

UNIVERSIDAD ANDINA SIMÓN BOLÍVAR
SEDE ECUADOR

AREA DE GESTIÓN

PROGRAMA DE MAESTRÍA EN FINANZAS Y
GESTIÓN DE RIESGOS

“GESTIÓN DE RIESGO DE CRÉDITO DE LA
COOPERATIVA DE AHORRO Y CRÉDITO
MAQUITA CUSHUNCHIC – ANÁLISIS Y
PREPARACIÓN ESTADÍSTICA DE VARIABLES
PARA EL DISEÑO DE UN MODELO CREDIT
SCORE DE CARTERA DE CONSUMO”

MILTON EFRAÍN GUAMÁN GUANOPATÍN

QUITO – ECUADOR

2011

UNIVERSIDAD ANDINA SIMÓN BOLÍVAR
SEDE ECUADOR

AREA DE GESTIÓN

PROGRAMA DE MAESTRÍA EN FINANZAS Y
GESTIÓN DE RIESGOS

“GESTIÓN DE RIESGO DE CRÉDITO DE LA
COOPERATIVA DE AHORRO Y CRÉDITO
MAQUITA CUSHUNCHIC – ANÁLISIS Y
PREPARACIÓN ESTADÍSTICA DE VARIABLES
PARA EL DISEÑO DE UN MODELO CREDIT
SCORE DE CARTERA DE CONSUMO”

SUPERVISOR: EC. VERONICA ALVARRACIN

MILTON EFRAÍN GUAMAN GUANOPATÍN

QUITO – ECUADOR

2011

RESUMEN

El presente estudio busca analizar y preparar estadísticamente un conjunto de variables para el diseño de un modelo de aprobación CREDIT SCORE de cartera de consumo, tipo probabilístico, que apoye al oficial de crédito en la toma de la decisión antes de conceder o no un crédito de consumo; para que la decisión no sea subjetiva sino objetiva, medible (*probabilística*) apoyada en una ecuación que contenga sustento teórico y empírico dado por la base de datos histórica de la Cooperativa de ahorro y crédito Maquita Cushunchic.

En la presente tesis partimos con fundamentación teórica que sustenta nuestro trabajo de investigación.

En el segundo capítulo se realiza un diagnóstico situacional de la empresa, donde se define la visión, la misión y definición de temas estratégicos.

En el tercer capítulo realizaremos un análisis y preparación estadística de variables para el diseño de un modelo de aprobación CREDIT SCORE de cartera de consumo, esta fase involucra *“la identificación de las fuentes de información, la verificación de cantidad y calidad de los datos, la determinación y tipificación de variables cualitativas, cuantitativas y demográficas, así como la verificación y cuantificación del poder discriminante de dichas variables respecto al objetivo planteado”*¹, todo esto se lo realiza con el objetivo de aplicar un concepto de sistemas de calificación de crédito que se define de la siguiente manera: *“la calificación supone que el desempeño de los préstamos futuros con características dadas, será parecido al desempeño de los préstamos pasados con características similares”*². Lo que se busca finalmente es que el modelo sea una herramienta de apoyo empírico que ayude a la toma de decisiones antes de conceder o no un crédito.

Finalmente termina este trabajo formulando conclusiones y recomendaciones.

¹ Freddy Carranza, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008.

² Paul Noboa, Diapositivas de Riesgo de Crédito, Universidad Andina Simón Bolívar, Enero 2009.

AGRADECIMIENTOS

A Dios mi pilar, mi roca firme, mi refugio, mi fortaleza y mi guía, con su misericordia y bendición no me desampara ni de noche ni de día.

A mis amados padre y madre (+), hermanos y hermanas, cuñados y cuñadas, sobrinos y sobrinas, por su cariño, confianza, amor, respeto y apoyo incondicional.

A todas las personas que de una u otra manera han contribuido al desarrollo de esta tesis.

Milton E. Guamán G.

DEDICATORIA

A Dios, por regalarme la vida, la salud, el conocimiento, la sabiduría y sus bendiciones.

A la dulce memoria de mi madre Beatriz, por cuyos sacrificios y oraciones soy lo que soy. Su cariño, amor y bendiciones han contribuido a mi formación personal y profesional. El cielo debe verse mejor contigo. Siempre te extrañare...

Y a todos (*as*) los que cumplen sus objetivos y metas, sólo ellos pueden llegar a entender el significado de alcanzar estrellas, sólo ellos pueden llegar a entender el significado de luchar por alcanzar objetivos tamaño elefantes y no hormigas.

Milton E. Guamán G.

Al presentar esta tesis como uno de los requisitos previos para la obtención del título de magister, autorizo al Centro de Información de la Universidad para que haga de este trabajo un documento disponible para su lectura según las normas de la institución.

También cedo a la Universidad Andina Simón Bolívar, los derechos de publicación de este trabajo o de partes de ella, manteniendo mis derechos de autor hasta por un período de 30 meses contados después de su aprobación

Milton E. Guamán G

Quito, 2011

CONTENIDO

RESUMEN

CAPÍTULO I

MARCO TEORICO

1.1	Tipo de variables	1
1.1.1	Variables cualitativas	1
1.1.2	Variables cuantitativas	2
1.2	Análisis exploratorio de datos	3
1.2.1	Definiciones Básicas	3
1.3	Análisis estadístico	5
1.3.1	Análisis univariante	5
1.3.2	Análisis bivariante	24
1.4	Econometría	30
1.4.1	Definiciones básicas	30
1.4.2	Naturaleza del análisis de modelos de regresión	32
1.4.3	Modelos estadísticos	33
1.4.3.1	Análisis discriminante	33
1.4.3.2	Modelo lineal	34
1.4.3.3	Modelo logit	34

CAPÍTULO II

DIAGNÓSTICO SITUACIONAL DE LA INSTITUCIÓN

2.1	Descripción de la empresa	36
2.1.1	Antecedentes	36

2.1.2	Misión y visión	37
2.1.3	Gobernabilidad y estructura organizacional	37
2.1.4	Principales productos y servicios	38
2.1.5	Factores críticos de éxito	38
2.1.6	Contexto macro-económico	39
2.1.7	El sector de las microfinanzas	40
2.1.8	Posicionamiento en el mercado	41
2.1.9	Impacto social	42

CAPÍTULO III:

ANALIZAR Y PREPARAR ESTADÍSTICAMENTE UN CONJUNTO DE VARIABLES PARA EL DISEÑO DE UN MODELO DE APROBACION CREDIT SCORE DE CARTERA DE CONSUMO

3.1	Descripción de las variables	45
3.2	Definición de incumplimiento	46
3.3	Variable dependiente " <i>tipo pagador</i> "	47
3.4	Análisis de las Variables explicativas	49
3.5	Preparar variables para un modelo credit score	50
3.5.1	Modelo credit score	50
3.5.2	Análisis de variables	51
3.5.3	Ajuste al modelo logístico	54
3.6	Test del modelo credit score	60
3.6.1	Test de significancia del modelo	60
3.6.2	Test de significancia de ajuste de datos y coeficientes del modelo	62
3.6.3	Backtesting del modelo	67

3.6.4	Categorización y puntos de corte	68
-------	----------------------------------	----

CAPÍTULO IV

	CONCLUSIONES	70
	RECOMENDACIONES	70
	BIBLIOGRAFÍA	72

ANEXOS

ANEXO 1: Análisis exploratorio de variables cuantitativas y cualitativas del conjunto de variables entregadas por la institución

ANEXO 2: Análisis exploratorio bivalente de variables cuantitativas y cualitativas

ANEXO 3: Regresión logística de variables

ANEXO 4: Test del modelo credit score - backtesting

CAPÍTULO I: MARCO TEORICO

1.1 TIPO DE VARIABLES

La construcción de un modelo de credit score¹ de gestión de riesgo de crédito², parte con la solicitud del diccionario de datos a tecnología a priori a la preparación de la base de datos; la misma que a la postre se convertirá en información útil en el desarrollo del modelo de aprobación y calificación del crédito a clientes. Los tipos de variables que aportaran al modelo serán variables de tipo cualitativas y cuantitativas.

1.1.1 VARIABLES CUALITATIVAS³

Son las variables que expresan distintas cualidades, características o modalidad. Cada modalidad que se presenta se denomina atributo o categoría y la medición consiste en una clasificación de dichos atributos.

Las variables cualitativas pueden ser dicotómicas cuando sólo pueden tomar dos valores posibles como sí y no, hombre y mujer o son politómicas cuando pueden adquirir tres o más valores. Las variables que a menudo encontraremos se clasifican en cuatro⁴ categorías generales: escala de proporción, escala de intervalo, escala ordinal y escala nominal.

Escala de proporción: Para la variable X , al tomar dos valores (X_1 y X_2), la proporción X_1/X_2 y la distancia ($X_1 - X_2$) son cantidades que tienen sentido. Así mismo, existe un ordenamiento natural (*ascendente o descendente*) de los valores a lo largo de la escala. Por consiguiente, comparaciones

¹ Modelo de credit score.- Es un sistema de evaluación automático, más rápido, más seguro y consistente para determinar el otorgamiento de créditos, que, en función de toda la información disponible, es capaz de predecir la probabilidad de no pago, asociada a una operación crediticia. Ayuda en el proceso de evaluación, comportamiento y cobranza, siendo capaz de analizar en un tiempo mínimo, gran cantidad de información, en forma homogénea.

FUENTE: Credit Scoring una forma de medir el riesgo de crédito, fecha acceso: Enero 2011, dirección: <http://maxsilva.bligoo.com/content/view/185004/Credit-Scoring-una-forma-de-medir-el-riesgo-de-credito.html>

² Riesgo de Crédito.- Es el riesgo de que un cliente o contraparte no pueda o no quiera cumplir con un compromiso que ha celebrado con un miembro o miembros de una Institución. FUENTE: HSBC México (*HBMX*), Riesgo de Crédito, Gaceta de Basilea II, Edición Número 5, 2007.

³ Definiciones básicas de variables cualitativas, fecha acceso: Enero 2011, dirección: http://maralboran.org/wikipedia/index.php/Variables_estadisticas

⁴ GUJARATI Damodar, Introducción a la Econometría, Cuarta edición, McGraw Hill, 2004, Pág. 29 – 30.

como $X_2 \leq X_1$, o $X_2 \geq X_1$, tienen sentido. La mayoría de las variables económicas pertenecen a esta categoría. Por consiguiente, resulta sensato preguntar cuán grande es el PIB de este año, en comparación con el del año anterior?.

Escala de intervalo: Una variable de escala de intervalo satisface las dos últimas propiedades de la variable de escala de proporción, pero no la primera. Por tanto, la distancia entre dos periodos, digamos entre 2000 -1995, es significativa, no así la proporción entre dichos periodos ($2000/1995$).

Escala ordinal: Una variable pertenece a esta categoría solo si satisface la tercera propiedad de la escala de proporción (*es decir, el orden natural*). Los ejemplos son sistemas de calificaciones (A, B, C) o clases de ingresos (*alto, medio, bajo*). Para estas variables, existe el orden pero las distancias entre las categorías no pueden cuantificarse. Por ejemplo las curvas de indiferencia entre dos bienes, en donde una curva superior de indiferencia señala un mayor nivel de utilidad, pero no se puede cuantificar en qué medida una curva de inferencia es mayor que otra.

Escala nominal: Las variables de esta categoría no tienen ninguna de las características de las variables de escala de proporción, variables como el género (*masculino, femenino*) y el estado civil (*casado, soltero, divorciado, separado*) simplemente denota categorías. Pregunta: Cual es la razón por la que dichas variables no pueden expresarse mediante las escalas de proporción de intervalo o de orden?.

Las técnicas econométricas que pudieran ser adecuadas para las variables de escala de proporción, tal vez no resulten pertinentes para las variables de escala nominal. En consecuencia, resulten importante tener en mente las diferencias entre los cuatro tipos de escalas de medición: proporción, intervalo, ordinal, nominal.

1.1.2 VARIABLES CUANTITATIVAS⁵

Son las variables que se expresan mediante cantidades numéricas. Las variables cuantitativas además pueden ser: variable discreta y variable continua.

Variable discreta: Es la variable que presenta separaciones o interrupciones en la escala de valores que puede tomar. Estas separaciones o interrupciones indican la ausencia de valores entre los distintos valores específicos que la variable pueda asumir. Por ejemplo el número de lavadoras producidas por una empresa en un año (100, 20, 3476, 14), el número de hijos de una pareja (1, 2, 3, 4, 5), el número de pagos a realizar por un préstamo o en una compra a plazo de un artículo o un bien (12, 24, 120).

Variable continua: Es la variable que puede adquirir cualquier valor dentro de un intervalo especificado de valores. Por ejemplo la masa (2.3 kg, 2.4 kg, 2.5 kg) o la altura (1.64 m, 1.65 m, 1.66 m), la temperatura (35.53°C, 121.3°F) que solamente está limitado por la precisión del aparato medidor, en teoría permiten que siempre exista un valor entre dos variables.

Las variables cuantitativas generalmente van acompañadas por medidas descriptivas, que permiten tener información complementaria de la variable y su distribución en sí, expresando usualmente por diagrama de barras.

1.2 ANALISIS EXPLORATORIO DE DATOS

1.2.1 DEFINICIONES BÁSICAS⁶

Al llevar a cabo mediciones de un experimento o tener observaciones de un fenómeno, se obtiene una serie de datos que poseen más información que las simples cifras de que se disponen. El cómo

⁵ Definiciones básicas de variables cuantitativas, fecha acceso: Enero 2011, dirección: http://es.wikipedia.org/wiki/Variable_estad%C3%ADstica

⁶ GALINDO Edwin, Probabilidad y Estadística para ingeniería y administración, Centro de Matemática – Universidad Central, 2004, Pág. 1-4.

conseguir esta información y su análisis e interpretación puede realizarse de muchas maneras, pero lo primero es tener una idea clara de las características de los datos obtenidos.

Los datos pueden ordenarse en tablas, estas sin embargo no muestran su comportamiento global. Su representación gráfica permite captar fácilmente tendencias y establecer modelos probabilísticos. Conjuntamente con el empleo de métodos numéricos sencillos, permiten presentar datos, resumir información, dando una respuesta rápida del comportamiento de los fenómenos de donde ellos provienen.

Al examinar un fenómeno se tiene una variedad de información que podemos obtener de él, entonces es preciso determinar cuál es la de interés para fines que tengamos, así mismo se debe tener una idea del número y tipo de observaciones que serán necesarias para disponer de una información confiable; aquí surgen dos definiciones población y muestra.

Población.- El conjunto total de mediciones de interés para determinado problema constituye la población o universo.

Muestra.- El subconjunto de la población del cual se obtienen las observaciones mediante un experimento (*u otro proceso*) constituye la muestra.

La distinción entre población y muestra es una muestra, es un número limitado de mediciones de una fuente mayor; mientras que una población es una gran fuente de mediciones a partir de la cual se toma la muestra.

Por ejemplo se interesa conocer el "*rating*" de sintonía de los canales de televisión de la ciudad. La población lo constituye los hogares que poseen televisión y una muestra se podría determinar los hogares de los alumnos de una Universidad Pública.

1.3 ANALISIS ESTADISTICO

"El propósito de la estadística es descubrir métodos para condensar la información relativa a un gran número de hechos relacionados en cortas y compendiosas expresiones adecuadas para su discusión". Francis Galton

El análisis estadístico se ha asociado de manera general con la investigación de corte experimental, ha dejado de ser únicamente la ciencia de recopilar datos y, tras fusionarse con la corriente de estudios sobre el cálculo de probabilidades, se ha constituido en una rama de la matemática aplicada, entendiendo ésta como el uso de principios y modelos matemáticos en diversos ámbitos de la ciencia o la técnica.

Las técnicas y métodos estadísticos existentes son la estadística descriptiva y estadística inferencial.⁷

Estadística descriptiva: Son procedimientos dirigidos a la organización y descripción de un conjunto de datos.

Estadística inferencial: Se orientan a realizar inferencias sobre una población a partir de las características conocidas para una muestra extraída de ella.

La descriptiva tendría únicamente la finalidad de proporcionar los índices a partir de los cuales se estimarán los parámetros, entonces es frecuente que se otorgue especial importancia a la estadística inferencial, considerándola el verdadero objetivo de la estadística.

Si adoptáramos únicamente la complejidad como criterio, basándonos en el número de variables implicadas, podríamos hablar de análisis estadístico univariante y bivariante.

⁷ Definiciones básicas de análisis estadístico descriptivo, análisis estadístico inferencia, fecha acceso: Enero 2011, dirección: http://ocw.us.es/metodos-de-investigacion-y-diagnostico-en-educacion/analisis-de-datos-en-la-investigacion-educativa/Bloque_I/page_09.htm

1.3.1 ANÁLISIS UNIVARIANTE ⁸

El análisis univariante de los datos consiste en hacer una revisión del valor de un sólo indicador, por ejemplo, la edad, nivel de instrucción o el sexo. Suele usarse cuando interesa conocer en profundidad qué significan por separado algunos indicadores. Es utilizado con fines descriptivos en las primeras etapas de los estudios realizados.⁹

El análisis permite analizar el comportamiento de las variables por separado, sin cruzar información de otras variables. Busca seleccionar aquellos indicadores más discriminantes del caso para que vayan de acuerdo a la realidad.

Variables cualitativas: Una de las herramientas más empleadas para representar este tipo de variables es la distribución de frecuencias, que consiste en una tabla que presenta las categorías de una variable y sus repeticiones. Si tenemos k clases o categorías, se representan con letras minúsculas los datos $n_1, n_2, n_3, \dots, n_k$ que aparecen en cada categoría k , bajo estos elementos se define:

- Frecuencia absoluta de la clase i -ésima (n_i): número de observaciones en la clase i .
- Frecuencia relativa de la clase i -ésima (f_i): es la proporción de datos en la clase i -ésima, es decir:

$$f_i = \frac{n_i}{N}$$

- La suma de las k frecuencias relativas es igual a la unidad: $f_1 + f_2 + \dots + f_k = 1$

La distribución de frecuencias permite comparar las frecuencias de las categorías en conjuntos de datos con distinto número de observaciones. Para entender mejor este concepto lo podemos

⁸ El análisis Univariante fue tomado de la tesis: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008.

⁹ Definiciones básicas de análisis univariante, bivariante y multivariante, fecha acceso: Enero 2011, dirección: <http://www.serviciosestadisticos.es/metodologia/metodologia.htm>

explicar de la siguiente manera. En un conjunto de datos de 20 clientes, tomamos la variable nivel de estudios, la misma que pueden ser las siguientes categorías:



Las observaciones se encuentran codificadas de la siguiente manera:

1 1 4 3 3 3 2 2 4 2 2 1 4 2 3 2 3 4 2 3

Frecuencias absolutas:

$$n_1= 3; n_2= 7; n_3= 6; n_4= 4$$

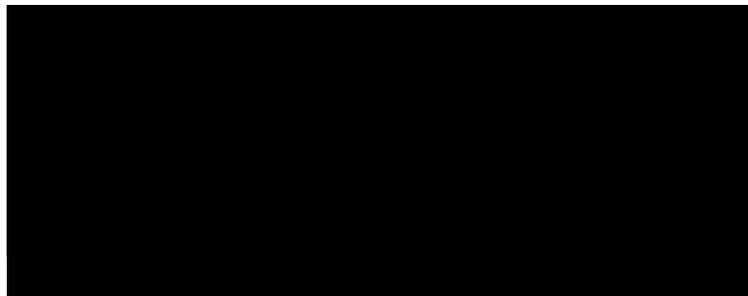
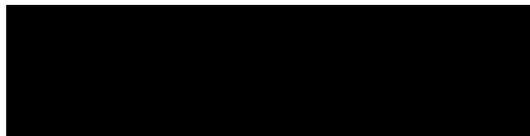
$$N = n_1 + n_2 + n_3 + n_4 = 3 + 7 + 6 + 4 = 20$$

Frecuencias relativas:

$$f_1 = \frac{3}{20} = 0.15; \quad f_2 = \frac{7}{20} = 0.35; \quad f_3 = \frac{6}{20} = 0.3; \quad f_4 = \frac{4}{20} = 0.2;$$

$$f_1 + f_2 + f_3 + f_4 = 0.15 + 0.35 + 0.3 + 0.2 = 1$$

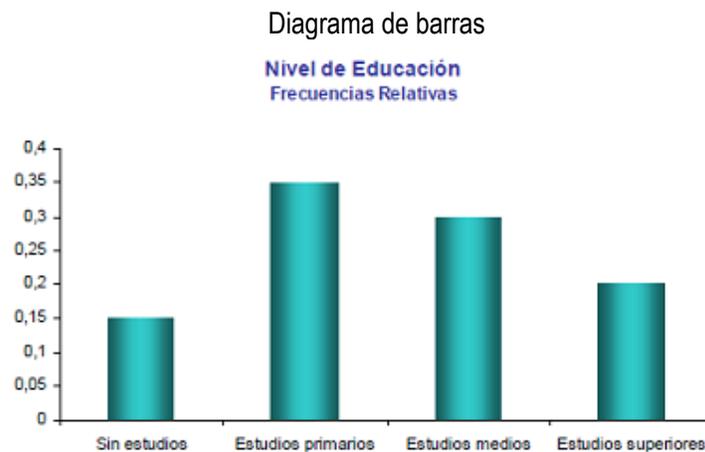
Distribución de frecuencias:



Según las frecuencias mostradas tenemos que en el grupo analizado la frecuencia más alta está asociada a la categoría estudios primarios; mientras que la menos frecuente es la categoría sin

estudios. Estas tablas de frecuencias tienen diversas maneras de ser representadas gráficamente. A continuación, mencionaremos las más utilizadas:

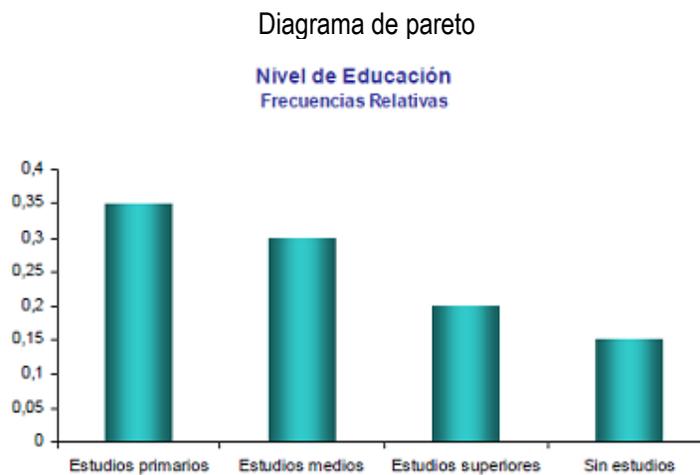
Diagrama de barras: Permite visualizar de forma sencilla la distribución de una variable cualitativa. Se dibuja sobre cada categoría una barra (*o rectángulo*) cuya altura coincide con la frecuencia absoluta o relativa de dicha categoría. Por ejemplo, para la variable analizada tendríamos:



Fuente: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008. Pag. 21.

GRAFICO 1.1

Diagrama de Pareto: Es como un diagrama de barras descrito en el párrafo anterior, pero en este caso se ordenan las categorías de mayor a menor frecuencia (*absoluta o relativa*).



Fuente: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008. Pag. 21.

GRAFICO 1.2

En la parte superior de la figura suele trazarse una línea que representa la suma de la frecuencia de cada categoría y las que la preceden:

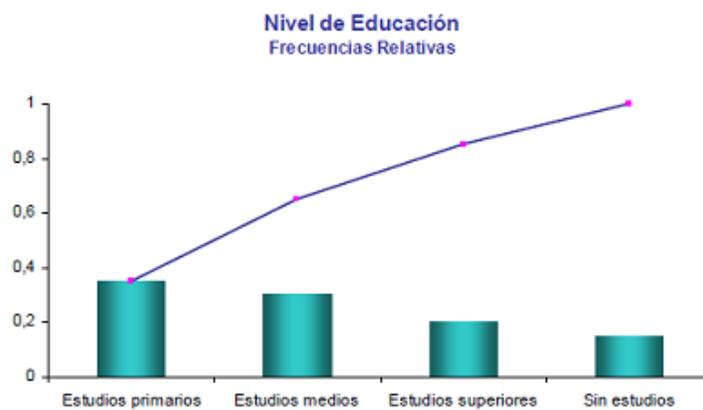
$$f_2 = 0.35$$

$$f_2 + f_3 = 0.35 + 0.30 = 0.65$$

$$f_2 + f_3 + f_4 = 0.35 + 0.30 + 0.20 = 0.85$$

$$f_2 + f_3 + f_4 + f_1 = 0.35 + 0.30 + 0.20 + 0.15 = 1$$

Diagrama de Pareto



Fuente: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008. Pag. 21.

GRAFICO 1.3

Pictograma: Consiste en un círculo en el que se representan sectores o porciones con áreas proporcionales a las frecuencias de cada una de las categorías.



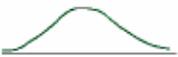
Fuente: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008. Pag. 22.

GRAFICO 1.4

Otro de los elementos utilizados para analizar variables cualitativas es la denominada “*moda*” o clase modal que corresponde al dato o categoría con mayor frecuencia o número de repeticiones; A pesar que no sirve como una medida de descripción de una variable, ya que la moda puede no ser única. En el ejemplo que se ha venido desarrollando vemos que los estudios primarios también corresponden a la moda de los datos analizados.

Variables cuantitativas: Para este tipo de variables, dependiendo si se trata de variables discretas, el análisis es muy similar al aplicado para las variables cualitativas, es decir mediante el uso de los diagramas de frecuencia que fueron descritos con anterioridad. Para el caso de variables continuas se puede incorporar la utilización de histogramas y polígonos de frecuencia, diagramas de caja, tallo, hoja. Adicionalmente este tipo de variables deben ser analizadas a través de medidas de posición central, dispersión, asimetría y curtosis ¹⁰.

¹⁰ Se define curtosis como la medida o grado de apuntamiento de una curva de distribución respecto a un estándar, mide la mayor o menor

concentración de datos alrededor de la media, pudiendo darse el caso de una curva muy puntiaguda o leptocurtica , mediamente puntiaguda o mesocurtica  y platicurtica o curva completamente aplastada .

Histogramas de frecuencias: Las clases o rangos de las variables continuas no están dados en forma implícita como en las variables cualitativas o en las discretas; por tanto, es necesario construirlas manualmente, para ello, se divide el conjunto de posibles valores de la variable en intervalos que no se intersequen o se solapen; aquí se puede identificar un punto central de cada intervalo (*marca de categoría c_i*), luego se puede proceder al igual que en las variables cualitativas. Por ejemplo, los datos que se muestran a continuación, corresponden a los egresos familiares mensuales a un determinado grupo de 75 clientes:

No. Obs.	Egreso										
	Familiar	No. Obs.									
1	81,86	16	142,68	31	90,5	46	531,10	61	152,08		
2	105,63	17	510,22	32	89,5	47	475,76	62	228,81		
3	110,69	18	158,83	33	466,9	48	316,50	63	76,92		
4	134,25	19	278,85	34	87,1	49	279,59	64	255,20		
5	226,18	20	168,62	35	309,8	50	48,59	65	241,99		
6	273,87	21	176,20	36	247,4	51	96,67	66	417,10		
7	142,38	22	179,11	37	427,8	52	256,55	67	752,44		
8	309,96	23	113,07	38	195,7	53	514,33	68	352,71		
9	101,43	24	876,16	39	257,6	54	161,60	69	259,47		
10	276,27	25	64,43	40	176,7	55	228,37	70	225,39		
11	662,80	26	112,35	41	285,9	56	638,37	71	174,34		
12	493,73	27	255,47	42	450,6	57	442,16	72	308,71		
13	308,79	28	321,31	43	56,3	58	65,06	73	455,13		
14	254,42	29	434,38	44	306,5	59	160,58	74	122,70		
15	172,93	30	707,44	45	156,8	60	197,39	75	479,79		

Elaboración: Propia
Fuente: Bdd APEVSC

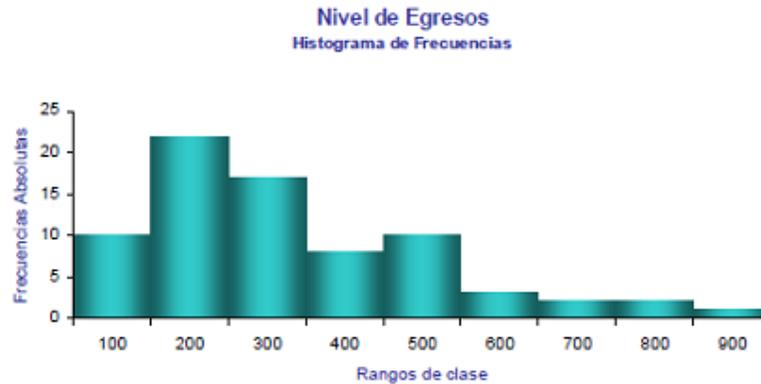
Se puede estructurar rangos o intervalos de tamaño igual a \$100 para construir la distribución de frecuencias, de este modo, el primer rango corresponderá a los egresos mensuales inferiores a \$100, cuya marca de categoría c_1 es igual a \$ 50, procediendo igual con el resto de datos se pueden encontrar nueve categorías diferentes ($k = 9$).

<i>Rangos de Clase</i> (Gasto en \$)	<i>Frecuencias absolutas</i> n_i	<i>Frecuencias absolutas acumuladas</i> N_i	<i>Frecuencias relativas</i> f_i	<i>Frecuencias relativas acumuladas</i> F_i
100	10	10	13,33%	13,33%
200	22	32	29,33%	42,67%
300	17	49	22,67%	65,33%
400	8	57	10,67%	76,00%
500	10	67	13,33%	89,33%
600	3	70	4,00%	93,33%
700	2	72	2,67%	96,00%
800	2	74	2,67%	98,67%
900	1	75	1,33%	100,00%

De la tabla de frecuencias generadas, se puede extraer aquellos clientes que tienen un egreso familiar inferior a \$200 es igual al 42.67%, mientras que aquellos que gastan más de \$600 mensuales representan apenas el 6.67% del total evaluado ($1 - 93.33\%$). La proporción de clientes que han declarado gastar entre \$100 y \$300 corresponde al 52%. Con los datos obtenidos, podemos elaborar un histograma, que no es más que la representación de las frecuencias mediante áreas, para ello, sobre cada rango o clase se levanta un rectángulo, cuya área representa la frecuencia o número de observaciones de esa clase.

- Cuando las clases (*o intervalos*) en que dividimos los datos son de distinta longitud, el eje vertical no tiene sentido. Como la frecuencia es el área de cada rectángulo, si dibujamos rectángulos con distinta base su mayor o menor altura no nos da información.
- Cuando las clases son de la misma longitud, las frecuencias son proporcionales a las alturas de los rectángulos. La altura nos informa sobre la densidad o concentración de datos en ese intervalo.
- Si los rectángulos son más altos hay más datos de la variable.
- Si los rectángulos son más bajos los datos de la variable son más escasos.

Histograma de frecuencias



Fuente: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008. Pag. 26.

GRAFICO 1.5

Los rectángulos se dibujan en forma contigua (*a diferencia del diagrama de barras o de pareto*) para transmitir la idea de variable continua. La forma del histograma es la misma si representamos frecuencias absolutas o relativas, sólo cambia la escala del eje vertical. La forma del histograma depende de:

- El ancho de las clases o tamaño de los intervalos
- Elección del punto donde empieza la primera clase

Para la selección del número de intervalos para un histograma, es preferible empezar con pocos grupos y a medida que se quieran aumentar éstos se debe verificar que exista un aumento de información. Si tenemos N observaciones se puede elegir el número de familias como el entero más próximo a \sqrt{N} . En nuestro ejemplo tenemos N=75 clientes, entonces $\sqrt{N} = \sqrt{75} = 8.6 \approx 9$

La forma del histograma refleja las propiedades y características que tiene la variable, las que se pueden apreciar visualmente. Por ejemplo, se puede ver si la variable presenta algún tipo de simetría, si presenta un solo valor máximo (*pico o moda*) o en su defecto, tiene varias modas; al respecto, podemos mencionar las formas más frecuentes de los histogramas.

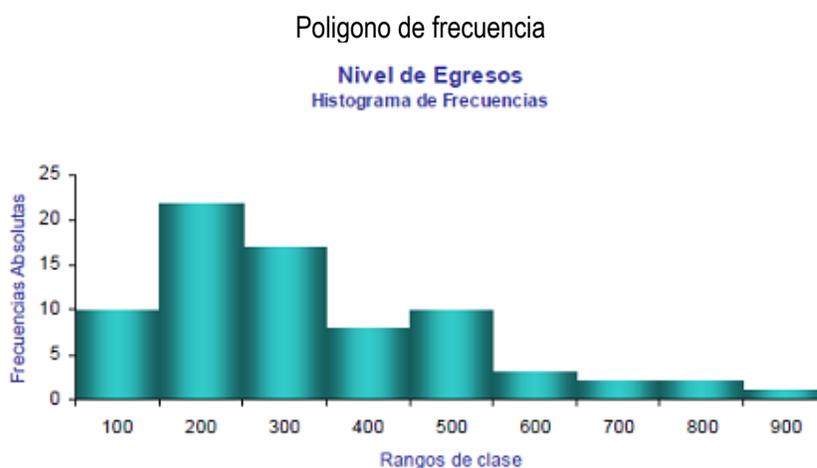
Unimodal simétrico: se suele dar en variables en las que hay una gran cantidad de observaciones con valores intermedios y algunos valores en ambos extremos (*notas, peso, altura, otros*).

Unimodal asimétrico a la derecha: se da en variables que tienen una gran cantidad de observaciones pequeñas o intermedias y algunos datos grandes (*gasto, ingreso, otros*).

Unimodal asimétrico a la izquierda: variables con muchas observaciones de valor alto o intermedio (*esperanza de vida en los distintos países*).

Bimodal simétrico: suele aparecer cuando los datos son de 2 grupos heterogéneos y conviene estudiarlos por separado (*un objeto que se hiciera en dos tamaños distintos en cantidades iguales*).

Polígono de Frecuencia: consiste en una representación gráfica de las frecuencias de una variable, similar al histograma y se obtiene al unir los centros de la base superior de los rectángulos del histograma.

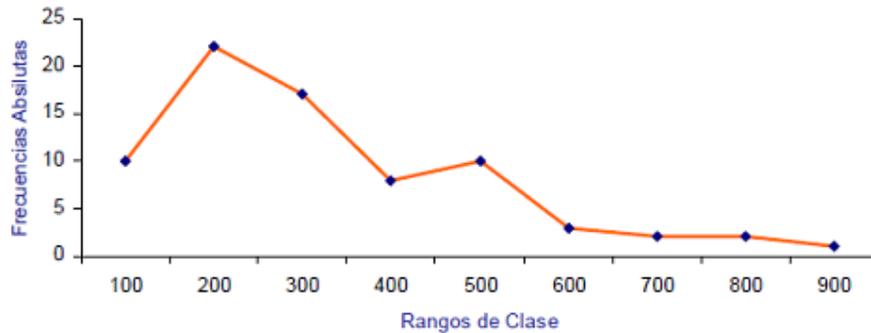


Fuente: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008. Pag. 27.

GRAFICO 1.6

Polígono de frecuencia

Nivel de Egresos Polígono de Frecuencias



Fuente: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008. Pag. 27.

GRAFICO 1.7

Tanto el histograma como el polígono de frecuencias pueden determinarse en forma acumulada, la diferencia fundamental de éstas dos representaciones es la forma más suavizada que proporciona el polígono.

Otra de las herramientas del análisis univariante que permiten conocer más sobre una variable cualitativa es el *diagrama de tallo y hojas*, el cual permite obtener en forma simultánea la distribución de frecuencias de la variable y su representación gráfica. Para construirlo hay que separar en cada dato el último dígito de la derecha (*la hoja*) del resto de las cifras (*el tallo*). De este modo los tallos aparecen a la izquierda de una línea vertical y a la derecha de cada uno anotamos las cifras finales (*hojas*) de todos los datos de cada rango o clase. Si al grupo de 75 clientes que se analizó los egresos familiares, tabulamos sus edades, tenemos la siguiente tabla:

No. Obs.	Edad								
1	18	16	18	31	75	46	19	61	76
2	80	17	84	32	74	47	76	62	64
3	76	18	71	33	67	48	31	63	62
4	31	19	32	34	61	49	35	64	54
5	35	20	36	35	57	50	29	65	53
6	28	21	28	36	56	51	28	66	48
7	23	22	25	37	52	52	27	67	46
8	23	23	23	38	47	53	61	68	42
9	68	24	67	39	43	54	65	69	42
10	40	25	40	40	40	55	41	70	35
11	44	26	44	41	39	56	42	71	30
12	45	27	46	42	35	57	47	72	29
13	57	28	57	43	28	58	57	73	28
14	57	29	51	44	25	59	55	74	27
15	51	30	52	45	19	60	53	75	19

Elaboración: Propia
Fuente: Bdd APEVSC

Aplicando el procedimiento señalado para los datos mostrados tenemos el siguiente diagrama de tallo y hojas:

1	8	8	9	9	9											
2	3	3	3	5	5	7	7	8	8	8	8	8	9	9		
3	0	1	1	2	5	5	5	5	6	9						
4	0	0	0	1	2	2	2	3	4	4	5	6	6	7	7	8
5	1	1	2	2	3	3	4	5	6	7	7	7	7	7		
6	1	1	2	4	5	7	7	8								
7	1	4	5	6	6	6										
8	0	4														

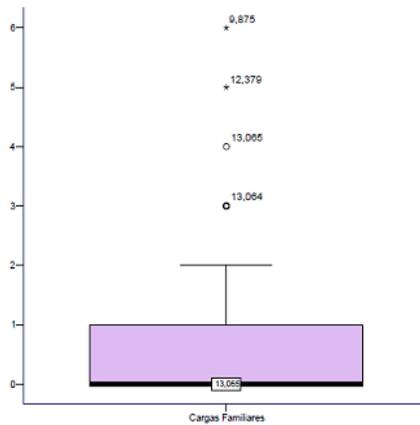
Al igual que el histograma o el polígono de frecuencias, este tipo de diagramas proporciona una impresión visual del número de observaciones de cada clase, con la ventaja de que al darnos un mayor detalle nos permite recuperar los datos, lo que no puede hacerse con el histograma o el polígono.

Diagrama de caja: Los diagramas de caja proporcionan información visual completa referente a la distribución de los datos. Pueden ser de gran utilidad como técnica de análisis exploratorio de datos, ya que nos proporcionan información sobre la mediana (*o media*), sobre el 50% y 90% de los datos, sobre la existencia de valores atípicos, así como, de la simetría de la distribución, su construcción se realiza de la siguiente manera:

- Se ordenan los datos de la muestra y se obtiene el valor mínimo, el máximo, y los tres cuartiles Q_1 , Q_2 y Q_3 .
- Dibujar un rectángulo cuyos extremos son Q_1 y Q_3 e indicar la posición de la mediana, Q_2 mediante una línea.
- Calcular unos límites admisibles inferior y superior, L_i y L_s , que identifiquen a los valores atípicos.
- Considerar como valores atípicos los situados fuera del intervalo (L_i, L_s) .
- Dibujar una línea que va desde cada extremo del rectángulo central hasta el valor más alejado no atípico, es decir, que está dentro del intervalo (L_i, L_s) .
- Identificar todos los datos que están fuera del intervalo (L_i, L_s) , marcándolos como atípicos.

Es posible introducir algunas variaciones en la construcción de estos diagramas, dependiendo del tipo de estudio y de la información disponible¹¹. La caja o rectángulo contiene un porcentaje de la muestra y puede construirse con diferentes rangos de variación, como el 80% de los datos y ser cortada por la media; sin embargo, lo más usual es que sea cortada por la mediana, de este modo se tiene de antemano conocimiento del comportamiento del 50% de la población en estudio sobre una variable específica. Los diagramas de caja proporcionan una idea intuitiva de la simetría de la distribución de los datos; si la media no está en el centro del rectángulo, eso significa que la distribución no es simétrica, conociendo además a qué lado se escora o desvía.

¹¹ CALVO-FLORES Antonio, ARQUEZ PEREZ Antonio, Modelos Estadísticos Teóricos, Facultad de Economía y Empresa, Universidad de Murcia, Pag. 12.



Complementariamente al análisis gráfico desarrollado para las variables cuantitativas, es conveniente estudiar los estadísticos descriptivos de las diferentes variables agrupadas en medidas de posición y de variabilidad. Las medidas de posición (*tendencia central*) forman parte de las medidas descriptivas numéricas, cuya función es darnos la orientación del conjunto de datos. Por su parte las medidas de variabilidad se encargan de proporcionarnos información correspondiente a la dispersión de los datos, puesto que varios conjuntos de datos pueden presentar iguales valores promedios pero diferente variabilidad.

Media Aritmética: o simplemente media, es la medida de posición más utilizada, representa el centro físico del conjunto de datos y se define como la suma de todos los posibles valores observados, ponderada por el total de observaciones registradas. Si x_1, x_2, \dots, x_n son n observaciones numéricas, entonces la media aritmética de dichas observaciones se define de la siguiente manera: Es decir, si la tabla de valores de una variable X es:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + X_n}{n}$$

O lo que es lo mismo:

$$\bar{x} = \frac{\sum_{i=1}^n X_i}{n}$$

Desviación Típica: es una medida de variabilidad de mayor uso se determina mediante la raíz cuadrada de la varianza, definida como la media de las diferencias cuadráticas de n observaciones respecto a su media aritmética y se calcula como:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

En la varianza al tomar el cuadrado de las desviaciones se obtienen unidades al cuadrado, para evitar que se magnifique dicha diferencia real se establece la desviación estándar (o *desvío típico*) como sigue:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Por ejemplo si tomamos los siguientes datos:

44, 59, 36, 55, 47, 61, 53, 32, 65, 51

Podemos obtener su media aritmética de la siguiente manera:

$$\bar{x} = \frac{44 + 59 + 36 + 55 + 47 + 61 + 53 + 32 + 65 + 51}{10}$$

$$\bar{x} = \frac{503}{10} = 50.3$$

La varianza de los datos presentados estará dada por:

$$S^2 = \frac{\sum_{i=1}^{10} (x_i - \bar{x})^2}{10}$$

$$(x_1 - \bar{x})^2 = (44 - 50.3)^2 = 39.69$$

$$(x_2 - \bar{x})^2 = (59 - 50.3)^2 = 75.69$$

$$(x_3 - \bar{x})^2 = (36 - 50.3)^2 = 204.49$$

$$(x_4 - \bar{x})^2 = (55 - 50.3)^2 = 22.09$$

$$(x_5 - \bar{x})^2 = (47 - 50.3)^2 = 10.89$$

$$(x_6 - \bar{x})^2 = (61 - 50.3)^2 = 114.49$$

$$(x_7 - \bar{x})^2 = (53 - 50.3)^2 = 7.29$$

$$(x_8 - \bar{x})^2 = (32 - 50.3)^2 = 334.89$$

$$(x_9 - \bar{x})^2 = (65 - 50.3)^2 = 216.09$$

$$(x_{10} - \bar{x})^2 = (51 - 50.3)^2 = 0.49$$

$$S^2 = \frac{1026.10}{10} = 102.61$$

La varianza de los datos presentados estará dada por:

$$S = \sqrt{102.61} = 10.13$$

Como se mencionó, tanto la media como la desviación estándar nos permiten conocer de mejor manera las características de un conjunto de datos, los cuales, en el caso de entidades financieras, podrían estar representados por los montos de los préstamos concedidos durante un tiempo específico; los plazos de concesión, las edades de los sujetos de crédito, la morosidad de dichos créditos, los ingresos y/o egresos de una persona dedicada al micro crédito que solicita un producto crediticio.

Cuando tratamos de datos es común hablar de las medidas descriptivas que los caracterizan, tales como la media, su desviación estándar y la moda. Pero cuando ya se empieza a hablar de variables aleatorias se procura homologar dichas medidas descriptivas hacia las distribuciones o el comportamiento que pueden seguir dichas variables.

Valor Esperado: Sea X una variable aleatoria (v.a.) discreta que toma los valores x_1, x_2, \dots, x_n y cuya función de probabilidad es p_1, p_2, \dots, p_n respectivamente. Se define el valor esperado de X , como:

$$\mu_x = E[X] = x_1 p_1 + x_2 p_2 + x_3 p_3 + \dots + x_n p_n$$

$$\mu_x = E[X] = \sum_{i=1}^n x_i P(x_i)$$

Si la v.a. X es de tipo continuo, con función de densidad $f(x)$, definimos el valor esperado $E(X)$, como:

$$\mu_x = E[X] = \int_{-\infty}^{+\infty} x f(x) . dx$$

La expresión de $E(X)$ en el caso que X sea una v.a. discreta, este valor es la media ponderada de los posibles valores que puede tomar la variable X , en donde los pesos o ponderaciones son las probabilidades, $P(x_i) = P(X = x_i)$, de ocurrencia de los posibles valores de X . Luego el valor esperado de X se interpreta como una media ponderada de los posibles valores de X , y no como el valor que se espera que tome X , pues puede suceder que $E(X)$ no sea uno de los posibles valores de X . En el caso de v.a. continua, $E(X)$ nos indica el centro de la función de densidad, es decir, nos indica el centro de gravedad de la distribución.

Propiedades:

- La esperanza de una constante es la propia constante. Es decir si k es una constante entonces:

$$E(k) = k$$

- Si una v.a. X y k una constante, entonces:

$$E(k.X) = k.E(X)$$

- Si una v.a. X y k una constante, entonces:

$$E(X + k) = E(X) + k$$

- Si una v.a. X esta acotada, es decir existen dos valores a y b tales que $a \leq X \leq b$, entonces se verifica que:

$$a \leq E(X) \leq b$$

- Si X y Y son variables aleatorias, entonces:

$$E(X + Y) = E(X) + E(Y)$$

- Sea X y Y variables aleatorias; a, b constantes, entonces:

$$E[a.X + b.Y] = a.E[X] + b.E[Y]$$

- No se cumple que:

$$E[X^2] = (E[X])^2$$

Varianza: Sea X una v.a. que toma los valores x_1, x_2, \dots, x_n y cuya función de probabilidad es p_1, p_2, \dots, p_n respectivamente. La varianza de una distribución se denota y define así:

$$\sigma_x^2 = \text{Var}[X] = E[(X - \mu_x)^2]$$

También puede escribirse como

$$\sigma_x^2 = \text{Var}[X] = E([X - E(X)]^2)$$

Es una medida de dispersión de los valores de la variable respecto de su media, y nos permite conocer el grado de separación de los valores de la distribución, pudiendo realizar comparaciones con otras distribuciones. La varianza se expresa en las mismas unidades que la variable X, pero al cuadrado. La desviación estándar o desviación típica, se expresa en las mismas unidades de medida que la variable X.

Propiedades:

- La varianza no puede ser negativa
- La varianza de una constante k es cero.

$$\text{Var}(k) = 0$$

Sea X una v.a. cuya varianza existe. Entonces:

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

- Sea X una v.a. cuya varianza existe, y k una constante cualquiera. Entonces:

$$\text{Var}(k.X) = k^2 \cdot \text{Var}(X)$$

- Sea X una v.a. cuya varianza existe y a, b dos constantes cualesquiera. Entonces:

$$\text{Var}(aX + b) = a^2 \cdot \text{Var}(X)$$

-Sean X e Y dos v.a. independientes cuyas varianzas existen, entonces se verifica que la varianza de la suma o de la diferencia de ambas v.a. independientes es igual a la suma de las varianzas. Es decir:

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$$

Si las v.a. no son independientes entonces:

$$E[(X-E(X))(Y-E(Y))] = \text{Cov}(X,Y)$$

Y se verificará que:

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X,Y)$$

En términos generales, el valor medio representa las concentraciones de los valores observados de una variable y permite a priori tener una idea de la composición de los datos. Además mediante la desviación estándar veremos el distanciamiento del resto de observaciones respecto al valor medio, es decir, si en la variable analizada existe distanciamiento o agrupamiento de datos. Estas dos primeras conclusiones se ven complementadas con los valores máximos y mínimos, pues a más de acotar el conjunto de valores de la variable nos dan un indicio si dichos valores (*máximo y mínimo*) deben o no ser considerados en el análisis a fin de evitar distorsiones por información mal recabada.

1.3.2 ANÁLISIS BIVARIANTE ¹²

El análisis bivalente relaciona dos o más indicadores de manera que se pueda estudiar una variable en función de otra determinada, por ejemplo, un típico análisis bivalente son las tablas cruzadas donde se refleja el conocimiento de una marca en función de la edad, los niveles de instrucción o el estrato socioeconómico de los entrevistados.¹³

Al igual que en el caso univariante, se puede distinguir el análisis a emplear, dependiendo si las variables son cualitativas o cuantitativas, es necesario mencionar que en esta parte del análisis deja de ser tan intuitivo como en el análisis preliminar.

Variables cualitativas: Una de las principales herramientas utilizadas para el análisis bivalente de variables cualitativas son las tablas de contingencia.

Tablas de Contingencia: Son tablas de frecuencias conjuntas en donde cada entrada representa un criterio de clasificación, que da como resultado que las frecuencias aparecen organizadas en casillas que contienen información sobre la relación existente entre los criterios que conforman la tabla. Las tablas de contingencia según la cantidad de variables que clasifique se denominan bidimensionales o de 2x2, cuando la cantidad de variables son dos, tridimensionales o de 3x3, cuando la cantidad es 3, hasta llegar a las multidimensionales o de n variables.

Si consideramos dos variables cualitativas, la notación estándar de las tablas de contingencias es la siguiente:

n_{ij} : frecuencia observada en la casilla formada por la combinación del nivel i de la variable X , el nivel j de la variable Y

$i: 1, 2, \dots, l$ ($l =$ número de categorías de la variable X)

¹² El análisis Bivalente fue tomado del la tesis: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008.

¹³ Definiciones básicas de análisis univariante, bivalente y multivariante, fecha acceso: Enero 2011, dirección: <http://www.serviciosestadisticos.es/metodologia/metodologia.htm>

$j: 1, 2, \dots, J$ ($J = \text{número de categorías de la variable } Y$)

Así por ejemplo, la frecuencia n_{12} se refiere a la frecuencia de la casilla resultante de combinar la categoría 1 de la variable X y la categoría 2 de la variable Y. La notación aquí utilizada corresponde a una tabla de contingencia bidimensional pero también resulta aplicable para tablas de más de dos variables. Por ejemplo, para una tabla tridimensional la notación será n_{231} y se refiere a la frecuencia resultante de combinar la categoría 2 de la variable X, la categoría 3 de la variable Y, así como la categoría 1 de la variable Z.

La utilidad de las tablas de contingencia no es únicamente la obtención de las frecuencias conjuntas de las variables analizadas, sino que permiten la aplicación de estadísticos para estudiar las posibles pautas de asociación existentes entre dichas variables.

El grado de asociación existente entre dos variables categóricas no puede ser establecido simplemente observando las frecuencias de una tabla de contingencia. Para determinar si dos variables se encuentran relacionadas necesitamos utilizar algún índice de asociación acompañado de su correspondiente prueba de significación. La prueba Chi-cuadrado (χ^2) de Pearson proporciona un estadístico que permite contrastar la hipótesis que dos criterios de clasificación utilizados sean independientes; es decir, podemos establecer si dos variables cualitativas son independientes entre sí. Para ello se comparan las frecuencias obtenidas (*frecuencias observadas*) con las frecuencias que teóricamente deberíamos haber encontrado en cada casilla si las dos variables fueran independientes.

Las frecuencias esperadas se estiman de la siguiente forma:

$$m_{ij} = \frac{n_i + n_j}{n}$$

Donde i se refiere a una fila cualquiera, j a una columna cualquiera e ij a una casilla cualquiera.

Una vez obtenidas las frecuencias esperadas para cada casilla el estadístico χ^2 se calcula de la siguiente forma:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \bar{m}_{ij})^2}{m_{ij}}$$

Donde n_{ij} se refiere a las frecuencias observadas y \bar{m}_{ij} se refiere a las esperadas. De la ecuación se desprende que el estadístico χ^2 valdrá cero cuando las variables sean completamente independientes, ya que las frecuencias esperadas coincidirán con las observadas, y que será tanto mayor cuanto mayor sea la discrepancia entre las frecuencias observadas y las esperadas, señalando la inexistencia de independencia entre las variables analizadas.

El estadístico χ^2 resulta de utilidad para establecer si existe asociación o no entre variables categóricas, pero no nos permite establecer el grado o fuerza de asociación entre dichas variables. Esta limitación es consecuencia directa de que su valor no solo depende del grado en que los datos se ajustan al modelo de independencia sino del número de casos que consta la muestra, ya que con tamaños muestrales muy grandes, diferencias relativamente pequeñas entre las frecuencias observadas y las esperadas pueden dar lugar a valores de Chi-cuadrado muy altos. Es por ello que para estudiar el grado de asociación entre dos variables categóricas se utilizan índices o medidas que intentan cuantificar ese grado de asociación, eliminando el efecto del tamaño muestral.

La selección de una medida de asociación concreta deberá tener en cuenta el tipo de variable analizada (*ordinal o nominal*) y la hipótesis que se intenta contrastar (*independencia*). Así, para analizar variables cualitativas nominales tenemos las siguientes medidas¹⁴:

Coeficiente de contingencia.

- Lambda.

- Tau.

¹⁴ AGUILERA DEL PINO Ana María, Tablas de Contingencia Bidimensional, Edición La Muralla, Madrid 2001, pag.20-22.

- Coeficiente de incertidumbre.

Con respecto a las variables cualitativas ordinales tenemos los siguientes estadísticos:

- Gamma.
- Tau-b de Kendall.
- Tau-c de Kendall.

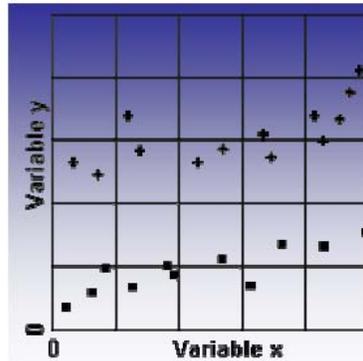
Variables cuantitativas: Para el caso de variables cuantitativas fundamentalmente nos apoyamos en el análisis de correlación y de covarianza existente entre las variables a analizar. Por ejemplo en educación se ha comprobado la relación entre las notas de lenguaje y matemática, la fortaleza de personas altas respecto a las de menor estatura, la relación entre los precios de venta de los productos de primera necesidad, respecto a su disponibilidad.

Correlación: mide la relación entre dos variables y su sentido (*si es directo o inverso*), cuando dicha relación es perfectamente lineal, dicho coeficiente vale 1 (ó -1), cuando el coeficiente tiene un valor próximo a cero, se puede afirmar que o bien, no existe relación entre las variables analizadas o bien, dicha relación no es lineal. La correlación está ligada directamente al concepto de covarianza, y podemos encontrar varias formas de escribir su ecuación entre las que tenemos:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X).Var(Y)}}$$

Correlación



$$r = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$$

Fuente: CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008. Pag. 43.

GRAFICO 1.8

De dichas ecuaciones vemos claramente que para la determinación de la correlación, primero hay que calcular la covarianza entre las variables, pese a que su significado intuitivo es más complejo que el del coeficiente de correlación.

Covarianza: La fórmula que expresa la covarianza entre dos variables es la siguiente:

$$\sigma_{xy}^2 = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Donde:

x_i es la observación de la variable X,

y_i corresponde a la observación de la variable Y,

Además de sus correspondientes valores medios (\bar{x}, \bar{y}) .

Una covarianza positiva significa que existe una relación lineal positiva entre las dos variables; es decir, valores bajos de la primera variable X se asocian con valores bajos de la segunda variable Y, y viceversa. Una covarianza negativa significa que existe una relación lineal inversa perfecta entre las dos variables, lo que significa que valores bajos en X se asocian con los valores altos en Y, mientras los valores altos en X se asocian con los valores bajos en Y.

Una covarianza igual a 0 se interpreta como la no existencia de una relación lineal entre las dos variables estudiadas. Una aproximación gráfica al significado de la covarianza es la que se muestra en la siguiente gráfica:

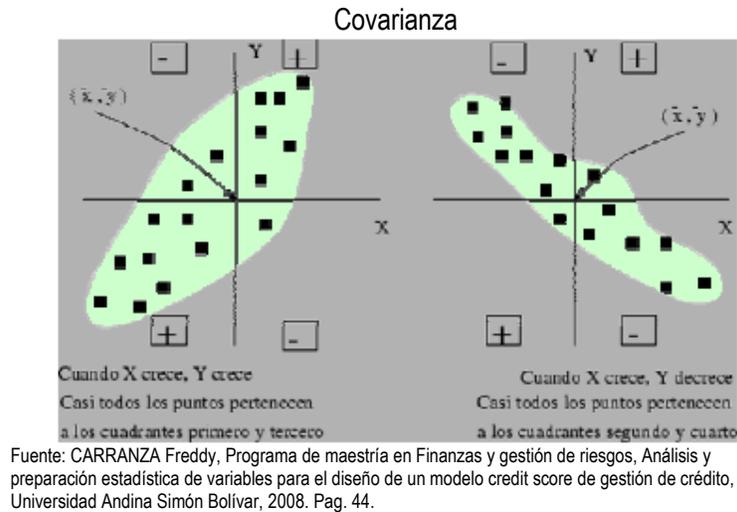


GRAFICO 1.8

Desde el punto de vista técnico la relación entre dos variables está dada por el coeficiente de correlación de Pearson que cumple las siguientes características:

- El índice de correlación de Pearson no puede valer menos de -1 ni más de +1.
- El índice de correlación de Pearson (*en valor absoluto*) no varía cuando se transforman linealmente las variables.
- Es importante señalar que correlación no implica causación, es decir que si dos variables están altamente correlacionadas no implica que X causa Y ni que Y causa X.
- Es importante indicar que el coeficiente de correlación de Pearson puede verse afectado por la influencia de terceras variables.
- Además el valor del coeficiente de Pearson depende en parte de la variabilidad de la muestra de trabajo o población en estudio.

1.4 ECONOMETRÍA

1.4.1 DEFINICIONES BASICAS¹⁵

Literalmente, econometría significa "*medición económica*". Sin embargo, si bien es cierto que la medición es una parte importante de la econometría, el enlace de esta disciplina es mucho más amplio, como puede deducirse de las siguientes citas:

La econometría, es el resultado de cierta perspectiva sobre el papel que juega la economía, consiste en la aplicación de la estadística matemática a la información económica para dar soporte empírico a los modelos construidos por la economía matemática y obtener resultados numéricos.¹⁶

La econometría, puede ser definida como el análisis cuantitativo de fenómenos económicos reales, basados en el desarrollo simultaneo de la teoría y la observación, relacionados mediante métodos apropiados de inferencia.¹⁷

La econometría, puede ser definida como la ciencia social en la cual las herramientas de la teoría económica, las matemáticas y la inferencia estadística son aplicadas al análisis de los fenómenos económicos.¹⁸

La econometría, tiene que ver con la determinación empírica de las leyes económicas.¹⁹

El arte del econométrista consiste en encontrar el conjunto de supuestos que sean suficientemente específicos y realistas, de tal forma que le permitan aprovechar de la mejor manera los datos que tiene a su disposición.²⁰

Los econométristas, son una ayuda en el esfuerzo por disipar la mala imagen pública de la economía (*cuantitativa o de otro tipo*) considerada como una materia en la cual se abren cajas vacías, suponiendo la existencia de abrelatas, para revelar un contenido que será interpretado por cada 10 economistas de 11 maneras diferentes.²¹

¹⁵ MONTGOMERY Wiley, Diseño y Análisis de Experimentos, EEUU, Segunda Edición, 2002, Pág. 1 – 3.

¹⁶ GERHARD Tintner, Methodology of Mathematical Economics and Econometrics, The University of Chicago Press, Chicago, 1968, Pág. 74.

¹⁷ SAMUELSON P.A, KOOPMANS T.C y STONE J.R.N., Report of the Evaluative Committee for Econometria, Econometrica, 1954, Pág. 141-146.

¹⁸ ARTHUR S, Econometric Theory, Jhon Wiley & Sons, New York. 1964. Pág. 1.

¹⁹ THEIL H., Principles of Econometrics, Jhon Wiley & Sons, New York. 1964, Pág. 1.

²⁰ MALINVAUD E., Statistical Methods of Econometrics, Rand Mc Nally, Chicago, 1966, Pág. 514.

²¹ DARNELL Adrian C. y EVANS J., The limits of Econometrics, Edward Elgar Publishing, Inglaterra, 1990, Pág. 54.

El método de la investigación econométrica busca esencialmente una conjunción entre la teoría económica y la medición real, utilizando como puente la teoría y la técnica de la inferencia estadística²².

La econometría combina elementos de: Teoría Económica, Matemáticas y Estadística

La econometría nos puede ayudar a comprender cuestiones como:

- Los efectos de un programa de entrenamiento.
- La variación de los rendimientos de estrategias diferentes de inversión.
- Los efectos de una campaña publicitaria.
- Predicción del comportamiento de variables macroeconómicas: tasa de interés, inflación, PIB, entre otros.
- Relaciones macroeconómicas: relación inflación desempleo, relación inflación masa monetaria.
- Relaciones macroeconómicas: relación educación salario, relación producción factores productivos.
- Finanzas: análisis de la volatilidad de los activos, modelos de valoración de activos.

Metodología de la Econometría²³: La metodología de la econometría se realiza dentro de los siguientes lineamientos:

1. Planeamiento de la teoría o de la hipótesis
2. Especificación del modelo matemático de la teoría
3. Especificación del modelo econométrico o estadístico de la teoría
4. Obtención de datos
5. Estimación de los parámetros del modelo econométrico
6. Prueba de hipótesis
7. Pronostico o predicción

²² HAAVELMO T., The probability Approach in Econometrics, Suplemento de Econometrica, 1994, Pág. Prefacio iii.

²³ GUJARATI Damodar, Introducción a la Econometría, Cuarta edición, McGraw Hill, 2004, Pág. 3 – 5.

8. Utilización del modelo para fines de control o de política

1.4.2 NATURALEZA DEL ANÁLISIS DE MODELOS DE REGRESIÓN²⁴

El análisis de regresión trata del estudio de la dependencia de la variable dependiente, respecto a una o más variables (*variables explicativas*), con el objeto de estimar y/o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (*en muestras repetidas*) de las últimas.

Terminología y notación

En la teoría económica los términos variable dependiente y variable explicativa están descritos de varias maneras:

variable dependiente	variable explicativa
variable explicada	variable independiente
predicha	predictora
regresada	regresora
respuesta	estimulo
endógena	exógena
resultado	covariante
variable controlada	variable de control

Si se está estudiando la dependencia de una variable en una única variable explicativa, como es el caso del gasto de consumo sobre el ingreso real, dicho estudio es conocido como el análisis de regresión simple, o con dos variables. Sin embargo, si se está estudiando la dependencia de una variable en más de una variable explicativa, tal como el producto de una cosecha, la lluvia, la temperatura, el sol y los fertilizantes, este se conoce como análisis de regresión múltiple. En otras

²⁴ GUJARATI Damodar, Introducción a la Econometría, Cuarta edición, McGraw Hill, 2004, Pág. 24.

palabras, en una regresión de dos variables solo hay una variable explicativa, mientras que la regresión múltiple hay más de una variable explicativa.

1.4.3 MODELOS ESTADISTICOS²⁵

Existen varios modelos que tratan de explicar el comportamiento de una serie de datos, existe una amplia clasificación y especificación de modelos, que se particularizan por su tratamiento y connotación. Por ejemplo, de acuerdo al número de alternativas u opciones posibles de respuestas de la variable endógena o dependiente (*modelos binarios de respuesta dicotómica y modelos de elección múltiple*), a la función empleada para la estimación de la probabilidad, (*modelo de probabilidad lineal, logit, poisson, multinomial*), de acuerdo al tipo de análisis de respuesta (*de conjunto multidimensional, correspondencias, cluster, canonico*), es decir, existen varias tecnologías de regresión o análisis multivariante que interpretan a una variable dependiente a través de otras independientes. Sin embargo, de las experiencias conocidas las más aplicables que permiten desarrollar modelos de scoring se concentran en tres modelos estadísticos constituidos a través de método paso a paso²⁶, estos tres modelos son:

- Análisis Discriminante
- Regresión Lineal
- Regresión Logística Binaria

1.4.3.1 ANÁLISIS DISCRIMINANTE

Permite clasificar a los sujetos por estratos de categorización específicos conocidos también como grupos (D), en función de una serie de características o variables predictivas, establecida como una combinación lineal, la ecuación del modelo es:

²⁵ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.

²⁶ Método paso a paso: O denominado step-wise, consiste en ir acondicionando variables a un modelo, usando el estadístico G (*Wald*), para saber si la nueva variable explica significativamente la variable respuesta. De igual manera, este procedimiento elimina variables ya incluidas que por el efecto de la inclusión de la variable, esta puede ser no significativa.

$$D = B_1X_1 + B_2X_2 + \dots + B_pX_p \quad (ec. 1)$$

1.4.3.2 MODELO LINEAL

Determina una relación entre la variable respuesta Y, bajo el supuesto de que esta última sigue una distribución normal, y un conjunto de variables predictivas, la ecuación es:

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_pX_p \quad (ec. 2)$$

1.4.3.3 MODELO LOGIT

Denominado modelo Logístico, proporciona la probabilidad de que una determinada observación pertenezca a un cierto grupo, conocidos los valores de las variables independientes para esa observación, la ecuación del modelo es la siguiente:

$$P_j = \frac{e^{Z_j}}{1 + e^{Z_j}}; z_j; \text{sabiendo que } Z_j = B_1Z_1 + B_2Z_2 + \dots + B_pZ_p \quad (ec. 3)$$

O lo que es lo mismo:

$$\log\left(\frac{P_j}{1 - P_j}\right) = B_1X_1 + B_2X_2 + \dots + B_pX_p \quad (ec. 4)$$

Donde: X_p = variables independientes, $p=1,2,\dots,n$

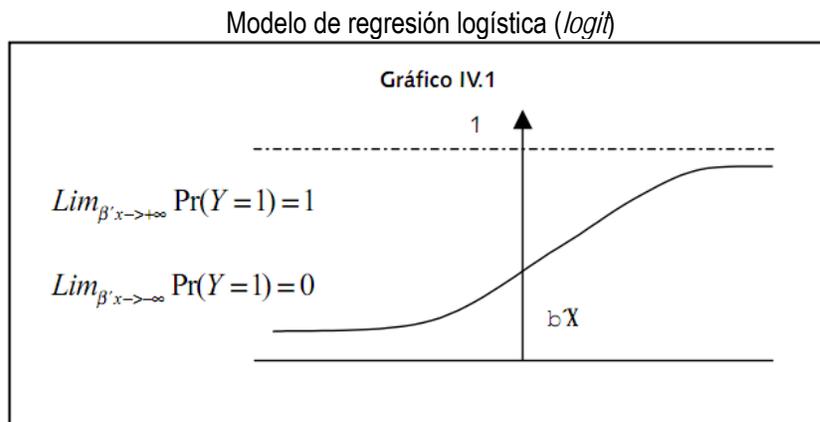
B_p = coeficientes estimados o pesos de cada variable, $p=1,2,\dots,n$

Z_j = puntuación del individuo o de la observación $j=1,2,\dots,k$

P_j = probabilidad del individuo j, de pertenecer a un grupo.

La función de distribución logística que modela la variable respuesta de la ecuación, cumple con los requisitos de ser una función continua, y que puede tomar cualquier valor comprendido entre 0 y 1. Además que la variable dependiente modelada es dicotómica. En cuanto a la interpretación de los parámetros estimados (*coeficientes*) el signo de los mismos indica la dirección en que se mueve la

probabilidad cuando aumenta la variable explicativa. Gráficamente, la representación del modelo logístico se ilustra a continuación.



Fuente: Los modelos logit y probit, fecha acceso: Marzo 2011, dirección:
<http://www1.inei.gob.pe/biblioineipub/bancopub/Est/Lib0515/Libro.pdf>

GRAFICO 1.9

CAPÍTULO II: DIAGNÓSTICO SITUACIONAL DE LA INSTITUCIÓN

2.1 DESCRIPCIÓN DE LA EMPRESA



2.1.1 ANTECEDENTES

MAQUITA CUSHUNCHIC tiene el significado "*DEMONOS LAS MANOS*", la Cooperativa nació como una respuesta a la necesidad de financiamiento de las mujeres pobres del Sur de Quito, la Cooperativa tiene 10 años en el mercado.²⁷

COOPERATIVA DE AHORRO Y CRÉDITO MAQUITA CUSHUNCHIC (*COAC*), fue fundada en 1998 por la Fundación MCCH²⁸ y con la participación de la Fundación Tierra Nueva, el FEPP²⁹ y las organizaciones de base de mujeres de la Quito Sur con la finalidad de brindar servicios financieros a sus socios en los barrios del sur de Quito.

Actualmente es una Cooperativa regulada bajo la Dirección Nacional de Cooperativas del Ministerio de Bienestar Social. En Junio 2009 cuenta con 21,988 socios, de los cuales 9,123 son ahorristas y 2,313 son prestatarios activos. En la fase de arranque empezó con una sola oficina en la zona sur de Quito; sucesivamente abrió cuatro agencias en otros barrios de la ciudad: Arcadia, La Comuna, Chillogallo y Centro, tres en la zona sur de Quito, una en el norte (*la Comuna*) y una en el centro³⁰.

El fondeo de la institución es representado por el ahorro captado de sus socios. Poco más de un quinto de los recursos financieros llegan como préstamos del exterior. MAQUITA CUSHUNCHIC se encuentra en un proceso de autorregulación con la Red Financiera Rural y de cumplimiento de

²⁷ Breve reseña histórica de la Cooperativa de ahorro y crédito Maquita Cushunchic, fecha acceso: Enero 2011, dirección: <http://www.uasb.edu.ec/UserFiles/363/File/pdfs/SUBSITIO%20FIDA/ENFOQUES%20Y%20EXPERIENCIAS%20DEL%20DESARROLLO%20RURAL/Cooperativa%20Maquita%20Cushunchic.pdf>

²⁸ Fundación Maquita Cushunchic Comercializando Como Hermanos (MCCH) es una institución privada sin fines de lucro que nació el 24 de marzo de 1985 por iniciativa de las Comunidades Eclesiales de Base del Sur de Quito; se constituyó legalmente para trabajar en la comercialización comunitaria a servicio de los sectores rurales y urbano marginales de menos recursos del país. fecha acceso: Enero 2011, dirección: http://www.ratinginitiative.org/uploads/tx_dbreports/COAC_Maquita_Cushunchic_informe_final_de_calificacion_Julio_2005.pdf

²⁹ Fondo Ecuatoriano Populorum Progressio.

³⁰ La agencia Centro, la última en orden de apertura (Julio del 2004), es parte de un proyecto co-financiado por la Municipalidad de Quito y la Comunidad Europea (PROQUITO-COSPE)

requisitos para pasar bajo control de la Superintendencia de Bancos y Seguros del Ecuador en el corto plazo.³¹

2.1.2 MISION Y VISION

Misión:³² Trabajamos para modificar la estructura y las relaciones de la sociedad y la economía con incidencia, en favor de las familias, comunidades y organizaciones de menos recursos económicos del Ecuador, mediante procesos productivos-comerciales y asociativos, de construcción de circuitos sostenibles de economía social y solidaria con productos estratégicos que permitan incrementar el ingreso familiar, mejorar su calidad de vida y practicar valores humanos y cristianos.

Visión³³: En el 2020 se convertirá en la organización articuladora de las cadenas de valor de los productos estratégicos seleccionados, logrando un alto impacto en la economía de las familias a través de redes, liderazgos y procesos de incidencia, con los principios de la economía social y solidaria.

2.1.3 GOBERNABILIDAD Y ESTRUCTURA ORGANIZACIONAL

MAQUITA CUSHUNCHIC es, por su propia naturaleza, propiedad de todos sus socios, quienes, a través de la Asamblea General, eligen cada dos años una Asamblea de Representantes (*se elige 26 representantes principales y 26 suplentes*)³⁴. El Consejo de Administración³⁵, que se reúne regularmente una vez por mes, desde el inicio, está en manos de las tres Fundaciones que apoyaron a la Cooperativa: la Fundación MCCH, la Fundación Tierra Nueva y el Fondo Ecuatoriano Populorum Progressio (*FEPP*). Esto ha permitido hasta la fecha mantener claro el enfoque social y la misión de la Cooperativa.

³¹ Por haber superado los límites de activo (*US\$ 1millon*) y de patrimonio (*US\$ 200.000*) la cooperativa debe por ley pasar a la supervisión bancaria.

³² Misión de la Cooperativa de ahorro y crédito Maquita Cushunchic, fecha acceso: Enero 2011, dirección: <http://www.fundmcch.com.ec/mision.php>

³³ Visión de la Cooperativa de ahorro y crédito Maquita Cushunchic, fecha acceso: Enero 2011, dirección: <http://www.fundmcch.com.ec/mision.php>

³⁴ Los Representantes han sido electos en Febrero 2005.

³⁵ Renovado en el año 2005

Con el crecimiento del número de socios, la gobernabilidad se vuelve más compleja, también por el hecho que en las nuevas agencias (*Centro y Norte de Quito*) el carisma y el reconocimiento moral de las tres instituciones fundadoras son menores que en la zona sur de la ciudad.

Los miembros se renuevan por un tercio cada tres años asegurando una cierta continuidad de las políticas y estrategias. En los últimos cuatro años la Gobernabilidad ha mejorado notablemente, logrando superar la crisis de gobernabilidad del 2000 y aportando positivamente al crecimiento de la institución. En cuanto a gobernabilidad es positivo el hecho que haya 6,810³⁶ socios ahorristas netos que tienen como interés primordial el de cuidar a la solvencia financiera de la Cooperativa.

2.1.4 PRINCIPALES PRODUCTOS Y SERVICIOS

El desarrollo de productos y servicios se realiza sobre un equilibrio REAL entre rentabilidad financiera y social. Ofrece servicios: crédito para microempresarios, crédito para el desarrollo familiar, créditos de consumo para consolidación de deudas, gastos de viaje, compra de electrodomésticos, remodelación de viviendas y mucho más³⁷.

2.1.5 FACTORES CRITICOS DE EXITO

Los factores críticos de éxito para el desarrollo³⁸ y las mejores prácticas son las siguientes:

- Liderazgo
- Equipo humano de alto nivel técnico, motivado y comprometido
- Tecnología de punta
- Gestión basada en mejores practicas
- Rendición de cuentas (*evaluación*)

³⁶ Es la diferencia entre 9.123 socios activos y 2.313 socios prestatarios activos (*al año 2005*)

³⁷ Principales productos y servicios de la Cooperativa de ahorro y crédito Maquita Cushunchic, fecha acceso: Octubre 2010, dirección: <http://www.coacmaquitacushunchic.com.ec/>

³⁸ Factores claves para el desarrollo de la Cooperativa de ahorro y crédito Maquita Cushunchic, fecha acceso: Enero 2011, dirección: <http://www.uasb.edu.ec/UserFiles/363/File/pdfs/SUBSITIO%20FIDA/ENFOQUES%20Y%20EXPERIENCIAS%20DEL%20DESARROLLO%20RURAL/Cooperativa%20Maquita%20Cushunchic.pdf>

- Interna: auditoria, evaluación de personal, consejos
- Externa: auditoria, evaluación de riesgo, Mix Market (*5 diamantes*), premisa a la transparencia financiera.
- Gestión integral de riesgos: liquidez, mercado, crédito y operativo.
- Observancia a indicadores propuestos por el ente regulador en cuanto a:
 - Clasificación de cartera
 - Calificación de cartera (*buro de crédito*)
 - Constitución de provisiones
 - Suficiencia patrimonial
 - Límites créditos vinculados
 - Límites riesgo de crédito individual

2.1.6 CONTEXTO MACRO-ECONOMICO

El Ecuador es un país pequeño con una de las economías más abiertas de la región. Es muy dependiente de pocos rubros de exportación³⁹ y de las remesas del exterior que representan la segunda fuente del PIB total. Luego de una fuerte crisis económica a finales del los '90, el país ha adoptado el dólar estadounidense como divisa legal (*Marzo 2000*). De ahí ha arrancado un proceso de lenta recuperación, sobre todo a los elevados precios de exportación del petróleo. Aun el PIB per-capita esté aumentando, la tasa de desempleo sigue creciendo y el 60% de la población ecuatoriana vive en condiciones de pobreza.

En el 2004 la economía ha crecido del 6,9% (*PIB*) en un contexto de inflación a la baja (*1,9% en Diciembre 2004*) y de tasas de interés, pasiva y activa, en disminución constante en el tiempo. Hay previsiones de un crecimiento del PIB en 2005 del 3,7%, en un contexto de inflación controlada.

³⁹ Petróleo y bananas representan casi el 65% de las exportaciones totales, en Diciembre 2004.

Desde el punto de vista político, es un contexto altamente inestable con casos de corrupción y frecuentes cambios de presidentes y gobiernos.

2.1.7 EL SECTOR DE LAS MICROFINANZAS

La oferta de microcrédito ha crecido rápidamente en los últimos años y está siempre más concentrada en las instituciones financieras reguladas.

Los principales actores son cuatro bancos (*Banco Solidario, Banco del Pichincha con Credifé, ProCredit y Banco Centromundo*⁴⁰) que cuentan con una cartera total para microempresa⁴¹ a Junio 2005 de US\$ 207,3 millones. Las 35 cooperativas de ahorro y crédito reguladas manejan una cartera de microempresa total de US\$ 248,7 millones⁴².

MAQUITA CUSHUNCHIC se encuentra aún entre las no reguladas. En tamaño de la cartera para microempresas sin embargo hoy en día se posicionaría entre las primeras 23 reguladas.

Cabe notar que está en acto un proceso de upgrading de las instituciones financieras: las sociedades financieras se vuelven Bancos (*Sociedad Financiera Ecuatorial transformada en Banco Procredit*), las Cooperativas reguladas aumentan de número, las ONG se transforman en reguladas (FINCA⁴³). Entre las ONG especializadas en microcrédito destacan por tamaño de la cartera activa el FED (*Fondo de Desarrollo Ecuatoriano*), D-MIRO de la Misión Alianza Noruega y ESPOIR con cartera entre US\$ 3,5 y 5 millones. Juntos a las COAC no reguladas ellas siguen jugando un papel importante en cuanto a cobertura geográfica y a alcance en profundidad del servicio de crédito.

En Quito empiezan a manifestarse los primeros síntomas de sobreendeudamiento, sobre todo por causa del boom del crédito al consumo, un microcrédito a sola firma, ofrecido por algunos Bancos.

⁴⁰ Banco Solidario con US\$ 78,7 millones, Credifé con US\$ 53,1 millones, ProCredit con 45,9 millones y Centromundo con 29,6 millones..

⁴¹ Según disposiciones de ley, otras categorías de cartera de crédito son "consumo", "comercial" y "vivienda", cada una con sus especificaciones y disposiciones de provisiones por incobrables

⁴² La cartera mayor, US\$ 20,7 millones, es de la cooperativa Mego.

⁴³ FINCA detiene una cartera de microempresas (*Bancos comunales principalmente*) de US\$ 13,2 millones

Cabe aquí señalar el riesgo relacionado a la “contaminación” del mercado del microcrédito y a la consecuente aparición de problemas de mora, como el caso boliviano enseña. Pese a que MAQUITA CUSHUNCHIC esté concentrada en la ciudad de Quito, su mercado es principalmente el de la zona sur de Quito, menos congestionada que otras zonas de la ciudad.

Un aspecto positivo que cabe señalar es la difusión de buró de créditos privados. Todos incluyen los datos de las instituciones financieras reguladas, solo algunos incorporan datos de empresas comerciales y solo uno datos de las microfinancieras no reguladas.

2.1.8 POSICIONAMIENTO EN EL MERCADO⁴⁴

La Cooperativa brinda servicios financieros a microempresarios de escasos recursos en el área urbana de Quito, que es el mercado más competitivo del país. En algunos barrios de la ciudad los microempresarios se encuentran sobre-endeudados por la elevada oferta de créditos de microempresa y consumo. El riesgo de sobreendeudamiento es solo parcialmente mitigado por el hecho que la Cooperativa trabaje sobretodo en los barrios de la zona sur, menos “congestionados” que en el norte y centro de la ciudad. Además la Cooperativa consulta sistemáticamente una central de riesgo (*CINFOCREDIT*) aunque la misma no incluya las informaciones de todas las microfinancieras no reguladas.

Los principales competidores son los Bancos especializados en microfinanzas: Banco Solidario, Banco Centromundo y Credifé del Banco del Pichincha. Otros competidores son la Fundación Ecuatoriana de Desarrollo (*FED*), FINCA y Unibanco. Cabe notar que las Cooperativas de ahorro y crédito de Quito no representan una competencia para MAQUITA CUSHUNCHIC pues los productos que ofrecen, con encaje elevado y requerimiento frecuente de garantía hipotecaria, no son atractivos para los microempresarios.

⁴⁴ Posicionamiento de la Cooperativa de ahorro y crédito Maquita Cushunchic en el mercado. fecha acceso: Enero 2011, dirección: http://www.ratinginitiative.org/uploads/tx_dbreports/COAC_Maquita_Cushunchic_informe_final_de_calificacion_Julio_2005.pdf

Las principales ventajas comparativas de MAQUITA CUSHUNCHIC frente a sus competidores son las siguientes:

- Baja tasa efectiva del crédito, elemento, que en un mercado casi maduro como el de Quito representa un factor competitivo;
- Plazo más largo que permite otorgar montos promedios elevados manteniendo cuotas de reembolso pequeñas; y
- Mayor rapidez de desembolso del crédito y no requerimiento de encaje, comparado con las cooperativas.

Las principales desventajas comparativas están relacionadas a una presencia limitada en el territorio nacional y a una limitada tipología de servicios financieros (*sobretudo comparado con grandes Bancos*). Además, los Bancos especializados en crédito al consumo son muy rápidos (*dos días*) en el desembolso de créditos.

La Cooperativa está orientándose hacia una nueva estrategia de expansión para alcanzar microempresarios de más escasos recursos y salir al menos en parte del mercado más atractivo para las microfinancieras reguladas.

2.1.9 IMPACTO SOCIAL⁴⁵

El impacto social en el sistema de medición de desempeño de la Cooperativa esta dado por:

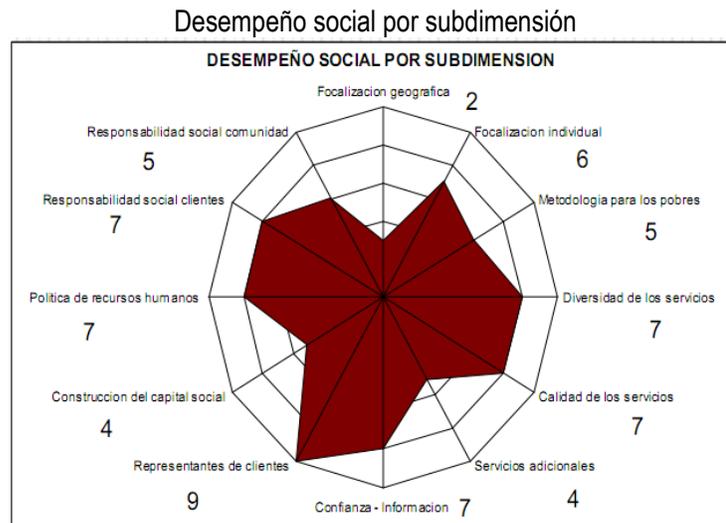
- El puntaje de desempeño social por subdimensión obtenido es de 74/100 puntos.
- El puntaje promedio de desempeño social por subdimensión de Cooperativas en América Latina obtenido es de 63/100 puntos.

El impacto social en los colaboradores esta dado por:

- Participación del personal en la toma de decisiones

⁴⁵ Impacto social: de medición de desempeño social de los colaboradores, la comunidad, socios y de la Cooperativa de ahorro y crédito Maquita Cushunchic, fecha acceso: Enero 2011, dirección: <http://www.uasb.edu.ec/UserFiles/363/File/pdfs/SUBSITIO%20FIDA/ENFOQUES%20Y%20EXPERIENCIAS%20DEL%20DESARROLLO%20RURAL/Cooperativa%20Maquita%20Cushunchic.pdf>

- Liderazgo creciente y estructura participativa
- Capacitación de calidad en todos los cargos
- Plan de carrera dentro de la Cooperativa
- Incentivos monetarios y no monetarios
- Reducción de brecha entre sueldo más alto y más bajo
- Oportunidades de desarrollo humano en el compromiso con el servicio a los más necesitados



Fuente: IMPACTO SOCIAL: Sistema de medición de desempeño social.
<http://www.uasb.edu.ec/UserFiles/363/File/pdfs/SUBSITIO%20FIDA/ENFOQUES%20Y%20EXPERIENCIAS%20DEL%20DESARROLLO%20RURAL/Cooperativa%20Maquita%20Cushunchic.pdf>

GRAFICO 2.1

El impacto social en la comunidad está dado por:

- Hospital Padre José Carollo. Un Canto a la Vida.
- Centro de Desarrollo Integral El Niño
- Programa de salud comunitaria (*atención médica a barrios marginales*)
- Programa de atención a adultos mayores
- Gestiones de la Cooperativa para obtener apoyo para estas obras por parte de otras empresas o personas
- Casa de la Familia. Quito Sur

El impacto social con socios esta dado por:

- Capacitación a directivos y representantes
- Tasas activas que permiten el desarrollo del negocio
- Tasas pasivas no dependen del monto, se premia el ahorro de todos por igual
- Existe subsidio cruzado de créditos de mayor monto a créditos de menor monto
- Existencia de seguro de desgravamen para todos los clientes con crédito
- Entrega de seguro de salud para premiar fidelidad de los clientes y para créditos de montos más bajos
- Evita el sobreendeudamiento a través de políticas de crédito
- Desarrollo de productos y servicios para mejorar profundización (*llegar a las personas de menores recursos*)
- Condiciones de crédito ajustadas a necesidades de los clientes
 - No progresividad en los créditos
 - Consideración del ingreso de la unidad familiar
 - Toma en cuenta el destino del crédito
- Gestión para convenios que permitan conceder crédito a emprendedores y emigrantes

CAPÍTULO III: ANALIZAR Y PREPARAR ESTADÍSTICAMENTE UN CONJUNTO DE VARIABLES PARA EL DISEÑO DE UN MODELO DE APROBACION CREDIT SCORE DE CARTERA DE CONSUMO

3.1 DESCRIPCION DE LAS VARIABLES

Uno de los más importantes elementos a tomar en cuenta al desarrollar un modelo de gestión de riesgo tipo credit score, es la calidad de la información que se encuentra almacenada en la base de datos de la Institución.

Los campos de las tablas de las bases de datos cuando estén en un archivo plano se convertirán en variables (*numéricas, cadenas, caracteres, binarias*) previo al análisis de un modelo de gestión de riesgo tipo credit score.

Tenemos 61 variables de la base de datos que nos fueron proporcionados por la Institución, las mismas que son mostradas en la siguiente tabla:

VARIABLES DE LA BASE DE DATOS DE CARTERA DE CONSUMO

No	VARIABLE	DESCRIPCION	TIPO
1	ALQ_PROP	valor del alquiler, se valida si es vivienda alquilada o propia	numérico
2	ANIO	año del crédito	numérico
3	BONIF_SUB	alguna bonificación del cliente	numérico
4	CART_CASTIG	cartera castigada	numérico
5	CART_JUDICIAL	cartera judicial	numérico
6	CART_NO_DEVEN	cartera no devenga	numérico
7	CART_POR_VENCER	cartera por vencer	numérico
8	CART_VENCIDA	cartera vencida	numérico
9	COMISIONES	alguna comisión extra del cliente	numérico
10	CTAS_C_RH_2_ULT	cuentas cerradas en los 2 últimos años	numérico
11	CTAS_C_RH_5_PEN	cuentas cerradas en los 5 últimos años	numérico
12	CTAS_C_RH_ULT	cuentas cerradas ultimo año	numérico
13	CTAS_C_VIG	cuentas cerradas vigentes (SIB)	numérico
14	CUOTA	valor de la cuota	numérico
15	CUOTAS_PAG	cuotas pagadas	numérico
16	DEUDAS_DIREC	deudas directas	numérico
17	DEUDAS_INDIREC	deudas indirectas	numérico

19	DIAS_MORA_MAX	días de mora máximo	numérico
20	DIAS_MORA_PROM	días de mora promedio	numérico
21	EDAD	edad	numérico
22	INGRESOS	ingresos	numérico
23	MIEMBROS_FAM	miembros de la familia	numérico
24	MONTO	monto del crédito	numérico
25	N_TRAMITE	numero del tramite	numérico
26	N_VMORA_M30	número de veces que cayó en mora mayor que 30 días	numérico
27	N_VMORA_M45	número de veces que cayó en mora mayor que 45 días	numérico
28	N_VMORA_M5	número de veces que cayó en mora mayor que 5 días	numérico
29	N_VMORA_M60	número de veces que cayó en mora mayor que 60 días	numérico
30	NUM_CUOTAS	numero de cuotas	numérico
31	PATRIMONIO	patrimonio	numérico
32	RECIPROCIDAD	reciprocidad con la Coop	numérico
33	RENTA_JUBILAR	renta jubilar	numérico
34	SALDO_OPER	saldo operativo	numérico
35	SUELDO_BASICO	sueldo básico	numérico
36	TASA	tasa	numérico
37	UTILIDADES	utilidades	numérico
38	ACTIVIDAD	actividad económica	string
39	CAL_C_R	calificación de la central de riesgos	string
40	DESTINO_CRED	destino del crédito	string
41	E_A_RES	extranjero años de residencia	string
42	EST_CIVIL	estado civil	string
43	ESTAB_TRAB_ACT	estabilidad del trabajo actual	string
44	ESTAB_TRAB_ANT	estabilidad del trabajo anterior	string
45	EXP_CRED_COOP	calificación de riesgo de crédito Coop.	string
46	GAR_EMP_DEU	garante que este empleado del deudor	string
47	GAR_M_NEG_DEU	garante mismo negocio del deudor	string
48	GENERO	genero	string
49	INGRESOS_ADIC	ingresos adicionales	string
50	N_PROT_INJUST	numero de protestos injustificados	string
51	N_PROT_JUST	numero de protestos justificados	string
52	NIVEL_EDUC	nivel de educación	string
53	POLIC_MIL	policía/militar	string
54	PROFESION	profesión	string
55	SECTOR	sector	string
56	SOLIC_PROP_NEG	solicitante propietario del negocio	string
57	TIEMPO_NEG	tiempo del negocio	string
58	TIPO_CREDITO	tipo de crédito	string
59	TIPO_VIVIENDA	tipo de vivienda	string
60	FECHA_CONS	fecha de concesión	Fecha
61	FECHA_VCTO	fecha de vencimiento	fecha

Fuente: Base de Datos proporcionado por la Cooperativa Maquita Cushunchic

TABLA 3.1

3.2 DEFINICION DE INCUMPLIMIENTO

Se define como incumplimiento a la falta de pago de un préstamo o cualquier otro tipo de violación de las condiciones de un contrato de préstamo. El incumplimiento es un cambio de estado cuando el prestamista falla en pagar su préstamo⁴⁶. Los modelos para gestionar el riesgo crédito se sustentan en el cálculo de probabilidades de incumplimiento, para lo cual se deben diseñar modelos de calificación y aprobación de riesgo tipo credit score que permitirán asignar la categoría de riesgo a la cual pertenece cada cliente, en función de variables cualitativas y cuantitativas referentes a las características y comportamiento de los sujetos de crédito. Estos sistemas parten de la definición del concepto de incumplimiento, es decir la Institución concibe que los clientes dejen de tener un comportamiento de pago voluntario, siendo necesario iniciar acciones de cobranza, para minimizar las posibles pérdidas que pueden presentarse⁴⁷.

El modelo permitirá establecer conjunto de variables propias de la Institución, el mismo que nos permita definir y/o determinar el incumplimiento para las operaciones de la cartera de consumo. El incumplimiento se lo maneja como una variable dependiente que nos ayudara a evidenciar los denominados “*clientes buenos*” y “*clientes malos*”.

3.3 VARIABLE DEPENDIENTE “TIPO PAGADOR”

Para la definición de cual es un “*buen cliente*” o un “*mal cliente*” es importante definir qué es un cliente...? Un cliente es la persona, empresa u organización que adquiere o compra de forma voluntaria productos o servicios que necesita o desea para sí mismo, para otra persona o para una

⁴⁶ Definición de incumplimiento, fecha acceso: Enero 2011, dirección: <http://es.mimi.hu/economia/incumplimiento.html>

⁴⁷ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010.

empresa u organización; por lo cual, es el motivo principal por el que se crean, producen, fabrican y comercializan productos y servicios ⁴⁸.

Un cliente puede llegar a ser “bueno” o “malo” luego de analizar el incumplimiento atado al análisis del número de días que el cliente a caído en “mora”. Para definir los clientes buenos y malos como variable objetivo, se establecerá una matriz de contingencia para las variables de mora máxima y mora promedio⁴⁹, esta nos permitirá identificar al “mal cliente” que será el cliente a quien la Institución ya no debería comercializar sus productos y servicios.

“Buen Cliente”: Aquellos clientes que registren durante la vigencia del crédito de cartera de consumo con mora o atrasos máximos de 0 a 15 días y mora o atrasos promedio de 0 a 15 días.

“Mal Cliente”: Aquellos clientes con atrasos promedio mayor o igual a los 16 días, mas los clientes con atraso máximo mayor o igual a 16 días.

La variable dependiente “tipo pagador” del modelo a desarrollar se define como se muestra:

La variable dependiente “tipo pagador” del modelo a desarrollar

Tipo	Definición en días
buen cliente	atraso promedio <= 15 días atraso máximo <= 15 días
mal cliente	Demás atrasos promedios y máximos

Fuente: TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010

TABLA 3.2

Matriz de contingencia para las variables de mora máxima y mora promedio para determinar la variable dependiente “tipo pagador”.

⁴⁸ Definición de que es un cliente?, fecha acceso: Enero 2011, dirección: <http://www.promonegocios.net/clientes/cliente-definicion.html>

⁴⁹ Los datos son los comportamientos que han tenido los clientes de este tipo de crédito, a fin de evaluar la peor calificación a la que ha llegado, es decir, cuantos días de retraso en el pago de una cuota han acumulado en un determinado mes, lo que determinara si ese socio es buen pagador o mal pagador.

Definición del "tipo pagador" de acuerdo a la morosidad del pago

mora promedio (días)	mora máxima (días)						
	0	1 - 15	16 - 30	31 - 60	61 - 90	91 - 120	> 120
0							
1 - 15							
16 - 30							
31 - 60							
61 - 90							
91 - 120							
> 120							

"buen cliente"
 "mal cliente"

Fuente: El Autor

TABLA 3.3

Con la variable "tipo pagador" se pueden categorizar y clasificar aquellos clientes considerados como buenos y malos, esta variable será el punto de partida para la construcción del modelo. Con el apoyo de esta variable se crea entonces la variable 62 que representara el "Incumplimiento" que la utilizaremos para el análisis y construcción del modelo de gestión de riesgo tipo credit score.

3.4 ANALISIS DE LAS VARIABLES EXPLICATIVAS

Todas las variables excepto el incumplimiento se agregaran individualmente o en ciertos casos como variables compuestas; siempre y cuando las variables permitan interpretar o explicar el comportamiento de pago del cliente durante la vida del crédito. Las variables que utilizaremos son: variables de perfil, variables demográficas y variables económicas del cliente. Utilizando análisis estadístico y apoyado con algún paquete estadístico, se tomara como pivote el incumplimiento y se tratara de determinar las variables que intervendrán en el modelo de gestión de riesgo tipo credit score. En el ANEXO 1 realizaremos un análisis exploratorio de variables cuantitativas y cualitativas del conjunto de variables entregadas por la Institución.

3.5 PREPARAR VARIABLES PARA UN MODELO CREDIT SCORE

3.5.1 MODELO CREDIT SCORE⁵⁰

Se logro determinar a los buenos clientes y malos clientes para la cartera de consumo en la variable *incumplimiento*, ahora hay que determinar su influencia en la administración del riesgo de crédito como una de las principales causas de pérdida. El modelo credit score se presenta de dos tipos: modelo de aprobación⁵¹ y modelo de calificación⁵². El análisis y preparación estadística de variables será orientado al diseño de un modelo de aprobación como técnica de regresión del modelo de crédito.

Parte del proceso del diseño del modelo es calcular la probabilidad de incumplimiento del sujeto de crédito, se realiza un análisis univariante, un análisis bivariante y la aplicación de una regresión logística⁵³ que nos permitirá obtener una función donde la variable dependiente es el *incumplimiento*. Se trabaja en un proceso de asociación, suponiendo que el ser humano tiene un comportamiento social parecido, es decir nosotros nos comportamos como nuestros pares, en nuestra base de datos trataremos de entender las particularidades y perfil de riesgo de la Institución asignado al mercado y a los productos que ofrece, el objetivo principal es desarrollar un modelo propio que se adapte a la Cooperativa.

⁵⁰ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.

⁵¹ Modelo de Aprobación: Tiene la característica de calificar por asociación al sujeto de crédito, que no lo conocemos, partiendo de que el ser humano tiene un comportamiento social, es decir nosotros nos comportamos como nuestros pares. El modelo también es conocido como reactivo, ya que responden a una solicitud de crédito por parte del cliente, y para medir el riesgo se sustentan básicamente en los datos que aporta el cliente de su información verificada en su solicitud de crédito, antecedentes históricos, y reportada en la central de riesgo, y son utilizados para evaluar al cliente como apto o no para otorgarle un crédito.

⁵² Modelo de calificaciones: Lo utilizamos para determinar un comportamiento de pago del sujeto de crédito, con una característica importante que es la calificación de riesgo, donde se reemplaza el cuadro de calificación de riesgo por un modelo de calificación que contendrá datos de operación, comportamiento. Este permitirá validar el comportamiento de pago en la institución y en el sistema, en definitiva tener el valor correcto de la provisión.

⁵³ Modelo Logit: O denominado modelo Logístico, proporciona la probabilidad de que una determinada observación pertenezca a un cierto grupo, conocidos los valores de las variables independientes para esa observación, la ecuación del modelo es la siguiente:

$$P_j = \frac{e^{z_j}}{1 + e^{z_j}}; z_j = B_1 Z_1 + B_2 Z_2 + \dots + B_p Z_p$$

Para el análisis y preparación estadística de variables para el diseño de un modelo credit score de cartera de consumo utilizaremos herramientas estadísticas como: SPSS v.15.0; Eviews v.5.1; Microsoft Excel.

Modelo propio que se adapte a la institución:⁵⁴ El desarrollo de un modelo propio que se adapte a la Institución, permite el cálculo de la pérdida esperada bajo un enfoque de Basilea II⁵⁵, además de la oportunidad que todos los componentes de la pérdida esperada sean calculados por la propia institución. Sin embargo estos modelos tienen que ser sometidos a un proceso de supervisión y evaluación por parte de la SBS, para que la Institución pueda implementar y emplear el cálculo de las provisiones de acuerdo a un resultado cuantitativo que entregaría el modelo.

3.5.2 ANALISIS DE VARIABLES⁵⁶

En la primera pasada (*ANEXO 1*) se realizó un análisis de variables, donde se determinó que existen variables que van a ser descartadas, debido a que no aportan al diseño del modelo de credit score, de 61 variables iniciales nos quedaron 26 variables independientes que NO fueron descartadas (*7 variables son cualitativas y 19 variables son cuantitativas*). Ahora estas variables serán candidatas a un siguiente análisis bivalente⁵⁷ tratando de buscar las variables que mejor expliquen el comportamiento entre los clientes buenos y malos, de acuerdo a una definición de incumplimiento.

⁵⁴ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.

⁵⁵ Es el segundo de los acuerdos de Basilea. Dichos acuerdos consisten en recomendaciones sobre la legislación y regulación bancaria y son emitidos por el Comité de supervisión bancaria de Basilea. El propósito de Basilea II, publicado inicialmente en junio de 2004, es la creación de un estándar internacional que sirva de referencia a los reguladores bancarios, con objeto de establecer los requerimientos de capital necesarios, para asegurar la protección de las entidades frente a los riesgos financieros y operativos. Estas se apoyan en los siguientes tres pilares:

Pilar I: el cálculo de los requisitos mínimos de capital

Pilar II: el proceso de supervisión de la gestión de los fondos propios

Pilar III: La disciplina de mercado

Fuente de consulta: Página de internet relacionado con el acuerdo de Basilea, Wikipedia. http://es.wikipedia.org/wiki/Basilea_II

⁵⁶ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.

⁵⁷ El análisis bivalente relaciona dos o más indicadores de manera que se pueda estudiar una variable en función de otra determinada, por ejemplo, un típico análisis bivalente son las tablas cruzadas donde se refleja el conocimiento de una marca en función de la edad, los niveles de instrucción o el estrato socioeconómico de los entrevistados

Definiciones básicas de análisis univariante, bivalente y multivariante, fecha acceso: Enero 2011, dirección: <http://www.serviciosestadisticos.es/metodologia/metodologia.htm>

El análisis bivariante se realiza a través de tablas de contingencia de la variable dependiente incumplimiento vs. las variables independientes. La capacidad de discriminación se observa a través del *porcentaje de buenos y malos* y complementando con el nivel de significancia (*probabilidad que la variable no explique el incumplimiento Ho*), todo el análisis realizado se muestra en el ANEXO 2.

**Variables de la base de datos de cartera de consumo
primera pasada (ANEXO 1) – segunda pasada (ANEXO 2)**

No	VARIABLE	DESCRIPCION	TIPO	1RA PASADA	2DA PASADA
1	ALQ_PROP	valor del alquiler, se valida si es vivienda alquilada o propia	numérico	DESCARTADA	
2	ANIO	año del crédito	numérico	DESCARTADA	
3	BONIF_SUB	alguna bonificación del cliente	numérico	DESCARTADA	
4	CART_CASTIG	cartera castigada	numérico		OK
5	CART_JUDICIAL	cartera judicial	numérico	DESCARTADA	
6	CART_NO_DEVEN	cartera no devenga	numérico		OK
7	CART_POR_VENCER	cartera por vencer	numérico		OK
8	CART_VENCIDA	cartera vencida	numérico		OK
9	COMISIONES	alguna comisión extra del cliente	numérico	DESCARTADA	
10	CTAS_C_RH_2_ULT	cuentas cerradas en los 2 últimos años	numérico	DESCARTADA	
11	CTAS_C_RH_5_PEN	cuentas cerradas en los 5 últimos años	numérico	DESCARTADA	
12	CTAS_C_RH_ULT	cuentas cerradas ultimo año	numérico	DESCARTADA	
13	CTAS_C_VIG	cuentas cerradas vigentes (SIB)	numérico	DESCARTADA	
14	CUOTA	valor de la cuota	numérico		OK
15	CUOTAS_PAG	cuotas pagadas	numérico		DESCARTADA
16	DEUDAS_DIRECT	deudas directas	Numérico		DESCARTADA
17	DEUDAS_INDIREC	deudas indirectas	Numérico		DESCARTADA
18	DIAS_MORA	días de mora	numérico	INCUMPLIMIENTO	
19	DIAS_MORA_MAX	días de mora máximo	numérico	INCUMPLIMIENTO	
20	DIAS_MORA_PROM	días de mora promedio	numérico	INCUMPLIMIENTO	
21	EDAD	edad	numérico		OK
22	INGRESOS	ingresos	numérico		OK
23	MIEMBROS_FAM	miembros de la familia	numérico		OK
24	MONTO	monto del crédito	numérico		OK
25	N_TRAMITE	numero del tramite	Numérico	DESCARTADA	
26	N_VMORA_M30	núm. veces que cayó en mora mayor que 30 días	Numérico	DESCARTADA	
27	N_VMORA_M45	núm. veces que cayó en mora mayor que 45 días	Numérico	DESCARTADA	
28	N_VMORA_M5	núm. veces que cayó en mora mayor que 5 días	numérico		OK
29	N_VMORA_M60	núm. veces que cayó en mora mayor que 60 días	Numérico	DESCARTADA	
30	NUM_CUOTAS	numero de cuotas	numérico		DESCARTADA
31	PATRIMONIO	Patrimonio	numérico		DESCARTADA
32	RECIPROCIDAD	reciprocidad con la Coop	numérico		OK
33	RENTA_JUBILAR	renta jubilar	numérico	DESCARTADA	
34	SALDO_OPER	saldo operativo	numérico		OK

35	SUELDO_BASICO	sueldo básico	numérico		DESCARTADA
36	TASA	Tasa	numérico		DESCARTADA
37	UTILIDADES	Utilidades	numérico	DESCARTADA	
38	ACTIVIDAD	actividad económica	string	DESCARTADA	
39	CAL_C_R	calificación de la central de riesgos	string		OK
40	DESTINO_CRED	destino del crédito	string	DESCARTADA	
41	E_A_RES	extranjero años de residencia	string	DESCARTADA	
42	EST_CIVIL	estado civil	string		OK
43	ESTAB_TRAB_ACT	estabilidad del trabajo actual	string		DESCARTADA
44	ESTAB_TRAB_ANT	estabilidad del trabajo anterior	string	DESCARTADA	
45	EXP_CRED_COOP	calificación de riesgo de crédito Coop.	string		OK
46	GAR_EMP_DEU	garante que este empleado del deudor	string	DESCARTADA	
47	GAR_M_NEG_DEU	garante mismo negocio del deudor	string	DESCARTADA	
48	GENERO	Genero	string		DESCARTADA
49	INGRESOS_ADIC	ingresos adicionales	string	DESCARTADA	
50	N_PROT_INJUST	numero de protestos injustificados	string	DESCARTADA	
51	N_PROT_JUST	numero de protestos justificados	string	DESCARTADA	
52	NIVEL_EDUC	nivel de educación	string		OK
53	POLIC_MIL	policía/militar	string	DESCARTADA	
54	PROFESION	Profesión	string	DESCARTADA	
55	SECTOR	Sector	string	DESCARTADA	
56	SOLIC_PROP_NEG	solicitante propietario del negocio	string	DESCARTADA	
57	TIEMPO_NEG	tiempo del negocio	string	DESCARTADA	
58	TIPO_CREDITO	tipo de crédito	string	DESCARTADA	
59	TIPO_VIVIENDA	tipo de vivienda	string	DESCARTADA	
60	FECHA_CONS	fecha de concesión	fecha	DESCARTADA	
61	FECHA_VCTO	fecha de vencimiento	fecha	DESCARTADA	

Fuente: Base de Datos proporcionado por la Cooperativa Maquita Cushunchic

TABLA 3.1

El análisis univariante y bivariante a las variables independientes de la base de datos de la Cooperativa han permitido descartar aquellas variables que “no aportan” al desarrollo del modelo de credit score de cartera de consumo, ya que su contribución es mínima, redundante y/o que la variable pueda ser explicada por otra variable mas explicativa; al final tendremos un conjunto de variables independientes “candidatas”, sin embargo tenemos que tomar en cuenta que nuestro estudio está enfocado a realizar *un modelo de aprobación*⁵⁸ de credit score de la cartera de

⁵⁸ Modelo de Aprobación: Tiene la característica de calificar por asociación al sujeto de crédito, que no lo conocemos, partiendo de que el ser humano tiene un comportamiento social, es decir nosotros nos comportamos como nuestros pares. El modelo también es conocido como reactivo, ya que

consumo, donde este modelo tiene el objetivo medir el riesgo y se sustentan básicamente en los datos que aporta el cliente de su información verificada en su solicitud de crédito, entonces las variables independientes que hasta ahora tenemos son las siguientes:

**VARIABLES DE LA BASE DE DATOS DE CARTERA DE CONSUMO
VARIABLES INDEPENDIENTES "CANDIDATAS"**

CAL_C_R: calificación de la central de riesgos
 EXP_CRED_COOP: calificación de riesgo de crédito Cooperativa
 EST_CIVIL: estado civil
 MIEMBROS_FAM: miembros de la familia, cargas familiares
 NIVEL_EDUC: nivel de educación
 EDAD: edad
 INGRESOS: ingresos
 RECIPROCIDAD: reciprocidad del cliente con la Cooperativa
 SALDO_OPER: saldo operativo en la Cooperativa
 MONTO: monto del crédito
 CUOTA: valor de la cuota

Fuente: Base de Datos proporcionado por la Cooperativa Maquita Cushunchic.- Variables de la base de datos de cartera de consumo, primera pasada (ANEXO 1) – segunda pasada (ANEXO 2).

TABLA 3.2

3.5.3 AJUSTE AL MODELO LOGISTICO⁵⁹

Con el conjunto de variables independientes "*candidatas*", ajustaremos a un modelo logístico con la incorporación paso a paso⁶⁰ de las variables explicativas con la variable dependiente a fin de obtener un modelo logístico que se ajuste a la ecuación Logístico:

$$P_j = \frac{e^{z_j}}{1 + e^{z_j}}; z_j; \text{sabiendo que } Z_j = B_0 + B_1Z_1 + B_2Z_2 + \dots + B_pZ_p \quad (\text{ec. 3})$$

responden a una solicitud de crédito por parte del cliente, y para medir el riesgo se sustentan básicamente en los datos que aporta el cliente de su información verificada en su solicitud de crédito, antecedentes históricos, y reportada en la central de riesgo, y son utilizados para evaluar al cliente como apto o no para otorgarle un crédito.

⁵⁹ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.

⁶⁰ Método paso a paso: O denominado step-wise, consiste en ir acondicionando variables a un modelo, usando el estadístico G (Wald), para saber si la nueva variable explica significativamente la variable respuesta. De igual manera, este procedimiento elimina variables ya incluidas que por el efecto de la inclusión de la variable, esta puede ser no significativa.

Se realizara entonces las corridas necesarias con diferentes combinaciones de variables y finalmente se escogerán las corridas con mayor significación estadística, tratando de utilizar la definición que el valor de p^{61} sea menor a 0.05 al 95% de confianza.

Luego de realizar varias corridas con las variables independientes “*candidatas*” se escoge el modelo de regresión donde los parámetros estadísticos, estadísticos de prueba revelen mejor ajuste. Se escoge el de mejor R^2 con menor sumatoria del error, las diversas corridas del modelo logístico se muestran en el (ANEXO 3), se demuestra también que las variables de la TABLA 3.2 son efectivamente las que mejor contribuyen a fin de obtener un modelo logístico tipo credit score de la cartera de consumo, que apoye a la gestión de riesgo de crédito de la Cooperativa de ahorro y crédito Maquita Cushunchic.

En la antepenúltima corrida se observa que la significancia de unas categorías de la variable “*CAL_CR_COOP_REC*” presenta un mal ajuste por lo que esta categoría se debe excluir categorizando con una nueva variable binaria donde: A=1; B=0; C=0, además se observa que la significancia de unas categorías de la variable “*CAL_CR_REC*” presenta un mal ajuste por lo que esta categoría se debe excluir categorizando con una nueva variable binaria donde: A=1; B=0; C=0.

Las variables independientes resultantes con sus respectivos coeficientes asociados luego de realizar las corridas con la regresión logística se muestran en la siguiente tabla.

variables independientes resultantes con sus respectivos coeficientes asociados luego de realizar las corridas con la regresión logística

No	VARIABLE	DESCRIPCION	COEFICIENTE	SIGNIFICANCIA
1	CAL_CR_COOP_REC3(2)	calificación de riesgo (B o C o D o E) de crédito Coop. al momento de la concesión.	0.757756	0.000020838446
2	CAL_CR_REC3(2)	calificación de riesgo (B o C) de central de riesgos al momento de la concesión.	0.443445	0.072936034109
3	CUOTA	valor de la cuota al solicitar el crédito	-0.004122	0.005320307179

⁶¹ Valor de p , se emplea con frecuencia en las pruebas de hipótesis y representan la probabilidad de que la estadística de prueba tomaría un valor al menos tan extremo como el valor observado en ella cuando H_0 es verdadera, es decir es el nivel de significancia más pequeño que conduciría al rechazo de H_0 .

4	SALDO_OPER	saldo operativo con la Coop. al momento de la concesión	-0.000218	0.000000000000
5	CAPACIDAD_PAGO	%capacidad de pago (ingresos-gastos/cuota)	0.010969	0.426692790862
6	EDAD	edad del cliente	-0.021469	0.000000000000
7	MONTO	monto del crédito	0.000292	0.000000711525
8	RECIPROCIDAD	reciprocidad del cliente con la Coop. al momento de la concesión	-0.002177	0.000000073817
9	INGRESOS2	ingresos del cliente	0.000374	0.004141810563
10	MIEMBROS_FAM	miembros de la familia del cliente	-0.132056	0.000000039748
11	C	constante del modelo	-0.981981	0.000000000031

Fuente: Base de Datos proporcionado por la Cooperativa Maquita Cushunchic.- (ANEXO 3)

TABLA 3.3

La estimación del modelo logístico viene dado por la siguiente función:

$$Z_j = -0.981981 + 0.757756 \cdot \text{CAL_CR_COOP_REC3}(B,C,D,E) + 0.443445 \cdot \text{CAL_CR_REC3}(B,C) - 0.004122 \cdot \text{CUOTA} - 0.000218 \cdot \text{SALDO_OPER} + 0.010969 \cdot \text{CAPACIDAD_PAGO} - 0.021469 \cdot \text{EDAD} + 0.000292 \cdot \text{MONTO} - 0.002177 \cdot \text{RECIPROCIDAD} + 0.000374 \cdot \text{INGRESOS2} - 0.132056 \cdot \text{MIEMBROS_FAM}$$

Donde Z_j representa la puntuación del sujeto de crédito, luego la probabilidad de que este sujeto de crédito cumpla sus obligaciones con la Cooperativa está definido por P_j así:

$$P_j = \frac{e^{Z_j}}{1 + e^{Z_j}}; z_j; \text{sabiendo que } Z_j = B_0 + B_1 Z_1 + B_2 Z_2 + \dots + B_p Z_p$$

Al calcular la probabilidad de P_j si el valor de la probabilidad sea más cercano a cero la posibilidad de impago del sujeto de crédito es menor que si el valor de la probabilidad sea más cercano a uno.

3.5.3.1 ANALISIS ECONOMICO DEL MODELO

La construcción de nuestro modelo de gestión del riesgo crédito se sustentan en el cálculo de probabilidades de *incumplimiento*, para lo cual se diseño un modelo de calificación de riesgo tipo

credit score que permitirán asignar la categoría de riesgo a la cual pertenece cada cliente en función de variables cualitativas y cuantitativas referentes a las características y comportamiento de los sujetos de crédito, en nuestro modelo se observa que existen variables independientes como por ejemplo: *cuota, saldo, saldo operativo, miembros de familia*, que producen el siguiente efecto: "*que mientras más grande es el valor de la variable, menor es el riesgo*" a pesar que estadísticamente el modelo credit score hallado es el que representó en las corridas del modelo logístico el "*mejor*" R^2 .

A continuación se realizará un análisis de las variables del modelo de calificación de riesgo tipo credit score:

(1) *CAL_CR_COOP_REC3(2)* vs $Z^{\beta 2}$: El valor del coeficiente de la variable independiente que representa la calificación de riesgo (*B o C o D o E*) de crédito de la Cooperativa al momento de la concesión actúa de la siguiente manera en el modelo: *cuando la calificación de riesgo sea "A" sumará al valor del coeficiente, caso contrario será cero*, este comportamiento se da porque es una variable dicotómica, esta variable fue aceptada al modelo porque su nivel de significancia es menor al 0.05 esperado.

(2) *CAL_CR_REC3(2)* vs Z_j : El valor de coeficiente de la variable independiente que representa la calificación de riesgo (*B o C*) de la central de riesgos al momento de la concesión actúa de la siguiente manera en el modelo: *cuando la calificación de riesgo sea "A" sumará al valor del coeficiente, caso contrario será cero*, este comportamiento se da porque es una variable dicotómica, esta variable fue aceptada al modelo porque su nivel de significancia que es menor al 0.05 esperado.

⁶² Donde Z_j representa la puntuación del sujeto de crédito, luego la probabilidad de que este sujeto de crédito cumpla sus obligaciones con la Cooperativa está definido por P_j .

(3) *CUOTA vs. Zj*: El valor de coeficiente de la variable independiente que representa el valor de la cuota al momento de solicitar el crédito actúa de la siguiente manera en el modelo: *mientras más grande es el valor de la CUOTA más pequeño es el valor que representa al riesgo*, lo cual va en contra de la lógica económica, pues cuotas altas pueden causar que el cliente caiga en mora al no poder completar parte de la cuota lo cual implicaría mayor riesgo y no menor.

(4) *SALDO_OPER vs. Zj*: El valor de coeficiente de la variable independiente que representa el valor del saldo operativo con la Cooperativa al momento de la concesión actúa de la siguiente manera en el modelo: *mientras más grande es el valor del SALDO OPERATIVO más pequeño es el valor que representa al riesgo*, lo cual va en contra de la lógica económica, si un cliente tiene mucho saldo operativo en el sistema puede ser un sujeto de alto riesgo porque tendría otra deuda que cubrir y posiblemente caiga en mora al no poder completar parte de la cuota de esta nueva deuda adquirida.

(5) *CAPACIDAD_PAGO vs. Zj*: El valor de coeficiente de la variable independiente que representa el porcentaje de la capacidad de pago actúa de la siguiente manera en el modelo: *mientras más grande es el valor de la CAPACIDAD_PAGO más grande es el valor que representa al riesgo*, lo cual va en contra de la lógica, pues mientras mayor es mi capacidad de pago, menor es el riesgo de que caiga en mora.

(6) *EDAD vs. Zj*: El valor de coeficiente de la variable independiente que representa la edad del cliente, mientras más años tenga el sujeto de crédito más pequeño es el valor que será incluido al total que representará el riesgo por Z_j , este efecto podría ser verdadero porque entre más años tiene el sujeto, menor podría ser el riesgo, lógicamente esta variable dependerá de otras variables del modelo y la política de edad que se considere al sujeto de crédito en la Cooperativa.

(7) *MONTO* vs. *Zj*: El valor de coeficiente de la variable independiente que representa el monto del crédito actúa de la siguiente manera en el modelo: *mientras mayor es el MONTO más grande es el valor que representa al riesgo*.

(8) *RECIPROCIDAD* vs. *Zj*: El valor de coeficiente de la variable independiente que representa el valor de la reciprocidad del cliente con la Cooperativa al momento de la concesión actúa de la siguiente manera en el modelo: *mientras más grande es el valor del RECIPROCIDAD más pequeño es el valor que representa al riesgo*, lo cual va en contra de la lógica, pues mientras mayor es el valor de la reciprocidad del sujeto con la Cooperativa se debería considerar que mas “*fiel*” es el sujeto de crédito a la Entidad.

(9) *INGRESOS* vs. *Zj*: El valor de coeficiente de la variable independiente que representa los ingresos del cliente actúa de la siguiente manera en el modelo: *mientras más grande es el valor de los INGRESOS más grande es el valor que representa al riesgo*, la cual se encuentra fuera de la lógica económica, pues mientras mayores son los ingresos de una persona, menor es la posibilidad de que caiga en mora.

(10) *MIEMBROS_FAM* vs. *Zj*: El valor de coeficiente de la variable independiente que representa el número de miembros de la familia del cliente actúa de la siguiente manera en el modelo: *mientras más MIEMBROS DE FAMILIA son, más pequeño es el valor que representa el riesgo*, la cual se encuentra fuera de la lógica económica, pues a mayor número de familia, los gastos son mayores, lo que puede ocasionar una mayor posibilidad de caer en mora.

Con el análisis anterior se puede determinar que existen variables que presentan inconsistencias en SIGNOS, las mismas que se están fuera de la lógica económica: CUOTA, SALDO_OPER,

RECIPROCIDAD, CAPACIDAD_PAGO, INGRESOS, MIEMBROS_FAM, sin embargo se correrá el test de significancia y el nivel de predictividad y discriminación.

3.6 TEST DEL MODELO CREDIT SCORE

Para la evaluación estadística de la confiabilidad del modelo propuesto se aplicara pruebas de significancia y pruebas de predictividad y discriminación.

3.6.1 TEST DE SIGNIFICANCIA DEL MODELO⁶³

El estadístico usado para comparar los modelos logísticos es el *logaritmo del cociente de verosimilitudes*⁶⁴. Con el conjunto de variables independientes “*candidatas*” se construyo un modelo logístico y luego con la incorporación paso a paso⁶⁵ de las variables explicativas con la variable dependiente (*ANEXO 3*) se encontró un modelo credit score. Se empieza con un modelo de una variable y luego se incorpora las demás variables, se elige el modelo que presente el menor cociente de verosimilitudes y que contrasta con el nivel de significancia elegido. Si es menor que el valor crítico... se para el proceso y se elige el modelo simple como mejor modelo, y si es mayor o igual que dicho valor crítico esa variable se excluye del modelo y se vuelve a calcular el logaritmo de cociente de verosimilitudes para la inclusión de las variables restantes, y así sucesivamente hasta que no se pueda incluir ninguna mas.

⁶³ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.

⁶⁴ En todos los modelos de regresión lineal simple y regresión múltiple, los coeficientes del mismo se estiman y los contrastes de hipótesis se realizan del mismo modo, incluso se puede hacer contrastes no solo sobre cada coeficiente, sino también sobre el modelo completo o para comparar modelos a través del estadístico F de Fisher; en el caso de una regresión logística el procedimiento es análogo, y se lo hace mediante el llamado logaritmo de cociente de verosimilitud (*log. likelihood ratio*).

⁶⁵ Método paso a paso: O denominado step-wise, consiste en ir acondicionando variables a un modelo, usando el estadístico G (*Wald*), para saber si la nueva variable explica significativamente la variable respuesta. De igual manera, este procedimiento elimina variables ya incluidas que por el efecto de la inclusión de la variable, esta puede ser no significativa.

Historial de iteraciones^{b,c,d}

Iteración	-2 log de la verosimilitud	Coeficientes												
		Constant	CAL_CR_COOP_REC3(1)	CAL_CR_COOP_REC3(2)	CAL_CR_REC3(1)	CAL_CR_REC3(2)	CUOTA	SALDO_OPER	CAPACIDAD_PAGO	EDAD	MONTO	RECIPROCIDAD	INGRESOS2	MIEMBROS_FAM
Paso 1	7716.046	-1.132	.187	.434	.074	.295	.000	.000	.013	-.010	.000	.000	.000	-.071
2	7458.750	-1.208	.330	.722	.127	.454	-.001	.000	.022	-.019	.000	-.001	.000	-.121
3	7427.189	-1.108	.361	.786	.142	.468	-.002	.000	.019	-.022	.000	-.001	.000	-.133
4	7418.418	-1.003	.345	.764	.142	.447	-.004	.000	.012	-.022	.000	-.002	.000	-.132
5	7417.828	-.982	.340	.758	.142	.444	-.004	.000	.011	-.021	.000	-.002	.000	-.132
6	7417.824	-.982	.339	.758	.143	.443	-.004	.000	.011	-.021	.000	-.002	.000	-.132

a. Método: Introducir

b. En el modelo se incluye una constante.

c. -2 log de la verosimilitud inicial: 7780.553

d. La estimación ha finalizado en el número de iteración 6 porque las estimaciones de los parámetros han cambiado en menos de .001.

La prueba Omnibus (*test de significancia de razón de verosimilitud*), brinda una prueba de ajuste global del modelo a través del estadístico Ji-cuadrado, contrastando la hipótesis nula. La prueba ayuda a determinar qué: si al introducir variables independientes se consigue un incremento significativo del ajuste global, este incremento se valora tomando como referencia un modelo nulo, el ajuste del modelo es fuertemente significativo con $p < 0.05$, incluso $p < 0.001$; por lo que se rechaza la hipótesis nula y se concluye que con las variables consideradas el ajuste mejora significativamente y además se ajusta a la capacidad predictiva del modelo.

Pruebas omnibus sobre los coeficientes del modelo

Paso 1		Chi-cuadrado	gl	Sig.
	Paso	362.728	12	.000
	Bloque	362.728	12	.000
	Modelo	362.728	12	.000

El test de logaritmo de la función de verosimilitud (*-2 veces el logaritmo de verosimilitud*) y el estadístico R^2 de Nagelkerke se muestran en la siguiente tabla.

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	7417.824 ^a	.038	.067

a. La estimación ha finalizado en el número de iteración 6 porque las estimaciones de los parámetros han cambiado en menos de .001.

3.6.2 TEST DE SIGNIFICANCIA DE AJUSTE DE DATOS Y COEFICIENTES DEL MODELO⁶⁶

El test de Wald, el test de K-S (*Kolmogorov Smirnov*), el test de Hosmer-Lemeshow, y el test de razón de ajuste (*AR-ROC-prueba de poder*) permiten evaluar el poder-discriminante del modelo como adecuado y significativo.

TEST DE WALD

El valor del estadístico de Wald en regresión logística, es una prueba de significancia estadística que testea la hipótesis nula que los coeficientes son iguales a cero y realiza las estimaciones por intervalos. Su valor para un coeficiente viene dado por el cociente entre el valor del coeficiente y su error estándar. La obtención de significación ($p < 0.05$) indica que dicho coeficiente es diferente de cero y tenemos que conservar en el modelo. Se comporta como una distribución Ji-cuadrado.

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1	CAL_CR_COOP_REC3			38.733	2	.000	
	CAL_CR_COOP_REC3(1)	.339	.064	28.191	1	.000	1.404
	CAL_CR_COOP_REC3(2)	.758	.178	18.111	1	.000	2.133
	CAL_CR_REC3			7.416	2	.025	
	CAL_CR_REC3(1)	.143	.065	4.797	1	.029	1.153
	CAL_CR_REC3(2)	.443	.247	3.216	1	.073	1.558
	CUOTA	-.004	.001	7.767	1	.005	.996
	SALDO_OPER	.000	.000	52.892	1	.000	1.000
	CAPACIDAD_PAGO	.011	.014	.632	1	.427	1.011
	EDAD	-.021	.003	65.537	1	.000	.979
	MONTO	.000	.000	24.584	1	.000	1.000
	RECIPROCIDAD	-.002	.000	28.962	1	.000	.998
	INGRESOS2	.000	.000	8.221	1	.004	1.000
	MIEMBROS_FAM	-.132	.024	30.162	1	.000	.876
	Constante	-.982	.148	44.130	1	.000	.375

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC3, CAL_CR_REC3, CUOTA, SALDO_OPER, CAPACIDAD_PAGO, EDAD, MONTO, RECIPROCIDAD, INGRESOS2, MIEMBROS_FAM.

TEST DE K-S (*KOLMOGOROV SMIRNOV*)

El valor del test Kolmogorov Smirnov (*K-S*) en regresión logística, es una prueba no paramétrica de bondad de ajuste a una ley continua, de dos funciones de probabilidad empíricas entre sí. La prueba está basada en medidas de separación de distribuciones sobre un conjunto de observaciones

⁶⁶ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.

cuantiles. Mide el grado de concordancia existente entre la distribución de un conjunto de datos y su distribución teórica.

Se calculan las frecuencias y frecuencias acumuladas para un determinado número de clases, se obtiene la máxima distancia K-S entre ambas y se compara el estadístico en la tabla de valores críticos de "D" bajo la hipótesis de distribución normal, para la decisión de aceptar o no la hipótesis.

Valores de la prueba Kolmogorov Smirnov (K-S)

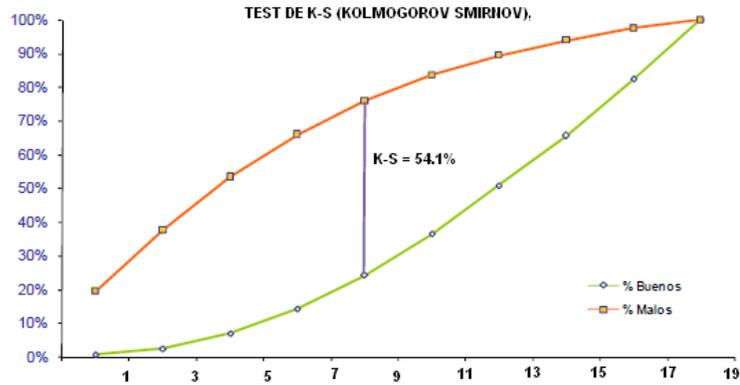
valores	% buenos	% malos	% buenos acumulado	% malos acumulado	diferencia K-S	% K-S
P(1/20)	0	0.123396	0	0.123396	0.123396	12.3
P(2/20)	0.014684	0.103653	0.014684	0.227048	0.212364	21.2
P(3/20)	0.019090	0.097730	0.033774	0.324778	0.291004	29.1
P(4/20)	0.023495	0.091807	0.057269	0.416584	0.359316	35.9
P(5/20)	0.027900	0.085884	0.085169	0.502468	0.417299	41.7
P(6/20)	0.032305	0.079961	0.117474	0.582428	0.464954	46.5
P(7/20)	0.036711	0.074038	0.154185	0.656466	0.502281	50.2
P(8/20)	0.041116	0.068115	0.195301	0.724580	0.529279	52.9
P(9/20)	0.045521	0.062192	0.240822	0.786772	0.540950	54.1
P(10/20)	0.056535	0.047384	0.297357	0.834156	0.536799	53.7
P(11/20)	0.060940	0.047384	0.358297	0.881540	0.523243	52.3
P(12/20)	0.065345	0.041461	0.423642	0.923001	0.499359	49.9
P(13/20)	0.069750	0.035538	0.493392	0.958539	0.465147	46.5
P(14/20)	0.074156	0.029615	0.567548	0.988154	0.420606	42.1
P(15/20)	0.078561	0.023692	0.646109	1.001846	0.355737	35.6
P(16/20)	0.082966	0.017769	0.729075	1.001846	0.272771	27.3
P(17/20)	0.087372	0.011846	0.816446	1.001846	0.185400	18.5
P(18/20)	0.091777	0.005923	0.908223	1.001846	0.093623	9.4
P(19/20)	0.091777	0.005923	1.000000	1.001846	0.001846	0.2
P(20/20)	0.091777	0.005923	1.000000	1.001846	0.001846	0.2

Fuente: Base de Datos proporcionado por la Cooperativa Maquita Cushunchic.- (ANEXO 4)⁶⁷

TABLA 3.4

Si la distancia K-S es comparativamente mayor al valor crítico se demuestra que en cada rango de funciones acumuladas existen comportamientos diferentes, por lo tanto existiría evidencia estadísticamente significativa para indicar que la discriminación del modelo es considerable.

⁶⁷ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.



El valor crítico D al 95% de confianza constituye el 0.237⁶⁸ (ANEXO 4). Siendo así el valor Crítico-D \leq que K-S, por lo tanto, no se rechaza Ho y se rechaza Ha, es decir las frecuencias observadas y las teóricas calculadas no difieren significativamente. Por lo tanto, las observaciones tienen una distribución normal.

TEST DE HOSMER-LEMESHOW

El valor del test Hosmer-Lemeshow en regresión logística, evalúa la bondad del ajuste de un modelo de regresión logística, si el ajuste es bueno, un valor alto de la p predicha se asociará con el resultado 1 de la variable binomial. Se trata de calcular para cada observación del conjunto de datos las probabilidades de la variable dependiente que predice el modelo, agruparlas y calcular, a partir de ellas, las frecuencias esperadas y compararlas con las observadas mediante la prueba X^2 .⁶⁹

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	4.936	8	.764

⁶⁸ Decisión: En virtud de lo anterior, el estadístico de Kolmogorov-Smirnov obtenido es menor que el crítico y su probabilidad mayor que 0.05, por lo tanto, se acepta Ho y se rechaza Ha. Interpretación: Las frecuencias observadas y las teóricas calculadas no difieren significativamente. Por lo tanto, las observaciones tienen una distribución normal, dirección: <http://members.fortunecity.com/bucker4/estadistica/pruebaks1m.htm>. El valor se lo calcula de la siguiente manera (ANEXO 4)

$$D_{\alpha} = \frac{c_{\alpha}}{k(n)}$$

⁶⁹ Test Hosmer-Lemeshow: Es otra prueba para evaluar la bondad del ajuste de un modelo de regresión logística, aunque su uso está más discutido que la anterior. La idea es si el ajuste es bueno, un valor alto de la p predicha se asociará (con un frecuencia parecida a la p) con el resultado 1 de la variable binomial. Se trata de calcular para cada observación del conjunto de datos las probabilidades de la variable dependiente que predice el modelo, agruparlas y calcular, a partir de ellas, las frecuencias esperadas y compararlas con las observadas mediante la prueba X^2 , dirección: http://www.hrc.es/bioest/Reglog_5.html

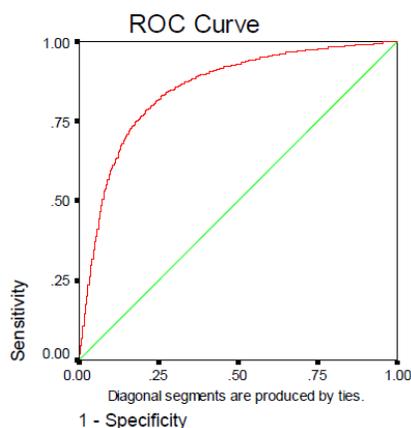
El estadístico es obtenido por aplicación de la prueba Ji-cuadrado en tablas de contingencia, donde la tabla de contingencia es construida por el cruce de la clasificación dicotómica de la variable dependiente con una variable de agrupación. Si los valores de la variable independiente para la observación j y j' pronosticado son los mismos entonces se dice que j y j' pronosticado pertenecen al mismo bloque.

Tabla de contingencias para la prueba de Hosmer y Lemeshow

		INCUMPLIMIENTO = .00		INCUMPLIMIENTO = 1.00		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	894	886.281	36	43.719	930
	2	855	855.105	75	74.895	930
	3	839	839.118	91	90.882	930
	4	830	824.127	100	105.873	930
	5	818	809.136	112	120.864	930
	6	801	793.037	129	136.963	930
	7	769	775.526	161	154.474	930
	8	753	753.757	177	176.243	930
	9	714	725.181	216	204.819	930
	10	650	661.731	275	263.269	925

TEST DE RAZÓN DE AJUSTE (AR-ROC-PRUEBA DE PODER)

El valor del test de razón de ajuste (*AR-ROC-prueba de poder*) en regresión logística, evalúa un aspecto de la validez del modelo (1) la calibración (*grado en que la probabilidad predicha coincide con la observada*), (2) el otro aspecto es la discriminación (*grado en que el modelo distingue entre casos en los que ocurre el evento y los que no*). Como medida de la discriminación se usa el área bajo la curva.⁷⁰



En el modelo desarrollado, la clasificación agrupa al 71.9% del total. La razón de ajuste medido como el área entre los modelos: estocástico y de perfecta previsión ocupada por el modelo

⁷⁰ Test area bajo la curva ROC: La prueba de Hosmer-Lemeshow evalúa un aspecto de la validez del modelo: la calibración (grado en que la probabilidad predicha coincide con la observada). El otro aspecto es la discriminación (grado en que el modelo distingue entre individuos en los que ocurre el evento y los que no). Como medida de la discriminación se usa el área bajo la curva ROC construida para la probabilidad predicha por el modelo, que representa, para todos los pares posibles de individuos formados por un individuo en el que ocurrió el evento y otro en el que no, la proporción de los que el modelo predice una mayor probabilidad para el que tuvo el evento, dirección: http://www.hrc.es/bioest/Reglog_5.html

estimado, es adecuada para esta especificación. El área bajo las curvas es el mejor indicador global de la precisión de una prueba diagnóstica y expresa el desempeño de esta mediante un número. Un área del 50% indica que no hay diferencia en la distribución de los valores de la prueba entre los dos grupos, mientras que una área de 100% es una discriminación perfecta, para nuestro caso, significa que un individuo seleccionado aleatoriamente del grupo de malos clientes tiene un valor de la prueba mayor que uno seleccionado del grupo de buenos en el 71.9% de las veces.

Área bajo la curva

VARIABLES RESULTADO DE CONTRASTE: Predicted probability

Área	Error típ. ^a	Sig. asintótica ^b	Intervalo de confianza asintótico al 95%	
			Límite inferior	Límite superior
,719	,052	,000	,618	,820

La variable (o variables) de resultado de contraste: Predicted probability tiene al menos un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Los estadísticos pueden estar sesgados.

a. Bajo el supuesto no paramétrico

b. Hipótesis nula: área verdadera = 0,5

3.6.3 BACKTESTING DEL MODELO⁷¹

Con una muestra de la data histórica que se considero inicialmente en el desarrollo del modelo credit score de la cartera de consumo se realizo un análisis retrospectivo (*análisis backtesting*⁷²). La muestra consta de 114 registros que se tomo de forma aleatoria del conjunto de registros correspondientes al año 2010. Con esta muestra se realizo la ejecución del modelo logístico, comprobando que este “*nuevo modelo*” mantiene la misma estructura de variables encontradas en el modelo original, la *significancia* del modelo original vs. el modelo backtesting muestran cumplimiento de la hipótesis planteada, además los *coeficientes* asociados a las variables independientes se asemejan al modelo original en los signos (+/-), se muestran en el (ANEXO 4).

⁷¹ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.

⁷² Backtesting.- Es el proceso de evaluación de una estrategia, teoría o modelo aplicándolo a los datos históricos. Un elemento clave de backtesting que lo diferencia de otras formas de pruebas históricas es que backtesting calcula cómo una estrategia habría realizado si hubiera sido realmente aplicadas en el pasado. Backtesting es un enfoque común y aceptado metodológicamente a la investigación, sin embargo, una alta correlación entre el éxito y los resultados históricos no pueden probar una teoría correcta, ya que los resultados pasados no necesariamente indica resultados futuros. dirección: <http://en.wikipedia.org/wiki/Backtesting>

modelo original vs. el modelo backtesting

No	VARIABLE	DESCRIPCION	modelo score original		modelo backtesting	
			COEFICIENTE	SIGNIFICANCIA	COEFICIENTE	SIGNIFICANCIA
1	CAL_CR_COOP_REC3(2)	calificación de riesgo (B o C o D o E) de crédito Coop. al momento de la concesión.	0.757756	0.000020838446	0.461697	0.015525505685
2	CAL_CR_REC3(2)	calificación de riesgo (B o C) de central de riesgos al momento de la concesión.	0.443445	0.072936034109	0.398017	0.046369288776
3	CUOTA	valor de la cuota al solicitar el crédito	-0.004122	0.005320307179	-0.000806	0.063902359608
4	SALDO_OPER	saldo operativo al momento de la concesión	-0.000218	0.000000000000	-0.000260	0.00009020274
5	CAPACIDAD_PAGO	%capacidad de pago (Ingresos-gastos/cuota)	0.010969	0.426692790862	0.023907	0.338293435296
6	EDAD	edad del cliente	-0.021469	0.000000000000	-0.013448	0.008785608945
7	MONTO	monto del crédito	0.000292	0.000000711525	0.000241	0.003774800744
8	RECIPROCIDAD	reciprocidad del cliente con la Coop. al momento de la concesión	-0.002177	0.000000073817	-0.005481	0.000179308256
9	INGRESOS2	ingresos del cliente	0.000374	0.004141810563	0.000110	0.066649378379
10	MIEMBROS_FAM	miembros de la familia del cliente	-0.132056	0.000000039748	-0.117619	0.016062565355
11	C	constante del modelo	-0.981981	0.000000000031	-1.874327	0.000000000001

Fuente: Base de Datos proporcionado por la Cooperativa Maquita Cushunchic.- (ANEXO 4)

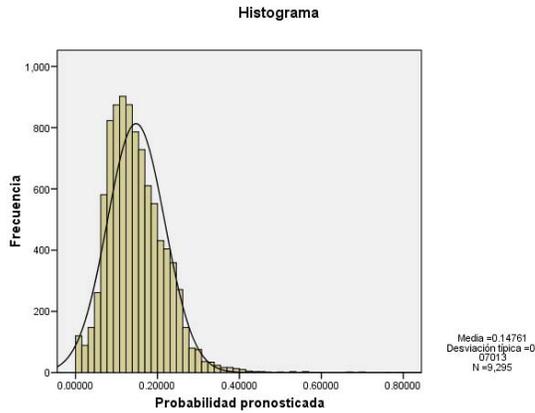
TABLA 3.5

3.6.4 CATEGORIZACION Y PUNTOS DE CORTE⁷³

Una vez encontrada la probabilidad logística con el modelo original se clasificó en 3 categorías: empezando por la categoría A (*aquellos valores cercanos a cero, considerados como "buen cliente"*) y a la categoría C (*aquellos valores cercanos a uno, considerados como "mal cliente"*). Se usa tres categorías porque en la data de la Cooperativa tenemos definidas tres variables de riesgos cuando el Oficial realiza la aprobación y concesión al potencial sujeto de crédito.

Luego se definieron puntos de corte tomando como límite al máximo y mínimo de las probabilidades del modelo logístico, luego se utilizó el método *prueba-error* hasta hallar una correlación "*adecuada*" entre el nuevo rango y el rango incumplimiento.

⁷³ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.



PUNTOS DE CORTE	
SEGMENTACION	PUNTOS DE CORTE
A	0 thru 0.3908
B	0.3909 thru 0.7250
C	0.7251 thru 1

El condicionante que se aplicó fue *que las categorías no deben presentar concentración*. La aceptación de los rangos de los puntos de corte dependerá del apetito al riesgo de la Cooperativa⁷⁴ y de las políticas existentes al momento de la asignación del crédito a un potencial sujeto de crédito.

Tabla de contingencia RANGO_SCORE * INCUMPLIMIENTO

		INCUMPLIMIENTO		Total
		1.00	.00	
RANGO_SCORE A	Recuento	491	4054	4545
	% de RANGO_SCORE	10.8%	89.2%	100.0%
	% de INCUMPLIMIENTO	12.9%	70.0%	48.8%
B	Recuento	1959	1629	3588
	% de RANGO_SCORE	54.6%	45.4%	100.0%
	% de INCUMPLIMIENTO	56.9%	27.8%	38.5%
C	Recuento	1042	120	1162
	% de RANGO_SCORE	89.7%	10.3%	100.0%
	% de INCUMPLIMIENTO	30.2%	2.2%	12.6%
Total	Recuento	3492	5803	9295
	% de RANGO_SCORE	36.9%	63.0%	100.0%
	% de INCUMPLIMIENTO	100.0%	100.0%	100.0%

⁷⁴ El apetito al riesgo se refiere por ejemplo a que si un sujeto de crédito cae en la categoría B se lo podría asignar el crédito pero con diferentes condiciones que a un sujeto de crédito categorizado con categoría A.

CAPITULO IV

CONCLUSIONES

Para el diseño de un modelo credit score se requiere de una serie de pasos o fases las cuales implican: la preparación de la base de datos, la clasificación de los tipos de variables tanto cualitativas como cuantitativas, la selección de la metodología, la selección de la muestra, la depuración de datos aberrantes y valores perdidos, las pruebas de control sobre el modelo propuesto, se concluye que sin lugar a dudas el paso o fase más importante es el análisis y preparación estadística de las variables que se incorporaran al modelo.

El modelo credit score de la cartera de consumo analizado y propuesto evidencia condiciones necesarias para presentar la validez metodológica en la obtención de la probabilidad de incumplimiento de pago de un sujeto de crédito, sin embargo la parte metodológica no es suficiente al momento de obtener un modelo de éste tipo, porque como observamos en nuestro modelo que a pesar de que el test de significancia y el nivel de predictividad y discriminación indique un aceptable nivel de confiabilidad existen variables independientes: *cuota*, *saldo_oper*, *reciprocidad*, *capacidad_pago*, *ingresos*, *miembros_fam* que presentan inconsistencias en SIGNOS, las mismas que se están fuera de la lógica económica. Al considerarse estas variables fundamentales ocasionan distorsión en el modelo, es decir, el modelo entregaría resultados erróneos y permitirá tomar malas decisiones al momento de conceder un crédito de consumo, entonces, aunque estadísticamente el modelo propuesto sea el que produce el "mejor" R^2 , económicamente el modelo no tiene validez, por lo tanto no se lo puede considerar como una herramienta que apoye a tomar decisiones.

Con un modelo de score de aprobación adecuado, mas el know-how de los expertos de crédito de la Cooperativa debería apoyar a la estimación de pérdidas esperadas de cada sujeto de crédito, esperando cuantificar la probabilidad de default, la exposición y la severidad, logrando que la información discrimine aquellas operaciones que no agreguen valor a la Cooperativa y enmarcado en un concepto financiero conservador al momento de conceder un crédito.

RECOMENDACIONES

El know-how de la Cooperativa no debería ser la única herramienta en el otorgamiento de un crédito, se recomienda diseñar un modelo credit score propio para sus productos que asignen crédito. Basilea II recomienda que cada Institución realice sus propios modelos, su propio análisis y preparación estadística de variables e incorporar la mayor cantidad de información posible en el modelo, evitando criterios subjetivos y percepciones.

La Cooperativa debería hallar mecanismos que garanticen la integridad de los datos al momento de ingresar la data en los sistemas informáticos por parte de todos los actores que interactúan en la Institución y la base de datos.

El algoritmo del modelo propuesto debería volver a ejecutarse con una *nueva muestra adquirida de por lo menos 3 meses*, ya que el modelo depende también de variables externas como: el comportamiento del cliente en el Sistema Financiero, la calificación en la Central de Riesgos que podrían cambiar en el tiempo y podría influir drásticamente en el comportamiento de pago del cliente, y luego de realizar el backtesting se debería implementar un proceso de control periódico.

BIBLIOGRAFIA

PHILIPPE Jorion, Value at Risk The New Benchmark for Managing Financial Risk (*Hardcover*), McGraw Hill, Tercera Edición, año 2007

GUITIAN Manuel, VARELA Felix, Sistemas Financieros ante la Globalización (*C-07730*), Ediciones Piramide, año 2000, pág. 219 – 223.

GUJARATI Damodar, Introducción a la Econometría, Cuarta edición, Ediciones McGraw Hill, año 2004.

BERNDT Ernst R, The Practice Econometrics Classic and Contemporary, Addison-Wesley Publishing Company, año 1991.

PINDYCK Robert y RUBINFELD Daniel, Econometría, Modelos y Pronósticos. Cuarta edición, Ediciones McGraw Hill, año 2001.

NOBOA Paul, Diapositivas de Riesgo de Crédito, Universidad Andina Simón Bolívar, Maestría en Finanzas y Riesgos, año 2009.

CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008.

TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010.

DIRECCIONES DE INTERNET

Una aproximación al riesgo de crédito en las Entidades Financieras: cómo analizar la morosidad.

http://www.navactiva.com/web/es/descargas/pdf/acyf/riesgo_credito.pdf

Breve reseña histórica de la Cooperativa de ahorro y crédito Maquita Cushunchic, se presenta un documento pdf.

<http://www.uasb.edu.ec/UserFiles/363/File/pdfs/SUBSITIO%20FIDA/ENFOQUES%20Y%20EXPERI>

[ENCIAS%20DEL%20DESARROLLO%20RURAL/Cooperativa%20Maquita%20Cushunchic.pdf](http://www.uasb.edu.ec/UserFiles/363/File/pdfs/SUBSITIO%20FIDA/ENFOQUES%20Y%20EXPERIENCIAS%20DEL%20DESARROLLO%20RURAL/Cooperativa%20Maquita%20Cushunchic.pdf)

http://www.ratinginitiative.org/uploads/tx_dbreports/COAC_Maquita_Cushunchic_informe_final_de_calificacion_Julio_2005.pdf

Acuerdos de Basilea, los 3 pilares de Basilea II y riesgo de crédito

http://es.wikipedia.org/wiki/Riesgo_de_cr%C3%A9dito

http://es.wikipedia.org/wiki/Basilea_II

ANEXOS

**ANEXO 1:
ANÁLISIS EXPLORATORIO DE VARIABLES CUANTITATIVAS Y CUALITATIVAS DEL
CONJUNTO DE VARIABLES ENTREGADAS POR LA INSTITUCIÓN**

Sobre la base de datos de 9,295 registros y 61 variables realizaremos el análisis exploratorio de variables cuantitativas y cualitativas del conjunto de variables entregadas por la Institución, luego se analizará la variable "*tipo pagador*" que representa el "*incumplimiento*".

Variables Cuantitativas: Comprende las variables que se expresan mediante cantidades numéricas. Las variables cuantitativas además pueden ser: variables continuas y variables discretas¹.

Variables de la Base de Datos de cartera de consumo

No	VARIABLE	DESCRIPCION	TIPO
1	ALQ_PROP	valor del alquiler, se valida si es vivienda alquilada o propia	numérico
2	ANIO	año del crédito	numérico
3	BONIF_SUB	alguna bonificación del cliente	numérico
4	CART_CASTIG	cartera castigada	numérico
5	CART_JUDICIAL	cartera judicial	numérico
6	CART_NO_DEVEN	cartera no devenga	numérico
7	CART_POR_VENCER	cartera por vencer	numérico
8	CART_VENCIDA	cartera vencida	numérico
9	COMISIONES	alguna comisión extra del cliente	numérico
10	CTAS_C_RH_2_ULT	cuentas cerradas en los 2 últimos años	numérico
11	CTAS_C_RH_5_PEN	cuentas cerradas en los 5 últimos años	numérico
12	CTAS_C_RH_ULT	cuentas cerradas último año	numérico
13	CTAS_C_VIG	cuentas cerradas vigentes (SIB)	numérico
14	CUOTA	valor de la cuota	numérico
15	CUOTAS_PAG	cuotas pagadas	numérico
16	DEUDAS_DIREC	deudas directas	numérico
17	DEUDAS_INDIREC	deudas indirectas	numérico
18	DIAS_MORA	días de mora	numérico
19	DIAS_MORA_MAX	días de mora máximo	numérico
20	DIAS_MORA_PROM	días de mora promedio	numérico
21	EDAD	edad	numérico
22	INGRESOS	ingresos	numérico
23	MIEMBROS_FAM	miembros de la familia	numérico
24	MONTO	monto del crédito	numérico
25	N_TRAMITE	numero del tramite	numérico
26	N_VMORA_M30	numero de veces que cayo en mora mayor que 30 días	numérico
27	N_VMORA_M45	numero de veces que cayo en mora mayor que 45 días	numérico
28	N_VMORA_M5	numero de veces que cayo en mora mayor que 5 días	numérico
29	N_VMORA_M60	numero de veces que cayo en mora mayor que 60 días	numérico
30	NUM_CUOTAS	numero de cuotas	numérico

¹ Definiciones básicas de variables cuantitativas, fecha acceso: Enero 2011, dirección: http://es.wikipedia.org/wiki/Variable_estad%C3%ADstica

31	PATRIMONIO	patrimonio	numérico
32	RECIPROCIDAD	reciprocidad con la Coop	numérico
33	RENTA_JUBILAR	renta jubilar	numérico
34	SALDO_OPER	saldo operativo	numérico
35	SUELDO_BASICO	sueldo básico	numérico
36	TASA	lata	numérico
37	UTILIDADES	utilidades	numérico

Fuente: Base de Datos proporcionado por la Cooperativa Maquita Cushunchic

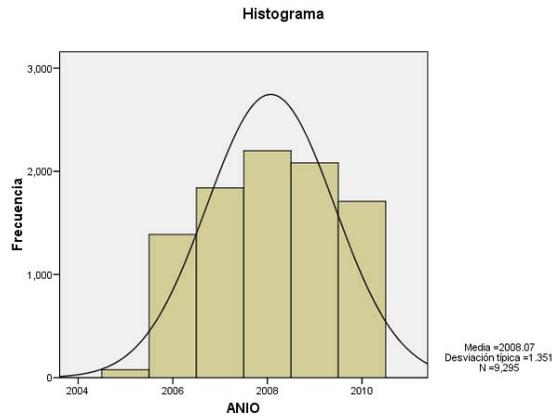
TABLA A1.1

A cada una de las variables cuantitativas se obtendrán las medidas de tendencia central tales como la media, mediana, mínimo, máximo, desviación estándar y los histogramas de frecuencia de cada una de las variables.

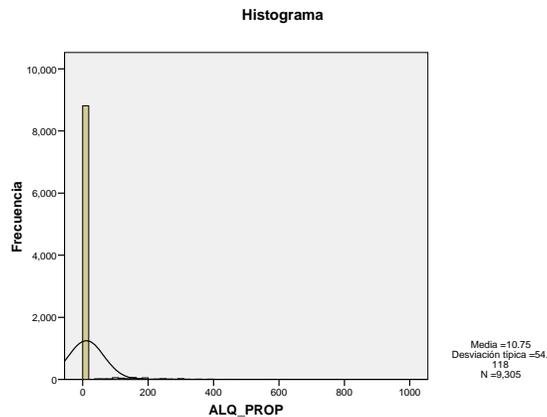
Estadísticos descriptivos

	N	Mínimo	Máximo	Suma	Media	Desv. típ.	Varianza
ANIO	9295	2005	2010	18665014	2008.07	1.351	1.824
ALQ_PROP	9295	0	900	99897	10.75	54.130	2930.031
BONIF_SUB	9295	0	2700	10742	1.16	41.849	1751.337
CART_CASTIG	9295	0	2062	126732	13.63	117.429	13789.617
CART_JUDICIAL	9295	0	0	0	.00	.000	.000
CART_NO_DEVEN	9295	0	7481	77241	8.31	166.022	27563.323
CART_POR_VENCER	9295	.00	20588.06	4736204	509.5432	1375.052	1890768
CART_VENCIDA	9295	0	1689	25590	2.75	43.367	1880.707
COMISIONES	9295	0	1319	115672	12.44	61.587	3792.899
N_VMORA_M5	9295	0	22	7228	.78	1.831	3.353
N_VMORA_M30	9295	0	6	698	.08	.380	.144
N_VMORA_M45	9295	0	6	454	.05	.293	.086
N_VMORA_M60	9295	0	13	1881	.20	1.071	1.146
CTAS_C_RH_2_ULT	9295	0	1	3	.00	.018	.000
CTAS_C_RH_5_PEN	9295	0	4	216	.02	.208	.043
CTAS_C_RH_ULT	9295	0	0	0	.00	.000	.000
CTAS_C_VIG	9295	0	4	210	.02	.178	.032
DEUDAS_DIRECT	9295	0	16487	7475673	804.27	2006.262	4025087
DEUDAS_INDIRECT	9295	0	130894	4484007	482.41	2198.117	4831719
EDAD	9295	19	78	359339	38.66	12.586	158.416
INGRESOS	9295	0	7850	4192230	451.02	340.355	115841.5
MIEMBROS_FAM	9295	1	9	23551	2.53	1.320	1.742
N_TRAMITE	9295	6890	35198	2E+008	21956.62	7926.039	6E+007
PATRIMONIO	9295	0	757000	96663564	10399.52	21093.940	4E+008
RECIPROCIDAD	9295	-274.23	18892.46	661839.76	71.2038	384.19858	147608.5
RENTA_JUBILAR	9295	0	1724	169846	18.27	81.470	6637.286
SALDO_OPER	9295	.00	20776.82	5155732	554.6780	1416.573	2006680
SUELDO_BASICO	9295	0	5000	3613387	388.75	315.325	99430.038
UTILIDADES	9295	0	7850	182686	19.65	122.331	14964.760
DIAS_MORA	9295	0	1566	169397	18.22	126.337	15960.928
DIAS_MORA_PROM	9295	.0	1359.0	60506.9	6.510	39.2828	1543.137
DIAS_MORA_MAX	9295	0	1566	242910	26.13	126.900	16103.713
CUOTA	9295	14.83	10658.58	1071688	115.2973	150.04850	22514.554
NUM_CUOTAS	9295	1	60	140007	15.06	5.900	34.814
CUOTAS_PAG	9295	0	37	122488	13.18	5.577	31.107
MONTO	9295	104.56	20911.75	2E+007	1796.3425	1974.034	3896811
TASA	9295	12.1	25.9	146995.8	15.815	3.0082	9.049
N válido (según lista)	9295						

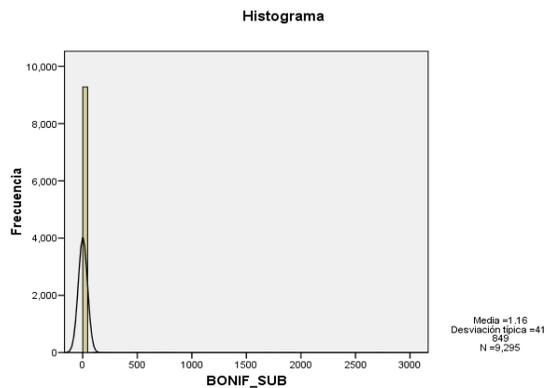
1. **ANIO:** Indica el año del crédito. Esta es una variable que no va aportar al modelo por lo que va a ser descartada.



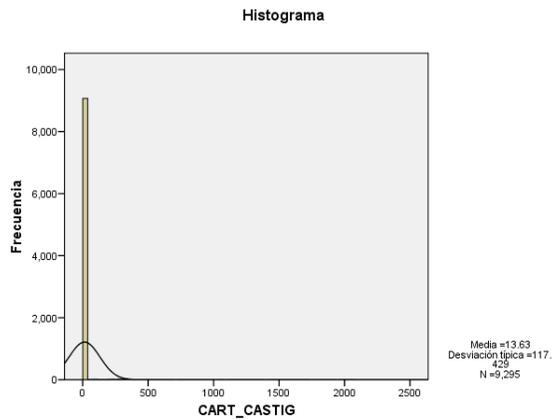
2. **ALQ_PROP:** Indica el valor que paga el sujeto de crédito cuando la vivienda es alquilada, el 94.97% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada.



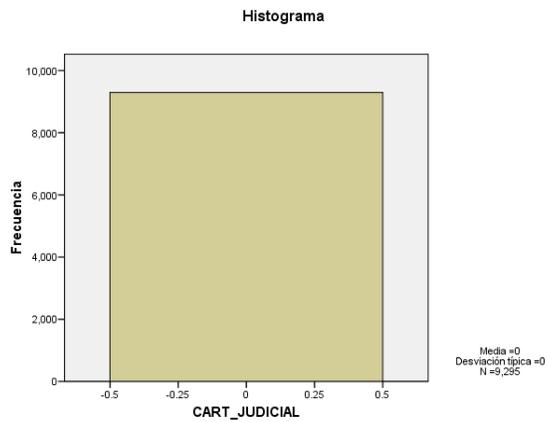
3. **BONIF_SUB:** Determina la bonificación del sujeto de crédito tiene, el 99.84% de los valores de esta variable es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada.



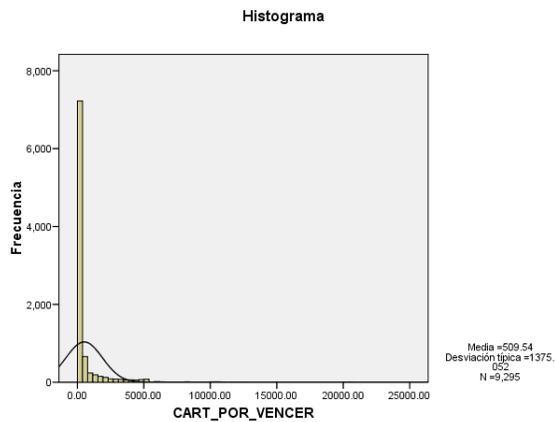
4. CART_CASTIG: Determina la cartera castigada del cliente.



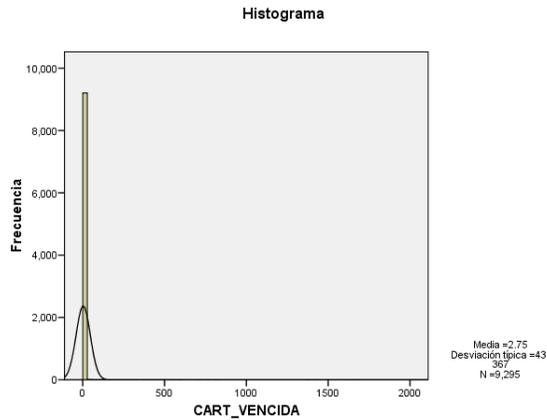
5. CART_JUDICIAL: Determina la cartera judicial del cliente, el 100% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada



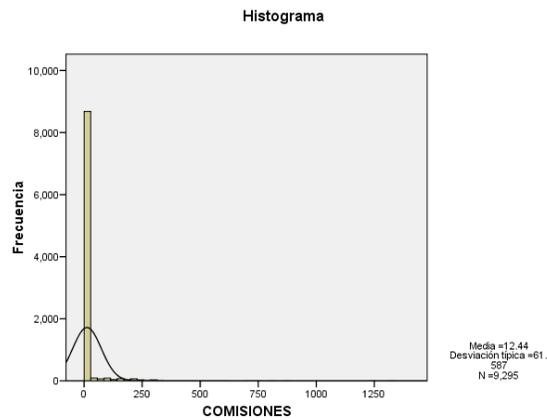
7. CART_POR_VENCER: Determina la cartera por vencer del cliente.



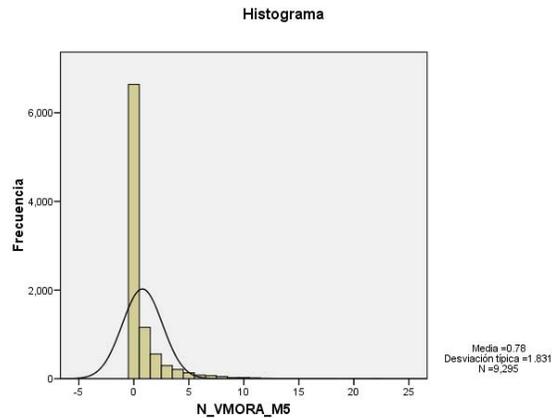
8. CART_VENCIDA: Determina la cartera vencida del cliente.



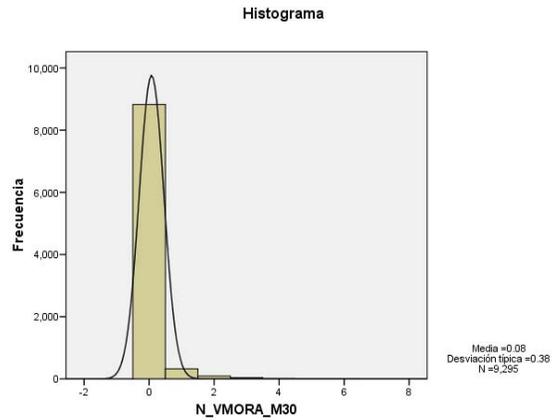
9. **COMISIONES:** Determina la comisión del cliente, el 93.17% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada.



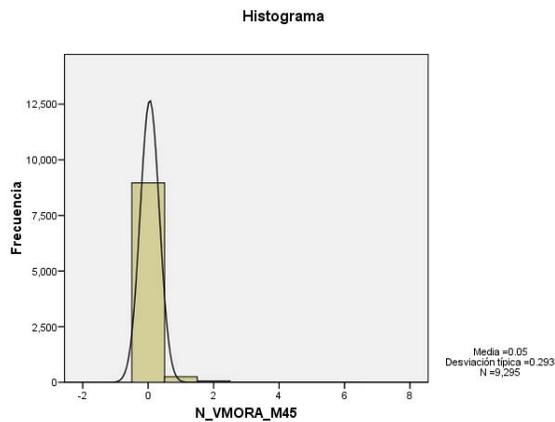
10. **N_VMORA_M5:** Indica el número de veces que cayó en mora mayor que 5 días.



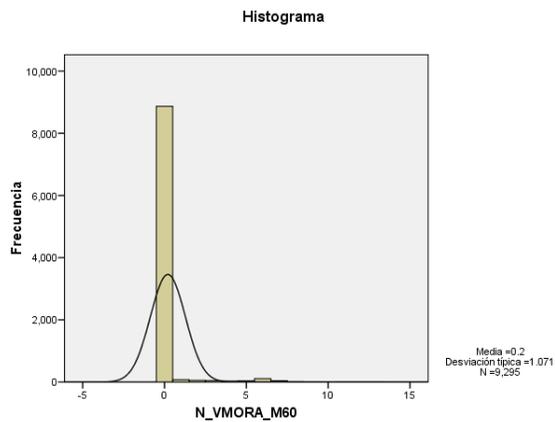
11. **N_VMORA_M30:** Indica el número de veces que cayó en mora mayor que 30 días, el 94.97% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada.



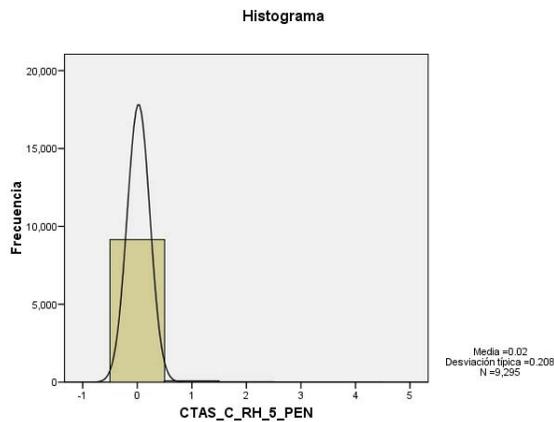
12. **N_VMORA_M45:** Indica el número de veces que cayó en mora mayor que 45 días, el 96.41% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada



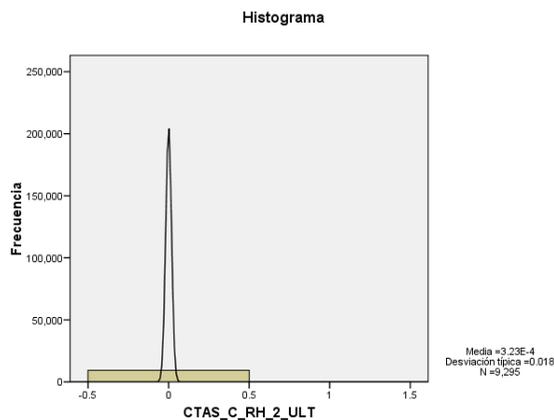
13. **N_VMORA_M60:** Indica el número de veces que cayó en mora mayor que 60 días, el 95.43% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada.



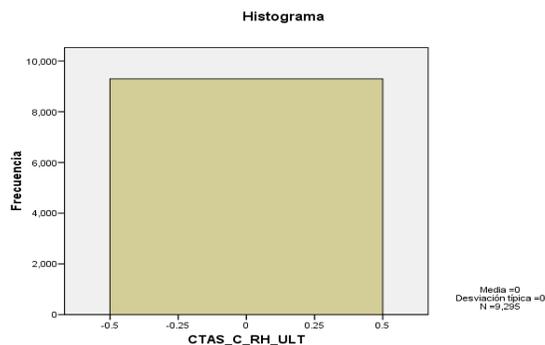
14. **CTAS_C_RH_5_PEN:** Determina las cuentas cerradas en los 5 últimos años, el 98.45% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada.



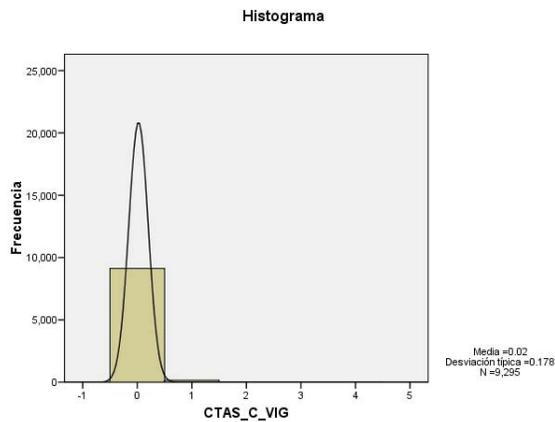
15. **CTAS_C_RH_2_ULT:** Indica cuentas cerradas en los 2 últimos años, el 100% el valor es uno, es decir existe una concentración en uno, es una variable que no va aportar al modelo y va a ser descartada.



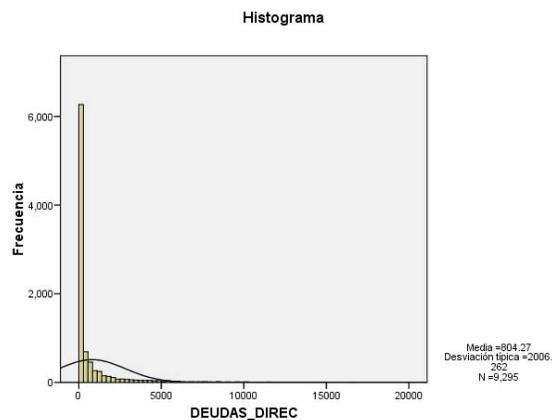
16. **CTAS_C_RH_ULT:** Indica las cuentas cerradas del último año, el 100% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada



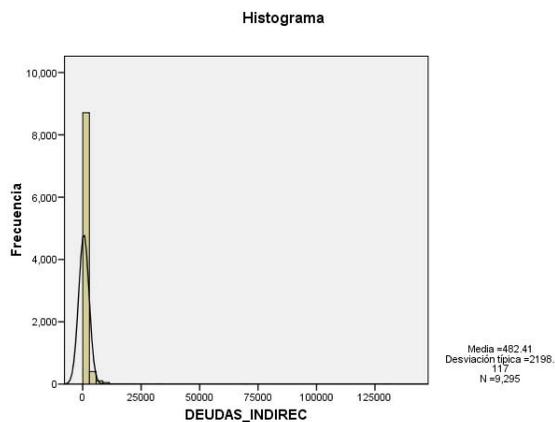
17. CTAS_C_VIG: Determina las cuentas cerradas vigentes (S/B), el 98.10% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada.



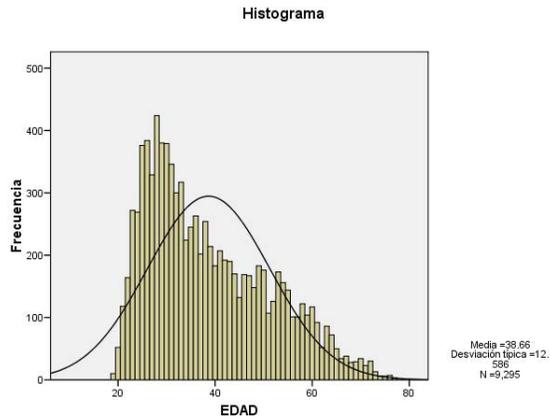
18. DEUDAS_DIREC: Determina las deudas directas del cliente.



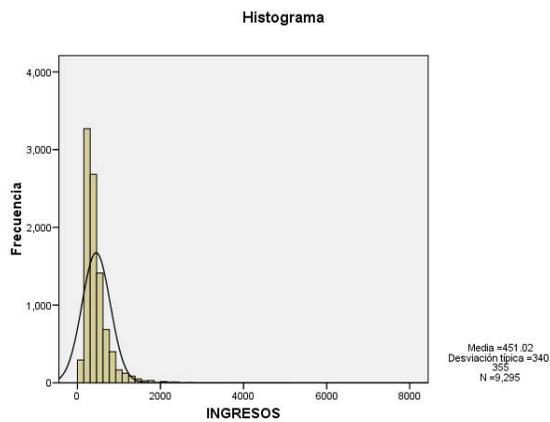
19. DEUDAS_INDIREC: Determina las deudas indirectas del cliente.



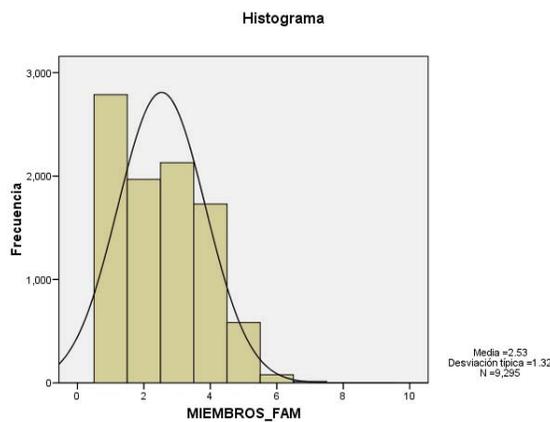
20. EDAD: Determina la edad del cliente.



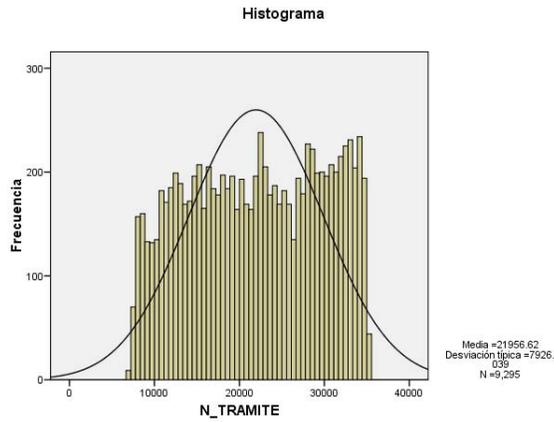
21. **INGRESOS**: Indica los ingresos del cliente.



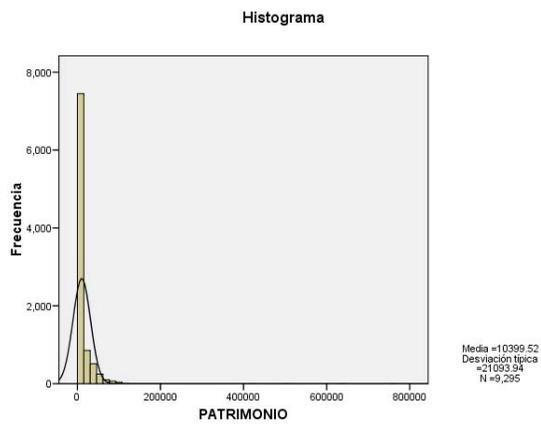
22. **MIEMBROS_FAM**: Determina los miembros de la familia del cliente.



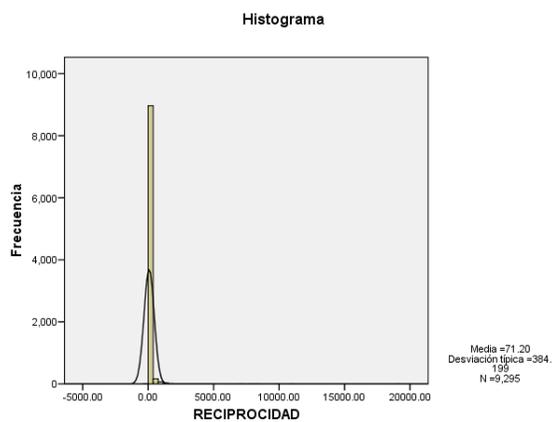
23. **N_TRAMITE**: Determina el número único de trámite, este es un secuencial de la base histórica, es una variable que no va aportar al modelo y va a ser descartada.



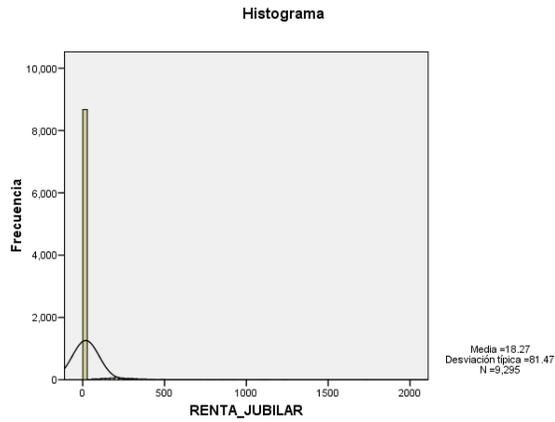
24. PATRIMONIO: Determina el patrimonio.



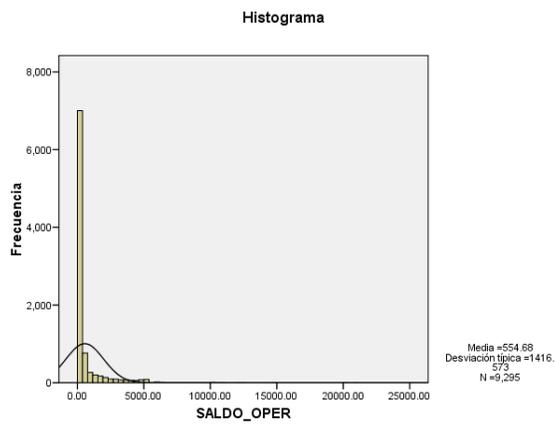
25. RECIPROCIDAD: Determina la reciprocidad del cliente.



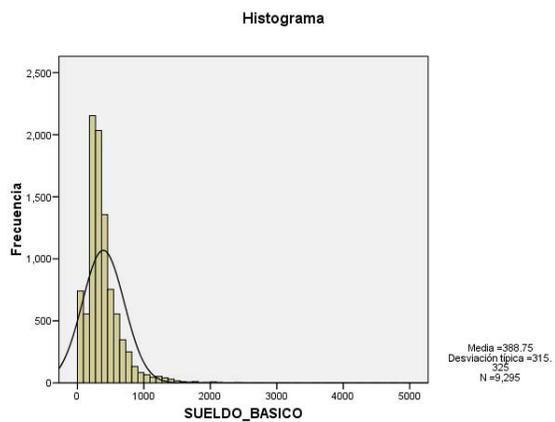
26. RENTA_JUBILAR: Determina la renta jubilar, el 93.92% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada.



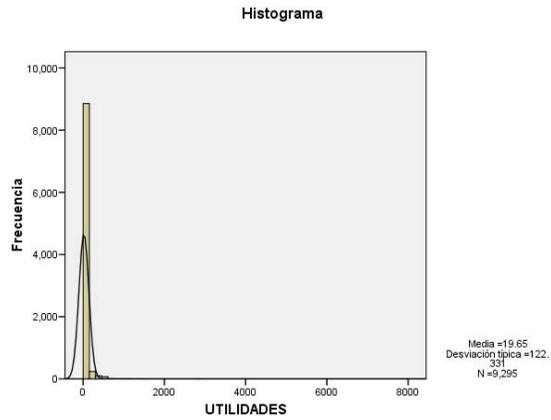
27. SALDO_OPER: Determina el saldo operativo.



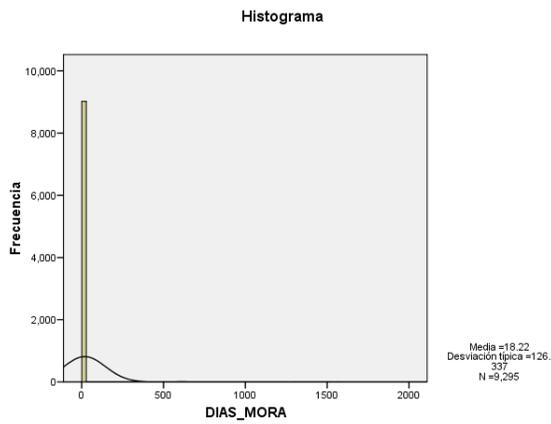
28. SUELDO_BASICO: Determina el sueldo básico.



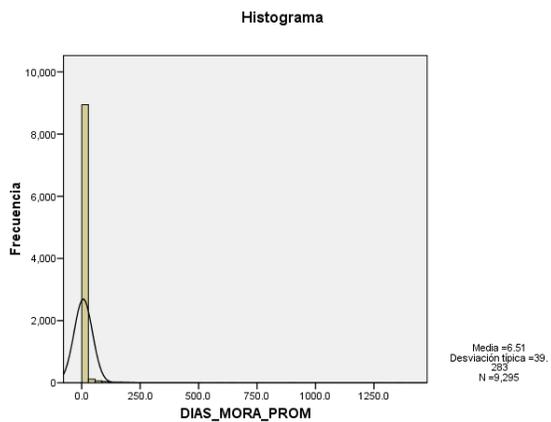
29. UTILIDADES: Determina las utilidades, el 93.71% el valor es cero, es decir existe una concentración en cero, es una variable que no va aportar al modelo y va a ser descartada.



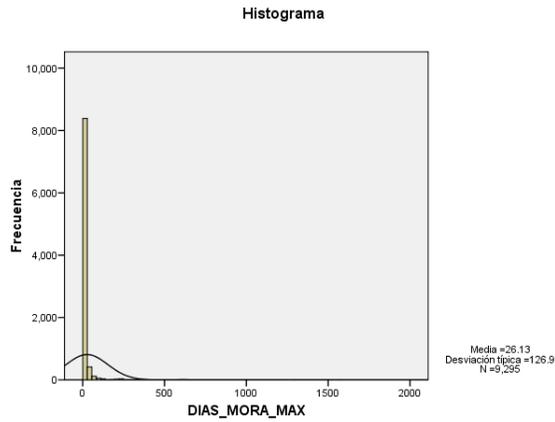
30. DIAS_MORA: Determina los días de mora de un crédito acreditado al cliente.



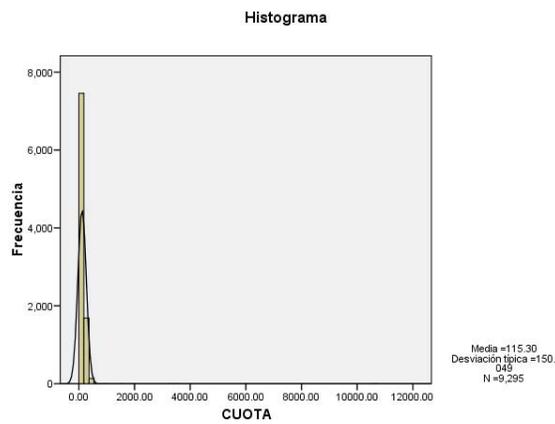
31. DIAS_MORA_PROM: Determina los días de mora promedio de un crédito acreditado al cliente.



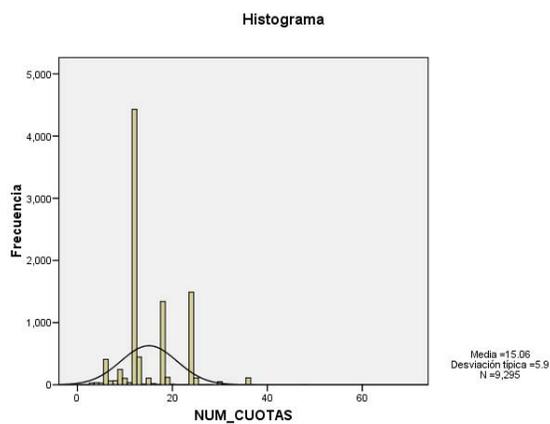
32. DIAS_MORA_MAX: Determina los días de mora máximo de un crédito acreditado al cliente.



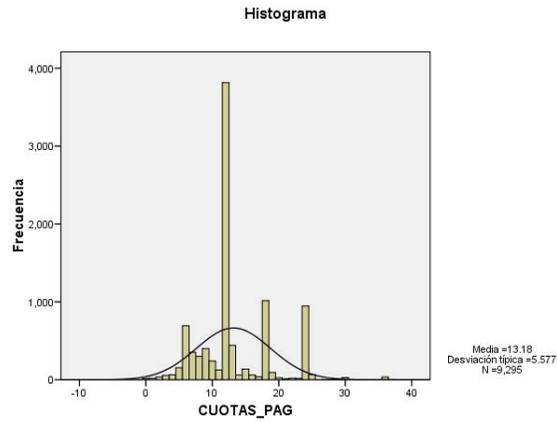
33. CUOTA: Determina la el valor de la cuota.



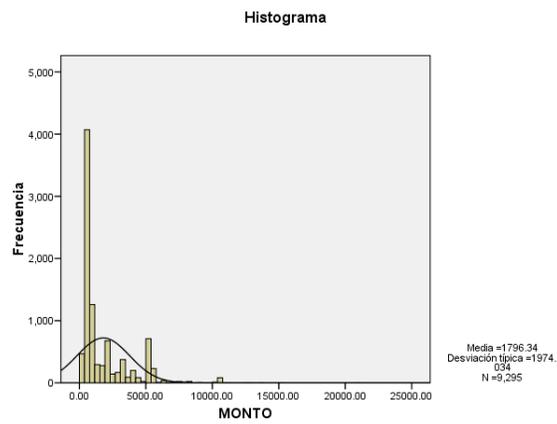
34. NUM_CUOTAS: Determina el número de cuotas del crédito.



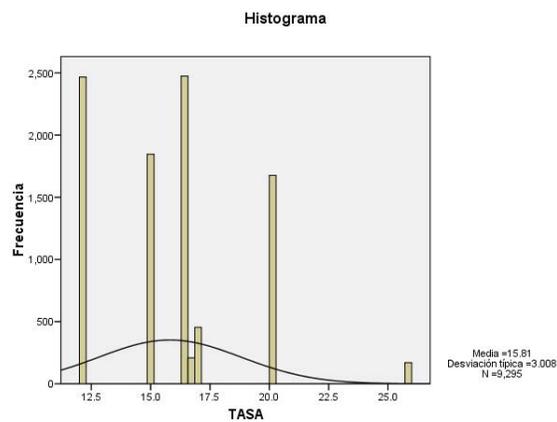
35. CUOTAS_PAG: Determina el número de cuotas que están pagadas.



36. **MONTO:** Determina el monto total del crédito.



37. **TASA:** Determina la tasa al cual fue asignado el crédito.



Variables Cualitativas: Cuando se hace referencia a este tipo de variables intuitivamente se relacionan con aquellas que brindan cierta "*cualidad*" del elemento a medir. Desde el punto de vista de la gestión de riesgo crediticio pueden asociarse a este grupo de variables como: genero, estado civil, nivel de educación, el comportamiento de los agentes económicos, oferta de trabajo, profesión de una persona, otros.²

Variables de la Base de Datos de Cartera de Consumo

No	VARIABLE	DESCRIPCION	TIPO
38	ACTIVIDAD	actividad económica	string
39	CAL_C_R	calificación de la central de riesgos	string
40	DESTINO_CRED	destino del crédito	string
41	E_A_RES	extranjero años de residencia	string
42	EST_CIVIL	estado civil	string
43	ESTAB_TRAB_ACT	estabilidad del trabajo actual	string
44	ESTAB_TRAB_ANT	estabilidad del trabajo anterior	string
45	EXP_CRED_COOP	calificación de riesgo de crédito Coop	string
46	GAR_EMP_DEU	garante que este empleado del deudor	string
47	GAR_M_NEG_DEU	garante mismo negocio del deudor	string
48	GENERO	genero	string
49	INGRESOS_ADIC	ingresos adicionales	string
50	N_PROT_INJUST	numero de protestos injustificados	string
51	N_PROT_JUST	numero de protestos justificados	string
52	NIVEL_EDUC	nivel de educación	string
53	POLIC_MIL	policia/militar	string
54	PROFESION	profesión	string
55	SECTOR	sector	string
56	SOLIC_PROP_NEG	solicitante propietario del negocio	string
57	TIEMPO_NEG	tiempo del negocio	string
58	TIPO_CREDITO	tipo de crédito	string
59	TIPO_VIVIENDA	tipo de vivienda	string
60	FECHA_CONS	fecha de concesión	fecha
61	FECHA_VCTO	fecha de vencimiento	fecha

Fuente: Base de Datos proporcionado por la Cooperativa Maquita Cushunchic

TABLA A1.2

² CARRANZA Freddy, Programa de maestría en Finanzas y gestión de riesgos, Análisis y preparación estadística de variables para el diseño de un modelo credit score de gestión de crédito, Universidad Andina Simón Bolívar, 2008.

1. **ACTIVIDAD:** Determina la actividad económica

Estadísticos

ACTIVIDAD

N	Válidos	9295
	Perdidos	0

Los tipos de la actividad económica son representados de alrededor de 166 tipos, en este sentido no es fácil de visualizar ni entender, por ello esta variable no mostramos medidas de tendencia. La cantidad de tipos hace que la variable no aporte al modelo y va a ser descartada.

2. **E_A_RES:** Determina que es extranjero con años de residencia en el país, el 69.9% existe concentración en el ítem "Mas de 5" y el 1% en el ítem "Menos de 5". Existe concentración en un solo ítem, entonces esta variable no va aportar al modelo y va a ser descartada.

E_A_RES

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Más de 5	6499	69.9	69.9	69.9
	Menos de 5	10	.1	.1	70.0
	no registra	2786	30.0	30.0	100.0
	Total	9295	100.0	100.0	

3. **CAL_C_R:** Determina la calificación de la central de riesgos

CAL_C_R

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	A	6065	65.3	65.3	65.3
	B	47	.5	.5	65.8
	C o menos	33	.4	.4	66.1
	C O MENOS	23	.2	.2	66.4
	no registra	3127	33.6	33.6	100.0
	Total	9295	100.0	100.0	

4. **DESTINO_CRED:** Determina el destino del crédito

Estadísticos

DESTINO_CRED

N	Válidos	9295
	Perdidos	0

Los tipos de destino crediticio son representados de alrededor de 104 tipos, en este sentido no es fácil de visualizar ni entender, por ello esta variable no mostramos medidas de tendencia. La cantidad de tipos hace que la variable que no aporte al modelo y va a ser descartada.

5. ESTAB_TRAB_ACT: Determina la estabilidad del trabajo actual
ESTAB_TRAB_ACT

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 o más	2	.0	.0	.0
	1 o más	7191	77.4	77.4	77.4
	Más de 12 meses	1115	12.0	12.0	89.4
	Menos de 1	231	2.5	2.5	91.9
	Menos de 12 meses	77	.8	.8	92.7
	no registra	679	7.3	7.3	100.0
	Total	9295	100.0	100.0	

6. ESTAB_TRAB_ANT: Determina la estabilidad del trabajo anterior, el 79.7% existe concentración en el ítem "24 meses o más" y el 1% en el ítem "menos de 24 meses". Existe concentración en un solo ítem, entonces esta variable no va aportar al modelo y va a ser descartada.

ESTAB_TRAB_ANT

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	24 meses o más	7411	79.7	79.7	79.7
	menos de 24 meses	13	.1	.1	79.9
	no registra	1871	20.1	20.1	100.0
	Total	9295	100.0	100.0	

7. EST_CIVIL: Determina el estado civil.

EST_CIVIL

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	CASADO	4478	48.2	48.2	48.2
	DIVORCIADO	501	5.4	5.4	53.6
	SOLTERO	3851	41.4	41.4	95.0
	UNION LIBRE	178	1.9	1.9	96.9
	VIUDO	287	3.1	3.1	100.0
	Total	9295	100.0	100.0	

8. EXP_CRED_COOP: Determina la calificación experiencia crédito de la cooperativa.

EXP_CRED_COOP

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	A	4433	47.6	47.6	47.6
	B	172	1.8	1.8	49.5
	C	23	.2	.2	49.7
	D	1	.0	.0	49.7
	E	11	.1	.1	49.9
	no registra	4665	50.1	50.1	100.0
	Total	9305	100.0	100.0	

9. **GAR_M_NEG_DEU:** Determina al garante mismo negocio del deudor, el 96.4% el valor del ítem es "*no registra*", es decir existe una concentración en este ítem, es una variable que no va aportar al modelo y va a ser descartada.

GAR_M_NEG_DEU

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	No	320	3.4	3.4	3.4
	no registra	8968	96.5	96.5	99.9
	Si	7	.1	.1	100.0
	Total	9295	100.0	100.0	

10. **GAR_EMP_DEU:** Determina al garante que este empleado del deudor, el 96.5% el valor del ítem es "*no registra*", es decir existe una concentración en este ítem, es una variable que no va aportar al modelo y va a ser descartada.

GAR_EMP_DEU

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	No	327	3.5	3.5	3.5
	no registra	8968	96.5	96.5	100.0
	Total	9295	100.0	100.0	

11. **GENERO:** Determina la genero

GENERO

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	FEMENINO	4877	52.4	52.4	52.4
	MASCULINO	4428	47.6	47.6	100.0
	Total	9305	100.0	100.0	

12. **INGRESOS_ADIC:** Determina los ingresos adicionales, el 99.5% el valor del ítem es "*no registra*", es decir existe una concentración en este ítem, es una variable que no va aportar al modelo y va a ser descartada.

INGRESOS_ADIC

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	ARRIENDOS U OTROS	2	.0	.0	.0
	FIJO Y ARRIENDOS	3	.0	.0	.1
	ING. FIJO	6	.1	.1	.1
	NINGUNO	32	.3	.3	.5
	no registra	9252	99.5	99.5	100.0
	Total	9295	100.0	100.0	

13. NIVEL_EDUC: Determina el nivel de educación

NIVEL_EDUC

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	BACHILLER	33	.4	.4	.4
	ESTUDIANTE SUPERIOR	710	7.6	7.6	8.0
	NINGUNA	377	4.1	4.1	12.0
	no registra	500	5.4	5.4	17.4
	PRIMARIA	2520	27.1	27.1	44.5
	SECUNDARIA	5165	55.5	55.5	100.0
	Total	9305	100.0	100.0	

14. N_PROT_INJUST: Determina el número de protestos injustificados, el 87.1% existe concentración en el ítem "0" y el 12.9% en el ítem "no registra". Existe concentración en un solo ítem, entonces esta variable no va aportar al modelo y va a ser descartada.

N_PROT_INJUST

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0	8106	87.1	87.1	87.1
	1 o más	4	.0	.0	87.2
	no registra	1195	12.9	12.9	100.0
	Total	9305	100.0	100.0	

15. N_PROT_JUST: Determina el número de protestos justificados, el 87.1% existe concentración en el ítem "0" y el 12.9% en el ítem "no registra". Existe concentración en un solo ítem, entonces esta variable no va aportar al modelo y va a ser descartada.

N_PROT_JUST

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0	8097	87.1	87.1	87.1
	no registra	1198	12.9	12.9	100.0
	Total	9295	100.0	100.0	

16. POLIC_MIL: Determina la policía/militar, el 77.5% existe concentración en el ítem "Mayor a cabo" y el 22.5% en el ítem "no registra". Existe concentración en un solo ítem, entonces esta variable no va aportar al modelo y va a ser descartada.

POLIC_MIL

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Mayor a cabo	4783	77.5	77.5	77.5
	no registra	1385	22.5	22.5	100.0
	Total	6168	100.0	100.0	

17. **PROFESION:** Determina la profesión del sujeto de crédito, el 94.6% el valor del ítem es "no registra", es decir existe una concentración en este ítem, es una variable que no va aportar al modelo y va a ser descartada.

PROFESION

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	ADMINIST. Y CONTABILIDAD	12	.1	.1	.1
	ARQUITECTO	6	.1	.1	.2
	BACHILLER	1	.0	.0	.2
	DR. LEYES	9	.1	.1	.3
	DR.CONTABILIDAD	9	.1	.1	.4
	DR.MEDICINA GENERAL	34	.4	.4	.8
	DR.ODONTOLOGIA	6	.1	.1	.8
	DR.PEDAGOGIA	7	.1	.1	.9
	DR.QUIMICA/FARMACIA	1	.0	.0	.9
	DR.VETERINARIA	1	.0	.0	.9
	ESTUDIANTE	9	.1	.1	1.0
	ING. ADM. DE EMPRESAS	22	.2	.2	1.3
	ING. AGRONOMIA	4	.0	.0	1.3
	ING. CIVIL	9	.1	.1	1.4
	ING. COMERCIAL	16	.2	.2	1.6
	ING. ELECTRONICO	9	.1	.1	1.7
	ING. FINANZAS/BANCA	11	.1	.1	1.8
	ING. INDUSTRIAL	1	.0	.0	1.8
	ING. MECANICO	4	.0	.0	1.8
	ING. MINAS Y PETROLEO	2	.0	.0	1.9
	ING. QUIMICO	2	.0	.0	1.9
	ING. SISTEMAS	28	.3	.3	2.2
	LCDO. ADMINISTRACION	19	.2	.2	2.4
	LCDO. CIENCIAS EDUCACION	99	1.1	1.1	3.4
	LCDO. CONTAB./AUDITORIA	39	.4	.4	3.9
	LCDO. FINANZAS/BANCA	5	.1	.1	3.9
	LCDO. LEYES	4	.0	.0	4.0
	MECANICA	9	.1	.1	4.1
	no registra	8805	94.6	94.6	98.7
	PROFESOR	54	.6	.6	99.3
	QUIMICA Y BIOLOGIA	2	.0	.0	99.3
	SECRETARIA EJECUTIVA	8	.1	.1	99.4
	SECRETARIADO	1	.0	.0	99.4
	SOCIALES	2	.0	.0	99.4
	TEC. ADMINISTRACION	4	.0	.0	99.5
	TEC. INFORMATICA	3	.0	.0	99.5
	TEC.ELECTRONICA	4	.0	.0	99.5
	TGNLO. ELECTRONICA	3	.0	.0	99.6
	TGNLO. MEDICO	32	.3	.3	99.9
	TGNLO. QUIMICA	1	.0	.0	99.9

TGNLO. SISTEMAS	8	.1	.1	100.0
Total	9305	100.0	100.0	

18. **SOLIC_PROP_NEG**: Determina la solicitante propietario del negocio, el 96.5% el valor del ítem es "no registra", es decir existe una concentración en este ítem, es una variable que no va aportar al modelo y va a ser descartada.

SOLIC_PROP_NEG

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	no registra	8968	96.5	96.5	96.5
	Si	327	3.5	3.5	100.0
	Total	9295	100.0	100.0	

19. **TIEMPO_NEG**: Determina el tiempo del negocio, el 99.5% el valor del ítem es "no registra", es decir existe una concentración en este ítem, es una variable que no va aportar al modelo y va a ser descartada.

TIEMPO_NEG

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	DE 0 A 24	19	.2	.2	.2
	DE 25 A 48	11	.1	.1	.3
	DE 49 A 72	8	.1	.1	.4
	DE 73 A 96	3	.0	.0	.4
	MAS DE 120	2	.0	.0	.5
	no registra	9262	99.5	99.5	100.0
	Total	9305	100.0	100.0	

20. **TIPO_VIVIENDA**: Determina al tipo de vivienda, el 47.7% representa al ítem "no registra", es decir existe una concentración en este ítem, es una variable que no va aportar al modelo por tener un gran porcentaje que no ha registrado esta variable y va a ser descartada.

TIPO_VIVIENDA

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	FAMILIAR	2464	26.5	26.5	26.5
	no registra	4443	47.7	47.7	74.2
	PROPIA	2398	25.8	25.8	100.0
	Total	9305	100.0	100.0	

21. **TIPO_CREDITO**: Determina la tipo de crédito. El tipo de crédito es de consumo, es una variable que no va aportar al modelo y va a ser descartada.

TIPO_CREDITO

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	CCA	1	.0	.0	.0
	CDF	1845	19.8	19.8	19.8
	CE	3	.0	.0	19.9

CI	596	6.4	6.4	26.3
CMV	371	4.0	4.0	30.3
CO	7	.1	.1	30.3
CS	18	.2	.2	30.5
CSG	1800	19.3	19.3	49.9
PF	4635	49.8	49.8	99.7
PH	20	.2	.2	99.9
VCH	9	.1	.1	100.0
Total	9305	100.0	100.0	

21. **SECTOR:** Determina el tipo de sector del crédito, el 99.9% el valor corresponde al ítem "P", es decir existe una concentración en este ítem, es una variable que no va aportar al modelo y va a ser descartada.

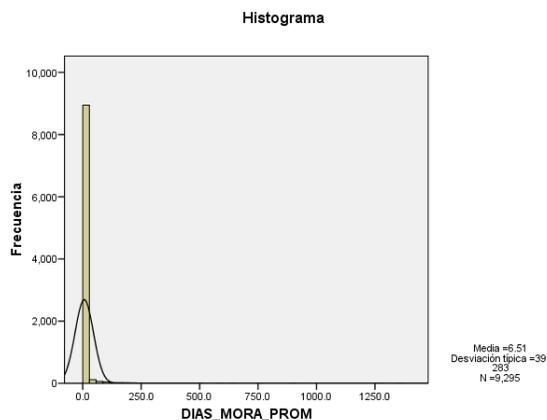
SECTOR

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	P	9291	100.0	100.0	100.0
	PR	4	.0	.0	100.0
	Total	9295	100.0	100.0	

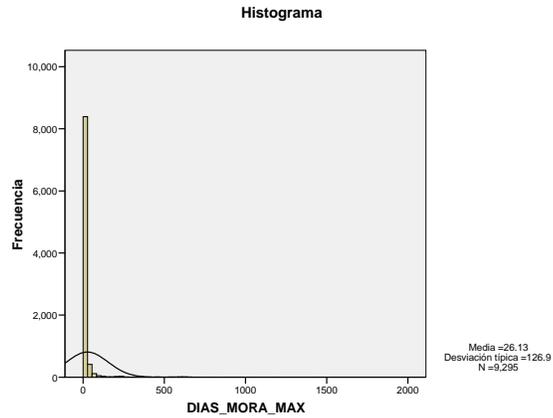
VARIABLE DEPENDIENTE "tipo pagador" que representa el "incumplimiento"

La variable dependiente "tipo pagador" o "incumplimiento", se constituye como una variable discreta que toma dos valores: "buen cliente" y "mal cliente".

A. DIAS_MORA_PROM: Representa los días de mora promedio, esta variable representa el número de días que un crédito en promedio estuvo en mora.



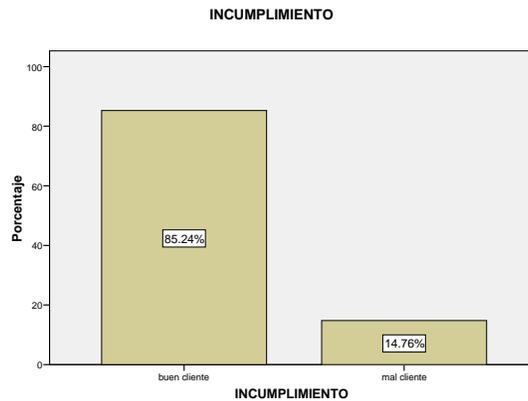
B. DIAS_MORA_MAX: Representa los días de mora máximo, esta variable representa el número máximo de días que un crédito estuvo en mora.



Utilizando una matriz de contingencia para la morosidad promedio y la morosidad máxima de la cartera de clientes de la cartera de consumo se obtiene una primera clasificación de "*clientes buenos*" y "*clientes malos*" al estructurar las intersecciones. Los resultados se muestran en la variable "*incumplimiento*".

INCUMPLIMIENTO

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0=buen cliente	7923	85.2	85.2	85.2
	1=mal cliente	1372	14.8	14.8	100.0
	Total	9295	100.0	100.0	



ANEXO 2: ANÁLISIS EXPLORATORIO BIVARIANTE DE VARIABLES CUANTITATIVAS Y CUALITATIVAS

Luego del análisis exploratorio de variables cuantitativas y cualitativas del conjunto de variables entregadas por la institución, en primera instancia se trato de considerar algunas variables que "aportan"³ y descartar aquellas que "no aportan". En este ANEXO trataremos de evaluar el comportamiento de cada una de las variables independientes vs. el incumplimiento. Se realizara mediante el *análisis bivalente* como sustento estadístico para priorizar la incorporación de variables al modelo, esperando que en la regresión final se queden las variables independientes que mejor expliquen el comportamiento entre clientes buenos y malos.

Análisis Bivalente ⁴: La metodología requiere realizar cruces de variables mediante tablas de contingencia, cruces entre las variables independientes vs. la variable incumplimiento que es la dependiente. Para medir el nivel de discriminación entre las variables independientes y la dependiente se contrasta la hipótesis de independencia con el estadístico de prueba Ji-cuadrado de Person (χ^2 en ingles *Chi-square*), donde, el estadístico permitirá determinar si la variable considerada explica o no el incumplimiento.

El estadístico Ji-cuadrado de Person para probar Ho es igual:

$$\chi^2 = \sum \frac{(n_{ij} - u_{ij})^2}{u_{ij}}$$

Donde, la hipótesis nula (Ho) define la probabilidad de que la variable considerada no explique el incumplimiento ($p=0$), y que la hipótesis alternativa (H1) si explique.

Análisis de Variables Cualitativas: El análisis bivalente permitirá medir el nivel de discriminación entre las variables independientes y el incumplimiento se contrasta la hipótesis de independencia con el estadístico de prueba Ji-cuadrado de Person.

CAL_C_R vs INCUMPLIMIENTO: Para la variable CAL_C_R (*calificación de la central de riesgos*) el porcentaje de discriminación de esta variable vs. el incumplimiento, presentan diferencias significativas, el mismo que se confirma con el nivel de significancia del estadístico Ji-cuadrado, en este sentido, la variable "aporta", es decir es considerada para el modelo.

Tabla de contingencia CAL_C_R * INCUMPLIMIENTO

		INCUMPLIMIENTO		Total
		mal cliente	buen cliente	
CAL_C_R A	Recuento	778	5287	6065
	% de INCUMPLIMIENTO	62.8%	65.6%	65.3%
B	Recuento	20	27	47
	% de INCUMPLIMIENTO	1.6%	.3%	.5%
C o menos	Recuento	3	30	33
	% de INCUMPLIMIENTO	.2%	.4%	.4%
C O MENOS	Recuento	4	19	23
	% de INCUMPLIMIENTO	.3%	.2%	.2%
no registra	Recuento	434	2693	3127
	% de INCUMPLIMIENTO	35.0%	33.4%	33.6%
Total	Recuento	1239	8056	9295
	% de INCUMPLIMIENTO	100.0%	100.0%	100.0%

³ Variable que aporta.- Contribuir con alguna cosa para el logro o la realización de algún objetivo común. Para nuestro caso es la suma de variables independientes aportaran en el modelo credit score de cartera de consumo.

⁴ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	37.725 ^a	4	.000
Razón de verosimilitudes	27.240	4	.000
N de casos válidos	9295		

a. 2 casillas (20.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 3.07.

ESTAB_TRAB_ACT vs INCUMPLIMIENTO: Para la variable ESTAB_TRAB_ACT (*estabilidad del trabajo actual*) el porcentaje, no presentan diferencias significativas, el mismo que se confirma con el nivel de significancia del estadístico Ji-cuadrado, en este sentido, la variable "no aporta", es decir es descartada para el modelo. Hay concentración en un ítem ("2=Mas de 1" con el 89.3%)

Tabla de contingencia ESTAB_TRAB_ACT_REC * INCUMPLIMIENTO

			INCUMPLIMIENTO		Total
			buen cliente	mal cliente	
ESTAB_TRAB_ACT_REC	no registra	Recuento	431	48	479
		% de INCUMPLIMIENTO	8.1%	5.5%	7.8%
	Menos de 1	Recuento	149	32	181
		% de INCUMPLIMIENTO	2.8%	3.7%	2.9%
	Mas de 1	Recuento	4720	788	5508
		% de INCUMPLIMIENTO	89.1%	90.8%	89.3%
Total		Recuento	5300	868	6168
		% de INCUMPLIMIENTO	100.0%	100.0%	100.0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	8.699 ^a	2	.013
Razón de verosimilitudes	9.191	2	.010
Asociación lineal por lineal	4.569	1	.033
N de casos válidos	6168		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 25.47.

EST_CIVIL vs INCUMPLIMIENTO: Para la variable EST_CIVIL (*estado civil*) el porcentaje de discriminación de esta variable vs. el incumplimiento, presentan diferencias significativas, el mismo que se confirma con el nivel de significancia del estadístico Ji-cuadrado, en este sentido, la variable "aporta", es decir es considerada para el modelo.

Tabla de contingencia EST_CIVIL_REC * INCUMPLIMIENTO

			INCUMPLIMIENTO		Total	
			buen cliente	mal cliente		
EST_CIVIL_REC	SOLTERO	Recuento	1950	389	2339	
		% de INCUMPLIMIENTO	36.8%	44.8%	37.9%	
	CASADO	Recuento	2899	399	3298	
		% de INCUMPLIMIENTO	54.7%	46.0%	53.5%	
	DIVORCIADO	Recuento	289	63	352	
		% de INCUMPLIMIENTO	5.5%	7.3%	5.7%	
	VIUDO	Recuento	162	17	179	
		% de INCUMPLIMIENTO	3.1%	2.0%	2.9%	
	Total		Recuento	5300	868	6168
			% de INCUMPLIMIENTO	100.0%	100.0%	100.0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	30.651 ^a	3	.000
Razón de verosimilitudes	30.568	3	.000
Asociación lineal por lineal	10.940	1	.001
N de casos válidos	6168		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 25.19.

EXP_CRED_COOP vs INCUMPLIMIENTO: Para la variable EXP_CRED_COOP (*calificación experiencia crédito Coop.*) el porcentaje de discriminación de esta variable vs. el incumplimiento, presentan diferencias significativas, el mismo que se confirma con el nivel de significancia del estadístico Ji-cuadrado, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Tabla de contingencia EXP_CRED_COOP * INCUMPLIMIENTO

		INCUMPLIMIENTO		Total
		buen cliente	mal cliente	
EXP_CRED_COOP A	Recuento	2774	360	3134
	% de INCUMPLIMIENTO	52.3%	41.5%	50.8%
B	Recuento	106	30	136
	% de INCUMPLIMIENTO	2.0%	3.5%	2.2%
C	Recuento	14	6	20
	% de INCUMPLIMIENTO	.3%	.7%	.3%
D	Recuento	0	1	1
	% de INCUMPLIMIENTO	.0%	.1%	.0%
E	Recuento	8	2	10
	% de INCUMPLIMIENTO	.2%	.2%	.2%
no registra	Recuento	2398	469	2867
	% de INCUMPLIMIENTO	45.2%	54.0%	46.5%
Total	Recuento	5300	868	6168
	% de INCUMPLIMIENTO	100.0%	100.0%	100.0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	47.483 ^a	5	.000
Razón de verosimilitudes	43.998	5	.000
N de casos válidos	6168		

a. 4 casillas (33.3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es .14.

GENERO vs INCUMPLIMIENTO: Para la variable GENERO el porcentaje, no presentan diferencias significativas, el mismo que se confirma con el nivel de significancia del estadístico Ji-cuadrado, en este sentido, la variable "*no aporta*", es decir es descartada para el modelo.

Tabla de contingencia GENERO * INCUMPLIMIENTO

		INCUMPLIMIENTO		Total
		buen cliente	mal cliente	
GENERO FEMENINO	Recuento	2716	437	3153
	% de INCUMPLIMIENTO	51.2%	50.3%	51.1%
MASCULINO	Recuento	2584	431	3015
	% de INCUMPLIMIENTO	48.8%	49.7%	48.9%
Total	Recuento	5300	868	6168
	% de INCUMPLIMIENTO	100.0%	100.0%	100.0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	.242 ^b	1	.623		
Corrección por continuidad ^a	.207	1	.649		
Razón de verosimilitudes	.242	1	.623		
Estadístico exacto de Fisher				.634	.325
N de casos válidos	6168				

a. Calculado sólo para una tabla de 2x2.

b. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 424.29.

NIVEL_EDUC vs INCUMPLIMIENTO: Para la variable NIVEL_EDUC (*nivel de educación*) el porcentaje de discriminación de esta variable vs. el incumplimiento, presentan diferencias significativas, el mismo que se confirma con el nivel de significancia del estadístico Ji-cuadrado, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Tabla de contingencia NIVEL_EDUC_REC * INCUMPLIMIENTO

			INCUMPLIMIENTO		Total
			buen cliente	mal cliente	
NIVEL_EDUC_REC	no registra	Recuento	339	58	397
		% de INCUMPLIMIENTO	6.4%	6.7%	6.4%
PRIMARIA		Recuento	211	17	228
		% de INCUMPLIMIENTO	4.0%	2.0%	3.7%
SECUNDARIA		Recuento	1328	159	1487
		% de INCUMPLIMIENTO	25.1%	18.3%	24.1%
SECUNDARIA		Recuento	2961	557	3518
		% de INCUMPLIMIENTO	55.9%	64.2%	57.0%
ESTUDIANTE SUPERIOR		Recuento	461	77	538
		% de INCUMPLIMIENTO	8.7%	8.9%	8.7%
Total		Recuento	5300	868	6168
		% de INCUMPLIMIENTO	100.0%	100.0%	100.0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	31.437 ^a	4	.000
Razón de verosimilitudes	33.671	4	.000
Asociación lineal por lineal	8.660	1	.003
N de casos válidos	6168		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 32.09.

Análisis Variables Cuantitativas: Realizaremos un análisis bivalente de medias y varianza (*ANOVA*) de cada variable independiente vs. el incumplimiento, que se contrasta con la prueba de independencia entre grupos o muestras al 95% de confianza. ⁵

CART_CASTIG: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

CART_CASTIG

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	.00	.000	.000
mal cliente	868	93.37	305.922	10.384
Total	6168	13.14	119.213	1.518

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
CART_CASTIG * INCUMPLIMIENTO	Inter-grupos (Combinadas)		6502637.2	1	6502637.2	494.142	.000
	Intra-grupos		81141107	6166	13159.440		
	Total		87643744	6167			

CART_NO DEVEN: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

⁵ TOALOMBO Franklin, Programa de maestría en Finanzas y gestión de riesgos, Diseño de un modelo de scoring para el segmento de microcrédito individual urbano para la Cooperativa de ahorro y crédito San Francisco de Asís, Universidad Andina Simón Bolívar, 2010. Esta tesis es apoyo para la construcción del modelo de evaluación credit score de la cartera de consumo de la Cooperativa Maquita Cushunchic.

Informe

CART_NO_DEVEN

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	.90	44.669	.614
mal cliente	868	72.92	511.368	17.357
Total	6168	11.03	197.750	2.518

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
CART_NO_DEVEN	Inter-grupos	(Combinadas)	3869276.3	1	3869276.3	100.543	.000
* INCUMPLIMIENTO	Intra-grupos		237291059	6166	38483.792		
	Total		241160335	6167			

CART POR VENCER: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

CART_POR_VENCER

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	730.5561	1646.540	22.61696
mal cliente	868	209.4700	898.78975	30.50691
Total	6168	657.2256	1573.505	20.03530

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
CART_POR_VENCER	Inter-grupos	(Combinadas)	202521068	1	2.0E+008	82.882	.000
* INCUMPLIMIENTO	Intra-grupos		1.51E+010	6166	2443474.3		
	Total		1.53E+010	6167			

CART VENCIDA: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

CART_VENCIDA

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	.03	1.349	.019
mal cliente	868	23.02	133.059	4.516
Total	6168	3.27	50.543	.644

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
CART_VENCIDA *	Inter-grupos	(Combinadas)	394183.617	1	394183.617	158.241	.000
INCUMPLIMIENTO	Intra-grupos		15359722	6166	2491.035		
	Total		15753906	6167			

CUOTA: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

CUOTA

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	126.9670	184.55396	2.53504
mal cliente	868	132.7460	100.60625	3.41480
Total	6168	127.7803	175.19484	2.23074

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
CUOTA * INCUMPLIMIENTO	Inter-grupos (Combinadas)		24909.073	1	24909.073	.812	.037
	Intra-grupos		189260257	6166	30694.171		
	Total		189285166	6167			

CUOTAS PAG: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que no existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "no aporta", es decir es descartada para el modelo.

Informe

CUOTAS_PAG

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	13.36	5.814	.080
mal cliente	868	14.44	6.642	.225
Total	6168	13.51	5.949	.076

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
CUOTAS_PAG * INCUMPLIMIENTO	Inter-grupos (Combinadas)		874.706	1	874.706	24.814	.647
	Intra-grupos		217350.588	6166	35.250		
	Total		218225.294	6167			

DEUDAS DIREC: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que no existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "no aporta", es decir es descartada para el modelo.

Informe

DEUDAS_DIREC

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	1175.00	2313.153	31.774
mal cliente	868	1364.14	2581.240	87.613
Total	6168	1201.62	2353.424	29.966

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
DEUDAS_DIREC * INCUMPLIMIENTO	Inter-grupos (Combinadas)		26681600	1	26681600	4.820	.028
	Intra-grupos		3.41E+010	6166	5535173.6		
	Total		3.42E+010	6167			

DEUDAS INDIRECT: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que no existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "no aporta", es decir es descartada para el modelo.

Informe

DEUDAS_INDIREC				
INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	640.87	2486.243	34.151
mal cliente	868	696.93	2262.419	76.792
Total	6168	648.76	2455.882	31.271

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
DEUDAS_INDIREC * INCUMPLIMIENTO	Inter-grupos (Combinadas)		2344158.5	1	2344158.5	.389	.533
	Intra-grupos		3.72E+010	6166	6031953.4		
	Total		3.72E+010	6167			

EDAD: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

EDAD				
INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	39.48	12.345	.170
mal cliente	868	36.88	11.976	.406
Total	6168	39.11	12.326	.157

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
EDAD * INCUMPLIMIENTO	Inter-grupos (Combinadas)		5050.978	1	5050.978	33.420	.000
	Intra-grupos		931902.485	6166	151.136		
	Total		936953.463	6167			

INGRESOS: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

INGRESOS				
INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	490.71	364.908	5.012
mal cliente	868	559.55	501.809	17.033
Total	6168	500.40	387.802	4.938

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
INGRESOS * INCUMPLIMIENTO	Inter-grupos (Combinadas)		3534273.6	1	3534273.6	23.587	.000
	Intra-grupos		923924078	6166	149841.725		
	Total		927458352	6167			

MIEMBROS_FAM: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

MIEMBROS_FAM				
INCUMPLIMIENTO	N	Media	Desv. tıp.	Error tıp. de la media
buen cliente	5300	2.65	1.324	.018
mal cliente	868	2.43	1.345	.046
Total	6168	2.62	1.329	.017

Tabla de ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
MIEMBROS_FAM *	Inter-grupos (Combinadas)	36.245	1	36.245	20.582	.000
INCUMPLIMIENTO	Intra-grupos	10858.493	6166	1.761		
	Total	10894.739	6167			

MONTO: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

MONTO				
INCUMPLIMIENTO	N	Media	Desv. tıp.	Error tıp. de la media
buen cliente	5300	2030.4674	2137.423	29.35976
mal cliente	868	2265.8333	2196.223	74.54466
Total	6168	2063.5896	2147.175	27.33979

Tabla de ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
MONTO *	Inter-grupos (Combinadas)	41317887	1	41317887	8.974	.003
INCUMPLIMIENTO	Intra-grupos	2.84E+010	6166	4604405.3		
	Total	2.84E+010	6167			

N_VMORA_M5: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

N_VMORA_M5				
INCUMPLIMIENTO	N	Media	Desv. tıp.	Error tıp. de la media
buen cliente	5300	.39	1.101	.015
mal cliente	868	3.35	3.334	.113
Total	6168	.81	1.914	.024

Tabla de ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
N_VMORA_M5 *	Inter-grupos (Combinadas)	6543.353	1	6543.353	2512.161	.000
INCUMPLIMIENTO	Intra-grupos	16060.405	6166	2.605		
	Total	22603.758	6167			

NUM_CUOTAS: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que no existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*no aporta*", es decir es descartada para el modelo.

Informe

NUM_CUOTAS

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	15.80	6.318	.087
mal cliente	868	16.20	5.957	.202
Total	6168	15.86	6.270	.080

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
NUM_CUOTAS * INCUMPLIMIENTO	Inter-grupos (Combinadas)		121.900	1	121.900	3.102	.078
	Intra-grupos		242295.101	6166	39.295		
	Total		242417.001	6167			

PATRIMONIO: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que no existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*no aporta*", es decir es descartada para el modelo.

Informe

PATRIMONIO

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	11518.03	24534.039	337.001
mal cliente	868	8588.14	17141.223	581.811
Total	6168	11105.72	23654.707	301.193

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
PATRIMONIO * INCUMPLIMIENTO	Inter-grupos (Combinadas)		6.40E+009	1	6.4E+009	11.462	.715
	Intra-grupos		3.44E+012	6166	5.6E+008		
	Total		3.45E+012	6167			

RECIPROCIDAD: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

RECIPROCIDAD

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	82.0265	454.53609	6.24353
mal cliente	868	32.2366	192.87591	6.54664
Total	6168	75.0197	427.84771	5.44775

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
RECIPROCIDAD * INCUMPLIMIENTO	Inter-grupos (Combinadas)		1848988.5	1	1848988.5	10.116	.001
	Intra-grupos		1.13E+009	6166	182783.483		
	Total		1.13E+009	6167			

SALDO_OPER: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*aporta*", es decir es considerada para el modelo.

Informe

SALDO_OPER

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	737.4114	1659.739	22.79827
mal cliente	868	517.0352	1268.563	43.05783
Total	6168	706.3987	1612.179	20.52774

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
SALDO_OPER * INCUMPLIMIENTO	Inter-grupos (Combinadas)		36222684	1	36222684	13.966	.000
	Intra-grupos		1.60E+010	6166	2593668.6		
	Total		1.60E+010	6167			

SUELDO BASICO: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que no existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*no aporta*", es decir es descartada para el modelo.

Informe

SUELDO_BASICO

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	419.94	339.472	4.663
mal cliente	868	478.60	459.522	15.597
Total	6168	428.20	359.338	4.575

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
SUELDO_BASICO * INCUMPLIMIENTO	Inter-grupos (Combinadas)		2565985.9	1	2565985.9	19.933	.816
	Intra-grupos		793741000	6166	128728.673		
	Total		796306986	6167			

TASA: El test de medias y varianza (*ANOVA*), muestra evidencia estadística que presume que no existe independencia entre los grupos de esta variable vs. el incumplimiento, además, los resúmenes de medias y desviación estándar de los grupos mantienen diferencias significativas, en este sentido, la variable "*no aporta*", es decir es descartada para el modelo.

Informe

TASA

INCUMPLIMIENTO	N	Media	Desv. típ.	Error típ. de la media
buen cliente	5300	15.834	2.7398	.0376
mal cliente	868	15.936	3.3361	.1132
Total	6168	15.848	2.8312	.0360

Tabla de ANOVA

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
TASA * INCUMPLIMIENTO	Inter-grupos (Combinadas)		7.791	1	7.791	.972	.324
	Intra-grupos		49426.432	6166	8.016		
	Total		49434.223	6167			

ANEXO 3: REGRESION LOGISTICA DE VARIABLES

El objetivo del modelo logístico encontrar la *función* que mejor discriminen a la población de "buenos clientes" y "malos clientes" con el conjunto de variables independientes "candidatas", la incorporación de variables se lo realizara paso a paso⁶ las variables explicativas con la variable dependiente a fin de obtener un modelo logístico que se ajuste a la ecuación Logit:

$$P_j = \frac{e^{Z_j}}{1 + e^{Z_j}}; z_j; \text{sabiendo que } Z_j = B_1Z_1 + B_2Z_2 + \dots + B_pZ_p \quad (\text{ec. 3})$$

Se realizara entonces las corridas necesarias con diferentes combinaciones de variables y finalmente se escogerán las corridas con mayor significación estadística, tratando de utilizar la definición que el valor de p⁷ sea menor a 0.05 al 95% de confianza.

- 1.- LOGISTIC REGRESSION VARIABLES INCUMPLIMIENTO
 /METHOD = ENTER CAL_CR_COOP_REC CAL_CR_REC
 /SAVE = PRED PGROUP RESID
 /PRINT = GOODFIT CORR SUMMARY
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	95.793	8	.000
	Bloque	95.793	8	.000
	Modelo	95.793	8	.000

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	7684.760 ^a	.010	.018

a. La estimación ha finalizado en el número de iteración 20 porque se han alcanzado las iteraciones máximas. No se puede encontrar una solución definitiva.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	4.282	4	.369

Tabla de clasificación^a

Observado		Pronosticado			
		INCUMPLIMIENTO		Porcentaje correcto	
		.00	1.00		
Paso 1	INCUMPLIMIENTO	.00	7923	0	100.0
		1.00	1371	1	.1
Porcentaje global					85.3

a. El valor de corte es .500

⁶ Método paso a paso: O denominado step-wise, consiste en ir acondicionando variables a un modelo, usando el estadístico G (*Wald*), para saber si la nueva variable explica significativamente la variable respuesta. De igual manera, este procedimiento elimina variables ya incluidas que por el efecto de la inclusión de la variable, esta puede ser no significativa.

⁷ Valor de p, se emplea con frecuencia en las pruebas de hipótesis y representan la probabilidad de que la estadística de prueba tomaría un valor al menos tan extremo como el valor observado en ella cuando Ho es verdadera, es decir es el nivel de significancia mas pequeño que conduciría al rechazo de Ho.

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1			75.547	5	.000	
CAL_CR_COOP_REC						
CAL_CR_COOP_REC(1)	-.493	.061	64.715	1	.000	.611
CAL_CR_COOP_REC(2)	.301	.188	2.550	1	.110	1.351
CAL_CR_COOP_REC(3)	.573	.493	1.349	1	.245	1.774
CAL_CR_COOP_REC(4)	21.799	40192.969	.000	1	1.000	3E+009
CAL_CR_COOP_REC(5)	.633	.686	.852	1	.356	1.883
CAL_CR_REC			13.508	3	.004	
CAL_CR_REC(1)	-.133	.062	4.619	1	.032	.876
CAL_CR_REC(2)	.875	.327	7.167	1	.007	2.398
CAL_CR_REC(3)	-.217	.388	.312	1	.576	.805
Constante	-1.471	.053	765.793	1	.000	.230

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC, CAL_CR_REC.

Matriz de correlaciones

	Constant	CAL_CR_COOP_REC(1)	CAL_CR_COOP_REC(2)	CAL_CR_COOP_REC(3)	CAL_CR_COOP_REC(4)	CAL_CR_COOP_REC(5)	CAL_CR_REC(1)	CAL_CR_REC(2)	CAL_CR_REC(3)
Paso 1									
Constant	1.000	-.391	-.107	-.033	.000	-.021	-.687	-.125	-.113
CAL_CR_COOP_REC(1)	-.391	1.000	.134	.047	.000	.035	-.093	-.029	.006
CAL_CR_COOP_REC(2)	-.107	.134	1.000	.017	.000	.013	-.059	-.024	-.006
CAL_CR_COOP_REC(3)	-.033	.047	.017	1.000	.000	.042	-.020	.001	-.258
CAL_CR_COOP_REC(4)	.000	.000	.000	.000	1.000	.000	.000	.000	.000
CAL_CR_COOP_REC(5)	-.021	.035	.013	.042	.000	1.000	-.020	.000	-.146
CAL_CR_REC(1)	-.687	-.093	-.059	-.020	.000	-.020	1.000	.122	.105
CAL_CR_REC(2)	-.125	-.029	-.024	.001	.000	.000	.122	1.000	.018
CAL_CR_REC(3)	-.113	.006	-.006	-.258	.000	-.146	.105	.018	1.000

2.- LOGISTIC REGRESSION VARIABLES INCUMPLIMIENTO
 /METHOD = ENTER CAL_CR_COOP_REC CAL_CR_REC CUOTA
 /SAVE = PRED PGROUP RESID
 /PRINT = GOODFIT CORR SUMMARY
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .

Pruebas omnibus sobre los coeficientes del modelo

	Chi-cuadrado	gl	Sig.
Paso 1			
Paso	95.809	9	.000
Bloque	95.809	9	.000
Modelo	95.809	9	.000

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	7684.743 ^a	.010	.018

a. La estimación ha finalizado en el número de iteración 20 porque se han alcanzado las iteraciones máximas. No se puede encontrar una solución definitiva.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	36.454	8	.000

Tabla de clasificación^a

Observado		Pronosticado		Porcentaje correcto
		INCUMPLIMIENTO .00	1.00	
Paso 1	INCUMPLIMIENTO	.00		
		7923	0	100.0
	1.00	1371	1	.1
	Porcentaje global			85.3

a. El valor de corte es .500

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1	CAL_CR_COOP_REC			75.555	5	.000	
	CAL_CR_COOP_REC(1)	-.493	.061	64.686	1	.000	.611
	CAL_CR_COOP_REC(2)	.300	.189	2.537	1	.111	1.350
	CAL_CR_COOP_REC(3)	.562	.501	1.258	1	.262	1.754
	CAL_CR_COOP_REC(4)	21.796	40192.969	.000	1	1.000	3E+009
	CAL_CR_COOP_REC(5)	.631	.686	.847	1	.357	1.880
	CAL_CR_REC			13.521	3	.004	
	CAL_CR_REC(1)	-.134	.062	4.628	1	.031	.875
	CAL_CR_REC(2)	.874	.327	7.147	1	.008	2.396
	CAL_CR_REC(3)	-.217	.388	.314	1	.575	.805
	CUOTA	.000	.000	.017	1	.896	1.000
	Constante	-1.473	.055	710.739	1	.000	.229

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC, CAL_CR_REC, CUOTA.

Matriz de correlaciones

		Constant	CAL_CR_COOP_REC(1)	CAL_CR_COOP_REC(2)	CAL_CR_COOP_REC(3)	CAL_CR_COOP_REC(4)	CAL_CR_COOP_REC(5)	CAL_CR_REC(1)	CAL_CR_REC(2)	CAL_CR_REC(3)	CUOTA
Paso 1	Constant	1.000	-.364	-.095	.016	.000	-.015	-.629	-.114	-.106	-.273
	CAL_CR_COOP_REC(1)	-.364	1.000	.135	.053	.000	.036	-.088	-.028	.007	-.043
	CAL_CR_COOP_REC(2)	-.095	.135	1.000	.021	.000	.013	-.056	-.023	-.005	-.027
	CAL_CR_COOP_REC(3)	.016	.053	.021	1.000	.000	.045	-.001	.005	-.252	-.173
	CAL_CR_COOP_REC(4)	.000	.000	.000	.000	1.000	.000	.000	.000	.000	.000
	CAL_CR_COOP_REC(5)	-.015	.036	.013	.045	.000	1.000	-.018	.000	-.146	-.019
	CAL_CR_REC(1)	-.629	-.088	-.056	-.001	.000	-.018	1.000	.123	.106	-.103
	CAL_CR_REC(2)	-.114	-.028	-.023	.005	.000	.123	1.000	.018	-.023	-.023
	CAL_CR_REC(3)	-.106	.007	-.005	-.252	.000	-.146	.106	.018	1.000	-.013
	CUOTA	-.273	-.043	-.027	-.173	.000	-.019	-.103	-.023	-.013	1.000

- 3.- LOGISTIC REGRESSION VARIABLES INCUMPLIMIENTO
 /METHOD = ENTER CAL_CR_COOP_REC CAL_CR_REC CUOTA SALDO_OPER
 /SAVE = PRED PGROUP RESID
 /PRINT = GOODFIT CORR SUMMARY
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	109.459	10	.000
	Bloque	109.459	10	.000
	Modelo	109.459	10	.000

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	7671.094 ^a	.012	.021

a. La estimación ha finalizado en el número de iteración 20 porque se han alcanzado las iteraciones máximas. No se puede encontrar una solución definitiva.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	36.356	8	.000

Tabla de clasificación^a

	Observado		Pronosticado		Porcentaje correcto
			INCUMPLIMIENTO .00	1.00	
Paso 1	INCUMPLIMIENTO	.00	7922	1	100.0
		1.00	1371	1	.1
	Porcentaje global				85.2

a. El valor de corte es .500

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1			74.585	5	.000	
CAL_CR_COOP_REC						
CAL_CR_COOP_REC(1)	-.489	.061	63.450	1	.000	.613
CAL_CR_COOP_REC(2)	.321	.189	2.896	1	.089	1.379
CAL_CR_COOP_REC(3)	.496	.502	.977	1	.323	1.643
CAL_CR_COOP_REC(4)	22.090	40192.969	.000	1	1.000	4E+009
CAL_CR_COOP_REC(5)	.690	.689	1.003	1	.317	1.994
CAL_CR_REC			11.051	3	.011	
CAL_CR_REC(1)	-.103	.063	2.731	1	.098	.902
CAL_CR_REC(2)	.868	.327	7.039	1	.008	2.383
CAL_CR_REC(3)	-.154	.388	.158	1	.691	.857
CUOTA	.000	.000	1.053	1	.305	1.000
SALDO_OPER	.000	.000	12.195	1	.000	1.000
Constante	-1.466	.055	714.506	1	.000	.231

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC, CAL_CR_REC, CUOTA, SALDO_OPER.

Matriz de correlaciones

	Constant	CAL_CR_COOP_REC(1)	CAL_CR_COOP_REC(2)	CAL_CR_COOP_REC(3)	CAL_CR_COOP_REC(4)	CAL_CR_COOP_REC(5)	CAL_CR_REC(1)	CAL_CR_REC(2)	CAL_CR_REC(3)	CUOTA	SALDO_OPER
Paso 1											
Constant	1.000	-.372	-.097	.008	.000	-.016	-.631	-.116	-.106	-.227	-.046
CAL_CR_COOP_REC(1)	-.372	1.000	.135	.050	.000	.035	-.088	-.027	.009	-.026	-.013
CAL_CR_COOP_REC(2)	-.097	.135	1.000	.019	.000	.013	-.052	-.024	-.004	-.014	-.029
CAL_CR_COOP_REC(3)	.008	.050	.019	1.000	.000	.042	-.010	.004	-.236	-.173	.034
CAL_CR_COOP_REC(4)	.000	.000	.000	.000	1.000	.000	.000	.000	.000	.000	.000
CAL_CR_COOP_REC(5)	-.016	.035	.013	.042	.000	1.000	-.015	.000	-.153	-.010	-.024
CAL_CR_REC(1)	-.631	-.088	-.052	-.010	.000	-.015	1.000	.121	.110	-.060	-.125
CAL_CR_REC(2)	-.116	-.027	-.024	-.004	.000	.000	.121	1.000	.018	-.021	.003
CAL_CR_REC(3)	-.106	.009	-.004	-.236	.000	-.153	.110	.018	1.000	-.006	-.039
CUOTA	-.227	-.026	-.014	-.173	.000	-.010	-.060	-.021	-.006	1.000	-.193
SALDO_OPER	-.046	-.013	-.029	.034	.000	-.024	-.125	.003	-.039	-.193	1.000

4. LOGISTIC REGRESSION VARIABLES INCUMPLIMIENTO
 /METHOD = ENTER CAL_CR_COOP_REC CAL_CR_REC CUOTA SALDO_OPER CAPACIDAD_PAGO
 /SAVE = PRED PGROUP RESID
 /PRINT = GOODFIT CORR SUMMARY
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .

Pruebas omnibus sobre los coeficientes del modelo

	Chi-cuadrado	gl	Sig.
Paso 1			
Paso	111.632	11	.000
Bloque	111.632	11	.000
Modelo	111.632	11	.000

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	7668.920 ^a	.012	.021

a. La estimación ha finalizado en el número de iteración 20 porque se han alcanzado las iteraciones máximas. No se puede encontrar una solución definitiva.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	4.145	8	.844

Tabla de clasificación^a

Observado	INCUMPLIMIENTO	Pronosticado		Porcentaje correcto
		.00	1.00	
Paso 1	INCUMPLIMIENTO	.00	1.00	100.0
	Porcentaje global	7922	1371	85.2

a. El valor de corte es .500

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1			71,401	5	.000	
CAL_CR_COOP_REC						
CAL_CR_COOP_REC(1)	-.478	.062	59,997	1	.000	.620
CAL_CR_COOP_REC(2)	.331	.189	3,063	1	.080	1,392
CAL_CR_COOP_REC(3)	.493	.503	.960	1	.327	1,637
CAL_CR_COOP_REC(4)	22,032	40192,969	.000	1	1,000	4E+009
CAL_CR_COOP_REC(5)	.708	.689	1,058	1	.304	2,031
CAL_CR_REC						
CAL_CR_REC(1)	-.113	.063	3,228	1	.072	.893
CAL_CR_REC(2)	.860	.327	6,895	1	.009	2,363
CAL_CR_REC(3)	-.154	.388	.158	1	.691	.857
CUOTA	.000	.000	1,562	1	.211	1,000
SALDO_OPER	.000	.000	10,539	1	.001	1,000
CAPACIDAD_PAGO	.013	.009	2,245	1	.134	1,013
Constante	-1,526	.068	501,313	1	.000	.217

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC, CAL_CR_REC, CUOTA, SALDO_OPER, CAPACIDAD_PAGO.

Matriz de correlaciones

	Constant	CAL_CR_COOP_REC(1)	CAL_CR_COOP_REC(2)	CAL_CR_COOP_REC(3)	CAL_CR_COOP_REC(4)	CAL_CR_COOP_REC(5)	CAL_CR_REC(1)	CAL_CR_REC(2)	CAL_CR_REC(3)	CUOTA	SALDO_OPER	CAPACIDAD_PAGO
Paso 1												
Constant	1,000	-.363	-.098	.012	.000	-.023	-.441	-.083	-.084	-.292	-.120	-.593
CAL_CR_COOP_REC(1)	-.363	1,000	.138	.050	.000	.037	-.099	-.029	.008	-.006	.003	.112
CAL_CR_COOP_REC(2)	-.098	.138	1,000	.019	.000	.014	-.056	-.024	-.004	-.009	-.025	.034
CAL_CR_COOP_REC(3)	.012	.050	.019	1,000	.000	.042	-.009	.005	-.236	-.178	.035	-.006
CAL_CR_COOP_REC(4)	.000	.000	.000	.000	1,000	.000	.000	.000	.000	.000	.000	.000
CAL_CR_COOP_REC(5)	-.023	.037	.014	.042	.000	1,000	-.017	.000	-.151	-.007	-.021	.018
CAL_CR_REC(1)	-.441	-.099	-.056	-.009	.000	-.017	1,000	.123	.110	-.079	-.138	-.108
CAL_CR_REC(2)	-.083	-.029	-.024	.005	.000	.000	.123	1,000	.018	-.025	.001	-.018
CAL_CR_REC(3)	-.084	.008	-.004	-.236	.000	-.151	.110	.018	1,000	-.006	-.039	-.001
CUOTA	-.292	-.006	-.009	-.178	.000	-.007	-.079	-.025	-.006	1,000	-.168	.186
SALDO_OPER	-.120	.003	-.025	.035	.000	-.021	-.138	.001	-.039	-.168	1,000	.145
CAPACIDAD_PAGO	-.593	.112	.034	-.006	.000	.018	-.108	-.018	-.001	.186	.145	1,000

5.- LOGISTIC REGRESSION VARIABLES INCUMPLIMIENTO
 /METHOD = ENTER CAL_CR_COOP_REC CAL_CR_REC CUOTA SALDO_OPER CAPACIDAD_PAGO
 EDAD MONTO RECIPROCIDAD EST_CIVIL_REC2 INGRESOS2 MIEMBROS_FAM NIVEL_EDUC_REC2
 /SAVE = PRED PGROUP RESID
 /PRINT = GOODFIT CORR SUMMARY
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .

Pruebas omnibus sobre los coeficientes del modelo

	Chi-cuadrado	gl	Sig.
Paso 1			
Paso	399,316	22	.000
Bloque	399,316	22	.000
Modelo	399,316	22	.000

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	7381,236 ^a	.042	.074

a. La estimación ha finalizado en el número de iteración 20 porque se han alcanzado las iteraciones máximas. No se puede encontrar una solución definitiva.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	6,710	8	.568

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1			39.245	5	.000	
CAL_CR_COOP_REC						
CAL_CR_COOP_REC(1)	-.348	.064	29.419	1	.000	.706
CAL_CR_COOP_REC(2)	.362	.194	3.495	1	.062	1.436
CAL_CR_COOP_REC(3)	.537	.552	.946	1	.331	1.712
CAL_CR_COOP_REC(4)	21.991	40192.969	.000	1	1.000	4E+009
CAL_CR_COOP_REC(5)	.667	.699	.909	1	.340	1.948
CAL_CR_REC			9.655	3	.022	
CAL_CR_REC(1)	-.142	.065	4.700	1	.030	.868
CAL_CR_REC(2)	.665	.346	3.691	1	.055	1.944
CAL_CR_REC(3)	-.146	.393	.137	1	.711	.865
CUOTA	-.004	.001	7.152	1	.007	.996
SALDO_OPER	.000	.000	53.138	1	.000	1.000
CAPACIDAD_PAGO	.010	.014	.531	1	.466	1.010
EDAD	-.017	.003	31.011	1	.000	.983
MONTO	.000	.000	24.682	1	.000	1.000
RECIPROCIDAD	-.002	.000	27.528	1	.000	.998
EST_CIVIL_REC2			12.407	3	.006	
EST_CIVIL_REC2(1)	-.023	.213	.012	1	.913	.977
EST_CIVIL_REC2(2)	.375	.237	2.495	1	.114	1.455
EST_CIVIL_REC2(3)	.213	.216	.971	1	.324	1.237
INGRESOS2	.000	.000	8.408	1	.004	1.000
MIEMBROS_FAM	-.082	.031	7.214	1	.007	.921
NIVEL_EDUC_REC2			19.637	3	.000	
NIVEL_EDUC_REC2(1)	.224	.130	2.981	1	.084	1.251
NIVEL_EDUC_REC2(2)	.409	.119	11.847	1	.001	1.505
NIVEL_EDUC_REC2(3)	.081	.157	.268	1	.605	1.085
Constante	-1.214	.302	16.148	1	.000	.297

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC, CAL_CR_REC, CUOTA, SALDO_OPER, CAPACIDAD_PAGO, EDAD, MONTO, RECIPROCIDAD, EST_CIVIL_REC2, INGRESOS2, MIEMBROS_FAM, NIVEL_EDUC_REC2.

Matriz de correlaciones

	Constant	CAL_CR_COOP_REC(1)	CAL_CR_COOP_REC(2)	CAL_CR_COOP_REC(3)	CAL_CR_COOP_REC(4)	CAL_CR_COOP_REC(5)	CAL_CR_REC(1)	CAL_CR_REC(2)	CAL_CR_REC(3)	CUOTA	SALDO_OPER	CAPACIDAD_PAGO	EDAD	MONTO	RECIPROCIDAD	EST_CIVIL_REC2(1)	EST_CIVIL_REC2(2)	EST_CIVIL_REC2(3)	INGRESOS2	MIEMBROS_FAM	NIVEL_EDUC_REC2(1)	NIVEL_EDUC_REC2(2)	NIVEL_EDUC_REC2(3)	
Paso 1	Constant	1.000	-.016	.000	.021	.000	.004	-.069	-.018	-.021	-.219	-.000	-.209	-.535	-.148	.000	-.862	-.591	-.791	-.139	-.000	-.000	-.000	-.000
	CAL_CR_COOP_REC(1)	-.016	1.000	.144	.047	.000	.009	-.086	-.022	.000	-.069	.001	.862	-.105	.061	-.114	.016	.022	.004	.002	-.004	.002	-.004	-.002
	CAL_CR_COOP_REC(2)	.000	.144	1.000	.017	.000	.018	-.041	-.019	-.002	-.015	-.024	.043	-.027	.023	-.003	-.007	-.017	-.004	-.031	-.001	-.004	-.031	-.001
	CAL_CR_COOP_REC(3)	.021	.047	.017	1.000	.000	.036	-.008	.005	-.263	.045	.052	-.017	-.022	-.086	.011	-.002	.008	-.010	.022	.002	-.010	.022	-.010
	CAL_CR_COOP_REC(4)	.000	.000	.000	.000	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	CAL_CR_COOP_REC(5)	.004	.039	.018	.006	.000	1.000	-.010	.005	-.122	-.013	-.017	.020	-.008	.017	.012	-.009	.007	.000	-.018	.000	-.018	.000	-.018
	CAL_CR_REC(1)	-.069	-.086	-.041	-.008	.000	-.010	1.000	.130	.113	-.047	-.058	-.081	-.018	.005	.030	-.015	-.012	.001	-.023	-.001	-.023	-.001	-.023
	CAL_CR_REC(2)	-.018	-.022	-.019	.005	.000	.005	.130	1.000	.019	.024	.025	.024	-.017	-.024	.017	.003	.018	.004	-.059	-.001	-.059	-.001	-.059
	CAL_CR_REC(3)	-.021	.000	-.002	.263	.000	-.122	.113	.019	1.000	-.006	-.048	.004	-.009	.013	.017	-.003	-.016	.002	-.003	-.003	-.003	-.003	-.003
	CUOTA	-.219	-.000	-.015	.045	.000	-.013	-.047	.024	-.006	1.000	.157	.439	.009	-.893	-.049	-.012	-.013	.000	-.418	.000	-.418	.000	-.418
	SALDO_OPER	-.000	.001	.024	.052	.000	-.017	-.058	.025	-.048	.157	1.000	-.012	.065	-.331	.002	.027	.010	.015	.002	.015	.002	.015	.002
	CAPACIDAD_PAGO	-.009	.862	.043	-.017	.000	.020	-.081	.024	.024	.439	-.012	1.000	-.030	-.134	-.023	.011	.010	.008	-.728	-.000	-.728	-.000	-.728
	EDAD	-.535	-.105	-.027	-.022	.000	-.008	-.018	-.017	-.009	.009	.065	-.030	1.000	-.020	-.042	.140	.061	.061	-.063	-.000	-.063	-.000	-.063
	MONTO	.148	.061	.023	-.086	.000	.017	.005	-.024	.013	-.893	-.331	-.134	-.020	1.000	.022	-.003	.007	-.003	-.107	-.001	-.107	-.001	-.107
	RECIPROCIDAD	.000	-.114	-.003	.011	.000	.012	.030	.017	.017	-.049	.002	-.023	-.042	.022	1.000	-.003	.015	.005	.029	.000	.029	.000	.029
	EST_CIVIL_REC2(1)	-.862	.016	-.007	-.002	.000	-.009	-.015	.003	-.003	-.012	.027	.011	.140	-.003	-.003	1.000	.821	-.914	-.016	-.016	-.016	-.016	-.016
	EST_CIVIL_REC2(2)	-.591	.022	-.017	-.008	.000	.007	-.012	.018	-.016	.013	.010	.010	.061	.007	.015	.821	1.000	.813	-.026	-.000	-.026	-.000	-.026
	EST_CIVIL_REC2(3)	-.791	.004	-.004	-.010	.000	.000	.001	.004	.002	.000	.015	.008	.283	-.003	.005	.914	.813	1.000	-.022	.000	-.022	.000	-.022
	INGRESOS2	.139	.002	-.031	.022	.000	-.018	-.023	-.059	-.003	-.418	.002	-.726	-.063	.107	.029	-.016	-.026	-.022	1.000	.000	-.022	.000	-.022
	MIEMBROS_FAM	-.008	-.044	-.027	-.038	.000	.010	-.029	-.007	.001	.000	.011	-.076	.071	-.021	.000	-.179	-.036	.073	.000	.000	.073	.000	.073
	NIVEL_EDUC_REC2(1)	-.356	-.017	.012	-.010	.000	-.005	.044	.009	.020	-.008	-.051	.012	-.001	.047	-.005	.019	.034	.032	.000	.000	.032	.000	.032
	NIVEL_EDUC_REC2(2)	-.411	-.022	-.013	-.019	.000	-.014	.004	-.004	.006	-.018	-.028	-.019	.131	.034	.009	.003	.004	.015	.067	.000	.015	.067	.000
	NIVEL_EDUC_REC2(3)	-.289	-.006	-.018	.014	.000	-.050	-.002	-.017	-.023	-.037	-.001	-.027	.075	.018	.007	-.004	-.011	-.001	.017	.000	.017	.000	.017

- 6.- LOGISTIC REGRESSION VARIABLES INCUMPLIMIENTO
 /METHOD = ENTER CAL_CR_COOP_REC CAL_CR_REC CUOTA SALDO_OPER CAPACIDAD_PAGO
 EDAD MONTO RECIPROCIDAD INGRESOS2 MIEMBROS_FAM
 /SAVE = PRED PGROUP RESID
 /PRINT = GOODFIT CORR SUMMARY
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .

Pruebas omnibus sobre los coeficientes del modelo

	Chi-cuadrado	gl	Sig.
Paso 1			
Paso	367.073	16	.000
Bloque	367.073	16	.000
Modelo	367.073	16	.000

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	7413.480 ^a	.039	.068

a. La estimación ha finalizado en el número de iteración 20 porque se han alcanzado las iteraciones máximas. No se puede encontrar una solución definitiva.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	4.976	8	.760

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1	CAL_CR_COOP_REC			39.117	5	.000	
	CAL_CR_COOP_REC(1)	-.341	.064	28.468	1	.000	.711
	CAL_CR_COOP_REC(2)	.388	.193	4.028	1	.045	1.474
	CAL_CR_COOP_REC(3)	.623	.544	1.312	1	.252	1.865
	CAL_CR_COOP_REC(4)	21.706	40192.969	.000	1	1.000	3E+009
	CAL_CR_COOP_REC(5)	.575	.700	.675	1	.411	1.778
	CAL_CR_REC			9.441	3	.024	
	CAL_CR_REC(1)	-.141	.065	4.695	1	.030	.868
	CAL_CR_REC(2)	.647	.345	3.519	1	.061	1.910
	CAL_CR_REC(3)	-.152	.392	.151	1	.698	.859
	CUOTA	-.004	.001	7.496	1	.006	.996
	SALDO_OPER	.000	.000	51.667	1	.000	1.000
	CAPACIDAD_PAGO	.012	.014	.706	1	.401	1.012
	EDAD	-.021	.003	65.341	1	.000	.979
	MONTO	.000	.000	23.971	1	.000	1.000
	RECIPROCIDAD	-.002	.000	28.978	1	.000	.998
	INGRESOS2	.000	.000	7.435	1	.006	1.000
	MIEMBROS_FAM	-.132	.024	30.122	1	.000	.876
	Constante	-.501	.133	14.235	1	.000	.606

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC, CAL_CR_REC, CUOTA, SALDO_OPER, CAPACIDAD_PAGO, EDAD, MONTO, RECIPROCIDAD, INGRESOS2, MIEMBROS_FAM.

Matriz de correlaciones

	Constant	CAL_CR_COOP_REC(1)	CAL_CR_COOP_REC(2)	CAL_CR_COOP_REC(3)	CAL_CR_COOP_REC(4)	CAL_CR_COOP_REC(5)	CAL_CR_REC(1)	CAL_CR_REC(2)	CAL_CR_REC(3)	CUOTA	SALDO_OPER	CAPACIDAD_PAGO	EDAD	MONTO	RECIPROCIDAD	INGRESOS2	MIEMBROS_FAM	
Paso 1	Constant	1.000	-.056	-.012	.007	.000	.000	-.137	-.036	-.032	-.505	-.079	-.478	-.565	.365	.021	.333	-.209
	CAL_CR_COOP_REC(1)	-.056	1.000	-.145	.046	.000	.039	-.094	-.023	.022	-.067	-.001	.056	-.107	.062	-.113	.005	-.037
	CAL_CR_COOP_REC(2)	-.012	.145	1.000	.018	.000	.016	-.045	-.022	-.005	-.017	-.023	.038	-.026	.021	-.002	-.030	-.037
	CAL_CR_COOP_REC(3)	.007	.046	.018	1.000	.000	.038	-.005	.006	-.254	.047	-.052	-.017	-.011	-.086	.011	.025	-.036
	CAL_CR_COOP_REC(4)	.000	.000	.000	.000	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	CAL_CR_COOP_REC(5)	.000	.039	.016	.038	.000	1.000	-.013	.002	-.133	-.016	-.012	.018	-.006	.016	.012	-.018	.000
	CAL_CR_REC(1)	-.137	-.094	-.045	-.005	.000	-.013	1.000	.129	.110	-.049	-.054	-.086	-.021	.001	.031	-.023	-.062
	CAL_CR_REC(2)	-.036	-.023	-.022	.006	.000	.002	.129	1.000	.018	.022	.027	.021	-.012	-.025	.016	-.055	-.011
	CAL_CR_REC(3)	-.032	.002	-.005	-.254	.000	-.133	.110	.018	1.000	-.007	-.044	.003	-.014	.012	.019	-.006	-.003
	CUOTA	-.505	-.067	-.017	.047	.000	-.016	-.049	.022	-.007	1.000	-.164	.440	.004	-.896	-.051	-.420	-.021
	SALDO_OPER	-.079	-.001	-.023	.052	.000	-.012	-.054	.027	-.044	.164	1.000	-.007	.071	-.333	-.002	.002	.030
	CAPACIDAD_PAGO	-.478	.058	.038	-.017	.000	.018	-.086	.021	.003	.440	-.007	1.000	-.020	-.137	-.020	-.729	-.087
	EDAD	-.565	-.107	-.026	-.011	.000	-.006	-.021	-.012	-.014	.004	.071	-.020	1.000	-.018	-.059	-.072	-.183
	MONTO	.365	.062	.021	-.086	.000	.016	.001	-.025	.012	-.896	-.333	-.137	-.018	1.000	.026	.106	-.024
	RECIPROCIDAD	.021	-.113	-.002	.011	.000	.012	.031	.016	.019	-.051	-.002	-.020	-.059	.026	1.000	.027	-.019
	INGRESOS2	.333	.005	-.030	.025	.000	-.018	-.023	-.055	-.006	-.420	.002	-.729	-.072	.106	.027	1.000	.057
	MIEMBROS_FAM	-.209	-.037	-.037	-.036	.000	.000	-.062	-.011	-.003	-.021	.030	-.087	-.183	-.024	-.019	.057	1.000

7. LOGISTIC REGRESSION VARIABLES INCUMPLIMIENTO
 /METHOD = ENTER CAL_CR_COOP_REC3 CAL_CR_REC3 CUOTA SALDO_OPER CAPACIDAD_PAGO
 EDAD MONTO RECIPROCIDAD INGRESOS2 MIEMBROS_FAM
 /SAVE = PRED PGROUP RESID
 /PRINT = GOODFIT CORR SUMMARY
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .

Codificaciones de variables categóricas

	Frecuencia	Codificación de parámetros	
		(1)	(2)
CAL_CR_REC3	3127	1.000	.000
0	103	.000	1.000
1	6065	.000	.000
CAL_CR_COOP_REC3	4663	1.000	.000
0	207	.000	1.000
1	4425	.000	.000

a. Esta codificación da como resultado coeficientes de indicador.

Pruebas omnibus sobre los coeficientes del modelo

	Chi-cuadrado	gl	Sig.
Paso 1	362.728	12	.000
Bloque	362.728	12	.000
Modelo	362.728	12	.000

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	7417.824 ^a	.038	.067

a. La estimación ha finalizado en el número de iteración 6 porque las estimaciones de los parámetros han cambiado en menos de .001.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	4.936	8	.764

Tabla de clasificación^a

Observado	Pronosticado	INCUMPLIMIENTO		Porcentaje correcto
		.00	1.00	
		Paso 1	INCUMPLIMIENTO	
		7914	9	.4
		1366	6	85.2
Porcentaje global				

a. El valor de corte es .500

Variables en la ecuación

Paso		B	E.T.	Wald	gl	Sig.	Exp(B)
1	CAL_CR_COOP_REC3			38.733	2	.000	
	CAL_CR_COOP_REC3(1)	.339	.064	28.191	1	.000	1.404
	CAL_CR_COOP_REC3(2)	.758	.178	18.111	1	.000	2.133
	CAL_CR_REC3			7.416	2	.025	
	CAL_CR_REC3(1)	.143	.065	4.797	1	.029	1.153
	CAL_CR_REC3(2)	.443	.247	3.216	1	.073	1.558
	CUOTA	-.004	.001	7.767	1	.005	.996
	SALDO_OPER	.000	.000	52.892	1	.000	1.000
	CAPACIDAD_PAGO	.011	.014	.632	1	.427	1.011
	EDAD	-.021	.003	65.537	1	.000	.979
	MONTO	.000	.000	24.584	1	.000	1.000
	RECIPROCIDAD	-.002	.000	28.962	1	.000	.998
	INGRESOS2	.000	.000	8.221	1	.004	1.000
	MIEMBROS_FAM	-.132	.024	30.162	1	.000	.876
	Constante	-.982	.148	44.130	1	.000	.375

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC3, CAL_CR_REC3, CUOTA, SALDO_OPER, CAPACIDAD_PAGO, EDAD, MONTO, RECIPROCIDAD, INGRESOS2, MIEMBROS_FAM.

Matriz de correlaciones

	Constant	CAL_CR_COOP_REC3(1)	CAL_CR_COOP_REC3(2)	CAL_CR_REC3(1)	CAL_CR_REC3(2)	CUOTA	SALDO_OPER	CAPACIDAD_PAGO	EDAD	MONTO	RECIPROCIDAD	INGRESOS2	MIEMBROS_FAM	
Paso 1	Constant	1.000	-.340	-.084	-.277	-.043	-.506	-.096	-.444	-.562	.357	-.017	.292	-.231
	CAL_CR_COOP_REC3(1)	-.340	1.000	.203	-.094	-.012	.066	.000	-.059	.107	-.061	.113	-.003	.037
	CAL_CR_COOP_REC3(2)	-.084	.203	1.000	.012	-.096	.020	-.013	.014	.011	-.025	.045	-.025	-.031
	CAL_CR_REC3(1)	-.277	-.094	.012	1.000	.082	.050	.054	.086	.021	-.002	-.031	.023	.062
	CAL_CR_REC3(2)	-.043	-.012	-.096	.082	1.000	.030	.001	.040	-.014	-.019	.022	-.046	.004
	CUOTA	-.506	.066	.020	.050	.030	1.000	.162	.443	.003	-.896	-.052	-.424	-.019
	SALDO_OPER	-.096	.000	-.013	.054	.001	.162	1.000	-.004	.071	-.329	-.002	-.001	.032
	CAPACIDAD_PAGO	-.444	-.059	.014	.086	.040	.443	-.004	1.000	-.020	-.142	-.020	-.729	-.087
	EDAD	-.562	.107	.011	.021	-.014	.003	.071	-.020	1.000	-.018	-.059	-.071	-.183
	MONTO	.357	-.061	-.025	-.002	-.019	-.896	-.329	-.142	-.018	1.000	.027	.112	-.026
	RECIPROCIDAD	-.017	.113	.045	-.031	.022	-.052	-.002	-.020	-.059	.027	1.000	.027	-.019
	INGRESOS2	.292	-.003	-.025	.023	-.046	-.424	-.001	-.729	-.071	.112	.027	1.000	.058
	MIEMBROS_FAM	-.231	.037	-.031	.062	.004	-.019	.032	-.087	-.183	-.026	-.019	.058	1.000

**ANEXO 4:
TEST DEL MODELO CREDIT SCORE - BACKTESTING**

1.- Test del modelo logístico

Matriz de correlaciones

	Constant	CAL_CR_COOP_REC3(1)	CAL_CR_COOP_REC3(2)	CAL_CR_REC3(1)	CAL_CR_REC3(2)	CUOTA	SALDO_OPER	CAPACIDAD_PAGO	EDAD	MONTO	RECIPROCIDAD	INGRESOS2	MIEMBROS_FAM	
Paso 1	Constant	1.000	-.340	-.084	-.277	-.043	-.506	-.096	-.444	-.562	.357	-.017	.292	-.231
	CAL_CR_COOP_REC3(1)	-.340	1.000	.203	-.094	-.012	.066	.000	-.059	.107	-.061	.113	-.003	.037
	CAL_CR_COOP_REC3(2)	-.084	.203	1.000	.012	-.096	.020	-.013	.014	.011	-.025	.045	-.025	-.031
	CAL_CR_REC3(1)	-.277	-.094	.012	1.000	.082	.050	.054	.086	.021	-.002	-.031	.023	.062
	CAL_CR_REC3(2)	-.043	-.012	-.096	.082	1.000	.030	.001	.040	-.014	-.019	.022	-.046	.004
	CUOTA	-.506	.066	.020	.050	.030	1.000	.162	.443	.003	-.896	-.052	-.424	-.019
	SALDO_OPER	-.096	.000	-.013	.054	.001	.162	1.000	-.004	.071	-.329	-.002	-.001	.032
	CAPACIDAD_PAGO	-.444	-.059	.014	.086	.040	.443	-.004	1.000	-.020	-.142	-.020	-.729	-.087
	EDAD	-.562	.107	.011	.021	-.014	.003	.071	-.020	1.000	-.018	-.059	-.071	-.183
	MONTO	.357	-.061	-.025	-.002	-.019	-.896	-.329	-.142	-.018	1.000	.027	.112	-.026
	RECIPROCIDAD	-.017	.113	.045	-.031	.022	-.052	-.002	-.020	-.059	.027	1.000	.027	-.019
	INGRESOS2	.292	-.003	-.025	.023	-.046	-.424	-.001	-.729	-.071	.112	.027	1.000	.058
	MIEMBROS_FAM	-.231	.037	-.031	.062	.004	-.019	.032	-.087	-.183	-.026	-.019	.058	1.000

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)	
Paso 1	CAL_CR_COOP_REC3		38.733	2	.000		
	CAL_CR_COOP_REC3(1)	.339	.064	28.191	1	.000	1.404
	CAL_CR_COOP_REC3(2)	.758	.178	18.111	1	.000	2.133
	CAL_CR_REC3		7.416	2	.025		
	CAL_CR_REC3(1)	.143	.065	4.797	1	.029	1.153
	CAL_CR_REC3(2)	.443	.247	3.216	1	.073	1.558
	CUOTA	-.004	.001	7.767	1	.005	.996
	SALDO_OPER	.000	.000	52.892	1	.000	1.000
	CAPACIDAD_PAGO	.011	.014	.632	1	.427	1.011
	EDAD	-.021	.003	65.537	1	.000	.979
	MONTO	.000	.000	24.584	1	.000	1.000
	RECIPROCIDAD	-.002	.000	28.962	1	.000	.998
	INGRESOS2	.000	.000	8.221	1	.004	1.000
	MIEMBROS_FAM	-.132	.024	30.162	1	.000	.876
	Constante	-.982	.148	44.130	1	.000	.375

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC3, CAL_CR_REC3, CUOTA, SALDO_OPER, CAPACIDAD_PAGO, EDAD, MONTO, RECIPROCIDAD, INGRESOS2, MIEMBROS_FAM.

Tabla de clasificación^a

Observado	Pronosticado	INCUMPLIMIENTO		Porcentaje correcto
		.00	1.00	
Paso 1 INCUMPLIMIENTO .00		7914	9	99.9
1.00		1366	6	.4
Porcentaje global				85.2

a. El valor de corte es .500

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	4.936	8	.764

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	7417.824 ^a	.038	.067

a. La estimación ha finalizado en el número de iteración 6 porque las estimaciones de los parámetros han cambiado en menos de .001.

Historial de iteraciones^{b,c,d}

Iteración	-2 log de la verosimilitud	Coeficientes												
		Constant	CAL_CR_COOP_REC3(1)	CAL_CR_COOP_REC3(2)	CAL_CR_REC3(1)	CAL_CR_REC3(2)	CUOTA	SALDO_OPER	CAPACIDAD_PAGO	EDAD	MONTO	RECIPROCIDAD	INGRESOS2	MIEMBROS_FAM
Paso 1	7716.046	-1.132	.187	.434	.074	.295	.000	.000	.013	-.010	.000	-.000	.000	-.071
2	7458.750	-1.208	.330	.722	.127	.454	-.001	.000	.022	-.019	.000	-.001	.000	-.121
3	7427.189	-1.108	.361	.786	.142	.468	-.002	.000	.019	-.022	.000	-.001	.000	-.133
4	7418.418	-1.003	.345	.764	.142	.447	-.004	.000	.012	-.022	.000	-.002	.000	-.132
5	7417.828	-.982	.340	.758	.142	.444	-.004	.000	.011	-.021	.000	-.002	.000	-.132
6	7417.824	-.982	.339	.758	.143	.443	-.004	.000	.011	-.021	.000	-.002	.000	-.132

a. Método: Introducir

b. En el modelo se incluye una constante.

c. -2 log de la verosimilitud inicial: 7780.553

d. La estimación ha finalizado en el número de iteración 6 porque las estimaciones de los parámetros han cambiado en menos de .001.

Pruebas omnibus sobre los coeficientes del modelo

Paso 1		Chi-cuadrado	gl	Sig.
Paso		362.728	12	.000
Bloque		362.728	12	.000
Modelo		362.728	12	.000

La estimación del modelo logístico viene dado por la siguiente función:

$$Z_j = -0.981981 + 0.757756 * \text{CAL_CR_COOP_REC3}(B,C,D,E) + 0.443445 * \text{CAL_CR_REC3}(B,C) - 0.004122 * \text{CUOTA} - 0.000218 * \text{SALDO_OPER} + 0.010969 * \text{CAPACIDAD_PAGO} - 0.021469 * \text{EDAD} + 0.000292 * \text{MONTO} - 0.002177 * \text{RECIPROCIDAD} + 0.000374 * \text{INGRESOS2} - 0.132056 * \text{MIEMBROS_FAM}$$

Donde Z_j representa la puntuación del sujeto de crédito, luego la probabilidad de que este sujeto de crédito cumpla sus obligaciones con la Cooperativa está definido por P_j así:

$$P_j = \frac{e^{Z_j}}{1 + e^{Z_j}}; z_j; \text{sabiendo que } Z_j = B_0 + B_1 Z_1 + B_2 Z_2 + \dots + B_p Z_p$$

Al calcular la probabilidad de P_j , si el valor de la probabilidad sea más cercano a cero la posibilidad de impago del sujeto de crédito es menor que si el valor de la probabilidad sea más cercano a uno. Es decir existe mayor riesgo de crédito del sujeto de crédito donde P_j es más cercano a uno.

2.- PRUEBA DE BONDAD DE AJUSTE DE KOLMOGOROV SMIRNOV.

Hipótesis a contrastar

H_0 : Los datos analizados siguen una distribución M.

H_1 : Los datos analizados no siguen una distribución M.

Estadístico de contraste:

$$D = \sup_{1 \leq i \leq n} \left| \hat{F}_n(X_i) - F_0(X_i) \right|$$

Así pues, D es la mayor diferencia absoluta observada entre la frecuencia acumulada observada $\hat{F}_n(X)$ y la frecuencia acumulada teórica $F_0(X)$, obtenida a partir de la distribución de probabilidad que se especifica como hipótesis nula.

Por tanto, el criterio para la toma de la decisión entre las dos hipótesis será de la forma:

Si $D \leq D_\alpha \Rightarrow$ Aceptar H_0

Si $D > D_\alpha \Rightarrow$ Rechazar H_0

Donde el valor D_α se elige de tal manera que:

$$P(\text{Rechazar } H_0 / H_0 \text{ es cierta}) = P(D > D_\alpha / \text{Los datos siguen la distribución M}) = \alpha$$

Siendo α el nivel de significación del contraste. A su vez, el valor de D_α depende del tipo de distribución a probar y se encuentra tabulado. En general es de la forma:

$$D_\alpha = \frac{c_\alpha}{k(n)}$$

Donde c_α y $k(n)$ se encuentran en las siguientes tablas:

c_α	α		
	0.1	0.05	0.01
Modelo			
General	1.224	1.358	1.628
Normal	0.819	0.895	1.035
Exponencial	0.990	1.094	1.308
Weibull n=10	0.760	0.819	0.944
Weibull n=20	0.779	0.843	0.973
Weibull n=50	0.790	0.856	0.988
Weibull n=∞	0.803	0.874	1.007

DISTRIBUCIÓN QUE SE CONTRASTA	k(n)
General. Parámetros desconocidos.	$k(n) = \sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}$
Normal	$k(n) = \sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}}$
Exponencial	$k(n) = \sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}$
Weibull	$k(n) = \sqrt{n}$

3.- BACKTESTING DEL MODELO

- LOGISTIC REGRESSION VARIABLES INCUMPLIMIENTO
 /METHOD = ENTER CAL_CR_COOP_REC3 CAL_CR_REC3 CUOTA SALDO_OPER CAPACIDAD_PAGO
 EDAD MONTO RECIPROCIDAD INGRESOS2 MIEMBROS_FAM
 /SAVE = PRED PGROUP RESID
 /PRINT = GOODFIT CORR SUMMARY
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .

Pruebas omnibus sobre los coeficientes del modelo

Paso		Chi-cuadrado	gl	Sig.
Paso 1	Paso	88.331	12	.000
	Bloque	88.331	12	.000
	Modelo	88.331	12	.000

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	2043.420 ^a	.023	.054

a. La estimación ha finalizado en el número de iteración 8 porque las estimaciones de los parámetros han cambiado en menos de .001.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	11.307	8	.185

Tabla de contingencias para la prueba de Hosmer y Lemeshow

		INCUMPLIMIENTO = .00		INCUMPLIMIENTO = 1.00		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	372	372.881	7	6.119	379
	2	365	361.392	14	17.608	379
	3	361	357.575	18	21.425	379
	4	359	354.445	20	24.555	379
	5	349	351.538	30	27.462	379
	6	352	348.043	27	30.957	379
	7	340	344.419	39	34.581	379
	8	325	340.386	54	38.614	379
	9	337	334.277	42	44.723	379
	10	324	319.046	56	60.954	380

Tabla de clasificación^a

Observado	INCUMPLIMIENTO		Pronosticado		Porcentaje correcto
			INCUMPLIMIENTO		
			.00	1.00	
Paso 1	INCUMPLIMIENTO	.00	3483	1	100.0
		1.00	306	1	.3
	Porcentaje global				91.9

a. El valor de corte es .500

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1	CAL_CR_COOP_REC3			2.040	2	.036	
	CAL_CR_COOP_REC3(1)	.051	.129	.158	1	.069	1.053
	CAL_CR_COOP_REC3(2)	.462	.325	2.020	1	.016	1.587
	CAL_CR_REC3			8.447	2	.015	
	CAL_CR_REC3(1)	.380	.133	8.163	1	.004	1.463
	CAL_CR_REC3(2)	.398	.543	.537	1	.046	1.489
	CUOTA	-.001	.002	.220	1	.064	.999
	SALDO_OPER	.000	.000	19.708	1	.000	1.000
	CAPACIDAD_PAGO	.024	.025	.917	1	.338	1.024
	EDAD	-.013	.005	6.866	1	.009	.987
	MONTO	.000	.000	8.389	1	.004	1.000
	RECIPROCIDAD	-.005	.001	14.036	1	.000	.995
	INGRESOS2	.000	.000	.186	1	.067	1.000
	MIEMBROS_FAM	-.118	.049	5.796	1	.016	.889
	Constante	-1.874	.261	51.537	1	.000	.153

a. Variable(s) introducida(s) en el paso 1: CAL_CR_COOP_REC3, CAL_CR_REC3, CUOTA, SALDO_OPER, CAPACIDAD_PAGO, EDAD, MONTO, RECIPROCIDAD, INGRESOS2, MIEMBROS_FAM.

4.- CATEGORIZACION Y PUNTOS DE CORTE

1. DESCRIPTIVES

VARIABLES=PRE_8

/STATISTICS=MEAN SUM STDDEV VARIANCE RANGE MIN MAX KURTOSIS SKEWNESS .

Estadísticos descriptivos

	N	Rango	Mínimo	Máximo	Suma	Media	Desv. tip.	Varianza	Asimetría		Curtosis	
	Estadístico	Error típico	Estadístico	Error típico								
Probabilidad pronosticada	9295	.78810	.00000	.78810	1372.00148	.1476064	.07013370	.005	1.100	.025	4.181	.051
N válido (según lista)	9295											

2. FRECUENCIAS

VARIABLES=PRE_8

/NTILES= 4

/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM SEMEAN MEAN MEDIAN MODE

SUM SKEWNESS SESKEW KURTOSIS SEKURT

/HISTOGRAM NORMAL

/ORDER= ANALYSIS .

Estadísticos

Probabilidad pronosticada		
N	Válidos	9295
	Perdidos	0
Media		.1476064
Error típ. de la media		.00072745
Mediana		.1379599
Moda		.00013 ^a
Desv. tip.		.07013370
Varianza		.005
Asimetría		1.100
Error típ. de asimetría		.025
Curtosis		4.181
Error típ. de curtosis		.051
Rango		.78810
Mínimo		.00000
Máximo		.78810
Suma		1372.001
Percentiles	25	.0977676
	50	.1379599
	75	.1892944

a. Existen varias modas. Se mostrará el menor de los valores.

Histograma

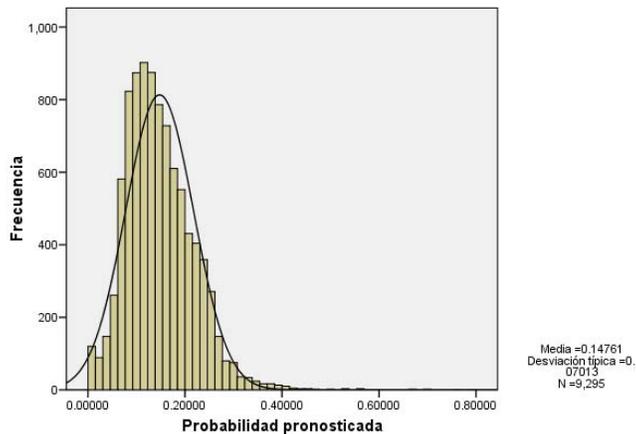


Tabla de contingencia RANGO_SCORE * INCUMPLIMIENTO

			INCUMPLIMIENTO		Total
			1.00	.00	
RANGO_SCORE	A	Recuento	491	4054	4545
		% de RANGO_SCORE	10.8%	89.2%	100.0%
		% de INCUMPLIMIENTO	12.9%	70.0%	48.8%
	B	Recuento	1959	1629	3588
		% de RANGO_SCORE	54.6%	45.4%	100.0%
		% de INCUMPLIMIENTO	56.9%	27.8%	38.5%
	C	Recuento	1042	120	1162
		% de RANGO_SCORE	89.7%	10.3%	100.0%
		% de INCUMPLIMIENTO	30.2%	2.2%	12.6%
Total	Recuento	3492	5803	9295	
	% de RANGO_SCORE	36.9%	63.0%	100.0%	
	% de INCUMPLIMIENTO	100.0%	100.0%	100.0%	

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	14.384 ^a	2	.001
Razón de verosimilitudes	10.948	2	.004
N de casos válidos	9295		

a. 2 casillas (33.3%) tienen una frecuencia esperada inferior a 5.
La frecuencia mínima esperada es 2.07.