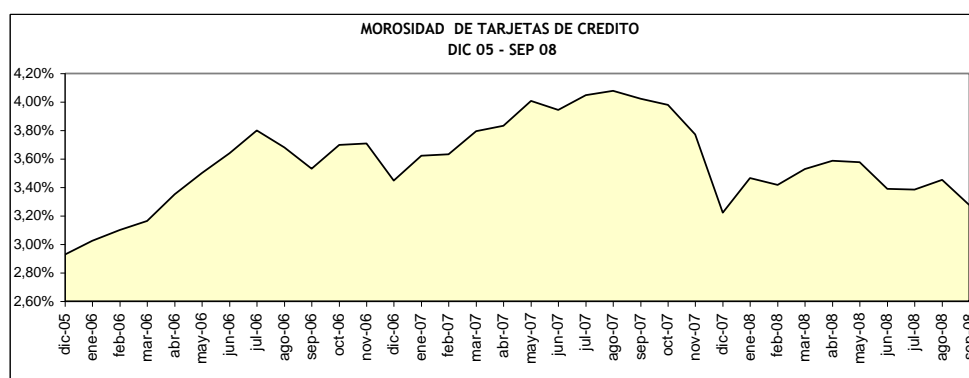
economía, no se vean afectados los bancos con una elevada morosidad, producto del impago de los créditos de consumo, entre los que destaca las tarjetas de crédito.

1.4 Morosidad de las Tarjetas de Crédito

Con respecto a la morosidad que se ha venido presentado en el sistema financiero, y específicamente las tarjetas de crédito, se puede ver un comportamiento positivo. En el gráfico 1.4 vemos un claro descenso de la morosidad, sobretodo en el último año. Se considera que existe correlación entre el desempeño de la economía de un país y la morosidad del sistema financiero, hay causalidad entre el PBI y el empleo, con esto se dice que un país que crece constantemente verá disminuir su nivel de desempleo, a su vez la variable empleo afecta directamente a los créditos de consumo, por una razón simple: mientras las personas cuenten con trabajo pueden cumplir con todas sus obligaciones financieras. Nuestra economía a mostrado un constante crecimiento a partir del 2002.

Otra variable que se considera también importante es el PBI per-capita, los ingresos de los ciudadanos peruanos han ido subiendo en los últimos años. Todo esto creemos a influenciado en la baja de la morosidad del sistema financiero.

GRAFICO 1.4



Fuente: Asbanc
Elaboración Propia

Adicionalmente a esto se puede decir también que los bancos han estado evaluando mejor a sus clientes al momento de darles crédito, esto aunado a que han tenido más instrumentos para tomar decisiones y una adecuada gestión de cobranza. Por ejemplo, el tener acceso a las centrales de riesgo, a sido decisivo para que los bancos puedan tener mayor conocimiento de las personas a las cuales les van a prestar plata.

1.5 Sobre endeudamiento²

En Agosto del 2008 La SBS publicó la resolución N° 6941-2008, que justamente aborda el tema de la administración del riesgo de sobre endeudamiento de deudores minoristas.

Lo que este reglamento busca es que las empresas adopten un sistema de administración del riesgo de sobreendeudamiento que permita reducir dicho riesgo antes y después del otorgamiento, que efectúe un seguimiento permanente de la cartera con el objeto de identificar a los deudores sobre endeudados, y que incluya la evaluación periódica de los mecanismos de control utilizados, así como de las acciones correctivas o mejoras requeridas, según sea el caso.

El sobreendeudamiento no es abordado en el desarrollo del modelo del presente trabajo de investigación, ya que por si solo significa un “nuevo modelo”, que se alimenta principalmente de la información obtenida de las centrales de riesgos. Las principales limitaciones que encontramos para poder incorporarlo en nuestro modelo de Scoring, es el costo que involucraría obtener la central de riesgos de cada cliente de nuestra base de datos (el banco que nos proporcionó la data no accedió a darnos esta información) y también es el tiempo para su desarrollo.

² Resolución S.B.S. N° 6941 – 2008 Reglamento de Sobre endeudamiento, Agosto 2008

CAPITULO II: MARCO TEÓRICO

Los modelos de Credit Scoring son algoritmos que de manera automática evalúan el riesgo de crédito de un solicitante de financiamiento o de alguien que ya es cliente de la entidad. Tiene una dimensión individual, ya que se enfoca en el riesgo de incumplimiento del individuo o empresa, independientemente de lo que ocurra con el resto de la cartera de préstamos.

El Scoring es una combinación de ciencia y arte. Las técnicas que se utilizan para calcularlo, en general ayudan a construir una scorecard que sea estadísticamente válido pero sin ningún sentido de los negocios. El arte se encuentra en comprender las necesidades del negocio y ajustar la información que se posee para ser considerada en el scorecard, de manera que todavía sea estadísticamente válida pero tenga la noción del negocio.

2.1 Desarrollo de un Scorecard

2.1.1 Scorecard (cuadro de mando integral)³

Un cuadro de mando integral (scorecard) es una fórmula para la asignación de puntos a las variables independientes con el fin de obtener un valor numérico que refleje qué tan probable es que un prestatario, en relación a otras personas, experimente un determinado evento o realice una determinada acción. Por ejemplo un scorecard puede decirnos si un solicitante en particular es más o menos probable que caiga en default en relación con otro solicitante.

La mayoría de tableros se construyen con la estimación de un modelo de regresión. Los modelos de regresión examinan la manera en que una variable particular (la variable resultado) se explica por otra variable o más

³ Mays, Elizabeth. Credit Scoring For Risk Managers, The Handbook for Lender. Thomson. 2004

generalmente, por un conjunto de otras variables. Los resultados obtenidos de un modelo de regresión es un conjunto de factores llamados coeficientes de regresión. Cada uno de estos se puede interpretar como la correlación entre la variable resultado que estamos tratando de predecir y de la variable explicativa o la característica. Los pesos de los puntajes para el scorecard se calculan utilizando una simple transformación matemática de los coeficientes de la regresión

2.1.2 Determinación de la performance de la ventana⁴

Una serie de decisiones importantes deben hacerse sobre el diseño de un Scoring antes de iniciar cualquier trabajo en él. El período resultado es el período de tiempo durante el cual se observa los préstamos en la muestra para clasificarlos como buenos o malos. A menudo, la elección del período de resultados se ve limitada por las limitaciones que hay con los datos. Aunque los prestamistas han llegado a comprender en los últimos años que la buena calidad de los datos es un gran activo empresarial, a veces data de calidad se han mantenido sólo reservado para los préstamos en el pasado reciente, digamos el último año o dos. Esto, por supuesto, limita las opciones para definir el período de resultados.

Según el autor Naeem Siddigi indica en su libro, algunos desarrolladores de Scoring prefieren utilizar el más largo período de resultados disponibles para cada préstamo, utilizando toda la información disponible para cada préstamo a pesar de que algunos préstamos se puede observar desde hace varios años y otros tal vez sólo para uno. Evidentemente, los préstamos de más edad tienen una mayor probabilidad de ir mal porque se han observado durante un período más largo.

⁴ Siddigi, Naeem. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. SAS Institute Inc. 2006

2.1.3 Determinación de la definición de buenos y malos⁵

La elección de la definición de malos depende que en el modelo que se va a usar se cuente con la cantidad de malos préstamos disponibles para la muestra de desarrollo y también de validación. Esto debe de ir de la mano con lo que la empresa interesada en desarrollar el Scoring considera un mal evento.

Según la autora Elizabeth Mays se cuentan con tres categorías para la variable resultado: los buenos, malos, e indeterminados. Indeterminados son los préstamos que tienen una performance peor que un buen préstamo, pero mejor que un mal préstamo.

Según indica la autora, la mayoría de desarrolladores de Scoring buscan eliminar los indeterminados de la muestra con la esperanza de que la eliminación pueda producir una puntuación que mejor pueda distinguir entre buenos y malos.

Sin embargo, también se menciona que a otros no le parece útil esta práctica. Su razonamiento es que todos los solicitantes deben ser considerados, no sólo aquellos cuyas características coincidan con los prestatarios cuyos préstamos se definen como buenos o malos.

2.1.4 Diseño de la muestra⁶

Es una práctica usual crear tanto una muestra de desarrollo como una muestra de validación. Como su nombre lo indica, la muestra de desarrollo se utiliza para construir los scorecard, mientras que la muestra de validación se utiliza para comprobar la exactitud del score basado en un conjunto de préstamos

⁵ Mays, Elizabeth. Credit Scoring For Risk Managers, The Handbook for Lender. Thomson. 2004

⁶ Anderson, Raymond. The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation. Oxford University Press Inc. 2007

que no tenga ninguna influencia sobre la generación de los pesos en los puntos.

Idealmente, la muestra de validación debe contener a un conjunto de préstamos que se originó durante un período de tiempo distinto que la de la muestra de desarrollo. La intención es asegurar que el resultado es robusto a través de diferentes períodos de tiempo. Si no se dispone de préstamos fuera de la muestra de desarrollo, la muestra de validación se obtendrá seleccionando al azar préstamos que formen parte de la muestra de desarrollo, es decir, préstamos que se originaron durante el mismo período de tiempo como los de la muestra de desarrollo. Lo que se busca es que tanto las muestras de desarrollo y validación contengan préstamos que son representativos de los tipos de préstamos en los que se utilizará el Scoring.

Según indica el autor Raymond Anderson, normalmente, el número de préstamos definidos como malos es el factor determinante. Asimismo indica Anderson que a pesar de que algunos desarrolladores de scorecard desarrollan el modelo con tan sólo 300, la mayoría se sienten más cómodos con al menos 1000 préstamos malos en la muestra de desarrollo, aunque unos pocos miles serían ideales. Para la muestra de validación, se debería contar con 300 préstamos malos, pero siempre más, es preferido.

A veces no hay suficientes créditos disponibles para desarrollar el scorecard y conservar una muestra para la validación. Existen algunas técnicas encaminadas a ayudar en la construcción de modelos para estas situaciones: jackknifing y bootstrapping.

2.1.5 Elección de características

Es muy importante para el desarrollador que deba tener un buen conocimiento de la empresa para el scorecard que está construyendo. Caso contrario, puede cometer errores en el trabajo con los datos y la interpretación de los resultados de los test que se vayan a realizar.

De acuerdo a la bibliografía examinada se recomienda que, si el scorecard va a ser construido por un tercero, es importante seleccionar un proveedor que tenga experiencia con el producto para el que se está generando el Scoring. Un miembro del personal interno con una excelente comprensión de los datos (preferentemente, con conocimientos de los métodos estadísticos) debe ser asignado a trabajar en estrecha colaboración con el proveedor para estar alerta para cualquier cosa que no tenga sentido, ya sea en los propios datos o en los resultados del modelo.

A veces la limitación de datos hace que las variables sean estadísticamente insignificantes. Esto a menudo ocurre cuando hay muy pocos préstamos con una característica particular para obtener un coeficiente significativo para esta característica. Este tipo de variable debe ser eliminada de las nuevas versiones del scorecard.

2.2 Análisis y Reducción de Variables⁷

El análisis de variable investiga la relación entre las variables que queremos poner a prueba para que mida la capacidad de predicción de la variable resultado. La reducción de variables reduce un gran número de variables que son candidatos para su inclusión en el scorecard a un número menor que puede ser más fácilmente analizado a partir de la estadística multivariante. No

⁷ Mays, Elizabeth. Credit Scoring For Risk Managers, The Handbook for Lender. Thomson. 2004

sólo se busca un menor número de variables, lo que se quiere es mantener sólo aquellas que son más predictivas de la variable resultado.

Según la autora Elizabeth Mays un scorecard debería contener entre 8 a 15 variables predictoras. Sin embargo, también indica que a menudo se debe probar decenas, incluso cientos de posibles variables para determinar cuáles son más predictivas.

2.2.1 Análisis de variable

Los tres estadísticos que se utilizan comúnmente para evaluar la relación entre los candidatos a variables predictoras y variable resultado son: estadístico chi-cuadrado, el estadístico de Spearman y el valor de la información.

2.2.1.1 El estadístico de Chi-cuadrado: El estadístico chi-cuadrado es un test que evalúa la asociación lineal entre una variable y la frecuencia con que los préstamos tienen un mal resultado en lugar de un buen resultado. Usualmente se recomienda que las variables con un valor de probabilidad (p) mayor a 0.5 sean eliminadas, por lo tanto el conjunto variables se reduciría mediante la exclusión de tales variables del análisis multivariante que seguiría a continuación. Se menciona que fijar un criterio diferente es siempre una opción.

2.2.1.2 Estadístico de correlación de Spearman: El estadístico de correlación de Spearman mide la correlación entre el ranking de la variable predictiva y la variable resultado. Es decir, en lugar de mirar a la asociación entre los valores reales predictivos y las variable resultado binaria, en este cálculo el rango asignado a cada valor de la variable predictiva reemplaza el valor real.

Una de las ventajas de la correlación de Spearman es que, a diferencia de los estadísticos chi-cuadrado y del valor de la información, es que muestra la dirección de la relación de una variable de predicción con la variable de resultado. Se debe evaluar si el efecto de cada variable sobre el resultado es

coherente con lo que se piensa acerca de cómo esa variable debe de afectar a los resultados.

2.2.1.3 Valores de Información: Los valores de información para las variables predictivas debe fluctuar entre 0 y 3. Los altos valores indican una fuerte relación con la variable resultado, indica que una variable tiene una gran cantidad de "información" para ayudar a predecir variable resultado, Lo que si hay que tomar en cuenta es que este estadístico no muestra la relación, ni la dirección de esta relación entre la variable de predicción y la variable resultado.

2.2.2 Reducción de variables

Después de analizar todas las posibles variables y eliminar aquellos que no tienen capacidad predictiva (con excepciones que dependen de la opinión del experto), Menciona la autora Mays que se puede reducir más variables que presenten información redundante. La importancia de esto, según indica, es porque las variables redundantes incluidas en el análisis multivariado pueden desestabilizar las estimaciones de los parámetros, confundir la interpretación de los coeficientes o aumentar el tiempo de cálculo del modelo.

Una vez que los grupos se especifican, es decir se han reducido variables, lo que sigue es seleccionar una sola variable para que represente a ese grupo. La variable seleccionada para representar al grupo deberá ser la más predictiva entre todas de la variable resultado. Esa es la razón por la que el análisis y reducción de variables deben ir de la mano, según indica la autora.

2.3 Métodos Estadísticos Empleados en la Construcción de Credit Scoring

2.3.1 Análisis discriminante⁸

El proceso de concesión de un crédito consiste en la elección entre dos acciones, una que trata de conceder un nuevo crédito y otra que implica la no concesión del crédito. El Credit Scoring trata de ayudar en la toma de esta decisión, buscando cual debería ser la mejor regla a aplicar a partir de una muestra de concesiones de crédito previas. La ventaja de esta es el conocimiento previo de la evolución de dichos préstamos.

Si hay solamente dos acciones posibles, aceptar o rechazar, entonces no hay ninguna ventaja en clasificar el desarrollo de estas concesiones de crédito conocidas en más de dos clases distintas de las de buenos y malos. Se consideran a un cliente bueno, según indica Thomas, cuando el desarrollo de la misma es aceptable para la entidad prestamista, y es malo cuando ante el conocimiento de la evolución posterior al préstamo la entidad desearía haberla rechazado.

En el anexo 2 se explica más detalladamente como se utiliza el análisis discriminante para clasificar una población en dos grupos (en nuestro caso, morosos y no morosos). También se está detallando otras dos formas de análisis discriminante: Separación en dos grupos y regresión lineal.

2.3.2 Regresión Logística⁹

Este método fue empleado por Wiginton (1980) en uno de los primeros resultados publicados de Credit Scoring. En la regresión logística, se toma el logaritmo de la probabilidad ampliada como una combinación lineal de las características de los clientes.

$$\ln(p_i / 1-p_i) = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_p x_p = \mathbf{w}'\mathbf{x} \quad (2.1)$$

⁸ Thomas, Lyn, David Edelman y Jonathan Crook. Credit Scoring and Its Applications. SIAM. 2002

⁹ Cabrer, Beranardi, Amparo Sancho y Guadalupe Serrano. Microeconometría y Decisión. Ediciones Pirámide. 2001

Si $p_i / 1-p_i$ toma valores entre 0 e ∞ , entonces, $\ln(p_i / 1-p_i)$ toma valores entre $-\infty$ y $+\infty$. Tomando exponenciales en ambos lados de (2.26) obtenemos la ecuación:

$$p_i = e^{w'x} / 1 + e^{w'x} \quad (2.2)$$

La única dificultad, comparada con la regresión ordinaria es que no es posible utilizar el método de los mínimos cuadrados ordinarios para calcular los coeficientes. Se utiliza el método de la máxima verosimilitud para estimar estos coeficientes y se emplea Newton-Raphson para resolver las ecuaciones que aparecen. Con el poder de los ordenadores modernos esto no es un problema, sobre todo para las grandes muestras que normalmente se utilizan cuando se construyen Credit Scoring.

Uno de los resultados sorprendentes que se obtienen es que, teóricamente, la regresión logística es mejor para aplicarla a las muchas clases de distribuciones que las resoluciones de clasificación de la regresión lineal. Y, sin embargo, cuando se realizan comparaciones de los desarrollos de Scoring utilizando los dos métodos con el mismo conjunto de datos, aparecen unas diferencias muy pequeñas en sus clasificaciones. En términos generales, la diferencia entre los métodos es, que la regresión lineal trata de ajustar la probabilidad p de impago como una combinación lineal de los atributos mientras la regresión logística trata de ajustar el $\ln(p/(1-p))$ como una combinación lineal de los atributos.

2.3.3 Otros métodos de regresión no lineal

Otras dos funciones no lineales se han utilizado para construir un Credit Scoring. La primera es la función probit, que fue usada para un Credit Scoring por Grablowsky y Talley (1981). En el análisis probit, si $N(x)$, representa la función de una distribución normal, esto es,

$$N(x) = 1/\sqrt{2\pi} \int_{-\infty}^x e^{-y^2/2} dy$$

Entonces la clave consistirá en calcular $N^{-1}(p_i)$ como una función lineal de las características de la aplicación, así consideraremos

$$N^{-1}(p_i) = \mathbf{w}'\mathbf{x}_i = w_0 + w_1 X_{1i} + w_2 X_{2i} + \dots + w_p X_{pi} \quad (2.3)$$

De nuevo p toma valores entre 0 y 1, y $N^{-1}(p_i)$ toma valores entre $-\infty$ y $+\infty$ y esto permite a la función lineal variar sobre todos los valores. Se podría pensar sobre el método probit como que la variable

$$W = w_0 + w_1 X_1 + w_2 X_2 + \dots + w_p X_p,$$

es una medida de la bondad de una aplicación, si tratamos de predecir si una nueva aplicación es buena o mala esto va a depender de si el valor de W es mayor o menor que una cierta barrera con un valor C . Como en el caso de la regresión logística, se utiliza la estimación de la máxima verosimilitud para obtener los valores de \mathbf{w} y de nuevo habría que utilizar técnicas iterativas para resolver estos valores. Hay situaciones en las que un proceso de iteración puede no converger, y entonces deberían probarse procesos iterativos alternativos.

El otro método usado en Credit Scoring y que es muy común en modelos económicos es el análisis tobit. La transformación tobit supone que se puede estimar p_i como

$$p_i = \max \{\mathbf{w}'\mathbf{x}_i, 0\} = \max \{w_0 + w_1 X_{1i} + w_2 X_{2i} + \dots + w_p X_{pi}, 0\} \quad (2.4)$$

En este caso, se trata de limitar los valores del método de regresión del análisis discriminante. Hay muchas situaciones económicas donde la variable que estamos interesados en mostrar su valor solo tiene significado si es positiva, los análisis estadísticos de las regresiones tobit sirven para ser aplicados en estos casos.

En el contexto de crédito Scoring encontramos algunas cosas insatisfactorias sobre la asimetría de la transformación tobit. Se termina con la dificultad de estimar probabilidades negativas pero no con probabilidades mayores de uno. Un modelo más simétrico sería

$$P_i = \min \{1, \max\{w'x_i, 0\}\}, \quad (2.5)$$

pero desafortunadamente el análisis y la estimación de los parámetros para esta transformación no es posible con los métodos disponibles.

Los métodos probit y tobit no encuentran muchos adeptos entre los estudiosos de Credit Scoring ya que estos están más preocupados por encontrar la decisión correcta para casos en que los valores de decisión no están muy claros que en el hecho de que la probabilidad de que un cliente sea malo sea claramente 0.05 ó -0.05. De todas formas, esto tiene que ser recordado por las técnicas estadísticas que no conocen esto, y su objetivo es minimizar la suma total de errores.

2.4 Medición de la Performance de un Scorecard

2.4.1 Estadísticas de Separación¹⁰

Cuando las personas hablan acerca de que tan bien un scorecard “separa”, esto es exactamente lo que están hablando de la separación de las distribuciones de préstamos buenos y malos.

2.4.1.1 Divergencia Estadística

¹⁰ Mays, Elizabeth. Credit Scoring For Risk Managers, The Handbook for Lender. Thomson. 2004

La divergencia estadística es muy sencilla. Es simplemente el cuadrado de la diferencia de la media de los buenos y la media de los malos, dividido por el promedio de la varianza de la distribución de las puntuaciones. La fórmula para la divergencia estadística es la siguiente:

$$D^2 = [(media_score_buenos) - (media_score_malos)]^2 / \sigma^2$$

Donde: $\sigma^2 = (\sigma_G^2 - \sigma_B^2) / 2$

Dada su simplicidad y facilidad de interpretación, la divergencia es, probablemente, una buena estadística para el informe. Una crítica que puede hacerse a ella es: no captura toda la información importante acerca de la forma de las distribuciones.

Según la autora Mays, es importante que se reconozca que las estadísticas que se usa para evaluar scorecard son simplemente medidas que de algún aspecto diferencia entre distribuciones de buenos y malos. Ellos no dicen todo lo que necesita saber acerca de cómo un scorecard está haciendo su trabajo.

2.4.1.2 El estadístico KS

La estadística KS es muy fácil de entender y de calcular. Es la máxima diferencia entre el porcentaje acumulado de una distribución buenos y el porcentaje acumulado de la distribución de malos.

Según menciona la autora Mays en su libro, no existe norma sobre lo que el valor esperado de la KS debería ser o cuán grande debe ser antes de que podamos estar seguros de que tenemos una buena puntuación. Un relativamente bajo KS no significa necesariamente que alguien ha hecho un pobre trabajo en la construcción del scorecard, quizá puede ser lo mejor que se puede obtener en esa situación en particular. Lo que se busca es que se pueda tener un KS que de confianza a los que desarrollan el modelo.

Si bien en la teoría el KS puede estar en el rango de 0 a 100, en la práctica, la gama es generalmente de alrededor de 20 a alrededor de 70. Si el KS es inferior a 20, sería razonable preguntarse si la puntuación vale la pena usar. Por encima de 70, es probablemente demasiado bueno para ser verdad y debemos sospechar de problemas con la forma en que se calcula o con el mismo score.

2.4.2 Estadísticos de Clasificación – Curva de ROC y el coeficiente GINI¹¹

Uno de los posibles inconvenientes de las estadísticas de separación es que hemos observado que se basan en un solo punto, en las distribuciones de buenas y malas.

El coeficiente GINI, por otra parte, utiliza la información acerca de todo el tramo de la distribución de buenos y malos y resume la capacidad del scorecard para asignar relativamente puntajes más bajos a los préstamos que eventualmente van mal. El coeficiente GINI se calcula como el área bajo la curva ROC. La curva de ROC es también conocido como la curva de trade-off porque muestra el trade-off entre buenos y malos, el porcentaje del total de malos que deben ser aceptadas con el fin de aceptar un determinado porcentaje del total de buenos.

No hay un valor teórico óptimo para el GINI, sin embargo en la bibliografía que se ha revisado existe cierto acuerdo en considerar que a partir de un 35%, se puede considerar aceptable.

2.4.3 Estadísticos de Error de Predicción

2.4.3.1 Bondad de Ajuste

¹¹ Thomas, Lyn, David Edelman y Jonathan Crook. Credit Scoring and Its Applications. SIAM. 2002

El estadístico de Hosmer Lemeshow (HL) es un estadístico de bondad de ajuste para los modelos de regresión logística. Para los modelos de Credit Scoring, mide cómo la predicción del modelo de un número de préstamos malos para un rango de score coincide con el número real de créditos malos en cada rango. A diferencia de los estadísticos de separación y clasificación, esta prueba no mide la capacidad del scorecard de asignar puntuaciones inferiores a los préstamos malos, sino más bien mide la capacidad del modelo para predecir el número real de los malos dentro de un determinado segmento de score.

No hay ningún rango típico para el estadístico HL que dice que el modelo es experto en predecir el número de préstamos malos, porque el estadístico varía para las diferentes muestras en función del ratio esperado de malos. Una crítica al estadístico de HL es que es sensible al número de rangos de score utilizado en su cálculo.

2.5 CUT-OFF o Puntos de Corte¹²

Hay una serie de consideraciones importantes que participan en la fijación de un punto de corte, entre ellos el número de préstamos que será aprobado y el rendimiento que se espera de estos préstamos para la institución bancaria. Un prestamista debe tener muy en claro su capacidad de operación a la hora de establecer el corte. La aprobación de demasiados préstamos, o la aprobación de un menor número de préstamos, bien podría causar sobrecarga de recursos, igual de importante, infrutilización de los recursos.

El autor Lyn Thomas considera que, la forma más simple de determinar los puntos de corte, es considerar una tasa de aprobación igual a la que se venía aprobando en la institución bancaria, esto hasta que se le vaya tomando más

¹² Thomas, Lyn, David Edelman y Jonathan Crook. Credit Scoring and Its Applications. SIAM. 2002

confianza al nuevo modelo. Sin embargo, indica el autor que esto debe cambiar en un corto tiempo ya que la inversión tanto en tiempo como en dinero, debe de reflejarse en un mayor porcentaje de aprobación y por consiguiente en una mayor rentabilidad.

2.6 Modelo a Aplicar en base al Marco teórico Desarrollado

De acuerdo a la teoría desarrollada en el presente capítulo, se ha procedido a considerar lo siguiente para la aplicación del modelo:

- El periodo de tiempo que se ha utilizado ha estado supeditado a la data que se ha podido obtener para la construcción del modelo. Asimismo, se ha tomado la definición solo de “Buenos” y “Malos”, eliminando la opción de “Indeterminados”, ya que lo se busca es poder discriminar y a su vez se ha estado supeditado, a la cantidad de data que se ha proporcionado.
- Respecto al diseño de la muestra, se ha dividido en un 80% para muestra de desarrollo y un 20% para la muestra de validación según indica la teoría. También la cantidad de características o variables ha estado supeditado a la data proporcionada. Para el análisis principalmente se ha usado la prueba Chi-Cuadrado.
- El modelo estadístico que se ha empleado es La Regresión Logística Multivariante, este modelo es preferido por los autores leídos en relación a la regresión lineal porque el tipo de variable dependiente cuenta con la naturaleza dicotómica (elección binaria), mientras que en la regresión lineal no pone restricciones pudiendo tomar cualquier valor, a su vez en la regresión lineal el término de error no es homocedástico, ya que la varianza condicional varía según las observaciones, por lo que las estimaciones de Beta no son eficientes.
- Para la medición de la performance del scorecard, se han utilizado las pruebas del KS y del coeficiente de GINI, las mismas que los autores sugieren utilizar. Si bien es cierto muchos de los autores que se ha

revisado no dan mucha referencia respecto al Backtesting, se ha decidido emplearlo como una prueba adicional de Respaldo.

- Finalmente se concluye con la determinación de los CUT-OFF. Se va a calcular los tramos en función a los porcentajes de aprobación y de mora (esto medido en número de operaciones, no de saldo).

CAPITULO III: DESARROLLO DEL MODELO Y ANÁLISIS DE LOS RESULTADOS

3.1 Datos a Utilizar Para el Modelo

Para el desarrollo del modelo se ha obtenido data de una cartera de tarjetas de créditos reales solicitadas a una institución bancaria la cual ha solicitado la confidencialidad de su nombre y de los datos de los titulares de las tarjetas. Esta cartera cuenta con data histórica que comprende los periodos julio del 2006 hasta junio del 2008. La misma contiene información amplia de 20,000 tarjetas de créditos, asimismo se cuenta con 17 variables entre cualitativas y cuantitativas las mismas que han ayudado para modelar el Scoring. Los datos corresponden a la cartera de clientes de Lima Metropolitana de la institución bancaria en mención.

3.2 Depuración de Datos

De los 20,000 registros con los que se cuentan, se han procedido a separar a aquellos expedientes que no se encuentran formalizados, es decir que fueron aprobados pero que no se llegaron a emitir por razones como que el cliente no aceptó o que no pudieron entregar la tarjeta ya que no encontraron la dirección, entonces la nueva población quedó establecida en 11,190 expedientes de tarjetas de crédito. Asimismo, mencionar que en la data obtenida, no se encontraron campos en blanco (todos los campos de las variables contenían información).

3.3 Definición de Buenos y Malos

La definición de Buenos y Malos, dado que se trata de un modelo personalizado de originación, depende del conocimiento del sistema de cartera en mora dentro de la institución bancaria y también del proceso de cobranza. Una definición de Malos se refiere a aquellas tarjetas de crédito, que dada su experiencia, la institución financiera no quiere seleccionar para que pasen a ser clientes. Dicho en otros términos, se refiere a aquellos casos que el Banco en mención, hubiera deseado no haber aprobado, de haber conocido lo que pasaría.

Para poder obtener una definición nos procedimos a reunir con dos especialistas del área de recuperaciones, para conocer como es el proceso de cobranza y recuperación de la institución financiera.

La gestión de cobranza se hace por tramos, los cuales se detallan a continuación:

- a) **Primer tramo**, este se da en los primeros 30 días de vencida la primera cuota. La gestión se realiza a través de call center (llamadas telefónica) que contrata la institución. Usualmente los clientes que se encuentran en este tramo son conocidos como “desordenados”, cuentan con el dinero para pagar pero muchas veces “no cuentan con el tiempo” suficiente para acercarse a una oficina a pagar su deuda. De acuerdo a las estadísticas que maneja la institución bancaria alrededor del 80% se logra cobrar y ponerse al día.
- b) **Segundo tramo**, este tramo comprende a aquellos clientes que ya presentan dos cuotas vencidas de su tarjeta. La gestión de recuperación la derivan a agencias de cobranza contratadas por la institución. Los clientes que llegan a este tramo ya evidencian problemas de pago, y por consiguiente requieren tener un seguimiento más de cerca para evitar que lleguen al tercer tramo. Las estadísticas que maneja la institución

indican que aproximadamente el 60% se logra cobrar y mejorar su clasificación en el sistema.

- c) Tercer tramo**, en este tramo están las personas que ya cuentan con 3 cuotas de tarjeta vencida. Aquí si hay un problema serio, la cobranza se comparte entre las agencias de cobranza y los estudios de abogados contratados por el banco. Las estadísticas muestran que entre un 25% a 30% de este tramo son recuperables, mejorando su clasificación.

En el modelo que se ha desarrollado, se ha definido como clientes malos, aquellos que se encuentran en el tercer tramo, tomando en cuenta dos aspectos: el primero es que la cartera que se puede recuperar de este tramo es baja y un segundo motivo es que los clientes que se encuentran en este tramo registran una clasificación de “Dudoso” en las centrales de riesgos, por lo tanto le corresponde una provisión del 60% del monto adeudado lo que se refleja en resultados negativos para la institución (deja de generarle utilidad para el negocio) . En la base de datos que se obtuvo los clientes que figuran con la etiqueta “Mora”, son aquellos que justamente se encuentran en ese tramo y que son 493 expedientes.

3.4 Determinación de la Muestra Madura

Un primer punto a considerar es si se cuenta con una población suficiente como para poder extraer una muestra madura que permita predecir el comportamiento de los futuros clientes de tarjeta de crédito.

Para esto se ha hecho un análisis de cosecha mensual, se muestra en el cuadro 3.1, donde se detalla el número de operaciones “Normales” y las operaciones que han presentado “Mora”. Se calcula el % de mora por segmento (cosecha) y lo que se busca es que se pueda contar con un periodo en el cual la mora se ha estabilizado, es decir que las tarjetas que debieron

caer en mora ya lo hicieron y las que han tenido un comportamiento de pago normal, es muy improbable que vayan a caer en impago.

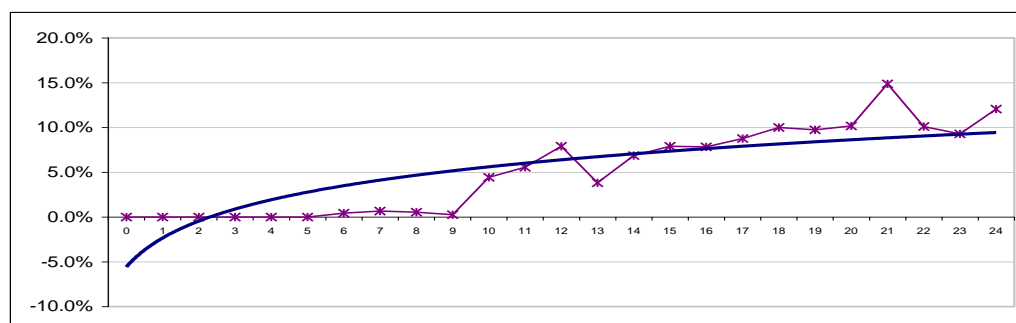
Esto se aprecia mejor gráficamente, cuando la curva deja de mostrar una pendiente positiva, más bien se observa que la curva tiende a estabilizarse en un determinado rango de mora (se adjunta línea de tendencia donde se observa esto). A continuación se muestra el grafico 3.1, en donde se puede observar este comportamiento descrito anteriormente.

CUADRO 3.1

fecha	Madurez	NORMAL	MORA	Total Operaciones	% mora por segmento
200807	0	10	0	10	0.0%
200806	1	814	0	814	0.0%
200805	2	594	0	594	0.0%
200804	3	532	0	532	0.0%
200803	4	501	0	501	0.0%
200802	5	441	0	441	0.0%
200801	6	451	2	453	0.4%
200712	7	438	3	441	0.7%
200711	8	1281	7	1288	0.5%
200710	9	781	2	783	0.3%
200709	10	259	12	271	4.4%
200708	11	390	23	413	5.6%
200707	12	280	24	304	7.9%
200706	13	151	6	157	3.8%
200705	14	353	26	379	6.9%
200704	15	548	47	595	7.9%
200703	16	318	27	345	7.8%
200702	17	302	29	331	8.8%
200701	18	198	22	220	10.0%
200612	19	213	23	236	9.7%
200611	20	495	56	551	10.2%
200610	21	521	91	612	14.9%
200609	22	160	18	178	10.1%
200608	23	469	48	517	9.3%
200607	24	197	27	224	12.1%

Elaboración propia

GRAFICO 3.1



Elaboración propia

Con lo anterior se puede concluir que existe estabilidad en el periodo que se está observando, la mora se ha ordenado (es decir, a más periodo de maduración mayor mora), por lo tanto concluimos que el periodo de 24 meses es adecuado para poder extraer la muestra madura.

Luego, se ha procedido a hallar el periodo que comprende la muestra madura, para esto es necesario hallar el porcentaje de mora acumulado en el periodo de 24 meses. En el cuadro 3.2 se detalla los cálculos obtenidos, así como en el gráfico 3.2 se observa el comportamiento de la mora acumulada. Se toma el porcentaje de mora acumulado porque de esta forma se evita ocultar información. Lo que se quiere decir es que si se considera la mora por cada cosecha, pueden existir meses en los que hubo un número de emisiones por encima del promedio, esto podría dar la impresión que la mora ha disminuido, lo cual es engañoso ya que no refleja la realidad, por ejemplo esto se ve en los meses de agosto y octubre del 2006, en donde se podría creer que la mora disminuyó, pero no fue así.

CUADRO 3.2

fecha	Madurez	NORMAL	MORA	Total Operaciones	% mora por segmento	Operaciones Acumulada	Morasas Acumulada	% MORA
200607	24	197	27	224	12.1%	224	27	12.1%
200608	23	469	48	517	9.3%	741	75	10.1%
200609	22	160	18	178	10.1%	919	93	10.1%
200610	21	521	91	612	14.9%	1531	184	12.0%
200611	20	495	56	551	10.2%	2082	240	11.5%
200612	19	213	23	236	9.7%	2318	263	11.3%
200701	18	198	22	220	10.0%	2538	285	11.2%
200702	17	302	29	331	8.8%	2869	314	10.9%
200703	16	318	27	345	7.8%	3214	341	10.6%
200704	15	548	47	595	7.9%	3809	388	10.2%
200705	14	353	26	379	6.9%	4188	414	9.9%
200706	13	151	6	157	3.8%	4345	420	9.7%
200707	12	280	24	304	7.9%	4649	444	9.6%
200708	11	390	23	413	5.6%	5062	467	9.2%
200709	10	259	12	271	4.4%	5333	479	9.0%
200710	9	781	2	783	0.3%	6116	481	7.9%
200711	8	1281	7	1288	0.5%	7404	488	6.6%
200712	7	438	3	441	0.7%	7845	491	6.3%
200801	6	451	2	453	0.4%	8298	493	5.9%
200802	5	441	0	441	0.0%	8739	493	5.6%
200803	4	501	0	501	0.0%	9240	493	5.3%
200804	3	532	0	532	0.0%	9772	493	5.0%
200805	2	594	0	594	0.0%	10366	493	4.8%
200806	1	814	0	814	0.0%	11180	493	4.4%
200807	0	10	0	10	0.0%	11190	493	4.4%

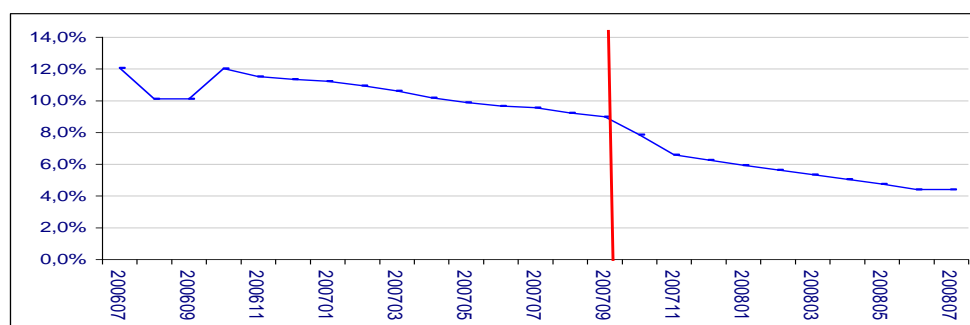
Elaboración propia

Para determinar el tamaño de la muestra madura se deben de considerar dos aspectos importantes: uno es el periodo de maduración del producto y otro es

el expertise del desarrollador del modelo. Normalmente para evaluar la madurez de una tarjeta de crédito de consumo se considera entre 6 a 12 meses, se podría considerar como indicador basado en el expertise que la mora en la muestra madura debería variar en un rango de 3% a 4%.

Observando tanto en el cuadro como en el gráfico un comportamiento más o menos estable de la mora acumulada en los periodos comprendidos entre julio del 2006 y setiembre del 2007, después este periodo se observa una clara tendencia a la baja de la morosidad acumulada, ya que son operaciones relativamente nuevas.

GRAFICO 3.2



Elaboración propia

Por lo anterior se llega a la conclusión que el periodo de muestra madura que se ha usado para estimar el modelo ha estado comprendido entre julio 2006 y setiembre del 2007.

3.5 Análisis de Variables

De acuerdo a la data obtenida, se ha contado con 17 variables, las cuales se detallan a continuación:

- **Variables Cualitativas:** Sexo, Estado Civil, Continuidad Laboral, Profesión, Cliente Banco, Cliente Nomina Banco, Procedencia Tarjeta.

- **Variables Cuantitativas:** Edad, Unidad Familiar, Antigüedad Empleo, Ingresos Totales, gastos Fijos, Gastos Hipotecarios, Valor Vivienda, Patrimonio, Saldo Medio, Antigüedad Cliente Banco.

Se ha realizado dos tipos de análisis previo a la corrida del modelo, estos son el univariante, que es un análisis descriptivo de cada variable, así como un bivariante que implica comparar cada variable independiente con la variable dependiente, esto con el fin de ver el grado de asociación entre ambas, separando aquellas que no han guardado asociación, por lo tanto, no han sido incluidas en el modelo final. La principal herramienta que se ha usado es el programa estadístico SPSS versión 15.

3.5.1 Análisis Univariante

El análisis estadístico ha consistido en el cálculo de medidas de tendencia central y de dispersión, para las variables cuantitativas. Para el caso de las variables cualitativas, se ha obtenido las frecuencias para cada categoría, así como el porcentaje correspondiente.

3.5.1.1 Variable Dependiente

Distribución y porcentaje de la variable MORA, la cual consta de dos categorías, se observa que la mayoría de los clientes, con un porcentaje de 91%, se clasificaron en la categoría Normal.

CUADRO 3.3

MORA					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Normal	4855	91.0	91.0	91.0
	Mora	479	9.0	9.0	100.0
	Total	5334	100.0	100.0	

3.5.1.2 Variables Independientes Cualitativas

Con los resultados obtenidos que pueden apreciarse en el anexo 3 y el CD adjunto tenemos que en las variables “Estado civil”, “Continuidad Laboral” y “Profesión”, se encontraron porcentajes con menos del 1%, por lo que se ha procedido a combinar algunas categorías. El criterio para agrupar las categorías se nota más claro al momento de cruzarlas con la variable dependiente.

3.5.1.3 Variables Independientes Cuantitativas

Se presentan estadísticos descriptivos tales como la media y la desviación estándar, las mayores desviaciones las presentan las variables Valor vivienda y Patrimonio, estas variables presentan un alto grado de dispersión respecto al valor promedio.

CUADRO 3.4

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. típ.
Edad	5334	19	71	37.66	10.085
Unidad Familiar	5334	1	9	1.91	1.209
Antigüedad Empleo	5334	0	42	6.03	6.903
Ingresos Totales	5334	1.00	52854.00	2219.0529	2092.15399
Gastos Fijos	5334	0	9379	159.33	372.086
Gastos Hipotecarios	5334	0	149600	63.93	2058.650
Valor Vivienda	5334	0	622000	4945.85	29918.667
Patrimonio	5334	0	621041	1939.09	20800.265
Saldo Medio en Cuenta	5334	0	74676	776.89	3460.593
Antigüedad Cliente Banco	5334	0	27	3.82	4.303
N válido (según lista)	5334				

3.5.2 Análisis Bivariante¹³

¹³ Siegel, Sidney. Estadística No Paramétrica. Editorial Trillas. 1991

Ante de realizar la regresión logística se ha hecho un análisis bivalente, esto para comprobar la existencia de las relaciones de la variable dependiente con cada una de las variables independientes.

- Para el caso de las variables independientes categóricas, el contraste se ha hecho a través de la prueba chi cuadrado, en la cual se evalúa la fuerza de asociación y se comprueba la significancia estadística del contraste asociado al estadístico.
- Para el caso de las variables cuantitativas el contraste se ha hecho mediante la prueba de Kolmogorov Smirnov, en el cual se evalúa la fuerza de asociación mediante la diferencia de medianas (ver si existen diferencias entre los grupos Normal y Mora).

3.5.2.1 Variables Cualitativas

En el caso de las variables dicotómicas se ha calculado el χ^2 con la corrección de continuidad de Yates y en los demás casos el χ^2 de Pearson.

Para un nivel de significancia igual a 0.25 (no se ha tomado 0.05 como nivel de significancia en este caso ya que se considera muy estricto, puesto que solo se ha visto la existencia del grado de asociación), la prueba de hipótesis es:

H_0 : No existe asociación entre las variables (entre la variable MORA y la independiente que se tome)

H_1 : Existe asociación entre las variables

Se rechaza H_0 si $p < 0.25$, caso contrario se acepta.

Por otro lado, antes de comprobar la significancia se debe observar que las frecuencias esperadas de la tabla de contingencia deben ser lo suficientemente grandes (frecuencias mayores que 5), ya que "la existencia de celdas casi

vacías puede dominar la solución del modelo”¹⁴, en caso no se cumpliera este requisito se hace necesario una agrupación de las categorías. El criterio que se ha tomado es agrupar las categorías de cada variable según el nivel de morosidad (lo que se busca es que las agrupaciones discriminen por morosidad). Esta es una parte muy importante del trabajo, ya que lo que se busca es perfilar aquellas variables que van a diferenciar mejor los morosos de los no morosos.

Los resultados de los cálculos realizados están presentados en el anexo 3 y en el CD, sin embargo se puede concluir que se ha encontrado asociación entre la variable dependiente “Mora” y las variables cualitativas “Sexo”, “Estado Civil”, “Profesión”, “Cliente Banco”, “Cliente Pago de Haberes” y “Procedencia”. La variable “Continuidad Laboral” después del análisis realizado se ha observado que no ayuda a discriminar entre morosos y no morosos, sin embargo, se ha dejado a dicha variable ya que agrupada con otra puede ayudar a discriminar entre morosos y no morosos.

Otro aspecto importante que saltó del análisis, es que las variables “Estado Civil” y “Profesión” han presentado en las tablas de contingencias celdas con frecuencias menores a 5%, lo que obliga necesariamente a agruparlas. En el caso de “Estado Civil” se agruparon en tres grupos y en el caso de “Profesión”, esta se ha agrupado en dos. La composición de estas se pueden apreciar en el anexo 3.

Asimismo se ha encontrado asociación entre las variables “Sexo”, “Estado Civil” y “Profesión” se ha pasado a realizar un análisis de correspondencias para ver si es posible agruparlas en una menor dimensión. Por lo que se ha obtenido los siguientes resultados.

CUADRO 3.5

¹⁴ Tabachnick y Fidell (1996), <http://www.tdx.cesca.es/TDX-1006103-132336>

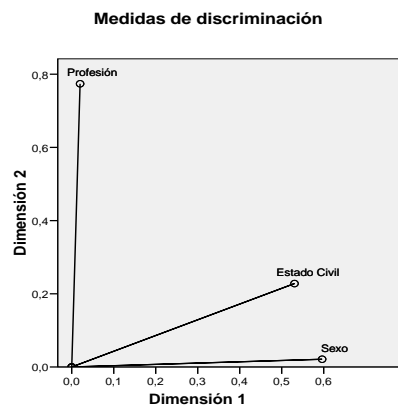
Medidas de discriminación

	Dimensión		Media
	1	2	
Sexo	.596	.021	.309
Estado Civil	.530	.228	.379
Profesión	.020	.774	.397
Total activo	1.146	1.023	1.085
% de la varianza	38.214	34.097	36.155

En el anexo 3 se muestra los resultados de la nueva agrupación entre las variables “Sexo” y “Estado Civil”. Los nuevos grupos conformados son los siguientes (está ordenado de menor a mayor riesgo):

-
-
-

GRAFICO 3.3



Normalización principal por variable.

Se ha aplicado una matriz de correlaciones para variables cualitativas observando que las variables “Procedencia” y “Pago de Haberes” presentan una alta correlación por lo que se ha visto conveniente quitar una de ellas. Se ha procedido a quitar del modelo la variable “Procedencia”, ya que la variable “Pago de Haberes” presenta mayor relevancia en este estudio.

Se ha visto que existe correlación aunque en menor grado entre las variables “Cliente Banco” y “Pago de Haberes”. A estas últimas dos variables se llaman variables de vinculación, y es lógico que haya cierto grado de correlación entre ellas, ambas se han quedado en el modelo ya que se considera que son importantes para ayudar a discriminar entre morosos y no morosos (opinión de experto).

CUADRO 3.6

Correlaciones de las Variables originales

	Sexo y Estado Civil	Profesión	Continuidad Laboral	Cliente Banco	Pago de Haberes Banco	Procedencia
Sexo y Estado Civil	1.000	.042	-.063	.042	-.057	-.045
Profesión	.042	1.000	-.137	.066	-.066	-.020
Continuidad Laboral	-.063	-.137	1.000	.016	.166	.052
Cliente Banco	.042	.066	.016	1.000	.292	.214
Pago de Haberes Banco	-.057	-.066	.166	.292	1.000	.461
Procedencia	-.045	-.020	.052	.214	.461	1.000

3.5.2.2 Variables Cuantitativas

Para cada variable cuantitativa se ha utilizado la prueba de Kolmogorov – Smirnov para ver si existen diferencias entre los grupos (Normal y Mora) de la variable dependiente Mora. Se ha utilizado esta prueba ya que ninguna de las variables sigue distribución normal y además por los tipos de variables cuantitativas que se encuentran en el estudio (variables continuas y discretas).

H_0 : No existen diferencias entre los grupos (medianas iguales).

H_1 : Existen diferencias entre los grupos (medianas distintas).

*se ha usado la mediana debido a que es una medida de tendencia central más robusta que la media, no es afectada por las observaciones outliers.

Se muestra en el cuadro siguiente la prueba para cada variable:

CUADRO 3.7

Estadísticos de contraste^a

		Edad	Unidad Familiar	Antigüedad Empleo	Ingresos Totales
Diferencias más extremas	Absoluta	.084	.059	.176	.053
	Positiva	.000	.002	.009	.012
	Negativa	-.084	-.059	-.176	-.053
Z de Kolmogorov-Smirnov		1.758	1.234	3.665	1.108
Sig. asintót. (bilateral)		.004	.095	.000	.171

^a Variable de agrupación: MORA

Estadísticos de contraste^a

		Gastos			Saldo Medio en Cuenta	Antigüedad Cliente Banco	
		Gastos Fijos	Hipotecarios	Valor Vivienda	Patrimonio		
Diferencias más extremas	Absoluta	.146	.032	.036	.012	.254	.201
	Positiva	.001	.002	.002	.002	.000	.000
	Negativa	-.146	-.032	-.036	-.012	-.254	-.201
Z de Kolmogorov-Smirnov		3.050	.672	.757	.242	5.312	4.204
Sig. asintót. (bilateral)		.000	.757	.616	1.000	.000	.000

^a Variable de agrupación: MORA

Para un nivel de significancia igual a 0.25, se rechaza H_0 para las variables “Edad”, “Unidad Familiar”, “Antigüedad del Empleo”, “Ingresos Totales”, “Gastos Fijos”, “Saldo Media en Cuenta” y “Antigüedad del Cliente”. Para las variables “Gastos Hipotecarios”, “Valor Vivienda” y “Patrimonio”, no existen diferencias significativas entre los grupos, por lo que la fuerza de asociación para con la variable dependiente es débil o no existe, quedando fuera del modelo.

3.5.3 Regresión Logística Simple

Se ha realizado una regresión logística simple para ver que variables resultan significativas y cuales entran en el modelo de regresión logística multivariante, para un nivel significancia igual a 0.1.

Para las variables continuas aparte de la significancia se ha evaluado el incremento o decremento unitario en sus valores para ver si tienen algún sentido un cambio en sus odds. En algunos casos ha sido mejor transformarla en una variable categórica a costa de perder un poco de información.

La variable Unidad Familiar con un $p=0.155$ no resulta significativa en el modelo, por lo que no ha ingresado a la regresión logística multivariante.

_____ (está ordenado de menor a mayor riesgo).

Temporal (contratado) e ingresos < 2300

Fijo e ingresos > 2300

Independiente e ingresos < 2300

Otros e ingresos < 2300

Otros e ingresos > 2300

_____ (en este caso está ordenado de mayor a menor riesgo).

Finalmente las variables que han ingresado al modelo son "Sexo_Estado Civil", "Profesión", "Pago de Haberes", "Cliente del Banco", "Antigüedad del empleo", "Antigüedad del Cliente", "Edad", "Ingresos_Continuidad Laboral", "Gastos Fijos" y "Saldo Media en Cuenta".

Se ha ingresado las variables en el modelo, para un nivel de significancia igual a 0.05 se ha pasado a retirar las variables "cliente Banco" y "Gastos Fijos", las variables "Edad" e "Ingreso_continuidad laboral", se quedaron en el modelo debido a que se ha considerado relevantes en el modelo (opinión de experto) aun cuando no presenten significancia en el.

Variables en la ecuación						
	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a						
SEXO_ESTCIVIL			43.317	2	.000	
SEXO_ESTCIVIL(1)	.626	.120	27.331	1	.000	1.869
SEXO_ESTCIVIL(2)	1.141	.218	27.391	1	.000	3.130
PROFESION_COD(1)	.668	.112	35.536	1	.000	1.950
CLIE_NOM(1)	.785	.259	9.205	1	.002	2.191
ANTEMP	-.053	.012	18.329	1	.000	.948
CLIE_ANT	-.050	.016	9.707	1	.002	.952
EDAD	.009	.007	1.625	1	.202	1.009
SMD_COD			33.824	2	.000	
SMD_COD(1)	-.278	.265	1.101	1	.294	.758
SMD_COD(2)	-1.467	.262	31.320	1	.000	.231
INGTOTAL_CONTILABO(1)	.227	.129	3.077	1	.079	1.255
Constante	-3.363	.391	73.863	1	.000	.035

^a. Variable(s) introducida(s) en el paso 1: SEXO_ESTCIVIL, PROFESION_COD, CLIE_NOM, ANTEMP, CLIE_ANT, EDAD, SMD_COD, INGTOTAL_CONTILABO.

3.6.1 Prueba de Bondad de Ajuste

Se ha realizado la prueba de bondad de ajuste para corroborar que el modelo calculado se ajusta efectivamente a los datos usados para estimarlo, se llevó a cabo mediante la prueba de Hosmer – Lemeshow.

La hipótesis a probar es:

H_0 : El modelo es adecuado

H_1 : El modelo no es adecuado

Si $p < 0.05$ se rechaza la H_0 caso contrario se acepta H_0 . En este caso tenemos que $p > 0.05$ por lo tanto se acepta H_0

CUADRO 3.10

Prueba de Hosmer y Lemeshow			
Paso	Chi-cuadrado	gl	Sig.
1	7.500	8	.484

p

3.6.2 Cálculo del Coeficiente GINI y el Estadístico KS

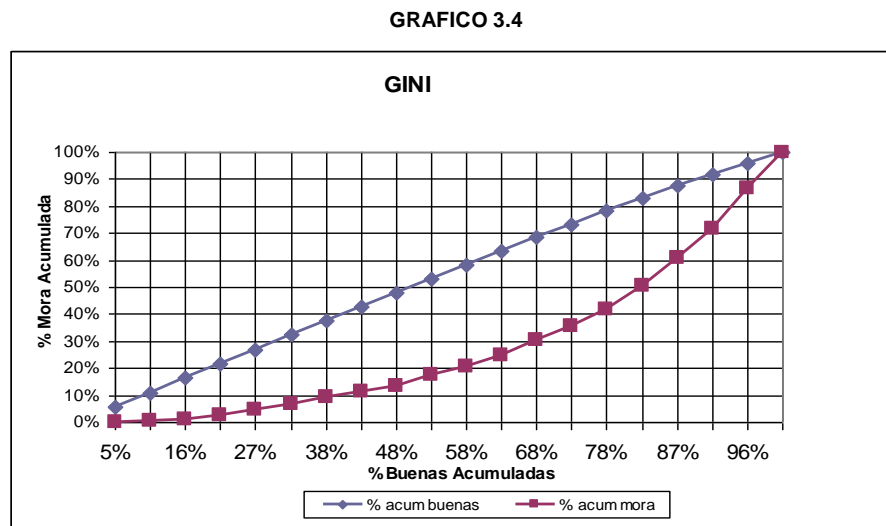
Una vez calculados los Beta, el siguiente paso es correr estos resultados a la muestra de desarrollo, con este paso se obtiene la probabilidad de mora de cada cliente de la muestra de desarrollo, los resultados se pueden apreciar en el CD adjunto.

El Coeficiente de GINI nos permite saber si el modelo ordena bien a los clientes, es decir, asigna probabilidades de mora baja a los buenos clientes (Normal) y probabilidades de mora alta a los malos clientes (Morosos). No existe un criterio teórico único que diga cual debería ser el GINI adecuado, sin embargo, en base a la experiencia de desarrolladores de modelos de Scoring con los cuales hemos podido conversar, así como en algunos documentos de trabajo, se indica que un GINI a partir del 35%, indica que se cuenta con un buen modelo que agrupa bien a los clientes. Para el cálculo de los GINI es necesario agrupar la muestra en percentiles, para este modelo se ha decidido agruparlos en percentiles de 5%. Es necesario obtener los porcentajes acumulados.

La prueba del KS nos permite saber que tan bien el modelo discrimina entre los buenos (Normal) y malos clientes (Morosos), tal como se comentaba en el marco teórico este valor debería fluctuar entre 20% y 70%, con lo cual tendríamos un modelo que discrimina bien. Este estadístico se halla como la mayor diferencia entre el porcentaje acumulado de buenos clientes y el porcentaje acumulado de malos clientes, para los distintos percentiles de la muestra desarrollo.

La muestra de desarrollo arroja un GINI de 50,62%, lo cual es bastante bueno para este tipo de modelos, sin embargo es importante calcular también el GINI para la muestra de validación y este debe ser cercano al GINI de la muestra de desarrollo.

El KS obtenido para el modelo de desarrollo es de 38,76%, el cual también se encuentra dentro del rango aceptable para este estadístico. Se muestra en el gráfico 3.4, donde se observan los resultados.



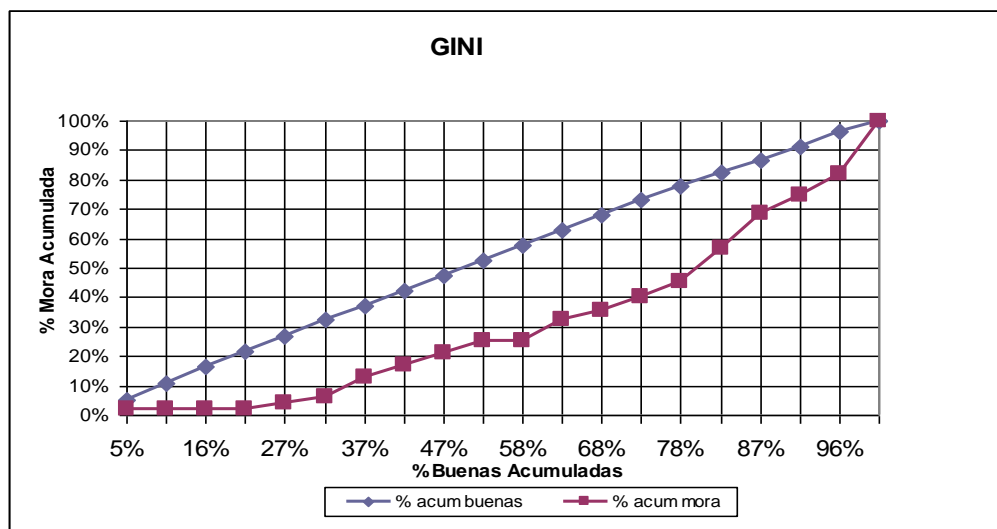
3.6.3 Cálculo del GINI y KS para la muestra de validación

Un aspecto muy importante de este tipo de modelos, es su validación. Para esto de la muestra madura se ha seleccionado un 20% de clientes en forma aleatoria, y con los Beta obtenidos de la muestra de desarrollo se ha calculado también las probabilidades de mora y con estos resultados se ha obtenido los estadísticos antes mencionados.

Para el caso del GINI, lo recomendable es que este no varíe más del 10%, lo ideal sería que tengan un igual GINI, sin embargo esto no se observa en la práctica, normalmente difieren estos dos valores. En el caso del KS, igual se busca obtener un valor que este entre 20% y 70%. Para nuestra muestra de validación se obtuvo un GINI del 43,29% lo cual es bueno, el KS obtenido es de 32,94%.

Se muestra el gráfico 3.5 con los resultados obtenidos.

GRAFICO 3.5



3.7 Backtesting

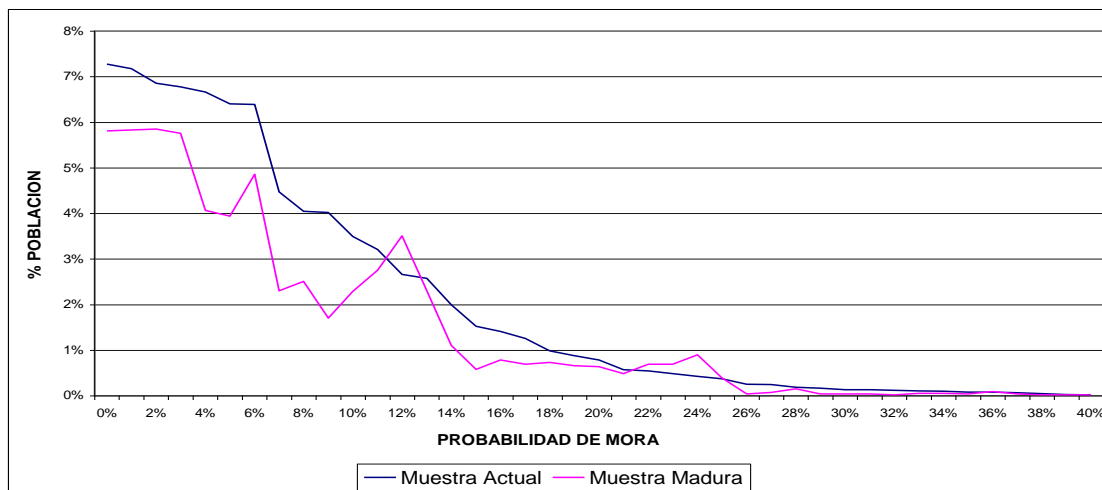
Una vez validado el modelo, se ha procedido a realizar un backtesting, que es una prueba de respaldo al modelo propiamente dicho.

Para este cálculo se ha tomado una muestra actual que corresponde al periodo octubre del 2007 hasta julio del 2008, en esta muestra contamos con 10,282 registros (en esta muestra se incluyen los casos no formalizados, cosa que no se ha hecho para la determinación de la muestra madura, lo que se busca en contar con mayor data para el Backtesting) y con los Betas calculados de la muestra de desarrollo, se calculan la probabilidad de mora para cada cliente de esta muestra actual. Luego esto es comparado con toda la muestra madura, y se observan como se comportan uno en relación al otro.

En caso presenten comportamientos distintos, esto indicaría que el modelo no está prediciendo para clientes nuevos, en términos de riesgo Crediticio, se podría afirmar que el perfil del cliente está cambiando con lo que sería necesario recalibrar el modelo, ya que no va a pronosticar bien con los clientes actuales.

Se muestra los resultados del Backtesting realizado en el gráfico 3.6, los cálculos de las probabilidades se pueden observar en el CD que se está adjuntando.

GRAFICO 3.6



Elaboración Propia

Tal como se puede apreciar en el gráfico la muestra actual y la muestra madura presentan comportamientos similares, por lo tanto podemos afirmar que el modelo discrimina bien para clientes nuevos.

3.8 Cálculo de los CUT-OFF o Puntos de Corte

Una vez ya hallado el modelo, confirmado los resultados con la muestra de validación y finalmente respaldado el modelo con un Backtesting, el siguiente paso tiene que ver con la política de riesgo de cada institución bancaria, es decir, de cuanto riesgo Crediticio está dispuesto a admitir.

Para el cálculo de los CUT-OFF es necesario seguir los siguientes pasos: Ordenar de menor a mayor probabilidad de mora tanto la muestra actual como la muestra madura, una vez realizado esto la pregunta que debe de hacerse es cuantas operaciones la institución está dispuesta a aprobar, suponiendo que sea un 50%, entonces se procede a ubicar la probabilidad de mora que

corresponde a la mitad de la muestra actual que está ordenada de menor a mayor probabilidad de mora, esto se compara con la muestra madura, y ahí es donde se hace el punto de corte. Hasta ese punto de corte se cuenta cuantos clientes morosos se tiene, y se calcula el porcentaje de mora que se estaría admitiendo.

Para este modelo se ha realizado los cálculos correspondientes, los cuales se encuentran detallados en el CD. A continuación, en el siguiente cuadro 3.11 se tiene el resumen de resultados obtenidos

CUADRO 3.11

% de Aprobación	Mora
30% DE APROBACION	2,65%
40% DE APROBACION	3,19%
50% DE APROBACION	3,64%
60% DE APROBACION	4,00%

Elaboración Propia

Del cuadro se puede observar que con una aprobación del 30% de todos los expedientes ingresados se tendría una morosidad del 2,65%, mientras que para una aprobación del 40% se tendría una morosidad del 3,19%, para una aprobación del 50% tendríamos una morosidad del 3,64, y para un 60% sería una morosidad del 4%. Tal como se indico la elección estará sujeta al nivel de riesgo que se desea admitir, pero también existen factores como la decisión del directorio de la institución bancaria en relación a la participación en el mercado de tarjetas de crédito, ya que si tiene una importante participación, podría buscar minimizar el riesgo, tomando un menor porcentaje de aprobación (30%), mientras que en el caso contrario que busque ganar participación tendrá que tomar un mayor riesgo, aprobando un mayor porcentaje (60%).

En nuestra opinión, una mora del 4% es aún una cifra manejable, teniendo en cuenta las altas tasas de interés que se cobran en el mercado de tarjetas de crédito, así como que la morosidad del sistema financiero actualmente es baja. Cabe aclarar que esta mora es en volumen, no en saldo. Aunque se debe esperar que la mora de saldo sea menor.

CONCLUSIONES Y RECOMENDACIONES

Una vez desarrollado y validado el modelo de Scoring en el presente trabajo de investigación, se extraen las siguientes conclusiones y recomendaciones:

- 1) El modelo Credit Scoring desarrollado en el presente trabajo, es una herramienta que agrega valor a la evaluación Crediticia realizada por cualquier institución financiera, sin embargo, por si sola no es suficiente para tomar la decisión final sobre la aprobación de una tarjeta de crédito (para nuestro caso), existen otros factores que no se explican tales como: la revisión en la central de riesgos del solicitante, el modulo de capacidad de pago que no es otra cosa que el disponible que le queda al solicitante para endeudarse (involucra el desarrollo de otro modelo, aunque bastante más sencillo que el Scoring del presente trabajo), las verificaciones domiciliarias y laborales (usuales para préstamos de consumo y tarjetas de crédito) y la validación de la documentación presentada por el solicitante.

- 2) Cualquier institución bancaria o financiera que no cuente aún con un Scoring, debería empezar con un modelo personalizado como el desarrollado en el presente trabajo, ya que este es el que sienta las bases para futuros modelos de Scoring. Este modelo puede convivir con otros modelos de Scoring tales como los modelos genéricos tan difundidos en los países desarrollados (ejemplo: Estados Unidos). La ventaja de un modelo personalizado en relación a los genéricos, es que en el primero se puede seleccionar el objetivo del modelo, así como la población objetivo, en el genérico esto no es posible.

- 3) Consideramos que este modelo puede ser utilizado como un modelo interno de medición de riesgo, según Basilea II. Ya que es posible determinar la probabilidad de default (PD) a partir de las probabilidades de mora que

calcula el modelo desarrollado. La PD es un componente importante de la pérdida esperada, que junto a la exposición y la severidad constituyen el modelo simple de Basilea II.

- 4) Resulta imprescindible que para la construcción del modelo este presente aparte del desarrollador del modelo, algún trabajador del banco que cuente con conocimientos estadísticos, así como con un amplio conocimiento del negocio bancario (tarjetas de crédito) para el que se construye el Scoring. Asimismo, es importante el adecuado llenado de la herramienta, ya que de lo contrario se obtendrán probabilidades de mora que no corresponden. Para esto es importante, educar a los encargados de llenar la herramienta, así también no debe de ser de conocimiento de la parte comercial cuales son las variables que más pesan en el modelo para evitar posibles manipulaciones de la información.
- 5) Se recomienda una revisión periódica del modelo, el periodo dependerá de cada institución financiera, para este modelo debe ser cada 6 meses, ya que esto permitirá saber si el perfil de la población ha cambiado (Backtesting), con lo que será necesario calibrar el modelo, esto es, re-estimar los Betas y volver a correr el modelo a la población. Esto con el objetivo de que no pierda poder de predicción.
- 6) Si bien al momento de colocar los CUT – OFF, se ha creado un intervalo de aprobación y de rechazo, se recomienda contar con un intervalo de “Duda” de tal forma que, Los “Rechazados” ya no llegan al analista, Los “Duda” son revisados por el analista pudiendo salvar algunos casos en base a su expertise y Los “Aprobados” continúan su proceso para la aprobación definitiva. El objetivo de esto es ir afinando la herramienta, es decir, que el criterio experto sea identificado en la herramienta, con lo que en el futuro la “Duda” disminuya.

BIBLIOGRAFÍA

- 1) Anderson, Raymond. The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation. Oxford University Press Inc. 2007
- 2) Cabrer, Beranardi, Amparo Sancho y Guadalupe Serrano. Microeconometría y Decisión. Ediciones Pirámide. 2001
- 3) Fisher, Ronald. The Use of Multiple Measurements in Taxonomic Problems, Ann. Eugenics. 1936.
- 4) Mays, Elizabeth. Credit Scoring For Risk Managers, The Handbook for Lender. Thomson. 2004
- 5) Pairazaman, Roberto, "El Crédito de Consumo en el Perú", CIUP, 2004
- 6) Siddigi, Naeem. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. SAS Institute Inc. 2006
- 7) Siegel, Sidney. Estadística No Paramétrica. Editorial Trillas. 1991
- 8) Thomas, Lyn, David Edelman y Jonathan Crook. Credit Scoring and Its Applications. SIAM. 2002
- 9) Altman, Edward, Anthony Saunders. Credit risk measurement: Developments over the last 20 years. Journal of Banking & Finance 21. 1998.

- 10) Gutiérrez, Matías. Modelos de Credit Scoring: Qué, Cómo, Cuándo y Para Qué. BCRA, 2007
- 11) Mesía, Mario, Eduardo Costa, Oscar Graham, Robert Soto y Alex Rabanal. El Costo del Crédito en el Perú revisión de la Evolución Reciente. Documentos de Trabajo BCRP. 2004
- 12) Van Vuuren, Gary, Krishnan Ramadurai. How much Credit in Credit Risk Models?. Fitchrating. 2007
- 13) Norma 18.10.017 Sobre Tarjetas de Crédito BBVA Banco Continental, Abril 2001
- 14) Resolución S.B.S. N° 6941 – 2008 Reglamento de Sobre endeudamiento, Agosto 2008
- 15) Tabachnick y Fidell (1996), <http://www.tdx.cesca.es/TDX-1006103-132336>
- 16) Asociación de Bancos del Perú, www.asbanc.com.pe
- 17) Banco de Central de Reserva del Perú www.bcrp.gob.pe
- 18) Superintendencia de Banca, Seguros y AFP www.sbs.gob.pe

ANEXO 1

1 Los Créditos de Consumo en el Perú

1.1 Definición

Los créditos de consumo se definen como un préstamo de corto o mediano plazo que se otorga para obtener dinero de libre disposición, el cual pueden ser manejados de acuerdo a sus ingresos mensuales. Existen dos clases de consumidores, uno de ellos es el que piensa que su dinero tiene un costo de oportunidad y el otro es el que sus ingresos no son suficientes para satisfacer sus necesidades, estas dos razones son los que motivan a las personas naturales a solicitar un crédito de consumo.

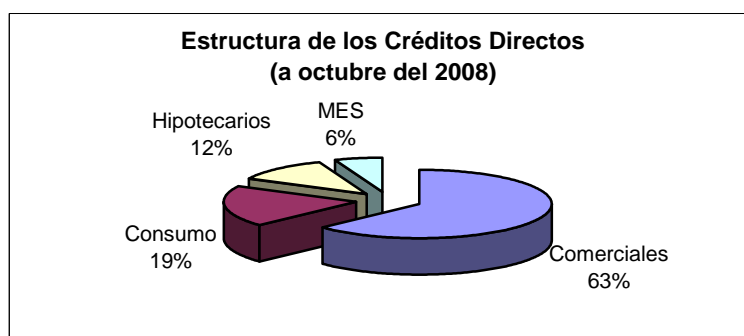
Los plazos para esta modalidad de préstamos, en general fluctúan entre 1 y 3 años, sin embargo, en nuestro país esto se ha ido extendiendo hasta llegar a ofrecer préstamos hasta en un plazo de 5 años, esto con el fin de que las cuotas a pagar sean más accesibles para los deudores, la contrapartida de esto es que el tener un plazo largo hace que las personas terminen pagando puro intereses y la amortización represente un monto mínimo de cada cuota. Sin embargo, cabe mencionar que según un estudio realizado por la empresa Inmark¹⁵ un 41% de los peruanos toma créditos personales para pagarlos hasta un plazo de 12 meses, mientras que un 27% lo cancela en un plazo entre uno y dos años.

En el gráfico 1.1 se puede observar la composición de los créditos directos en el Perú, en donde el 19% del total es representado por los créditos de consumo con una cifra de S/ 16,844 millones y con un crecimiento respecto a Octubre del 2007 del 46%. Tal como es de conocimiento público, los créditos directos están presentando un crecimiento sostenido, principalmente explicado por el

¹⁵ Estudio realizado por la empresa Inmark

buen desempeño de la economía peruana a pesar de la crisis internacional que actualmente se vive. De los tipos de crédito el que presenta un mayor dinamismo es el de créditos de consumo, esto a pesar de las medidas aplicadas por el BCRP quien a buscado desacelerar este crecimiento, se considera que este crecimiento del consumo se explica por contar con una demanda interna en alza, así como en una mayor confianza en el consumidor peruano (sus expectativas son positivas).

GRÁFICO 1.1



Fuente: Asbanc
Elaboración: Propia

1.2 Historia¹⁶

Haciendo un poco de historia, en la década de los años 90s, el sistema financiero del Perú solo atendían créditos comerciales y muy pocas veces colocaban créditos personales y si lo hacían eran solo para a aquellos sectores de niveles socioeconómicos A Y B.

A finales del año 1994 el sistema financiero tiene un rumbo diferente y empieza a fijarse en los créditos de consumo pero dirigidos a sectores socioeconómicos más bajos de nuestro país que son C y D.

¹⁶ Pairazaman, Roberto, “El Crédito de Consumo en el Perú”, CIUP, 2004

Dichos sectores se encontraban en total abandono, pero muy ansiosos por adquirir bienes y servicios, es por este motivo que los grandes bancos como (Banco de Crédito, Interbank, Banco de Trabajo, entre otros) empezaron a crear financieras para atender dichos sectores.

Pero no fue sencillo, porque las entidades financieras cobraban altas tasas de interés el cual generaba una alta morosidad en el pago, esto debido que las personas de crédito no contaban con un historial Crediticio.

Este esquema de penetración de mercado (productos masivos, estandarizados, poco flexibles y elevados precios) resultó eficaz cuando existe un mercado totalmente desatendido y ansioso por consumir, es decir, cuando el negocio esta en sus inicios. Sin embargo, con el transcurrir del tiempo, las condiciones van cambiando, los clientes poco a poco, ganan una posición en el mercado y logran obtener cierto poder de negociación.

Esto generó que los usuarios del consumo se sobre endeudaran, provocando que la cartera pesada se incremente de una manera abismal. Sin embargo, la presencia de entidades calificadoras de riesgo como Infocorp, Certicom ha permitido un mayor control, de igual forma le ha permitido a las entidades financieras reducir el problema de asimetría de la información que enfrentan, en sus dos formas: selección adversa (ex - ante) cuando la entidad financiera no puede distinguir entre clientes con diferentes niveles de riesgo, con lo que podría seleccionar a personas con alto riesgo de incumplimiento o riesgo moral (ex – post) cuando la persona a la que se le dio el crédito pueda desconocer el compromiso asumido. Todo esto se ha venido reflejando en una baja de la mora en consumo.

1.3 Principales Productos disponibles

En el mercado existen variedades de productos, pero los más importantes son: Libre Disponibilidad, Vehicular y Tarjetas de Crédito. En lo que respecta a Libre Disponibilidad, es un préstamo lo cual permite al usuario disponer de dinero en efectivo. Pero con la finalidad de utilizarlo para financiar compra de bienes y/o servicios, en algunos casos el usuario lo utiliza para satisfacer sus necesidades de efectivo. En el caso de Vehiculares, este tipo de sistema de crédito financia la compra de vehículos nuevos o usados, principalmente se financian autos nuevos, ya que los usados son financiados mayormente como préstamos de libre disponibilidad, esto debido a que en un auto usado esta de por medio una tasación cuyo costo usualmente es transferido a el cliente, desanimándolo de la compra mediante esta modalidad por el costo que esto implica. Estos créditos son amortizados en cuotas fijas mensuales. Dentro de esta modalidad de crédito el bien queda prendado. A nombre de la entidad financiera como garantía, hasta el momento que el acreedor cancele el integro del crédito vehicular. La tercera modalidad es La Tarjeta de Crédito, que es un medio de pago cuyas características se pasa a detallar a continuación.

1.3.1 Características Tarjetas de Crédito

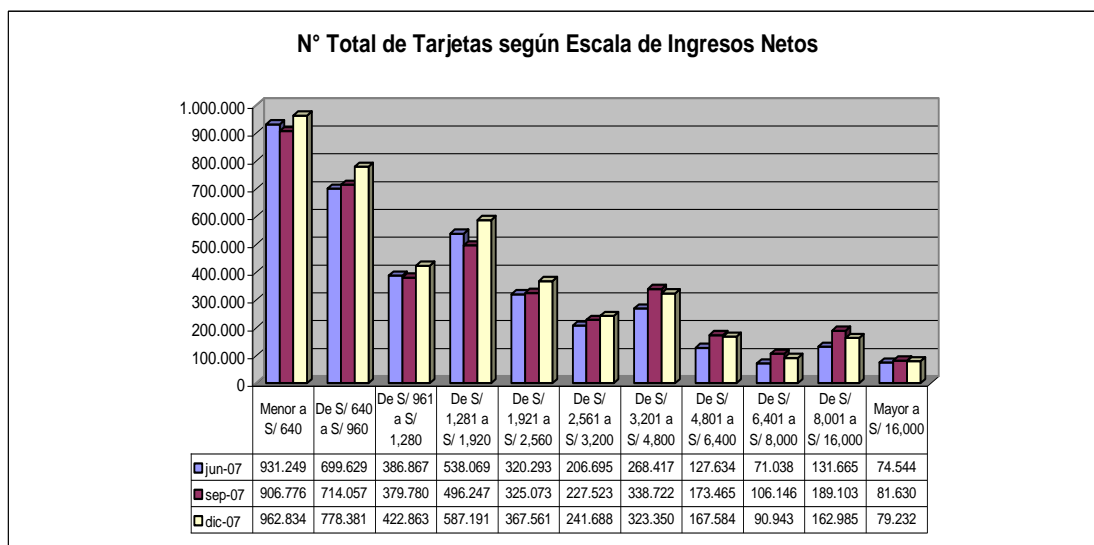
- Son un medio de pago internacional o nacional, que permite adquirir bienes y/o servicios en una gran cantidad de establecimientos afiliados a la marca emisora
- Se emiten por un plazo determinado (fecha de caducidad). Fecha de caducidad no mayor a lo establecido por el Reglamento de la Superintendencia de Banca y Seguros; y es la misma tanto para la tarjeta titular del contrato como para los adicionales (si los tuviera).
- Todas las Tarjetas de crédito cuentan con una clave secreta (número de cuatro dígitos), que le permite efectuar según el tipo de tarjeta, compras y operaciones de manera segura. En el extranjero, dependerá de los procedimientos establecidos en el País.

1.4 Sobre endeudamiento

Se muestra unos cuadros que si bien es cierto refleja información a diciembre del 2007, develan un problema potencial que se puede desencadenar en una alta morosidad para el sistema financiero. Se ha tomado información que corresponde a tarjetas de crédito, que es lo que interesa por el tema que se ha desarrollado.

En el gráfico 1.2 se puede ver como se distribuyen el número de tarjetas por nivel de ingresos, lo que debería llamar poderosamente la atención, es que por los menos el 46% de los plásticos se concentra en las personas con ingresos de hasta S/ 960.

GRAFICO 1.2



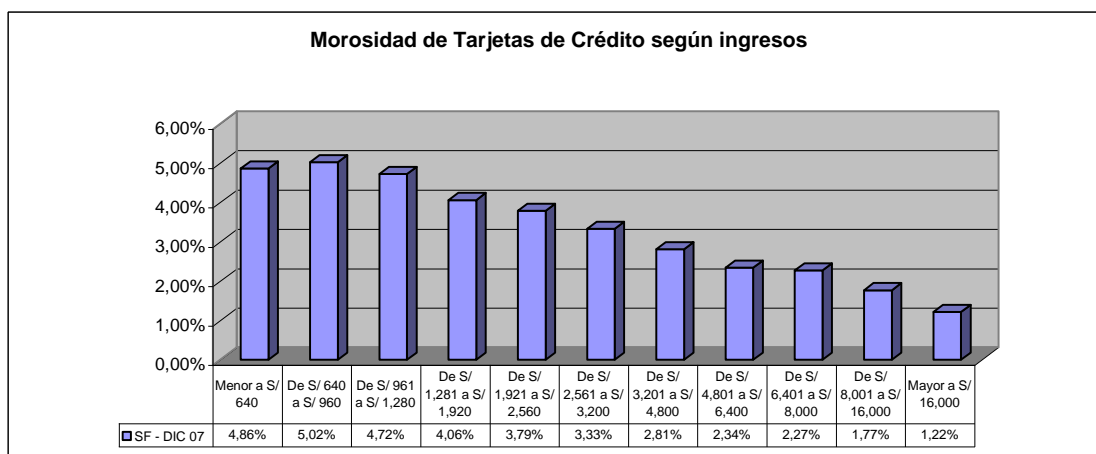
Menor a S/ 640	De S/ 640 a S/ 960	De S/ 961 a S/ 1,280	De S/ 1,281 a S/ 1,920	De S/ 1,921 a S/ 2,560	De S/ 2,561 a S/ 3,200	De S/ 3,201 a S/ 4,800	De S/ 4,801 a S/ 6,400	De S/ 6,401 a S/ 8,000	De S/ 8,001 a S/ 16,000	Mayor a S/ 16,000
24,8%	18,6%	10,3%	14,3%	8,5%	5,5%	7,1%	3,4%	1,9%	3,5%	2,0%
24,1%	19,0%	10,1%	13,2%	8,7%	6,1%	9,0%	4,6%	2,8%	5,0%	2,2%
25,6%	20,7%	11,3%	15,6%	9,8%	6,4%	8,6%	4,5%	2,4%	4,3%	2,1%

Fuente: Asbanc

Elaboración: Propia

A continuación se muestra el gráfico 1.3 que muestra como se reparte la morosidad según el nivel de ingresos.

GRAFICO 1.3



Fuente: Asbanc

Elaboración: Propia

Como se puede ver la mayor morosidad se concentra en los 3 niveles de más bajos ingresos, lo cual si que es preocupante y que es a nuestro entender lo que ha motivado a la SBS a publicar el reglamento.

Queda claro, que los distintos bancos e instituciones financieras en por lo menos estos dos últimos años se han abocado en dar tarjetas de crédito en los segmentos C y D, también llamados segmentos de bajo valor. Sin embargo, al parecer estos no estarían respondiendo como se esperaba, por otro lado, una pregunta que habría que hacerse, es si los bancos contaban o cuentan con herramientas de evaluación adecuadas para segmentos de bajo valor.

Definitivamente ha sido de nuestro interés poder elaborar un Scoring de tarjetas para los segmentos de bajo valor, pero las limitaciones que hemos encontrado para acceder bases de datos de estos segmentos no nos ha permitido esto.

ANEXO 2

2.1 Historia del Credit Scoring

Mientras la historia de los créditos se remonta a 5000 años atrás en la antigua Babilonia, la historia de los Credit Scoring tiene tan solo 50 años de antigüedad. Un Credit Scoring es esencialmente un método para diferenciar grupos de población cuando no se pueden ver claramente las características que definen los diferentes grupos. El primer método para resolver este problema de identificación de grupos en una población fue introducido en estadística por Fisher¹⁷ (1936), lo utilizó para resolver un problema de biología. En 1941, Durand fue el primero en reconocer que se pueden utilizar las mismas técnicas para discriminar entre clientes solventes e insolventes frente a un préstamo, este trabajo lo desarrolló como un proyecto de investigación para el U.S. National Bureau of Economic Research pero no estaba enfocado con una intención predictiva, solo de clasificación.

La primera empresa en elaborar un Credit Scoring fue creada por Bill Fair y Earl Isaac en San Francisco a principios de la década de los 50, y sus clientes eran principalmente empresas de ventas por correo y las denominadas casas de finanzas.

La llegada de las tarjetas de crédito a finales de los 60, hace que las entidades bancarias y otras empresas emisoras de tarjetas se planteen la utilidad de los Credit Scoring, pero el número de personas a las que diariamente se les adjudicaba una tarjeta de crédito era tan alto que era imposible en términos económicos y de mano de obra aplicar un método de decisión de crédito, y es

¹⁷ Fisher, R, The use of multiple measurements in taxonomic problems, Ann. Eugenics. 1936

solo el desarrollo de los ordenadores lo que hace factible esta aplicación. Las entidades que deciden aplicar Credit Scoring buscan que estos sean el mejor método de predicción y que los ratios de impago disminuyan al menos en un 50% ó más.

En la década de los 80, el éxito de la aplicación de los Credit Scoring en la concesión de tarjetas de crédito hizo que los bancos comenzaran a utilizarlos para otros productos, como préstamos personales y en los últimos años de esta década para préstamos hipotecarios y pequeñas empresas. En los 90, se comenzó a aplicar en marketing para la clasificación del ratio de respuestas a las campañas publicitarias. De hecho, este fue uno de los primeros usos en los años 50, cuando Sears utilizó un Scoring para decidir a quien enviaba sus catálogos.

2.2 Papel del Credit Scoring en los Créditos de Consumo

En los últimos 19 años se ha producido una revolución en la industria de los servicios financieros, las instituciones financieras se han dedicado a buscar modelos que den velocidad a las decisiones de aprobar o no un préstamos y la gestión del riesgo de crédito (los prestamistas se han apresurado en adoptar las decisiones automatizadas), teniendo los Credit Scoring una participación preponderante en los préstamos de consumo, así como también en los préstamos hipotecarios y préstamos para pequeñas empresas. Podríamos resumir en tres puntos la importancia de esta herramienta:

- Promueve la eficiencia y gran ahorro de tiempo en el proceso de aprobación del préstamo, con lo cual se reduciría el costo de evaluación del crédito.
- Reduce la subjetividad en el proceso de aprobación del préstamo. El criterio humano esta sujeto de error, asimismo este puede variar entre las personas (puede haber préstamos que para un analista es viable, mientras que para otro no lo es).

- Una herramienta experta como el Scoring, puede tomar en cuenta más variables que un humano podría considerar e interrelacionarlas para la decisión de aprobación.

2.2.1 Uso del Credit Scoring en el Otorgamiento de Préstamos

Hoy en día, la mayoría de prestamistas cuentan con sus propios modelos personalizados de Credit Scoring, esto es un score construido para un producto específico, este usualmente contendrá las características sobre la base de la solicitud de crédito, tales como ingresos del prestatario o relación deuda-ingreso, estado civil, edad u otros. También existen los Scoring genéricos, que son herramientas que no se centran en un producto específico, ni tampoco en un tipo de prestatario en particular, sino en datos que provienen de la experiencia de pago de un grupo de instituciones financieras.

2.2.2 Scoring de Gestión de Cuenta¹⁸

Los Scoring también pueden ser usados en cada etapa del ciclo de vida de un préstamo de consumo. A estos modelos también se les conoce como Scoring comportamental, este se distingue de los Scoring que sirven para seleccionar a un cliente, en que incluye las características que representan al deudor del préstamo en relación a como ha venido pagando su préstamo. Se toman en cuenta variables como el número de veces que un prestatario ha tenido retrasos de pago, la gravedad del retraso, entre otros. Debido a que incluye variables relacionadas con el prestatario demostrando la voluntad y la capacidad de pago del préstamo en cuestión, el Scoring comportamental tiende a ser aún más predictivo que el Scoring inicial para otorgar un préstamo, que, por supuesto, se basan únicamente en los datos disponibles.

Según indica la autora Mays, el Scoring comportamental se utilizan también para venta cruzada de productos a los actuales prestatarios. Por ejemplo, un

¹⁸ Mays, Elizabeth. Credit Scoring For Risk Managers, The Handbook for Lender. Thomson. 2004

prestamista puede utilizar una puntuación de comportamiento generados por su cartera hipotecaria para seleccionar los prestatarios que podrían acceder a una oferta de tarjetas de crédito.

El Scoring comportamental también es usado para estimar la probabilidad de que la totalidad o una parte de una deuda incobrable se recuperarán. Después de que un préstamo ha caído en serios retrasos, el Scoring comportamental se puede generar para clasificar los préstamos por la probabilidad de que algunos de ellos se podrán cobrar. Esta herramienta va a permitir a los prestamistas escoger con que cartera de cobranza se pueden quedar para gestionarlas ellos mismos y que cartera debe de ser cedida a las empresas que se dedican a la cobranza de este tipo de deudas.

2.2.3 Nuevos Métodos de Credit Scoring¹⁹

En los últimos años muchas técnicas para la elaboración de Credit Scoring se han probado. La regresión logística es una técnica estadística muy popular en la construcción de scorecard. Hay estudios realizados en relación a usos de herramientas no estadísticas utilizadas para la elaboración de scorecard, tales como: redes neuronales, algoritmos genéticos, programación lineal entre otros.

Una serie de estudios se han llevado a cabo para determinar si alguna de estas técnicas de hacer Scoring es la mas poderosa. Aunque la mayoría de los métodos que se han utilizado han arrojado resultados aceptables, ninguno de ellos ha demostrado ser una técnica superior a las demás en lo que respecta a la clasificación del riesgo de crédito de los prestatarios. Sin embargo, habría que mencionar que algunas técnicas pueden tener ventajas sobre otras.

Por ejemplo, técnicas de regresión tienen una ventaja sobre redes neuronales, ya que hacen más fácil de interpretar y explicar el efecto de las características

¹⁹ Thomas, Lyn, David Edelman y Jonathan Crook. Credit Scoring and Its Applications. SIAM. 2002

de la variable explicada, ya que es fácil determinar el efecto incremental de un determinado cambio en cada una de las características. Redes neuronales, por otra parte, se han vuelto bastante común su uso en modelos de fraude.

Actualmente los Credit Scoring están siendo utilizados para predecir tasas de incumplimiento de una cartera de crédito e incluso para estimar la rentabilidad de un préstamo individual. Los prestamistas están dispuestos a asumir más riesgo y ofrecer productos que no lo hubieran hecho en el pasado, en la medida que se sientan cómodos con el riesgo que involucra ello. Por lo tanto, la comprensión del riesgo por parte de los prestamistas requiere que estos tengan muy en claro cuanto es la rentabilidad que podrían obtener por cada unidad de riesgo que están dispuestos a asumir.

Últimamente se han venido desarrollando nuevas técnicas que puedan vincular la rentabilidad con cada préstamo individual, esto es relativamente nuevo para el mundo de los Credit Scoring. Existen dos trabajos que se han venido realizando en relación a esto:

La primera es el modelo de cadena de Markov, el autor Thomas considera como una alternativa a los modelos de Scoring comportamental y como una forma de estimar la rentabilidad de un único préstamo. El modelo Markov predice la probabilidad de como un préstamo pasará de un estado particular a otro en cada período de tiempo futuro.

La idea es segmentar la cartera en grupos homogéneos de los préstamos con similares características de riesgo de manera que la probabilidad de pasar de un estado particular es el mismo para todos los préstamos en el grupo. Ciertamente, los préstamos debe ser segmentado por su actual estado (con retrasos de pago o no), otros importantes impulsores o indicadores de resultados futuros se puede encontrar mediante el uso de árboles de decisión.

Un segundo trabajo que puede ser útil para la estimación de la rentabilidad es el Análisis de Supervivencia (survival analysis), este ha tenido una amplia aplicación en otros campos, tales como la bio-médica. Es una técnica de regresión como la regresión logística pero en lugar de predecir la probabilidad que un evento ocurra, predice el tiempo que transcurrirá hasta que el evento ocurra. Con este enfoque, se prevé la probable performance de los préstamos en el futuro (generalmente con una periodicidad de un mes), desde su origen hasta un punto dado en el tiempo.

Se señala que el análisis de supervivencia tiene una serie de ventajas sobre el modelo de regresión logística para la construcción de modelos de Scoring. Una de ellas es que evita la inestabilidad provocada por tener que elegir un periodo determinado como 18 meses para determinar si un préstamo es bueno o malo. Ciertas características pueden llegar a dominar un sistema de puntuación en los primeros 18 meses de la vida de un préstamo, mientras que otros pueden tener una mayor influencia posterior en el plazo de préstamo.

Con el análisis de supervivencia también se puede incorporar el efecto de los cambios en la economía durante la vida del préstamo. Además de las características del prestatario, las variables económicas pueden ser muy importantes predictores de los préstamos que van mal.

2.2.4 El Futuro de los Credit Scoring

Los próximos años prometen ser dinámicos en el ámbito de los Credit Scoring. Los gestores de riesgo se están convirtiendo más y más a las técnicas cuantitativas para entender y gestionar los riesgos, sus carteras. El Credit Scoring está ganando cada vez más terreno, y nuevas y mejores técnicas están siendo ejercidas cada año. Cada vez más, los prestamistas están encontrando en casa la experiencia para construir y usar modelos que reconocen las ventajas competitivas que pueden obtenerse de la utilización de

primera mano de la información, como del modelo. Los prestamistas que reconocen estas ventajas y el uso temprano de las herramientas expertas verán mayores beneficios para su institución.

2.3 Modelos de Scoring Genéricos y Personalizados²⁰

Desde que se introdujeron los modelos genéricos a mediados de los años 80, el crecimiento en su uso y aceptación ha sido asombroso. Actualmente, las decisiones de crédito se ven más afectados por modelos genéricos que por modelos personalizados de Scoring, a pesar de que muchos acreedores usan ambos métodos. Los modelos de Scoring personalizado se han desarrollado para el uso de un prestamista en particular, el prestamista influye en el actual modelo de Scoring por participar en la toma de muestras, selección característica, y las decisiones de aplicación. Mientras que los modelos genéricos se venden en el mercado para su uso por múltiples instituciones, son generalmente sobre la base de una muestra de las experiencias pasadas de varios prestamistas.

Cabe precisar que la venta de modelos genéricos en el mercado se da principalmente en los EE UU. En nuestro país solo se desarrollan modelos personalizados, tenemos conocimiento que Equifax ofrece el servicio de score, pero que se los vende principalmente a instituciones o empresas (este se considera un modelo genérico).

Normalmente, un modelo personalizado se basa en los datos proporcionados por la experiencia de pago de préstamos de un solo de un prestamista, mientras que un modelo genérico se basa en los datos proporcionados por la experiencia de pago de préstamo de un grupo de prestamistas.

2.3.1 Modelos Genéricos Versus Modelos Personalizados

²⁰ Mays, Eizabeth. Credit Scoring For Risk Managers, The Handbook for Lender. Thomson. 2004

Conceptualmente, un Scoring personalizado debe ser más preciso que un genérico. Un Scoring personalizado es hecho a medida de los acreedores, tomado de la propia experiencia del pasado del prestamista y acorde a los objetivos del mismo. Sin embargo, hay situaciones en las que el desarrollo y la aplicación de un Scoring personalizado, o bien no son viables o no es la alternativa más apropiada. Es importante tomar en consideración los siguientes aspectos al momento de elegir uno de estos dos modelos:

Viabilidad

Pocas situaciones de crédito son absolutamente perfectas para el modelado. Por lo tanto, el equilibrio entre lo que sería ideal y lo que puede hacerse debe de considerarse en la decisión de escoger entre Scoring personalizados y genéricos.

- *La experiencia histórica del préstamo:* Dado que el desarrollo de un scorecard requiere el análisis de las últimas decisiones, el prestamista debe de haber ofrecido préstamos en el pasado. Por lo tanto, el no tener data histórica es igual a no poder desarrollar un Scoring personalizado. Idealmente, el Credit Scoring se debe utilizar para el mismo producto, área de negocios, medio ambiente económico que generó la experiencia histórica. Los modelos genéricos se basan en la información existente en los archivos de la oficina de créditos (se hace referencia a EEUU. En Perú sería el equivalente a las centrales de riesgos) que contienen una gran riqueza en lo que respecta a experiencia de crédito.
- *Retención de datos:* La información utilizada para apoyar decisiones adoptadas en el pasado debe haber sido conservado en una forma utilizable con el fin de construir un Scoring personalizado. Resulta necesario conservar la información por periodos largos de tiempo. Desde comienzos del decenio de 1990, las oficinas de crédito han archivado la totalidad de

sus informes. Estos registros se conservan durante largos períodos de tiempo y se utilizan para desarrollar y validar Scoring genéricos.

- *Conocer los resultados de decisiones anteriores:* Los resultados de decisiones anteriores deberá estar disponible en una forma cuantificable. La historia de pago puede utilizarse para clasificar los resultados entre buenos o malos préstamos. En este caso los modelos personalizados tienen en mayor detalle la información del historial de pago, que los modelos genéricos. Por lo tanto si se quisiera elaborar un Scoring Comportamental resulta más complejo la clasificación de resultados para un modelo genérico que para un personalizado.
- *Edad de la decisión:* Las decisiones deben tener edad suficiente para permitir la debida clasificación y medición de los resultados. Por ejemplo, los préstamos aprobados hace tres meses, no tienen la edad suficiente como para clasificarlos como buenos o malos, mientras que si estos cuenta con una antigüedad de 24 meses, es seguro que sí. En el otro extremo, operaciones desembolsadas hace 10 años resulta muy vieja como para ser usada al momento de discriminar entre buenos y malos, ya que no reflejarían los perfiles actuales. Según indica Mays: “Los desarrolladores de modelos deben de definir un periodo de tiempo para seleccionar una muestra que sea madura”.
- *Tamaño de la muestra:* El número de decisiones de crédito debe haber sido lo suficientemente grande como para permitir un adecuado tamaño de la muestra. A menudo al pedir una muestra, se prefiere por lo menos contar con 4500 casos (lo ideal sería 1500 buenos, 1500 malos y 1500 rechazados). Este problema se puede presentar con pequeños prestamistas cuando quieren desarrollar modelos personalizados. Las muestras disponibles para la elaboración de modelos genéricos, suelen ser enormes. Se utilizan muestra de decenas a cientos de miles de expedientes y en algunos casos de millones. Algunos de los primeros genéricos se

centraron en modelos de quiebra, ya que las instituciones no tenían tanta historia como para construir un modelo personalizado.

Hechos relacionados al desarrollo del modelo

Durante el desarrollo de cualquier modelo de Credit Scoring, se toman las decisiones que afectarán a su rendimiento y aplicación. Muchas de las diferencias entre modelos genéricos y personalizados son el resultado de la interacción Individual entre el prestamista con el desarrollador del modelo. Estas interacciones no se producen en el desarrollo de modelos genéricos.

- *Objetivo del modelo:* En el desarrollo de un modelo personalizado usualmente el prestamista selecciona el objetivo del modelo, así como la población objetivo. La idea es influir en las decisiones que van desde la definición de los resultados hasta su implementación. Los modelos genéricos tienden a tener objetivos generales que no pueden ser modificados por los prestamistas individuales. Los diferentes modelos genéricos tienen diferentes objetivos generales, los prestamistas podrán convenir con estos objetivos o buscarán sentirse cerca de ellos.
- *Población objetivo:* se refiere a los solicitantes a los que serán evaluados por el modelo. Para un modelo personalizado, los solicitantes que no se ajustan a la población objetivo pueden ser eliminado de la muestra de desarrollo. Los modelos genéricos sólo serán desarrollado para las grandes poblaciones. Si el prestamista se dirige a un grupo relativamente pequeño, un modelo personalizado puedes ser más apropiado.
- *Desarrollo de la muestra y datos:* Los prestamistas podrán conservar la experiencia histórica de crédito, ya sea en forma manual, tales como papel o en microficha, o automatizada a partir de ficheros informáticos, o ambas cosas. El desarrollo de cualquier modelo de Scoring exige que los datos estén en un ordenador. En los modelos genéricos los datos son totalmente informatizado. La base de datos es enorme, con millones de registros que

contienen todas las informaciones detalladas acerca de las investigaciones sobre el solicitante de crédito, cuentas de crédito y pagos. Sin embargo, no contiene la información demográfica detallada.

- *Definición de variable Independiente:* Las variables independientes son las características que determinan el valor de la puntuación del crédito, en un modelo personalizado para el nuevo solicitante, las variables independientes son generalmente tomadas de la solicitud de crédito y el informe Crediticio. En los modelos genéricos, las variables independientes son seleccionados por las agencias de crédito y sus desarrolladores, poco o nada es provista por los prestamistas. De hechos, las oficinas de crédito, no divulgan las definiciones exactas de las variables que has usado para elaborar sus modelos genéricos.
- *Definición de variable dependiente:* La variable dependiente es el resultado, la variable dependiente tradicional para un nuevo modelo de Scoring es si la performance de pago es buena o mala. En los modelos personalizados, los prestamistas pueden especificar las definiciones para cumplir sus objetivos. Incluso, esta posibilidad le permite a los prestamistas poder tener definiciones complejas de la variable dependiente. En el caso de los modelos genéricos, la definición de la variable dependiente depende de lo que se encuentre en los ficheros de deudores de la oficina de créditos.
- *El tiempo de desarrollo:* Se puede tomar de tres a 12 meses para desarrollar un modelo personalizado de Scoring. Su implementación involucra más tiempo, que van desde un mes al año. En los modelos genéricos ya los sistemas de Scoring se encuentran en el mercado (cabe aclarar que esto se da en los EEUU, en Perú aún no), están disponibles para su uso en un plazo relativamente breve. A veces, una necesidad del prestamista es tan inmediata que los modelos generales son la única alternativa viable.

Hechos relacionados con la implementación

El prestamista debe ser capaz de implementar con éxito el sistema de Scoring. La implementación puede ser tan importante como la exactitud de predicción de los modelos. Si no se puede implementar, no desarrollar o comprar.

- *Interpretación de la información y entrada de datos:* El costo de la entrada de datos es una función del número de solicitantes, la cantidad de información a introducir y la cantidad de la información que se tiene que interpretar. La exacta y coherente interpretación de la información puede ser bastante difícil. Si bien se puede lograr un alto nivel de precisión y coherencia para el desarrollo de una muestra, es más difícil y costoso obtener los mismos resultados en un ambiente de producción. Los modelos personalizados a menudo requieren de una amplia entrada de datos, los genéricos solo requieren de un ingreso mínimo (el documento de identificación).
- *Performance del pronóstico:* Es relativamente sencillo para los modelos personalizados calcular la performance de los pronósticos, el desarrollador obtiene los puntajes de una muestra de resultados conocidos que usualmente es la muestra de desarrollo, luego realiza lo mismo con una muestra de validación, o una muestra de un plazo determinado o región geográfica. En el caso de los modelos genéricos se basan en grandes muestras de archivos con definiciones de resultados estándar. En estos casos, los prestamistas no tienen la flexibilidad que tienen los que pueden contar con su propia data y sus propios modelos.
- *Monitoreo / Validación:* Es necesario realizar un monitoreo a los modelos, esto con el fin de saber si el poder predictivo de los modelos cambian, en la medida que haya cambios en las relaciones entre las variables. Asimismo, a fin de validar el desempeño de cualquier modelo, el score en el momento de la decisión del crédito debe de conservarse. Con un modelo personalizado, la retención es a menudo un subproducto natural del sistema. Esta

información puede fácilmente ser comparada con la performance de pagos de determinados resultados. El score, la información de entrada y los resultados están usualmente disponibles para el análisis. En el caso de los modelos genéricos se requiere de un esfuerzo adicional para conservar los score y poder combinarlos con los pagos.

- *Compartir experiencias y asesoramiento:* Dado que cada modelo de Scoring personalizado es único, los prestamistas no pueden discutir sus experiencias con otros que están utilizando el mismo modelo de puntuación. Sin embargo, los muchos prestamistas que utilizan exactamente el mismo modelo genérico puede compartir experiencias con el fin de aprender unos de otros. Además, las oficinas de crédito mantienen personal para asesorar sobre el uso de los modelos.

2.4 Modelos de Regresión

2.4.1 Regresión Logística frente a la regresión lineal

Para los modelos de Scoring, no se está hablando de un caso en que la variable dependiente toma una amplia gama de valores. Más bien estamos trabajando con una variable binaria, que toma sólo dos valores, por ejemplo, si un préstamo es impago o no (o si un consumidor respondió a un correo o no).

En estos casos normalmente se codifica a la variable dependiente con un valor de 1 para los préstamos que se han incumplido y 0 para aquellos que no lo han hecho. Lo que el modelo lineal podría estimar en el caso de una variable 0 / 1 es la probabilidad de que el préstamo será incumplido o el consumidor responderá. Aunque a veces la regresión lineal se utiliza para estimar los modelos de Scoring, la regresión logística es generalmente preferida porque está diseñado específicamente para el caso de que la variable dependiente es binaria.

Un problema con la regresión lineal es que puede predecir probabilidades que son superiores a 1 o inferiores a 0, lo que no tiene mucho sentido. En el modelo logístico tomando el logaritmo natural de las probabilidades, se previene la predicción de probabilidades negativas.

La regresión logística, dado que es una técnica estadística multivariante, es capaz de aislar el efecto independiente de cada X en las probabilidades de que el evento se produzca .

2.4.2 Escala de Puntuación y la asignación de los pesos a los puntos

Una vez estimado el modelo de regresión logística, se puede poner las variables X en valores para un determinado préstamo y calcular una puntuación usando la ecuación:

$$\text{SCORE} = B_1X_1 + \dots + B_nX_n$$

La puntuación resultante, sin embargo, presenta dificultad para interpretar la escala del logaritmo natural. Una práctica habitual en el mundo de Scoring es transformar el score a una escala lineal, donde un número determinado de puntos se traducirá en una duplicación de las probabilidades de que el evento va a suceder.

Para hacer la transformación, se multiplica el score por un factor igual al número de puntos para los que queremos duplicar las probabilidades de que ocurran, y luego dividir por $\ln(2)$.

Normalmente, un número determinado de puntos base, se añade al score. Todos los solicitantes recibirán este número de puntos además de los puntos relacionados con los valores de sus características particulares. La asignación de puntos de base se hace sólo para razones cosméticas para asegurarse de que la puntuación de la distribución tiene un rango deseable. Por ejemplo,

dependiendo de cómo las características del score se construyeron, el score podría tener valores negativos, que van desde, por ejemplo, -133 a 333. La asignación de puntos de base de 200 puntos dará posibles valores de 87 a 533.

2.5 Segmentación de la población²¹

La segmentación divide la población en grupos y permite construir un scorecard para cada grupo. Se han identificado tres tipos de razones para la segmentación de los scorecard: estratégicas, operacionales, y estadística.

Un prestamista puede querer segmentar su scorecard para orientarse estratégicamente hacia determinados segmentos de clientes, tales como los prestatarios que ya tienen un préstamo con el prestamista. Quiere un scorecard de este grupo porque quiere tratarlos de manera diferente.

Una razón operativa de segmentación surge cuando se dispone de diferentes datos para los diferentes segmentos de clientes. Por ejemplo, las diferentes solicitudes de préstamo pueden ser utilizados para la solicitud de los prestatarios en una sucursal bancaria, los teléfono en un call center, o las que se aplican a través de un sitio Web. Esto podría significar que ciertas características de predicción están disponibles para algunos segmentos del solicitante, pero no para otros, necesitando un scorecard segmentado.

Por último, una razón estadística de segmentación surge cuando las características que afecta a la variable resultado (dependiente) difieren para algunos subgrupos de la población que para otros.

²¹ Mays, Eizabeth. Credit Scoring For Risk Managers, The Handbook for Lender. Thomson. 2004

La segmentación puede o no dar lugar a una serie de scorecards que son más predictivos que un único cuadro de mando integral, esto está en función de cada situación en particular, debe quedar claro que la performance de la segmentación debe ser significativamente mejor que no hacerlo.

2.6 Análisis Discriminante

Se define $X = (X_1, X_2, \dots, X_p)'$ como el conjunto de p variables, o también llamadas características, aleatorias que describen la información disponible en una valoración para la asignación de crédito. El valor actual de las variables para una aplicación particular se denota como $x = (x_1, x_2, \dots, x_p)'$. En la terminología de Credit Scoring, los diferentes valores posibles o respuestas, x_i , de la variable X_i se denominan atributos de una característica. Así, si una característica típica es el tipo de vivienda, entonces sus atributos serían: propia, familiar, propia con hipoteca, alquilada, u otros. Distintas entidades pueden tener distintos grupos de atributos de la misma característica.

Volviendo a la decisión que la entidad prestamista debe tomar, denotaremos por A al conjunto de todos los valores posibles que la variable $X = (X_1, X_2, \dots, X_p)'$ puede tomar, es decir, el conjunto de todas las posibles respuestas que se pueden obtener. El objetivo es encontrar una regla que divida el conjunto A en dos subconjuntos A_G y A_B que clasifica las aplicaciones cuyas respuestas están en A_G como buenos y aceptarlas mientras que aquellas cuyas respuestas están en A_B son malos y rechazarlas minimizando el coste esperado para el prestamista. Los dos tipos de coste corresponden con los dos tipos de error que pueden surgir de esta decisión. Se puede clasificar a alguien que es bueno como malo y rechazar a esta persona. Por lo tanto, incurrimos en un coste por rechazar la concesión de préstamos a clientes que hubieran sido solventes.

Asumimos por ahora, que el coste de no conceder estos préstamos es el mismo para cada uno de ellos y lo denotaremos por L . El segundo error es

clasificar un malo como bueno y aceptar el crédito. En este caso se incurriría en un débito cuando el cliente no pagara el crédito.

Se plantea asumir las siguientes hipótesis de trabajo

- La pérdida esperada por préstamo concedido es el mismo para todos los clientes y la denotaremos por D .
- Denotaremos por $p(G)$ (respectivamente, $p(B)$) la proporción de asignaciones que no producen impago (respectivamente, que producen impago).
- Las características de una asignación de crédito tiene un número finito y discreto de atributos, así A es finito y solo tiene un número finito de atributos diferentes x .

Denotaremos por $p(x|G)$ a la probabilidad de que tenga los atributos x teniendo una buena asignación de crédito, de forma similar, definiremos $p(x|B)$ como la probabilidad de que tenga los atributos x sabiendo que tiene una mala asignación de crédito. Si $p(G|x)$ denota la probabilidad de que tenga una buena asignación de crédito teniendo como atributos x , con $p(G)=p(\text{asignación buena})$ y $p(B)=p(\text{asignación mala})$, el Teorema de Bayes nos dice

$$p(G|x) = p(x|G) p(G) / p(x). \quad (2.1)$$

Un resultado similar lo encontramos para $p(x|B)$, la probabilidad de que un individuo tenga los atributos x sabiendo que tiene una mala asignación de crédito, y lo expresamos mediante

$$p(B|x) = p(x|B) p(B) / p(x). \quad (2.2)$$

Notamos que (2.1) y (2.2) se pueden expresar de la siguiente manera:

$$p(G|x) / p(B|x) = p(x|G) p(G) / p(x|B) p(B). \quad (2.3)$$

El coste de equivocarnos en la concesión si aceptamos las asignaciones con atributos en A_G y rechazamos aquellas con atributos en A_B es:

$$L \sum_{x \in A_B} p(x|G)p(G) + D \sum_{x \in A_G} p(x|B)p(B) = L \sum_{x \in A_B} p(G|x)p(x) + D \sum_{x \in A_G} p(B|x)p(x) \quad (2.4)$$

La regla que minimiza este coste esperado es sencilla. Considerando que L y D están dados, si tomamos un particular $x=(x_1, x_2, \dots, x_p)'$, o bien está en A_G o bien en A_B . Si está en A_G , entonces existe un coste si este incurre en impago, en cuyo caso el coste esperado es $Dp(x|B)p(B)$. Si x está clasificado en A_B existe un coste si este es solvente, y el coste esperado es $Lp(x|G)p(G)$. Así, se caracteriza la pertenencia de x en A_G mediante la desigualdad si:

$$D p(x|B) p(B) \leq L p(x|G) p(G)$$

De esta manera la regla de decisión que minimiza el coste esperado viene dada por:

$$A_G = \{x : D p(x|B) p(B) \leq L p(x|G) p(G)\} = \{x : (D/L) \leq p(x|G) p(G) / p(x|B) p(B)\} \\ = \{x : (D/L) \leq p(G|x) / p(B|x)\}. \quad (2.5)$$

Donde la última expresión la obtenemos a partir de (2.3).

Una crítica al criterio anterior es que el coste esperado depende de los valores de L y D , los cuales pueden no ser conocidos. Así, en lugar de minimizar el coste esperado, se podría intentar minimizar la probabilidad de cometer un tipo de error mientras mantenemos la probabilidad de cometer el otro tipo de error en un valor dado. En el contexto de la concesión de créditos, lo obvio es minimizar el nivel de impago mientras mantenemos el porcentaje de aplicaciones aceptadas en un valor dado. El último requerimiento es equivalente a mantener la probabilidad de rechazar buenas aplicaciones en algún nivel fijado.

Suponemos que queremos que el porcentaje de aplicaciones aceptadas (el ratio de aceptación) sea a . Entonces A_G debe satisfacer:

$$\sum_{x \in A_G} p(x) = \sum_{x \in A_G} p(x|G)p(G) + \sum_{x \in A_G} p(x|B)p(B) = a \quad (2.6)$$

Mientras que al mismo tiempo se minimiza el ratio de impago

$$\sum_{x \in A_G} p(x|B)p(B)$$

Si definimos para cada x e A ,

$$b(x) = p(x|B) p(B)$$

Entonces nuestro objetivo es encontrar el conjunto A_G tal que podamos resolver el programa

$$\min \sum_{x \in A_G} b(x) = \sum_{x \in A_G} (b(x)/p(x))p(x) \quad (2.7)$$

$$\text{Sujeto a } \sum_{x \in A_G} p(x) = a$$

Empleando el método de los multiplicadores de Lagrange, se puede ver que este debe ser el conjunto de atributos x , que satisfacen

$$b(x) / (p(x)) \leq c$$

Donde c es elegido tal que la suma de las $p(x)$ satisface la restricción de ser igual a "a". Entonces

$$\begin{aligned} A_G &= \{x : b(x) / (p(x)) \leq c\} = \{x : p(B|x) \leq c\} = \\ &= \{x : (1-c) / c \leq p(x|G) p(G) / (p(x|B) p(B))\}, \end{aligned} \quad (2.8)$$

Y donde la segunda desigualdad se obtiene a partir de las definiciones de $p(x)$ y $b(x)$.

Así la forma de la regla de la decisión bajo este criterio es la misma que en (2.5), para $L = c$ y $D = 1 - c$.

El mismo análisis se podría plantear asumiendo que las características de la asignación son continuas y no variables aleatorias discretas. La única diferencia sería que las funciones de distribución condicional $p(x|G)$, $p(x|B)$ son reemplazadas por funciones de densidad condicionales $f(x|G)$ y $f(x|B)$ y las sumatorias son reemplazadas por integrales. Así el coste esperado si se divide

el conjunto A en los conjuntos A_G y A_B y se aceptan solo aquellos que estén en A_G sería

$$L \int_{x \in A_B} f(x/G) p(G) dx + D \int_{x \in A_G} f(x/B) p(B) dx \quad (2.9)$$

Y la regla de decisión que minimiza es análoga a (2.5), y la expresamos como:

$$\begin{aligned} A_G &= \{x : D f(x|B) p(B) \leq L f(x|G) p(G)\} = \\ &= \{x : D p(B) / (L p(G) \leq f(x|G) / (f(x|B))\}. \end{aligned} \quad (2.10)$$

2.6.1 El caso de una Normal Univariante

Consideramos el caso posible más simple donde solo hay una variable característica continua X y su distribución condicional respecto a los buenos $f(x|G)$ es una normal con media μ_G y varianza σ^2 , mientras la distribución respecto a los malos es una normal de media μ_B y varianza σ^2 .

Entonces:

$$f(x|G) = (2\pi)^{-1/2} \exp(-(x - \mu_G)^2/2\sigma^2),$$

Y así la regla de (2.10) se transforma en:

$$\begin{aligned} f(x|G) / f(x|B) &= \exp(-(x - \mu_G)^2/2\sigma^2) / \exp(-(x - \mu_B)^2/2\sigma^2) = \\ &= \exp(-(x - \mu_G)^2 + (x - \mu_B)^2/2\sigma^2) \geq D p(B) / L p(G). \end{aligned} \quad (2.11)$$

Entonces

$$x(\mu_G - \mu_B) \geq (\mu_G^2 - \mu_B^2) / 2 + \sigma^2 \ln(Dp(B) / Lp(G)) \Leftrightarrow x \in A_G$$

2.6.2 El caso de una Normal Multivariante con covarianza común

Un ejemplo más realista lo tenemos cuando hay p características (variables) en la información obtenida para la aplicación y los resultados tanto de los buenos clientes como de los malos forman una distribución normal multivariante. Asumimos que la media de la distribución de los buenos es μ_G y la de los malos μ_B con una matriz de covarianzas común Σ . La correspondiente función de densidad en este caso es

$$f(\mathbf{x}|G) = (2\pi)^{-p/2} (\det \Sigma)^{-1/2} \exp(-(\mathbf{x} - \mu_G)' \Sigma^{-1} (\mathbf{x} - \mu_G) / 2) \quad (2.12)$$

donde $(\mathbf{x} - \boldsymbol{\mu}_G)$ es un vector columna con p filas y $(\mathbf{x} - \boldsymbol{\mu}_G)'$ denota su transpuesta. Siguiendo lo obtenido en (2.11), tenemos:

$$f(\mathbf{x}|G) / f(\mathbf{x}|B) \geq D p(B) / L p(G).$$

En consecuencia,

$$\mathbf{x}' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_G - \boldsymbol{\mu}_B) \geq (\boldsymbol{\mu}_G' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_G - \boldsymbol{\mu}_B' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_B) / 2 + \sigma^2 \text{Ln}(Dp(B) / Lp(G)) \quad (2.13)$$

La parte izquierda de la expresión (2.13) es una suma ponderada de los valores de las variables, que expresamos como, $x_1 w_1 + x_2 w_2 + \dots + x_p w_p$, mientras la parte derecha de la expresión es una constante. Así, la expresión (2.13) nos lleva hacia una regla lineal de Scoring, la cual es conocida como una función lineal discriminante.

En el ejemplo anterior se asume que las medias y las covarianzas de la distribución son conocidas. Este caso se da raramente, y es más normal reemplazarlas por los estimadores, conocidos como medias muestrales, y que denotaremos por \mathbf{m}_G y \mathbf{m}_B y la matriz de covarianzas muestrales que denotaremos por \mathbf{S} . Entonces la regla de decisión (2.13) quedaría expresada por:

$$\mathbf{x}' \mathbf{S}^{-1} (\mathbf{m}_G - \mathbf{m}_B) \geq [(\mathbf{m}_G' \mathbf{S}^{-1} \mathbf{m}_G - \mathbf{m}_B' \mathbf{S}^{-1} \mathbf{m}_B) / 2] + \sigma^2 \text{Ln}(Dp(B) / Lp(G)) \quad (2.14)$$

2.6.3 El caso de una Normal Multivariante con diferentes matrices de covarianzas

Otra restricción obvia en el caso anterior es que las matrices de covarianzas son iguales para la población de buenos y malos clientes. Suponemos que la matriz en la población de buenos clientes es $\boldsymbol{\Sigma}_G$ y para los malos $\boldsymbol{\Sigma}_B$. En este caso (2.14) se convierte en:

$$f(\mathbf{x}|G) / f(\mathbf{x}|B) \geq Dp(B) / Lp(G) \quad (2.15)$$

Entonces,

$$\exp\left\{-1/2\left[(x-\mu_G)' \Sigma_G^{-1}(x-\mu_G) - (x-\mu_B)' \Sigma_B^{-1}(x-\mu_B)\right]\right\} \geq Dp(B)/Lp(G) \quad (2.16)$$

Es decir,

$$\left[x'(-\Sigma_G^{-1} + \Sigma_B^{-1})x + 2x'(\Sigma_G^{-1}\mu_G - \Sigma_B^{-1}\mu_B)\right] \geq (\mu_G' \Sigma_G^{-1}\mu_G - \mu_B' \Sigma_B^{-1}\mu_B) + 2Ln(Dp(B)/Lp(G))$$

(2.17)

Nótese que el lado izquierdo de la anterior desigualdad es una forma cuadrática en las variables x_1, x_2, \dots, x_p . Esta parece ser una regla de decisión más general y se podría esperar que este desarrollo fuese mejor que la regla lineal. En la práctica, sin embargo, se tiene que estimar el doble de parámetros, esto es, Σ_G y Σ_B . La incertidumbre extra que envuelve a estos estimadores hace que la regla de decisión cuadrática sea menos robusta que la lineal. Esto se puede confirmar en los trabajos de Reichert, Cho y G.M. Wagner.

2.6 Análisis discriminante: Separación en dos grupos

En el trabajo original de Fisher (1936), en la que se introducía la función lineal discriminante, la clave era encontrar la combinación de variables que mejor separaran dos grupos cuyas características eran observables. Estos dos grupos podían ser subespecies de una planta y las características son las medidas físicas, o podía ser quien sobrevive o sucumbe a algún trato traumático y las características son las respuestas a varias pruebas. En el contexto del Credit Scoring, los dos grupos son clasificados por el prestamista en buenos y malos clientes y las características son la información de la oficina de crédito y la aplicación de otros detalles.

Consideramos $Y = w_1 X_1 + w_2 X_2 + \dots + w_p X_p$ una combinación lineal de las características $\mathbf{X} = (X_1, X_2, \dots, X_p)'$. Parece obvio que medir la separación es como medir cuánto de diferentes son los valores de las medias de Y para los dos diferentes grupos de buenos y malos en la muestra. Así, buscamos

elegir los valores de w_i , de forma que $\sum_i w_i=1$, y maximicen la diferencia entre $E[Y|G]$ y $E[Y|B]$. El valor que obtenemos es similar si los datos de cada grupo están dispersos, que si están reagrupados uniendo los buenos y los malos clientes previamente cuando queremos discutir su separación y este segundo método facilita el cálculo y por lo tanto se procede de esta forma.

Fisher sugirió que si asumimos que los dos grupos tienen varianzas comunes, entonces una medida sensible de separación es:

$M = (\text{distancia entre las medias muestrales de ambos grupos} / \text{varianza muestral de ambos grupos})^{1/2}$

Se divide entre la raíz cuadrada de la varianza de la muestra para hacer la medida escalar independiente, esto es, si se transforma la variable Y en cY , entonces la medida de M no cambia.

Al igual que en la anterior sección denotaremos las medias muestrales por m_G y m_B para los clientes buenos y malos, respectivamente, y S denotará la varianza muestral común. Entonces la distancia de separación correspondiente M sería:

$$M = w(m'_G - m'_B / w' S w)^{1/2} \quad (2.18)$$

Diferenciando M con respecto a w y tomando la derivada igual a cero, obtenemos que el valor de M sea mínimo cuando

$$(m'_G - m'_B / w' S w)^{1/2} - ((w'(m_G - m_B))(S w) / w' S w) = 0$$

Es decir,

$$(m_G - m_B)(w' S w) = (w'(m_G - m_B)) S w \quad (2.19)$$

El hecho de que la segunda derivada de M con respecto a w sea una matriz definida positiva garantiza que el valor obtenido es efectivamente un mínimo.

$$(w' S w) / (w'(m_G - m_B))$$

Es un escalar λ de forma que

$$w = \lambda(S^{-1}(m_G - m_B)) \quad (2.20)$$

Estos valores son los mismos que hemos obtenido en la expresión (2.14), solo que esta vez no hemos asumido una distribución normal. Este es el mejor separador de los buenos y malos clientes bajo el criterio de que no conocemos sus distribuciones. Este resultado se obtiene para todas las distribuciones de varianza finita, porque la medida de la distancia M considera solamente la media y la varianza de las distribuciones, y esta da el mismo resultado para todas las distribuciones con la misma media y varianza.

2.7 Análisis discriminante: Una forma de regresión lineal

Es otro modelo de Credit Scoring, que también llega a una función lineal discriminante, es la regresión lineal. En este modelo se trata de encontrar la mejor combinación lineal de las características

$$w_0 + w_1X_1 + w_2X_2 + \dots + w_pX_p = (\mathbf{w}^*)' \mathbf{X}^*$$

Que explique la probabilidad de impago. Siendo

$$\mathbf{w}^* = (w_1, w_2, \dots, w_p)', \quad \mathbf{X}^* = (1, X_1, X_2, \dots, X_p)'$$

Así, si p_i es la probabilidad de que la aplicación i en la muestra tenga impago, se quiere encontrar la mejor aproximación de \mathbf{w}^* tal que

$$p_i = w_0 + w_{i1}X_1 + w_{i2}X_2 + \dots + w_{ip}X_p \quad \text{para todo } i \quad (2.21)$$

Suponemos una muestra con n elementos tal que $n_G + n_B = n$, donde n_B son malos, para facilitar la notación asumimos que son los primeros de la muestra y así $p_i = 1$ para $i=1, \dots, n_B$. Los restantes n_G de la muestra con $i=n_G+1, \dots, n_G + n_B$ son buenos, así para ellos $p_i = 0$. Buscamos los coeficientes w tales que minimicen el error cuadrático medio. Este viene dado aproximadamente por:

$$\sum_{i=1}^{n_B} \left(1 - \sum_{j=0}^p w_j x_{ij}\right)^2 + \sum_{i=n_B+1}^{n_B+n_G} \left(\sum_{j=0}^p w_j x_{ij}\right)^2 \quad (2.22)$$

Si para obtener resoluciones explicativas consideramos los valores

esperados muestrales iguales a los valores poblacionales y denotamos por \mathbf{S} a la matriz de covarianzas de la muestra, se tendrá que

$$X_G' X_G + X_B' X_B = nS + n_G m_G m_G' + n_B m_B m_B' \quad (2.23)$$

Finalmente obtenemos lo siguiente

$$S\mathbf{w} = c(m_G + m_B) \quad (2.24)$$

Donde: $c = n_G n_B (1 - (m_G - m_B)' \mathbf{w})$

Así, (2.24) da la mejor elección de $\mathbf{w}=(w_1, w_2, \dots, w_p)$ para los coeficientes de la regresión lineal. Este es el mismo \mathbf{w} que obteníamos en (2.20) con la función lineal discriminante. Este método muestra, que se pueden obtener los coeficientes del Credit Scoring por el método de los mínimos cuadrados de la regresión lineal.

Hemos considerado en la ecuación de la regresión (2.21), que en el lado izquierdo de esta, los buenos clientes tienen valor 1 y los malos valor 0.

Esto da un conjunto de constantes, que calificamos como $\mathbf{w}(1,0)^*$. Si tomamos otros valores tales que los buenos clientes quedan a la izquierda de \mathbf{g} y los malos a la izquierda de \mathbf{b} , entonces los coeficientes en la regresión $\mathbf{w}(\mathbf{g},\mathbf{b})^*$ difieren solo en el término constante w_0 tal que:

$$\mathbf{w}(\mathbf{g},\mathbf{b})^* = \mathbf{b} + (\mathbf{g} - \mathbf{b})\mathbf{w}(1,0)^* \quad (2.25)$$

ANEXO 3

3.1 Análisis Univariante

3.1.1 Variables Cualitativas

CUADRO 1

Sexo

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Masculino	3331	62.4	62.4	62.4
	Femenino	2003	37.6	37.6	100.0
	Total	5334	100.0	100.0	

CUADRO 2

Estado Civil

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Soltero	3137	58.8	58.8	58.8
	Casado	1854	34.8	34.8	93.6
	Viudo	40	.7	.7	94.3
	Divorciado	45	.8	.8	95.2
	Separado Judicial	3	.1	.1	95.2
	Conviviente	255	4.8	4.8	100.0
	Total	5334	100.0	100.0	

CUADRO 3

Continuidad Laboral

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	F	4656	87.3	87.3	87.3
	I	557	10.4	10.4	97.7
	O	119	2.2	2.2	100.0
	T	2	.0	.0	100.0
	Total	5334	100.0	100.0	

CUADRO 4**Cliente Banco**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Cliente del Banco	3163	59.3	59.3	59.3
	No cliente del Banco	2171	40.7	40.7	100.0
	Total	5334	100.0	100.0	

CUADRO 5**Profesión**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	AA	16	.3	.3	.3
	AB	8	.1	.1	.4
	AD	17	.3	.3	.8
	AF	118	2.2	2.2	3.0
	AH	2	.0	.0	3.0
	AI	204	3.8	3.8	6.8
	AJ	11	.2	.2	7.0
	AK	177	3.3	3.3	10.4
	AL	1	.0	.0	10.4
	AO	3	.1	.1	10.4
	EA	163	3.1	3.1	13.5
	EB	316	5.9	5.9	19.4
	EC	48	.9	.9	20.3
	ED	395	7.4	7.4	27.7
	EE	73	1.4	1.4	29.1
	EF	1729	32.4	32.4	61.5
	EG	40	.7	.7	62.3
	EH	150	2.8	2.8	65.1
	EI	71	1.3	1.3	66.4
	EJ	45	.8	.8	67.2
	EN	26	.5	.5	67.7
	EZ	1602	30.0	30.0	97.8
	OD	2	.0	.0	97.8
	OE	54	1.0	1.0	98.8
	OF	48	.9	.9	99.7
	OZ	15	.3	.3	100.0
	Total	5334	100.0	100.0	

El significado de cada componente de la variable profesión es el siguiente:

AA: Médico, Dentista, Psicólogo	ED: Rentista
---------------------------------	--------------

AB: Abogado	EE: Vendedor, Comisionista
AD: Arquitecto, Ingeniero	EF: Encargado, Capataz
AF: Veterinario	EG: Obrero Especializado
AH: Artista, Deportista	EH: Obrero
AI : Supervisor	EI : Empleado Público
AJ: Profesor	EJ: Militar, Seguridad
AK: Analista	EN: Asistente
AL: Programador	EZ: Mecánico, Electricista
AO: Empleado Privado	OD: Economista
EA: Gerente, Alto Cargo, Ejecutivo	OE: Otros
EB: Técnico Superior	OF: Jubilado
EC: Mando Intermedio	OZ: Suboficial

CUADRO 6

Pago de Haberes Banco

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Si	1614	30.3	30.3	30.3
	No	3720	69.7	69.7	100.0
	Total	5334	100.0	100.0	

CUADRO 7

Procedencia

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Oficinas	2242	42.0	42.0	42.0
	Canales	3092	58.0	58.0	100.0
	Total	5334	100.0	100.0	

3.2 Análisis Bivariante

3.2.1 Variables Cualitativas

Variable Sexo: Observamos que los hombres presentan mayor porcentaje de morosidad respecto a las mujeres. Por otro lado, en la prueba de Chi cuadrado ($\chi^2=12.939$, corrección de continuidad de Yates) con un $p=0.000$, y para un nivel de significancia de 0.25, rechazamos H_0 , entonces decimos que la variable SEXO presenta asociación con MORA.

CUADRO 8

Tabla de contingencia Sexo * MORA

		MORA			
		Normal	Mora	Total	
Sexo	Masculino	Frecuencia esperada	3031.9	299.1	3331.0
		% de Sexo	89.9%	10.1%	100.0%
	Femenino	Frecuencia esperada	1823.1	179.9	2003.0
		% de Sexo	92.9%	7.1%	100.0%
Total		Frecuencia esperada	4855.0	479.0	5334.0
		% de Sexo	91.0%	9.0%	100.0%

CUADRO 9

Pruebas de chi-cuadrado

	Valor	gl	Sig. asíntótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	13.298 ^b	1	.000		
Corrección por continuidad ^a	12.939	1	.000		
Razón de verosimilitudes	13.695	1	.000		
Estadístico exacto de Fisher				.000	.000
Asociación lineal por lineal	13.295	1	.000		
N de casos válidos	5334				

^a. Calculado sólo para una tabla de 2x2.

^b. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 179.87.

Variable Estado Civil: En esta variable observamos celdillas con frecuencias esperadas menores a 5; para la agrupación de las categorías nos fijaremos en el nivel (porcentaje) de morosidad de cada una de ellas.

CUADRO 10

Tabla de contingencia Estado Civil * MORA

		MORA			
			Normal	Mora	Total
Estado Civil	Soltero	Frecuencia esperada	2855.3	281.7	3137.0
		% de Estado Civil	89.9%	10.1%	100.0%
	Casado	Frecuencia esperada	1687.5	166.5	1854.0
		% de Estado Civil	93.9%	6.1%	100.0%
	Viudo	Frecuencia esperada	36.4	3.6	40.0
		% de Estado Civil	90.0%	10.0%	100.0%
	Divorciado	Frecuencia esperada	41.0	4.0	45.0
		% de Estado Civil	95.6%	4.4%	100.0%
	Separado Judicial	Frecuencia esperada	2.7	.3	3.0
		% de Estado Civil	100.0%	.0%	100.0%
	Conviviente	Frecuencia esperada	232.1	22.9	255.0
		% de Estado Civil	83.9%	16.1%	100.0%
Total		Frecuencia esperada	4855.0	479.0	5334.0
		% de Estado Civil	91.0%	9.0%	100.0%

Agencia de Estadística y Censos del Perú

- Grupo 1: Casado, Divorciado, Separado
- Grupo 2: Soltero, Viudo
- Grupo 3: Conviviente

CUADRO 11

Tabla de contingencia Estado Civil * MORA

			MORA		Total
			Normal	Mora	
Estado Civil	Grupo 1	Recuento	1786	116	1902
		% de Estado Civil	93.9%	6.1%	100.0%
	Grupo 2	Recuento	2855	322	3177
		% de Estado Civil	89.9%	10.1%	100.0%
	Grupo 3	Recuento	214	41	255
		% de Estado Civil	83.9%	16.1%	100.0%
Total		Recuento	4855	479	5334
		% de Estado Civil	91.0%	9.0%	100.0%

Por otro lado, en la prueba de Chi cuadrado ($\chi^2=40.224$) obtenemos un $p=0.000$ y para un nivel de significancia de 0.25, rechazamos H_0 , entonces decimos que la variable Estado Civil presenta asociación con MORA.

CUADRO 12

Pruebas de chi-cuadrado

	Valor	gl	Sig. asíntotica (bilateral)
Chi-cuadrado de Pearson	40.224 ^a	2	.000
Razón de verosimilitudes	39.566	2	.000
Asociación lineal por lineal	39.438	1	.000
N de casos válidos	5334		

^a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 22.90.

Variable Profesión: En esta variable observamos también categorías vacías o con porcentajes bajos respecto al nivel de morosidad.

CUADRO 13

Tabla de contingencia Profesión * MORA

% de Profesión		MORA		
		Normal	Mora	Total
Profesión	AA	93.8%	6.3%	100.0%
	AB	100.0%		100.0%
	AD	94.1%	5.9%	100.0%
	AF	83.1%	16.9%	100.0%
	AH	100.0%		100.0%
	AI	91.2%	8.8%	100.0%
	AJ	72.7%	27.3%	100.0%
	AK	96.0%	4.0%	100.0%
	AL	100.0%		100.0%
	AO	100.0%		100.0%
	EA	92.0%	8.0%	100.0%
	EB	93.7%	6.3%	100.0%
	EC	95.8%	4.2%	100.0%
	ED	93.7%	6.3%	100.0%
	EE	91.8%	8.2%	100.0%
	EF	87.9%	12.1%	100.0%
	EG	95.0%	5.0%	100.0%
	EH	87.3%	12.7%	100.0%
	EI	93.0%	7.0%	100.0%
	EJ	97.8%	2.2%	100.0%
	EN	88.5%	11.5%	100.0%
	EZ	92.9%	7.1%	100.0%
	OD	100.0%		100.0%
	OE	94.4%	5.6%	100.0%
	OF	91.7%	8.3%	100.0%
	OZ	80.0%	20.0%	100.0%
Total		91.0%	9.0%	100.0%

CUADRO 14

Tabla de contingencia Profesión * MORA

		MORA			
		Normal	Mora	Total	
Profesión	Grupo 1	Recuento	3064	221	3285
		% de Profesión	93.3%	6.7%	100.0%
	Grupo 2	Recuento	1791	258	2049
		% de Profesión	87.4%	12.6%	100.0%
Total		Recuento	4855	479	5334
		% de Profesión	91.0%	9.0%	100.0%

Donde:

- Grupo 1: AB, AH, AL, AO, OD, EJ, AK, EC, EG, OE, AD, AA, EB, ED, AI, EA, EE, EZ, EI, OF.
- Grupo 2: EN, EF, EH, AF, OZ, AJ

Mediante la prueba Chi cuadrado comprobamos que existe un grado de correlación entre ambas variables.

CUADRO 15

Pruebas de chi-cuadrado

	Valor	gl	Sig. asíntotica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	53.087 ^b	1	.000		
Corrección por continuidad ^a	52.372	1	.000		
Razón de verosimilitudes	51.553	1	.000		
Estadístico exacto de Fisher				.000	.000
Asociación lineal por lineal	53.077	1	.000		
N de casos válidos	5334				

a. Calculado sólo para una tabla de 2x2.

b. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 184.00.

Variable Continuidad Laboral: En esta variable observamos que el nivel de morosidad en cada categoría presenta porcentajes similares, excepto la última categoría T (temporal) que no presenta datos (porcentaje 0%) en la categoría MORA. Aun no deberíamos descartar esta variable puesto que al combinarla

con otra variable podría discriminar mejor respecto al nivel de morosidad, más adelante se verá con que variable se ha de combinar.

CUADRO 16

Tabla de contingencia Continuidad Laboral * MORA

		MORA			
		Normal	Mora	Total	
Continuidad Laboral	F	Frecuencia esperada	4237.9	418.1	4656.0
		% de Continuidad Laboral	91.0%	9.0%	100.0%
	I	Frecuencia esperada	507.0	50.0	557.0
		% de Continuidad Laboral	91.0%	9.0%	100.0%
	O	Frecuencia esperada	108.3	10.7	119.0
		% de Continuidad Laboral	91.6%	8.4%	100.0%
	T	Frecuencia esperada	1.8	.2	2.0
		% de Continuidad Laboral	100.0%	.0%	100.0%
Total	Frecuencia esperada	4855.0	479.0	5334.0	
	% de Continuidad Laboral	91.0%	9.0%	100.0%	

Variable Cliente Banco: Observamos que los que no son clientes del banco presentan mayor porcentaje de morosidad respecto a los que son clientes. Por otro lado, en la prueba de Chi cuadrado ($\chi^2 = 72.829$, corrección de continuidad de Yates) con un $p=0.000$, y para un nivel de significancia de 0.25, rechazamos H_0 , entonces decimos que la variable Cliente Banco presenta asociación con la variable MORA.

CUADRO 17

Tabla de contingencia

			MORA		Total
			Normal	Mora	
Cliente Banco	Cliente del Banco	Frecuencia esperada	2879.0	284.0	3163.0
		% de Cliente Banco	93.8%	6.2%	100.0%
	No cliente del Banco	Frecuencia esperada	1976.0	195.0	2171.0
		% de Cliente Banco	87.0%	13.0%	100.0%
Total	Frecuencia esperada		4855.0	479.0	5334.0
	% de Cliente Banco		91.0%	9.0%	100.0%

CUADRO 18

Pruebas de chi-cuadrado

	Valor	gl	Sig. asíntótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	73.663 ^b	1	.000		
Corrección por continuidad ^a	72.829	1	.000		
Razón de verosimilitudes	72.139	1	.000		
Estadístico exacto de Fisher				.000	.000
Asociación lineal por lineal	73.649	1	.000		
N de casos válidos	5334				

a. Calculado sólo para una tabla de 2x2.

b. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 194.96.

Variable Pago de Haberes: Observamos que los de la categoría No Pago de Haberes presenta mayor porcentaje de morosidad respecto a los que son Si Pagan Haberes. Por otro lado, en la prueba de Chi cuadrado ($\chi^2=89.900$, corrección de continuidad de Yates) con un $p=0.000$, y para un nivel de significancia de 0.25, rechazamos H_0 , entonces decimos que existe asociación entre las variables Pago de Haberes y Mora.

CUADRO 19

Tabla de contingencia

		MORA			
			Normal	Mora	Total
Pago de Haberes Banco	Si	Frecuencia esperada	1469.1	144.9	1614.0
		% de Pago de Haberes Banco	96.7%	3.3%	100.0%
	No	Frecuencia esperada	3385.9	334.1	3720.0
		% de Pago de Haberes Banco	88.6%	11.4%	100.0%
Total		Frecuencia esperada	4855.0	479.0	5334.0
		% de Pago de Haberes Banco	91.0%	9.0%	100.0%

CUADRO 20

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	89.886 ^b	1	.000		
Corrección por continuidad ^a	88.900	1	.000		
Razón de verosimilitudes	106.001	1	.000		
Estadístico exacto de Fisher				.000	.000
Asociación lineal por lineal	89.869	1	.000		
N de casos válidos	5334				

^a. Calculado sólo para una tabla de 2x2.

^b. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 144.94.

Variable Procedencia: Observamos que los de la categoría Canales presenta mayor porcentaje de morosidad respecto a la categoría Oficinas. Por otro lado, en la prueba de Chi cuadrado ($\chi^2 = 59.998$, corrección de continuidad de Yates) con un $p=0.000$, y para un nivel de significancia de 0.25, rechazamos H_0 , entonces decimos que existe asociación entre las variables Procedencia y Mora.

CUADRO 21

Tabla de contingencia

			MORA		Total
			Normal	Mora	
Procedencia	Oficinas	Frecuencia esperada	2040.7	201.3	2242.0
		% de Procedencia	94.6%	5.4%	100.0%
	Canales	Frecuencia esperada	2814.3	277.7	3092.0
		% de Procedencia	88.4%	11.6%	100.0%
Total	Frecuencia esperada		4855.0	479.0	5334.0
	% de Procedencia		91.0%	9.0%	100.0%

CUADRO 22

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	60.752 ^b	1	.000		
Corrección por continuidad ^a	59.998	1	.000		
Razón de verosimilitudes	64.163	1	.000		
Estadístico exacto de Fisher				.000	.000
Asociación lineal por lineal	60.741	1	.000		
N de casos válidos	5334				

a. Calculado sólo para una tabla de 2x2.

b. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 201.33.

CUADRO 23

Tabla de contingencia Sexo y Estado Civil * MORA

			MORA		Total
			Normal	Mora	
Sexo y Estado Civil	Soltera	Frecuencia esperada	1644.7	162.3	1807.0
		% de Sexo y Estado Civil	87.9%	12.1%	100.0%
	Casada	Frecuencia esperada	1170.5	115.5	1286.0
		% de Sexo y Estado Civil	93.9%	6.1%	100.0%
	Viudo	Frecuencia esperada	12.7	1.3	14.0
		% de Sexo y Estado Civil	85.7%	14.3%	100.0%
	Divorciado	Frecuencia esperada	10.0	1.0	11.0
		% de Sexo y Estado Civil	81.8%	18.2%	100.0%
	Conviviente	Frecuencia esperada	193.9	19.1	213.0
		% de Sexo y Estado Civil	83.6%	16.4%	100.0%
	Soltera	Frecuencia esperada	1210.6	119.4	1330.0
		% de Sexo y Estado Civil	92.6%	7.4%	100.0%
	Casada	Frecuencia esperada	517.0	51.0	568.0
		% de Sexo y Estado Civil	93.7%	6.3%	100.0%
	Viuda	Frecuencia esperada	23.7	2.3	26.0
		% de Sexo y Estado Civil	92.3%	7.7%	100.0%
	Divorciada	Frecuencia esperada	30.9	3.1	34.0
		% de Sexo y Estado Civil	100.0%	.0%	100.0%
	Separada Judicial	Frecuencia esperada	2.7	.3	3.0
		% de Sexo y Estado Civil	100.0%	.0%	100.0%
	Conviviente	Frecuencia esperada	38.2	3.8	42.0
		% de Sexo y Estado Civil	85.7%	14.3%	100.0%
Total		Frecuencia esperada	4855.0	479.0	5334.0
		% de Sexo y Estado Civil	91.0%	9.0%	100.0%

CUADRO 24

Tabla de contingencia Sexo y Estado Civil * MORA

		MORA			
		Normal	Mora	Total	
Sexo y Estado Civil	Grupo 1	Frecuencia esperada	2955.4	291.6	3247.0
		% de Sexo y Estado Civil	93.4%	6.6%	100.0%
	Grupo 2	Frecuencia esperada	1695.7	167.3	1863.0
		% de Sexo y Estado Civil	87.8%	12.2%	100.0%
	Grupo 3	Frecuencia esperada	203.9	20.1	224.0
		% de Sexo y Estado Civil	83.5%	16.5%	100.0%
Total	Frecuencia esperada	4855.0	479.0	5334.0	
	% de Sexo y Estado Civil	91.0%	9.0%	100.0%	

CUADRO 25

Pruebas de chi-cuadrado

	Valor	gl	Sig. asíntotica (bilateral)
Chi-cuadrado de Pearson	61.076 ^a	2	.000
Razón de verosimilitudes	58.157	2	.000
Asociación lineal por lineal	60.797	1	.000
N de casos válidos	5334		

^a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 20.12.

3.3 Regresión Logística Simple

3.3.1 Variables cualitativas.

CUADRO 26

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a SEXO_ESTCIVIL	.586	.076	59.051	1	.000	1.796
Constante	-3.204	.131	595.542	1	.000	.041

^a. Variable(s) introducida(s) en el paso 1: SEXO_ESTCIVIL.

CUADRO 27**VARIABLES EN LA ECUACIÓN**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a CONTLABO	-.026	.116	.049	1	.825	.975
Constante	-2.287	.141	261.133	1	.000	.102

a. Variable(s) introducida(s) en el paso 1: CONTLABO.

CUADRO 28**VARIABLES EN LA ECUACIÓN**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a PROFESION_COD	.692	.096	51.533	1	.000	1.997
Constante	-3.321	.154	462.656	1	.000	.036

a. Variable(s) introducida(s) en el paso 1: PROFESION_COD.

CUADRO 29**VARIABLES EN LA ECUACIÓN**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a CLIE_SN	.819	.097	70.652	1	.000	2.269
Constante	-3.537	.161	484.409	1	.000	.029

a. Variable(s) introducida(s) en el paso 1: CLIE_SN.

CUADRO 30**VARIABLES EN LA ECUACIÓN**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a CLIE_NOM	1.315	.148	79.311	1	.000	3.726
Constante	-4.679	.282	276.079	1	.000	.009

a. Variable(s) introducida(s) en el paso 1: CLIE_NOM.

3.3.2 Variables Cuantitativas**CUADRO 31****VARIABLES EN LA ECUACIÓN**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a EDAD	-.021	.005	17.487	1	.000	.979
Constante	-1.541	.187	67.599	1	.000	.214

a. Variable(s) introducida(s) en el paso 1: EDAD.

CUADRO 32**VARIABLES EN LA ECUACIÓN**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a NUNIFA1T	-.058	.041	2.019	1	.155	.943
Constante	-2.206	.090	606.273	1	.000	.110

a. Variable(s) introducida(s) en el paso 1: NUNIFA1T.

CUADRO 33**Variables en la ecuación**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a INGRTOTAL	.000	.000	3.295	1	.069	1.000
Constante	-2.190	.083	700.876	1	.000	.112

a. Variable(s) introducida(s) en el paso 1: INGRTOTAL.

CUADRO 34**Variables en la ecuación**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a GTOSFIJ	-.001	.000	10.898	1	.001	.999
Constante	-2.221	.054	1701.084	1	.000	.109

a. Variable(s) introducida(s) en el paso 1: GTOSFIJ.

CUADRO 35**Variables en la ecuación**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a CLIE_SMD	.000	.000	12.621	1	.000	1.000
Constante	-2.237	.050	1989.956	1	.000	.107

a. Variable(s) introducida(s) en el paso 1: CLIE_SMD.

CUADRO 36**Variables en la ecuación**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a CLIE_ANT	-.110	.014	64.192	1	.000	.896
Constante	-1.980	.058	1156.827	1	.000	.138

a. Variable(s) introducida(s) en el paso 1: CLIE_ANT.

3.4 Agrupación de la variable Saldo Medio en Cuenta e Ingresos Totales:

Se ha procedido a categorizar la variable *Saldo Medio*, observando el histograma y los percentiles de la variable, se tomo en cuenta el porcentaje de morosidad en cada intervalo. Por otro lado vemos que la prueba de Chi cuadrado nos dice que existe asociación entre ambas variables.

CUADRO 37

Tabla de contingencia Saldo media cuenta * MORA

		MORA			
			Normal	Mora	Total
Saldo media cuenta	Hasta 20 soles	Frecuencia esperada	2919.9	288.1	3208.0
		% de Saldo media cuenta	87.6%	12.4%	100.0%
	De 20 a 200 soles	Frecuencia esperada	633.5	62.5	696.0
		% de Saldo media cuenta	93.4%	6.6%	100.0%
	Más de 200 soles	Frecuencia esperada	1301.6	128.4	1430.0
		% de Saldo media cuenta	97.6%	2.4%	100.0%
Total	Frecuencia esperada	4855.0	479.0	5334.0	
	% de Saldo media cuenta	91.0%	9.0%	100.0%	

CUADRO 38

Pruebas de chi-cuadrado

	Valor	gl	Sig. asíntótica (bilateral)
Chi-cuadrado de Pearson	127.972 ^a	2	.000
Razón de verosimilitudes	152.734	2	.000
Asociación lineal por lineal	127.487	1	.000
N de casos válidos	5334		

^a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 62.50.

Para la variable *Ingresos Totales* se ha obtenido los siguientes intervalos, siguiendo el mismo criterio para la variable Saldo Medio en Cuenta.

CUADRO 39

Tabla de contingencia Ing. Totales * MORA

		MORA			
			Normal	Mora	Total
Ing. Totales	Hasta 2300 soles	Frecuencia esperada	3402.3	335.7	3738.0
		% de Ing. Totales	90.4%	9.6%	100.0%
	Más de 2300 soles	Frecuencia esperada	1452.7	143.3	1596.0
		% de Ing. Totales	92.4%	7.6%	100.0%
Total	Frecuencia esperada	4855.0	479.0	5334.0	
	% de Ing. Totales	91.0%	9.0%	100.0%	

CUADRO 40

Pruebas de chi-cuadrado

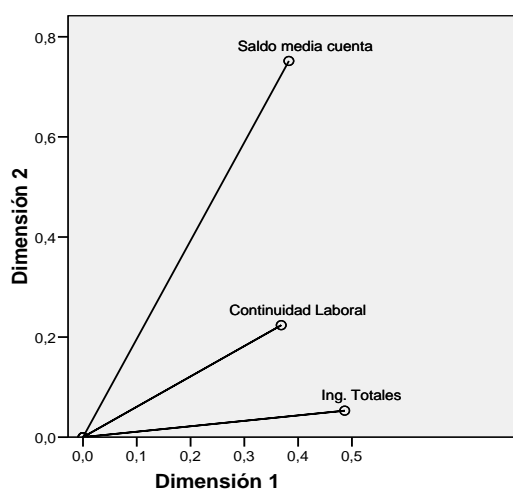
	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	5.451 ^b	1	.020		
Corrección por continuidad ^a	5.209	1	.022		
Razón de verosimilitudes	5.614	1	.018		
Estadístico exacto de Fisher				.021	.010
Asociación lineal por lineal	5.450	1	.020		
N de casos válidos	5334				

^a. Calculado sólo para una tabla de 2x2.

^b. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 143.32.

Al comprobar que existe cierta asociación entre las variables Continuidad Laboral, Ingresos Totales y Saldo Medio, ~~_____~~

Medidas de discriminación



Normalización principal por variable.

Podemos agrupar en la dimensión 1 a la variable Ingresos Totales y Continuidad Laboral y en la dimensión 2 a la variable Saldo Medio.

Agrupando las variables Ingresos y Continuidad Laboral y comparándola con la variable Mora, obtenemos los siguientes resultados:

CUADRO 41

Tabla de contingencia Ing. y Cont. Laboral * MORA

			MORA		Total
			Normal	Mora	
Ing. y Cont. Laboral	1	Frecuencia esperada	3060.1	301.9	3362.0
		% de Ing. y Cont. Laboral	90.2%	9.8%	100.0%
	2	Frecuencia esperada	264.9	26.1	291.0
		% de Ing. y Cont. Laboral	92.1%	7.9%	100.0%
	3	Frecuencia esperada	75.5	7.5	83.0
		% de Ing. y Cont. Laboral	91.6%	8.4%	100.0%
	4	Frecuencia esperada	1.8	.2	2.0
		% de Ing. y Cont. Laboral	100.0%	.0%	100.0%
	5	Frecuencia esperada	1177.8	116.2	1294.0
		% de Ing. y Cont. Laboral	93.0%	7.0%	100.0%
	6	Frecuencia esperada	242.1	23.9	266.0
		% de Ing. y Cont. Laboral	89.8%	10.2%	100.0%
	7	Frecuencia esperada	32.8	3.2	36.0
		% de Ing. y Cont. Laboral	91.7%	8.3%	100.0%
Total		Frecuencia esperada	4855.0	479.0	5334.0
		% de Ing. y Cont. Laboral	91.0%	9.0%	100.0%

Teniendo en cuenta el nivel de morosidad en cada categoría, finalmente nos quedaría:

CUADRO 42

Tabla de contingencia Ingresos Totales y Continuidad Laboral * MORA

			MORA		Total
			Normal	Mora	
Ingresos Totales y Continuidad Laboral	Grupo 1	Frecuencia esperada	1552.8	153.2	1706.0
		% de Ingresos Totales y Continuidad Laboral	92.7%	7.3%	100.0%
	Grupo 2	Frecuencia esperada	3302.2	325.8	3628.0
		% de Ingresos Totales y Continuidad Laboral	90.2%	9.8%	100.0%
Total		Frecuencia esperada	4855.0	479.0	5334.0
		% de Ingresos Totales y Continuidad Laboral	91.0%	9.0%	100.0%

CUADRO 43

Pruebas de chi-cuadrado

	Valor	gl	Sig. asíntotica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	8.990 ^b	1	.003		
Corrección por continuidad ^a	8.685	1	.003		
Razón de verosimilitudes	9.305	1	.002		
Estadístico exacto de Fisher				.002	.001
Asociación lineal por lineal	8.989	1	.003		
N de casos válidos	5334				

^a. Calculado sólo para una tabla de 2x2.

^b. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 153.20.

3.5 Regresión Logística Multivariante

Resultados previos del SPSS del modelo final,

CUADRO 44

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a SEXO_ESTCIVIL	.589	.090	42.907	1	.000	1.802
PROFESION_COD	.651	.113	33.460	1	.000	1.917
ANTEMP	-.052	.013	17.013	1	.000	.950
CLIE_NOM	.581	.226	6.589	1	.010	1.787
SMD_COD	-.749	.130	33.070	1	.000	.473
EDAD	.009	.007	1.903	1	.168	1.009
INGTOTAL_CONTILABO	.188	.131	2.061	1	.151	1.207
CLIE_ANT	-.073	.026	7.994	1	.005	.930
GTOSFIJ	.000	.000	3.003	1	.083	1.000
CLIE_SN	-.222	.188	1.406	1	.236	.801
Constante	-3.918	.774	25.634	1	.000	.020

^a. Variable(s) introducida(s) en el paso 1: SEXO_ESTCIVIL, PROFESION_COD, ANTEMP, CLIE_NOM, SMD_COD, EDAD, INGTOTAL_CONTILABO, CLIE_ANT, GTOSFIJ, CLIE_SN.

Para las variables categóricas se crearon las variables indicadoras Dummy, en la siguiente tabla se presentan sus respectivas codificaciones.

CUADRO 45

Codificaciones de variables categóricas

		Frecuencia	Codificación de parámetros	
			(1)	(2)
Sexo y Estado Civil	Grupo1	2607	.000	.000
	Grupo 2	1496	1.000	.000
	Grupo 3	183	.000	1.000
Saldo media cuenta	Hasta 20 soles	2565	.000	.000
	De 20 a 200 soles	576	1.000	.000
	Más de 200 soles	1145	.000	1.000
Profesión	Grupo 1	2641	.000	
	Grupo 2	1645	1.000	
Ingresos Totales y Continuidad Laboral	Grupo 1	1374	.000	
	Grupo 2	2912	1.000	
Pago de Haberes Banco	Si	1312	.000	
	No	2974	1.000	

3.6 GINI Muestra de Desarrollo y de Validación

3.6.1 Gini de Muestra de Desarrollo

CUADRO 46

Percentiles	Buenas	Morosas	Total	Acum. Buenas	Acum. Morosas	Acum. Total	% acum buenas	% acum mora	DIF % ACUM buenas, morosas
5%	214	0	214	214	0	214	5.49%	0.00%	5.49%
10%	212	2	214	426	2	428	10.93%	0.52%	10.41%
15%	212	2	214	638	4	642	16.38%	1.04%	15.33%
20%	209	5	214	847	9	856	21.74%	2.34%	19.40%
25%	205	9	214	1052	18	1070	27.00%	4.69%	22.31%
30%	206	8	214	1258	26	1284	32.29%	6.77%	25.52%
35%	205	9	214	1463	35	1498	37.55%	9.11%	28.44%
40%	205	9	214	1668	44	1712	42.81%	11.46%	31.35%
45%	207	7	214	1875	51	1926	48.13%	13.28%	34.85%
50%	198	16	214	2073	67	2140	53.21%	17.45%	35.76%
55%	203	12	215	2276	79	2355	58.42%	20.57%	37.85%
60%	198	16	214	2474	95	2569	63.50%	24.74%	38.76%
65%	193	21	214	2667	116	2783	68.45%	30.21%	38.25%
70%	193	21	214	2860	137	2997	73.41%	35.68%	37.73%
75%	190	24	214	3050	161	3211	78.29%	41.93%	36.36%
80%	181	33	214	3231	194	3425	82.93%	50.52%	32.41%
85%	174	40	214	3405	234	3639	87.40%	60.94%	26.46%
90%	172	42	214	3577	276	3853	91.81%	71.88%	19.94%
95%	158	56	214	3735	332	4067	95.87%	86.46%	9.41%
100%	161	52	213	3896	384	4280	100.00%	100.00%	0.00%

KS 38.76%

Coefficiente GINI: 50.6%

3.6.2 Gini de Muestra de Validación

CUADRO 47

Percentiles	Buenas	Morosas	Total	Acum. Buenas	Acum. Morosas	Acum. Total	% acum buenas	% acum mora	DIF % ACUM buenas, morosas
5%	50	2	52	50	2	52	5.35%	2.11%	3.24%
10%	51	0	51	101	2	103	10.80%	2.11%	8.70%
15%	51	0	51	152	2	154	16.26%	2.11%	14.15%
20%	51	0	51	203	2	205	21.71%	2.11%	19.61%
25%	50	2	52	253	4	257	27.06%	4.21%	22.85%
30%	49	2	51	302	6	308	32.30%	6.32%	25.98%
35%	46	6	52	348	12	360	37.22%	12.63%	24.59%
40%	47	4	51	395	16	411	42.25%	16.84%	25.40%
45%	48	4	52	443	20	463	47.38%	21.05%	26.33%
50%	47	4	51	490	24	514	52.41%	25.26%	27.14%
55%	52	0	52	542	24	566	57.97%	25.26%	32.70%
60%	44	7	51	586	31	617	62.67%	32.63%	30.04%
65%	49	3	52	635	34	669	67.91%	35.79%	32.12%
70%	47	4	51	682	38	720	72.94%	40.00%	32.94%
75%	47	5	52	729	43	772	77.97%	45.26%	32.70%
80%	40	11	51	769	54	823	82.25%	56.84%	25.40%
85%	41	11	52	810	65	875	86.63%	68.42%	18.21%
90%	45	6	51	855	71	926	91.44%	74.74%	16.71%
95%	45	7	52	900	78	978	96.26%	82.11%	14.15%
100%	35	17	52	935	95	1030	100.00%	100.00%	0.00%
KS									32.94%
Coefficiente GINI:									43.3%

NOTA BIOGRÀFICA

Victor Antonio David Tineo

Victor Antonio David Tineo (Lima, 1974), es Ingeniero Economista graduado de la Universidad Nacional de Ingeniería, tercio superior en los años de estudios del pre-grado. Asimismo, participó en el Curso de Extensión Universitaria organizado por BCRP, en el año 1999. Obtuvo el primer premio “Robert Maes” en el año 2006.

Actualmente labora en el BBVA Banco Continental, desempeñándose como Analista Senior del Área de Riesgos Personas. Cuenta con 7 años de experiencia en la evaluación de créditos de consumo, especialmente en los productos: Tarjetas de Crédito, Préstamos Vehiculares y Préstamos de Libre Disponibilidad.

Edwin Chuquipul Ruiz

Edwin Chuquipul Ruiz (Lima, 1970) Licenciado En Contabilidad graduado de la Universidad Nacional Federico Villarreal.