

# ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

CARRERA DE INGENIERIA MATEMÁTICA

SELECCIÓN DE PERFILES DE CLIENTES MEDIANTE REGRESIÓN  
LOGÍSTICA PARA MUESTRAS DESPROPORCIONADAS, VALIDACIÓN,  
MONITOREO Y APLICACIÓN EN LA PROYECCIÓN DE PROVISIONES

PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO  
MATEMÁTICO

CARLOS ALEJANDRO IÑIGUEZ SALAS

([ciniguez@pichincha.com](mailto:ciniguez@pichincha.com))

MARÍA GABRIELA MORALES ARIAS

([gmorales@cth.fin.ec](mailto:gmorales@cth.fin.ec))

DIRECTOR: Ec. CARLOS BAMBINO CONTRERAS

([cbambino@bgr.com.ec](mailto:cbambino@bgr.com.ec))

2009

## **DECLARACIÓN**

Nosotros, Carlos Alejandro Iñiguez Salas y María Gabriela Morales Arias, declaramos que el trabajo aquí descrito es de nuestra autoría, que no ha sido previamente presentado para ningún grado o calificación profesional y que se han consultado en las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional, puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley, Reglamento de Propiedad Intelectual y por la normativa institucional vigente.

---

Carlos Alejandro Iñiguez Salas

---

María Gabriela Morales Arias

## **CERTIFICACIÓN**

Certifico que el presente trabajo fue desarrollado por Carlos Alejandro Iñiguez Salas y María Gabriela Morales Arias bajo mi supervisión.

---

Ec. Carlos Bambino Contreras  
DIRECTOR DEL PROYECTO

# ÍNDICE

ÍNDICE .....	iv
RESUMEN.....	vii
ABSTRACT .....	vii
CAPITULO I.....	1
INTRODUCCIÓN A LOS ACUERDOS INTERNACIONALES DE BASILEA Y A LA NORMATIVA VIGENTE EN EL SISTEMA FINANCIERO ECUATORIANO. 1	
1.1. ADMINISTRACION INTEGRAL DEL RIESGO .....	1
1.2. CLASIFICACION DEL RIESGO .....	3
1.2.1. RIESGO DE MERCADO: .....	3
1.2.1.1. VaR Empírico .....	5
1.2.1.2. VaR Paramétrico.....	5
1.2.1.3. VaR por Simulación .....	5
1.2.2. RIESGO DE LIQUIDEZ:.....	5
1.2.3. RIESGO OPERATIVO: .....	7
1.2.4. RIESGO DE CRÉDITO:.....	9
1.3. ACUERDO DE BASILEA.....	11
1.3.1. PRIMER PILAR.....	12
1.3.2. SEGUNDO PILAR .....	23
1.3.3. TERCER PILAR .....	24
1.4. CREDIT SCORING .....	26
1.5. MODELOS SCORING EN EL CICLO DE CREDITO .....	27
1.5.1. PREVENTA .....	27
1.5.2. EVALUACION .....	28
1.5.3. SEGUIMIENTO .....	30
1.5.4. COBRANZA .....	31
CAPITULO II .....	34
SELECCIÓN DE PERFILES DE CLIENTES MEDIANTE REGRESIÓN LOGÍSTICA PARA MUESTRAS DESPROPORCIONADAS .....	34
2.1. ETAPAS DEL SCORING DE EVALUACION .....	36
2.2. SELECCIÓN DE LA VENTANA DE MUESTREO .....	39
2.2.1. CONSTRUCCIÓN DEL INDICADOR DE MORA.....	40
2.3. DEFINICION DE LA VARIABLE DEPENDIENTE DEL MODELO .....	46
2.4. ANALISIS DESCRIPTIVO DE LA BASE DE DATOS .....	49
2.4.1. ANÁLISIS DESCRIPTIVO .....	50
2.5. SELECCIÓN DE LAS VARIABLES EXPLICATIVAS DEL MODELO .	56
2.5.1. ANÁLISIS DE CORRELACIONES.....	56
2.6. MODELIZACION .....	65
2.6.1. CONSIDERACIONES PRELIMINARES.....	66
2.6.2. EL PROBLEMA DE MUESTRAS DESPROPORCIONADAS .....	69
2.6.3. ESTIMACIÓN DEL MODELO.....	73
2.6.4. ETAPA DE VALIDACIÓN.....	73
2.6.5. PRESENTACIÓN DE RESULTADOS.....	76

2.6.6.	PRUEBAS DE EFICIENCIA.....	82
2.7.	VALIDACIÓN DEL MODELO CON DATOS DE PRUEBA (CONTROL DEL MODELO).....	85
2.8.	CONSTRUCCION DE PERFILES DE CLIENTES .....	86
2.8.1.	DETERMINACIÓN DEL PUNTO DE CORTE (CUTOFF) UTILIZANDO LA CURVA ROC.....	86
2.8.2.	DETERMINACIÓN DEL PUNTO DE CORTE UTILIZANDO MATRICES DE CONFUSIÓN.....	91
2.8.3.	CONSTRUCCIÓN DE PERFILES UTILIZANDO LA PÉRDIDA COMO INDICADOR DE CORTE .....	92
	CAPITULO III.....	95
3.1.	INTRODUCCIÓN.....	95
3.2.	MAPA ESTRATEGICO DE MONITOREO .....	98
3.2.1.	MONITOREO DE LOS RESULTADOS DEL MODELO SCORING:.....	101
3.2.1.1.	Resultados por Perfil: .....	101
3.2.1.2.	Tendencia de la Razón de Rechazo (Reporte R2):.....	103
3.2.1.3.	Eficiencia (Reporte R3):.....	105
3.2.2.	MONITOREO ESTRUCTURAL DEL MODELO SCORING:.....	106
3.2.2.1.	Correlación de las variables:.....	106
3.2.2.2.	Estabilidad de las Distribuciones:.....	110
3.2.2.3.	Delta Approach (Enfoque Delta) (Análisis A3):.....	114
3.3.	IMPLEMENTACION .....	121
3.3.1.	PANTALLA PRINCIPAL .....	123
3.3.2.	SALIDA DE RESULTADOS .....	125
3.3.2.1.	Reportes de los resultados del Modelo.....	126
3.3.2.2.	Reportes Estructurales del Modelo.....	128
	CAPITULO IV.....	134
	PROYECCION DE LA VARIACION MENSUAL DE PROVISIONES POR PERFIL DE CLIENTE .....	134
4.1.	INTRODUCCIÓN.....	134
4.2.	PÉRDIDA ESPERADA .....	137
4.2.1.	PROBABILIDAD DE INCUMPLIMIENTO (default) .....	139
4.2.1.1.	Matrices de transición.....	141
4.2.2.	EXPOSICIÓN .....	147
4.2.3.	SEVERIDAD DE LA PÉRDIDA .....	148
4.3.	AFECTACION DE LAS PROVISIONES EN EL ESTADO DE RESULTADOS 150	
4.4.	MODELO DE ESTIMACION DE LAS PROVISIONES GENERADAS MENSUALMENTE .....	151
4.4.1.	DESCRIPCION DE LA ESTRUCTURA DEL MODELO .....	152
4.4.2.	CORRIDA DE REGRESIONES LINEALES.....	154
4.4.3.	MEDICIÓN DEL ERROR EN LA ESTIMACIÓN.....	156
4.4.4.	PROYECCIÓN DE LA PROVISIÓN GENERADA.....	157
4.4.4.1.	Pérdida por cosecha .....	157
4.4.4.2.	Proyección de la provisión generada.....	158
	CAPITULO V .....	161
	CONCLUSIONES Y RECOMENDACIONES.....	161

ANEXO I .....	169
TEORIA .....	169
ÁRBOLES DE DECISIÓN .....	169
PRUEBA DE INDEPENDENCIA CHI CUADRADO .....	172
REGRESIÓN LOGÍSTICA.....	173
CURVA ROC.....	183
INDICADOR DE INTERCORRELACIÓN KMO.....	184
INDICE DE ESTABILIDAD DE LA POBLACIÓN PSI .....	185
CADENAS DE MARKOV .....	186
ANEXO II .....	192
TABLAS Y GRÁFICOS.....	192
ÁRBOLES DE DECISIÓN .....	192
ANÁLISIS DESCRIPTIVO DE VARIABLES EXPLICATIVAS .....	201
FUNCIONES LINEALES DE PARA LOS MESES DE ANTIGÜEDAD DEL 2 AL 11 .....	210
ANEXO III.....	214
CÓDIGO FUENTE PANEL DE MONITOREO .....	214
REFERENCIAS BIBLIOGRÁFICAS.....	265

## **RESUMEN**

Los modelos de medición de riesgo se han convertido en piezas fundamentales en la administración de las instituciones financieras. El presente trabajo describe en detalle la metodología para la realización de un modelo scoring para la fase de iniciación, como método discriminante se usó la regresión logística para muestras desproporcionadas, posteriormente se identificaron los puntos de corte óptimos mediante dos métodos distintos y adicionalmente se construyeron perfiles de clientes, en base a varios puntos de corte determinados en función de la pérdida probable. El modelo original debe ser constantemente monitoreado durante todo su tiempo de aplicación, con ese objetivo se diseña una metodología para el monitoreo del mismo. Mediante macros de Excel se automatizaron los reportes que constituyen la metodología. Finalmente los perfiles de clientes construidos son aplicados para proyectar las provisiones mensuales de una institución financiera.

## **ABSTRACT**

The measurement risk models have become in basic pieces in the Financial Institutions. The present job describes in detail the methodology to do a scoring model for the beginning stage, as a discriminant method we used the logistic regression for separate samples, next we identified the optimal cutoff points using two different methods, besides, we built the customer's profiles based to several cutoff points, they was defined depending on the probable loss. The original model must be constantly monitored during this entire lifetime; because of that objective we design a methodology to check the model. The reports, which constitute the methodology, were automated, by means of Excel macros. Eventually, the customer's profiles are applied to cast Financial Institution monthly provisions.

## **PRESENTACION**

Durante los últimos años la regulación de las instituciones financieras por parte de entes de control estatales se ha ido incrementando, esto ha permitido la aplicación y desarrollo de modelos matemáticos, estadísticos y financieros que permitan identificar, medir y controlar eventos de riesgo. Los acuerdos de capital de Basilea plantean la administración integral del riesgo al reconocer distintos tipos de riesgos a los que las instituciones financieras se ven expuestas en sus actividades diarias. Uno de los objetivos de los acuerdos de Basilea buscan establecer provisiones por los riesgos de Liquidez, Mercado, Operativo y Crédito de tal manera que cuando ocurran los eventos de riesgo las instituciones se encuentren preparadas para cubrir las pérdidas que posiblemente se generen.

Una buena administración del riesgo permite a las instituciones financieras emitir créditos de una manera segura, eficiente y rápida; además, permite proyectar las pérdidas resultantes del ejercicio financiero. Actualmente la rapidez en la concesión de créditos se ha convertido, para una entidad financiera, en una pieza fundamental de su posicionamiento en el mercado. Sin embargo el objetivo no debe ser simplemente la agilidad sino también la búsqueda de buenos clientes (calidad en la aprobación), esto se consigue con la aplicación de modelos estadísticos.

El crédito sigue un ciclo que tiene las siguientes etapas: iniciación, evaluación, seguimiento y cobranza. Cada una de las etapas tiene propiedades que la caracterizan, estas propiedades develan nueva información al avanzar a lo largo del ciclo. El modelo de regresión logística se realizará con información de la etapa de iniciación, dicha etapa se refiere a la admisión de un individuo como cliente de crédito de un banco.

En este marco se desarrolló el presente estudio, en el Primer Capítulo se realiza una introducción detallada de los términos, metodologías y regulaciones



mencionadas en los párrafos anteriores; así como a los avances de la regulación y la normativa aplicadas en Ecuador.

En capítulo segundo se realiza un modelo de regresión logística aplicado sobre muestras desproporcionadas para identificar a los buenos y malos clientes. Una vez construido el modelo discriminante es necesario encontrar el punto sobre el cual rechazar a los clientes (punto de corte), para lo cual se presentan dos metodologías usando la Curva ROC, primero identificando el punto en el que la distancia de la recta identidad a la curva es máxima y por otro lado el punto en el que la pérdida esperada es mínima. También se hallarán puntos de corte usando matrices de confusión. Dado que el modelo de segmentación se realiza con información netamente externa a la institución, la determinación de un solo punto de corte no es completamente eficiente, por ello se presentará una alternativa para la segmentación de la población, para lo cual se construirán perfiles de clientes, particionando el score y asignando un porcentaje de pérdida a cada perfil así construido.

Luego de la implementación del modelo y en el transcurso del tiempo éste debe ser controlado para verificar su efectividad, por ello es importante establecer políticas de monitoreo de los modelos discriminantes. En el tercer capítulo se diseña una metodología sistematizada para realizar el monitoreo de modelos de evaluación de clientes, construidos en base a regresión logística. La metodología sugerida es una fusión de herramientas estadísticas de medición de la calidad del modelo y su automatización a través de Macros de Excel.

La estimación de provisiones que se tendrán que realizar en un horizonte de tiempo determinado, es necesaria para la presupuestación y planeación estratégica de cuánto y cómo colocar en función de la pérdida que la entidad esté dispuesta a asumir. En el cuarto capítulo se utilizan los perfiles de clientes para determinar el flujo de provisión causada en un horizonte de tiempo, considerando la estrategia de colocación dada por el banco. Adicionalmente se menciona las técnicas usadas para el cálculo de la pérdida esperada sugeridas por los comités de Basilea, se explica claramente su metodología y fundamento teórico.

# **CAPITULO I**

## **INTRODUCCIÓN A LOS ACUERDOS INTERNACIONALES DE BASILEA Y A LA NORMATIVA VIGENTE EN EL SISTEMA FINANCIERO ECUATORIANO.**

### **1.1. ADMINISTRACION INTEGRAL DEL RIESGO**

El riesgo puede ser entendido como la incertidumbre o aleatoriedad en la obtención de un resultado al desarrollar determinada actividad, el concepto de riesgo tiene dos elementos, la probabilidad de que un evento ocurra y las consecuencias de tal evento.

Existe una relación directa entre el grado de riesgo asumido por una institución y el potencial rendimiento a ser generado, en particular, en las actividades financieras el riesgo es ineludible, por ello en este tipo de instituciones se debe mantener un adecuado juicio del riesgo. Se dice que el riesgo es ineludible porque éste es inherente a toda acción emprendida dentro de una entidad dedicada a la intermediación, por ejemplo, al conceder un crédito siempre existe un nivel de riesgo, por tal motivo la institución financiera tendrá que determinar el nivel de riesgo que está dispuesta a asumir.

El proceso de administración integral de riesgos tiene como objetivo el identificar los riesgos a que se encuentra expuesta una institución, medirlos, hacer seguimiento de su impacto en la operación y controlar sus efectos sobre los rendimientos y el valor del capital, mediante la aplicación de estrategias y mecanismos que permitan realizar las operaciones con niveles acordes con su respectivo capital global y capacidad operativa, integrando la cultura de riesgos en la operación diaria.

La administración integral de riesgos debe asegurar la homogeneidad de las herramientas, estructuras organizativas, procesos y sistemas adecuados a la dimensión de la actividad realizada; de tal manera que permita facilitar la gestión global de todos los riesgos que se asuman. Esta homogeneidad se refiere a que todos los puntos mencionados mantengan similitud y se complementen entre sí, de manera que los controles y filtros establecidos en un área sean respetados en otra.

Una adecuada gestión del riesgo tiene un impacto positivo en la rentabilidad de una institución, puesto que se controla la exposición a pérdidas, aunque en un inicio estas medidas provoquen la constitución de provisiones y otros cambios en el Estado de Situación General de la entidad, dichos cambios influyen en los indicadores financieros, principalmente en los de rentabilidad y eficiencia.

La búsqueda de la maximización de utilidades, asume ciertos niveles de riesgo que tienen una relación directa con la productividad y la generación eficiente de utilidades en la institución, así mientras más riesgo implica una operación su rentabilidad debe ser mayor, de forma que la alta exposición al riesgo es compensada. Cada institución debe definir el grado de riesgo que está dispuesta a asumir, si una institución acepta bajos niveles de riesgo se dice que es adversa al riesgo, por otro lado si los niveles de riesgo aceptados son altos la institución es apetitosa al riesgo.

En los últimos años las instituciones financieras han ido incorporando a su estructura organizativa las Unidades de Riesgos, encargadas de la administración integral de riesgos, estas se han convertido en unidades estratégicas en el soporte a la toma de decisiones.

La administración integral de riesgos debe incluir:

- La identificación y valuación de los distintos tipos de riesgo.
- El establecimiento de políticas, procedimientos y límites de riesgo.
- Monitoreo y reporte del cumplimiento de los límites establecidos.
- Delineación del capital asignado y de la administración de la cartera.

- Guías para el desarrollo de nuevos productos y la inclusión de nuevas exposiciones al riesgo dentro de la estructura existente.
- Aplicación de nuevos métodos de medición a los productos existentes.

Una institución no debe aceptar la introducción de un nuevo producto mientras no se tenga un entendimiento profundo del mismo y que éste haya sido integrado a los sistemas de medición y control de riesgo establecidos.

Todas las inversiones en mercados de dinero, capital, cambios y productos derivados son diferentes de institución a institución, éstas variaciones surgen de las diferencias en tamaño, estructura organizacional, sofisticación de los sistemas de cómputo, la estrategia de la institución, las ganancias esperadas, el historial financiero, y la experiencia en cuanto a pérdidas, riesgos y tipos de productos que se están operando. Como resultado de lo anterior las prácticas, políticas y procedimientos de una institución en lo que se refiere a la administración de riesgos no necesariamente aplican en otras.

En el desarrollo de este capítulo se incluirán puntos que por su relevancia en la administración integral del riesgo deben ser tomados en cuenta.

## **1.2. CLASIFICACION DEL RIESGO**

Una extensa clasificación del riesgo podría ser considerada, sin embargo en el enfoque de este estudio se tratará una clasificación que contiene cuatro tipos de riesgos, planteados en los Acuerdos de Capital de Basilea I y II<sup>1</sup> y de mayor impacto en las instituciones financieras, que se definen a continuación:

### **1.2.1. RIESGO DE MERCADO:**

El riesgo de mercado se define como la pérdida potencial por cambios en los factores de riesgo que inciden sobre la valuación o sobre los resultados esperados, como tasas de interés, tipos de cambio, precios de mercado, índices

---

<sup>1</sup> A cerca de los Acuerdos de Capital de Basilea I y II se trata a partir de la página 12 del presente Capítulo

y otros factores de riesgo en los mercados de dinero, cambios, y productos derivados a los que se encuentra expuesto. Su valuación correcta requiere de la oportunidad y calidad de la información sobre el valor de mercado actual de los activos, pasivos y elementos de cuentas de orden de una institución.

El riesgo de mercado puede, por tanto, subdividirse en:

- Riesgo de tipos de interés
- Riesgo de tipo de cambio
- Riesgo de precios bursátiles
- Riesgo de precios de las mercancías

Las mediciones del riesgo de mercado pueden ser globales, relativas a todos estos subriesgos, o específicas, de cada una de estas categorías de riesgo. Sin embargo, estos factores no son independientes entre sí, sino que están relacionados de manera que no resulta posible la fragmentación del riesgo de mercado en cuatro diferentes.

La metodología de Valor en Riesgo (VaR, por su siglas del inglés Value at Risk) se ha consolidado hasta convertirse en el método de medición de riesgos de mercado más comúnmente utilizado, el VaR se suele definir como la pérdida máxima esperada en un horizonte de tiempo dado y con cierto nivel de confianza. El Valor en Riesgo está directamente relacionado con la volatilidad del valor del portafolio, el cual se ve afectado por los cambios en los factores que inciden en el valor de las posiciones que componen el portafolio. Una definición más completa indica que el VaR es un número que representa la caída de valor de la cartera, correspondiente a un percentil determinado de la variable aleatoria rentabilidad futura de dicha cartera, en un horizonte temporal determinado.

En todo caso se trata de estimar el valor del percentil del VaR del rendimiento futuro de una cartera. Para lo cual se distinguen tres métodos:

### **1.2.1.1. VaR Empírico**

Es el percentil estimado a partir de la distribución empírica de rendimientos de la cartera sin hacer suposición alguna sobre el proceso seguido por el valor de la cartera a lo largo del tiempo.

### **1.2.1.2. VaR Paramétrico**

Es el percentil estimado a partir de asumir que la distribución de los rendimientos futuros de la cartera en el horizonte temporal considerado pertenece a una familia paramétrica determinada.

### **1.2.1.3. VaR por Simulación**

Es el percentil estimado a partir de la distribución de rendimientos obtenidos a partir de la simulación de los factores de riesgo que influyen en el valor de la cartera, bien mediante Montecarlo o mediante valores históricos.

Para complementar la metodología de riesgo de mercado se utilizan pruebas de sensibilidad, simulando variaciones en los factores de riesgo que afectan el valor de las posiciones. A su vez, se llevan a cabo pruebas retrospectivas (back testing) para verificar la validez del modelo, comparando los resultados que genera el modelo, contra los resultados efectivamente observados. Adicionalmente el modelo debe ser sometido a supuestos extremos ("stress testing"), es decir, seguir un proceso para identificar y gestionar situaciones que podrían causar pérdidas extraordinarias.

## **1.2.2. RIESGO DE LIQUIDEZ:**

El riesgo de liquidez se define como la pérdida potencial por la imposibilidad o dificultad de renovar pasivos en condiciones normales para la institución, o por la venta anticipada o forzosa de activos a descuentos inusuales.

Los riesgos de liquidez de una entidad financiera se derivan de desfases en los flujos de las operaciones de captación, crédito y negociación como son:

- Pasivos a la vista

- Vencimientos de depósitos a plazo
- Disposición de líneas de crédito
- Liquidación de operaciones con fines de negociación y con instrumentos derivados.
- Gastos operativos.

En la medida en que la institución tenga la capacidad de obtener recursos de fuentes de fondeo alternas que tengan un costo aceptable, el riesgo de liquidez se reduce.

Entre los elementos que intervienen en la estrategia aplicada en la gestión de la liquidez están:

- Diferenciar el tratamiento de Activos, Pasivos e ítems fuera de balance.
- Controlar las brechas de vencimientos de activos y pasivos,
- Diversificar las fuentes de captación de fondos,
- Diversificar los vencimientos de activos y pasivos,
- Establecer límites prudentes y garantizar el acceso inmediato a los activos líquidos.

La medida más utilizada para el control de este riesgo es el análisis de brechas de liquidez (gap de liquidez).

Este gap de liquidez proporciona información sobre los desfases entre las entradas y salidas de fondos de forma diaria y por periodos de hasta 12 meses de todos los ítems de balance y fuera de balance que producen flujos de caja en las fechas reales en las que se producen.

De forma diaria se realiza un seguimiento de la liquidez a corto plazo. Las brechas se calculan en intervalos de tiempo definidos, por ejemplo intervalos de 1 a 7 días, de 8 a 15 días, de 15 a 30 días, de 30 a 60, de 60 a 90, de 90 a 180, de 180 a 360 y más de 360 días.

En Ecuador, las instituciones financieras reportan a la Superintendencia de Bancos y Seguros (SBS) mensualmente el análisis de Brechas de liquidez en los intervalos de tiempo mencionados en el ejemplo anterior, los reportes se realizan en tres escenarios distintos; el contractual, que considera los vencimientos, en las entradas y salidas, de flujos tal cual fueron pactados, es decir, sin considerar ningún evento que difiera los vencimientos. En el escenario esperado, se incluyen supuestos para tomar en cuenta los eventos que difieren los vencimientos, por ejemplo la morosidad de la cartera de créditos o la renovación de una parte de las obligaciones financieras. El último escenario es el dinámico, que considera los movimientos que posiblemente ocurran en el futuro, como por ejemplo, el incremento en los activos como consecuencia de un incremento en los pasivos, con el objetivo de observar el efecto que esas posibles variaciones tendrían en las posiciones de liquidez de la institución.

Otra de las metodologías para el control, que las instituciones financieras deben reportar a la SBS, es identificar a los cien mayores depositantes; es decir, los clientes que mantienen los montos más altos en depósito a la vista en la institución, esto se hace con la finalidad de medir el impacto de la deserción de uno o más de los clientes en este grupo.

La metodología LaR (Liquidity at Risk) es la más frecuentemente utilizada para la medición del riesgo de liquidez, se determina por el nivel de bursatilidad de cada uno de los instrumentos que conforman la posición, obteniéndose una medida de VaR ajustado por liquidez, dicha metodología consiste en adicionar al VaR de mercado el costo que representaría no poder vender el instrumento por falta de liquidez en el mercado.

### **1.2.3. RIESGO OPERATIVO:**

El riesgo operativo se define como el riesgo de pérdida debido a la inadecuación o fallos de los procesos, el personal y los sistemas internos o bien a causa de acontecimientos externos.<sup>2</sup>

---

<sup>2</sup> Nuevo Acuerdo de Capitales, Comité de Basilea II junio 2004.



Entre las actividades generadoras de riesgo operativo se encuentran: el Outsourcing de procesos, la integración de sistemas por fusiones o adquisiciones, las prácticas comerciales agresivas y el crecimiento de servicios bancarios a través del internet.<sup>3</sup> Los eventos de riesgo operativo generan pérdidas indirectas y pérdidas directas, estas últimas se contabilizan como gasto. A continuación se presentan ejemplos de pérdidas indirectas por riesgo operativo:

- Transacciones no realizadas por falta de reemplazo de personal o por falta de documentación.
- Horas – Hombre gastadas en resolver fallas diarias, realizar operaciones mecánicas, seguir procesos ineficientes.
- Tiempo consumido en reprocesos.
- Costos de oportunidad en general.

Para la identificación del riesgo operativo en una entidad se realiza un inventario de procesos, que permita establecer las etapas críticas de cada uno de los procesos seguidos en la institución, así como vincular los procesos críticos con sus respectivas áreas o departamentos. Es decir se requiere identificar líneas de negocio y tipos de eventos.

Entre los métodos utilizados para la medición del riesgo operativo, y consecuentemente para determinar la pérdida esperada o provisión<sup>4</sup>, se tienen:

- Método Básico: Establecer el 15% de los ingresos netos como provisión.
- Método Estándar: Se establecen líneas de negocio, y en cada una de ellas se consideran factores de riesgo (porcentaje de pérdida). La provisión obtenida es la suma de los ingresos en cada línea de negocio multiplicada por su respectivo factor de riesgo.
- Métodos Avanzados: cada entidad según su experiencia en riesgo operativo determina el método que mejor se ajuste a su caso, entre los métodos conocidos están:

---

<sup>3</sup> Scalar Consulting, Aspectos Cuantitativos y Cualitativos de Riesgo Operativo.

<sup>4</sup> Más adelante se presenta una explicación a cerca de las provisiones.

- Score card: método cualitativo que permite medir el porcentaje de cumplimiento de los procesos en plazo y efectividad.
- VaR<sup>5</sup> Operativo: se determina la frecuencia con que ocurren pérdidas por riesgo operativo según el tipo de evento, y se cruzan con la severidad de la pérdida (medida económica, porcentaje de la pérdida total). Requiere de la construcción de una matriz de riesgo, matriz de frecuencia de doble entrada que considera Línea de negocio vs Tipo de evento. Una vez que se determina el tipo de distribución que sigue el modelo completo (frecuencia y severidad) se realiza la simulación; la pérdida resultante será la peor pérdida observada en las simulaciones a un cierto nivel de confianza. El modelo considera la pérdida esperada y la no esperada.

#### **1.2.4. RIESGO DE CRÉDITO:**

El riesgo de crédito se refiere a la pérdida potencial en la que incurre quien otorga un crédito, debido a la posibilidad de que la contraparte no cumpla con sus obligaciones (probabilidad de no-pago).

Esta definición simplificada esconde varios riesgos, la cantidad de riesgo es el saldo insoluto del crédito otorgado. La calidad resulta tanto de la probabilidad de que ocurra el incumplimiento, como de las garantías que reducen la pérdida, debido a la recuperación potencial que se puede hacer del crédito, lo que depende de cualquier elemento que mitigue el riesgo, tales como las garantías reales, los avales, la capacidad de negociación con el acreditado, entre otros que permiten identificar la pérdida en el evento del default.

El incumplimiento es un elemento incierto y por otro lado la exposición al riesgo de crédito al momento del incumplimiento no se conoce. Así mismo la recuperación que se pueda hacer de un crédito tampoco se conoce de antemano.

---

<sup>5</sup> Value at Risk

Los factores que influyen en el riesgo de crédito pueden resumirse en los que se detallan a continuación:

- La economía: un buen crecimiento económico implica menor desempleo y mejor índice de calidad de cartera.
- El segmento de mercado: No siempre sigue el mismo camino de la economía.
- La actividad económica del asociado o cliente: Factores socioeconómicos, por ejemplo cuando es empleado su situación económica está muy ligada a la salud financiera de la empresa donde labora.

La administración del riesgo de crédito incluye fijar límites de crédito, con el fin de restringir las pérdidas en caso de incumplimiento. Antes de la toma de cualquier decisión de crédito debe existir un proceso de evaluación, mismo que debe establecer el monto máximo en riesgo que se está dispuesto a asumir con un cliente actual o futuro.

Los principios para establecer límites persiguen los objetivos siguientes:

- Evitar que la pérdida en un solo crédito ponga en peligro a la institución.
- Diversificar los compromisos de otorgamiento de crédito en varias dimensiones (por cliente, por sector económico, por región o zona geográfica).
- Evitar otorgar crédito a cualquier persona o grupo por una cantidad tal que exceda su capacidad de endeudamiento.

Las técnicas desarrolladas para la medición del riesgo de crédito y la medición de la probabilidad de incumplimientos son muy variadas, a continuación se nombran brevemente algunas de ellas<sup>6</sup>:

- Técnicas econométricas: análisis lineal y discriminante, regresiones múltiples, modelos binarios para estimar la probabilidad de incumplimiento como variable dependiente, cuya varianza es explicada

---

<sup>6</sup> Martha Galicia Romero, nuevos Enfoques de Riesgo de Crédito.

por un conjunto de variables independientes. Las variables independientes deben estar relacionadas con el acreditado.

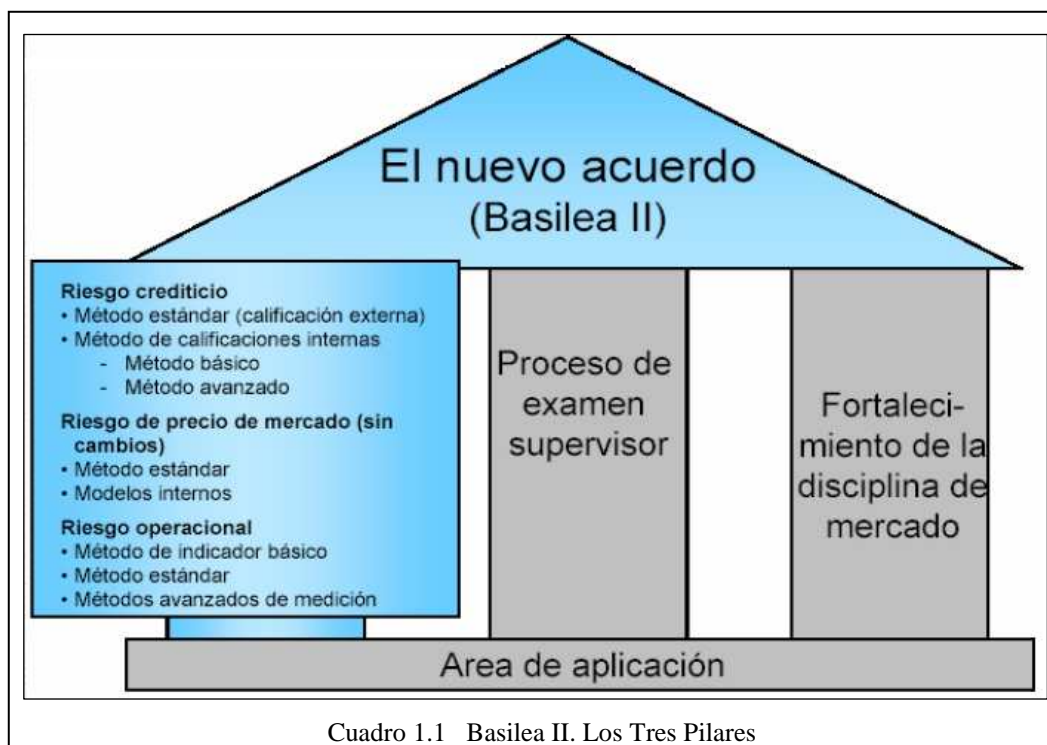
- Redes neuronales: utilizan los mismos datos que las técnicas econométricas pero crean un modelo de decisión a través de emular una red de neuronas (unidades de decisión) interconectadas.
- Modelos de Optimización: herramientas matemáticas de programación que buscan optimizar la relación entre el acreditado y los atributos del crédito para minimizar el incumplimiento y maximizar la utilidad de la institución.
- Sistemas expertos: Se utilizan para tratar de replicar de manera estructurada el proceso que un analista experto realiza para tomar una decisión de crédito, se caracterizan por establecer un grupo de reglas de decisión.
- Sistemas Híbridos (Sistemas de cómputo, Estimaciones y Simulaciones): Buscan relaciones directas causales de incumplimiento a través de la estimación de parámetros y la elaboración de matrices de probabilidad de migración para predecir la tendencia de un crédito a migrar a una mejor o peor condición.

### **1.3. ACUERDO DE BASILEA**

La administración integral del riesgo es una preocupación del sistema financiero mundial, por lo que se volvió necesario establecer políticas de regulación comunes a todas las entidades dedicadas al negocio de la banca, lo que dio la pauta para el Acuerdo de Capitales para la Banca, conocido como Acuerdo de Basilea de 1988, el cual transformó significativamente los mercados financieros, al establecer medidas de riesgo a través de ponderaciones de los diferentes tipos de activos. Posteriormente se realizó el Acuerdo de Basilea II, que se trata de un armazón regulatorio más sensible al riesgo financiero y que representa métodos que incluyen disciplinas avanzadas.

El Acuerdo de Basilea I sirve como la base sobre la cual las instituciones financieras determinan sus requerimientos de capital, a partir de la medición del

riesgo crediticio, el requerimiento se establece mediante el denominado método estandarizado, el cual implica valorar a los activos que mantiene una institución financiera, dándoles un peso en el valor total del activo según los riesgos asumidos al conservar dichos activos en balance bancario.



El Acuerdo de Basilea II incorpora dos componentes adicionales a Basilea I, resultando en tres pilares que se refuerzan mutuamente, los mismos que se describen a continuación:

### 1.3.1. PRIMER PILAR

Este pilar se ocupa de la determinación del coeficiente de solvencia (proporción de capital a activos), para cuyo efecto se hace referencia al cálculo de los requerimientos mínimos de capital para los riesgos de crédito, de mercado y operativo, este último no se consideraba en el primer acuerdo de Basilea. El objetivo principal es determinar una cobertura para posibles pérdidas.

Desde Basilea I se acordó que el coeficiente de capital (o de solvencia), en ningún caso podrá ser inferior al 8% del capital total, el cual se obtiene utilizando la definición de capital regulatorio de tres formas distintas: por el método estandarizado que se establece mediante ponderaciones de activos externas

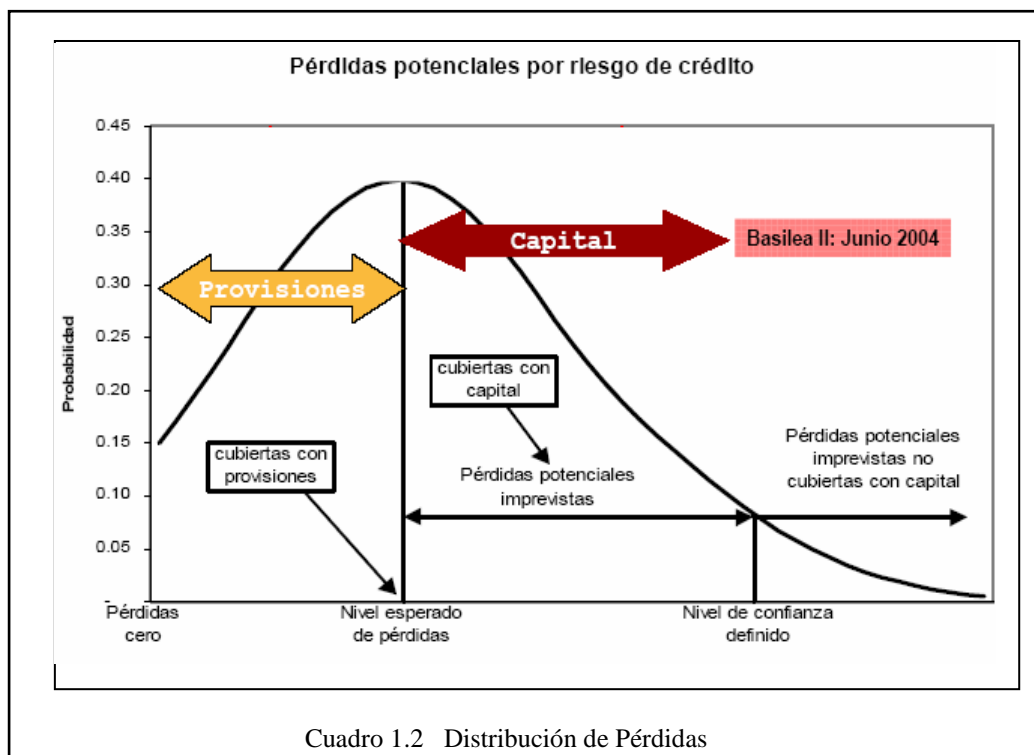
(dadas por el ente regulador), y por métodos de calificación interna, los conocidos FIRB (Fundamental Internal Ratings Based), es decir, Método fundamental basado en Calificaciones Internas y el AIRB (Advanced Internal Ratings Based) que es el método avanzado.

En los siguientes enunciados se profundiza en los términos introducidos en este párrafo:

- **Pérdida**

Se considera como pérdida a la porción de capital de una cartera que no se recupera. La pérdida que se produce como consecuencia de situaciones previsibles, constituye la pérdida esperada. Los perjuicios ocasionados por adversidades posibles pero no previsibles, constituyen la pérdida inesperada.

El cuadro 1.2 muestra un ejemplo de distribución de probabilidad para la pérdida, la esperanza de la variable pérdida es la pérdida esperada de la cartera, para cubrirla las instituciones financieras establecen provisiones por un valor igual al de la pérdida esperada, de tal manera que al darse el evento de la pérdida la institución está preparada para asumirla, en la parte izquierda del cuadro se señala la parte cubierta por provisiones. La pérdida mayor a la esperada es la pérdida inesperada, en el evento de que ocurra debe cubrirse con capital, por lo tanto, quita valor a la institución; en el cuadro 1.2 la pérdida inesperada se ubica a la derecha de la pérdida esperada. Obsérvese que en la cola derecha de la distribución se ubican pérdidas inesperadas con un bajo nivel de probabilidad de ocurrencia que al darse no se cubren ni siquiera con capital.



Fuente AIS

La provisión se convierte en un activo improductivo para la institución financiera y en consecuencia le resta rentabilidad.

- **Capital Regulatorio y Capital Económico**

Se entiende por capital regulatorio el nivel de capital mínimo exigido por el regulador, y por capital económico, el nivel de capital que elegirían los accionistas de una entidad financiera en ausencia de regulación.

El capital regulatorio y el capital económico no dependen totalmente de las mismas variables: el regulatorio (pero no el económico) depende del nivel de confianza exigido por el regulador, mientras que el económico (pero no el regulatorio) depende del margen de intermediación y del coste del capital bancario. Además, el capital regulatorio y el capital económico no reaccionan de la misma manera ante cambios en las variables que afectan a ambos, como son

la probabilidad de incumplimiento, la pérdida en caso de incumplimiento y la correlación entre los impagos de distintos acreditados.<sup>7</sup>

De acuerdo con Basilea II, el capital mínimo de cada banco debe determinarse sobre la base de tres tipos de riesgo - riesgo de crédito, riesgo de mercado y riesgo operativo – de tal forma que su capital efectivo sea siempre igual o superior a la suma de:

- 8% del valor de los activos de la institución ponderado cada uno por su nivel riesgo (riesgo de crédito), y
- Capital mínimo asociado a los riesgos de mercado y riesgo operativo

Para verificar el cumplimiento de las exigencias patrimoniales mínimas se ha establecido un indicador denominado Coeficiente de Capital, el cual no puede ser inferior a 8%<sup>8</sup> y se define de la siguiente forma:

$$\text{Coeficiente de Capital} = \frac{\text{Capital Regulatorio}}{\text{Activos Ponderados por Riesgo}} \geq 8\%$$

$$\text{Activos Ponderados por Riesgo} = \text{Capital RC} + 12.5 * (\text{Capital RM} + \text{Capital RO})$$

Donde:

*Capital RC* ; es el capital mínimo requerido por riesgo de crédito.

*Capital RM* ; es el capital mínimo requerido por riesgo de mercado.

*Capital RO* ; es el capital mínimo requerido por riesgo operativo.

El Activo Ponderado por Riesgo se entiende como la exposición de cada tipo de activo a pérdidas no esperadas y por tanto al requerimiento de capital regulatorio que se estima suficiente para cubrir esas pérdidas no esperadas.

---

<sup>7</sup> ELIZALDE Abel y REPULLO Rafael, Capital Regulatorio y Capital Económico Un Análisis de sus Determinantes.

<sup>8</sup> En América Latina el límite es 9%



El riesgo de mercado y el riesgo operacional se ponderan por 12,5 (que es el inverso de 8%) con el objeto de crear un vínculo numérico entre el cálculo de capital por riesgo crediticio, que es efectivamente 8% de los activos ponderados por riesgo, y los requisitos de capital por riesgo operativo y de mercado, que son equivalentes a la cuantificación de este tipo de riesgos.

La modificación de Basilea II en relación con Basilea I se refiere a los mecanismos de determinación del riesgo crediticio y a la incorporación del riesgo operativo en la medición del Coeficiente de Capital.

En Ecuador, se ha ido paulatinamente implementando el capital mínimo requerido para riesgo de crédito, a través de las metodologías estipuladas en el acuerdo de Basilea, otras requeridas por el Supervisor y en algunas instituciones a través de modelos estadísticos que permiten medir la probabilidad de incumplimiento. En cuanto a los requerimientos de capital por Riesgo de Mercado y Operativo, su implementación ha sido más lenta, dándole mayor importancia al riesgo operativo. La SBS ha dado plazo hasta octubre de 2008, a las instituciones del sistema financiero nacional para implementar al menos el método estándar, propuesto por Basilea, para la medición de riesgo operativo, por lo tanto hasta esa fecha las instituciones financieras ecuatorianas deben identificar sus procesos operativos según la línea de negocio correspondiente, así como estimar la pérdida que probablemente generen estos procesos. La medición de riesgo de mercado ha quedado un poco relegada, sujeta a la implementación de las metodologías de riesgo operativo, debido a la relativa estabilidad que la dolarización le ha dado al mercado ecuatoriano, sobre todo en lo que se refiere a la volatilidad cambiaria, tasas de interés e inflación.

De ahí que en Ecuador, el coeficiente de capital regulatorio aún está incompleto y solo se tendrá el coeficiente definitivo en cuanto se implementen las metodologías para el cálculo de los riesgos operativo y de mercado.

- **Método Estandarizado**

Utilizado para determinar el requerimiento de capital por riesgo de créditos. Bajo el método estándar se establecen ponderaciones fijas según las categorías establecidas, a aplicar a las exposiciones (netas de provisiones específicas), cuyo riesgo es evaluado por calificadoras de riesgo externas u organismos de crédito a la exportación<sup>9</sup> admitidos por el supervisor. El método estándar incluye los lineamientos para que los supervisores puedan determinar la elegibilidad de una agencia de calificación. La calificación externa es opcional para los créditos corporativos y si un banco opta por no hacerlo, la ponderación de riesgo en ese caso es del 100%.

Las ponderaciones se realizan según la categoría de exposición o riesgo y la calificación otorgada por el ente externo, estas han sido definidas en el acuerdo y se presentan en la tabla 1.1:

Categoría de Exposición	AAA a AA-	A+ a A-	BBB a BBB-	BB+ a B-	Inferior a B-	No calificado	Observación
Soberanos y Bcos. Centrales	0%	20%	50%	100%	150%	100%	
Bancos	20%	50%	100%	100%	150%	100%	Opción 1
	20%	50%	50%	100%	150%	50%	Opción 2
	20%	20%	20%	50%	150%	20%	Opción 2 Corto Plazo
Créditos a Empresas*	20%	50%	100%	150%	150%	100%	
Crédito Hipotecario Residencial						35%	
Crédito Hipotecario Comercial						100%	
Otros Créditos al por menor						75%	
Créditos en Mora						100% o 150%	Depende de la Provisión
Otros Activos						100%	
Activos Fuera de Balance						20% o 50%	Depende del FCC**

Tabla 1.1: Ponderaciones sugeridas por Basilea II

A continuación se realiza una breve explicación de cada una de las categorías de exposición:

- **Soberanos y Bancos Centrales:** Se refiere a los títulos emitidos por el Estado (soberanos) y los Bancos Centrales de un país.
- **Bancos:** Se refiere a los activos relacionados con los bancos del sistema financiero local. Existen dos opciones para la ponderación en esta categoría; el regulador tiene la atribución de escoger cual de las dos

<sup>9</sup> En Basilea II, las Agencias de Crédito a la Exportación (ECAs) son instituciones financieras públicas que apoyan la exportación de productos, servicios y la realización de negocios en países en desarrollo y mercados emergentes, otorgando Créditos respaldados por el gobierno, Garantías, avales y seguros al sector productivo.

opciones aplicar. Los activos relacionados a entidades del sector público o a Bancos de desarrollo multilateral también son ponderados por una de las dos opciones. Bajo la opción 2 se aplica una preferencia a las operaciones de corto plazo, por denominar de esta forma a las operaciones con plazos de tres meses o menos.

- *Créditos a Empresas*: representa a los créditos corporativos, en esta categoría se incluyen a las aseguradoras.
- *Crédito Hipotecario Residencial*: Son los créditos de vivienda; que por su naturaleza se garantizan a través del bien adquirido que se califica como residencial. La ponderación por riesgo para este tipo de créditos es 35%, debido a la existencia de la garantía.
- *Crédito Hipotecario Comercial*: Corresponde a los créditos avalados por un bien comercial. Debido a los problemas, ocurridos en varios países, con este tipo de créditos su ponderación por riesgo es de 100%.
- *Créditos en Mora*: La porción en mora por más de 90 días de un crédito se debe ponderar de la siguiente manera:
  - 150%, si la provisión constituida es menor o igual al 20% del saldo insoluto del crédito.
  - 100%, si la provisión constituida es mayor al 20% del saldo insoluto del crédito.
- *Otros Activos*: Activos diferentes a los mencionados anteriormente y que se encuentren dentro del balance deben ser ponderados al 100% de la exposición.
- *Activos Fuera de Balance*: son activos que representan compromisos de una entidad, estos se contabilizan fuera de los Activos y Pasivos normales de la institución; se convierten en exposiciones de crédito equivalentes a través de Factores de Conversión de Crédito (FCC)<sup>10</sup>. Un compromiso con plazo menor a un año se pondera al 20% de su FCC; mientras que un compromiso con plazo mayor a un año lo hace a un 50% de su FCC. Y los compromisos cancelables en cualquier momento se ponderarán al 0%.

---

<sup>10</sup> Son los porcentajes asignados para convertir las Operaciones Fuera de Balance a Activos equivalentes de carácter crediticio dentro del Balance General. Dichos porcentajes se asignan por Calificación de Riesgo.

Basilea II prevé una gama más amplia de garantías admisibles que Basilea I a los fines de asignar los ponderadores de riesgo. Las garantías deben ser vinculantes para las partes, legalmente exigibles, irrevocables e incondicionales. Además remarca el principio de que no debería existir correlación positiva entre la calidad del crédito y de la garantía. Existen dos enfoques para el tratamiento de las garantías:

*Enfoque simple:* La ponderación por riesgo de la parte colateralizada estará sujeta a un piso del 20%, excepto si es efectivo en depósito 0% y Soberanos o Entidades del Sector Público 0%. Al resto del crédito se le asignará la ponderación por riesgo correspondiente a la contraparte (categorías de riesgo indicadas en la tabla 1.1). Es decir se pondera el riesgo de contraparte en función del riesgo de la garantía.

*Enfoque Integral:* permite una mayor cobertura del riesgo con las garantías. En este las entidades deben ajustar el valor del crédito y de la garantía según su volatilidad (mediante el empleo de aforos que mitiguen la probabilidad de pérdida de su valor, el cual también deberá considerar fluctuaciones de moneda, si la garantía está emitida en otra moneda y también contemplar si la volatilidad de la exposición es diferente a la del colateral). Los aforos pueden ser calculados por las entidades utilizando su propio método de cálculo de volatilidad, el cual debe ser previamente aprobado por el supervisor. El monto del aforo deberá tener en cuenta el tipo de instrumento y la periodicidad de la cuenta del margen de garantía.

$$E^* = \max\{0, [E(1 + He) - C(1 - Hc - Hfx)]\}$$

Donde :

$E^*$  = valor de la posición tras la cobertura del riesgo

$E$  = valor corriente de la posición

$He$  = descuento correspondiente a la posición

$C$  = valor corriente del colateral recibido

$Hc$  = descuento correspondiente al colateral

$Hfx$  = descuento correspondiente a la discordancia de divisas entre el colateral y la posición

- **Métodos Basados en calificaciones Internas (IRB)**

La principal diferencia entre el IRB y el método estándar radica en que las evaluaciones internas de los principales factores de riesgo estimadas por los bancos actúan como argumentos determinantes para el cálculo de la exigencia de capital mínimo.

Basilea II introduce el cálculo de la pérdida esperada mediante el siguiente modelo:

$$EL = PD * LGD * EAD$$

Cada uno de los términos de la ecuación se define según el Comité de Basilea II en la forma siguiente:

*EL* (Expected Loss)= *Pérdida Esperada*.

*PD* (Probability of default)= *Probabilidad de incumplimiento*; mide la probabilidad de que el prestatario incumpla en un horizonte temporal determinado.

*LGD* (Loss given default) = *Pérdida en caso de incumplimiento*; Calcula la proporción de las exposiciones que se perderían si se produjera el incumplimiento. La entidad financiera, a la hora de estimar la LGD, debe ser consistente con la pérdida económica, teniendo en cuenta todos los factores relevantes, tales como los costes directos e indirectos vinculados al cobro de la exposición. Por tanto la LGD no debería recoger sólo la pérdida contable sino también la económica. Para ello, el banco tiene que tener en cuenta su propia experiencia en cuanto a sus logros a la hora de los recobros de las exposiciones, aunque estas estimaciones han de ser conservadoras, por lo menos hasta que se cuente con suficiente evidencia empírica.

*EAD* (Exposure at default) = *Exposición al riesgo de crédito*; estima la exposición al riesgo antes de producirse el impago. La EAD de una partida, tanto de dentro

de balance como fuera, se podría definir como la exposición bruta esperada de la posición si se produce el incumplimiento del deudor.

El Comité de Basilea permite que la entidad financiera estime estas variables con sus propios modelos, pero no que determine todos los elementos necesarios para calcular sus requerimientos de capital ya que la ponderación de riesgo y, por tanto, las exigencias de capital se establecen combinando los valores proporcionados por el banco y la formulación específica dada por el Comité.

Una vez que el banco haya adoptado el método IRB para alguna de sus exposiciones deberá ir, con el tiempo, extendiéndolo a todo el grupo bancario, aunque puede ser difícil, sobre todo para el cálculo de la LGD y EAD en algunas carteras que adolezcan de falta de datos, aunque esto se podría ir solucionado con el paso del tiempo. Lo que si recoge el Acuerdo es que, cuando se opte por éste método, en cualquiera de sus versiones y para una clase de activos, se debe aplicar a todos los activos que corresponden a esa clase.

A la hora de aplicar el método IRB a los activos, el Acuerdo de Basilea ha agrupado conjuntamente a las exposiciones provenientes de empresas, emisores soberanos y bancarios, definiéndolas de la siguiente forma:

- *Exposiciones frente a empresas (Corporate exposure)*. Se definen como la obligación de una empresa, sociedad o propiedad de saldar una deuda.
- *Exposiciones interbancarias* Cubren las exposiciones frente a bancos, sociedades de valores y empresas del sector público.
- *Exposiciones soberanas* Cubren todas las exposiciones tratadas como soberanas en el método estándar. Se incluyen las exposiciones frente a emisores soberanos y sus bancos centrales y frente a las empresas del sector público. Igualmente se incluirán en esta categoría el Banco Multilateral de Desarrollo cuando satisfaga los criterios que den derecho a una ponderación por riesgo del 0% en el método estándar.

La formulación para estos tipos de exposiciones, en función de *PD*, *LGD* y *EAD*, se muestra a continuación:

$$\text{Correlación} = R = 0.12 \left( \frac{1 - e^{-50PD}}{1 - e^{-50}} \right) + 0.24 \left( 1 - \frac{1 - e^{-50PD}}{1 - e^{-50}} \right)$$

$$\text{Ajuste por Vencimiento} = b = [0.11852 - 0.05478 * \log(PD)]^2$$

$$\text{Requirimiento de Capital} = K = LGD \left[ N \left( (1 - R)^{-0.5} * G(PD) + \left( \frac{R}{1 - R} \right)^{0.5} * G(0.999) \right) * \left( \frac{1 + (M - 2.5)b}{1 - 1.5b} \right) \right]$$

$$\text{Activos Ponderados por Riesgo} = RWA = 12.5 * K * EAD$$

Donde *M* es la madurez efectiva<sup>11</sup>, *N*(*z*) representa a la distribución normal y *G*(*x*) es la inversa de la distribución normal.

Esta formulación se aplicará tanto para el **Método Básico (FIRB)** como para el **Avanzado (AIRB)**. La diferencia radica en el cálculo de las variables; en el método básico, la entidad sólo calcula la *PD*, siendo el resto de los valores proporcionados por el regulador, y en el avanzado, la entidad estimará todas las variables.

Esta formulación representa la pérdida inesperada, “usando el modelo de Merton, donde existe un único factor de riesgo sistemático y el valor de los activos del prestatario se supone sigue una distribución lognormal”.

Por tanto se parte de la idea de que las posibles pérdidas por riesgo de crédito, a las que tendría que hacer frente una entidad financiera, se representan mediante una función de distribución de probabilidad. A partir de esta función de probabilidad se determina el nivel de confianza especificado, aquel que pudiera cubrir las pérdidas en que se han incurrido. Este modelo, donde existe un único factor, es denominado “modelo asintótico unifactorial”, el cual permite replicar comportamientos de impago en los distintos acreditados teniendo en cuenta la correlación existente entre los mismos, sobre la base de su grado de dependencia

---

<sup>11</sup> Se refiere a la antigüedad temporal de los activos considerados.

de un único factor común que, generalmente, se asocia a la actividad económica<sup>12</sup>.

Independientemente de la versión elegida, el banco siempre calculará la Probabilidad de Incumplimiento (PD). Para ello utilizará sus propios métodos, no obstante el Comité de Basilea establece un límite a estas estimaciones: “La PD se corresponderá al mayor valor entre la PD anual asociada a la calificación interna del prestatario al que fue asignada la exposición y 0,03%”<sup>13</sup>. Con este límite, el Acuerdo de Basilea asegura también un límite mínimo de ponderación de riesgo y recoge la dificultad a la que se enfrentan los bancos a la hora de validar estimaciones de PD tan pequeñas.

De lo anterior se extrae que la metodología IRB exige a las entidades diseñar un sistema de rating que les permita clasificar y ordenar a los acreditados en distintos grados de calidad crediticia según su nivel de riesgo (perfiles de clientes). El segundo paso será estimar los factores de riesgo de cada grado según el enfoque que se aplique, con todo ello se le asignará un valor de PD a cada uno de los grados crediticios.

### **1.3.2. SEGUNDO PILAR**

El segundo pilar son estándares supervisados y revisados para asegurar capital adecuado, la supervisión bancaria es necesaria, puesto que evita la disminución de rentabilidad y las crisis financieras, que hace poco tiempo fueron mundialmente comunes. En Ecuador el ente encargado de la supervisión es la Superintendencia de Bancos y Seguros (SBS), encargada de regular y monitorear el sistema financiero, así como los mecanismos que obligan a las instituciones financieras a mantener una provisión de capital destinada a cubrir pérdidas esperadas.

---

<sup>12</sup> Para mayor explicación ver El método IRB en el Acuerdo de Basilea, de Reyes Samaniego Medina y Martín Marín José Luis, Universidad Pablo de Olavide.

<sup>13</sup> BASEL COMMITTEE ON BANKING SUPERVISION (1999): “Credit Risk Modelling: Current Practices and Applications”. *abril*.



Los objetivos de la regulación son; entre los más importantes, que los bancos conozcan sus propios riesgos: al guardar la información de sus clientes por 5 años como mínimo, según Basilea II, al compartir tanto información positiva como negativa y al aplicar políticas de incentivos a sus buenos clientes; además, el regulador desea que los clientes sepan que las instituciones bancarias tienen y comparten información de su historial financiero.

### **1.3.3. TERCER PILAR**

El tercer pilar es la introducción de una disciplina de mercado a través de requerimientos de revelación pública. El Sistema Financiero formal, en la actualidad, comparte su información crediticia por medio de las Centrales de Información o Buros de Crédito, que son empresas constituidas como Sociedades de Información Crediticia, orientadas a integrar y proporcionar datos previos al otorgamiento del crédito y durante el plazo del mismo. Cada institución financiera entrega reportes periódicos a la Superintendencia de Bancos (SBS), los mencionados reportes son depurados y filtrados por la Superintendencia para posteriormente ser entregados a los Buros de Crédito donde la información de la totalidad de clientes del sistema financiero es procesada y transformada a un formato común que luego es vendido a las entidades bancarias.

La información es compartida con los siguientes fines; conocer la deuda total de cada cliente en el sistema financiero, comparar el comportamiento de un deudor con distintos bancos, identificar préstamos vinculados, entre otros. Una cultura crediticia se desarrolla como consecuencia de la existencia y transparencia de la información relacionada; por ejemplo, se evita que un cliente en mora en cierta entidad acceda a un nuevo crédito en otra entidad, se incentiva el pago a tiempo por la no aprobación de créditos futuros.

En Ecuador de acuerdo a la normativa de la Superintendencia de Bancos y Seguros (SBS), toda persona que opte por un préstamo bancario, tarjeta de crédito u otra operación crediticia pasa inmediatamente al buró de crédito denominado Central de Riesgos. Por tanto constar en este registro no es un

asunto que deba preocupar tanto como la calificación que allí recibe el cliente o deudor, pues esta determina si es un buen, regular o mal pagador, en una escala de cinco categorías (desde la A hasta la E). La Central de Riesgos ha permitido mejorar el nivel de morosidad en los bancos, las categorías de calificación se detallan a continuación:

**A:** Riesgo Normal: aquella persona que cuenta con suficientes ingresos para pagar el capital e intereses y lo hace puntualmente.

**B:** Riesgo Potencial: clientes que todavía demuestran que pueden atender sus obligaciones pero que no lo hacen a su debido tiempo.

**C:** Riesgo Deficiente: personas con ingresos deficientes para cubrir el pago del capital y sus intereses en las condiciones pactadas.

**D:** Riesgo Dudoso: igual que las personas que tienen calificación C, pero donde se tiene que ejercer la acción legal para su cobro y, generalmente renegocian el préstamo bajo otras condiciones.

**E:** Riesgo de Pérdida. Aquí están las personas que se declaran insolventes o en quiebra y no tienen medios para cancelar su deuda.

El objetivo de los párrafos anteriores ha sido describir brevemente las generalidades inmersas en los tres pilares que conforman los Acuerdos de Basilea, mismos que comprometen a las entidades financieras a llevar un adecuado control del negocio de la intermediación. La necesidad de medir los riesgos implícitos al sector ha impulsado el desarrollo de métodos estadísticos y matemáticos que permiten conseguir el objetivo propuesto, varios son los modelos conocidos para la medición de los diferentes tipos de riesgos –algunos de ellos ya fueron mencionados en la clasificación del riesgo -.

El objeto del presente trabajo es presentar una de las herramientas estadísticas para la medición de la probabilidad de incumplimiento, principal componente del riesgo de crédito individual, esta herramienta es el Credit Scoring –así denominado en el idioma inglés-, en español no se conoce una traducción simple que contenga todo el significado de los términos mencionados, por tal motivo en adelante se definirá y detallará lo que es un Credit Scoring.

## 1.4. CREDIT SCORING

Credit scoring es el conjunto de modelos de decisión y sus técnicas relacionadas que ayudan a los prestamistas en la asignación de crédito. Estas técnicas soportan la decisión de quién debe obtener un crédito, qué monto de crédito debería obtener, y qué estrategias operacionales incrementarían la rentabilidad de los clientes.

Las técnicas del credit Scoring asignan el riesgo de crédito a un cliente en particular. No obstante, la técnica no puede asignar a un consumidor la categoría de 'sujeto de crédito', pues éste no es un atributo o característica inherente al individuo, como el peso, la altura o incluso el nivel de ingresos. La consideración 'sujeto de crédito' es una valoración del que acredita respecto al acreditado y refleja las circunstancias en las cuales ambos se encuentran, así como la percepción del primero sobre los futuros escenarios económicos. Por lo tanto, no es saludable considerar 'no sujeto de crédito' a un consumidor cuyo perfil de riesgo no se acopla al requerido por la institución financiera. Resulta menos agravante y refleja mejor el estado de la realidad decir que la solicitud de crédito del prestatario representa un riesgo que no se está dispuesto a asumir.<sup>14</sup>

Las entidades financieras deben tomar decisiones en cualquier etapa del ciclo de crédito -se entiende por ciclo de crédito a las actividades secuenciales directamente relacionadas con la concesión de créditos-, por ejemplo, se debe decidir si otorgarle o no un crédito a un cliente nuevo, y en una etapa más avanzada la institución financiera requiere decidir cómo tratar a clientes ya conocidos e incluso si debe incrementar o disminuir sus límites de crédito o cupos de crédito.

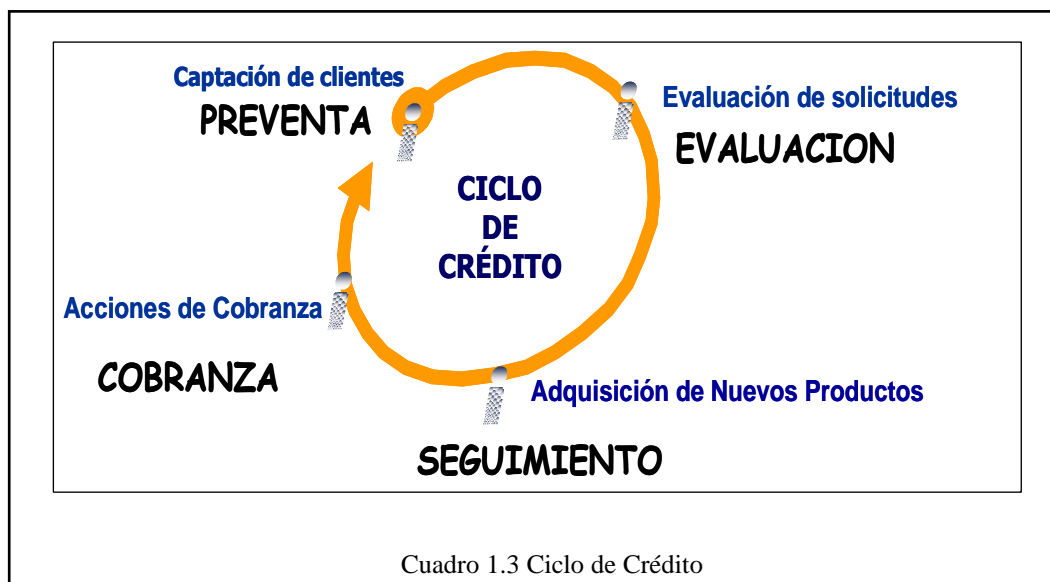
En las siguientes líneas se describen las etapas del ciclo de crédito de los clientes y los tipos de modelos scoring que pueden ser aplicados en cada etapa:

---

<sup>14</sup> Credit Scoring and its Applications, Lyn Tomas, David Edelman and Jonathan Crook

## 1.5. MODELOS SCORING EN EL CICLO DE CREDITO

El siguiente gráfico describe los procesos que conforman el ciclo de crédito, de acuerdo a la visión de LiSim<sup>15</sup>.



En cada etapa del ciclo de crédito es posible aplicar un modelo scoring distinto, con objetivos específicos para la etapa y en base a las diversas variables que se desarrollan mientras avanza el ciclo. Los modelos scoring tienen el objetivo de segmentar la población de clientes en grupos que permitan definir su comportamiento, según la etapa del ciclo de crédito en la cual se aplique.

A continuación se describen cada una de las etapas consideradas en el gráfico y la gestión a realizarse, así como el uso de modelos de segmentación (scoring) adecuados a cada fase:

### 1.5.1. PREVENTA

En los mercados actuales las empresas buscan y compiten por encontrar nuevos clientes. Las entidades financieras deben ser precisas y eficientes en la

<sup>15</sup> Firma de consultoría Internacional dedicada al diseño de soluciones inteligentes de negocios con estrategias de mercado específicas.

identificación y obtención de los clientes más apropiados para su negocio, ya que los recursos de tiempo y dinero son muy valorados y escasos.

Lo esencial en esta etapa es definir cómo y a qué clientes atraer, una de las herramientas para conseguir este objetivo es aplicar un scoring de iniciación, este modelo está construido exclusivamente con bases externas<sup>16</sup>, las bases externas a considerar son de tipo sociodemográfico y adicionalmente se incluyen bases con información del comportamiento de pago en otras instituciones del sistema financiero (bases de la Central de Riesgos), con esta información se calcula un puntaje que indica la probabilidad de que un cliente no cumpla con las obligaciones pactadas, cabe indicar que los clientes objetivos en esta etapa del ciclo del crédito son clientes nuevos, individuos que no han tenido ninguna operación de crédito con el banco anteriormente.

Una vez desarrollado el modelo, se procede a calificar a toda la base del registro civil, con la calificación obtenida mas la consideración del mercado objetivo se selecciona a los potenciales clientes. El mercado objetivo depende del enfoque que cada institución mantiene, los criterios son variados por ejemplo el estrato económico o la actividad desempeñada por el cliente.

El siguiente paso es establecer un contacto con los potenciales clientes, lo cual se realiza por medio de llamadas telefónicas (Call Center), la forma de atraer a estos clientes va a depender directamente de las gestiones estratégicas que se haga sobre cada cliente, es decir el proceso de CRM<sup>17</sup>.

### **1.5.2. EVALUACION**

Una vez identificado el cliente potencial, se debe realizar rápidamente el cierre de venta y establecerlo como una nueva cuenta dentro de la cartera de clientes.

---

<sup>16</sup> Base obtenida de una fuente diferente a la institución financiera.

<sup>17</sup> CRM, Customer Relationship Management, estrategia de negocio que gestiona la relación con los mejores clientes para optimizar su valor.

Posteriormente el cliente debe presentar los requisitos necesarios para verificar si es o no sujeto de crédito. La información recopilada permite tomar la decisión de la aprobación del crédito, y de ser así, determinar el cupo disponible para el cliente, además del plazo al que se otorgará el préstamo. Esta etapa concluye con el desembolso del monto aprobado.

Actualmente esto puede realizarse en tiempo real tomando las mejores decisiones, mediante la aplicación de modelos matemáticos y estadísticos adecuados para la etapa, Un buen ejemplo es la aplicación de un scoring de iniciación o aprobación.

Sí el posible nuevo cliente llegó a la institución por sus propios medios o fue atraído por la llamada telefónica (gestión del call center), el mencionado individuo tiene ya una calificación, resultado de la corrida del modelo, lo cual permite una agilización del proceso; puesto que solo se evalúan las solicitudes de los clientes, que el modelo determinó, adecuados al nivel de riesgo aceptado por la institución.

La evaluación de solicitudes incluye la verificación de la información proporcionada por el cliente; se comprueban direcciones, lugar de trabajo, antigüedad laboral, ingresos, referencias bancarias, deudas en el sistema financiero y el patrimonio declarado, así como la información del cónyuge, de ser el caso, y las cargas familiares. La confirmación se realiza mediante llamadas telefónicas, visitas al solicitante por parte de un oficial de crédito y mediante el acceso a bases de datos públicas, cómo la base del registro civil dónde se confirman datos sociodemográficos básicos, en la base del Seguro Social se ratifica la trayectoria y antigüedad laboral de quienes son afiliados, en las bases de matriculación y del municipio se verifica el patrimonio declarado y se consultan la bases de la central de riesgo para confirmar referencias bancarias y deudas en otras instituciones. Parte de la fase de verificación se facilita con la aplicación del scoring, puesto que una buena proporción de los datos mencionados son variables incluidas en el modelo. Por lo tanto el mayor énfasis se debe poner en la certificación de variables adicionales y en la actualización de la información considerada en el modelo.

La aplicación del modelo de evaluación automática permite un crecimiento en la prestación de servicios, teniendo un nivel exacto sobre el nivel de riesgos que desea manejar una entidad. El proceso de evaluación de solicitudes se realiza en forma objetiva y no ligada a la subjetividad de un analista, aunque el analista puede tener una amplia experiencia, no deja de tener sesgos en su apreciación.

### **1.5.3. SEGUIMIENTO**

Los cambios en las condiciones del mercado hacen que las entidades financieras deban monitorear y revisar constantemente la cartera de clientes existentes. Analizando la información y la historia de pago, es posible identificar y manejar oportunidades que permitan minimizar el riesgo.

En esta etapa el modelo scoring apropiado es uno de seguimiento, mismo que permite realizar la clasificación y segmentación de los clientes existentes para oferta de nuevos productos y desarrollo de estrategias de mercadeo, permitiendo un adecuado seguimiento y control de los créditos.

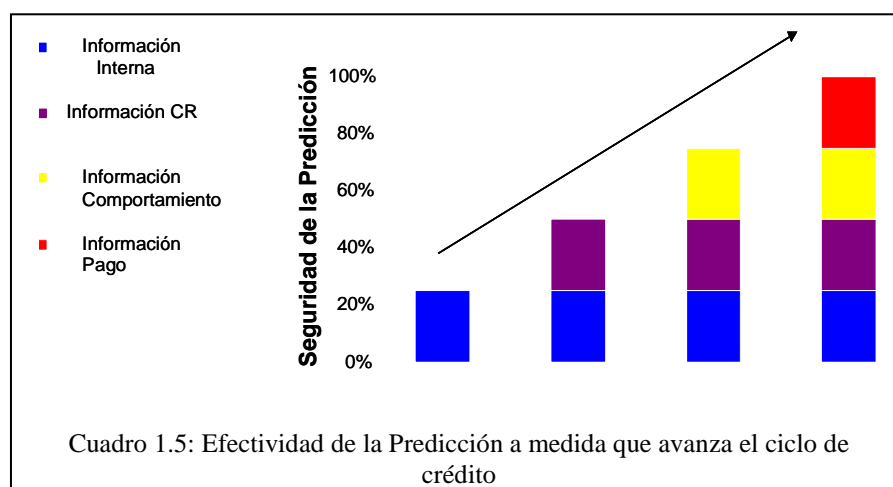
Mediante el desarrollo del modelo mencionado se determinan los clientes cuyo comportamiento amerita seguirlos manteniendo como tales, el objetivo es fidelizar a los buenos clientes y maximizar su rentabilidad, en otras palabras se desea optimizar el valor de un cliente a largo plazo. Con esta finalidad las entidades emprenden estrategias de venta dirigidas a clientes conocidos, ofreciéndoles, vía telefónica o por correo, productos complementarios al que ya tienen con beneficios adicionales (cross selling); por ejemplo a un cliente con un buen registro de pago en crédito de consumo, la institución financiera le ofrece una tarjeta de crédito, que es un producto revolvente, es decir, que a medida que el cliente paga su crédito la entidad financiera le sigue prestando según su cupo mes a mes.

Otra de las estrategias aplicadas para incrementar las operaciones y los servicios con los clientes antiguos es el conocido up selling, que es una estrategia que apunta más a la venta de mayor cantidad de un mismo servicio o producto, o a la

venta de productos / servicios de una “misma familia”, asociados naturalmente a los productos y servicios comercializados con un producto anterior; un ejemplo de esta estrategia aplicada a la banca es ofrecer a un buen pagador un crédito en la misma línea del anterior pero de un monto mayor y a un plazo más amplio.

En la etapa de seguimiento las instituciones deben también analizar las proporciones de deserción de sus clientes, sobre todo de aquellos que presentan un buen comportamiento de pago, es posible aplicar un scoring de deserción que permita calificar y segmentar todos los clientes en rangos, cuyos puntajes estiman la probabilidad de que un cliente quiera desertar o no de la entidad; de la misma manera permite actuar eficientemente en la retención de clientes leales y diseñar tácticas para elevar los perfiles de clientes neutrales, con el fin de que las entidades sean más rentables al mantener los clientes e inviertan menos en la vinculación de estos.

La efectividad de los modelos scoring se incrementa a medida que avanza el ciclo de crédito, puesto que en cada etapa del ciclo se incorpora mayor información acerca del cliente, con lo cual se tienen más criterios para definir si un cliente es bueno o malo y por lo tanto, la separación de las distribuciones de sus respectivas poblaciones es cada vez más amplia. El cuadro 1.5 ilustra esta última aseveración.





#### **1.5.4. COBRANZA**

Los deudores morosos pueden ser una amenaza para el sistema bancario. En la actualidad los bancos se focalizan en la administración de los gastos, por lo que es necesario realizar acciones de cobranza a clientes que tienen alta probabilidad de no pago.

En esta parte del ciclo de crédito la aplicación de un scoring de cobranza consiste en establecer un modelo estadístico para segmentar y pronosticar el riesgo de morosidad de los clientes, con el objeto de desarrollar estrategias de cobranza diferenciadas por tipo de cliente. Mediante este modelo se determinan e implementan estrategias de prevención y recuperación de cartera vencida para los deudores de la entidad, de acuerdo con pronósticos estadísticos de riesgo.

La gestión de cobranzas tiene el objetivo de incrementar los índices de recuperación de cartera, en otras palabras disminuir las pérdidas por no pago, cuando la mora es temprana la gestión de cobranzas debe ser personalizada, de tal manera que facilite la interacción con el cliente moroso con el fin de disminuir el monto en mora, ofrecerle facilidades para los pagos o en última instancia proceder a cambiar las condiciones de pago futuras (reestructuración de créditos). Cuando la mora ha pasado a niveles mayores se requiere de una gestión de cobranza judicial o ejecución de garantías si es el caso.

Aun con cobranza preventiva (sin mora), los costos de gestión disminuyen por la mayor eficiencia de recursos y mejores resultados por cobranza personalizada. Sólo por medio de una cobranza proactiva, que use eficientemente la información disponible y que diferencie a los clientes de acuerdo a su nivel de riesgo, se podrá maximizar la recuperación minimizando los costos asociados a la cobranza. Esto mejora la relación costo beneficio de la cobranza y reduce la cartera de impagados de la entidad. Las estrategias de cobranza a utilizarse deben ser cautelosas, para no desperdiciar los recursos utilizados por la entidad para construir imagen, promocionar productos, venderlos y fidelizar a los clientes, sobre todo en cobranza preventiva.

La ejecución de estos modelos scoring en cada etapa del ciclo de crédito, implica la existencia de una amplia muestra de clientes de la institución financiera, de la cual se tenga información de su aplicación, así como de su comportamiento en todas las etapas del ciclo de crédito. Los modelos buscan usar la muestra para identificar las conexiones entre las características de los clientes y cuan “bueno” o “malo” es su desempeño histórico, de tal suerte que se pueda segmentar la muestra según el desempeño de los clientes.

El Capítulo II se enfocará en la construcción de un modelo scoring para la fase de iniciación o evaluación, mismo que permitirá asignar a las características, o variables, incluidas una calificación (score), de tal forma que la suma de estas calificaciones conduzca a determinar si la probabilidad que tiene un cliente de ser malo es tan grande que no pueda ser aceptada. El modelo se construirá haciendo uso de la regresión logística para muestras desproporcionadas y el objetivo final será la obtención de perfiles de clientes en base a la pérdida probable que generen. El conjunto de las características (o variables) con su respectiva calificación (score) se conoce en el idioma inglés como scorecard. La suma de los scores en la scorecard servirá para determinar la probabilidad de incumplimiento de un cliente que aplica por primera vez a un crédito o de un cliente que la institución financiera busca atraer.

## **CAPITULO II**

### **SELECCIÓN DE PERFILES DE CLIENTES MEDIANTE REGRESIÓN LOGÍSTICA PARA MUESTRAS DESPROPORCIONADAS**

La construcción de un modelo de evaluación de clientes de crédito en fase de iniciación, en adelante se denominará scoring de iniciación, se basa en la mayor cantidad de información que se pueda disponer distinta a la relacionada al pago o al comportamiento crediticio dentro de la institución acreedora interesada en el otorgamiento del crédito, dado que esta última es inexistente porque es un potencial cliente, ante esto se requiere de información que permita visualizar el perfil social, demográfico y económico del cliente y también información del comportamiento en otras instituciones financieras.

Las variables sociodemográficas son de carácter general como: género, edad, nivel de instrucción, actividad, etc. Esta información puede provenir de dos fuentes diferentes, de la solicitud de crédito y de una fuente netamente externa a la institución financiera. Según la fuente, variará la eficiencia del modelo, si el modelo utiliza información de la solicitud se obtendrán mejores resultados, siempre que haya pasado por un proceso de control y verificación.

Para la ejecución del modelo scoring se utilizará información sociodemográfica netamente externa a la institución e información del comportamiento crediticio del cliente en otras instituciones del sistema financiero, para el análisis solo se considerarán los datos de clientes que solicitaron crédito por primera vez, las bases de datos utilizadas provienen de las fuentes que se detallan a continuación:

- Registro Civil; de donde se obtiene información a cerca de la edad, estado civil, información del cónyuge si es el caso, ciudad de nacimiento, profesión o nivel de educación.

- Seguro Social, de donde se obtiene datos referentes a lo laboral, si es afiliado o no, tipo de empresa, antigüedad en el trabajo actual, años de afiliación y sueldo.
- Comisión Nacional de tránsito se consigue información concerniente a si el cliente posee vehículos y el número de vehículos que posee.
- Servicio de Rentas Internas, que otorga datos a cerca de si el cliente posee RUC o no, la antigüedad de este, el tipo de actividad, entre otras.
- Empresas de Telecomunicaciones, de donde se conoce si el cliente tiene o no teléfono fijo.
- Buró de Crédito, de donde se obtiene el historial crediticio de la persona que aplica, por ejemplo si tiene o no deuda y la calificación de riesgo asignada en cada tipo de crédito.

Esta metodología está compuesta por un conjunto de modelos de decisión que ayudan al prestamista en la concesión de créditos. Las técnicas subyacentes a estos modelos permiten decidir no sólo quién recibirá el dinero, sino cuánto se le debe entregar y cuál debe ser la estrategia operacional que mejorará la rentabilidad de los consumidores. No obstante, la técnica no puede asignar a un consumidor la categoría de 'sujeto de crédito', pues éste no es un atributo o característica inherente al individuo, como el peso, la altura o incluso el nivel de ingresos. La consideración 'sujeto de crédito' es una valoración del que acredita respecto al acreditado y refleja las circunstancias en las cuales ambos se encuentran, así como la percepción del primero sobre los futuros escenarios económicos. Por lo tanto, no es saludable considerar 'no sujeto de crédito' a un consumidor cuyo perfil de riesgo no se acopla al requerido por la institución financiera. Resulta menos agravante y refleja mejor el estado de la realidad decir que la solicitud de crédito del prestatario representa un riesgo que no se está dispuesto a asumir.<sup>18</sup>

En las siguientes páginas se describirá la metodología utilizada para el desarrollo del modelo scoring de iniciación, mencionando las herramientas manejadas,

---

<sup>18</sup> Credit Scoring and its Applications, Lyn Tomas, David Edelman and Jonathan Crook

dejando para la sección de anexos la descripción detallada de la teoría matemática y estadística involucradas.

## **2.1. ETAPAS DEL SCORING DE EVALUACION**

Luego de la identificación de los potenciales clientes, el cierre de la negociación debe realizarse lo más rápido posible, pero la agilidad no es el principal objetivo, se requiere que el riesgo en la operación sea el mínimo deseado por la institución, considerando que la información disponible al momento de la evaluación se trata de inferir un comportamiento de pago a través de variables indirectas, solo se tiene información sociodemográfica, obtenida del registro civil por ejemplo, y la información crediticia disponible en la central de riesgos.

Construir un modelo scoring de evaluación abarca las siguientes etapas:

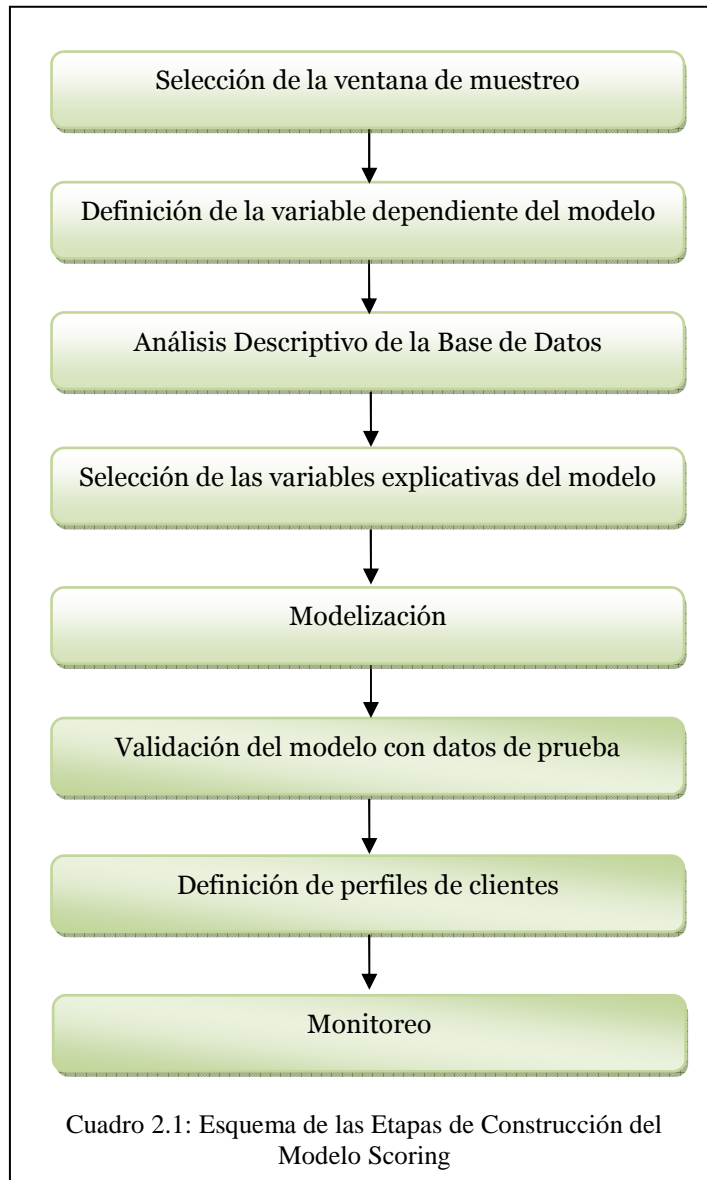
### **SELECCIÓN DE LA VENTANA DE MUESTREO**

En esta etapa se selecciona una muestra de los clientes, que permita la construcción del modelo sin sesgar el objetivo (evaluar a clientes nuevos para el banco), es decir seleccionar una ventana de tiempo y sobre esta solo tomar en cuenta a clientes que registraron por primera vez una operación en el banco. Para la selección de la muestra deben considerarse tres criterios: estabilidad, representatividad y madurez. La ventana de muestreo se dividirá en dos partes, una para la construcción del modelo y la otra para el control del mismo.

### **DEFINICION DE LA VARIABLE DEPENDIENTE DEL MODELO**

La variable dependiente será una variable con dos categorías bueno y malo, para definir esta variable se construye el indicador de buenos y malos clientes en base a las variables Atraso Promedio, Atraso Máximo, y la rentabilidad o pérdida que se asocia a los atrasos.

El promedio y máximo se miden en un periodo de tiempo (a partir de la fecha de desembolso) que permita considerar al cliente como maduro. (Periodo sobre el cual un cliente revela su real comportamiento, es decir si es bueno o malo).



## **ANÁLISIS DESCRIPTIVO DE LA BASE DE DATOS**

Realizar un control de calidad de la base de datos, mediante un análisis univariante de todas las variables con las que se cuenta. Los objetivos en esta etapa son identificar posibles errores en la base y de ser factible corregirlos,

también conocer el estado de las variables en la base, es decir identificar características de los datos.

El análisis univariante debe diferenciar las variables nominales de las continuas, para las primeras bastará con un estudio de frecuencias y para las segundas una identificación de máximos, mínimos y promedios dará una primera impresión de la calidad de la base de datos.

Como resultado del análisis descriptivo se debe obtener una base de datos depurada.

### **SELECCIÓN DE LAS VARIABLES EXPLICATIVAS DEL MODELO**

Se verifica que las variables en la base de datos depurada expliquen a la variable dependiente del modelo, es decir que estén correlacionadas con la misma. Con este objetivo es posible incluso realizar cruces de variables, tal que las variables resultantes tengan correlación con la dependiente y sean representativas de la población. La representatividad significa que se escojan cruces de variables que identifiquen características de un porcentaje significativo de la población.

### **MODELIZACION**

Con las variables o combinaciones de variables explicativas seleccionadas se procede a la estimación de un modelo discriminante, en este caso se usará regresión logística, pues es adecuado ya que la variable dependiente que se define es dicotómica.. Esta etapa incluye la validación estadística del modelo.

### **VALIDACIÓN DEL MODELO CON DATOS DE PRUEBA**

Esta etapa consiste en probar el modelo obtenido con la muestra de construcción, corriéndolo sobre la muestra de prueba o control, básicamente el objetivo que se desea conseguir es determinar la eficiencia del modelo sobre una muestra

distinta. Los índices que se considerarán son los mismos que se realizaron para las pruebas de eficiencia, excepto la divergencia.

## **DEFINICIÓN DE PERFILES DE LOS CLIENTES**

Con los resultados del modelo (Score) calculados para cada cliente se definen los perfiles (partición del score) en base a la pérdida promedio; es decir, determinar los puntos de corte para el modelo. Los perfiles así contruidos definen las características de pago del grupo de clientes que se asignen a cada partición.

## **MONITOREO DEL MODELO**

Desarrollar metodologías para monitorear los resultados del modelo scoring de iniciación, considerando los siguientes puntos: eficiencia, cambios en las distribuciones originales, impacto financiero (pérdida por perfil) y estadístico (correlación de las variables). Como resultado del monitoreo se determina la necesidad de ajustar o no el modelo, los ajustes son por ejemplo, corrección de los coeficientes del modelo, redefinición de los perfiles o la reingeniería total del mismo.

## **2.2. SELECCIÓN DE LA VENTANA DE MUESTREO**

La selección de la ventana de muestreo adecuada debe realizarse bajo tres consideraciones, representatividad, estabilidad y madurez; estas se definirán más adelante.

Para medir la estabilidad de la muestra se requiere en primera instancia definir una variable de interés en la población. Puesto que lo que se desea es diferenciar a los buenos y malos clientes para la institución financiera, se decide que la mora es una variable apropiada para caracterizar el comportamiento de los clientes. La mora es la diferencia en días entre la fecha de vencimiento y la fecha de pago, en caso de que la fecha de pago este antes que la fecha de vencimiento la mora s'era igual a cero. Es entonces coherente construir un Indicador de mora:



## 2.2.1. CONSTRUCCIÓN DEL INDICADOR DE MORA

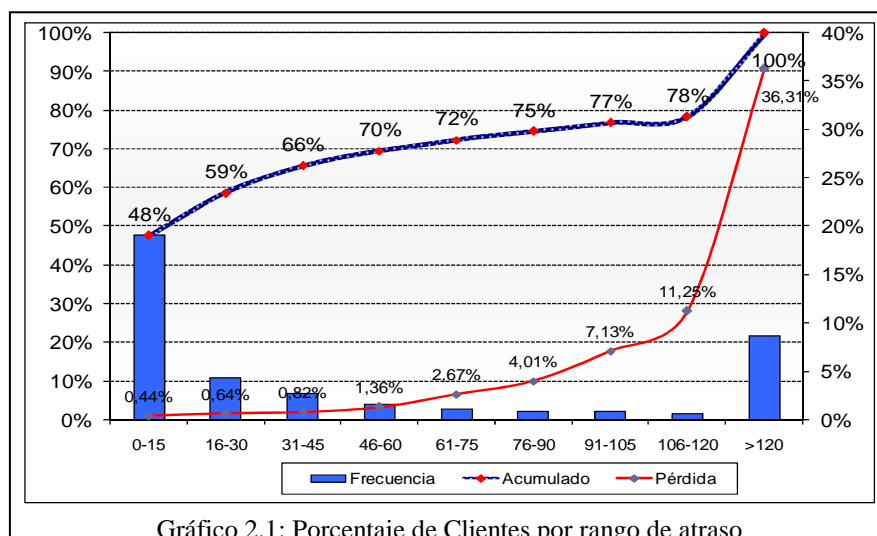
El indicador de mora se calcula por mes de desembolso, lo que se conoce en el argot financiero como análisis por cosechas, cada cosecha representa un mes de desembolso, entre los beneficios de analizar por cosechas se cuentan, el poder reconocer inmediatamente a qué grupo de clientes corresponde un alto o bajo indicador de mora, además elimina la estacionalidad que existe en este tipo de índices. El numerador del indicador de mora se establece en base al número de clientes que sobrepasan un atraso máximo  $W$ , este se denominará. La fórmula utilizada para el cálculo se muestra a continuación:

$$\text{Indicador Mora}_j = \frac{\text{Número clientes del mes de desembolso } j \text{ cuyo atraso máximo } > W \text{ días}}{\text{Número clientes del mes de desembolso } j}$$

De tal forma que el indicador representa el porcentaje de clientes cuyo crédito se desembolsó en el mes  $j$  y que sobrepasaron el umbral  $W$  de atraso máximo, en relación al total de clientes desembolsados en el mes  $j$ .

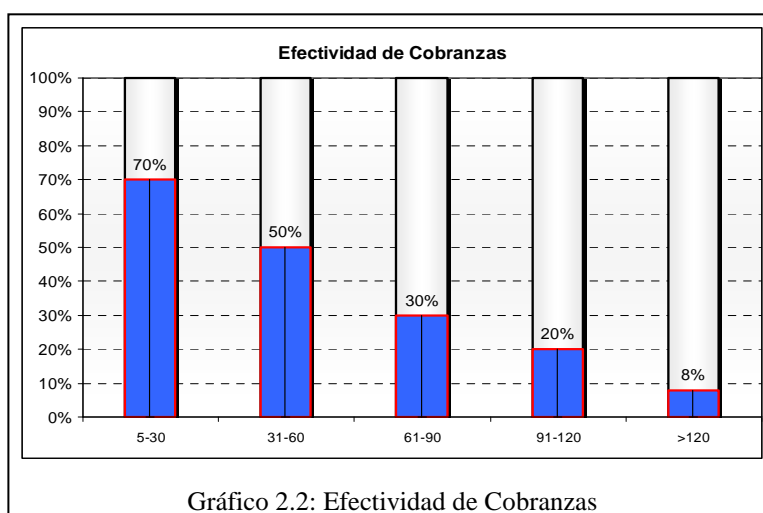
Es ahora necesario definir de una manera objetiva el umbral  $W$ , lo cual se realizará determinando hasta qué punto la mora se convierte en una pérdida significativa, lo que se conoce como mora dura.

A continuación se presenta un gráfico en el que se describe el porcentaje de los clientes que caen en cada rango de atraso, con la pérdida relacionada a cada uno de ellos:



Hasta el segundo rango de atraso en promedio solo el 59% de los clientes ha cancelado sus cuotas, esto representa una pérdida de 0.48% (pérdida ponderada en los primeros rangos). El restante 41% representa una pérdida ponderada del 20.48% en total.

La forma en que se recauda la cartera en mora va de la mano con la gestión de cobranzas, la efectividad de esta gestión se mide por rango de mora, puesto que para cada rango varía la frecuencia e intensidad de la gestión, a continuación se presenta un gráfico sobre la efectividad de cobranzas por rango de mora, para el tipo de cartera analizada:



El gráfico 2.2 indica que de la porción de cartera que cae en mora entre 5 y 30 días, la gestión de cobranzas logra recaudar el 70%, el 30% restante cae a un segundo rango de mora, donde la efectividad de cobranzas es menor, por ejemplo en el rango de 31 a 60 días de mora la efectividad es de 50%, y así sucesivamente.

Por lo tanto el umbral W será de 30 días y el indicador de mora queda definido entonces por:

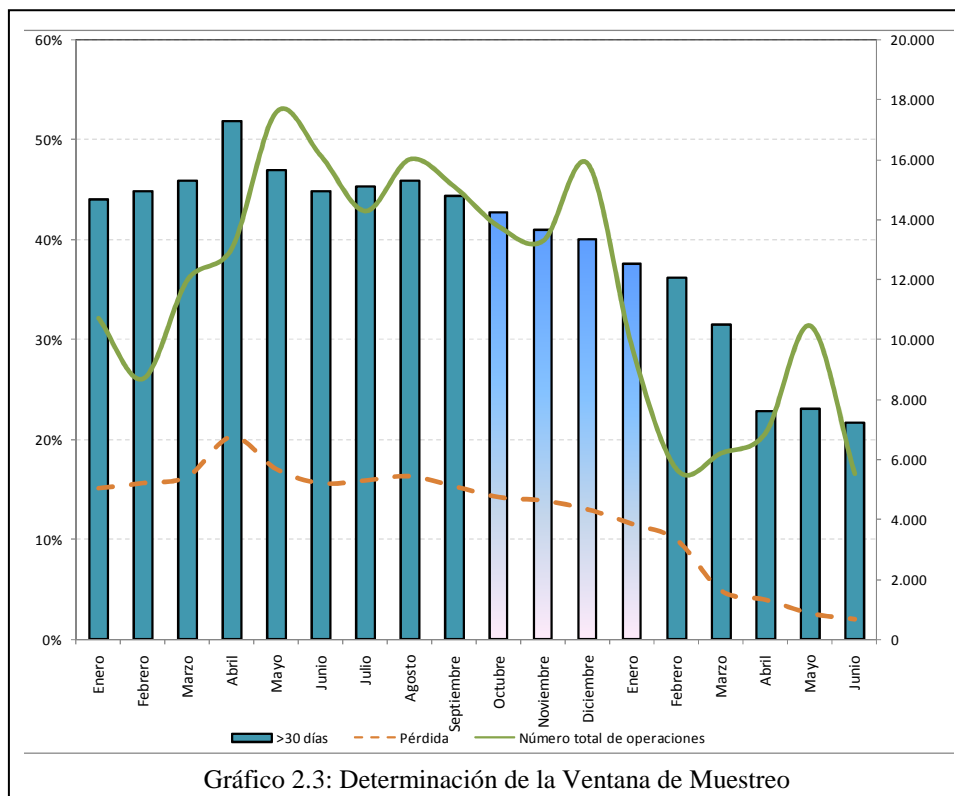
$$\text{Indicador Mora}_j = \frac{\text{Número clientes del mes de desembolso } j \text{ cuyo atraso máximo} > 30 \text{ días}}{\text{Número clientes del mes de desembolso } j}$$

El cálculo del indicador de mora para cada mes desde enero del 2005 a junio de 2006, consta en la tabla 2.1. Se puede apreciar con claridad el cálculo del indicador de mora y las pérdidas asociadas a cada mes de desembolso.

Año	Mes	Indicador Mora (N)		Número total de operaciones	Indicador Mora (%)		Pérdida
		<=30 días	>30 días		<=30 días	>30 días	
2005	Enero	5.998	4.719	10.717	55,97%	44,03%	15,11%
	Febrero	4.804	3.899	8.703	55,20%	44,80%	15,62%
	Marzo	6.500	5.494	11.994	54,19%	45,81%	16,28%
	Abril	6.293	6.759	13.052	48,21%	51,79%	20,23%
	Mayo	9.335	8.243	17.578	53,11%	46,89%	17,00%
	Junio	8.893	7.213	16.106	55,22%	44,78%	15,61%
	Julio	7.824	6.456	14.280	54,79%	45,21%	15,89%
	Agosto	8.666	7.335	16.001	54,16%	45,84%	16,30%
	Septiembre	8.400	6.684	15.084	55,69%	44,31%	15,30%
	Octubre	7.889	5.872	13.761	57,33%	42,67%	14,21%
	Noviembre	7.854	5.441	13.295	59,07%	40,93%	13,87%
	Diciembre	9.500	6.347	15.847	59,95%	40,05%	12,99%
2006	Enero	6.068	3.648	9.716	62,45%	37,55%	11,54%
	Febrero	3.626	2.056	5.682	63,82%	36,18%	9,93%
	Marzo	4.267	1.962	6.229	68,50%	31,50%	4,89%
	Abril	5.336	1.576	6.912	77,20%	22,80%	3,99%
	Mayo	8.056	2.407	10.463	77,00%	23,00%	2,63%
	Junio	4.349	1.197	5.546	78,42%	21,58%	2,05%
<b>Total</b>		123.658	87.308	210.966			12,41%

Tabla 2.1. Construcción Indicador de Mora por Mes

En base a la tabla anterior se realizó el gráfico 2.3, en el que se señala en color verde los meses escogidos para el desarrollo del modelo scoring, la ventana escogida va de octubre de 2005 a enero de 2006, con un total de 52.619 clientes.



Los clientes seleccionados en la muestra se distribuyen mes a mes de la forma indicada en la tabla 2.2.

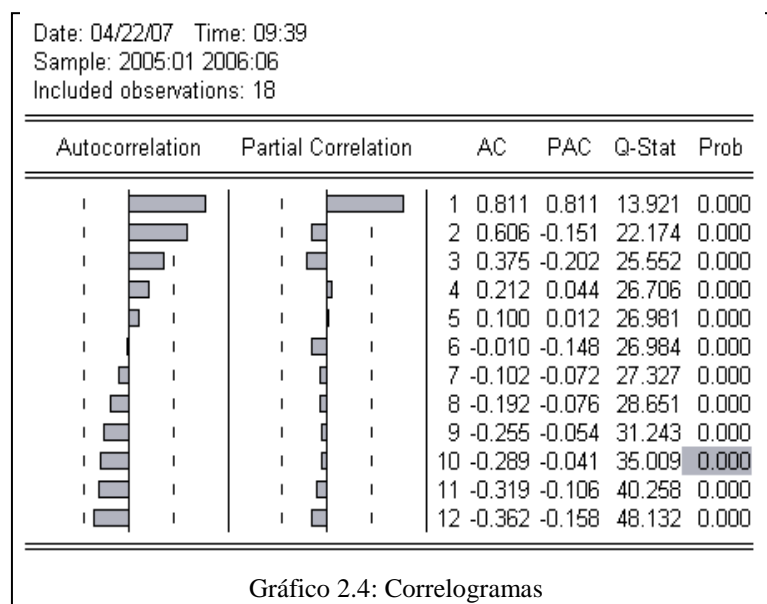
Año desembolso	Mes desembolso	Número Clientes
2005	Octubre	9,716
	Noviembre	13,761
	Diciembre	13,295
2006	Enero	15,847
<b>Total General</b>		<b>52,619</b>

Tabla 2.2: Tamaño Ventana de Muestreo

La selección de la ventana de muestreo, se debe realizar considerando tres criterios:

- Estabilidad, la cartera seleccionada para la muestra debe ser estable en el sentido de que el indicador de mora se mueva dentro de una franja no muy amplia, para lo cual en la serie formada por el indicador de mora en la muestra no debería existir tendencia, por lo tanto se debe probar que la mencionada serie se ajusta a un modelo ARMA(p,q), sin tener que diferenciarla ni una sola vez.

Las funciones de autocorrelación y autocorrelación parcial obtenidas en Eviews muestran que la serie construida con el indicador de mora es estacionaria y se ajusta, desde el punto de vista de las series temporales, con un modelo autoregresivo de orden uno, conocido como AR(1):

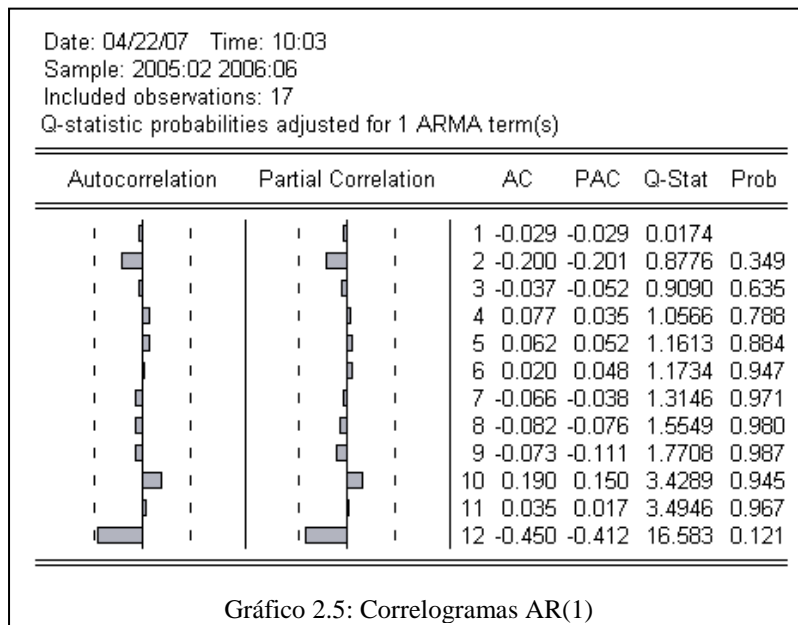


El resultado obtenido al modelar la serie con un AR(1) es el siguiente:

Dependent Variable: IND				
Method: Least Squares				
Date: 04/22/07 Time: 09:45				
Sample(adjusted): 2005:02 2006:06				
Included observations: 17 after adjusting endpoints				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.575490	0.235312	2.445645	0.0273
AR(1)	1.077364	0.096911	11.11701	0.0000
R-squared	0.891766	Mean dependent var		0.391588
Adjusted R-squared	0.884550	S.D. dependent var		0.092147
S.E. of regression	0.031310	Akaike info criterion		-3.979648
Sum squared resid	0.014704	Schwarz criterion		-3.881623
Log likelihood	35.82701	F-statistic		123.5880
Durbin-Watson stat	2.025772	Prob(F-statistic)		0.000000
Inverted AR Roots	1.08	Estimated AR process is nonstationary		

Tabla 2.3: Resultado modelación AR(1)

Las funciones de autocorrelación y autocorrelación parcial son estadísticamente cero, como se aprecia en el siguiente gráfico, las funciones se encuentran dentro de las bandas de confianza especificadas para la nulidad.



- Representatividad, la muestra escogida debe ser representativa de la población actual del banco, de tal manera que de los resultados obtenidos se pueda inferir con un buen nivel de confianza acerca de la población, además la muestra debe contener información suficiente de los distintos tipos de clientes y sus características, de tal manera que se pueda identificar y caracterizar, en base a la muestra, a los potenciales clientes de la institución financiera.

Madurez, el criterio de madurez en la cartera significa que la información que se tiene de un cliente sea suficiente en tiempo como para poder determinar sus características de pago, es decir, es el tiempo suficiente para que un cliente muestre si su comportamiento es bueno o malo.

El criterio de madurez se ratifica en el gráfico 2.3 donde se observa que los niveles de pérdida son significativos en la ventana de muestreo escogida, es decir, la cartera mala ya se ha revelado y el indicador se mantiene estable, esta cartera pasa a un proceso de recuperación que permite bajar los niveles de pérdida. Puesto que la extracción de la muestra se realizó con corte a noviembre 2006, la madurez mínima de la ventana de muestreo es 10 meses para los créditos originados en enero 2006, esta cantidad no es una definición general sino que dependerá del comportamiento de la cartera y del tipo de crédito en estudio.

La muestra seleccionada tiene que dividirse en dos partes: una para la construcción del modelo, y otra para realizar las pruebas retrospectivas o control del mismo. Las pruebas retrospectivas pueden ser realizadas escogiendo una muestra diferente, de cosechas pasadas, sin embargo, en el desarrollo no se contaba con la información de cosechas anteriores, pero dado que la finalidad de este trabajo es desarrollar una metodología se mostrará la realización de las pruebas retrospectivas con la división de la ventana de muestreo.

La división de la muestra se realizó por medio de un muestreo sistemático con  $k=3$ , empezando desde el primer cliente, se asegura que el proceso sea aleatorio porque no se alteró el orden que presentaban los clientes en la muestra, la

submuestra resultante corresponde a la que se utilizará para el control del modelo.

La representatividad que tiene las muestras de control y construcción debe mantenerse en las variables, por ejemplo en la variable ciudad se mantuvo, tanto en las muestras de control y construcción como en la ventana de muestreo total, según se verifica en la tabla 2.4:

Ciudad	Distribución de clientes en la muestra de control	Distribución de clientes en la muestra de construcción	Distribución de clientes en la Ventana de Muestreo
Ambato	3.49%	3.48%	3.48%
Cuenca	4.42%	4.27%	4.32%
Guayaquil	54.47%	54.67%	54.60%
Ibarra	4.31%	4.20%	4.24%
Latacunga	0.63%	0.62%	0.63%
Loja	1.47%	1.46%	1.46%
Manta	5.47%	5.27%	5.34%
Portoviejo	0.36%	0.33%	0.34%
Quito	24.19%	24.48%	24.38%
Riobamba	1.19%	1.21%	1.20%
Total	100.00%	100.00%	100.00%

Tabla 2.4: Representatividad de la variable ciudad en las muestras.

La primera columna corresponde a la muestra de control, la segunda a la muestra de construcción y la tercera a la ventana de muestreo total, se puede apreciar en la tabla que las muestras de construcción y control guardan las proporciones de la ventana de muestreo, por lo que las inferencias que se hagan no tendrán problemas de proporcionalidad en referencia a la ciudad.

De este modo la construcción del modelo se realizará con 34.523 clientes, y el testeo o control con 18.102 clientes, esto se resume en la tabla 2.5:

Muestra	No. Clientes
Control	18,096
Construcción	34,523
Total	52,619

Tabla 2.5. Partición de la Muestra

### 2.3. DEFINICION DE LA VARIABLE DEPENDIENTE DEL MODELO

En este punto el objetivo es obtener una partición de la cartera de clientes, mediante un criterio adecuado que permita identificar a los buenos y malos

clientes, a partir de lo que se obtendrá la variable dependiente del modelo; por lo tanto, una correcta clasificación a priori dará como resultado una correcta clasificación a posteriori.

Una de las técnicas más utilizadas para la construcción del indicador buenos/malos (indicador B/M), considera el atraso promedio y el atraso máximo en el que han incurrido los clientes, mediante el cruce de las mencionadas variables en una matriz, como se presenta en la tabla 2.6:

		Atraso Máximo										
		0	1-15	16-30	31-45	46-60	61-75	76-90	91-105	106-120	>120	Total General
A t r a s o	0	6254 (0,31%)	7786 (0,35%)	17 (0,27%)								14057 (0,33%)
	1-15		11245 (0,40%)	5964 (0,52%)	3074 (0,73%)	1047 (0,99%)	334 (1,66%)	131 (1,18%)	31 (1,60%)	10 (0,22%)	6 (20,65%)	21842 (0,54%)
	16-30			45 (0,67%)	471 (1,51%)	909 (2,04%)	847 (3,03%)	639 (3,57%)	459 (5,67%)	267 (6,32%)	267 (17,06%)	3904 (4,17%)
	31-45				7 (2,18%)	91 (3,23%)	276 (5,83%)	341 (7,53%)	448 (15,45%)	356 (16,86%)	1180 (29,26%)	2699 (19,24%)
	46-60						24 (5,81%)	90 (10,45%)	186 (20,38%)	178 (22,91%)	2405 (51,33%)	2883 (45,93%)
	61-75							10 (7,36%)	32 (17,07%)	70 (25,01%)	2143 (63,03%)	2255 (60,95%)
	76-90									3 (12,67%)	1015 (60,44%)	1018 (60,30%)
	91-105										587 (46,99%)	587 (46,99%)
	105-120										526 (50,92%)	526 (50,92%)
	>120										2848 (61,36%)	2848 (61,36%)
Total General		6254 (0,31%)	19031 (0,38%)	6026 (0,52%)	3552 (0,84%)	2047 (1,56%)	1481 (3,29%)	1211 (4,97%)	1156 (12,03%)	884 (15,34%)	10977 (53,59%)	52619 (12,26%)

Tabla 2.6 Matriz Atraso Promedio / Atraso Máximo

En la tabla 2.6 se distribuyen los clientes según el cruce que les corresponde por rango de atraso promedio (filas) y rango de atraso máximo (columnas), por ejemplo; existen 5.964 clientes cuyo atraso promedio está entre 1 y 15 días y con atraso máximo entre 16 y 30 días; el valor en paréntesis en cada casilla representa la pérdida  $(\frac{\text{provisión}}{\text{Monto Colocado}})$ . En base a la tabla anterior se determinan los valores de atraso máximo y promedio y la pérdida que generan. La entidad financiera debe establecer un criterio del valor de pérdida que es bueno o malo para la institución. Considerando los niveles de pérdida establecidos por la entidad se concluyó que los clientes con atraso promedio menor o igual a 15 días y atraso máximo de hasta 30 días son clientes buenos, casillas en color azul. Mientras que los clientes con atraso promedio entre 1 y 15 días, pero con atrasos máximos entre 31 y 105 días no pueden ser determinados como buenos o malos



clientes, al igual que aquellos con atraso promedio entre 16 y 30 días, y cuyo atraso máximo está entre 16 y 60 días; a este grupo se le denominará clientes indeterminados, puesto que no se tiene la información suficiente a cerca de ellos como para ubicarlos en un grupo definido, los clientes indeterminados están representados por la casillas en color naranja. Los clientes restantes son los clientes clasificados por el criterio como malos, casillas verdes. La tabla 2.7 a continuación muestra los porcentajes de clientes en cada categoría:

	Número	Porcentaje
<b>Buenos</b>	31.266	59,42%
<b>Indeterminados</b>	6.042	11,48%
<b>Malos</b>	15.311	29,10%

Tabla 2.7 Distribución Clientes

Se aprecia que hay una alta incidencia de clientes malos, que representan un 29.10% de la cartera, los clientes indeterminados constituyen el 11.48% y el 59.42% restante son los clientes buenos, por lo que se intuye la necesidad de establecer un mejor método de selección de la cartera. A nivel de consultoras se ha definido un estándar para la definición de buenos y malos, en el que se indica que la distribución es adecuada si mantiene un porcentaje de clientes indeterminados no mayor a 12%. Adicionalmente, la definición de buenos y malos dependerá de la acidez con que se quiera realizar el modelo.

Excluyendo a los clientes indeterminados, la relación de buenos a malos clientes es aproximadamente de 2 a 1.

La distribución de buenos y malos en la muestra de construcción se mantiene, como se puede observar en la tabla 2.8

	Número	Porcentaje
<b>Buenos</b>	20.570	59,58%
<b>Indeterminados</b>	3.904	11,31%
<b>Malos</b>	10.049	29,11%
<b>Total</b>	<b>34.523</b>	<b>100,00%</b>

Tabla 2.8 Distribución Clientes

Puesto que el objetivo final del modelo es diferenciar a los buenos de los malos clientes, se excluirán de la construcción a los clientes indeterminados, ya que estos sesgarían el modelo. Una vez obtenido el indicador de buenos y malos (indicador B/M), este se constituye como la variable dependiente del modelo. La variable dependiente es entonces dicotómica, donde los buenos se representarán con 0 (cero) y los malos con 1 (uno).

## **2.4. ANALISIS DESCRIPTIVO DE LA BASE DE DATOS**

Se desea escoger las variables que aportan información al modelo, en una primera aproximación se realizará el análisis exploratorio de los datos.

Las bases de datos con las que se trabajará son completamente externas, como atañe a un scoring de iniciación, como ya se mencionó las bases provienen de varias y diversas fuentes, que permitirán poseer información de un cliente, del que el evaluador apenas conoce su número de identificación.

La información utilizada es personal, laboral, económica y crediticia, esta última obtenida de la central de riesgos, que corresponde exclusivamente al comportamiento del cliente en instituciones diferentes a la entidad para la que se realiza el modelo. En la tabla 2.9 se muestra una descripción de las variables con las que se cuenta inicialmente.

Alias	Descripción	Tipo de Dato
codigocredito	Código del Crédito	Texto
codigocliente	Código del Cliente	Texto
numerodocumento	Número del Documento	Texto
fechadesembolso	Fecha de Desembolso	Fecha
añodesembolso	Año de Desembolso	Númerico
atrasomax	Atraso Máximo	Númerico
atrasoprom	Atraso Promedio	Númerico
desciudad	Descripción Ciudad	Texto
fechaactivacion	Fecha Activación	Fecha
masprox_fechaaprobacion	Fecha de Aprobación	Fecha
perdida	Pérdida	Númerico
estadocivil	Estado Civil	Texto
niveleduccion	Nivel de Educación	Texto
edadesembolso	Edad del cliente a la fecha de Desembolso	Númerico
genero	Género	Texto
indafiliacionIESS	Indica si el cliente es Afiliado al IESS	Binario
sueldo	Sueldo	Númerico
indcarros	Indica si el cliente tiene o no un auto	Binario
totalcarros	Número total de autos que tiene un cliente	Númerico
vehiculoanio	Año modelo del auto si lo posee	Númerico
numero propiedades	Número De Propiedades	Númerico
indcelulares	Indica si el cliente tiene Celular	Binario
indmedidor	Indica si el cliente tiene registrado un medidor de consumo de Energía Eléctrica	Binario
tipoplancelular	Tipo de Plan Celular (Prepago o Plan) si el cliente tiene celular	Texto
indfonovienda	Indica si el cliente tiene registrada una línea telefónica a su nombre	Binario
profesion	Profesión	Texto
indruc	Indica si el cliente tiene RUC	Binario
antiguedadRUC	Antigüedad RUC	Númerico
RUC	RUC	Binario
categoriaempresa	Categoría de la Empresa	Texto
actividadempresa	Actividad de la Empresa	Texto
subactividadempresa	Subactividad de la Empresa	Texto
numentidades	Número de Entidades	Númerico
inddeudacr	Indica si el cliente tiene deuda en la Central de Riesgos	Binario
inddeudaA	Indica si el cliente tiene deuda con calificación A en la Central de Riesgos	Binario
inddeudaBC	Indica si el cliente tiene deuda con calificación B o C en la Central de Riesgos	Binario
inddeudaDE	Indica si el cliente tiene deuda con calificación D o E en la Central de Riesgos	Binario
inddeudacomercial	Indica si el cliente tiene deuda de tipo Comercial en la Central de Riesgos	Binario
inddeudaconsumo	Indica si el cliente tiene deuda de tipo Consumo en la Central de Riesgos	Binario
inddeudaavivienda	Indica si el cliente tiene deuda de tipo Vivienda en la Central de Riesgos	Binario
inddeudaBC6M	Indica si el cliente tiene deuda con calificación B o C en la Central de Riesgos hace 6 meses	Binario
inddeudacr6M	Indica si el cliente tiene deuda en la Central de Riesgos hace 6 meses	Binario
inddeudaA6M	Indica si el cliente tiene deuda con calificación A en la Central de Riesgos hace 6 meses	Binario
inddeudaDE6M	Indica si el cliente tiene deuda con calificación D o E en la Central de Riesgos hace 6 meses	Binario
inddeudacomercial6M	Indica si el cliente tiene deuda de tipo Comercial en la Central de Riesgos hace 6 meses	Binario
inddeudaconsumo6M	Indica si el cliente tiene deuda de tipo Consumo en la Central de Riesgos hace 6 meses	Binario
inddeudaavivienda6M	Indica si el cliente tiene deuda de tipo Vivienda en la Central de Riesgos hace 6 meses	Binario
inddeudatitular	Indica si el cliente tiene registrada una deuda en la Central de Riesgo de Tipo Titular	Binario
inddeudasolidario	Indica si el cliente tiene registrada una deuda en la Central de Riesgo de Tipo Solidario	Binario
inddeudacodeudor	Indica si el cliente tiene registrada una deuda en la Central de Riesgo de Tipo Codeudor	Binario
saldoCR	Saldo del cliente en la Central de Riesgos	Númerico
saldoCR6M	Saldo del cliente en la Central de Riesgos hace 6 meses	Númerico

Tabla 2.9: Descripción de las variables en la base de datos

## 2.4.1. ANÁLISIS DESCRIPTIVO

El análisis descriptivo de los datos consiste en la aplicación de una serie de herramientas que permiten observar los datos, para obtener una impresión inicial de los mismos. De esta manera se busca una cierta familiarización con las bases de datos. Entre los objetivos de este análisis se pueden mencionar:

- Describir los datos e investigar la calidad de los mismos.
- Buscar posibles estructuras o comportamiento en los datos, sin realizar ninguna hipótesis matemática a cerca de la estructura de las variables.
- Obtener un conocimiento básico de los datos y de las relaciones entre las variables.

El análisis descriptivo de datos es necesario para la posterior aplicación de técnicas estadísticas multivariantes que requieren grandes conjuntos de datos e hipótesis más complejas que las técnicas del análisis descriptivo, puesto que proporciona una perspectiva razonada para la interpretación de los resultados.

La aplicación del análisis descriptivo requiere la identificación de los tipos de variables con los que se va a trabajar, a continuación se resumen algunas clasificaciones para las variables:

**Clasificación por el nivel de medición o clasificación de Stevens:**

- Variables Nominales: son aquellas que indican diferencia en categoría, clase, calidad o tipo, por lo que poseen categorías representadas por nombres. Algunos ejemplos incluyen lugar de nacimiento, marca de automóvil, carrera académica, género, raza, etc.
- Variables Ordinales: designan categorías, que pueden ser clasificadas desde la menor hasta la mayor; es decir, que existe un orden natural entre las categorías. Las variables ordinales comunes incluyen, entre otras, clasificación de clase social (alta, media, baja), calidad de vivienda (estándar, insuficiente, en ruinas).
- Variables de Intervalo: estas variables identifican las diferencias en monto, cantidad, grado o distancia, y se les asignan puntuaciones numéricas muy útiles. Los ejemplos incluyen la temperatura (registrada al grado térmico más cercano) y el coeficiente de inteligencia (CI), que va desde cero hasta 200 puntos. Con las variables de intervalo, los intervalos o distancias entre las puntuaciones son las mismas entre cualquier par de puntos en la escala de medición.
- Variables de Razón: estas variables poseen las características de las variables de Intervalo y un punto cero verdadero, donde una puntuación cero significa ninguno. Peso, altura, edad, ingreso, distancia duración de tiempo y promedio son variables de razón.

**Clasificación por la periodicidad de medición:**

- Variables Longitudinales: se observan a lo largo del tiempo (series de tiempo)
- Variables Transversales o de Corte Transversal: se observan en un instante de tiempo dado.
- Variables de tipo Panel: caso mixto, considera variables longitudinales y variables transversales.

### Clasificación por la posibilidad de ser cuantificadas numéricamente

- Variables Cualitativas o no métricas: describen cualidades de un objeto.
- Variables Cuantitativas o métricas: utilizan unidades de medida.

Para la determinación de la estructura y la calidad de la información se debe investigar la posible presencia de errores, puntos atípicos y datos perdidos. A continuación se presentarán, a manera de ejemplo, los resultados obtenidos para algunas de las variables, la totalidad del análisis univariante se presentará en el Anexo II.

Para el análisis descriptivo de los datos se utilizará el programa estadístico SPSS 11.5, que ofrece el análisis de frecuencias, con sus respectivos gráficos, para las variables nominales y el análisis descriptivo para las variables cuantitativas.

La tabla 2.10 muestra el análisis de frecuencias de la variable ciudad, en el que se aprecia que las ciudades de mayor concentración son Guayaquil y Quito, mientras que las ciudades de Latacunga y Portoviejo tienen muy poca representatividad. No se encuentran valores perdidos.

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos				
AMBATO	1.090	3,5	3,5	3,5
CUENCA	1.292	4,2	4,2	7,7
GUAYAQUIL	16.905	54,4	54,4	62,0
IBARRA	1.320	4,2	4,2	66,3
LATACUNGA	183	0,6	0,6	66,9
LOJA	454	1,5	1,5	68,3
MANTA	1.621	5,2	5,2	73,6
PORTOVIEJO	108	0,3	0,3	73,9
QUITO	7.736	24,9	24,9	98,8
RIOBAMBA	378	1,2	1,2	100,0
Total	31.087	100,0	100,0	

Tabla 2.10: Análisis de frecuencias variable ciudad.

La

variable estado civil presenta 357 valores perdidos, existe concentración en las categorías estado civil casado y soltero, mientras que en las categorías viudo y separado se encuentran pocos casos en comparación al tamaño de la muestra, el detalle se puede apreciar en la tabla 2.11

		estadocivil			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	CASADO	12.020	38,7	39,1	39,1
	SEPARADO	285	0,9	0,9	40,0
	SOLTERO	17.584	56,6	57,2	97,3
	VIUDO	841	2,7	2,7	100,0
	Total	30.730	98,9	100,0	
Perdidos		357	1,1		
<b>Total</b>		<b>31.087</b>	<b>100,0</b>		

Tabla 2.11: Análisis de frecuencias variable Estado Civil.

La variable profesión, obtenida de la base del registro civil, fue un caso especial debido a la inmensa diversidad de categorías encontradas, se encontró que un 44.3% de la muestra de construcción son estudiantes, mientras que en las categorías restantes, 290 en total, los casos se distribuyen con poca representatividad.

		profesion		
		Frecuencia	Porcentaje	Porcentaje válido
Válidos	ESTUDIANTE	13.778	44,3	44,6
	QUEHACERES DOMESTICOS	3.752	12,1	12,1
	EMPLEADO	3.008	9,7	9,7
	JORNALERO	1.314	4,2	4,3
	EMPLEADO PRIVADO	1.025	3,3	3,3
	OBRERO	926	3,0	3,0
	COMERCIANTE	705	2,3	2,3
	CHOFER	649	2,1	2,1
	AGRICULTOR	542	1,7	1,8
	EMPLEADO PARTICULAR	486	1,6	1,6
	ALBAÑIL	302	1,0	1,0
	COSTURERA	262	0,8	0,8
	BACHILLER	238	0,8	0,8
	MILITAR	225	0,7	0,7
	MECANICO AUTOMOTRIZ	195	0,6	0,6
	PROFESOR EN GENERAL	195	0,6	0,6
	POLICIA	194	0,6	0,6
	EMPLEADO PUBLICO	185	0,6	0,6
	MECANICO	168	0,5	0,5
	AYUDANTE	147	0,5	0,5
OTRAS	2.596	8,35	8,40	
Total		30.892	99,4	100,0
Perdidos		1	0,6	
<b>Total</b>		<b>31.087</b>	<b>100,0</b>	

Tabla 2.12: Análisis de frecuencias variable Profesión

La tabla 2.13 resume los resultados obtenidos con el análisis descriptivo, para cada variable se presenta el porcentaje de valores perdidos, el porcentaje de observaciones de las categorías con poca representatividad y el nombre de las mismas. En la primera parte de la tabla se encuentran las variables con un porcentaje de datos perdidos despreciable, se dirá despreciable a un porcentaje de datos perdidos menor a 10% del total de observaciones de la variable correspondiente. Siguiendo en la tabla se puede apreciar las variables con categorías de poca representatividad, es decir aquellas categorías que se observaron menos del 5% del total de observaciones, además estas variables tienen un porcentaje de datos perdidos mínimo. Por último se encuentran las variables con un porcentaje alto de datos perdidos; es decir, aquellas que no aportan mayor información, se encontró incluso el caso de 100% de datos perdidos, como es el caso de la variable catastral número de propiedades, por lo que es imposible considerar en el estudio esta variable, el mismo caso se da con la variable año del vehículo.

En una primera aproximación se podrían considerar las variables con un porcentaje de datos perdidos despreciable, también las variables con categorías de alta representatividad y desechar las variables con alto porcentaje de datos perdidos.

Variable	% Perdidos	% Poca Representatividad	Categorías Poca Representatividad
buenosmalos	0%		
indafiliacionIESS	0%		
indcelulares	0%		
indeudaA	0%		
indeudaA6M	0%		
indeudaconsumo	0%		
indeudaconsumo6M	0%		
indeudacr	0%		
indeudacr6M	0%		
indeudatitular	0%		
indfonovivienda	0%		
indruc	0%		
profesion	1%		
genero	1,1%		
indeudaDE	0%	0,29%	1
indmedidor	0%	0,38%	1
indeudavivienda	0%	0,46%	1
totalcarros	0%	0,48%	>2 Carros
indeudavivienda6M	0%	0,51%	1
indeudaDE6M	0%	0,56%	1
indeudaBC6M	0%	0,8%	1
indeudaBC	0%	0,86%	1
indeudacomercial	0%	0,86%	1
numentidades	0%	0,9%	>2
desciudad	0%	0,94%	PORTOVIEJO, LATACUNGA
indeudacomercial6M	0%	1,00%	1
niveleduccion	1,1%	1,4%	ANALFABETO, ESPECIAL
tipoclienteorigen	0,19%	1,99%	R1 R4 R5
indeudacodeudor	0%	2,10%	1
estadocivil	1,2%	3,62%	SEPARADO, VIUDO
indeudasolidario	0%	3,80%	1
indcarros	0%	6,68%	1
tipoplancelular	84%		
actividadempresa	85,8%		
categoriaempresa	86,6%		
subactividadempresa	98%		
numeropropiedades	100%		
vehiculoanio	100%		

Tabla 2.13: Resumen del Análisis Exploratorio de las variables Nominales.

Para el análisis de las variables cuantitativas se encontraron los estadísticos descriptivos mínimo, máximo, media y desviación típica; las variables cuantitativas son solo cinco: sueldo, edad de desembolso, antigüedad del RUC, saldo crédito y saldo crédito a 6 meses. La tabla 2.14 resume los estadísticos calculados en el programa SPSS.

Estadísticos Descriptivos					
Variable	N	Mínimo	Máximo	Media	Desv. Estándar
sueldo	11.985	\$ 150,00	\$ 3.000,00	\$ 184,40	\$ 150,10
edadesembolso	31.073	18 años	55 años	34 años	11 años
antiguedadRUC	6.484	0,00 años	10,00 años	3,90 años	5,10 años
saldoCR	6.089	\$ 0,01	\$ 103.063,22	\$ 1.599,92	\$ 3.432,40
saldoCR6M	31.087	\$ 1,00	\$ 1,00	\$ 1,00	\$ 0,00

Tabla 2.14: Resumen del Análisis Univariante para las Variables Cuantitativas.



Como se puede observar en la tabla 2.14 la variable saldo en la central de riesgo hace 6 meses (saldoCR6M) está dañada, ya que es imposible que todos los clientes que tenían saldo hace 6 meses tenga tan solo 1 dólar, es decir a esta variable no la consideramos para los análisis posteriores.

## **2.5. SELECCIÓN DE LAS VARIABLES EXPLICATIVAS DEL MODELO**

Compete ahora seleccionar las variables que explican a la variable dependiente, es decir aquellas que definen a un cliente como bueno o malo. Es entonces parte crucial en la modelización, dado que las variables escogidas deben ser suficientes para explicar a la variable dependiente, pero no deben ser demasiadas tal que compliquen el modelo, es decir se aplica el principio de parsimonia.

### **2.5.1. ANÁLISIS DE CORRELACIONES**

Luego del estudio descriptivo de las variables, se debe realizar un análisis de las variables en conjunto para conocer cada uno de sus atributos, el objeto es identificar el grado con el que pueden contribuir en el modelo para discriminar entre buenos y malos. Además en esta etapa se reducirá el número de categorías por variable, esto último se explica por dos razones, cada una asociada al tipo de variable (categórica o continua):

- En el caso de las variables categóricas, contar con demasiados atributos puede agotar la muestra para cada respuesta y quitarle robustez al análisis; además pueden existir atributos que no contienen información necesaria para suponer que su razón buenos/malos se reproduce en el total poblacional. Los atributos deben ser representativos en la variable, según se indicó anteriormente se asumirá como no significativa a una representatividad del 5%.
- Dado que el objetivo del credit scoring es predecir el riesgo en lugar de explicarlo, para las variables continuas es preferible contar con un sistema

en el cual el riesgo no sea monótona en estas variables, siempre que esto permita una mejor predicción del mismo. Puesto que no siempre es posible decir que al crecer una variable continua el riesgo es mayor o menor, muchas veces el riesgo se relaciona a agrupaciones de la variable, por ejemplo, en la variable edad, se puede caracterizar a los clientes menores a 25 años como malos, a los clientes entre 25 y 55 años como buenos, y a los mayores a 65 nuevamente como malos; en tal caso no existe una relación creciente o decreciente definida. Las relaciones que permiten predecir el riesgo, en base a variables continuas, son en su mayoría como la explicada en el ejemplo anterior. Por ello es más eficiente identificar las posibles agrupaciones de la variable y la asignación bueno/malo esperada, a incluir directamente la variable continua.

Las variables explicativas a ser incluidas en el modelo discriminante deben estar altamente correlacionadas con la variable dependiente del modelo (Indicador B/M); por tal razón es necesario determinar si la variable indicador buenos/Malos depende o no de las variables explicativas, para tal fin se utilizarán árboles de decisión, conocidos también como algoritmos de partición recursiva.

Un árbol de decisión no utiliza un modelo estadístico formal y es más bien un algoritmo para clasificar utilizando particiones binarias sucesivas de los valores de una variable cada vez, esta partición recursiva se traduce en una organización jerárquica del espacio de representación que puede modelarse mediante una estructura de tipo árbol. Cada nodo interior contiene una pregunta sobre un atributo concreto (con un hijo por cada posible respuesta) y cada nodo hoja se refiere a una decisión (clasificación). Las decisiones a tomar en la construcción de los árboles de decisión, para este caso en específico son:

- La selección de las variables y de sus puntos de corte para hacer las divisiones, regla de partición.
- La asignación de los nodos terminales a categoría Bueno o Malo, decisión de asignación. Con esto se determina cual es el signo esperado de la variable.

La decisión de asignación se toma en función de la proporcionalidad de las categorías Bueno-Malo en el nodo inicial, normalmente se asigna el nodo como bueno, si la proporcionalidad se mantiene o se concentra en la categoría de Buenos; por ejemplo si la proporcionalidad inicial es 70/30, un nodo será considerado como bueno si existe una concentración en la categoría de Buenos mayor a la original.

Existen varios algoritmos para construir árboles de decisión, las diferencias principales entre los distintos algoritmos radican en las estrategias de poda o terminación y en la regla adoptada para particionar nodos. El algoritmo a utilizarse para la consecución del fin mencionado, será el algoritmo CHAID, las siglas CHAID corresponden al término inglés Chi- squared Automatic Interaction Detector (detector automático de interacciones mediante chi-cuadrado). Es una técnica estadística desarrollada por Kass (1980) muy eficaz para segmentar o generar árboles. El método CHAID, que utiliza la significación de una prueba estadística como criterio, evalúa todos los valores de una variable predictora potencial. Funde los valores considerados estadísticamente homogéneos respecto a la variable criterio y conserva inalterados todos los valores heterogéneos.

A continuación, el algoritmo selecciona la mejor variable predictora para formar la primera rama del árbol de decisión, de forma que cada nodo esté compuesto por un grupo de valores homogéneos de la variable seleccionada. Este proceso se repite hasta que el árbol se ha desarrollado por completo. La prueba estadística utilizada depende del nivel de medida de la variable criterio.

No se trata de un método binario, es decir, puede generar más de dos categorías en cualquier nivel del árbol. Por lo tanto tiende a crear un árbol más ancho que otros métodos. Así también, aprovecha los valores perdidos, tratándolos como una categoría válida individual.

CHAID funciona con todos los tipos de variables continuas y categóricas. Sin embargo, las variables predictoras continuas se categorizan automáticamente para el análisis. Algunas de las opciones mencionadas a continuación se pueden

definir mediante las opciones avanzadas de CHAID. Estas opciones incluyen la posibilidad de seleccionar la prueba de Chi-cuadrado de Pearson o la de la razón de verosimilitud, el nivel de  $\alpha_{merge}$  y el nivel  $\alpha_{split}$ . A continuación se describen, en una forma básica, los pasos del algoritmo:

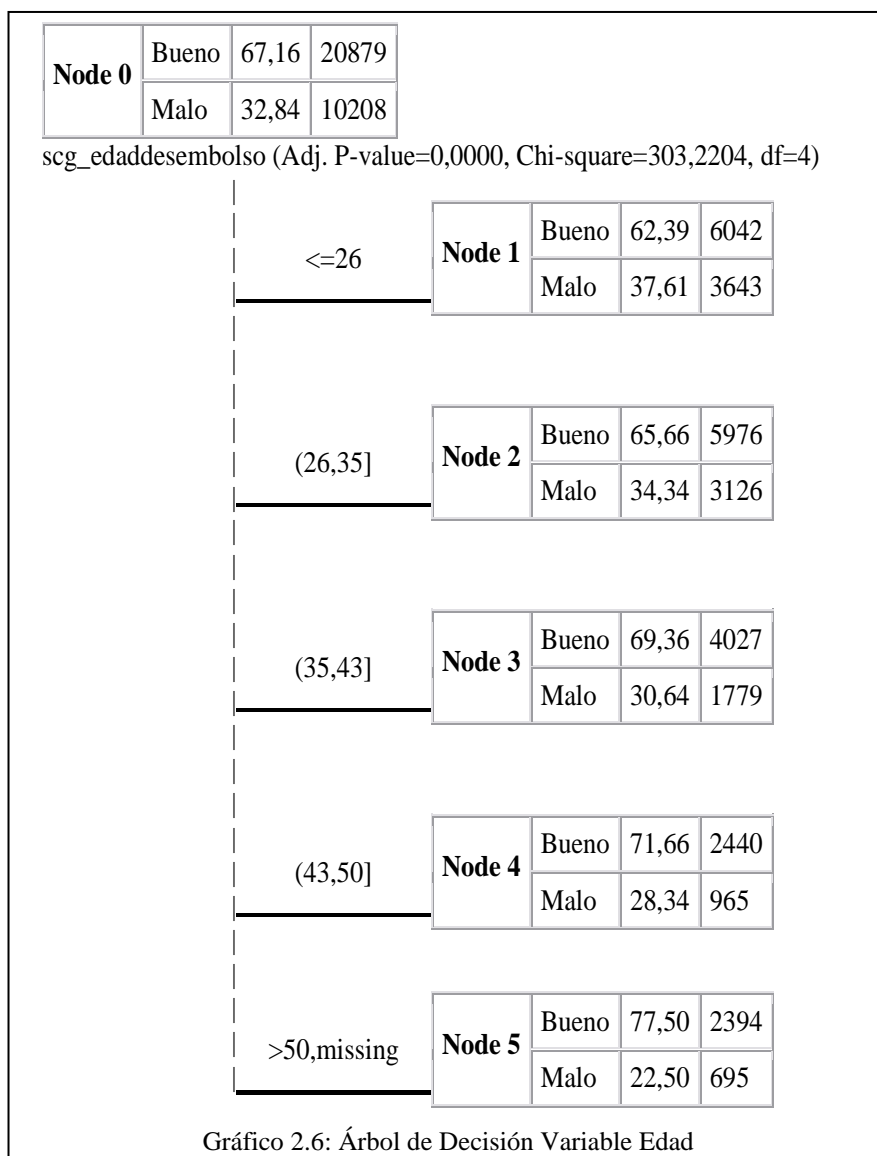
1. Para cada variable predictora X, buscar el par de categorías de X cuya diferencia presenta la menor significación (es decir, la que presente el mayor nivel crítico) respecto a la variable criterio Y. El método usado para calcular el nivel crítico depende del nivel de medida de Y.
  - Si Y es continua, utilizar la prueba F
  - Si Y es nominal, se forma una tabla de contingencia con las categorías de X como filas y las categorías de Y como columnas. Utilizar la prueba de chi-cuadrado de Pearson o la prueba de la razón de verosimilitud.
  - Si Y es ordinal, se ajusta un modelo de asociación de Y (Clogg y Eliasson, 1987; Goodman, 1979 y Madgison, 1992). Utilizar la prueba de la razón de verosimilitud.
2. Para el par de categorías de X con el mayor nivel crítico, se compara el valor del nivel crítico con el nivel alfa preestablecido,  $\alpha_{merge}$ .
  - Si el nivel crítico es mayor que  $\alpha_{merge}$ , fundir este par en una sola categoría compuesta, como resultado, se forma un nuevo conjunto de categorías de X y el proceso se vuelve a iniciar desde el paso 1.
  - Si el nivel crítico es menor que  $\alpha_{merge}$  ir al paso 3.
3. Calcular el nivel crítico corregido para el conjunto de las categorías de X y las categorías de Y, mediante la corrección de Bonferroni correspondiente.
4. Seleccionar la variable predictora X cuyo nivel crítico corregido sea el menor (la que sea más significativa). Comparar su nivel crítico con el nivel alfa preestablecido  $\alpha_{split}$ .

- Si el nivel crítico preestablecido es menor o igual que  $\alpha_{split}$ , dividir el nodo conforme al número de categorías de X.
- Si el nivel crítico preestablecido es mayor que  $\alpha_{split}$ , no dividir el nodo. Este nodo es un nodo terminal.

5. Continuar con el proceso de desarrollo del árbol hasta que se cumpla una de las reglas de parada.

La variable criterio a usarse para la construcción de los árboles de decisión es la variable Tipo Cliente, en cada fase del análisis, CHAID realiza una prueba de independencia Chi Cuadrado entre la variable dependiente y la variable predictora escogida.

Se hallan varios programas para la construcción de árboles de decisión, se utilizará Answer Tree 3.1. En el gráfico 2.6 se presentan los resultados obtenidos para la variable Edad:



Se aprecia que el estadístico chi-cuadrado obtenido para la variable Edad es de 303.22, en consecuencia el valor p es 0, menor que 5%, por lo tanto se rechaza la hipótesis de independencia a un nivel del 95% de confianza.

Se observa que el número de clientes buenos va incrementándose a medida que se incrementa la edad en los grupos obtenidos; es decir, por ejemplo, los clientes mayores de 50 son mejores clientes que aquellos con edades entre 35 y 43 años, la relación encontrada indica que la variable edad es candidata para ser ingresada al modelo como variable continua, aunque también se podría utilizar los rangos de

edades encontrados para cruzar con otras variables y encontrar nichos de clientes bueno ó malos,

La partición incluso ha permitido discretisar la variable continua edad en 5 categorías, de tal manera que esté correlacionada con la variable Indicador B/M.

Los grupos obtenidos son los descritos en la tabla 2.15 a continuación:

Grupo Edad	Bueno	Malo
<=26	62,39%	37,61%
(26,35]	65,66%	34,34%
(35,43]	69,36%	30,64%
(43,50]	71,66%	28,34%
>50,perdidos	77,50%	22,50%

Tabla 2.15: Discretización Variable Edad

La edad debe ser parte de las variables explicativas del modelo, debido a que el Indicador B/M es dependiente de la misma.

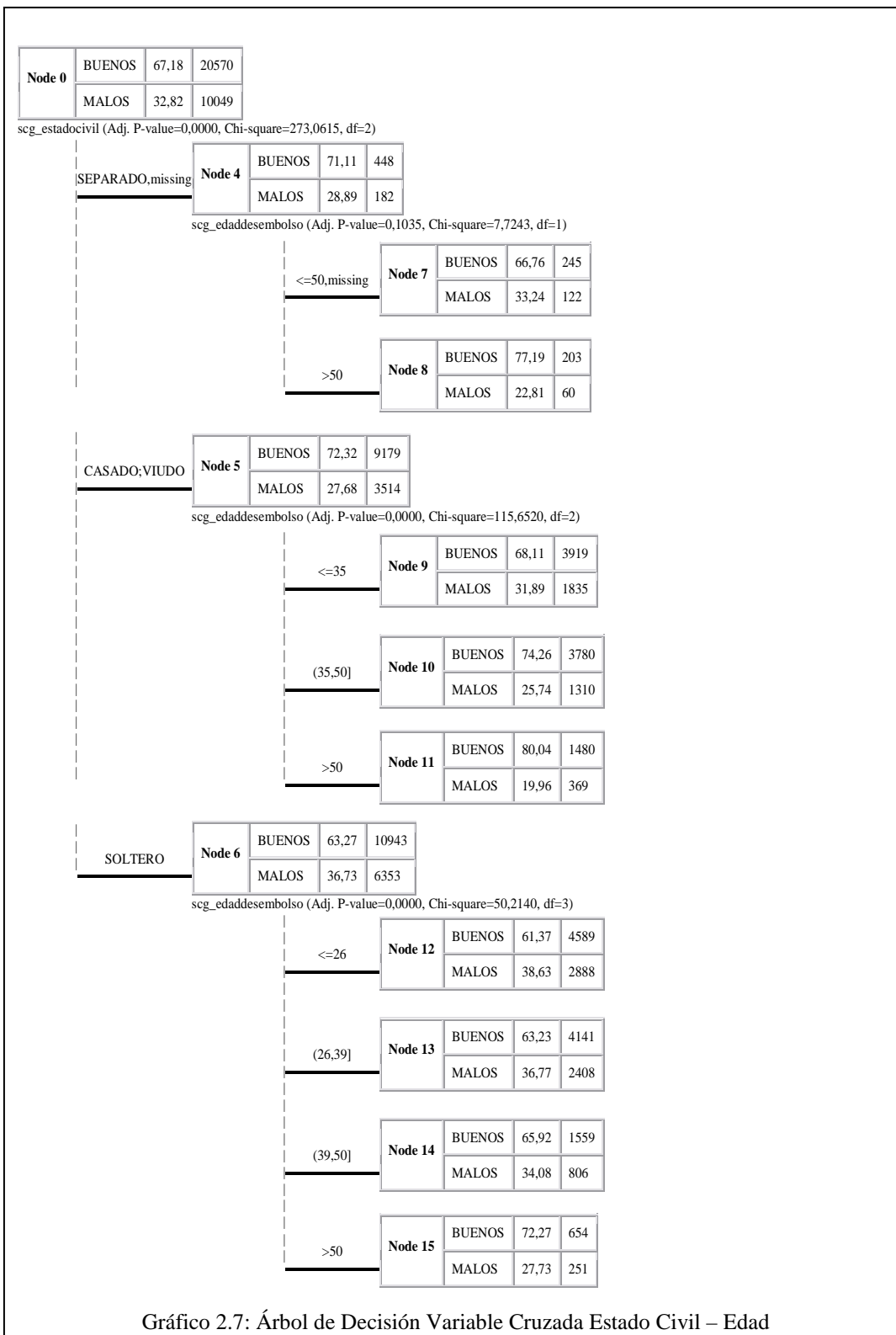
En ocasiones se piensa que la influencia de una de las variables sobre la probabilidad de que ocurra el hecho se modifica en función del valor de otra de las variables y es necesario incluir en el modelo una tercera que sea el producto de las anteriores. Estos son los conocidos como términos de interacción o cruces de variables que pueden incluir 2 o más variables. Los cruces de variables se introducen cuando existen razones para suponer que la influencia de una de las variables varía en función del valor que asume otra de las variables incluidas en el modelo; o sea, si la influencia de X1 varía en función del valor que toma X2, incluimos en el modelo un término que represente la interacción de X1 y X2.

Para el análisis se deben cruzar las variables, de tal manera que se identifiquen características de discriminación conjuntas, el cruce de variables permite no caer en la paradoja de Simpson<sup>19</sup>. Los árboles de decisión permiten conseguir los cruces de variables que estén más correlacionados con la variable indicador B/M, determinando si el cruce es independiente mediante la prueba Chi cuadrado.

<sup>19</sup> La paradoja de Simpson describe el cambio de sentido de una asociación entre dos variables cuando se controla el efecto de una tercera.

El gráfico 2.7 muestra el árbol de decisión para el cruce de las variables Estado Civil y Edad, el primer nivel de partición se lo hace con la variable Estado civil para el cual se obtiene un valor de 273,0615 para el estadístico Chi – Cuadrado; mientras que en el segundo nivel de partición se cruzan cada uno de los rangos de la variable Edad. Un cruce dependiente y significativo debe ser incluido en el modelo, así por ejemplo, en el grupo formado por casados y viudos mayores a 35 años la prueba de independencia Chi – Cuadrado se rechaza, además el porcentaje de clientes buenos que se concentra en dicho cruce es significativamente mayor con respecto al nodo inicial, también hay que observar que el número de clientes que pertenecen a este cruce son representativos con respecto a la población, razones por las cuales es factible incluir este cruce como variable candidata para el modelo. Por otra parte en el cruce de la categoría separado con la variable edad la prueba de independencia no se rechaza dado que el valor p es de 10.35%, y por último el cruce de la categoría soltero con la edad, si bien se rechaza la independencia se puede observar claramente que el número de clientes no es representativo en el nodo donde se aprecia una concentración de buenos clientes (solteros mayores a 50 años).





Las variables resultantes del análisis de árboles se presentan en la tabla 2.16:

Alias	Variables que intervienen	Descripción	Signo Esperado
V 1	Ciudad	Quito, Riobamba, Latacunga, Ambato, Ibarra, Portoviejo	Buenos
V 2	Edad	Variable continua	Buenos
V 3	Ind_Teléfono	Si tiene teléfono	Buenos
V 4	Ind_Carro	Si tiene carro	Buenos
V 5	Estado Civil - Nivel Educación	Soltero, Primaria o Analfabeto	Malos
V 6	Estado Civil - Nivel Educación - Ciudad	Soltero, Secundaria, Cuenca o Guayaquil	Malos
V 7	Estado Civil - Nivel Educación - Ciudad	Soltero, Secundaria, no Cuenca y no Guayaquil	Buenos
V 8	Estado Civil - Edad	Casado y mayor a 35 años	Buenos
V 9	Nivel Educación	Universitarios	Buenos
V 10	Nivel Educación - Edad	Primaria o Analfabeto, Menor a 35 años	Malos
V 11	Nivel Educación - Edad	Primaria o Analfabeto, Mayor a 43 años	Buenos
V 12	Nivel Educación - Profesión	Secundaria, no Obrero, no Estudiante, no Comerciante.	Buenos
V 13	IEES - Sueldo	Si IEES y sueldo mayor a 133 dólares	Buenos
V 14	IEES - Profesión	No IEES, no estudiante, no comerciante	Buenos
V 15	IEES - Profesión - Ciudad	No IEES, Estudiante o Comerciante, Guayaquil, Cuenca	Malos
V 16	IEES - Profesión - Ciudad	No IEES, estudiante o comerciante, no Guayaquil, no Cuenca	Buenos
V 17	Genero - Edad	Femenino menor o igual a 28 años	Malos
V 18	Genero - Edad	Femenino mayor a 35 años	Buenos
V 19	Ind_deuda_cr Deuda tipo A	Si Ind_deudacr=1 e Ind_deuda_A=1	Buenos
V 20	Ind_deuda_cr Ciudad	No deudacr, Quito, Ambato, Ibarra, Latacunga, Manta, Portoviejo.	Buenos
V 21	Ind_deuda titular - Ind_deuda consumo 6M	Si es titular y tiene deuda de consumo hace 6 meses	Buenos
V 22	Ind_RUC - Antigüedad RUC	Si tiene RUC y su antigüedad es mayor a 2 años	Buenos
V 23	Tipo de Cliente	Si el tipo de cliente es R2	Malos
V 24	Ind_deuda 6M	Ind_deuda 6M =1	Buenos
V 25	Ind_deuda 6M y Saldo cr hace 6M	Ind_deuda 6M =1 y saldo en CR hace 6 meses mayor \$1.311	Buenos
V 26	Estado Civil y Edad	Soltero de 39 años o menos	Malos
V 27	Estado Civil y Nivel de Educación	Casado, Viudo y Universitario o Especial	Buenos
V 28	Estado Civil y Nivel de Educación	Soltero y Universitario o Especial	Buenos
V 29	Género Edad	Masculino y de 26 años o menos	Malos
V 30	Género Edad	Femenino y R2	Buenos
V 31	Género Tipo Cliente	Masculino y R3	Buenos
V 32	Género Tipo Cliente	Masculino y R2	Malos
V 33	Género Tipo Cliente	Si IEES, no CR, R3 y Cuenca o Guayaquil	Buenos
V 34	IESS Ind_deudacr Tipo Cliente - Ciudad	Si IEES, no CR, R2	Malos
V 35	IESS Ind_deudacr Tipo Cliente	No CR, Cuenca o Guayaquil	Buenos
V 36	Ind_deudacr Ciudad	No Central Cuenca o Guayaquil	Malos
V 37	Ind_deudacr Ciudad	No Central Quito, Portoviejo, Ibarra, Ambato, Manta y Latacunga	Buenos
V 38	IESS	Si IEES	Buenos
V 39	RUC IEES Ciudad	No RUC, no IEES, Cuenca, Guayaquil, Loja	Malos
V 40	RUC IEES Ciudad	No RUC, no IEES Quito, Portoviejo, Ibarra, Ambato, Manta y Riobamba	Buenos
V 41	RUC IEES Tipo Cliente	No RUC, si IEES, R2	Buenos
V 42	Tipo Cliente Ciudad	R3, Guayaquil o Cuenca	Malos
V 43	Tipo Cliente Ciudad	R3, Quito, Portoviejo, Ibarra, Manta, Ambato y Riobamba	Buenos
V 44	Tipo Cliente IEES Ciudad	R3, no IEES, Cuenca, Guayaquil	Malos
V 45	Tipo Cliente IEES	R2, si IEES	Buenos
V 46	IESS Antigüedad Laboral	Si IEES y Antigüedad Laboral menor a un año	Malos
V 47	IESS Antigüedad Laboral	Si IEES y Antigüedad Laboral mayor o igual a 3 años	Buenos
V 48	IESS Categoría Empresa	Si IEES categoría Empresa pequeña	Buenos

Tabla 2.16 Variables Construidas

## 2.6. MODELIZACION

Se usará como modelo de discriminación la Regresión Logística, que se ha constituido como una de las técnicas estadísticas multivariadas de más frecuente uso en modelos de decisión.

En este caso la regresión lineal no es adecuada pues el modelo lineal presenta problemas cuando la variable dependiente es categórica, pues no se cumplen los supuestos básicos de la regresión lineal, por ejemplo:

- Los errores no siguen una distribución normal, sino que solo puede tomar tantos valores como categorías tenga la variable dependiente. Si bien este supuesto no resulta estrictamente necesario para aplicar mínimos cuadrados ordinarios (MCO), sí es fundamental a la hora de realizar

cualquier inferencia posterior sobre el modelo (intervalos de confianza para los parámetros estimados, contrastes de hipótesis, etc)

- Los errores no cumplen la hipótesis de heterocedasticidad.
- La variable dependiente puede tomar valores, por ejemplo 0 o 1, al ajustar un modelo lineal para explicar esta variable se puede obtener valores menores que cero (0) o mayores que uno (1).
- La determinación de las categorías sería dependiente de una combinación lineal de las variables explicativas, esto implica que el efecto marginal es constante para cualquiera de las variables explicativas, y por tanto la variación de la variable dependiente no dependería del valor actual de la variable explicativa. Este no es un supuesto adecuado para modelar una variable categórica.

En esta etapa de la construcción se mostrará la aplicación práctica del modelo, mientras que la explicación formal se presentará en el Anexo I.

### 2.6.1. CONSIDERACIONES PRELIMINARES

Se usará un modelo de regresión logística para explicar el comportamiento de la variable dependiente (Indicador Buenos Malos) mediante las variables explicativas determinadas con los árboles de decisión, (véase tabla 2.16), se notará el modelo logístico de la manera siguiente:

$$y_i^* = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \quad (2.1)$$

donde no se observa  $y_i^*$ . Ésta se conoce comúnmente como variable "latente".

Lo que se observa es una variable indicadora  $y_i$  definida por:

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases} \quad (2.2)$$

La diferencia entre el logit y otros modelos binarios difieren en la especificación de la distribución del término de error  $\varepsilon_i$ .

A partir de (2.2), observe que al multiplicar  $y_i^*$  por cualquier constante positiva,  $y_i$  no se modifica. Por lo tanto, si observamos  $y_i$ , podremos estimar las  $\beta_j$  de (2.1) y múltiplos positivos. Así, es costumbre suponer que  $Var(\varepsilon_i)=1$ . Esto fija la escala de  $y_i^*$ . Con base en las relaciones (2.1) y (2.2), obtenemos:

$$P_i = \Pr ob(y_i = 1) = \Pr ob\left[\varepsilon_i > -\left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij}\right)\right] = 1 + F\left[-\left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij}\right)\right]$$

donde F es la función de distribución acumulada de  $\varepsilon$ .

Si la distribución de  $\varepsilon$  es simétrica,  $1 - F(-W) = F(W)$ , es posible escribir:

$$P_i = -F\left[\left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij}\right)\right] \quad (2.3)$$

Puesto que las  $y_i$  observadas son solo realizaciones de un proceso binomial cuyas probabilidades están dadas por (2.3) y que varían de un ensayo a otro (dependiendo de  $x_{ij}$ , es posible escribir la función de verosimilitud con

$$L = \prod_{y_i=1} P_i \prod_{y_i=0} (1 - P_i) \quad (2.4)$$

La forma funcional para F en (2.3) dependerá de la suposición entorno al término de error  $\varepsilon$ . Si la distribución acumulada de los  $\varepsilon_i$  es logística, se tiene el llamado modelo Logit. En este caso,

$$F(W_i) = \frac{\exp(W_i)}{1 + \exp(W_i)} \quad (2.5)$$

Por lo tanto

$$\ln\left(\frac{P_i}{1 - P_i}\right) = \beta_o + \sum_{j=1}^k \beta_j x_{ij}$$

El primer miembro de esta ecuación se conoce como razón logarítmica de momios. Así, la razón logarítmica de momios es una función lineal de las variables explicativas.

Para una mejor interpretación de los coeficientes  $\beta_j$  (cualquiera de los coeficientes) es necesario referirse al concepto de riesgo relativo. El riesgo relativo de un suceso se define como la razón entre la probabilidad de que dicho suceso ocurra ( $p$ ) y la probabilidad de que no ocurra ( $1-p$ ), conocido también como *odds ratios* o *ratios de probabilidades*, los odds indican cuanto se modifican las probabilidades por unidad de cambio en las variables explicativas. El exponencial de los  $\beta_j$  se corresponde con el riesgo relativo, es decir, es una medida de la influencia de las variables  $x_{ij}$  sobre el riesgo de que ocurra ese hecho; suponiendo que el resto de las variables del modelo permanezcan constantes. Un intervalo de confianza para el exponencial de un  $\beta_j$  que contenga al 1 indica que la variable no tiene una influencia significativa en la ocurrencia del suceso y, por el contrario, valores más alejados de este indican una mayor influencia de la variable.<sup>20</sup> Una vez estimados los valores de los  $\beta_j$ , podemos determinar la probabilidad del suceso para distintos valores de los  $x_{ij}$ .

Las variables explicativas de índole nominal, con más de dos (2) categorías (politómicas), deben recibir un tratamiento especial para incluirlas en el modelo. Si

---

<sup>20</sup> Interpreting multiple logistic regression coefficients in prospective observational studies, Abbot RD, Carroll RJ.

estamos en presencia de una variable nominal con C categorías, debe incluirse en el modelo de regresión logística como variable categórica, de manera que a partir de ella se crean C-1 variables dicotómicas, llamadas *dummy* o ficticias. Al crear las variables *dummy* se debe precisar con cuál de las categorías de la variable original interesa comparar el resto y esa será la llamada categoría de referencia. En general, el exponencial del coeficiente correspondiente a una de las variables *dummy* estima la magnitud en que varía el riesgo de que ocurra el suceso.

Una de las ventajas de la regresión logística es que permite el uso de múltiples variables con relativamente pocos casos, sin embargo, hay que tener en cuenta algunas precauciones. Se ha sugerido que el número de sujetos para poder usar esta técnica estadística sin problemas debe ser superior a  $10(k+1)$  donde k es el número de variables explicativas; por tanto, si se introducen interacciones o variables *dummy*, el número de elementos en la muestra debe aumentar. Además se ha sugerido que si una de las variables dicotómicas (en especial si es la de respuesta) no tiene al menos 10 casos en cada uno de sus 2 valores posibles, entonces las estimaciones no son confiables.<sup>21</sup>

Para que la regresión logística tenga un sentido claro, tiene que existir una relación monótona entre las variables explicativas y la de respuesta, esto significa que el aumento de las unas se acompañe del aumento o la disminución aproximadamente constante de la otra, para todo el rango de valores estudiados.

## **2.6.2. EL PROBLEMA DE MUESTRAS DESPROPORCIONADAS**

El tratamiento de muestras desproporcionadas es propuesto por G.S. Maddala, según se cita textualmente en su libro Introducción a la Econometría: “En muchas aplicaciones de los modelos logit, probit o de probabilidad lineal, el número de observaciones en uno de los grupos es mucho más pequeño que en el otro. Por ejemplo, en un análisis de cartera, el número de clientes malos será mucho más

---

<sup>21</sup> Hsieh FY , “Sample size tables for logistic regression”, artículo para revista, Inglaterra, 1989.

pequeño que el de clientes buenos, esto se debe a razones obvias, pues si se diera tan solo la igualdad el banco quebraría. Por lo tanto, o tendremos que obtener un conjunto grande de datos o debemos muestrear los dos grupos con diferentes tamaños de muestra, para el caso anterior, se considerarían todos los clientes malos y se realizaría un muestreo sobre los clientes buenos con el fin de conseguir una proporción 50/50. En tales casos, surge la pregunta de cómo se deberían analizar los datos. Por lo general, se sugiere el uso de un modelo logit ponderado (o probit o de probabilidad lineal), similar al de los mínimos cuadrados ponderados. Ponderar las observaciones es el procedimiento correcto si existe un problema de heterocedasticidad. No hay razón por la cual las proporciones desiguales de muestreo deban provocar un problema de heterocedasticidad. Por lo tanto, es claro que ponderar las observaciones no es una solución correcta. Si el interés principal radica en examinar cuáles variables son significativas, es necesario no modificar el coeficiente estimado para el modelo logit. En lo que respecta a la estimación de los coeficientes de las variables explicativas, si se utiliza el modelo logit, los diferentes tamaños de muestra para los dos grupos no afectan los coeficientes”<sup>22</sup>

Para ver como el caso de muestras desproporcionadas es manejado, suponga que hay solo dos grupos, se puede extender al caso de varios grupos. Sea  $p_1$  y  $p_2$  las proporciones de las muestras de los dos grupos.

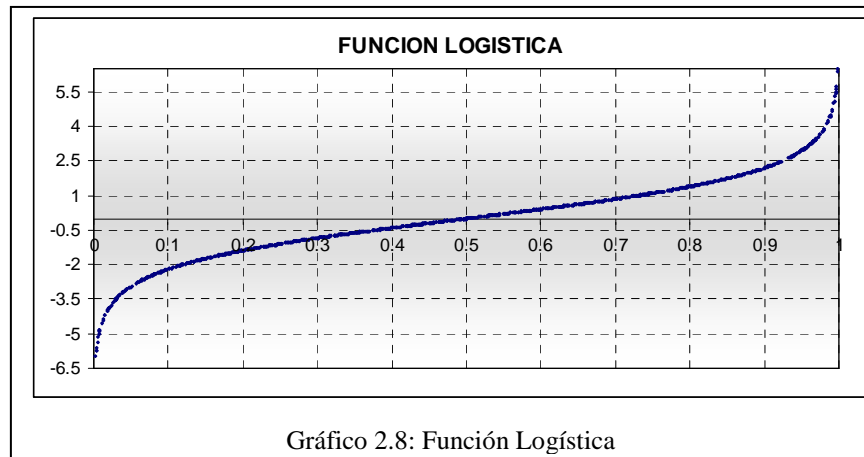
Puesto que  $p_1$  es la probabilidad de que una observación que pertenezca al primer grupo sea seleccionada, y  $p_2$  es la probabilidad de que una observación que pertenezca al segundo grupo sea seleccionada, cuando las muestras son desproporcionadas la función logística sufre un desplazamiento de la siguiente forma:

$$\ln\left(\frac{p}{1-p} * \frac{p_2}{p_1}\right) = \beta' x_i \quad (2.6)$$

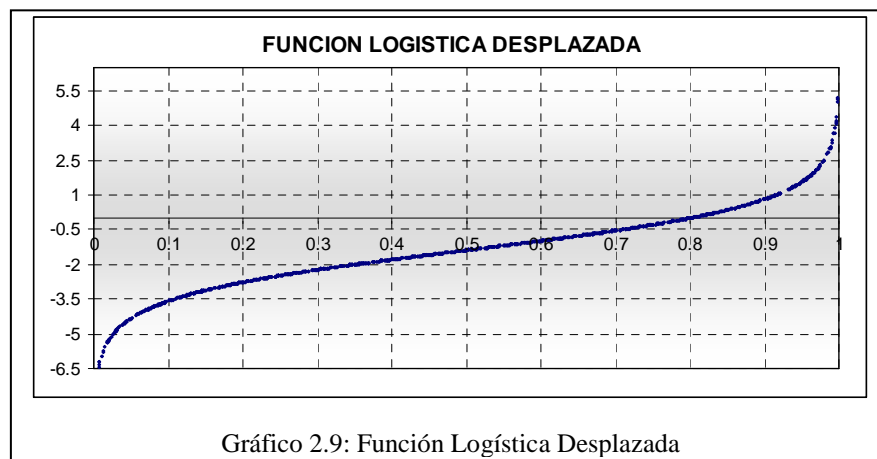
---

<sup>22</sup> G.S. Maddala, Introducción a la Econometría, segunda edición, Prentice-Hall Hispanoamericana, México, 1996, págs. 378 - 379

A continuación se presenta el gráfico de la función logística conocida,  $\ln\left(\frac{p}{1-p}\right)$ , se aprecia que en el gráfico la curva corta el eje de las abscisas en 0.5:



Al realizar el desplazamiento indicado, se verifica para los valores de  $p_1 = 0.2$  y  $p_2 = 0.8$ , que la curva corta el eje de las abscisas en 0.8, como se aprecia en el gráfico 2.9:



Se verifica además, que si  $p_1 = p_2$ , el cociente  $\frac{p_2}{p_1}$  es uno (1), haciendo uso de las propiedades de la función logarítmica el desplazamiento sería igual a cero, y la función logística no se altera.

A manera de aclaración, cabe indicar que las proporciones  $p_1$  y  $p_2$  no deben ser confundidas con el valor  $p$ , puesto que los primeros son valores dados, mientras



que el segundo es un valor observado, para este caso es la probabilidad de ser malo.

En las siguientes líneas se demostrará que la utilización de muestras desproporcionadas en la regresión logística, solo afecta al término constante del modelo:

Volviendo a 2.6, nótese lo siguiente:

$$\ln\left(\frac{p}{1-p} * \frac{P_2}{P_1}\right) = \beta' x_i$$

$$\frac{p}{1-p} * \frac{P_2}{P_1} = e^{\beta' x_i}$$

$$p + p\left(e^{\beta' x_i} * \frac{P_1}{P_2}\right) = e^{\beta' x_i} * \frac{P_1}{P_2}$$

$$p\left(1 + e^{\beta' x_i} * \frac{P_1}{P_2}\right) = e^{\beta' x_i} * \frac{P_1}{P_2}$$

$$p = \frac{e^{\beta' x_i} * \frac{P_1}{P_2}}{1 + e^{\beta' x_i} * \frac{P_1}{P_2}} ; \quad \text{de donde :}$$

Sea:

$y_i = 1$  Si la observación pertenece al primer grupo

$y_i = 0$  Caso contrario

$$p = \text{Pr ob}(Y_i = 1) = \frac{P_1 e^{\beta' x_i}}{P_2 + P_1 e^{\beta' x_i}} , \quad y;$$

$$1 - p = \text{Pr ob}(Y_i = 0) = \frac{P_2}{P_2 + P_1 e^{\beta' x_i}}$$

Definiendo  $p' = \frac{p_2}{p_1}$ . Entonces :

$$\Pr ob(Y_i = 1) = \frac{e^{\beta X_i}}{p' + e^{\beta X_i}}$$

$$\Pr ob(Y_i = 0) = \frac{p'}{p' + e^{\beta X_i}}$$

Sea  $\gamma = -\ln(p')$   $p' = e^{-\gamma}$ . Entonces se tiene :

$$\Pr ob(Y_i = 1) = \frac{e^{\gamma + \beta X_i}}{1 + e^{\gamma + \beta X_i}}$$

$$\Pr ob(Y_i = 0) = \frac{1}{1 + e^{\gamma + \beta X_i}}$$

Con una muestra equiproporcionada, se tiene  $p' = 1$  y  $\gamma = 0$ , se tiene entonces el modelo logístico usual. Así, el resultado de extraer muestras separadas de dos poblaciones es que solo el término constante de la función logística cambia. Este se reduce por  $\lambda = \ln p_1 - \ln p_2$ . Pero el resto de coeficientes ninguno se altera.

### 2.6.3. ESTIMACIÓN DEL MODELO

Para la estimación del modelo se escogió el método de pasos hacia Adelante de Wald, utilizando el software estadístico SPSS versión 11.5, la variable Indicador B/M se codificó de la manera siguiente:

Valor Original	Valor Codificado
BUENOS	0
MALOS	1

Tabla 2.17: Codificación Variable Dependiente

Esta codificación implica que  $P_i$  va a ser la probabilidad de ser mal cliente

### 2.6.4. ETAPA DE VALIDACIÓN

Para la validación estadística se realizarán varias pruebas al modelo:

- **Prueba Omnibus**

Esta prueba es útil para analizar la significancia conjunta de las componentes del modelo, donde:

$$H_o : \beta_o = \beta_1 = \beta_2 \dots = \beta_j = \dots = \beta_k = 0$$

$$H_1 : \exists j \in \{0,1,2,\dots,k\} \text{ tal que } \beta_j \neq 0$$

Pruebas omnibus sobre los coeficientes del modelo				
		Chi-cuadrado	gl	Sig.
Paso 19	Paso	5,31252939	1	0,0211726
	Bloque	1862,80753	19	0
	Modelo	1862,80753	19	0

Tabla 2.18: Resultados de la Prueba Omnibus

La tabla 2.18 muestra el resultado para la prueba omnibus obtenido en SPSS, puesto que el p-value en todos los casos es menor a 0.05 se rechaza la hipótesis nula.

- **Prueba de Hosmer – Lesmeshow**

Esta es una prueba de bondad de ajuste, las hipótesis a contrastarse son:

$$H_o : \text{El modelo está bien ajustado}$$

$$H_1 : \text{Falta ajuste en el modelo}$$

La tabla a continuación (2.19) muestra la salida en SPSS para la prueba de Hosmer y Lesmeshow:

Prueba de Hosmer y Lemeshow			
Paso	Chi-cuadrado	gl	Sig.
1	0	0	.
2	0,54803699	2	0,76031801
3	1,52989746	3	0,6753874
4	4,98949954	6	0,54516059
5	5,83689198	7	0,55891871
6	4,08584286	7	0,76983862
7	7,68523025	7	0,36116637
8	9,29056658	7	0,23246229
9	9,70891275	7	0,20567792
10	6,7329661	8	0,56569854
11	11,1167443	8	0,19517166
12	8,00659383	7	0,33201282
13	10,0205455	8	0,26358685
14	8,4548033	8	0,39034945
15	6,41178313	8	0,60120841
16	13,9196179	8	0,08388476
17	12,7899339	8	0,11928462
18	7,19636142	8	0,51560271
19	8,56462186	8	0,3803429

Tabla 2.19: Resultados de la Prueba

Esta prueba indica que los datos se encuentran bien modelados por medio de una regresión logística, es decir el modelo está bien calibrado, como el p-value es de 0.38 no se rechaza la hipótesis nula, es decir el modelo se encuentra bien ajustado.

- **Tabla de Clasificación**

También llamada matriz de confusión, es una forma sencilla de evaluar el ajuste del modelo de regresión logística, no es tan objetiva pero se usa como indicador de bondad de ajuste. Es una tabla sencilla de 2x2, en la cual se muestra la distribución de los clientes buenos y malos, es decir cuando  $y = 0$  e  $y = 1$  respectivamente, conjuntamente con la clasificación a cualquiera de las dos categorías según la probabilidad estimada. La interpretación se hace mediante el porcentaje de objetos bien clasificados, esto es, aquellos que mediante la probabilidad estimada permanecen en su respectiva categoría. También se interpreta el porcentaje de objetos mal clasificados, esto es, aquellos que mediante la probabilidad estimada se asignan a categorías diferentes de la cual fueron observados. Para construir la matriz de confusión en la tabla 2.20, se considera un punto de corte igual a 0.5, a partir del cual se realizan las clasificaciones de buenos y malos.

Observado	Pronosticado		
	BUENOS	MALOS	
BUENOS	19033	1167	94,22%
MALOS	8622	1271	12,85%
Porcentaje global			67,47%

El valor de corte es ,500

Tabla 2.20: Matriz de Confusión

En la tabla 2.20 se puede apreciar que el porcentaje de clientes bien clasificados es del 67.47 %, es importante tomar en cuenta que 12,85% de los clientes que son pronosticados malos en realidad son malos; esto no debe preocupar pues no se va a construir un scoring restrictivo, lo que se va a construir son perfiles de clientes, y probablemente la gran mayoría de estos clientes van a estar dentro de perfiles donde se les va a pedir más garantías y se dará mayor importancia a la información de la solicitud, previa verificación de la veracidad de la misma.

## 2.6.5. PRESENTACIÓN DE RESULTADOS

A continuación se presenta las variables resultantes de la última iteración (19), obtenida en SPSS, por el método paso a paso hacia adelante de Wald:

Paso 19(s)	Variables en la ecuación					
	B	E.T.	Wald	gl	Sig.	Exp(B)
V_1	-0,25812948	0,0526252	24,0595573	1	0,00	0,7724952
V_3	-0,46848621	0,04628222	102,46257	1	0,00	0,6259491
V_7	-0,45122165	0,06123391	54,2995087	1	0,00	0,63684967
V_8	0,20805705	0,03746464	30,8405285	1	0,00	1,23128341
V_12	-0,31069364	0,07749881	16,0721763	1	0,00	0,73293838
V_13	0,18843683	0,03923247	23,0696103	1	0,00	1,20736081
V_14	-0,26902268	0,04884432	30,3353992	1	0,00	0,76412592
V_16	-0,30384561	0,0379313	64,1668034	1	0,00	0,7379748
V_18	0,15124376	0,03619804	17,4576097	1	0,00	1,16328019
V_20	0,12484863	0,04821638	6,70468628	1	0,01	1,13297694
V_21	-0,21801305	0,04821652	20,4443503	1	0,00	0,80411495
V_24	-0,27959607	0,05468039	26,1456207	1	0,00	0,75608909
V_27	-0,10519783	0,0492992	4,55337907	1	0,03	0,90014643
V_32	-0,25079471	0,10923491	5,27125317	1	0,02	0,7781821
V_36	0,18273407	0,04353765	17,6160759	1	0,00	1,20049512
V_46	-0,19265465	0,07279951	7,00328932	1	0,01	0,82476676
V_51	0,20955575	0,04299256	23,7581502	1	0,00	1,23313012
V_52	-0,15915511	0,05050509	9,93049607	1	0,00	0,85286406
V_53	-0,12931437	0,0432182	8,95282889	1	0,00	0,87869768
Constante	-0,5052258	0,04725352	114,314873	1	0,00	0,60336932

Tabla 2.21: Iteración 19 del método paso a paso hacia adelante de Wald, todos los coeficientes son estadísticamente significativos.

- **Prueba de Wald**

Esta prueba se utiliza para evaluar la significancia estadística de cada variable explicativa o regresora.

Sin pérdida de generalidad seleccionamos  $\beta_j$  como el parámetro de interés.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

El estadístico de la prueba es:

$$W = \frac{\beta_j^2}{\sigma^2(\beta_j)}$$

Donde  $\sigma^2(\hat{\beta}_j)$  es el error estándar del parámetro  $\hat{\beta}_j$ , el estadístico Wald (W) sigue una distribución chi cuadrado con 1 grado de libertad y se tienen un nivel de significancia  $\alpha = 0.05$ .

En la tabla 2.21 se aprecia que las significancias de cada uno de los coeficientes es menor al 5%, por lo que se rechaza la hipótesis nula con un 95% de confianza.

- **Scorecard**

Puesto que los coeficientes son estadísticamente no nulos, se procederá a construir la scorecard<sup>23</sup>, teniendo en cuenta que hay que realizar la corrección a la constante por el problema de muestras desproporcionadas:

$$\tilde{\beta}_0 = \beta_0 - (\ln p_1 - \ln p_2)$$

$$\tilde{\beta}_0 = -0.5052 - (-0.72)$$

$$\tilde{\beta}_0 = 0.210$$

La tabla 2.22 presenta el scorecard resultante de la regresión, las variables predictoras con el valor de su respectivo coeficiente, adicionalmente se

---

<sup>23</sup> La scorecard es una tabla dónde se presentan las características (variables) resultantes de la aplicación del modelo logístico, con el valor respectivo del parámetro ajustado para representar el valor de score con el que aporta cada una de las características.

encuentra el signo esperado de la variable, el signo que se espera tenga la variable, se refiere a la clasificación como variable buena o mala especificada con los árboles de decisión.

Coeficiente	Descripción de la Variable	Signo Esperado
-0.258	Quito, Riobamba, Latacunga, Ambato, Ibarra, Manta o Portoviejo	Buenos
-0.468	Si tiene teléfono	Buenos
-0.451	Si tiene carro	Buenos
0.208	Soltero y Primaria o Analfabeto	Malos
-0.311	Universitarios	Buenos
0.188	Primaria o Analfabeto y Menor a 35 años	Malos
-0.269	Primaria o Analfabeto y Mayor a 43 años	Buenos
-0.304	Si IESS y Sueldo Mayor a 133 dólares	Buenos
0.151	No IEES, Estudiante o Comerciante y Guayaquil Cuenca	Malos
0.125	Femenino y menor o igual a 28 años	Malos
-0.218	Femenino y mayor a 35 años	Buenos
-0.280	No deuda CR y Quito Ambato Ibarra Latacunga Manta Portoviejo	Buenos
-0.105	Si el Tipo de Cliente es R2	Buenos
-0.251	Soltero y Universitario o Especial	Buenos
0.183	Masculino y R3	Malos
-0.193	No RUC, Si IESS y R2	Buenos
0.210	Si IESS y Antigüedad Laboral Menor igual a 1 años	Malos
-0.159	Si IESS y Antigüedad Laboral Mayor igual a 3 años	Buenos
-0.129	Si IESS Categoría Empresa Pequeña	Buenos
0.210	Constante	

Tabla 2.22: Scorecard

Otra forma de validar la calidad del modelo es comparando los signos esperados y los observados, es decir si se califica como buena a una variable (mayor concentración de clientes buenos) y puesto que se construye un modelo que estima la probabilidad de ser malo, el coeficiente que acompaña a la variable debería ser menor que cero, y cuando calificamos de mala una variable, el signo debería ser positivo.

La forma como se calculan los puntajes para cada uno de los clientes se detalla a continuación:

$$Z = \sum_{i=1}^n \beta_i x_i + \beta_0$$

Donde:

$\beta_i$ : estimador del coeficiente de la variable  $x_i$

$\beta_0$ : estimador de la constante del modelo

Ambos estimadores están detallados en la scorecard

$$p = \frac{\exp(z)}{1 + \exp(z)}$$

y de aquí se deduce:

$$\text{Score} = \lceil p \times 1000 \rceil$$

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

Donde  $p$  es la probabilidad de ser malo.

- **Interpretación de los Coeficientes**

Los coeficientes de la regresión logística, no son tan fáciles de interpretar como los de una regresión lineal, mientras los  $\beta_j$  son útiles para realizar las pruebas de validación del modelo, el  $\text{Exp}(\beta_j)$  es más fácil de interpretar.

$\text{Exp}(\beta_j)$ , representa el cambio en la razón de odds por cada cambio en una unidad de la variable  $x_j$ .

$\text{Exp}(\beta_{v_1}) = 0.772$ , indica que la razón de odds de los clientes cuyas ciudades de residencia son Quito, Riobamba, Latacunga, Ambato, Ibarra, Manta o Portoviejo es 0.772 veces mayor que la razón para los clientes que residen en otra ciudad, si el resto de variables se mantiene constante.

$\text{Exp}(\beta_{v_3}) = 0.626$ , indica que la razón de odds de los clientes que tienen teléfono es 0.626 veces mayor que la razón para los clientes que no lo tienen, si el resto de variables se mantiene constante.



$Exp(\hat{\beta}_{v_{-7}})=0.637$ , indica que la razón de odds de los clientes que tienen carro es 0.637 veces mayor que la razón para los clientes que no lo tienen, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{-8}})=1.231$ , indica que la razón de odds de los clientes que que son solteros y su nivel de educación es de primaria o analfabeto es 1.231 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{-12}})=0.733$ , indica que la razón de odds de los clientes que su nivel de educación es universitario es 0.733 veces mayor que la razón para los clientes que tienen cualquier otro nivel de educación, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{-13}})=1.207$ , indica que la razón de odds de los clientes que son menores a 35 años y su nivel de educación es de primaria o analfabeto es 1.207 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{-14}})=0.764$ , indica que la razón de odds de los clientes que son mayores a 43 años y su nivel de educación es de primaria o analfabeto es 0.764 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{-16}})=0.738$ , indica que la razón de odds de los clientes que que son afiliados al IESS y su sueldo es mayor a USD 133 es 0.738 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{-18}})=1.163$ , indica que la razón de odds de los clientes que no son afiliados al IESS, cuya actividad es comerciante o estudiante y residen en

Guayaquil y Cuenca es 1.163 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{20}}) = 1.133$ , indica que la razón de odds de los clientes que son mujeres y menor o igual a 28 años es 1.133 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{21}}) = 0.804$ , indica que la razón de odds de los clientes que son mujeres y mayor a 35 años es 0.804 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{24}}) = 0.756$ , indica que la razón de odds de los clientes que no registran deuda en la central de riesgos y residen en Quito, Latacunga, Ambato, Ibarra, Manta o Portoviejo es 0.756 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{27}}) = 0.900$ , indica que la razón de odds de los clientes que tienen la calificación R2 en la institución es 0.900 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{27}}) = 0.778$ , indica que la razón de odds de los clientes solteros y con nivel de educación universitaria o especial es 0.778 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\hat{\beta}_{v_{32}}) = 1.200$ , indica que la razón de odds de los clientes que son hombres y tienen calificación R3 en la institución es 1.200 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\beta_{v_{46}}) = 0.825$ , indica que la razón de odds de los clientes que no tienen RUC, están afiliados al IESS y tienen calificación R2 en la institución es 0.825 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\beta_{v_{51}}) = 1.233$ , indica que la razón de odds de los clientes que están afiliados al IESS y tienen antigüedad laboral menor a 1 año es 1.233 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\beta_{v_{52}}) = 0.853$ , indica que la razón de odds de los clientes que están afiliados al IESS y tienen antigüedad laboral mayor a 3 años es 0.853 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

$Exp(\beta_{v_{53}}) = 0.879$ , indica que la razón de odds de los clientes que están afiliados al IESS y la categoría de la empresa donde trabajan es pequeña es 0.879 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

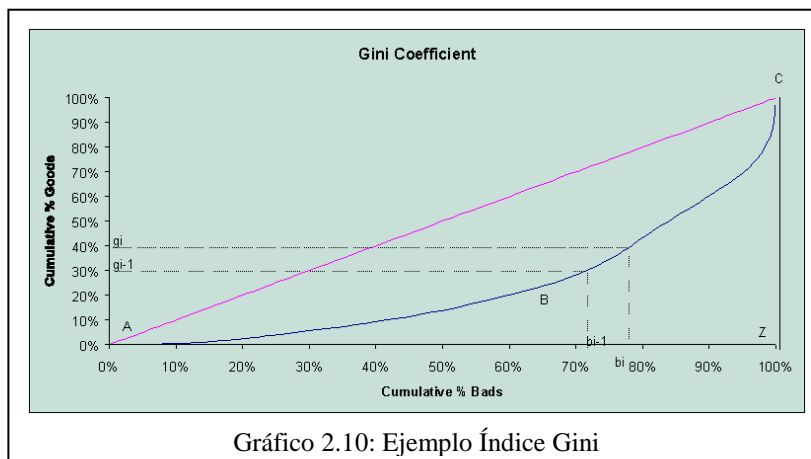
$Exp(\beta_{constante}) = 0.603$ , indica que la razón de odds de los clientes que están afiliados al IESS y la categoría de la empresa donde trabajan es pequeña es 0.603 veces mayor que la razón para los clientes que no tienen esta característica, si el resto de variables se mantiene constante.

#### 2.6.6. PRUEBAS DE EFICIENCIA

- **Coefficiente Gini:**

Mide la eficacia del modelo. Hace esto comparando el porcentaje de los buenos contra el porcentaje de los malos clientes para los mismos puntajes. Si el porcentaje de malos se traza contra el porcentaje buenos para una serie de

bandas de puntajes, el resultado es una curva (ABC). El coeficiente Gini es el área entre la curva (ABC) y la línea de la eficiencia nula (AC) establecida como un porcentaje del área del triángulo (ACZ).



El coeficiente Gini = área ABC / área del triángulo AZC

El coeficiente Gini se calcula como sigue:

$$Gini = 1 - \sum_{i=1}^n ((b_i - b_{i-1}) \times (g_i + g_{i-1})), \quad b_0 = g_0 = 0$$

Donde:

$g_i$  = % acumulado de buenos en un puntaje dado

$g_{i-1}$  = % acumulado de buenos en el puntaje anterior a  $g_i$

$b_i$  = % acumulado de los malos en un puntaje dado

$b_{i-1}$  = % acumulado de los malos en el puntaje anterior a  $b_i$

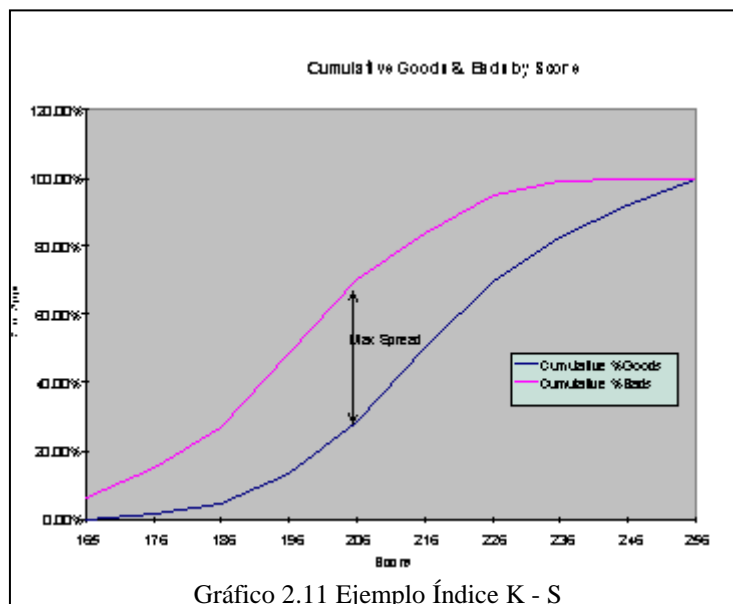
En la práctica  $i$  representa al percentil  $i$ , es decir se divide a toda la población en percentiles y para cada percentil se calcula el % acumulado de buenos y malos clientes.

Según Lisim<sup>24</sup>, de su experiencia empírica, un coeficiente Gini debe estar alrededor del 30% para un modelo de aprobación que incluye información de la solicitud, y estará generalmente más cercano al 60% para un modelo de comportamiento, dependiendo de las limitaciones sobre los datos.

<sup>24</sup> Consultora colombiana experta en temas de Scoring.

- **Coefficiente K-S:**

Es la medida de la máxima diferencia absoluta entre las distribuciones acumuladas de buenos y malos. El gráfico 2.11 presenta un ejemplo de el índice K – S.



Según Lisim el K-S debe ser mayor de 20%, esto para sistemas que se construyen con información externa e información de la solicitud.

Para nuestro modelo se calculó el coeficiente Gini y el K-S obteniéndose los siguientes resultados:

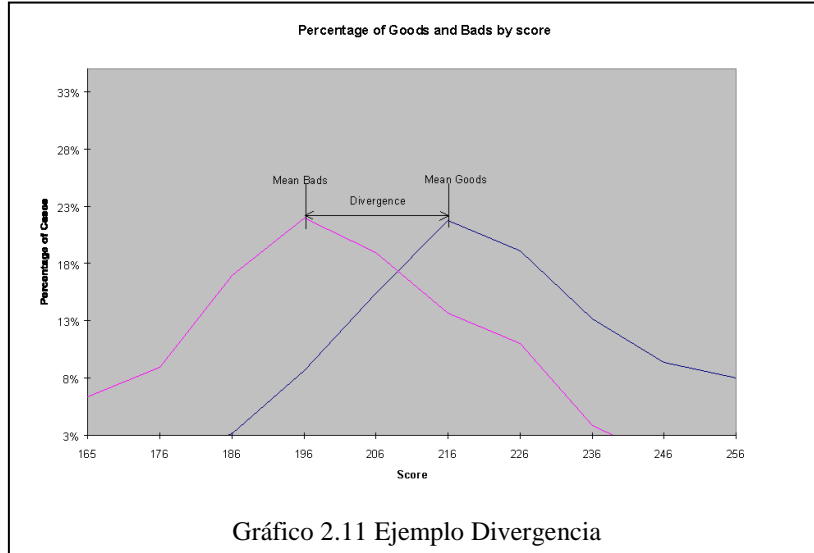
Rango Score	% Buenos	% de Buenos acumulados	% de Malos	% de Malos acumulados	Probabilidad de ser malo	Odds de ser Bueno	K-S	Gini
<304	13.00%	13.00%	4.10%	4.10%	13.39%	6 :1	8.9%	0.534%
305-366	12.53%	25.54%	6.14%	10.24%	19.34%	4 :1	15.3%	1.798%
367-419	11.16%	36.70%	7.62%	17.86%	25.07%	3 :1	18.8%	3.136%
420-463	13.06%	49.76%	11.23%	29.09%	29.63%	2 :1	20.7%	6.134%
464-496	7.26%	57.02%	6.85%	35.94%	31.62%	2 :1	<b>21.1%</b>	4.720%
497-535	11.44%	68.46%	12.07%	48.01%	34.07%	2 :1	20.4%	9.605%
536-573	8.11%	76.57%	9.95%	57.96%	37.51%	2 :1	18.6%	8.599%
574-618	9.45%	86.02%	14.18%	72.14%	42.37%	1 :1	13.9%	12.289%
619-645	6.89%	92.91%	12.24%	84.38%	46.52%	1 :1	8.5%	10.786%
>646	7.09%	100.00%	15.62%	100.00%	51.90%	1 :1	0.0%	13.071%
Total	100.00%		100.00%					<b>29.3%</b>

Tabla 2.23: Cálculo del K-S y del Gini

- **La divergencia:**

Mide con eficacia la diferencia entre la media de las distribuciones de buenos y malos, y se calcula como sigue.

$$Divergencia = 2 \times \frac{(\bar{X}_{Buenos} - \bar{X}_{Malos})^2}{(Var_{Buenos} + Var_{Malos})}$$



Según Lisim la divergencia debe ser mayor de 0.95; si se construye el modelo incluyendo información de la solicitud.

Para nuestro modelo la divergencia calculada es de 93.89%, debe considerarse que se usó información totalmente externa a la institución financiera.

## 2.7. VALIDACIÓN DEL MODELO CON DATOS DE PRUEBA (CONTROL DEL MODELO)

Esta etapa consiste en probar o testear el modelo obtenido con la muestra de construcción, corriéndolo sobre la muestra de prueba o control, básicamente el objetivo que se desea conseguir es determinar la eficiencia del modelo sobre una muestra distinta. Los índices que se considerarán son los mismos que se realizaron para las pruebas de eficiencia, excepto la divergencia.

A continuación se presenta la tabla 2.24 con el cálculo del coeficiente K – S y el Gini en la muestra de testeo.

Rango Score	Bueno	% Buenos	% de Buenos acumulados	Malo	% de Malos	% de Malos acumulados	Probabilidad de ser malo	Odds de ser Bueno	K-S	Gini
<304	1368	13.62%	13.62%	238	4.83%	4.83%	14.82%	6 :1	8.8%	0.659%
305-366	1191	11.86%	25.48%	322	6.54%	11.38%	21.28%	4 :1	14.1%	1.922%
367-419	1096	10.91%	36.40%	372	7.56%	18.93%	25.34%	3 :1	17.5%	3.308%
420-463	1053	10.49%	46.88%	416	8.45%	27.38%	28.32%	3 :1	19.5%	4.856%
464-496	1058	10.54%	57.42%	496	10.08%	37.46%	31.92%	2 :1	<b>20.0%</b>	6.831%
497-535	953	9.49%	66.91%	492	9.99%	47.45%	34.05%	2 :1	19.5%	8.058%
536-573	959	9.55%	76.46%	580	11.78%	59.23%	37.69%	2 :1	17.2%	10.188%
574-618	921	9.17%	85.63%	659	13.39%	72.62%	41.71%	1 :1	13.0%	12.093%
619-645	721	7.18%	92.81%	599	12.17%	84.79%	45.38%	1 :1	8.0%	11.301%
>646	722	7.19%	100.00%	749	15.21%	100.00%	50.92%	1 :1	0.0%	13.286%
<b>Total</b>	<b>10042</b>	<b>100.00%</b>		<b>4923</b>	<b>100.00%</b>					<b>27.5%</b>

Tabla 2.24: Cálculo del K-S y del Gini en la muestra de testeo

Se observa que tanto el K – S como el Gini son menores a los obtenidos con la muestra de construcción, sin embargo se aprecia que la discriminación del modelo en la muestra de control es muy similar a la discriminación observada en la muestra de construcción.

## 2.8. CONSTRUCCION DE PERFILES DE CLIENTES

En este punto se presenta dos metodologías para encontrar puntos de corte, en base a los cuales es posible definir perfiles de clientes.

### 2.8.1. DETERMINACIÓN DEL PUNTO DE CORTE (CUTOFF) UTILIZANDO LA CURVA ROC

A cada cliente que pertenece a la muestra de construcción, el modelo le asigna un puntaje (Score), que va de 0 a 1000 puntos, el problema se centra ahora en determinar el puntaje sobre el cual se rechazan a los clientes, es decir en términos de probabilidad , sería la máxima probabilidad aceptada de que un cliente sea malo, por ejemplo si determinamos que la máxima probabilidad aceptada es 75% (750 puntos) y tenemos un cliente con probabilidad de ser malo mayor a 75% automáticamente se le negaría cualquier operación.

Para identificar un punto de corte adecuado, se puede utilizar la curva ROC (Receiver Operating characteristic curve – Curva de características operativas del receptor), su origen se dio en la estimación de errores en la transmisión de mensajes, actualmente es muy usada en estadística aplicada a la medicina. Es

llamada también diagrama de Lorentz En el eje de las abscisas se ubica la proporción de malos con score  $s$  y en el eje de las ordenadas la proporción de buenos con score  $s$ .

Otras definiciones relacionadas con la curva ROC, son la sensibilidad y la especificidad:

- Sensibilidad (Se): si los buenos son los casos que se desean detectar, es la proporción de buenos predichos como buenos.
- Especificidad (Sp): es la proporción de malos predichos como malos.

Para el modelo en estudio, donde los casos sobre el punto de corte (cutoff) son clasificados como malos, se tiene:  $Sp = 1 - P_M(s)$  y  $Se = P_B(s)$ , así la curva ROC sería un gráfico  $1 - Sp$  vs  $Se$ .

La curva ROC describe la propiedad de clasificación de la scorecard, cuando el punto de corte varía.

El score que maximiza el K – S, por ejemplo, corresponde al punto en la curva ROC cuya distancia horizontal desde el eje es máxima. Este es el punto C en el gráfico 2.12. De esto se sigue que el punto es  $(P_M(s), P_B(s))$ , donde:

$$p_M(s) = \frac{n_M(s)}{n_M} = \frac{\text{Clientes malos con score igual a } s \text{ en una muestra de tamaño } n}{\text{Clientes malos en la muestra de tamaño } n}$$

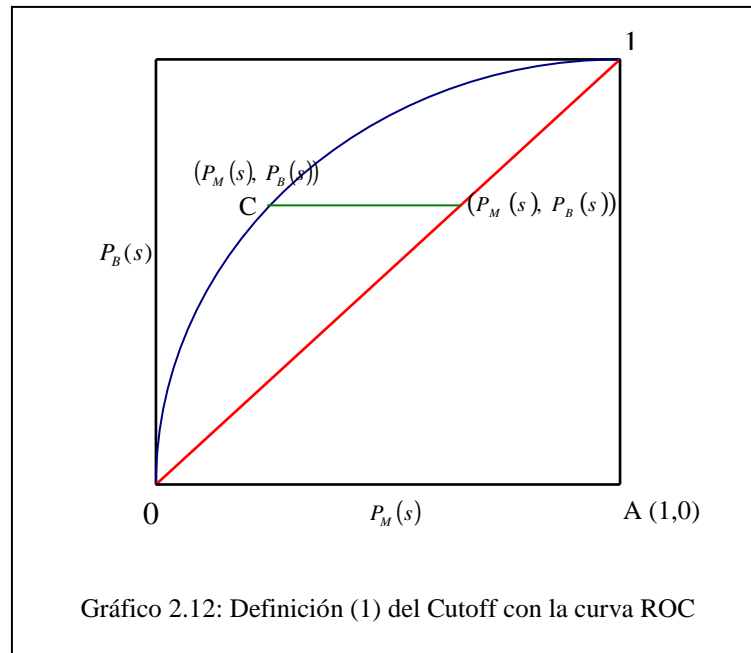
$$p_B(s) = \frac{n_B(s)}{n_B} = \frac{\text{Clientes buenos con score igual a } s \text{ en una muestra de tamaño } n}{\text{Clientes buenos en la muestra de tamaño } n}$$

y las distribuciones acumuladas se notarán por:

$$P_M(s) = \sum_{S \leq s} p_M(s) \quad \text{y} \quad P_B(s) = \sum_{S \leq s} p_B(s)$$

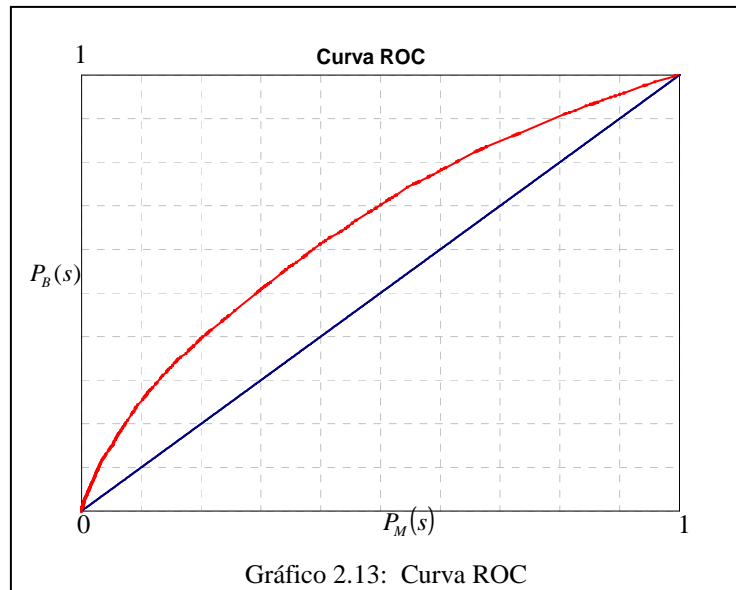
Entonces esta distancia horizontal es  $(P_M(s) - P_B(s))$ , la explicación de lo anterior se detalla en el gráfico.





En la práctica la construcción de la curva requiere la realización de los siguientes pasos:

1. Ordenar el score en forma ascendente.
2. Determinar el porcentaje de clientes buenos y malos que comparten el mismo score ( $p_B(s)$  y  $p_M(s)$ ).
3. Calcular el porcentaje acumulado ( $P_B(s)$  y  $P_M(s)$ ).
4. Calcular las diferencias ( $P_M(s) - P_B(s)$ ) entre porcentajes acumulados por score entre buenos y malos.
5. Identificar el score que produce la máxima diferencia.



Siguiendo los pasos descritos anteriormente, la máxima diferencia se produce en el score igual a 490 puntos.

Otra forma de encontrar un punto de corte adecuado es usar el concepto de pérdida por mala clasificación. Asumiendo que  $p_B$  y  $p_M$  son las proporciones de buenos y malos en la población, respectivamente, y que  $L$  y  $D$  son los costos por mala clasificación, donde  $L$  es la pérdida de rentabilidad por clasificar un bueno como malo, mientras que  $D$  es la pérdida esperada por clasificar un malo como bueno,  $D$  es considerablemente más grande que  $L$ . Así la pérdida esperada si se tiene un punto de corte (cut-off)  $s$  es:

$$EL = DP_M(s)p_M + L(1 - P_B(s))p_B$$

Para minimizar la función de pérdida esperada, supongamos una curva ROC  $f$ , esto es como minimizar:

$$EL = D x p_M + L(1 - f(x))p_B$$

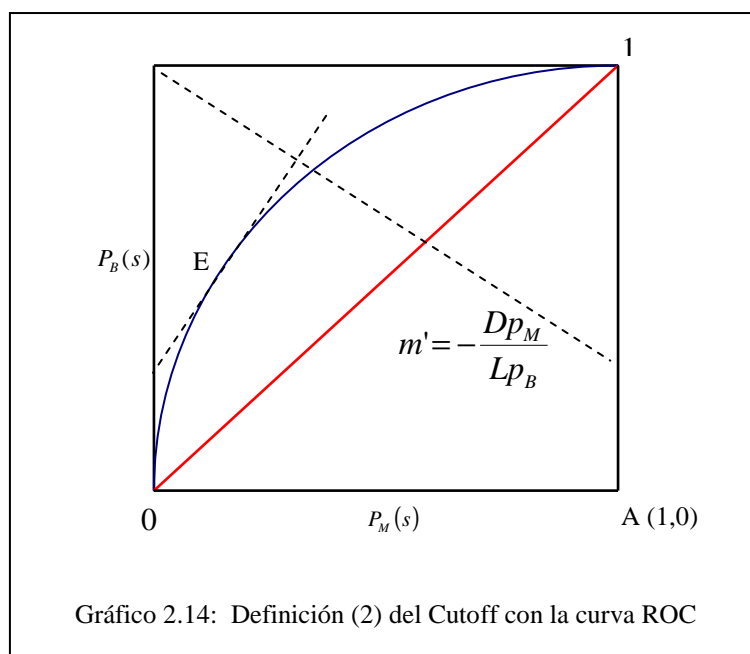
Lo cual ocurre cuando:  $Dp_M - L \frac{df(x)}{dx} p_B = 0$

Por lo tanto al mínimo, la pendiente de la recta tangente a la curva ROC es  $\frac{df(x)}{dx}$

, que satisface  $\frac{df(x)}{dx} = \frac{Dp_M}{Lp_B}$ , una manera de identificar este punto en la curva es

graficar la recata con pendiente  $-\frac{Dp_M}{Lp_B}$  que pase por el punto (0,1) y proyectar la

curva ROC en esta recta, el punto en la curva ROC que permite la proyección más cercana al punto (0,1) corresponde al punto de corte (cut-off) buscado  $E$ .



Para realizar el cálculo de este punto, se requiere la pérdida  $D$  y la rentabilidad  $L$ , una de las formas de medir la pérdida es dividiendo la provisión sobre el monto bruto, y la rentabilidad casi siempre se la mide por medio del cálculo de la TIR, una de las limitantes que se encontró en la base de datos proporcionada por la entidad bancaria fue la falta de variables para el cálculo de la TIR, mientras que sí se posee información para el cálculo de la pérdida, por lo que no es posible calcular el punto de corte por este método.

Algo importante de mencionar es que la medición de la rentabilidad usando la TIR representa un esfuerzo computacional demasiado pesado para una institución que maneja una cartera mediana o grande, por lo que se recomendaría buscar una forma alternativa de medir la rentabilidad.

## 2.8.2. DETERMINACIÓN DEL PUNTO DE CORTE UTILIZANDO MATRICES DE CONFUSIÓN.

Otra forma de determinar un punto de corte, es usar matrices de confusión y sobre esta extraer indicadores que nos pueden ayudar a elegir un punto de corte adecuado.

Los indicadores que se extrajeron de la matriz de confusión son:

- **Sensibilidad:** Es la proporción de malos, que son predichos como malos.
- **Especificidad:** Es la proporción de buenos que son predichos como buenos.
- **Razón de Aceptación Malos:** Es la cantidad de malos aceptados por modelo sobre el número de malos.
- **Razón de Rechazo:** Es la cantidad de clientes que el modelo clasifica como malos, sobre el número total de clientes.
- **Clasificación Correcta:** Es la suma de buenos clasificados como buenos y malos clasificados como malos sobre el total de clientes.
- **Clasificación Incorrecta:** Es la suma de buenos clasificados como malos y malos clasificados como buenos sobre el total de clientes.

Punto de Corte	Sensibilidad	Especificidad	Razón Acept. Malos	Razón de Rechazo	Clasificación Correcta	Clasificación Incorrecta
400	85.70%	31.45%	14.30%	77.13%	58.57%	41.43%
410	84.25%	33.81%	15.75%	75.22%	59.03%	40.97%
420	82.08%	36.48%	17.92%	72.80%	59.28%	40.72%
430	80.72%	38.28%	19.28%	71.22%	59.50%	40.50%
440	79.19%	40.12%	20.81%	69.53%	59.65%	40.35%
450	76.52%	42.97%	23.48%	66.77%	59.74%	40.26%
460	75.32%	44.32%	24.68%	65.50%	59.82%	40.18%
470	69.90%	50.82%	30.10%	59.54%	60.36%	39.64%
480	68.38%	52.24%	31.62%	58.07%	60.31%	39.69%
490	65.25%	55.55%	34.75%	54.85%	60.40%	39.60%
500	62.90%	58.01%	37.10%	52.45%	60.46%	39.54%
510	60.16%	61.04%	39.84%	49.56%	60.60%	39.40%
520	58.31%	62.67%	41.69%	47.82%	60.49%	39.51%
530	54.74%	65.96%	45.26%	44.39%	60.35%	39.65%
540	51.54%	68.90%	48.46%	41.32%	60.22%	39.78%
550	48.88%	71.29%	51.12%	38.80%	60.09%	39.91%
560	47.36%	72.59%	52.64%	37.39%	59.97%	40.03%
570	43.59%	75.46%	56.41%	34.06%	59.52%	40.48%
580	41.26%	77.08%	58.74%	32.09%	59.17%	40.83%
590	37.23%	80.23%	62.77%	28.50%	58.73%	41.27%
600	33.80%	82.97%	66.20%	25.42%	58.38%	41.62%

Tabla 2.25: Resultados Matrices de Confusión

En la tabla 2.25 se presentan los indicadores obtenidos en las matrices de confusión generadas para cada punto de corte indicado en la misma tabla.

Es posible tomar como punto de corte, aquel que permite la máxima clasificación correcta (60.60%), en este caso es el score 510, en este punto se logra una

sensibilidad de 60.16%, una especificidad de 61.04%, acepta el 39.84% de malos, rechaza al 49.56% de la población.

El score que maximiza el K-S (490), también tiene un buen indicador de clasificación correcta (60.40%) y un porcentaje de rechazo de 54.85%.

Como es posible notar, todos los métodos hasta ahora vistos para determinar el punto de corte tienen un indicador de rechazo bastante alto, lo que limita el crecimiento sostenido del negocio, pues talvez a una gran parte de clientes rechazados se los puede controlar por medio de garantías, limitaciones de cupos, cobranza preventiva, y otras políticas de gestión de riesgo de crédito. Además la decisión depende del criterio de pérdida con el que se maneja cada institución financiera, y del tipo de crédito que desea comercializar.

Este fenómeno se produce debido a la naturaleza del modelo, ya que es bastante difícil lograr una mejor precisión con información externa y más aún si esta presenta problemas de actualización o consistencia (no se utiliza información de la solicitud).

En la práctica la solución a este problema se basa en la segmentación del score por algún criterio que la gestión de riesgo de crédito trate de controlar, para nuestro caso vamos a tomar la pérdida como criterio.

### **2.8.3. CONSTRUCCIÓN DE PERFILES UTILIZANDO LA PÉRDIDA COMO INDICADOR DE CORTE**

Se llama perfiles a las agrupaciones de score, cada perfil se identificará por su pérdida esperada, lo que daría un mejor criterio para generar las políticas de control.

Inicialmente la construcción de los perfiles requirió de la segmentación del score en 23 grupos por percentiles, luego en cada grupo determinado se identificó la pérdida generada por los clientes en dichos rangos de score. Posteriormente se

procede a una nueva agrupación que permita determinar los porcentajes de pérdida deseados para la institución financiera, es decir que esta sea aceptable tal que el negocio continúe siendo rentable, de este modo se obtuvieron los siguientes perfiles por pérdida esperada:

Perfiles	Pérdida Generada	Proporción de Clientes
<b>P1</b>	4%	8%
<b>P2</b>	7%	17%
<b>P3</b>	12%	36%
<b>P4</b>	18%	21%
<b>P5</b>	24%	18%
<b>Total general</b>	<b>14%</b>	<b>100.00%</b>

Tabla 2.26: Perfiles por pérdida esperada

Dependiendo del porcentaje de pérdida que la institución establezca se podría excluir uno o más perfiles de aquellos que generan más pérdida; por ejemplo, podría excluirse el perfil P5 que genera en promedio un 24% de pérdida, esto significaría que el modelo scoring tiene una tasa de rechazo de aproximadamente 18%.

Cada perfil representa a los rangos de score siguientes:

Perfiles	Rango de Score
<b>P1</b>	0 - 293
<b>P2</b>	294 - 401
<b>P3</b>	402 - 538
<b>P4</b>	539 - 631
<b>P5</b>	> 631

Tabla 2.27: Rango de Score por perfil

Las políticas de selección y control incrementarán su rigurosidad a medida que se avanza en los perfiles. Hasta el perfil 3, se abarca un 61% de la población y una pérdida promedio del 10 %, por tanto en estos perfiles habrá más colocación, mientras que en los últimos existirá más limitación y control.

Cabe mencionar que la definición de los perfiles por pérdida, depende de el enfoque que cada institución financiera tiene, por ejemplo existirán entidades para

las cuales los perfiles definidos en este trabajo sean considerados de alto riesgo y otras que consideren lo contrario. Por tanto la construcción de los perfiles, en el modo indicado, dependerá del riesgo que cada institución esté dispuesta a asumir.

Las políticas de gestión del riesgo de crédito son muy variadas, su tratamiento puede ser muy amplio, pero no es el enfoque del presente estudio, en el siguiente capítulo se realizará un acercamiento a los temas que competen al monitoreo del sistema scoring construido.

## CAPITULO III

# METODOLOGÍA PARA EL MONITOREO DEL MODELO DE REGRESIÓN LOGÍSTICA

### 3.1. INTRODUCCIÓN

Un scoring debe ser constantemente monitoreado durante todo su tiempo de aplicación. Len McCahill, de la compañía Management Decision Systems, solía enfatizar este punto mostrando una diapositiva con un reloj en blanco en ella. Aplicar un score sin monitoreo, él decía, fue como usar un reloj sin manecillas.

El presente capítulo tiene el objetivo de desarrollar una metodología para monitorear el desempeño de un sistema scoring, se requiere determinar si el modelo está trabajando según se diseñó en la etapa de desarrollo, para lo cual se usarán medidas estadísticas que permitan valorar la eficiencia del modelo en el tiempo. La realización del monitoreo de los resultados y la eficiencia del modelo permitirá responder a las siguientes preguntas:

- ¿Se están calificando correctamente las aplicaciones?
- ¿Son las características de los clientes actuales las mismas que se tuvieron por ejemplo, hace un semestre, o comparado con la muestra de desarrollo?
- ¿La razón de aceptación actual es la misma que se tuvo anteriormente, o en la muestra de desarrollo?

Estas preguntas hacen referencia a una cadena de posibles eventos que podrían afectar el desempeño del modelo. Por ejemplo, la primera pregunta se refiere al posible problema que puede darse en la implementación de un Scoring en relación a la información utilizada. Cuando se construye un modelo scoring, además del equipo



de estadísticos interviene un equipo de extracción de la información, por lo general estos equipos interactúan mediante requerimientos, estos deben ser claros y específicos; así, si se pide al equipo de extracción de la información la variable edad, tiene que estar claro el hecho que se necesita el cálculo de la edad al momento del desembolso (que es el punto más cercano a la aplicación de un crédito), mas no al momento del requerimiento, puesto que la diferencia entre la fecha de desembolso y la fecha de desarrollo, como se explicó en el capítulo II, es tan amplia como se requiera madurez en la muestra. Una vez que se empieza a utilizar el modelo para calificar a los clientes, por lo general el proceso de extracción de la información pasa a ser trabajo de un nuevo equipo que genera la información de forma continua y automatizada (Equipo de Producción), este cambio de responsables en el proceso de extracción puede generar errores, puesto que el equipo de producción podría no considerar todos los parámetros bajo los cuales se realizó la extracción inicial, los parámetros a tomar en cuenta son: la actualización de la información y la forma de construir las variables, con respecto a la extracción inicial.

Para ilustrar las aseveraciones anteriores consideremos la variable (indica si el cliente está afiliado al IESS), asumiendo que en el desarrollo del modelo dicha variable se construyó con información desactualizada, la afiliación al IESS se podría considerar como verdadera si en alguno de los tres últimos meses disponibles al momento del desarrollo el cliente registra afiliación, una vez implementado el modelo puede considerarse la necesidad de una actualización de la información, y en el ejemplo de la variable del Ind\_IESS, la actualización de la información puede ocasionar una diferencia en los porcentajes de poblamiento de la variable (distribución de afiliados y no afiliados), si la variación es significativa se puede incurrir en un problema de calificación errada, puesto que el modelo fue construido bajo una distribución muy distinta en dicha variable.

El problema mencionado anteriormente, es susceptible de medirse estadísticamente en el sentido que afecta al modelo, puesto que un cambio de distribución en las variables actuales con respecto de las originales ocasionará que el sistema tenga

problemas de discriminación. Determinar las posibles causas para el cambio de distribución en las variables es tarea de cada analista, ya que es inherente a cada sistema y depende de su construcción.

La segunda pregunta planteada tiene relación con el efecto de la aplicación misma del sistema scoring, ya que dada la definición de perfiles de clientes, resultantes del modelo, se generan políticas y estrategias para el tratamiento de cada perfil, que pueden llevar a un cambio en las características de los clientes aprobados, cabe recalcar que la población sobre la que se realizará el monitoreo contiene información filtrada por el scoring; por ejemplo, como resultado del modelo se espera que los clientes casados sean buenos, por lo tanto se empezará a rechazar mayormente a los clientes solteros, pero seguramente entre los clientes casados que fueron aceptados durante el periodo a monitorear existen clientes casados que son malos, y por lo tanto el modelo por su modo de construcción no tiene poder de discriminación sobre estos clientes, el efecto es mayor en sistemas de iniciación donde la información es limitada. En el caso que el monitoreo indique este resultado se debe ajustar para mejorar el poder de discriminación.

El efecto de la aplicación del modelo scoring es la parte interna que hace que las características de los clientes varíen con respecto al periodo de desarrollo, es importante además conocer la incidencia de factores externos en dichos cambios, esto último tiene estrecha relación con la razón de aceptación de clientes, con lo cual se dará solución a la tercera pregunta que el monitoreo debe responder. Considérese el caso de que la razón de aceptación ha tendido a presentar una disminución en los últimos meses del periodo de monitoreo, es necesario entonces determinar el motivo que ocasiona dicha disminución, esto es, si se debe a un rechazo manual (Políticas) o a un Rechazo Automático (Modelo Scoring), en el caso de que se identifiquen fuentes de negocio interesantes para la institución se deben redefinir las políticas, el score o ambas.

Es importante indicar que un buen modelo no es el que más clientes rechaza, sino el que discrimina mejor; una baja tasa de aceptación influirá directamente en la rentabilidad de la institución financiera, es por ello imprescindible analizar las causas que originan una baja tasa de aceptación y buscar si entre los clientes rechazados, existen fuentes de negocio rentables para la institución.

Es decir, el monitoreo se utilizará para comparar el comportamiento inicial del modelo con el comportamiento actual del mismo. La frecuencia de su realización dependerá de las necesidades y recursos de la entidad financiera, por supuesto que a mayor frecuencia se podrá corregir a tiempo las anomalías que posiblemente se detecten, por otro lado realizar el monitoreo requiere de la utilización de los recursos de la entidad, por ejemplo, horas de extracción de información y análisis de resultados. De ahí la necesidad de construir una metodología de monitoreo, de tal manera que pueda automatizarse el proceso de generación de reportes y la labor mayor sea la del análisis.

El monitoreo debe aplicarse sobre una muestra escogida de la misma manera en que se seleccionó la ventana de muestreo para la construcción del modelo, es decir considerando madurez, estabilidad y representatividad, bajo los mismos criterios presentados en el capítulo anterior.

### **3.2. MAPA ESTRATEGICO DE MONITOREO**

En el gráfico 3.1 se presenta el mapa estratégico de monitoreo, en el cual se describe la metodología propuesta para la realización del monitoreo del sistema scoring.

La metodología que se propone está compuesta por varios reportes y análisis que permiten controlar el desempeño del modelo a nivel de los resultados que brinda y en relación a su comportamiento estadístico, la notación de los reportes empezará con la letra R, mientras que los análisis iniciaran con la letra A. La diferencia entre un

reporte y un análisis radica principalmente en que los primeros son susceptibles de ser automatizados, mientras que los segundos requieren de una decisión objetiva según el caso.

A continuación se explicará al detalle el mapa estratégico de monitoreo, cada una de las etapas y procesos que componen la metodología, luego de cada conjunto de reportes se presentará una ficha técnica de la etapa, en la cual se resumen los objetivos de cada reporte, la frecuencia propuesta para la realización de los reportes, la ventana de muestreo que se podría utilizar medida en meses, adicionalmente se nombran las variables que intervienen para elaborar los reportes implicados.

La metodología tiene dos (2) partes fundamentales, el monitoreo de los resultados por un lado, y por otro el monitoreo estructural del modelo; estas partes se describen en las siguientes hojas:

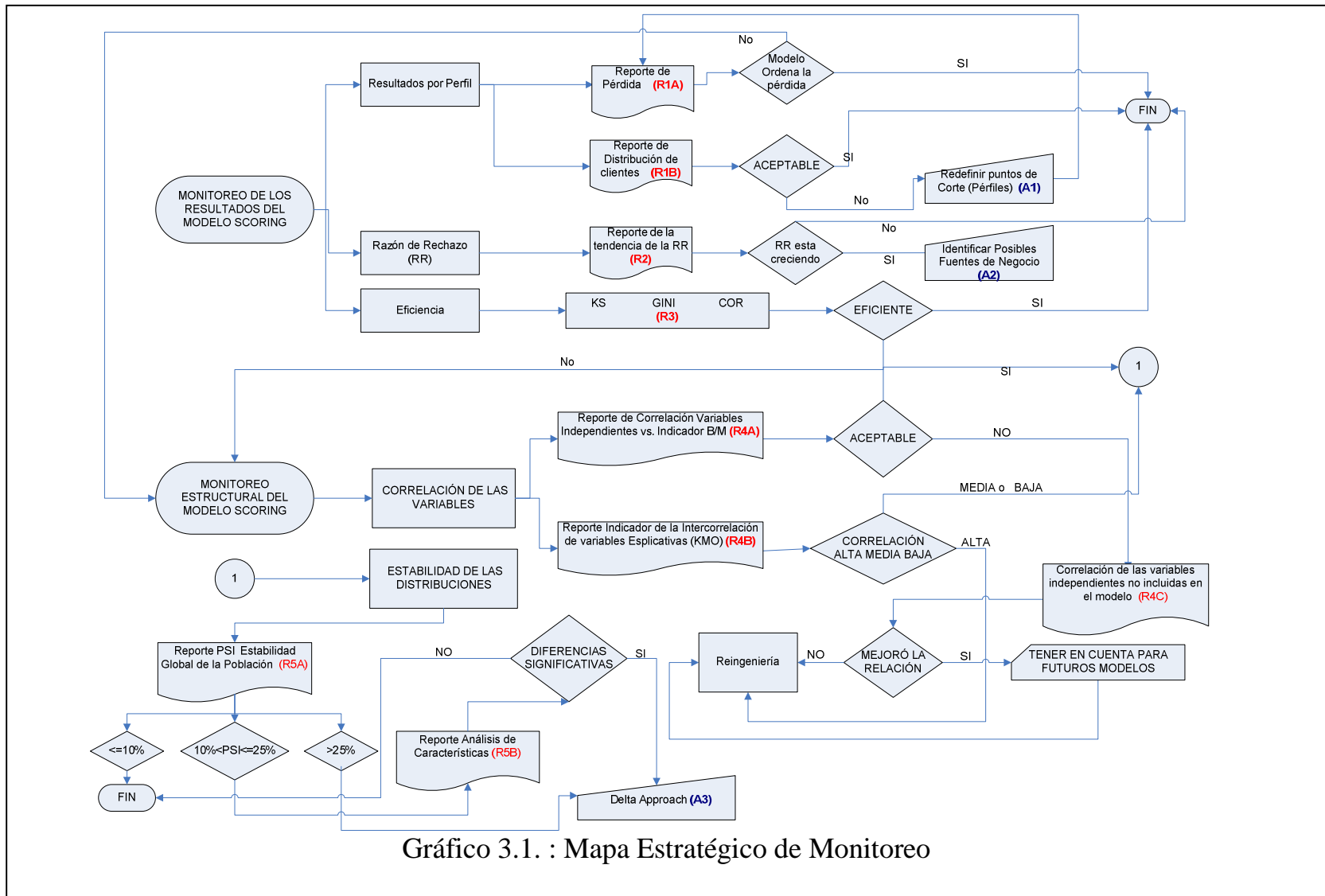


Gráfico 3.1. : Mapa Estratégico de Monitoreo

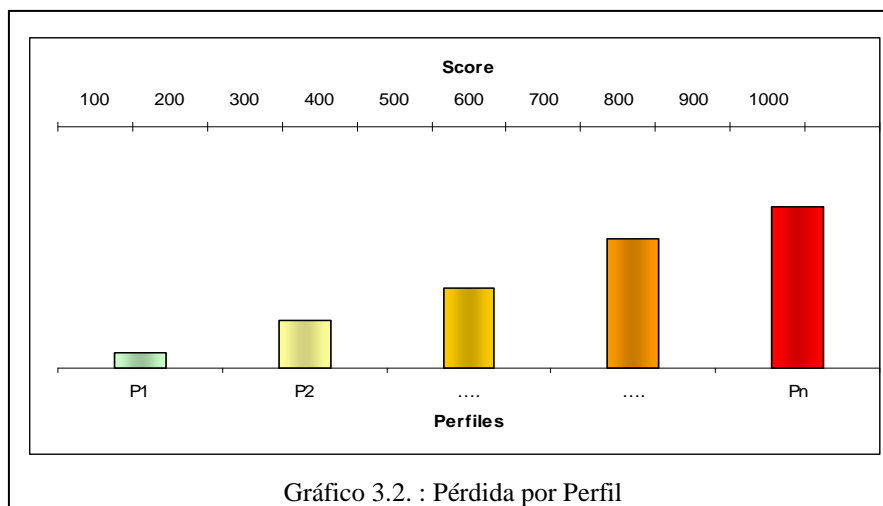
### **3.2.1. MONITOREO DE LOS RESULTADOS DEL MODELO SCORING:**

En este punto se medirá el impacto que ha causado el modelo en la institución financiera, además se concluirá acerca de la eficiencia de la clasificación que está realizando, así como la tendencia experimentada por la razón de rechazo, en el gráfico 3.1 Resultados por Perfil, Eficiencia y Razón de Rechazo, respectivamente. Sin importar los resultados que se obtengan en los reportes que conforman esta primera parte del monitoreo, sean estos positivos o negativos, es prudente analizar cada punto del grupo de reportes antes de pasar a la segunda parte del monitoreo, debido a que estos reportes son de alerta temprana y pueden ayudar al analista a encontrar posibles debilidades en el sistema.

#### **3.2.1.1. Resultados por Perfil:**

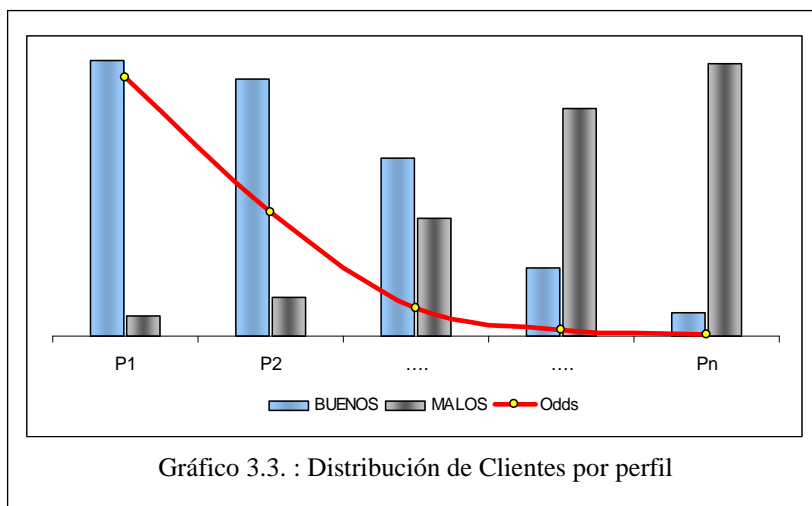
La aplicación de un modelo scoring tiene como consecuencia un cambio financiero, se esperaría en primer lugar controlar la pérdida; la eficiencia del modelo en la parte financiera se analizará considerando el desempeño de los perfiles de clientes en relación a la pérdida observada en el modelo de desarrollo. El método utilizado para calcular la pérdida ya fue explicado en el segundo capítulo. Los dos reportes que conforman este punto se detallan a continuación:

- **Análisis de Pérdida (Reporte R1A):** Se estudia la pérdida por Perfil; suponiendo por ejemplo, que la definición de orden de pérdida por perfil en el desarrollo se mantenga en la actualidad, entonces a medida que el score disminuye, la pérdida debería ir disminuyendo, es decir el modelo arroja la probabilidad de ser malo (1=Malo, 0=Bueno); como se muestra en el gráfico 3.2:



Si el supuesto se cumple, se dice que el modelo ordena bien la pérdida, además los valores de pérdida en cada rango no deberían variar significativamente de lo obtenido en el desarrollo del sistema, la conclusión en este punto es subjetiva, pues depende del analista así como de los requerimientos de cada institución. Si el modelo ordena la pérdida de forma estrictamente creciente y la pérdida no se ha deteriorado, el monitoreo continúa con la realización del Reporte R1B, en caso contrario se deberá estudiar la estructura modelo como tal (punto 3.2.2).

- **Análisis de distribución de clientes (Reporte R1B):** En esta parte se estudia la distribución de clientes por los puntos de corte establecidos en la construcción del modelo, es decir los porcentajes de clientes en cada perfil, si el porcentaje de clientes varía sustancialmente es necesario determinar si se debe a una falta de discriminación en el modelo o al efecto de las políticas aplicadas en cada perfil. Adicionalmente se analiza que la razón de buenos a malos (odds) en cada perfil sea lógica, es decir, que los mejores perfiles tengan mayoría de clientes buenos, mientras que a medida que disminuye la calidad de los perfiles aumente la cantidad de clientes malos;



Si las distribuciones de clientes varían significativamente es necesario redefinir los puntos de corte de cada perfil y realizar el análisis de pérdida en los perfiles redefinidos (Reporte R1A), en tal caso las pérdidas también se redefinen. La realización del reporte R1A en los perfiles redefinidos, dará la pauta para determinar si se debe o no pasar a la segunda parte del monitoreo. Si las distribuciones de clientes no cambian sustancialmente el análisis termina, y se pasa a la observación de la tendencia de la razón de rechazo (reporte R2).

GRUPO DE REPORTES	RESULTADOS POR PERFIL
<b>OBJETIVOS</b>	<b>R1A ANALISIS DE PERDIDA:</b> Observar cómo se ordena la pérdida en cada perfil y en general. <b>R1B ANALISIS DE DISTRIBUCION DE CLIENTES:</b> Observar la distribución de clientes buenos y malos en cada perfil y verificar la necesidad de redefinir los puntos de corte.
<b>FRECUENCIA</b>	Trimestral
<b>VENTANA DE MUESTREO</b>	Créditos emitidos hace 6 a 9 meses
<b>VARIABLES QUE INTERVIENEN</b>	Score, indicador de buenos y malos, rangos de atraso promedio

Ficha Técnica 3.1 Reportes Monitoreo: Resultados por Perfil

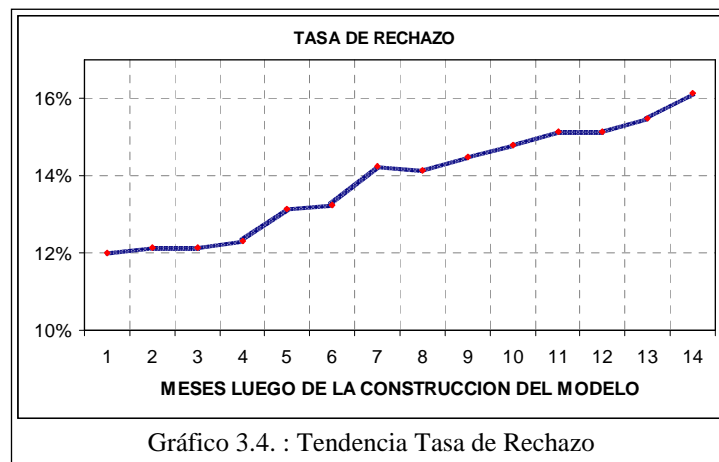
### 3.2.1.2. Tendencia de la Razón de Rechazo (Reporte R2):

El objetivo es obtener un gráfico de la tendencia de la razón de rechazo medida mensualmente durante el periodo a monitorear, el análisis de la razón de rechazo



tiene como finalidad determinar si existe o no un cambio en las características de los clientes que aplican para un crédito en la institución financiera.

En el gráfico 3.4 se presenta, a manera de ejemplo, una tasa de rechazo creciente, se aprecia con claridad la existencia de tendencia al crecimiento en los meses posteriores a la construcción del modelo.



Si la tendencia de la razón de rechazo es al crecimiento se decide identificar posibles fuentes de negocio (Análisis A1), es decir ubicar las características que el modelo no está considerando, pero que se presentan de forma adicional en los clientes que son aceptados por el modelo, por ejemplo, la combinación de la aplicación de políticas en conjunto con el modelo podrían generar la aceptación de un cliente profesional y de 25 años o menor, considerando que la aceptación de tal cliente se origina en sus otras características y no en la mencionada, si en el monitoreo se identifica como bueno, podría entonces determinarse como oportunidad de negocio a clientes que cumplen con la característica profesional y de 25 años o menor, lo cual se aplicaría como una política a considerar para la aprobación manual de un aplicante, sin la necesidad de ajustar el modelo scoring.

GRUPO DE REPORTES	TENDENCIA DE LA RAZON DE RECHAZO
OBJETIVOS	<b>R2 TENDENCIA DE LA RAZON DE RECHAZO:</b> Observar la tendencia de la razón de rechazo, si es creciente, identificar posibles fuentes de negocio.
FRECUENCIA	Trimestral
VENTANA DE MUESTREO	Créditos emitidos de 6 a 9 meses antes del monitoreo
VARIABLES QUE INTERVIENEN	Razón de Rechazo, clientes aceptados que poseen características malas en el scorecard.

Ficha Técnica 3.2 Reportes Monitoreo: Tendencia de la Razón de Rechazo

Si la razón de rechazo muestra una evolución normal en términos de que sea la deseada por la institución, el monitoreo continúa con el punto siguiente.

### 3.2.1.3. Eficiencia (Reporte R3):

Se mide la eficiencia del modelo, capacidad de discriminación, a través de los índices KS, Ginni y analizando la curva ROC, son los mismos índices calculados en el desarrollo del modelo, pero aplicados en la muestra de monitoreo. Si los índices no son aceptables tal que se pueda concluir que el modelo discrimina bien, se deben analizar las causas de la ineficiencia, estudiando la estructura del modelo discriminante (Punto 3.2.2).

GRUPO DE REPORTES	EFICIENCIA DEL MODELO
OBJETIVOS	<b>R3 EFICIENCIA:</b> Medir la eficiencia del modelo scoring, mediante el cálculo de los índices KS, Ginni, Área bajo la curva COR.
FRECUENCIA	Trimestral
VENTANA DE MUESTREO	Créditos emitidos de 6 a 9 meses antes del monitoreo
VARIABLES QUE INTERVIENEN	Score, indicador Buenos/Malos, Distribuciones de buenos y malos

Ficha Técnica 3.3 Reportes Monitoreo: Eficiencia del Modelo

Si todos los puntos descritos anteriormente llevan a conclusiones positivas, el monitoreo finaliza; de otro modo se requiere analizar con mayor énfasis el modelo discriminante, para hallar las causas que originan conclusiones negativas en el análisis de los resultados del modelo (Punto 3.2.2).

### **3.2.2. MONITOREO ESTRUCTURAL DEL MODELO SCORING:**

Esta es la segunda parte fundamental del monitoreo; su objetivo es encontrar la raíz de los problemas de eficiencia y ordenamiento de la pérdida, al analizar las variables que conforman el modelo, así como las variables que fueron excluidas en la construcción del mismo. Se analiza además la estabilidad de las distribuciones (buenos – malos) y de la población global de la muestra de desarrollo respecto a la muestra de monitoreo; lo último, debido a que el perfil de los nuevos aplicantes es susceptible de cambio en relación a la población de la que se tomó la muestra de construcción. En los próximos párrafos se detallan las etapas a considerarse en esta segunda parte del monitoreo:

#### **3.2.2.1. Correlación de las variables:**

El objetivo en esta etapa es medir la correlación de las variables independientes del modelo discriminante entre sí y con la variable dependiente.

- Correlación Variables Independientes en la scorecard vs Indicador B/M (Reportes R4A): Se requiere determinar si las variables incluidas en el modelo siguen correlacionadas con el indicador buenos-malos, con este fin se realiza una prueba de independencia Chi – Cuadrado de Pearson, esta prueba la hipótesis de independencia entre la variable dependiente y cada una de las variables explicativas – una explicación detallada de la prueba se presenta en el Anexo 1.

El fin de este reporte es determinar la existencia de correlación entre las variables explicativas y el indicador B/M, y adicionalmente verificar que el sentido de la correlación sea el mismo que se obtuvo en la muestra de construcción, esto se conoce como alineación, una variable estará alineada si cumple una de las dos afirmaciones siguientes, en el caso que el score estime la probabilidad de ser malo:

- Si una variable tiene coeficiente positivo y la razón de buenos a malos en la variable es menor o igual que la razón en la población.
- Si una variable tiene coeficiente negativo y la razón de buenos a malos en la variable es mayor o igual que la razón en la población.

En cualquier otra combinación de signo de coeficiente con la razón de buenos a malos respectiva, se dirá que la variable está desalineada.

La tabla a continuación presenta un posible resultado para el análisis de correlación de variables, con el resultado de la prueba de independencia, su respectivo valor crítico (p – value), el valor del coeficiente, la razón de buenos a malos en la variable (odds) y la valoración de alineación o desalineación de cada variable en la scorecard:

Variable	P-Value	Correlación	% Clientes	% Buenos	% Malos	Coefficiente	Odds	Alineación
EDAD	0,00	Correlacionado	100,00%	91,43%	8,57%	-0,03	10,67:1	Alineado
CARGAS	0,77	Independiente	100,00%	91,43%	8,57%	0,20	10,67:1	Alineado
FEME_UNI	0,00	Correlacionado	6,64%	95,09%	4,91%	-1,46	19,36:1	Alineado
CASADO	0,00	Correlacionado	20,99%	92,94%	7,06%	-0,59	13,17:1	Alineado
UNIONLIB	0,00	Correlacionado	1,91%	87,15%	12,85%	-0,73	6,78:1	Desalineado
MENOS1J	0,00	Correlacionado	6,44%	86,88%	13,12%	0,63	6,62:1	Alineado
CUENCA	0,90	Independiente	4,11%	91,55%	8,45%	1,06	10,83:1	Desalineado
LOJA	0,00	Correlacionado	4,01%	87,07%	12,93%	0,48	6,73:1	Alineado
QUITO	0,00	Correlacionado	42,39%	94,45%	5,55%	0,46	17,02:1	Desalineado
RIOBAMBA	0,64	Independiente	3,42%	90,92%	9,08%	0,34	10,02:1	Alineado
CALIF1	0,00	Correlacionado	33,56%	92,61%	7,39%	-1,02	12,53:1	Alineado
CALIF2	0,91	Independiente	53,10%	91,45%	8,55%	-0,82	10,7:1	Alineado
CALIF7	0,92	Independiente	0,33%	91,80%	8,20%	1,57	11,2:1	Desalineado
CALIF9	0,01	Correlacionado	1,86%	95,10%	4,90%	-0,82	19,41:1	Alineado
PROPIA	0,00	Correlacionado	26,10%	92,46%	7,54%	-0,47	12,27:1	Alineado
FAM_EMP	0,09	Independiente	40,62%	91,85%	8,15%	-0,34	11,27:1	Alineado
PROP_IND	0,68	Independiente	1,17%	92,20%	7,80%	0,68	11,82:1	Desalineado
NOIESS_EM	0,00	Correlacionado	19,21%	88,84%	11,16%	0,66	7,96:1	Alineado
COM_PRIM	0,01	Correlacionado	3,63%	88,81%	11,19%	0,61	7,93:1	Alineado
COM_SEC	0,01	Correlacionado	4,64%	89,17%	10,83%	0,88	8,23:1	Alineado
MASCULIN	0,00	Correlacionado	60,28%	90,63%	9,37%	-0,88	9,68:1	Desalineado
FEME_PRI	0,06	Independiente	5,52%	93,03%	6,97%	-1,20	13,35:1	Alineado
FEME_SEC	0,09	Independiente	27,01%	92,00%	8,00%	-1,15	11,5:1	Alineado

Tabla 3.1. : Correlación de Variables. Odds Población 10.67 : 1

Si existe correlación de todas las variables explicativas con la variable dependiente, se dice que las correlaciones son adecuadas, y por lo tanto se puede pasar a estudiar la intercorrelación de las variables explicativas del modelo.

Si una variable esta correlacionada y desalineada el análisis nos indica que la variable sigue explicando el comportamiento de los clientes en el tiempo pero que el signo del coeficiente que le acompaña cambió, es decir si la variable premiaba ahora tiene que castigar y viceversa, de ser este el caso el análisis que posteriormente se mencionara (Delta Approach) corregirá este problema. Para el caso de variables con estado independiente, se entiende que las variables ya no ayudan a explicar el futuro comportamiento del cliente por lo que se tiene que pensar en una posible reingeniería del modelo.

- Correlación Variables Independientes no incluidas vs Indicador B/M (Reportes R4B): si el reporte R4A muestra que existen variables en la scorecard que no están correlacionadas con el indicador B/M o si están desalineadas, es necesario conocer si en la muestra de monitoreo se encuentran variables no incluidas que estén correlacionadas con el indicador B/M. Las variables que estén correlacionadas se tendrán en cuenta para la realización de futuros modelos.

- Indicador de la Intercorrelación KMO (Reportes R4C): Mediante la medida de adecuación de la muestra, conocida como KMO (Kaiser-Meyer-Olkin), se requiere determinar el grado en que las variables incluidas en el modelo están correlacionadas entre sí. Este índice permite comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial. Un KMO medio o bajo, indica que la intercorrelación entre las variables no es grande y, por tanto, la aplicación de una regresión logística, sobre las variables, es apropiada; de ser ese el caso se pasa al análisis de estabilidad de las distribuciones (Punto 3.2.2.2). Por otro lado de encontrarse que la correlación está en un nivel alto es necesario hacer una reingeniería total del modelo.

Una alta correlación entre las variables independientes indica la existencia de un problema de multicolinealidad en la regresión logística, la

metodología propuesta no considera la posibilidad de corregir la multicolinealidad, puesto que en la práctica puede ser un trabajo largo y que no lleve a un resultado favorable en función de los objetivos del monitoreo, es necesario considerar que si la metodología llega al análisis de correlaciones es porque se detectaron problemas en los resultados por perfil del modelo, sobre todo en cuanto a eficiencia y ordenamiento de la pérdida se refiere.

Las correcciones al problema de multicolinealidad en la regresión logística no se abarcarán en este trabajo, pero es importante mencionar dos de los métodos de corrección. El primero conocido como “Penalidad Ridge” que elimina la degeneración numérica que causa la multicolinealidad, puesto que los coeficientes se hacen grandes e inestables bajo la existencia de multicolinealidad, un pequeño cambio en los datos produce coeficientes estimados muy diferentes, este método produce errores de clasificación más bajos que otros métodos pero con la debilidad que ninguno de los coeficientes de las variables incluidas se hace cero, lo que dificulta la interpretación. El segundo método se denomina “Penalidad Lasso”, el cual presenta estabilidad en los coeficientes y además por su estructura hay coeficientes que se anulan y por tanto su interpretación y aplicación es más factible.

GRUPO DE REPORTES	CORRELACION DE LAS VARIABLES
OBJETIVOS	<b>R4A PRUEBAS INDEPENDENCIA VARIABLES EXPLICATIVAS vs INDICADOR B/M:</b> Determinar las variables explicativas, incluidas en el modelo, para las cuales no se acepta la hipótesis de independencia de la variable indicador B/M. Determinar las variables de la scorecard que no están alineadas.
	<b>R4B CORRELACION DE LAS VARIABLES INDEPENDIENTES NO INCLUIDAS EN EL MODELO vs INDICADOR B/M:</b> Verificar si existen variables explicativas no incluidas en el modelo original, que en el monitoreo están correlacionadas con el indicador B/M
	<b>R4C INDICADOR DE LA INTERCORRELACION (KMO):</b> Medir la correlación entre las variables explicativas incluidas en el modelo.
FRECUENCIA	Trimestral
VENTANA DE MUESTREO	Créditos emitidos de 6 a 9 meses antes del monitoreo
VARIABLES QUE INTERVIENEN	Variables explicativas incluidas y no incluidas en el modelo, indicador de buenos y malos, Distribuciones de clientes y razón de buenos a malos en el desarrollo y el monitoreo

Ficha Técnica 3.4 Reportes Monitoreo: Correlación de las Variables

### 3.2.2.2. Estabilidad de las Distribuciones:

Los puntos siguientes se refieren a los posibles cambios de la población que afectan el desempeño del modelo, estas variaciones pueden ser causadas debido a que las personas que actualmente aplican para un crédito son de un tipo diferente a los clientes en la muestra de desarrollo. Adicionalmente a las razones presentadas en la introducción del capítulo, la diferencia puede ser ocasionada por un cambio en la estrategia de mercadeo de la institución, o se podrían haber cambiado las políticas de tratamiento solo en una parte de su población objetivo, por ejemplo la institución podría realizar una campaña para atraer estudiantes o dirigir una campaña solo para clientes exclusivos. Aun cuando se puede proyectar las consecuencias de la aplicación de una determinada estrategia de mercadeo, se dan casos en que la respuesta a la estrategia genera cambios que no se consideraron. Múltiples son las causas que originan cambios en la población y se podría nombrar algunas más, sin embargo lo que interesa a la metodología propuesta es proporcionar herramientas que permitan medir la estabilidad de la población, los reportes que se ocupan de este objetivo se tratan a continuación:

- PSI (Population stability index) Estabilidad Global de la Población (Reporte R5A): El análisis de estabilidad de la población registra la distribución de los clientes por rango de score; cada determinado tiempo. Este es un indicador temprano, pues permite reconocer potenciales problemas, puesto que muestra si el tipo de clientes que aplican actualmente para un crédito es el mismo que cuando se desarrolló el modelo. El supuesto es que la distribución por rango de score en la muestra de monitoreo comenzará a sesgarse de la muestra de desarrollo si la población que aplica para un crédito cambia. Por ejemplo, si el número de aplicantes más jóvenes, posiblemente más riesgosos, se incrementa debido a una campaña de venta, esto ocasionaría que el número de clientes en los scores más altos se incremente, lo que se traduce en un cambio de distribución.

El indicador utilizado en este análisis se conoce como PSI, Population Stability Index, el índice de estabilidad de la población, evalúa el cambio en la población en el transcurso del tiempo. Teóricamente este índice se basa en la entropía relativa utilizada en teoría de la información, denominada información de Kullback-Leibler (ver anexo 1). El PSI mide la separación (divergencia) entre las distribuciones de score del periodo de desarrollo y el periodo de validación. Si existen en el transcurso del tiempo cambios en la población, estos pueden resultar en cambios en los scores. Un valor pequeño del índice es representativo de una población estable, y esto sugiere que la población correspondiente al monitoreo es similar a la población usada para el desarrollo del scoring.

Estadísticamente el PSI se calcula por la fórmula siguiente:

$$PSI = \sum_{i=1}^n (O_i - E_i) * \ln\left(\frac{O_i}{E_i}\right)$$

Donde  $O_i$  es la frecuencia observada para la característica  $i$  y  $E_i$  es la frecuencia esperada para la característica  $i$ , los pasos para el cálculo del PSI se describen en los puntos a continuación:



- Se divide el score por deciles luego se calcula el porcentaje de malos, buenos y la distribución total por cada rango de score, estas serían las frecuencias observadas, las frecuencias esperadas se calculan de la misma forma pero estas provienen del periodo donde se desarrollo el scoring.
- Se aplica la fórmula y se decide si la población es estable o no para cada característica (malos, buenos o en general).

En el campo de aplicación de estos modelos, se suele usar una regla del pulgar, hasta cierto punto empírica, que dice que un PSI menor a 10% indica que la población actual es bastante similar a la población estudiada durante el desarrollo del modelo scoring. Un índice entre 10% y 25% sugiere que se debe hacer un análisis adicional, el análisis de características (Reporte R5B), para identificar las características específicas que producen inestabilidad en la población total. Un cambio mayor a 25% indica cambios significantes en la distribución global del score, por lo tanto sería necesario pasar directamente al delta approach que es un análisis más profundo (A3). La tabla 3.1 contiene datos que ejemplifica el cálculo del PSI para la población total (buenos y malos), si se divide al score original en deciles, se tiene que en cada uno de los rangos de score obtenidos en el desarrollo del modelo se ubica el 10% de la población, pero en los mismos rangos de score la distribución de la población del monitoreo es distinta:

Percentil	De	Hasta	% de Clientes Desarrollo (E)	% de Clientes Monitoreo (O)	(A) = (O- E)	(B) = Ln(O/E)	(A) * (B)
1	0	267	10%	12%	2%	19%	0,00
2	267	348	10%	14%	4%	37%	0,02
3	348	412	10%	13%	3%	30%	0,01
4	412	471	10%	13%	3%	26%	0,01
5	471	522	10%	10%	0%	3%	0,00
6	522	574	10%	10%	0%	1%	0,00
7	574	626	10%	8%	-2%	-23%	0,00
8	626	684	10%	7%	-3%	-29%	0,01
9	684	753	10%	6%	-4%	-46%	0,02
10	753	987	10%	5%	-5%	-70%	0,04
Total ----->>>			100%	100%			0,10

Tabla 3.2: Cálculo del PSI

El resultado en este caso (entre 10% y 25%) nos sugiere que se debe realizar el análisis de características.

- **Análisis de Características (Reporte R5B):** Si en el reporte anterior se obtiene un PSI mayor a 10% y menor o igual al 25%, es necesario determinar las variables que están influyendo en el cambio. El reporte de análisis de características toma cada atributo de una característica (variable explicativa) y considera las diferencias en las proporciones de cada atributo entre la muestra de desarrollo y la muestra actual. Entonces, en este reporte se calcula el efecto en el score por la característica determinada. La tabla 3.3 muestra el análisis de características para la variable edad, los atributos relacionados a esta característica son los rangos de edad presentados:

	Muestra Monitoreo %	Muestra Desarrollo %	Score	Diferencia % * Score
<b>&lt;=25 años</b>	12%	13%	668	-9
<b>&gt;=46 años</b>	21%	29%	738	-53
<b>26_45 años</b>	67%	58%	832	71
<b>Score Promedio</b>				9

Tabla 3.3.: Análisis de Características Edad

Así para la característica edad, en promedio el score se ha incrementado en 9 puntos. Se pueden encontrar otras características con score promedio más grande o menor que esta, el total de los cambios en todas las características de la scorecard determina el cambio total del score entre la muestra de desarrollo y la muestra de monitoreo.

El análisis de características puede ser realizado con dos (2) distintos objetivos, el primero para examinar individualmente cada una de las características en el scorecard y el segundo para hacer un análisis similar en el score total. Para las variables no incluidas en el modelo el análisis se reduce al cálculo del PSI, por lo tanto los reportes de estabilidad global de la población y el análisis de características son complementarios.

GRUPO DE REPORTES	ESTABILIDAD DE LAS DISTRIBUCIONES
OBJETIVOS	<p><b>R5A PSI REPORTE ESTABILIDAD DE LA POBLACION:</b> Determinar la estabilidad global de la población, mediante el cálculo del índice PSI.</p> <p><b>R5B ANALISIS DE CARACTERISTICAS:</b> Si el PSI indica un cambio en la población, el objetivo es reconocer de manera específica el efecto de cada característica en la inestabilidad de la población.</p>
FRECUENCIA	Trimestral
VENTANA DE MUESTREO	Créditos emitidos de 6 a 9 meses antes del monitoreo
VARIABLES QUE INTERVIENEN	Variables explicativas incluidas en el modelo, distribuciones de buenos, malos y odds, scorecard

Ficha Técnica 3.5 Reportes Monitoreo: Estabilidad de las distribuciones

Si el análisis de características indica que hay diferencias relevantes entre población actual y la población en el desarrollo del modelo, se debe pasar al Análisis A3, denominado Delta Approach.

### 3.2.2.3. Delta Approach (Enfoque Delta) (Análisis A3):

Todos los reportes y análisis anteriores se constituyen en alertas para el analista, generan dudas acerca del desempeño del modelo, pero ninguno presenta una solución a los posibles problemas encontrados, en este punto se exhibirá una sugerencia que permite llegar a una conclusión objetiva de la necesidad de una reingeniería total o un ajuste del modelo scoring, de optarse simplemente por el ajuste, esta herramienta tiene la ventaja de otorgar un valor estadísticamente adecuado para la rectificación del modelo.

El objetivo del delta approach es identificar dónde y por qué el modelo estuvo clasificando inapropiadamente, se constituye en una técnica que identifica las diferencias entre la forma en que el scorecard se comporta en la práctica y como fue previsto que se comporte. Puede ser usado para decidir si la scorecard es satisfactoria o necesita ser completamente reconstruida o si ajustar uno o dos scores es suficiente. Se debe considerar además que la técnica se aplica solo en sistemas basados en métodos de regresión, en este caso la scorecard original fue construida usando regresión logística.

A continuación se presenta la metodología sugerida por Lyn C. Thomas en su libro *Credit Scoring and its Applications*<sup>25</sup>, la cual se aplica a modelos que pronostican la probabilidad de ser bueno, sin embargo si se ha construido la regresión logística para determinar la probabilidad de ser malo, la herramienta aún puede ser aplicada, pero siempre y cuando se considere el cambio pertinente. El cambio referido consiste básicamente en multiplicar todos los coeficientes de la scorecard (incluyendo la constante) por menos uno (-1), continuación de demuestra este último hecho.

$$P(M) = 1 - P(B)$$

$$P(M) = \frac{e^z}{1 + e^z} \Rightarrow P(B) = 1 - \frac{e^z}{1 + e^z}$$

$$P(B) = \frac{1}{1 + e^z}, \text{ si se multiplica el numerador y denominador por } e^{-z}$$

$$P(B) = \frac{e^{-z}}{1 + e^{-z}}, \text{ donde } z = \beta_0 + \sum_{i=1}^n \beta_i X_i$$

Una vez demostrado la transformación del modelo, procederemos a la explicación formal del Delta Approach.

La scorecard original fue construida asumiendo un modelo en el cual:

$$P(B | \text{Score} = s) = F(s), \text{ donde:}$$

$\text{Score} = w_0 + w_1 x_1 + \dots + w_p x_p$  y  $F$  es una función monótona, tradicionalmente se desea probar si  $X_i = 1$ , es decir tiene el correcto score, a través de

$$P(B | s_1 < \text{Score} \leq s_2) = F(s_2) - F(s_1) = P(B | s_1 < \text{Score} \leq s_2 \text{ y } X_i = 1)$$

---

<sup>25</sup> Thomas C. Lyn, Edelman David B., Crook Jonathan, *Credit Scoring and Its Applications*, Delta Approach, pág 117

Hacer esto directamente puede consumir mucho tiempo puesto que existen muchas elecciones de  $s_1, s_2, i$  y valores para  $X_i$ . Un enfoque alternativo es explotar el conocimiento de  $F(s)$ , en la regresión logística se tiene:

$$F(s) = \frac{e^{\alpha + \beta s}}{1 + e^{\alpha + \beta s}}$$

entonces:

$$\log\left(\frac{P(B|Score = s)}{P(M|Score = s)}\right) = \log(Odds(B : M|s)) = \alpha + \beta s.$$

Lo que se desea es graficar el logaritmo de la actual razón de Buenos a Malos (Odds)  $\log(Odds(B : M|s))$  en diferentes rangos de score y para diferentes atributos de la característica en consideración. Supongamos que se está considerando la característica Dueño de número telefónico, S significa que un cliente tiene número telefónico y N significa que no lo tiene.

El gráfico 3.5 muestra las razones de Buenos a Malos para dicha característica cuando el score se ha dividido en cuatro rangos, se grafican entonces las razones de Buenos a Malos correspondientes al punto medio (score medio) de cada rango de score, si la scorecard fuera perfecta entonces todos los puntos estarían en la línea  $\alpha + \beta s$ . Para el ejemplo, los dueños de número telefónico tienen mejores razones de buenos a malos que las predichas en todos los rangos de score y los que no tienen número telefónico tienen peores razones en dos de los cuatro rangos de score.

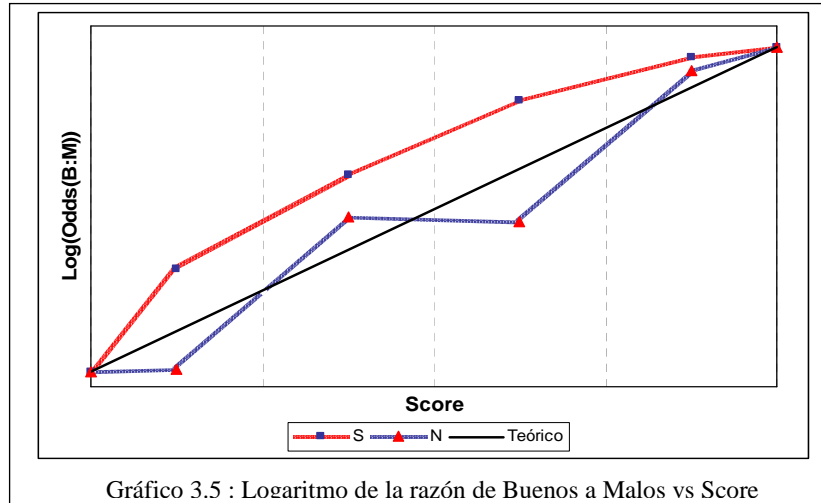


Gráfico 3.5 : Logaritmo de la razón de Buenos a Malos vs Score

Pero el delta approach busca definir una medida no en el  $\log(Odds(B:M|s))$  (eje Y), sino en el score mismo, para lo cual se debe definir el “verdadero score” en el que se alcanzan las razones de buenos a malos en una variable dada, para el caso de la variable Dueño de número telefónico se tiene gráficamente:

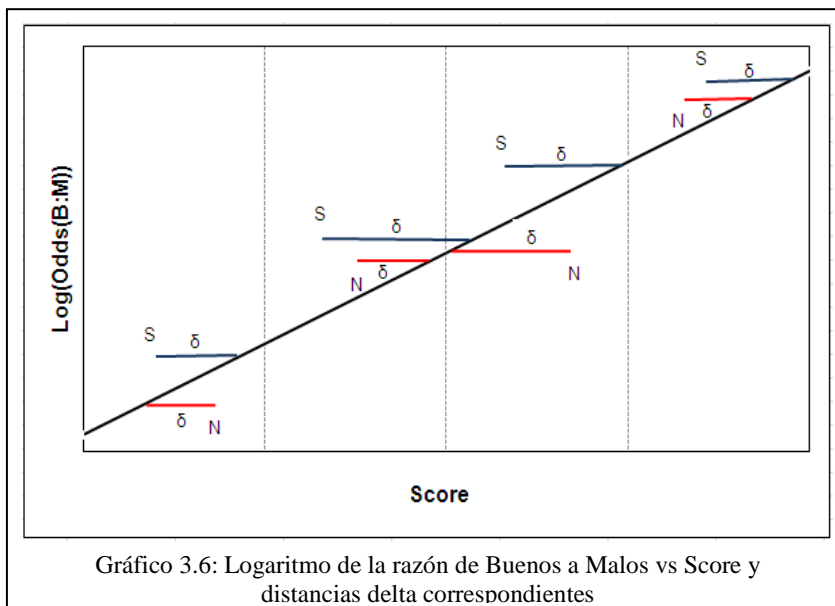


Gráfico 3.6: Logaritmo de la razón de Buenos a Malos vs Score y distancias delta correspondientes

El delta score en cada punto es la cantidad que se necesita adherir al score para que el punto coincida con la línea teórica  $(\alpha + \beta s)$ .

$$\delta = \frac{\log(actual\ odds) - \log(Predicted\ odds)}{\beta} = \frac{\log\left(\frac{actual\ odds}{Predicted\ odds}\right)}{\beta}$$

Si se tiene el conjunto de datos del monitoreo, cuáles clientes son buenos y cuáles malos, y se conoce su score original (de la solicitud) ( y el  $\beta_0$  en la regresión original). Sea  $b_{ij}$  ( $m_{ij}$ ) el número actual de buenos (malos) en la  $i$ -ésima banda de score con atributo  $j$  para una característica dada, y sea  $b_i$  ( $m_i$ ) el número total de buenos (malos) en la  $i$ -ésima banda de score. El delta approach (enfoque delta) consiste en aplicar otra regresión logística a los datos mencionados para estimar para cada score original la probabilidad predicha de que un cliente con ese score sea malo, usando los datos actuales.

$$P(B|Score = s) = \left( \frac{1}{1 + \exp(\hat{\alpha} + \hat{\beta}s)} \right)$$

Sea  $\hat{b}_{ij}$ ,  $\hat{b}_i$  la suma de los malos predicha en la  $i$ -ésima banda de score con atributo  $j$  y en la  $i$ -ésima banda de score en total, respectivamente.

Estimar la razón de buenos a malos actual (odds) por:

$$r_{ij} = \frac{b_{ij} + \frac{1}{2}}{m_{ij} + \frac{1}{2}}, \quad r_i = \frac{b_i + \frac{1}{2}}{m_i + \frac{1}{2}}$$

Estos son mejores estimadores de  $\log(r_{ij})$  que si se tomara  $r_{ij} = \frac{b_{ij}}{m_{ij}}$ ; facilita el caso  $m_{ij} = 0$ .

Estimar la razón de buenos a malos predicha por:

$$\hat{r}_{ij} = \frac{b_{ij} + m_{ij} - \hat{m}_{ij}}{m_{ij}}, \quad \hat{r}_i = \frac{b_i + m_i - \hat{m}_i}{m_i}$$

Es posible que  $\hat{r}_{ij}$  y  $r_{ij}$  difieren considerablemente sin embargo todos los  $\hat{r}_{ij}$  son consistentes con  $r_{ij}$ . Esto significa que toda la curva ha cambiado algo los scores para los atributos bajo consideración son consistente, para continuar se define un predictor escalado de la razón de buenos a malos:

$$\bar{r}_{ij} = \frac{\hat{r}_{ij} r_i}{\hat{r}_i}$$

Este ajuste asegura que el centroide de la línea predictora en la  $i$ -ésima banda concuerda con el centroide de la línea obtenida en la regresión logística original.

Entonces se define  $\delta$  como sigue:

$$\delta_{ij} = \frac{\log\left(\frac{n_{ij}}{\bar{n}_{ij}}\right)}{\beta}$$

Se puede mostrar que  $\delta_{ij}$  tiene un error estándar aproximado de:

$$\sigma_{ij} = \frac{\sqrt{\frac{b_i - b_{ij}}{b_i \left(b_{ij} + \frac{1}{2}\right)} + \frac{m_i - m_{ij}}{m_i \left(m_{ij} + \frac{1}{2}\right)}}}{\beta}$$

Si se toma la media ponderada de los delta scores en la región,

$$\bar{\delta}_i = \frac{\sum_j \frac{\delta_{ij}}{\sigma_{ij}^2}}{\sum_j \frac{1}{\sigma_{ij}^2}}$$

Se puede aplicar una prueba de  $\chi^2$ :

$$e^i = \sum_j \frac{(\delta_{ij} - \bar{\delta}_i)}{\sigma_{ij}^2}$$

Para revisar si la  $i$ -ésima banda de score está mal alineada. La suma de todas las  $e^i$ 's da una medida del alineamiento global de la scorecard.

Si se desea ajustar los para mejorar la alineación se suman los delta scores agregados para el atributo  $j$ ,  $\delta_j$  al score original de  $j$ , donde:

$$\delta_j = \frac{\sum_i \frac{\delta_{ij}}{\sigma_{ij}^2}}{\sum_i \frac{1}{\sigma_{ij}^2}}$$

Esta sería una forma de conseguir un ajuste para cada atributo, pero el  $\delta_{ij}$  da en mayor detalle entendimiento de cuales son los errores causados por ese atributo en diferentes rangos de la scorecard.

Continuando con la metodología de monitoreo, la realización del delta approach abarca los pasos siguientes:



1. División de la scorecard en bandas, rangos de score, adecuadas. No deberían ser muy pocas, pero tampoco demasiadas, esta metodología sugiere que un número adecuado de bandas se encuentra entre 4 y 10. Sin embargo mientras más bandas de score se tengan más preciso será el análisis, pero esto dificultaría la aplicación del delta approach.
2. Calcular el número de buenos y malos que se tienen en la muestra de monitoreo, para cada rango de score definido en el punto anterior  $b_i$  y  $m_i$ , y para cada característica en la scorecard en los rangos de score  $b_{ij}$  y  $m_{ij}$ .
3. Hacer una regresión logística de la variable Indicador B/M obtenida en la muestra de monitoreo<sup>26</sup> con los scores obtenidos con el modelo original (calificación asignada al cliente al momento de la aplicación), para estimar la probabilidad de que un cliente con ese score sea malo.
4. Estimar la actual razón de buenos a malos, obtenida en el monitoreo. ( $n_j$  y  $n$ )
5. Estimar las razones de buenos a malos pronosticadas, usando los resultados de la regresión logística del paso 3. ( $\hat{n}_j$  y  $\hat{n}$ )
6. Ajustar la razón de buenos a malos pronosticada, es decir hallar  $\bar{n}_j$ , para cada banda de score y para cada atributo.
7. Calcular los  $\delta$  y su error estándar,  $\delta_{ij}$  y  $\sigma_{ij}$  respectivamente, para cada banda de score y para cada atributo.
8. Calcular el promedio ponderado de los delta scores en cada rango de score,  $\bar{\delta}_i$ .
9. Realizar las pruebas de  $\chi^2$  para probar la alineación en todas las bandas de score, sobre el estadístico  $e^i$ .
10. Ajustar los scores para cada característica en la scorecard, por medio de:

$$\delta_j = \frac{\sum_i \frac{\delta_{ij}}{\sigma_{ij}}}{\sum_i \frac{1}{\sigma_{ij}}}$$

Finalmente la aplicación del delta approach debe ayudar en la decisión de reconstruir totalmente el modelo o no hacerlo, de todas maneras esta herramienta

---

<sup>26</sup> La variable Indicador B/M se determina en el monitoreo de la manera indicada en el capítulo II

permite ajustar el modelo de una manera objetiva. El modelo debería pasar a la reingeniería si está desalineado en la mayoría o en todas las bandas de score.

### **3.3. IMPLEMENTACION**

La aplicación eficiente del monitoreo requiere la automatización de los reportes incluidos en la metodología, esto permitirá facilitar el trabajo del analista, puesto que el tiempo ahorrado en la construcción de reportes puede ser utilizado en el análisis de los mismos.

Para la automatización se construirá un Panel de Monitoreo utilizando programación VBA en Excel, podrían utilizarse otros medios para la programación, pero el uso de Excel está generalizado a todas las empresas y entidades financieras que podrían aplicar sistemas scoring a sus actividades de selección, sin embargo, pese a no representar gastos reales adicionales, tiene implícito un costo de aprendizaje y mantenimiento asociado al riesgo de rotación del personal analítico. En las siguientes líneas se realiza una breve introducción a las macros:

- **¿Qué es una macro?**

Una macro es un conjunto de instrucciones que sirven para automatizar procesos. Refiriéndonos a excel, supongamos que realizamos frecuentemente la acción de seleccionar un rango para aplicarle negrita, cambio de fuente y centrado. En lugar de hacer estas acciones manualmente, se puede elaborar una macro e invocarla para que ejecute los tres procesos automáticamente.

- **Programación Orientada a Objetos o Programación Basada en Objetos.**

Hay una sutil diferencia entre las definiciones del título. Programación orientada a Objetos, significa que el programador trabaja con objetos fabricados por él mismo, es decir, el programador es quien implementa las clases (conjunto de objetos) para luego crear objetos a partir de ellas. Lo que haremos nosotros, por el

momento, será utilizar objetos ya definidos por la aplicación Excel (WorkSheets, Range,...) sin implementar ninguno de nuevo, por lo que en nuestro caso es más correcto hablar de programación basada en objetos. Esta es una de las grandes ventajas de la programación basada en objetos, utilizar objetos definidos por alguien sin tener que conocer nada sobre su implementación, sólo debemos conocer sus propiedades, métodos y utilizarlos de forma correcta.

- Editor de Visual Basic.

El editor de visual basic es la aplicación que utilizaremos para construir las macros que interactuarán junto con los libros de trabajo.

Para entrar en el editor de Visual Basic, ejecute los pasos siguientes.

1. Active opción Herramientas/ Macro/ Editor de Visual Básic. Se abrirá la ventana en gráfico 3.7.
2. Insertar un nuevo módulo. Un módulo sirve para agrupar procedimientos y funciones. El procedimiento y la función son entidades de programación que sirven para agrupar instrucciones de código que realizan una acción concreta. Para insertar un módulo active opción del menú Insertar/ Módulo. Se activará una nueva ventana, si aparece demasiado pequeña, maximícela.
3. Insertar un procedimiento. Un procedimiento es un bloque de instrucciones de código que sirven para llevar a cabo alguna tarea específica.

Un procedimiento empieza siempre con la instrucción

**Sub** Nombre\_Procedimiento

Y termina con la instrucción

**End Sub**

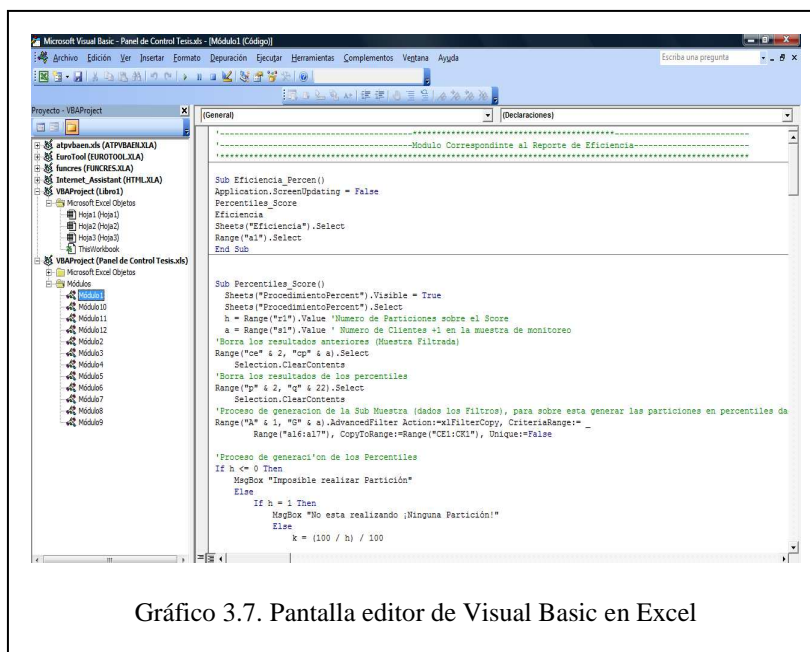


Gráfico 3.7. Pantalla editor de Visual Basic en Excel

Las pantallas con las salidas respectivas a cada reporte de la metodología se presentan a continuación, mientras que el código fuente puede ser consultado en el Anexo 3.

### 3.3.1. PANTALLA PRINCIPAL

Esta se constituye a la vez en una pantalla de presentación y de menú del Panel de Monitoreo. El menú está dividido en dos partes que representan a cada una de las partes fundamentales del monitoreo.

El menú a la izquierda corresponde al Monitoreo de los Resultados del modelo, en el cual existen iconos que permiten acceder a los reportes que conforman esta primera parte de la metodología, los iconos corresponden a los reportes según se indica a continuación:



Resultados por Perfil, a través de este icono se procesan los reportes R1A y R1B.



Razón de Rechazo, se ejecuta el reporte R2.



Índices de Eficiencia, corresponde al reporte R3.

El menú a la derecha se refiere al Monitoreo Estructural del modelo, de la misma forma los iconos que se ubican en este grupo son el acceso a los reportes de correlación de variables y de estabilidad de la población, que componen la segunda parte del monitoreo. Los iconos son:



Correlación de variables independientes vs Indicador B/M, reporte R4A



Indicador de Interrelación KMO, reporte R4B



Correlación de Variables no incluidas en el modelo vs Indicador B/M.



Estabilidad Global de la Población, PSI, reporte R5A



Análisis de Características, corresponde al reporte R5B

Adicionalmente, se tienen otros iconos que cumplen funciones de alimentación de información para ejecutar cada uno de los reportes:



Base Monitoreo, es el acceso para cargar las variables en la muestra de monitoreo, según se especificó en la fichas técnicas de los reportes.



Base de Desarrollo, a través de este icono se ingresa las variables de la muestra de desarrollo que son necesarias para el monitoreo, las mismas que se nombró en las fichas técnicas.



Scorecard, se ingresa la ecuación del modelo.

El gráfico 3.8 muestra la pantalla principal del Panel de Monitoreo:

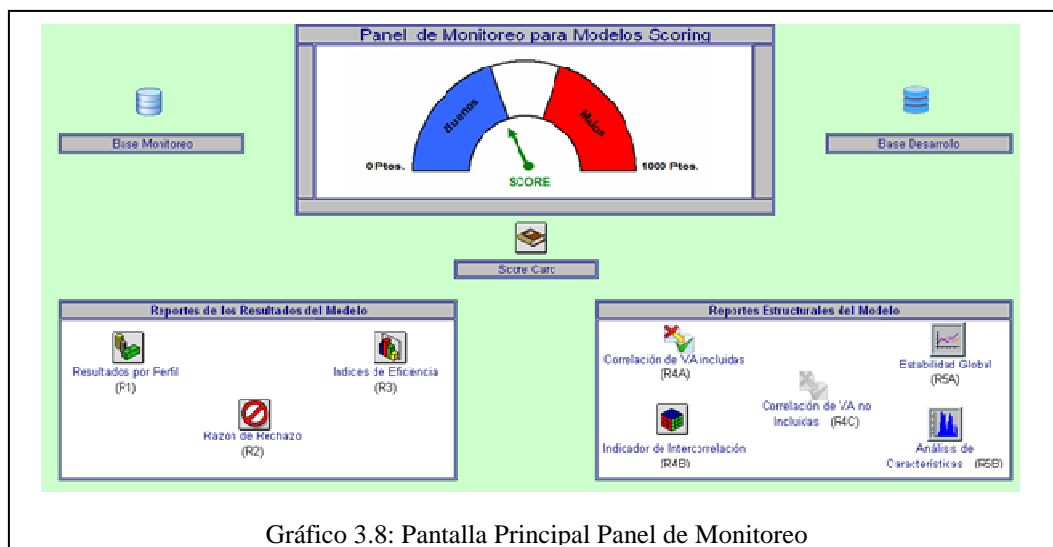


Gráfico 3.8: Pantalla Principal Panel de Monitoreo

### 3.3.2. SALIDA DE RESULTADOS

La salida de los resultados de la programación, son los reportes requeridos para el análisis del desempeño del modelo scoring, en todas las pantallas de salida se tiene un cuadro de controles con las opciones siguientes:



Regresar, permite volver a la pantalla principal



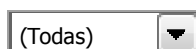
Ayuda, es un acceso a la teoría que respalda al reporte que corresponda.



Ejecutar, procesa la información nuevamente.



Imprimir.



Filtros, dan la opción al usuario de ejecutar el reporte en una sección exclusiva de la población, según se escoja en los filtros. Se pueden establecer restricciones por ciudad, por cosecha, etc. Luego de escoger los filtros se manda a ejecutar el programa.

### 3.3.2.1. Reportes de los resultados del Modelo

El conjunto de reportes que se presentan en esta parte del menú del panel de monitoreo dan resultados gráficos que permiten una fácil interpretación para el analista y facilitan la comprensión a nivel gerencial. Las pantallas que se generan tienen una distribución similar.

- Resultados por Perfil: la ejecución del reporte de Resultados por Perfil se muestra en el gráfico 3.9:

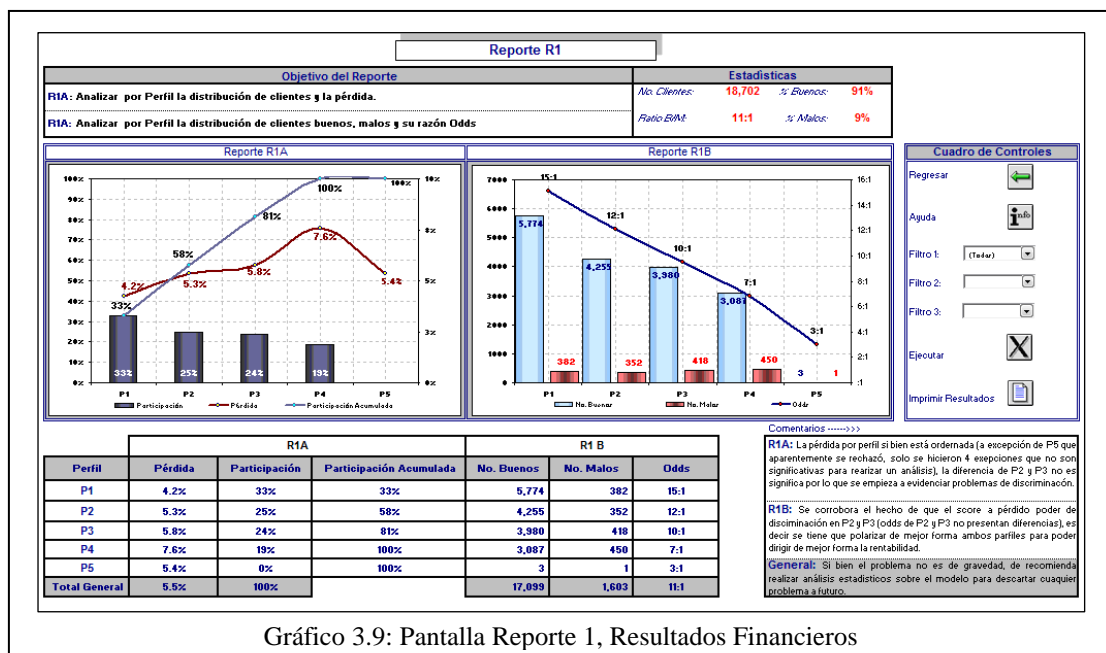
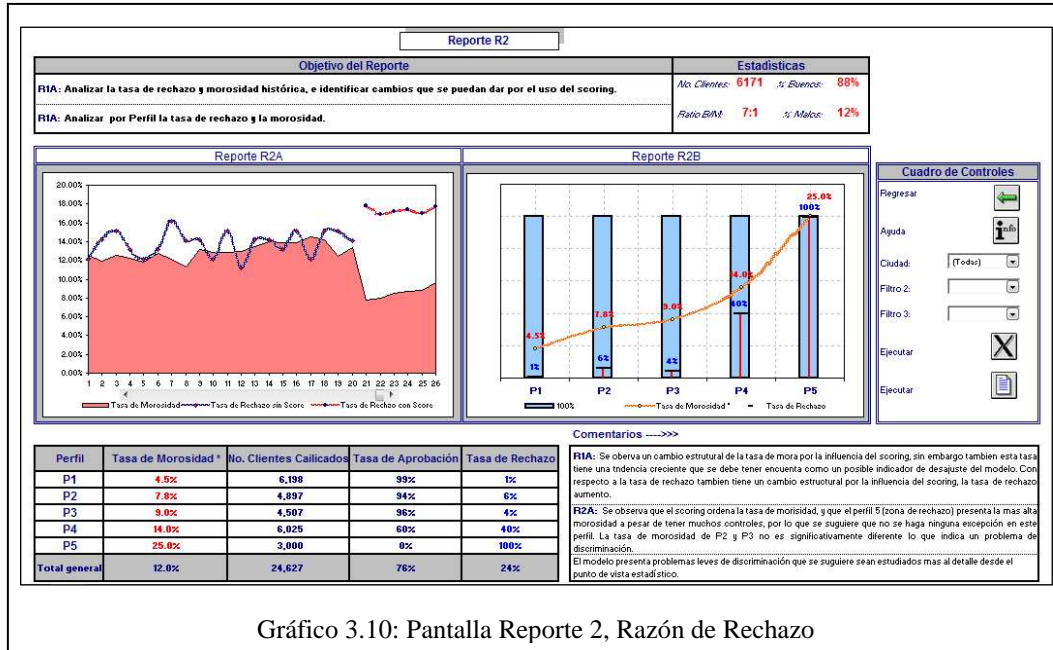


Gráfico 3.9: Pantalla Reporte 1, Resultados Financieros

El reporte R1A corresponde a la distribución de pérdida en los perfiles, se incluye la participación y la participación acumulada de clientes en cada perfil.

El reporte R1B es la distribución de buenos y malos en los perfiles, adicionalmente se puede apreciar la razón de buenos a malos (Odds).

- Razón de Rechazo: El gráfico 3.10 contiene la pantalla de resultados para el reporte en mención.



El reporte R2A corresponde a la tendencia en el tiempo de la tasa de rechazo, se ha incluido la observación de la tasa de morosidad, se contrasta la evolución de estos indicadores antes y después de la aplicación del scoring.

Por otro lado el Reporte R2B contiene los mismos datos que el anterior, con la salvedad de que la observación se realiza como una ponderación por perfil.

- Índices de Eficiencia: el reporte 3 de la metodología de monitoreo se refiere al análisis de la eficiencia del modelo, en el gráfico 3.11 se presenta la pantalla que arroja el panel en esta etapa:



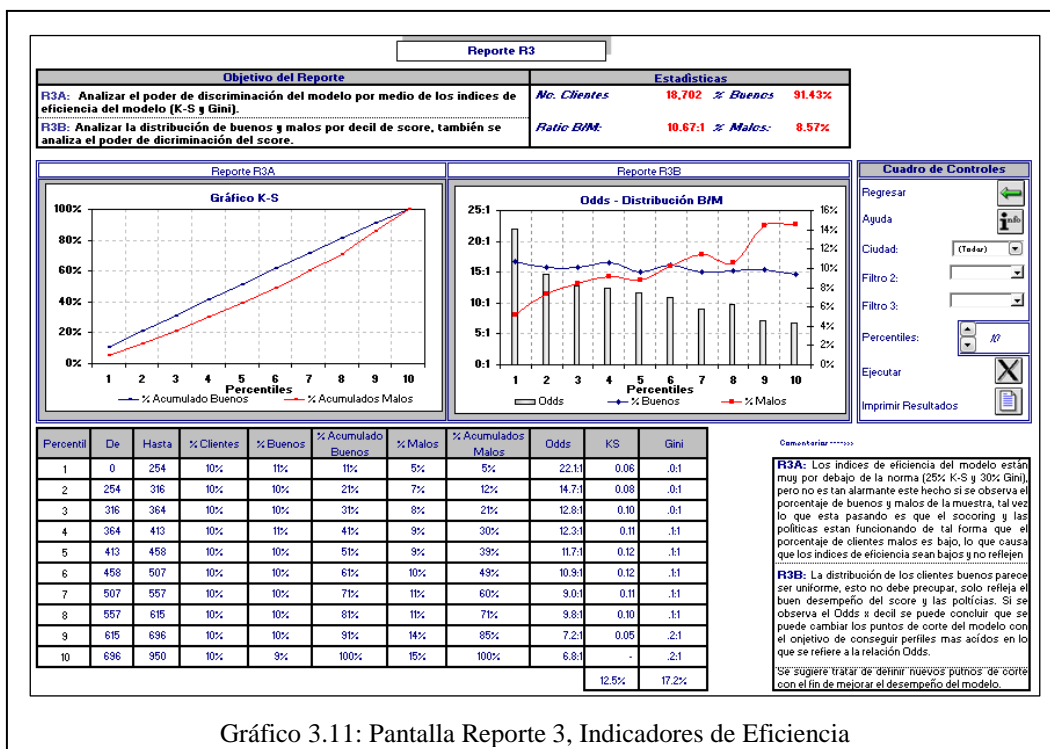


Gráfico 3.11: Pantalla Reporte 3, Indicadores de Eficiencia

El reporte R3A es el gráfico del índice K – S, la ejecución se realiza en el número rangos de score que el analista seleccione.

El segundo reporte (R3B) presenta la distribución de buenos, malos y la razón de buenos a malos (Odds) en los rangos de score seleccionados en R3A.

La tabla bajo los gráficos resume el resultado numérico obtenido en la corrida del reporte, así como el cálculo de los índices K – S y Gini.

### 3.3.2.2. Reportes Estructurales del Modelo

El grupo de reportes de la estructura del modelo tienen el objeto de identificar específicamente las razones por las que los resultados del modelo difieren de los esperados, en su mayoría estos reportes presenta la información por medio de tablas que prmiten visualizar los resultados de los análisis propuestos anteriormente:

- Correlación de Variables Incluidas en el Modelo vs. Indicador B/M: El panel de monitoreo permite ejecutar el reporte general de correlación de variables, las

pruebas de independencia se realizan al seleccionar el icono respectivo, la pantalla generada se presenta en el gráfico 3.12

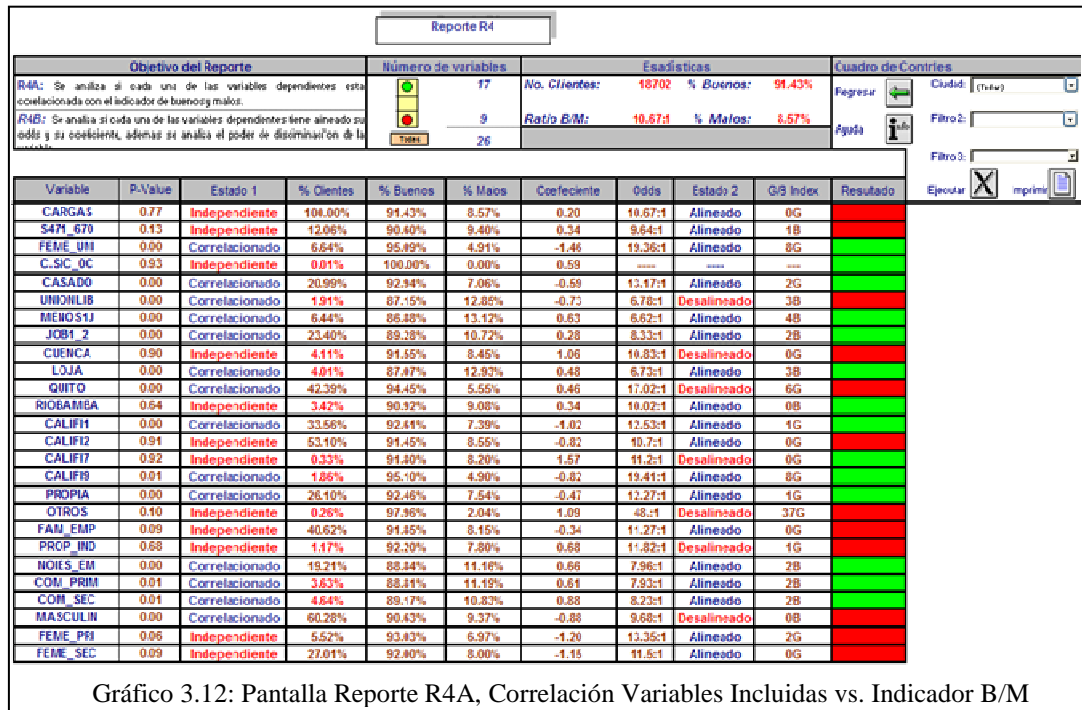


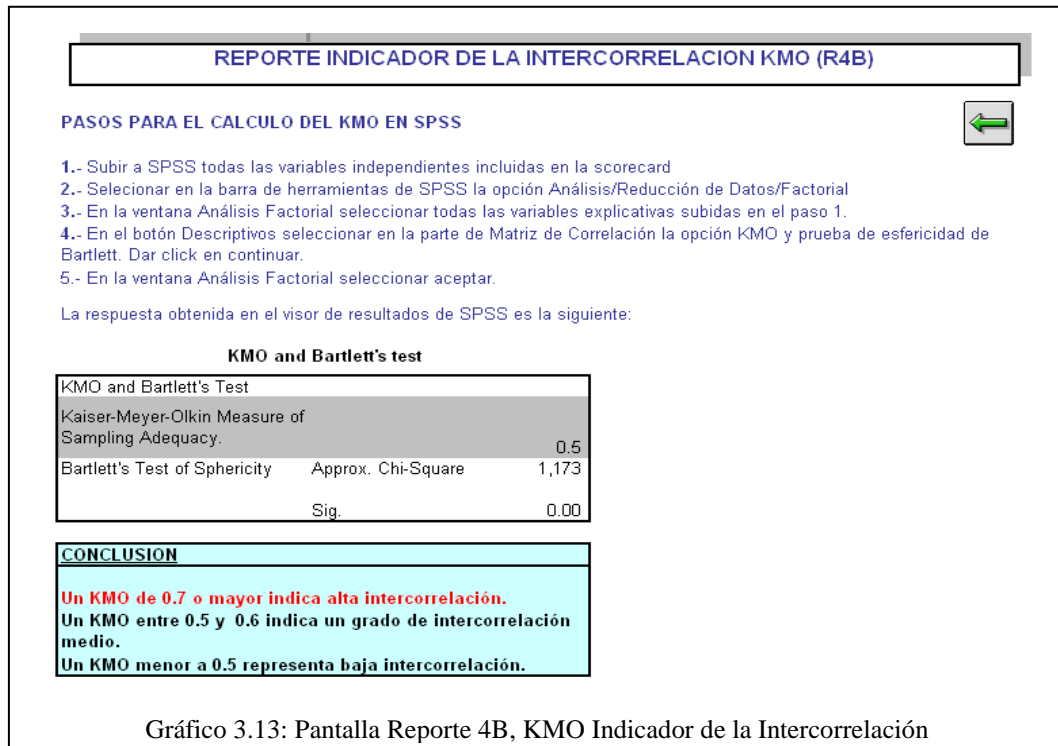
Gráfico 3.12: Pantalla Reporte R4A, Correlación Variables Incluidas vs. Indicador B/M

La salida del reporte R4A consiste en una tabla resumen en la que se presenta el resultado de la prueba de Independencia chi-cuadrado, así como los porcentajes de clientes buenos y malos relativos a la variable, la comparación entre el signo del coeficiente y la razón de buenos a malos permite identificar el estado de alineación o desalineación de las variables.

Adicionalmente el reporte da la opción de filtrar los resultados según la gravedad que implican para los supuestos del modelo, mediante la figura de un semáforo, donde la luz verde indica las variables se mantienen correlacionadas con el Indicador B/M y están alineadas en el monitoreo. La luz roja se identifica con las variables que deben ser revisadas con más detenimiento (indican deterioro en el modelo), es decir, las variables que están desalineadas o muestran independencia con el indicador B/M.

- Indicador de la Interrelación de variables Explicativas: el reporte R4B consiste en realizar el cálculo del KMO para todas las variables independientes incluidas en el modelo, la programación de la prueba es

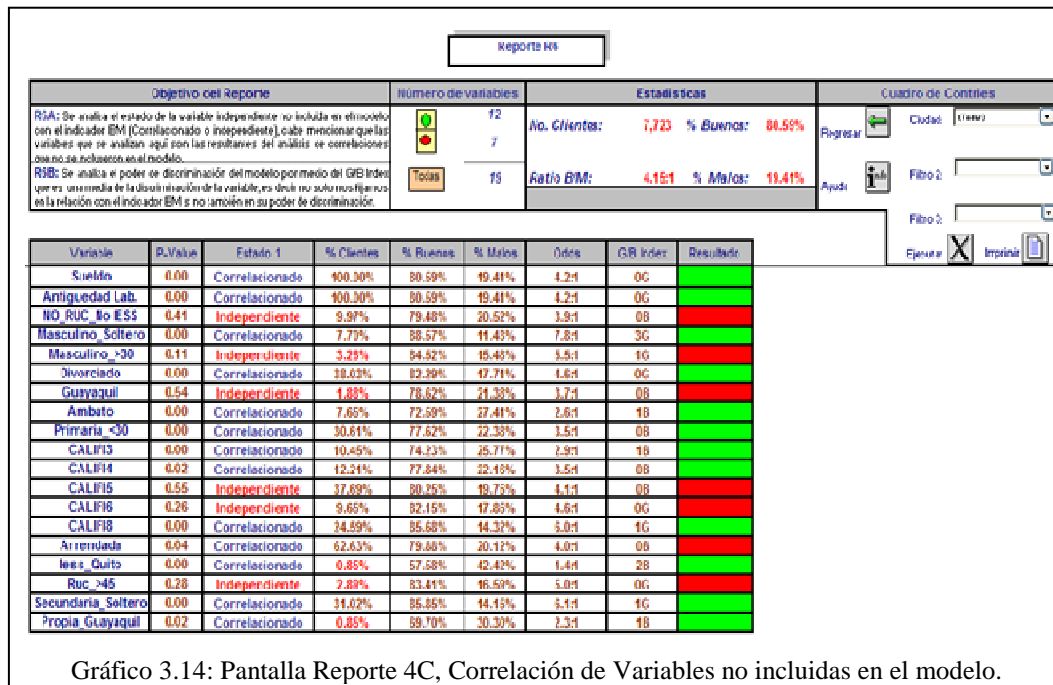
complicada de realizar en Excel, pero el resultado se obtiene fácilmente usando un programa estadístico, como SPSS. La pantalla presentada en el panel puede ser apreciada en el gráfico 3.13:



En este reporte el panel se limita a presentar los pasos que deben realizarse en SPSS para obtener el coeficiente KMO. Lo interesante para el analista es conocer el valor de este indicador para poder concluir a cerca de la intercorrelación de las variables independientes, por tanto el esfuerzo informático que implica el cálculo del KMO es irrelevante, cuando se cuenta con una herramienta como SPSS que permite conocer rápidamente el valor deseado para el análisis.

- **Correlación de Variables No Incluidas en el Modelo vs. Indicador B/M:** la implementación en el panel de monitoreo para el reporte R4C es muy similar a lo realizado para el reporte R4A, la información del estudio de correlaciones se presenta en una tabla, junto con el valor crítico obtenido en la prueba de independencia, además se presentan las distribuciones de buenos, malos y la razón de buenos a malos en cada variable. En este reporte no es posible

realizar conclusiones a cerca de la alineación puesto que los coeficientes de estas variables son cero, por cuanto no están incluidas en la scorecard.



- Estabilidad Global: el reporte R5A corresponde al cálculo del PSI, en el panel este reporte está conformado por una parte gráfica y por una parte informativa. Los gráficos contienen información de las distribuciones de buenos y malos y la razón de buenos a malos tanto en el desarrollo como en el monitoreo, según se aprecia en el gráfico 3.15. La parte informativa se muestra en una tabla que incluye los datos de las distribuciones de clientes en la muestra de desarrollo, así como en el monitoreo, con tales datos se realiza el cálculo del PSI para la población total, para los clientes buenos y para los malos.

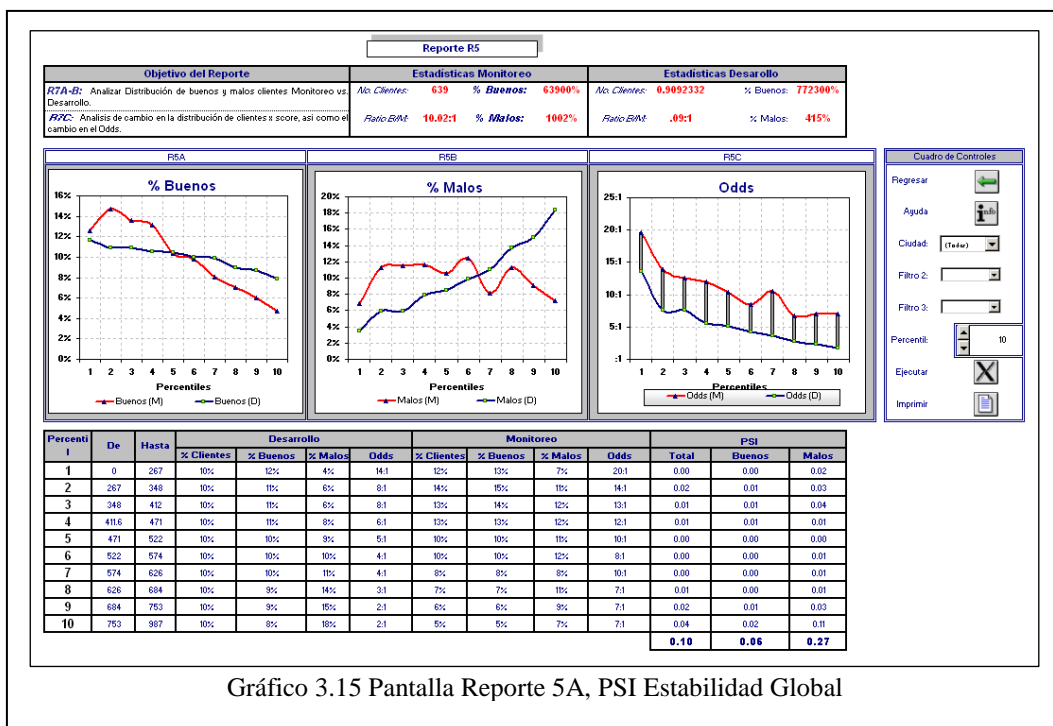


Gráfico 3.15 Pantalla Reporte 5A, PSI Estabilidad Global

- **Análisis de Características:** La pantalla en el gráfico 3.16 atañe a la automatización reporte R5B, para tal efecto se construye una tabla con todas las variables de la scorecard tal como se indicó en el punto 3.2.2.2. El panel permite realizar filtros, por ejemplo, por ciudad y por variable para permitir al analista la identificación de las variables y ciudades que causan mayor impacto en las diferencias de score que se encuentran en la muestra de monitoreo.

Todos los reportes considerados en la metodología se han automatizado con el Panel de Monitoreo, en lo posible la implementación considera una interfaz gráfica simple para la observación y el entendimiento a todo nivel. El análisis de fuentes de negocio, así como el delta approach se dejan para la aplicación de cada analista según se ha propuesto en este trabajo.



## **CAPITULO IV**

### **PROYECCION DE LA VARIACION MENSUAL DE PROVISIONES POR PERFIL DE CLIENTE**

#### **4.1. INTRODUCCIÓN**

Este capítulo pretende dar un enfoque general de la metodología para el cálculo de provisiones en una entidad financiera, y dejar en claro la necesidad de su proyección para presupuestar el gasto de provisiones de una entidad, tomando como base la normativa de la Superintendencia de Bancos y Seguros del Ecuador y la definición de perfiles de clientes.

El crecimiento descontrolado de la morosidad de una cartera puede ocasionar fuertes pérdidas en las instituciones financieras. Con el fin de prevenir estas posibles pérdidas Basilea II sugiere mantener provisiones que permitan cubrir el capital que podría perderse si tal evento ocurre, las provisiones constituyen un capital que la entidad financiera no puede considerar como parte integrante de su activo total. Las provisiones se establecen durante el tiempo de vida de un crédito, si el crédito cae en mora, las provisiones se mantienen hasta su cancelación total o hasta que la institución declare el crédito como capital perdido.

El objetivo de las provisiones es no sobreestimar el valor los activos, son aplicables a cualquier tipo de empresa, pero son de especial relevancia en las instituciones financieras, donde el principal activo es la cartera de créditos, de la cual no se tiene la certeza absoluta de recuperarla en su totalidad, por tal motivo no debe ser considerada en el estado de situación como un activo completo. De ahí la importancia de estimar las provisiones de una manera adecuada de modo que no sean insuficientes tal que sobrestimen el valor de los activos, y que no sean excesivas tal que afecten a la rentabilidad de la institución.

Según se explicó en el Capítulo I de este proyecto, las entidades deben, hasta desarrollar modelos eficientes para la proyección de pérdidas, aplicar la metodología que el ente supervisor haya señalado, en general los supervisores establecen porcentajes de provisión distintos según el deterioro de morosidad que presente el crédito, así en Ecuador los porcentajes de provisión cambian según varía la calificación de riesgo<sup>27</sup>, los porcentajes de provisión mínimos en cada calificación se aprecian en la tabla 4.1:

<b>Calificación</b>	<b>% de Provisión</b>
<b>A</b>	1%
<b>B</b>	5%
<b>C</b>	20%
<b>D</b>	50%
<b>E</b>	100%

Tabla 4.1: Porcentajes de Provisión Mínimos SBS

La altura de mora (Días de retraso en el pago) determina la calificación, las consideraciones de mora son distintas según el crédito sea de consumo, micro crédito, comercial o de vivienda. La aplicación de estos porcentajes de provisión implica que la institución financiera debe provisionar al menos 1% del total de su cartera aún cuando ninguno de sus clientes haya incurrido en mora. Finalmente el monto de provisión total corresponde a la suma de los montos que se encuentren en cada calificación ponderada por los porcentajes de provisión mencionados. Con este objetivo las instituciones financieras están obligadas a calificar mensualmente su cartera de créditos, lo que se denomina calificación de activos de riesgo.

De la misma manera que se tienen porcentajes mínimos de provisión, existen también porcentajes máximos que una entidad puede establecer como provisión, puesto que la provisión se constituye como un gasto deducible para las entidades, un exceso de provisiones podría ser usado como un medio para reducir pago de impuestos y utilidades; por tal razón la SBS del Ecuador ha determinado un techo

---

<sup>27</sup> La definición de las calificaciones de A a E se presentó en el Capítulo I



para la provisión en cada una de las calificaciones de riesgo, según se muestra en la tabla 4.2:

<b>Calificación</b>	<b>% de Provisión</b>
<b>A</b>	4%
<b>B</b>	19%
<b>C</b>	49%
<b>D</b>	99%

Tabla 4.2: Porcentajes de Provisión  
Máximos SBS

La medición de las provisiones adecuadas depende directamente de la forma de medir el riesgo de crédito, lo cual ha variado sustancialmente con la incorporación de metodologías de cuantificación de riesgo y la aplicación de sistemas de información más avanzados.

Actualmente, la medición del riesgo de crédito se realiza a través de la cuantificación de la pérdida esperada e inesperada, como síntesis del comportamiento histórico de un portafolio específico. La cuantificación de las pérdidas permite a las entidades financieras que la medición de la rentabilidad de sus clientes y de su cartera sea más apropiada, a través de aplicar políticas que aseguren la cobertura de pérdidas, es decir estimar las provisiones que deben ser mantenidas de acuerdo al riesgo asumido por cada entidad en la concesión de créditos Así como los requerimientos de capital para cubrir las pérdidas inesperadas

Mediante el cálculo de la Pérdida Esperada, se pretende poder anticipar la pérdida real en que incurrirá una entidad a un plazo determinado. Este cálculo permitirá a la entidad adoptar las medidas oportunas para afrontarla (provisiones), o para que en el futuro el monto a provisionar disminuya, contrastando con las exigencias del ente supervisor e incrementado la rentabilidad.

## 4.2. PÉRDIDA ESPERADA

Antes de iniciar con una explicación más a fondo del tema de la pérdida esperada, es necesario conocer los términos que se definen a continuación:

- Pérdida Esperada ( $PE$ ): Es el valor esperado de pérdida por riesgo crediticio en un horizonte de tiempo determinado.
- Probabilidad de No Pago (default) ( $PD$ ): Es la posibilidad de que un deudor no esté dispuesto o no pueda pagar sus obligaciones, se atribuye también al no cumplimiento de alguna de las condiciones del contrato de crédito firmado entre el deudor y el prestamista.
- Severidad de la pérdida ( $SP$ ): Es el porcentaje no recuperado al incumplir el acreditado, una vez tomados en cuenta todos los costos implicados en dicha recuperación.
- Valor Expuesto al riesgo de crédito ( $VE$ ): Es el valor presente al momento de producirse el incumplimiento de los flujos (cuotas) que se espera recibir de las operaciones crediticias.

Si un cliente cae en no pago de sus obligaciones, la entidad financiera perdería el capital que le es adeudado, así como los intereses que preveía ganar con la operación incumplida, además de otros gastos que se generan por el evento de no pago. Desde tal punto de vista se tiene que la pérdida sigue una distribución binomial de la forma siguiente:

$$Pérdida = \begin{cases} 0 & \text{con probabilidad } 1 - PD \\ 1 - r & \text{con probabilidad } PD \end{cases}$$

Dónde  $PD$  es la probabilidad de incumplimiento y  $r$  es la tasa de recuperación.

La esperanza y la varianza de la pérdida se definen de la forma siguiente:

$$E(P) = 0 \times (1 - PD) + (1 - r) \times PD$$

$$E(P) = (1 - r) \times PD$$

$$V(P) = E(P^2) - E(P)^2$$

$$V(P) = 0^2 \times (1 - PD) + (1 - r)^2 \times PD - (1 - r)^2 PD^2$$

$$V(P) = (1 - r)^2 \times PD \times (1 - PD)$$

Por lo tanto el monto de pérdida esperada ( $PE_i$ ) se calcula de manera individual para cada crédito ( $i$ ) con la siguiente ecuación:

$$PE_i = PD_i \times (1 - r_i) \times VE_i$$

En resumen:

<p>La PE considera tres elementos</p>	<p>Probabilidad de Default (<math>PD_i</math>)</p> <p><b>Cuál es la probabilidad de no pago?</b></p>	<p>Severidad (<math>1 - r_i</math>)</p> <p><b>Cuánto se va a recuperar?</b></p>	<p>Exposición (<math>VE_i</math>)</p> <p><b>Cuánto va a caer en default?</b></p>
---------------------------------------	--	---	--

Para calcular las pérdidas extremas (no esperadas), es importante precisar que éstas corresponden a una determinada función de distribución de pérdidas, a un horizonte temporal determinado y a un nivel de confianza establecido. El análisis de varios años de los eventos de no pago (default) asociados a cada calificación de riesgo mostrarán variaciones de año en año; esta volatilidad del riesgo es la pérdida no esperada de un crédito. En consecuencia, las pérdidas extremas comúnmente se refieren a  $x$  veces la desviación estándar de la función de pérdidas, el número  $x$  dependerá del nivel de confianza con el que se trabaje.

Si suponemos que la exposición crediticia y la tasa de recuperación son fijas e independientes de la función de distribución de pérdidas, entonces la desviación estándar de la pérdida  $DE(P)$  se calcula por:

$$\begin{aligned}
 DE(P) &= \sqrt{V(P)} = \sqrt{(1 - r)^2 \times PD \times (1 - PD)} \\
 &= (1 - r) \sqrt{PD \times (1 - PD)}
 \end{aligned}$$

Así la desviación estándar del monto de pérdida ( $DP$ ) calculada para cada crédito  $i$  sería:

$$DP = \sqrt{PD \times (1 - PD)} \times (1 - r) \times VE$$

De donde la pérdida no esperada puede obtenerse de:

$$\text{Pérdida no esperada} = xDP$$

Sin embargo la aplicación de esta metodología para el cálculo de la pérdida no esperada supone que la distribución de pérdidas es o puede ser aproximada a una normal, para poder relacionar  $x$  veces la desviación estándar con el nivel de confianza del intervalo en el que se mueve la pérdida.

El cálculo de la pérdida esperada exige una metodología rigurosa, que permita obtener una estimación adecuada de cada uno de los términos involucrados, para lo cual se aplican técnicas estadísticas y por tanto es necesario disponer de una información histórica suficiente y consistente. Es decir que los parámetros de la ecuación se obtienen de la aplicación de un modelo específico, a continuación se describe las consideraciones que deben ser hechas en el cálculo de los términos.

#### **4.2.1. PROBABILIDAD DE INCUMPLIMIENTO (default)**

Antes de tratar el tema de la probabilidad de no pago (default) es necesario aclarar el significado del evento de no pago; el incumplimiento es cualquier evento que obliga al prestamista a declarar de plazo vencido a una operación determinada, declarar de plazo vencido a una operación es anticipar el vencimiento, la anticipación del vencimiento puede generarse por incumplimiento de alguna de las cláusulas consideradas en el contrato de crédito, o por la falta de pago durante un tiempo establecido previamente; generalmente, la evolución de un crédito sufre alteraciones a medida que transcurre el tiempo hasta su completa terminación. En otras palabras, existen incumplimientos que son simples retrasos por pequeños desajustes de liquidez o por eventos extraordinarios de los deudores. Por lo anterior, los incumplimientos comienzan a ser relevantes después de cierto tiempo desde el impago de la cantidad establecida contractualmente. Además, este tiempo es diferente según la clase de crédito y

puede cambiar según las circunstancias económicas. Por lo tanto, inicialmente es imprescindible determinar el número de días de mora (altura de mora) en el cual el cliente no vuelve a cumplir sus obligaciones financieras, tiempo de no pago, es decir que al pasar un día determinado de mora se considerará que el crédito ya no se recuperará (tiempo establecido previamente, comúnmente se usan matrices de transición).

Metodológicamente el concepto de probabilidad de default surge de una variable aleatoria  $X$  asociada a cada crédito en donde se indica el estado en el que se encontrará el crédito luego de haber transcurrido un tiempo determinado. Solo puede tener dos valores, uno asociado al evento de default y otro al evento de no default, así:

$$X_i = \begin{cases} 0 & \text{no default} \\ 1 & \text{default} \end{cases}$$

Para  $i = 1, 2, 3, \dots, n$ ,  $n$  es el número total de créditos.

La variable aleatoria  $X$  se modela mediante una distribución de Bernoulli cuya función de densidad es:

$$f(x) = p^{X_i} (1 - p)^{1 - X_i}$$

Y los momentos de primer y segundo orden son:

$$E(X_i) = p \quad V(X_i) = p(1 - p)$$

Donde  $p$  es la probabilidad de default o incumplimiento.

Un estimador de la probabilidad se puede conseguir para una muestra de créditos de tamaño  $m$ , de la forma siguiente:

$$\bar{X} = \frac{\sum_{i=1}^m X_i}{m} = \frac{\text{Número de eventos de default en la muestra}}{\text{Tamaño de la muestra}}$$

Que para el caso binomial coincide con el valor esperado de la variable aleatoria  $X$ , entonces  $E(X)$  puede estimarse con  $\bar{X}$ <sup>28</sup>.

Para calcular la probabilidad de incumplimiento se usan herramientas como credit scoring o matrices de transición. El credit scoring puede desarrollarse de la misma forma descrita en el Capítulo II, considerando que la variable dependiente debe ser un Indicador de default y no default construido en base a  $\bar{X}$ , en lugar del Indicador de Mora con el que se desarrolló el Indicador de Buenos y Malos de el modelo presentado en este proyecto. Puesto que la definición de malo, no necesariamente es la de un cliente que caerá en no pago de sus obligaciones (caer en default). Al ser el Indicador de default y no default la variable dependiente, el modelo scoring arroja la probabilidad de no pago (default)  $PD$  para cada uno de los clientes.

Una breve introducción a las matrices de transición así como la metodología para el cálculo de la probabilidad de default utilizando matrices de transición se resume en las líneas a continuación. Para mayor detalle ver Transition Probability Estimation in a Delinquency Markov Chain Model de Scott D. GRIMSHAW y William P. ALEXANDER.:

#### 4.2.1.1. Matrices de transición

Se define la probabilidad de transición  $p_{ij}$  como la posibilidad de que un prestatario de un crédito con calificación crediticia  $i$  (o rango de mora  $i$ ) pueda migrar o moverse a otra calificación crediticia  $j$  (o rango de mora  $j$ ) en un horizonte de tiempo dado.

Por lo anterior, es posible construir una matriz de transición  $A$  con  $i$  filas y  $j$  columnas, de tal manera que satisfagan las siguientes condiciones:

- Todos los elementos de la matriz son no negativos, es decir,  $p_{ij} > 0$ .
- La suma de los elementos de cada fila es igual a la unidad, es decir,  
$$\sum_j p_{ij} = 1 \text{ para todo } i.$$

---

<sup>28</sup> Ver Requerimiento Mínimo de Capital, paper de Latoli

Lo anterior se deriva de un concepto matemático muy utilizado llamado cadenas de Markov; en el Anexo 1 se presentan algunos detalles de esta definición matemática.

Si  $A$  es una matriz de transición anual, entonces las entradas de ésta,  $p_{ij}$ , representan la fracción de créditos con calificación  $i$  (o rango de mora  $i$ ) que tienen un año después calificación  $j$  (o rango de mora  $j$ ).

Para la estimación de las probabilidades de transición se conocen dos metodologías, la discreta y la continua, mismas que se detallan en los puntos a continuación.

- **Método Discreto**

En este método, las probabilidades de transición son estimadas como:  $p_{ij} = N_{ij} / N_i$  para todo  $i, j$  donde  $N_{ij}$  es el número de créditos que comenzaron al inicio del periodo en la calificación  $i$  (rango de mora  $i$ ) y terminaron al finalizar el periodo en la calificación  $j$  (rango de mora  $j$ ), y  $N_i$  es el número de créditos que estaban en la calificación  $i$  al comienzo del periodo.

Una vez que se calculan las  $p_{ij}$  para cada momento del tiempo de la muestra analizada, se calculan unas  $\overline{p_{ij}}$  que representan las probabilidades de transición promedio de todo el periodo analizado. Es decir:

$$\overline{p_{ij}} = \sum_{t=0}^T w_t p_{ij}(t)$$

donde  $w_t$  es la ponderación para cada momento del tiempo analizado.

- **Método Continuo**

El punto de partida para estimar cadenas de Markov en tiempo continuo consiste en asumir homogeneidad en el tiempo para un período corto. En las estimaciones que se presentan en esta parte se asume homogeneidad para un período de un año; posteriormente se presentan los resultados de las estimaciones sin hacer ese supuesto.

Supongamos que se observa la calificación para  $N$  créditos entre el tiempo 0 y el tiempo  $T$ ; además, que el espacio de estados es finito, siendo la categoría 1 la mejor calificación y  $z$  la peor. Sea  $P(t)$  la matriz de transición para un período de tiempo determinado. La matriz de transición puede expresarse en términos de las intensidades de transición, que representan las probabilidades instantáneas de migración entre los diferentes estados (ej. Cambio de la calificación  $i$  (rango de mora  $i$ ) a la calificación  $j$  (rango de mora  $j$ )). En este sentido,

$$P(t) = \exp(\Lambda t), t \geq 0$$

donde  $\Lambda$  representa la matriz generadora, cuyos elementos son las intensidades de transición. La ventaja de poder expresar la matriz de transición en términos de la matriz generadora radica en que, teniendo las probabilidades instantáneas como insumo, es posible obtener la matriz de transición correspondiente para el período de tiempo que se desee, escalando la matriz generadora por el tiempo. De esta forma se soluciona el problema de la arbitrariedad en la definición del período de las estimaciones en tiempo discreto.

Dado que para cualquier  $t$  la matriz de migraciones es una función monótona de la matriz generadora, es posible obtener estimadores de máxima verosimilitud para las probabilidades de migración encontrando, primero, estimadores de máxima verosimilitud para las intensidades de migración, y luego, escalando por el período de tiempo apropiado.

Los estimadores de máxima verosimilitud de los elementos de la matriz generadora están dados por:

$$\hat{\lambda}_{ij} = \frac{N_{ij}(T)}{\int_0^T Y_i(s) ds}, \text{ para } i \neq j$$

donde  $N_{ij}(T)$  representa el total de migraciones del estado  $i$  al estado  $j$  ocurridas entre el tiempo 0 y el tiempo  $T$ , y  $Y_i(s)$  representa el número de créditos con calificación  $i$  en el tiempo  $s$ . Los elementos diagonales de la matriz generadora están dados por  $\hat{\lambda}_{ii} = -\sum_{i \neq j} \hat{\lambda}_{ij}$ . El denominador tiene en cuenta cada



crédito que ha obtenido la calificación  $i$  en algún momento entre el tiempo 0 y el tiempo  $T$ .

Una ventaja de este método es que considera también transiciones indirectas entre estados, y esto soluciona el problema de la subestimación de las probabilidades de eventos infrecuentes. En particular, la estimación de una transición es estrictamente positiva si durante el período considerado hay una secuencia de migraciones entre categorías intermedias, incluso si no hubo ninguna migración directa y ningún crédito experimentó esa secuencia de migraciones intermedias. Por ejemplo, si estamos interesados en estimar la probabilidad de un evento raro, tal vez la probabilidad de migración de la categoría 1 a la categoría  $s$  en un año, pero no hubo ningún crédito que experimentara directamente dicha transición, aún podemos estimar una probabilidad positiva si hubo siquiera un crédito que migró de 1 a 2, otro que lo hizo de 2 a 3, ... y otro de  $s - 1$  a  $s$ , durante dicho período.

Frecuentemente al construir matrices de transición se supone que las probabilidades provienen de cadenas de Markov homogéneas; en otras palabras, se supone que estas probabilidades no dependen del tiempo ni del ciclo económico. Sin embargo, los desarrollos recientes han mostrado que las probabilidades de transición cambian con el ciclo económico. Por ejemplo, durante recesiones o cuando la economía comienza a desacelerarse, las probabilidades de incumplimiento tienden a aumentar; mientras que en las etapas de auge o cuando la economía comienza a recuperarse luego de la recesión, las probabilidades de incumplimiento tienden a ser menores. Los distintos escenarios se definen en función de las variaciones del PIB en los distintos periodos.

Los ciclos económicos pueden considerarse como escenarios, por ejemplo:

Escenario 1: La economía está en recesión ( $PIB_t - PIB_{t-1} < 0$ )

Escenario 2: La economía estuvo en recesión, pero comienza a recuperarse ( $PIB_t - PIB_{t-1} > 0$ , pero  $PIB_{t-1} - PIB_{t-2} < 0$ ).

Escenario 3: La economía estuvo en expansión, pero comienza a desacelerarse ( $PIB_t - PIB_{t-1} < 0$ , pero  $PIB_{t-1} - PIB_{t-2} > 0$ ).

Escenario 4: La economía está en expansión ( $PIB_t - PIB_{t-1} > 0$ ).

El objetivo es determinar mediante una prueba de hipótesis si las matrices condicionadas son diferentes de las no condicionadas y así comprobar en qué medida el ciclo económico afecta la calificación o rango de mora de los acreditados.

En general, debe hacerse hincapié en cómo la calidad de los acreditados es inferior en las contracciones que en las expansiones, ya que las probabilidades de no pago y las migraciones a categorías peores de riesgo aumentan durante años de recesión y se reducen durante las expansiones. Por lo tanto, se puede concluir que los acreditados migran a través de las categorías a lo largo del tiempo y esta migración efectivamente va a depender del estado en que se encuentre la economía. Teniendo en cuenta este hecho, el siguiente paso consistirá en cuantificar qué porcentaje de estas migraciones puede atribuirse al ciclo económico, y cómo afecta a los requerimientos de capital. El objetivo es medir el impacto del factor común en la solvencia de los acreditados bancarios y sus repercusiones en términos de variaciones de capital a lo largo del tiempo cuando dicho capital se calcula con un sistema de clasificación interno de acreditados.

La tabla 2.1 a continuación muestra la estructura de una matriz de transición que considera la migración entre calificaciones y entre escenarios económicos.

		Al final del año															
		Escenario 1				Escenario 2				Escenario 3				Escenario 4			
		A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Al inicio del año	Escenario 1	A															
		B															
		C															
		D															
	Escenario 2	A															
		B															
		C															
		D															
	Escenario 3	A															
		B															
		C															
		D															
	Escenario 4	A															
		B															
		C															
		D															

Tabla 4.3 : Matriz de transición de la cartera sujeta al ciclo económico

En general una matriz de transición se construye siguiendo los pasos que se describen en las líneas siguientes:

- Crear rangos de días de mora y realizar un conteo de los créditos en cada uno de los estados.
- Definir la probabilidad de transición  $p_{ij}$  como la posibilidad de que un cliente con una categoría  $i$ , pueda migrar o moverse a otra categoría  $j$  en un horizonte de tiempo dado (mensual).
- Seleccionar un método de estimación discreto o continuo.
- Obtener matrices de transición en el período de madurez seleccionado y determinar  $\hat{p}_{ij}$
- Ajustar la estimación del default en función de las variaciones de la economía, debido a que es necesario condicionarlo al ciclo económico. Además, se debe ajustar el rezago que existe entre el crecimiento económico agregado y el sector financiero.

- Finalmente, a partir de las probabilidades se determina el punto de no retorno que se define como aquella mora cuya probabilidad de empeorar o permanecer es mayor a la probabilidad de mejorar o permanecer (Definición de default).

#### 4.2.2. EXPOSICIÓN

Es la cantidad de dinero que el banco arriesga en el caso de que el cliente entre en mora, en otras palabras, la máxima cantidad que podría ser perdida, asumiendo que no hubiera ningún tipo de recuperación. Para productos donde el repago del capital tiene lugar en fechas y cantidades determinadas, por ejemplo préstamos, es relativamente fácil calcular la exposición, ya que ésta equivale, aproximadamente, al capital pendiente de pago. Sin embargo para otro tipo de productos por ejemplo líneas de crédito, la exposición a la entrada en mora no puede ser anticipada y contiene una componente de incertidumbre.

En estos productos el cliente en teoría puede utilizar la totalidad de su cupo, además un empeoramiento de la calidad crediticia del cliente conllevará una mayor utilización de fondos con el fin de evitar la mora, es decir, realiza avances en efectivo. Aunque el banco pueda intentar reducir la posible exposición, cancelando las líneas no utilizadas o reduciendo los límites que el cliente utiliza, las instituciones financieras son en general más lentas que el propio cliente en notar el empeoramiento de su condición y a menudo las cláusulas contractuales impiden restringir el acceso a líneas ya concedidas.

La exposición se calcula como:

$$EAD = C_{exp} + IC + COper + VUso$$

Donde:

$C_{exp}$  : Capital expuesto al momento del default

$IC$  : Intereses corrientes.

*CO<sub>per</sub>* : Costos Operativos.

*V<sub>Uso</sub>* : Valor Esperado del uso del cupo de crédito remanente.

En resumen la exposición en el evento del default debe considerar el Capital expuesto, además los intereses devengados hasta la recuperación total de la deuda, los costos operativos del mantenimiento del crédito y el valor del uso programado para los flujos de la recaudación en condiciones normales, es decir, el monto de las cuotas no recuperadas en los periodos que dure la recaudación total del crédito no pagado.

Por ejemplo para créditos amortizables, los valores expuestos varían según el tiempo de antigüedad, - al momento de la aprobación del crédito se debe establecer la provisión adecuada que el crédito del cliente en mención podría generar;, entonces la pregunta en este caso es ¿cuál será el saldo insoluto de capital del crédito al momento que este cliente caiga en default?. El saldo de capital al momento del default depende del tiempo de antigüedad del crédito en tal momento, puesto que la amortización depende del número de periodos transcurridos.

La determinación del tiempo de default dependerá de la disponibilidad de información que tenga cada institución, el modelo con el cual se trabaje debe ser el que mejor se adapte a la información disponible. Es por ello que el proceso de transición de cálculo de provisiones con el método impuesto por el supervisor al cálculo de las provisiones con un método desarrollado por la propia entidad ha sido bastante largo, puesto que se debe mantener la información adecuada para poder construir los modelos adecuados.

#### **4.2.3. SEVERIDAD DE LA PÉRDIDA**

El capital pendiente de pago al momento del incumplimiento no necesariamente se pierde en su totalidad, puesto que la institución financiera da inicio a un

proceso de recuperación para reducir o cancelar el crédito incumplido. La porción no recuperable disminuye según las garantías con las que se concedió el crédito.

La severidad de la pérdida se define como el valor presente de los flujos en los cuales se incurre en el caso de default dividido por el valor de exposición de dicho evento; es decir, representa el ratio entre las pérdidas efectivas incurridas, como consecuencia de la entrada en mora (incluyendo todos los gastos asociados con recuperaciones), y el monto pendiente de pago al momento del default. Adicionalmente, cabe aclarar que la severidad es el complemento del porcentaje de recuperación de la deuda.

Para el cálculo es necesario pasar todos los flujos asociados a la recuperación, tanto positivos (pagos, venta de activos) como negativos (costos internos, judiciales), al valor presente en el día de entrada en mora.

$$LGD = (1 - \%R)$$
$$\%R = \frac{VpR - CR}{EAD}$$

Donde:

*VpR*: Valor presente de recuperaciones.

*CR*: Valor presente de los Costos de recuperación.

La tasa de recuperación puede obtenerse en base a la información histórica de pago de de clientes que entraron en default, clasificándolos por tipo de producto (segmentación), para cada crédito se traen a valor presente todos los flujos a la tasa de contrato, este valor se divide por el total de exposición al momento del incumplimiento, obteniendo la tasa de recuperación. Luego, por tipo de bien o por tipo de clasificación se toman las tasas de recuperación y se realiza una prueba de bondad de ajuste para obtener el estimador de recuperación con mejores propiedades.

Según se ha explicado las metodologías detrás de cada uno de los términos de la ecuación de pérdida esperada tienen un estricto fundamento teórico y requieren de las bases de datos con la información suficiente para realizar cada uno de los modelos. El objetivo de este proyecto no es modelar la pérdida esperada sino detallar el modelo para obtener uno de los términos de la ecuación, la probabilidad de default. La obtención del resto de términos implican un estudio adicional que no consta en el alcance del presente análisis.

En el párrafo siguiente se describe cómo se contabilizan las provisiones y el efecto en el balance de resultados.

#### **4.3. AFECTACION DE LAS PROVISIONES EN EL ESTADO DE RESULTADOS**

El monto de provisiones al igual que la cartera de créditos de una institución financiera es un activo en negativo que consta en los Estados Financieros. La variación de las provisiones entre el mes  $i - 1$  y el mes  $i$  se denominará provisión generada del mes  $i$ ; en un mes determinado la provisión por un crédito específico puede incrementarse si el cliente incurre en una mora mayor o disminuir si el cliente cancela sus obligaciones en mora, entonces la provisión generada resume estas variaciones en las provisiones por todos los clientes de una cartera de créditos.

En una institución financiera las provisiones constan en el estado de resultados como un gasto, aún cuando estos valores no sean un flujo de caja, bajo tal consideración una provisión generada con signo positivo representa un gasto para la entidad financiera y por tanto disminuye su rentabilidad; mientras que si su signo es negativo en un mes determinado esta se contabiliza como un ingreso (reversión de provisiones).

La aplicación contable de las provisiones obedece a un principio básico de la contabilidad, conocido como *principio de prudencia* que textualmente implica

reconocer las posibles pérdidas tan pronto como se tenga conocimiento de ellas; mientras que los posibles beneficios no se reconocen hasta que no se realizan.

En el plan de cuentas que la Superintendencia de Bancos exige mantener a las instituciones financieras bajo su control, las provisiones se contabilizan en una cuenta de gasto, cuyo código es 44. Mensualmente esta cuenta se alimenta de la provisión generada en el mes. Si esta fuera negativa, se reversarían los valores que se encuentran en la cuenta de provisiones (44) y estas se enviarían a una cuenta de ingreso denominada Utilidades Financieras (53).

#### **4.4. MODELO DE ESTIMACION DE LAS PROVISIONES GENERADAS MENSUALMENTE**

Las entidades financieras, al igual que otro tipo de instituciones, realizan su presupuesto anual, el presupuesto considera la proyección del balance general y del estado de pérdidas y ganancias de al menos el próximo año de ejercicio. Por mencionar las partes principales a proyectarse se considerarán, montos de colocación de créditos según la línea (Activo – cartera de créditos), así como las fuentes de financiamiento de este crecimiento (Pasivo - Incremento de Depósitos a la vista y a largo plazo, endeudamiento con instituciones Nacionales y Extranjeras), en cuanto al balance general. La diferencia de precios entre el activo productivo y el pasivo con costo permiten determinar el margen bruto de la intermediación. Como se mencionó en 4.3; las provisiones son también parte del estado de pérdidas y ganancias; por tal razón también es necesario proyectarlas para la presupuestación.

Cabe aclarar que la proyección de las provisiones permite tener cifras más claras en cuanto a utilidades antes y después de participaciones e impuestos, y por ende índices de rentabilidad cuyo pronóstico es más confiable, que si no se considerara la cuenta de provisiones.



En los puntos anteriores se ha indicado el motivo por el cual las instituciones financieras realizan provisiones de su cartera de crédito, y adicionalmente se ha mencionado a breves rasgos la metodología para establecer provisiones, sugerida en los acuerdos de Basilea. En ella el desarrollo de un modelo scoring es una parte constitutiva, pues es un instrumento que permite el cálculo de la probabilidad de default, sin embargo la obtención de los otros términos de la ecuación de pérdida esperada conllevan la estimación de modelos individuales que fácilmente pueden ser el objeto de estudio de un trabajo posterior.

Considerando el enfoque práctico con el que se ha realizado el presente trabajo en este apartado se presentará una metodología sencilla para la estimación de las provisiones generadas mensualmente para una cosecha determinada. Por supuesto el modelo inicial dependerá del método de cálculo de provisiones con el que trabaje la entidad al momento de desarrollar el modelo de estimación.

#### **4.4.1. DESCRIPCION DE LA ESTRUCTURA DEL MODELO**

Para estimar las provisiones se correrán regresiones lineales simples para determinar la provisión generada como porción del monto bruto, usando como variable explicativa a la pérdida en cada cosecha<sup>29</sup>. La pérdida por cosecha es la proporción que se estima se perderá por la colocación realizada en una cosecha determinada, esta estimación será proporcionada por la entidad financiera.

La variable dependiente se calculó con la variación de provisiones observada en los últimos veinte y dos (22) meses de información, enero 2005 a octubre 2007, esta variación se expresó como un porcentaje del monto colocado en cada cosecha.

Los modelos de regresión utilizados describen el comportamiento de la provisión generada en el tiempo de antigüedad de la cartera, explicada por su respectiva

---

<sup>29</sup> La notación de cosechas se introdujo en el Capítulo II, en el apartado 2.1.1 Construcción del Indicador de Mora

pérdida de cosecha, con este objetivo se corren varias regresiones para las provisiones generadas, a un mes de antigüedad de la cartera, para dos meses de antigüedad, para tres meses y así sucesivamente. La provisión generada por la cartera colocada en el mismo mes de observación es de 1%, puesto que su calificación es A<sup>30</sup>, ya que dicha cartera no ha tenido oportunidad de caer en mora, pues aún no experimenta ni siquiera su primer vencimiento.

La información con la que se trabaja se dispone de forma similar a una matriz triangular superior, las filas corresponden al mes de cosecha y las columnas a los meses de observación posteriores, la tabla 2.2 ejemplifica la matriz construida, el primer elemento ( $a_{11}$ ) indica la provisión generada en ene-2005 para la cartera de la cosecha ene-2005; la siguiente posición corresponde a la provisión generada en feb-2005, por la cartera cosechada en ene-2005; de tal manera que la diagonal principal indica el porcentaje de provisión generada en el mismo mes de cosecha, la diagonal principal (primera diagonal) muestra la provisión generada a un mes de cosechada la cartera, la segunda diagonal superior la provisión generada a dos meses de la cosecha y así progresivamente. Puesto que el objetivo es usar las provisiones generadas de manera mensual para considerarlas en el presupuesto anual, se considerará hasta el mes doce de antigüedad, es decir se tomarán doce diagonales.

En la matriz, la diagonal resaltada en azul corresponde a la provisión generada al cuarto mes de antigüedad de la cosecha.

---

<sup>30</sup> La definición de las calificaciones de A hasta E se presentó en el Capítulo I

	ene-05	feb-05	mar-05	abr-05	may-05	jun-05	jul-05	ago-05	sep-05
ene-05	a <sub>11</sub>	a <sub>12</sub>	a <sub>13</sub>	a <sub>14</sub>	a <sub>15</sub>	a <sub>16</sub>	a <sub>17</sub>	a <sub>18</sub>	a <sub>19</sub>
feb-05	0	a <sub>22</sub>	a <sub>23</sub>	a <sub>24</sub>	a <sub>25</sub>	a <sub>26</sub>	a <sub>27</sub>	a <sub>28</sub>	a <sub>29</sub>
mar-05	0	0	a <sub>33</sub>	a <sub>34</sub>	a <sub>35</sub>	a <sub>36</sub>	a <sub>37</sub>	a <sub>38</sub>	a <sub>39</sub>
abr-05	0	0	0	a <sub>44</sub>	a <sub>45</sub>	a <sub>46</sub>	a <sub>47</sub>	a <sub>48</sub>	a <sub>49</sub>
may-05	0	0	0	0	a <sub>55</sub>	a <sub>56</sub>	a <sub>57</sub>	a <sub>58</sub>	a <sub>59</sub>
jun-05	0	0	0	0	0	a <sub>66</sub>	a <sub>67</sub>	a <sub>68</sub>	a <sub>69</sub>
jul-05	0	0	0	0	0	0	a <sub>77</sub>	a <sub>78</sub>	a <sub>79</sub>
ago-05	0	0	0	0	0	0	0	a <sub>88</sub>	a <sub>89</sub>
sep-05	0	0	0	0	0	0	0	0	a <sub>99</sub>

Tabla 4.4 : Ejemplo matriz de datos provisión generada

Las regresiones se corren tomando como variable independiente cada una de las diagonales.

La información de la provisión generada estuvo disponible de ene-2006 a octubre 2007, las cosechas estudiadas fueron de ene-2005 a oct-2007. Se dispuso los datos en la matriz de la manera indicada anteriormente, dejando nulas las posiciones para las que no existía información. Las diagonales se consideraron solo desde el primer mes con información (ene-2006); con ello cada una de las diagonales tiene 22 observaciones.

#### 4.4.2. CORRIDA DE REGRESIONES LINEALES

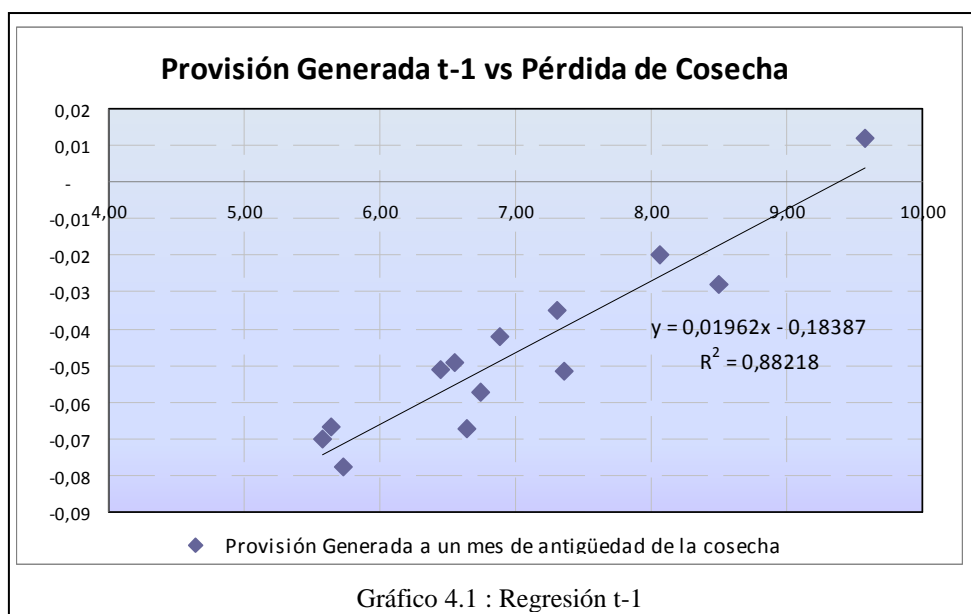
El objetivo en esta etapa es realizar 12 regresiones, del primer al doceavo mes de antigüedad de la cosecha, con ellos se verificará los residuos que deje la regresión y se proyectará la provisión generada mensualmente para los siguientes doce meses.

Una vez obtenidas las diagonales, cada una de ellas se relaciona con su respectiva pérdida de cosecha, de tal forma que la matriz de regresión tiene en la

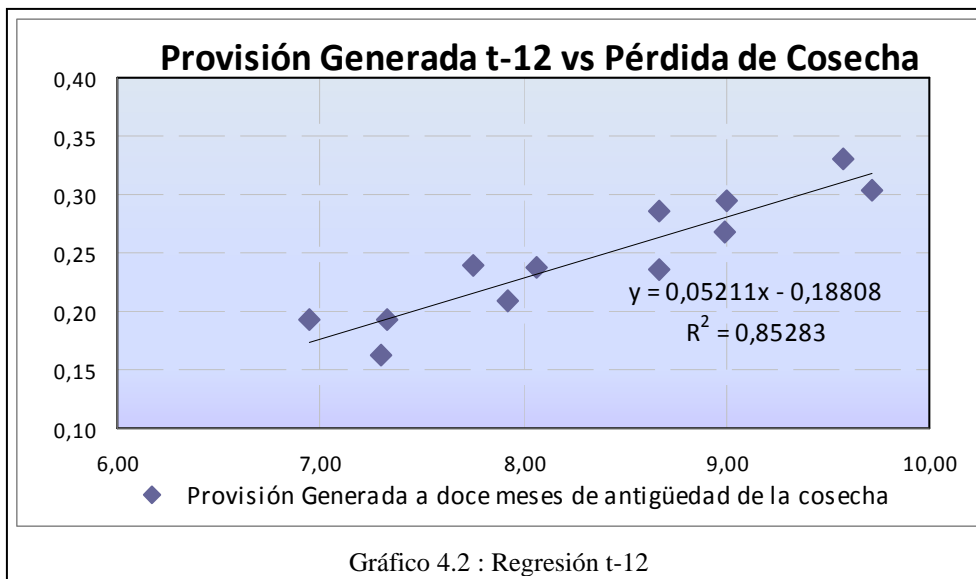
primera columna la provisión generada durante el mes  $j$  por la cosecha  $i$  y en la segunda columna la pérdida de la cosecha  $i$ . La provisión generada puede ser positiva o negativa, pues corresponde a la variación de un mes a otro.

Se denominará por  $t$  a la cosecha y  $t - j$ ;  $j = 1, 2, \dots, 12$ , a los meses de antigüedad relativos a la cosecha.

A continuación se presentarán las regresiones  $t - 1$  y  $t - 12$ ; mientras que los resultados de las regresiones intermedias constarán en el Anexo 2.



El gráfico indica que a un mes de generada una cosecha de, por ejemplo, pérdida estimada de 8% reducirá la provisión aproximadamente en el 0.03% del monto colocado en la cosecha. Los aportes a la provisión generada son negativos en el primer mes de antigüedad de la cosecha, mientras que en los meses posteriores los aportes pasan a ser positivos.



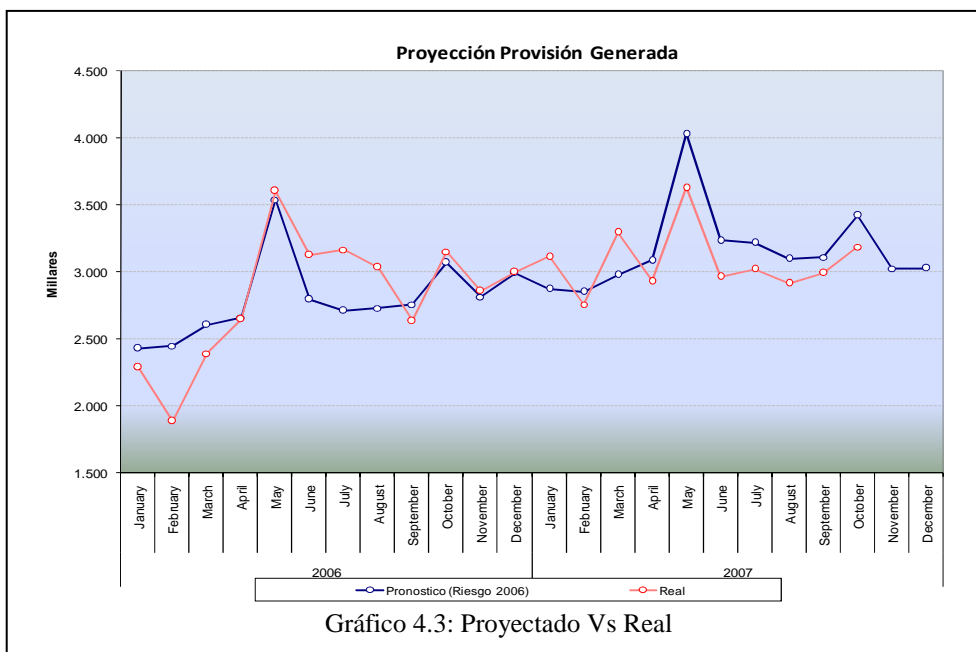
Bajo estas consideraciones se corrieron las doce (12) regresiones, obteniéndose las siguientes funciones generadoras de provisión:

	Función	R cuadrado
t	1,00	1,000
t-1	$y = 0,0196x - 0,1839$	0,882
t-2	$y = 0,1155x - 0,0071$	0,940
t-3	$y = 0,0920x - 0,2350$	0,840
t-4	$y = 0,1249x - 0,0004$	0,869
t-5	$y = 0,1282x + 0,0835$	0,913
t-6	$y = 0,0945x + 0,2016$	0,938
t-7	$y = 0,0628x + 0,3140$	0,921
t-8	$y = 0,0918x - 0,1214$	0,835
t-9	$y = 0,0686x - 0,0714$	0,925
t-10	$y = 0,0568x - 0,0595$	0,844
t-11	$y = 0,0900x - 0,4155$	0,944
t-12	$y = 0,0521x - 0,1881$	0,853

Tabla 4.5: Funciones generadoras de provisión

#### 4.4.3. MEDICIÓN DEL ERROR EN LA ESTIMACIÓN

Para determinar el error de la estimación se compararán lo realmente observado con lo obtenido mediante las funciones generadoras, para el periodo de estudio ene-2006 a oct-2007; en el gráfico 4.3 se presenta la comparación:



A simple vista se aprecia que lo obtenido con las funciones generadoras (proyectado), es muy similar a lo observado realmente, el error absoluto promedio de la estimación es 7.1%.

#### 4.4.4. PROYECCIÓN DE LA PROVISIÓN GENERADA

Para proyectar la provisión se requiere determinar la variable explicativa, en este caso la pérdida por cosecha. Esta variable se determinará haciendo uso de los perfiles de clientes obtenidos a partir del modelo scoring construido en el capítulo II.

La pérdida por cosecha dependerá entonces de los montos a colocarse por cada perfil, por supuesto esto depende exclusivamente de la estrategia que la institución financiera desee aplicar mensualmente en función de los objetivos a cumplir. Cabe recordar a cada perfil se asoció un porcentaje de pérdida, por lo cual la pérdida de la cosecha será el monto de colocación en cada perfil ponderado por la pérdida asociada al perfil respectivo.

##### 4.4.4.1. Pérdida por cosecha

La definición de montos de colocación por perfil permite encontrar la pérdida de cosecha, la tabla a continuación resume la proyección de colocación de un año

entero con los porcentajes estimados a colocar por cada perfil. Los meses de mayo y diciembre son diferentes y permiten más pérdida que el resto del año.

		Perfiles	P1	P2	P3	P4	P5		
		Pérdida por Perfil	4%	7%	12%	18%	24%		
		Proporción de Clientes Inicial	8%	17%	36%	21%	18%	Total Venta	Pérdida por Cosecha
ESTRATEGIA DE COLOCACIÓN	ENERO	40%	47%	13%	0%	0	100,00%	6,45%	
	FEBRERO	40%	47%	13%	0%	0	100,00%	6,45%	
	MARZO	40%	47%	13%	0%	0	100,00%	6,45%	
	ABRIL	40%	47%	13%	0%	0	100,00%	6,45%	
	MAYO	39%	45%	16%	0%	0	100,00%	6,63%	
	JUNIO	40%	47%	13%	0%	0	100,00%	6,45%	
	JULIO	40%	47%	13%	0%	0	100,00%	6,45%	
	AGOSTO	40%	47%	13%	0%	0	100,00%	6,45%	
	SEPTIEMBRE	40%	47%	13%	0%	0	100,00%	6,45%	
	OCTUBRE	40%	47%	13%	0%	0	100,00%	6,45%	
	NOVIEMBRE	40%	47%	13%	0%	0	100,00%	6,45%	
	DICIEMBRE	38%	44%	18%	0%	0	100,00%	6,76%	

Tabla 4.6 : Funciones generadoras de provisión

La estrategia que plantea la institución, junto a la identificación de perfiles de clientes, permite disminuir la pérdida de 14%<sup>31</sup> del monto colocado a un 7.99% (promedio ponderado de pérdida).

#### 4.4.4.2. Proyección de la provisión generada

La proyección se realizará para los próximos doce meses, en la tabla 4.4 la proyección corresponde a lo resaltado en color verde, para lo cual se usará la pérdida de cosecha determinada en el punto anterior.

<sup>31</sup> Ver Tabla 2.26 Perfiles por pérdida esperada

	ene-08	feb-08	mar-08	abr-08	may-08	jun-08	jul-08	ago-08	sep-08	oct-08	nov-08	dic-08	Pérdida de Cosecha
ene-07	t-12												10,63%
feb-07	t-11	t-12											9,56%
mar-07	t-10	t-11	t-12										8,98%
abr-07	t-9	t-10	t-11	t-12									8,05%
may-07	t-8	t-9	t-10	t-11	t-12								7,92%
jun-07	t-7	t-8	t-9	t-10	t-11	t-12							7,30%
jul-07	t-6	t-7	t-8	t-9	t-10	t-11	t-12						6,90%
ago-07	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12					6,94%
sep-07	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12				7,36%
oct-07	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12			6,64%
nov-07	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12		6,70%
dic-07	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12	6,64%
ene-08	to	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	6,45%
feb-08		to	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	6,45%
mar-08			to	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	6,45%
abr-08				to	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	6,45%
may-08					to	t-1	t-2	t-3	t-4	t-5	t-6	t-7	6,63%
jun-08						to	t-1	t-2	t-3	t-4	t-5	t-6	6,45%
jul-08							to	t-1	t-2	t-3	t-4	t-5	6,45%
ago-08								to	t-1	t-2	t-3	t-4	6,45%
sep-08									to	t-1	t-2	t-3	6,45%
oct-08										to	t-1	t-2	6,45%
nov-08											to	t-1	6,45%
dic-08												to	6,76%

Tabla 4.7: Funciones generadoras de provisión

Aplicando las funciones generadoras de provisión se obtienen los valores requeridos en la matriz de datos, la tabla 4.5 muestra la proyección de los porcentajes de provisión generada por las cosechas ene-08 a dic-08 en cada uno de los meses posteriores del mismo año.

La provisión generada en cada uno de los meses del año a presupuestar (2008) se calcula por:

$$PG_i = 1\% C_i + PG_i(t-1) * C_{i-1} + PG_i(t-2) * C_{i-2} + \dots + PG_i(t-12) * C_{i-12}$$

Dónde:

$PG_i$ , provisión generada en el mes  $i$ .

$C_i$ , Monto colocado en el mes o cosecha  $i$ .

$PG_i(t-1)$ , provisión generada en el mes  $i$  por la cosecha un mes anterior al mes  $i$ .



$PG_i(t-2)$ , provisión generada en el mes  $i$  por la cosecha dos meses anterior al mes  $i$ .

$PG_i(t-12)$ , provisión generada en el mes  $i$  por la cosecha doce meses anterior al mes  $i$ .

Se consideró hasta la cartera con doce meses de antigüedad, puesto que hasta este nivel de madurez donde se tiene un aporte significativo a la provisión generada.

Utilizando la técnica indicada se ha conseguido un modelo que estima las provisiones de una manera sencilla, práctica y sobre todo con un factor de error aceptable, permitiendo a la entidad financiera proyectar mes a mes los valores a provisionar por su cartera de créditos.

## **CAPITULO V**

### **CONCLUSIONES Y RECOMENDACIONES**

- La administración integral de riesgos juega actualmente un papel esencial en la gestión de las instituciones financieras. Los departamentos de riesgos se encargan de identificar los riesgos, medirlos, hacer seguimiento de su impacto en la operación y controlar sus efectos. Los acuerdos de capital de Basilea consideran los riesgos de mercado, liquidez, operativo y de crédito. Una adecuada administración de riesgos implica la contabilización de la exposición que se tiene por tipo de riesgos, y esto a su vez nos lleva a poder provisionar el dinero necesario para poder hacer frente a los siniestros cuando puedan suceder, así como establecer mecanismos para poder mitigarlos.
- Las políticas sugeridas por Basilea se están implementando paulatinamente en Ecuador, mediante el organismo de control (Superintendencia de Bancos y Seguros). Actualmente los requerimientos de la superintendencia se refieren principalmente a los riesgos de liquidez y de crédito, para los cuales algunas de las instituciones financieras ecuatorianas han desarrollado modelos estadísticos. En cuanto a la medición de los riesgos de mercado y operativo, su implementación ha sido más lenta, dándole mayor importancia al riesgo operativo.
- El crédito sigue un ciclo considerando la etapas de preventa, evaluación, seguimiento y cobranza; en cada una de ellas los análisis a realizar son diferentes debido a la información disponible al avanzar en el ciclo. Es factible aplicar modelos scoring exclusivos a cada etapa para aprovechar la información disponible y alcanzar objetivos propios de la fase de aplicación. Cabe nombrar la paradoja de la información relativa al ciclo de crédito, la cual menciona que al inicio del ciclo se cuenta con poca información y muchos clientes, mientras que al final la información es abundante pero los clientes

disminuyen. El modelo scoring desarrollado se enfoca en la preventa y evaluación.

- La aplicación de un modelo scoring consigue: mejorar la rentabilidad incrementando la tasa de aceptación y/o la reducción de la morosidad, ahorrar recursos, mejorar el servicio al cliente y apoyar a la toma de decisiones sin eliminar el criterio del analista de riesgo y la experiencia de los asesores de crédito.
- Se recomienda escoger una ventana de muestreo bajo los parámetros siguientes: estabilidad, madurez y representatividad. Para medir la estabilidad se recomienda la construcción de un indicador de comportamiento, se puede considerar una variable relevante en función de los objetivos planteados, en el presente estudio se utilizó como variable la mora, el indicador se analiza por cosechas de tal forma de obtener un periodo estable del indicador para poder hacer inferencias sobre los clientes. La madurez debe garantizar que la cartera a analizar provea la suficiente información en relación a su tiempo de vida, como para concluir a cerca del comportamiento crediticio del cliente, es decir, si es bueno o malo. La representatividad hace referencia al tamaño de muestra adecuado.
- Para definir la variable dependiente se usó una matriz cruzando las variables atraso promedio y atraso máximo, teniendo en cuenta la pérdida generada en cada rango de atraso considerado en la matriz. En base a la matriz y los porcentajes de pérdida asociados se clasificó a los clientes en buenos, malos e indeterminados, considerando los niveles de pérdida sugeridos por la institución financiera. Finalmente se consideran solo las definiciones de buenos y malos para construir la variable dependiente del modelo.
- La selección de variables independientes se realizó mediante árboles de decisión, específicamente aplicando el algoritmo CHAID. Tomando en cuenta siempre que las variables explicativas estén correlacionadas con la variable dependiente, así como la representatividad de la característica en la población.

Este tipo de herramientas además de identificar las variables independientes que más se correlacionan con el indicador de buenos y malos nos ayudan a encontrar combinaciones de variables donde la concentración de buenos o malos clientes aumenta significativamente, es decir el trabajo que se realiza con los árboles es de carácter multivariante; mejorando la capacidad de discriminación que se obtiene al solo analizar una variable, el potencial del análisis realizado radica en la utilización de los arboles para encontrar las variables o combinaciones de variables que mejor discriminen y se correlacionen con el indicador de buenos y malos, una vez encontradas las variables independientes se usa la regresión logística para encontrar un modelo para estimar la probabilidad de que un cliente sea bueno ó malo.

- Tradicionalmente los modelos de regresión logística binaria se han usado para muestras con 50% - 50% de observaciones en cada una de las dos categorías de la variable dependiente. Sin embargo, en procesos que consideran características como la de cliente bueno y malo, lo natural no es una muestra de iguales proporciones, la cantidad de clientes buenos es mucho más alta que la de malos con el fin de mantener la estabilidad del negocio. La solución a este problema ha sido el uso de modelos de regresión logística ponderada. Otra solución muy utilizada es considerar en el análisis a todos los clientes malos en la muestra, y usar el mismo número de clientes buenos; es decir se reduce la cantidad de clientes buenos, el problema que presenta esta solución es la posible exclusión de algunas características de clientes buenos. El aporte de este trabajo es la utilización de la muestra tal cual se seleccionó, es decir no se realizó ninguna corrección a priori, sino que se demostró que la utilización de muestras desproporcionadas en la corrida del modelo logístico, solo afecta al término constante que necesita reducirse por medio de  $\ln p_1 - \ln p_2$ , donde  $p_1$  y  $p_2$  son las proporciones de las observaciones elegidas de ambos grupos, para las cuales  $y = 1$  y  $0$  respectivamente.
- Se recomienda, antes de iniciar con la construcción de un modelo scoring, realizar una auditoría informática sobre los datos a ser utilizados; con el fin de garantizar la fidelidad y la consistencia de la información, así como levantar

una bitácora que indique el estado de las bases tanto internas como externas, por ejemplo, la bitácora debería incluir la fecha de la última actualización de la base. Los resultados del proceso de auditoría tienen un impacto en la implementación en tiempo real del modelo scoring.

- Una recomendación relevante para las instituciones financieras es que dentro de la estructura organizacional del departamento de riesgo se incluya un departamento de generación de información (datawarehouse), que trabaje de manera conjunta con el departamento de riesgo, pues sin información no se pueden realizar análisis ni modelos para la administración de riesgos.
- La prueba omnibus útil para analizar la significancia conjunta de los parámetros arrojados por el modelo, prueba la hipótesis de nulidad de todos los parámetros, contra la hipótesis alternativa de existencia de al menos un parámetro distinto de cero. Para el modelo realizado se obtuvo un valor crítico de 0.021, menor a 0.05, por lo que a un 95% de confianza no se rechazó la hipótesis nula.
- La prueba de Hosmer and Lemeshow arrojó un p-value de 0.38 por lo que no se rechazó la hipótesis nula. Esta es una prueba de bondad de ajuste, la hipótesis nula se refiere a que los datos se encuentran bien modelados, mientras que la hipótesis alternativa es una falta de ajuste en el modelo. Los resultados obtenidos indican entonces que el modelo está bien ajustado por medio de una regresión logística.
- La matriz de confusión construida con un punto de corte de 0.5 indica que el 67,47% de los clientes están bien clasificados cuando el punto de corte es el mencionado. El valor obtenido está dentro de los parámetros aceptables para un scoring de iniciación (65% de buena clasificación).
- Se recomienda, aplicar el algoritmo Boosting, que consiste en identificar los casos mal clasificados, darles mayor peso dentro de la muestra, y construir un nuevo modelo considerando los casos ponderados. Al realizar repetidamente

el algoritmo se consiguen modelos distintos, una manera de verificar si los modelos obtenidos son mejores al inicial es observar, mediante matrices de confusión, si el poder de discriminación se incrementa en los modelos posteriores.

- Para medir la eficiencia del modelo se utilizaron los índices Gini, K-S y la divergencia, obteniéndose 29.3%, 21.1% y 93.89% respectivamente. Los valores resultantes se encuentran dentro de los parámetros empíricamente establecidos por consultoras internacionales, a pesar de que el modelo se construyó con información exclusivamente externa a la institución financiera, y los parámetros se han determinado en base a modelos que consideran información de la solicitud de crédito (con un proceso de verificación). Se recomienda usar los valores empíricos considerados por Lisim para los modelos en etapa de preventa y evaluación, según la mencionada consultora un K-S aceptable debe ser mayor al 20%, mientras que el índice Gini debe ubicarse alrededor del 30% y para la divergencia sugiere un valor mayor a 0.95.
- La validación del modelo con datos de prueba tiene el objetivo de verificar la transportabilidad del modelo de una muestra a otra. Con la muestra de control se corrió el modelo y se observó que los índices de eficiencia resultaron muy similares a los obtenidos con la muestra de desarrollo. Se recomienda que la muestra con la que se realice el proceso de control del modelo incluya cosechas diferentes a las seleccionadas para el desarrollo, sin embargo, en el presente trabajo no se pudo realizar en esa forma, ya que no se disponía de la información necesaria.
- La selección de puntos de corte puede ser realizada utilizando la curva ROC o por medio de matrices de confusión. La curva ROC permite dos opciones para determinar los puntos de corte; la primera determinando en la curva el score que maximiza el índice K-S; la segunda forma de encontrar un punto de corte adecuado es usar el concepto de pérdida por mala clasificación, se consideran la pérdida de rentabilidad por clasificar un bueno como malo y la pérdida

esperada por clasificar un malo como bueno, para aplicar esta opción no se contó con los datos suficientes a cerca de las rentabilidades requeridas. La aplicación de matrices de confusión incluye analizar parámetros de relevancia para seleccionar puntos de corte, entre estos parámetros se cuentan: sensibilidad, especificidad, razón de rechazo, porcentajes de clasificación correcta e incorrecta. Las técnicas mencionadas anteriormente son excluyentes y en la práctica resultarían en una alta razón de rechazo, lo cual no es conveniente para la institución financiera, por ese motivo se decidió usar más de un punto de corte, estos puntos se determinaron en base a la pérdida promedio generada por los clientes ubicados entre un punto y otro. En base a este criterio se definieron perfiles de clientes con porcentajes de pérdida y participación acordes a los deseados por la institución financiera. La definición de perfiles permite a la institución determinar flujos de crédito y políticas de riesgo propias a las características de cada perfil, por ejemplo, para los perfiles de más bajo riesgo los flujos de crédito pueden ser inmediatos, mientras que a medida que el riesgo incrementa en los perfiles los flujos de crédito pueden requerir garantías o información adicional, incluso se podría excluir alguno de los perfiles si la pérdida promedio generada no es una oferta interesante para la entidad. Para los datos analizados la exclusión del último perfil obtenido implica una tasa de rechazo del scoring del 18%, manteniendo el resto de perfiles la pérdida promedio de la institución sería del 12%, se sugiere implementar aprobaciones de crédito instantáneas para el primer y segundo perfil, para el tercer perfil la aprobación previa presentación de solicitud y su respectiva verificación, en el cuarto perfil además de la verificación de la información de la solicitud se propone requerir garantías que avalen la operación de crédito.

- Un modelo scoring debe ser monitoreado en el tiempo, pues este puede sufrir cambios en su eficiencia y estructura, el monitoreo debe permitir determinar si las aplicaciones se están calificando correctamente, si la población actual es diferente a la de construcción, si la razón de rechazo difiere o no de lo inicialmente pensado; si el modelo ordena la cartera generada, es decir, el

monitoreo se utilizará para comparar el comportamiento inicial del modelo con el comportamiento actual del mismo.

- La metodología desarrollada en el capítulo III permite analizar al modelo scoring desde dos puntos de vista los resultados del modelo y la estructura del mismo. Dentro de los resultados del modelo se analiza la pérdida y la distribución de clientes por perfil, la razón de rechazo y los índices de eficiencia. Si en esta primera etapa del monitoreo se aprecia que los resultados no son aceptables se sugiere continuar con el monitoreo estructural del modelo. En la estructura del modelo se analizan las correlaciones de las variables independientes vs la dependiente, la alineación de los coeficientes de cada una de las variables, la intercorrelación de las variables independientes (para detectar posibles problemas de multicolinealidad) y la estabilidad de la población sobre la que se ha aplicado el modelo, es decir si las características de los nuevos clientes, referentes a las variables que intervienen en el modelo, se mantienen con respecto a la muestra de desarrollo.
- Los reportes mencionados anteriormente son indicadores de alertas de posibles problemas en el modelo scoring, por otro lado el Enfoque Delta (Delta Approach) permite determinar mediante pruebas de hipótesis de chi cuadrado la alineación por bandas de score, y además estima una medida de corrección del modelo, al establecer un valor de desviación (un delta) que permita alinear los coeficientes.
- Se recomienda, puesto que el monitoreo se debe realizar periódicamente, realizar una implementación automática de los reportes de monitoreo sugeridos en la metodología, para dedicar mayor tiempo al análisis que a la construcción de los reportes. Cabe recalcar que la realización del monitoreo requiere una estructura de base de datos adecuada para evitar problemas de extracción de información, esto implica disponer de un datawarehouse exclusivo para el área de riesgos.



- Las entidades financieras están obligadas por los entes de control a establecer provisiones por los riesgos a los que se enfrenta en su actividad diaria. Por ejemplo por riesgo de crédito se establecen provisiones de la cartera de créditos, es decir que la cartera no se contabiliza como un activo completo. La provisión generada de un mes a otro se contabiliza como un gasto, y por tanto reduce la rentabilidad de la institución. Es por ello necesario establecer modelos estadísticos que permitan provisionar adecuadamente, lo suficiente como para enfrentar la pérdidas esperadas pero no demasiadas tal que la rentabilidad se vea afectada.
- La estimación de las provisiones a generarse mensualmente son importantes para la realización del presupuesto anual de la entidad, pues esta partida podría ser variada en el tiempo de presupuestación en base a un modelo estadísticamente verificable, permitiendo así obtener resultados más cercanos a la realidad. Se construyó un modelo de estimación de la provisión generada mensualmente (variación mensual de la provisión) compuesto por una matriz de funciones lineales que permiten establecer una proyección de la provisión a generarse mensualmente en un horizonte temporal determinado. Los datos requeridos para el modelo de estimación se obtienen de la estrategia de colocación por perfil establecida por la institución financiera y de la pérdida de cosecha estimada para las futuras colocaciones. La pérdida de cosecha es un valor calculado por la entidad.

# ANEXO I

## TEORIA

### ÁRBOLES DE DECISIÓN

Un árbol de decisión no utiliza un modelo estadístico formal y es más bien un algoritmo para clasificar utilizando particiones binarias sucesivas de los valores de una variable cada vez, esta partición recursiva se traduce en una organización jerárquica del espacio de representación que puede modelarse mediante una estructura de tipo árbol. Cada nodo interior contiene una pregunta sobre un atributo concreto (con un hijo por cada posible respuesta) y cada nodo hoja se refiere a una decisión (clasificación). Las decisiones a tomar en la construcción de los árboles de decisión, para este caso en específico son:

- La selección de las variables y de sus puntos de corte para hacer las divisiones, regla de partición.
- La asignación de los nodos terminales a categoría Bueno o Malo, decisión de asignación. Con esto se determina cual es el signo esperado de la variable.

La decisión de asignación se toma en función de la proporcionalidad de las categorías Bueno-Malo en el nodo inicial, normalmente se asigna el nodo como bueno, si la proporcionalidad se mantiene o se concentra en la categoría de Buenos; por ejemplo si la proporcionalidad inicial es 70/30, un nodo será considerado como bueno si existe una concentración en la categoría de Buenos mayor a la original.

- **Algoritmo Chaid:**

1. Existen varios algoritmos para construir árboles de decisión, las diferencias principales entre los distintos algoritmos radican en las estrategias de poda

o terminación y en la regla adoptada para particionar nodos. El algoritmo a utilizarse para la consecución del fin mencionado, será el algoritmo CHAID, las siglas CHAID corresponden al término inglés Chi- squared Automatic Interaction Detector (detector automático de interacciones mediante chi-cuadrado). Es una técnica estadística desarrollada por Kass (1980) muy eficaz para segmentar o generar árboles. El método CHAID, que utiliza la significación de una prueba estadística como criterio, evalúa todos los valores de una variable predictora potencial. Funde los valores considerados estadísticamente homogéneos respecto a la variable criterio y conserva inalterados todos los valores heterogéneos.

A continuación, el algoritmo selecciona la mejor variable predictora para formar la primera rama del árbol de decisión, de forma que cada nodo esté compuesto por un grupo de valores homogéneos de la variable seleccionada. Este proceso se repite hasta que el árbol se ha desarrollado por completo. La prueba estadística utilizada depende del nivel de medida de la variable criterio.

No se trata de un método binario, es decir, puede generar más de dos categorías en cualquier nivel del árbol. Por lo tanto tiende a crear un árbol más ancho que otros métodos. Así también, aprovecha los valores perdidos, tratándolos como una categoría válida individual.

CHAID funciona con todos los tipos de variables continuas y categóricas. Sin embargo, las variables predictoras continuas se categorizan automáticamente para el análisis. Algunas de las opciones mencionadas a continuación se pueden definir mediante las opciones avanzadas de CHAID. Estas opciones incluyen la posibilidad de seleccionar la prueba de Chi-cuadrado de Pearson o la de la razón de verosimilitud, el nivel de  $\alpha_{merge}$  y el nivel  $\alpha_{split}$ . A continuación se describen, en una forma básica, los pasos del algoritmo:

6. Para cada variable predictora  $X$ , buscar el par de categorías de  $X$  cuya diferencia presenta la menor significación (es decir, la que presente el

mayor nivel crítico) respecto a la variable criterio Y. El método usado para calcular el nivel crítico depende del nivel de medida de Y.

- Si Y es continua, utilizar la prueba F
- Si Y es nominal, se forma una tabla de contingencia con las categorías de X como filas y las categorías de Y como columnas. Utilizar la prueba de chi-cuadrado de Pearson o la prueba de la razón de verosimilitud.
- Si Y es ordinal, se ajusta un modelo de asociación de Y (Clogg y Eliasin, 1987; Goodman, 1979 y Madgidson, 1992). Utilizar la prueba de la razón de verosimilitud.

7. Para el par de categorías de X con el mayor nivel crítico, se compara el valor del nivel crítico con el nivel alfa preestablecido,  $\alpha_{merge}$ .

- Si el nivel crítico es mayor que  $\alpha_{merge}$ , fundir este par en una sola categoría compuesta, como resultado, se forma un nuevo conjunto de categorías de X y el proceso se vuelve a iniciar desde el paso 1.
- Si el nivel crítico es menor que  $\alpha_{merge}$  ir al paso 3.

8. Calcular el nivel crítico corregido para el conjunto de las categorías de X y las categorías de Y, mediante la corrección de Bonferroni correspondiente.

9. Seleccionar la variable predictora X cuyo nivel crítico corregido sea el menor (la que sea más significativa). Comparar su nivel crítico con el nivel alfa preestablecido  $\alpha_{split}$ .

- Si el nivel crítico preestablecido es menor o igual que  $\alpha_{split}$ , dividir el nodo conforme al número de categorías de X.
- Si el nivel crítico preestablecido es mayor que  $\alpha_{split}$ , no dividir el nodo. Este nodo es un nodo terminal.

10. Continuar con el proceso de desarrollo del árbol hasta que se cumpla una de las reglas de parada.

La variable criterio a usarse para la construcción de los árboles de decisión es la variable Tipo Cliente, en cada fase del análisis, CHAID realiza una prueba de independencia Chi Cuadrado entre la variable dependiente y la variable predictora escogida.

## PRUEBA DE INDEPENDENCIA CHI CUADRADO

Consistente en comprobar si dos características cualitativas están relacionadas entre sí (por ejemplo: ¿el color de ojos está relacionado con el color de los cabellos?). Este tipo de contraste se aplica cuando deseamos comparar una variable en dos situaciones o poblaciones diferentes, i.e., deseamos estudiar si existen diferencias en las dos poblaciones respecto a la variable de estudio.

Supongamos que de  $n$  elementos de una población se han observado dos características  $X$  e  $Y$ , obteniéndose una muestra aleatoria simple bidimensional  $(X_1, Y_1); (X_2, Y_2); \dots; (X_n, Y_n)$ . Sobre la base de dichas observaciones se desea contrastar si las características poblacionales  $X$  e  $Y$  son independientes o no. Para ello se dividirá el conjunto de posibles valores de  $X$  en  $k$  conjuntos disjuntos  $A_1, A_2, \dots, A_k$ ; mientras que el conjunto de posibles valores  $Y$  será descompuesto en  $r$  conjuntos disjuntos:  $B_1, B_2, \dots, B_r$ . Al clasificar los elementos de la muestra, aparecerá un cierto número de ellos,  $n_{ij}$ , en cada una de las  $k \times r$  clases así constituidas, dando lugar a una tabla de contingencia de la forma:

	$A_1$	$A_2$	...	$A_k$	Total
$B_1$	$n_{11}$	$n_{12}$		$n_{1k}$	$n_{1.}$
$B_2$	$n_{21}$	$n_{22}$		$n_{2k}$	$n_{2.}$
...					
$B_r$	$n_{r1}$	$n_{r2}$		$n_{rk}$	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.k}$	$n$

El estadístico del contraste será

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Sigue una distribución  $\chi^2$  con  $(k-1)(r-1)$  grados de libertad.

Donde:  $e_{ij} = \frac{n_i * n_j}{n}$

## REGRESIÓN LOGÍSTICA

La regresión logística o modelo Logit se utiliza cuando se desea predecir un resultado binario, por ejemplo, quiebra vs no quiebra y se sabe que existen varios factores que pueden incidir sobre tal resultado. Ésta regresión binaria es un tipo de análisis de regresión donde la variable dependiente es una variable dummy, por ejemplo: código 0 (Buen Cliente) o 1 (Mal Cliente).

- **Distribución Logística y Función Logística:**

La distribución logística se utiliza en el estudio del crecimiento temporal de variables, en particular, demográficas. Depende de dos parámetros la media  $\alpha$  y la desviación estándar  $\beta$ . Los valores de la variable aleatoria son no negativos, mientras que el dominio de la distribución se mueve en todo el eje real.

$$F(x) = \frac{\exp\left(-\frac{x-\alpha}{\beta}\right)}{\beta\left(1 + \exp\left(-\frac{x-\alpha}{\beta}\right)\right)^2}$$

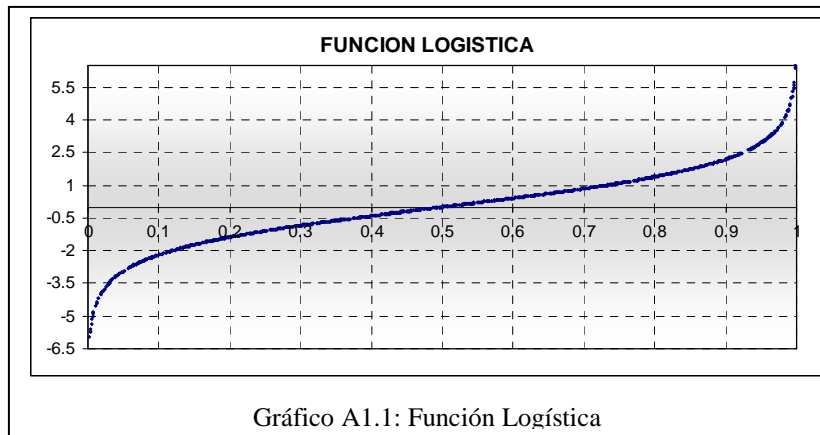
Si se hace el cambio de variable  $y = \frac{x-\alpha}{\beta}$ , se tiene la distribución logística estándar con función de densidad dada por:

$$F(x) = \frac{\exp(-y)}{(1 + \exp(-y))^2}$$

Por otro lado la función logística estándar para una variable  $x$ , está dada por:

$$f(x) = \ln\left(\frac{x}{1-x}\right)$$

El gráfico de la función logística se presenta a continuación:



- **Supuestos:**

El modelo logit o logístico se aplica a una amplia gama de situaciones donde las variables explicativas no tienen una distribución conjunta normal multivariante. Por ejemplo, si algunas son categóricas, podemos introducirlas en el modelo logit mediante variables ficticias, como se hace en el modelo de regresión estándar. En particular, si todas las variables son binarias independientes y llamamos  $p_1=(p_{11},\dots,p_{1p})$  a los parámetros de la primera población y  $p_2=(p_{21},\dots,p_{2p})$  a los de la segunda y observamos un elemento  $x_i = (x_{1p},\dots,x_{ip})$ , tendremos que, suponiendo que las probabilidades a priori son las mismas:

$$P(y = 1|x_i) = \frac{P(x_i|y = 1)P(y = 1)}{P(x_i)} = c \prod_{j=1}^p p_{1j}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}}$$

donde  $c=P(y=1)/P(x_i)$ . La transformación logística será

$$g_i = \log \frac{p(y = 1|x_i)}{1 - P(y = 1|x_i)} = \sum x_{ij} \log \left( \frac{p_{1j}}{p_{2j}} \right) + \sum (1 - x_{ij}) \log \left[ \frac{(1 - p_{1j})}{(1 - p_{2j})} \right]$$

que es una función lineal en las variables, que podemos escribir

$$g_i = \beta_0 + \beta_1' x_i$$

donde 
$$\beta_o = \sum \log \left[ \frac{(1-p_{ij})}{(1-p_{2j})} \right] \text{ y } \beta_1 = \sum \log \left[ \frac{p_{1j}(1-p_{2j})}{p_{2j}(1-p_{1j})} \right]$$

Par tanto, esperamos que este modelo se comporte bien cuando todas las variables de clasificacion son binarias y aproximadamente independientes.

Una ventaja adicional de este modelo es que cuando las variables son normales también verifican el modelo logístico. En efecto, supongamos dos poblaciones normales multivariantes con distinta media pero la misma matriz de varianzas covarianzas. Suponiendo las probabilidades a priori de ambas poblaciones iguales:

$$p_i = P(y = 1|x_i) = \frac{f_1(x_i)}{f_1(x_i) + f_2(x_i)} \quad (1.8)$$

utilizando la transformación logit, (1.7):

$$g_i = \log \frac{f_1(x_i)}{f_2(x_i)} = -\frac{1}{2}(x_i - \mu_1)'V^{-1}(x_i - \mu_1) + \frac{1}{2}(x_i - \mu_2)'V^{-1}(x_i - \mu_2)$$

y simplificando

$$g_i = \frac{1}{2}(\mu_2 V^{-1} \mu_2 - \mu_1 V^{-1} \mu_1) + (\mu_1 - \mu_2)' V^{-1} x_i$$

Par tanto,  $g_i$  es una función lineal de las variables  $x$ , que es la característica que define el modelo logit. Comparando con (1.7) la ordenada en el origen,  $\beta_o$ , es igual

$$\beta_o = \frac{1}{2}(\mu_2 V^{-1} \mu_2 - \mu_1 V^{-1} \mu_1) = -\frac{1}{2} w'(\mu_1 + \mu_2)$$

donde  $w = V^{-1}(\mu_1 - \mu_2)$  y el vector de pendientes

$$\beta_1 = w$$



Observemos que la estimación de  $\hat{w}$  mediante el modelo logístico no es eficiente en el caso normal. En efecto, en lugar de estimar los  $p(p+1)/2$  términos de la matriz  $\hat{V}$  y los  $2p$  de las medias  $\bar{x}_1$  y  $\bar{x}_2$ , con el modelo logístico estimamos únicamente  $p+1$  parámetros  $\beta_0, \beta_1, \dots, \beta_p$ . En el caso de normalidad se obtiene un mejor procedimiento con la regla de Fisher, que estima,  $\hat{V}, \bar{x}_1$  y  $\bar{x}_2$  la distribución completa de las  $x$ , mientras que el modelo logístico estima solo los  $p+1$  parámetros de la distribución de  $y$  condicionada a  $x$ . Como:

$$f(x, y) = f(y|x)f(x)$$

Perdemos información al considerar solo la condicionada  $f(y|x)$  -como hace el modelo logístico- en lugar de la conjunta  $f(x, y)$ . Efron (1975) demostró que cuando los datos son normales multivariantes y estimamos los parámetros en la muestra, la función de discriminación lineal de Fisher funciona mejor que regresión logística

En resumen, en el caso de normalidad la regla discriminante es mejor que el modelo logístico. Sin embargo, la función logística puede ser más eficaz cuando las poblaciones tengan distinta matriz de covarianzas o sean marcadamente no normales. En el campo de la concesión automática de créditos (Credit Scoring) existen numerosos estudios comparando ambos métodos. La conclusión general es que ninguno supera al otro de manera uniforme. Rosenberg y Gleit (1994) y Hand y Henley (1997) han presentado estudios sobre este problema.

- **Interpretación del modelo**

Los parámetros del modelo son  $\beta_0$ , la ordenada en el origen, y  $\beta = (\beta_1, \dots, \beta_p)$  las pendientes. A veces se utilizan también como parámetro  $\exp(\beta_0)$  y  $\exp(\beta_i)$ , que se denominan los odds ratios o ratios de probabilidades, e indican cuanto se modifican las probabilidades por unidad de cambio en las variables  $x$ . En efecto, de (1.7) deducimos que

$$O_i = \frac{p_i}{1 - p_i} = \exp(\beta_0) \prod_{j=1}^p \exp(\beta_j)^{x_j}$$

Supongamos dos elementos, i, k, con todos los valores de las variables iguales excepto la variable h y  $X_{ih} = X_{kh} + 1$ . El cociente de los ratios de probabilidades (odds ratio) para estas dos observaciones es:

$$\frac{O_i}{O_k} = e^{\beta_h}$$

e indica cuanto se modifica el ratio de probabilidades cuando la variable  $X_h$  aumenta una unidad. Sustituyendo  $\hat{p}_i = 0.5$  en el modelo logit, entonces,

$$\log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} = 0,$$

es decir,

$$x_{i1} = -\frac{\beta_0}{\beta_1} - \sum_{j=2}^p \frac{\beta_j x_{ij}}{\beta_1},$$

y  $x_{i1}$  representa el valor de  $x_1$  que hace igualmente probable que un elemento, cuyas restantes variables son  $x_{i2}, \dots, x_{ip}$ , pertenezca a la primera o la segunda población.

- **Estimación**

Supondremos una muestra aleatoria de datos  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ . La función de probabilidades para una respuesta  $y_i$  cualquiera es:

$$P(y_i) = p_i^{y_i} (1 - p_i)^{1 - y_i} \quad y_i = 0, 1$$

y para la muestra:

$$P(y_1, \dots, y_n) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1 - y_i}$$

Tomando logaritmos:

$$\log P(y) = \sum_{i=1}^n y_i \log \left( \frac{p_i}{1 - p_i} \right) + \sum \log(1 - p_i) \quad (1.9)$$

la función soporte (de verosimilitud en logaritmos) puede escribirse como

$$\log P(\beta) = \sum_{i=1}^n (y_i \log p_i + (1 - y_i) \log(1 - p_i)) \quad (1.10)$$

donde  $\beta = (\beta_0, \beta_1, \dots, \beta_p)$  es un vector de  $p+1$  componentes, incluyendo la constante  $\beta_0$  que determina las probabilidades  $p_i$ . Maximizar la verosimilitud puede expresarse como minimizar una función que mide la desviación entre los datos y el modelo. Se define la desviación de un modelo mediante  $D(\theta) = -2L(\theta)$  y, por tanto, la desviación del modelo será:

$$D(\beta) = -2 \sum_{i=1}^n (y_i \log p_i + (1 - y_i) \log(1 - p_i)) \quad (1.11)$$

y hablaremos indistintamente de maximizar el soporte o minimizar la desviación del modelo. Se define la desviación (deviance) de cada dato por:

$$d_i = -2(y_i \log p_i + (1 - y_i) \log(1 - p_i)) \quad (1.12)$$

y mide el ajuste del modelo al dato  $(y_i, x_i)$ . En efecto, observemos, en primer lugar, que como los  $\hat{p}_i$  son menores que uno, sus logaritmos son negativos, por lo que la desviación es siempre positiva. Además, en el cálculo de la desviación solo interviene uno de sus dos términos, ya que  $y_i$  solo puede valer cero o uno. Entonces:

Si  $y_i = 1$ , y la observación pertenece a la segunda población, el segundo término de la desviación es nulo y  $d_i = -2 \log p_i$ . La observación tendrá una desviación grande si la probabilidad estimada de pertenecer a la segunda población,  $p_i$ , es pequeña, es decir, cuando esta observación está mal explicada por el modelo.

Si  $y_i = 0$ , y la observación pertenece a la primera población, solo interviene el segundo término de la desviación  $d_i = -2 \log(1 - p_i)$ . La desviación será grande si  $p_i$  es grande, lo que indica que la probabilidad de pertenecer a la verdadera población es pequeña y el modelo ajusta mal dicho dato.

Para maximizar la verosimilitud, expresando  $p_i$  en función de los parámetros de interés,  $\beta$ , en (1.9), obtenemos la función soporte:

$$L(\beta) = \sum_{i=1}^n y_i x_i \beta - \sum_{i=1}^n \log(1 + e^{x_i \beta}) \quad (I.13)$$

que derivaremos para obtener los estimadores MV. Escribiendo el resultado como vector columna:

$$\frac{\partial L(\beta)}{\partial \beta} = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \left( \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} \right) \quad (I.14)$$

e igualando este vector a cero y llamando  $\hat{\beta}$  a los parámetros que satisfacen el sistema de ecuaciones:

$$\sum_{i=1}^n y_i x_i = \sum_{i=1}^n x_i \left( \frac{1}{1 + e^{-x_i \beta}} \right) = \sum_{i=1}^n \hat{y}_i x_i \quad (I.15)$$

Estas ecuaciones establecen que el producto de los valores observados por las variables explicativas debe ser igual al de los valores previstos. También, que los residuos del modelo,  $e_i = y_i - \hat{y}_i$  deben ser ortogonales a las variables x. Esta condición es análoga a la obtenida en el modelo de regresión estándar, pero ahora el sistema (I.15) resultante no es lineal en los parámetros  $\beta$ . Para obtener el valor  $\hat{\beta}_{MV}$  que maximiza la verosimilitud, acudiremos a un algoritmo tipo Newton-

Raphson. Desarrollando el vector  $\left( \frac{\partial L(\beta)}{\partial \beta} \right)$  alrededor de un punto  $\beta_a$ , se tiene

$$\frac{\partial L(\beta)}{\partial \beta} = \frac{\partial L(\beta_a)}{\partial \beta} + \frac{\partial^2 L(\beta_a)}{\partial \beta \partial \beta'} (\beta - \beta_a);$$

para que el punto  $\beta_a$  corresponda al máxima de la verosimilitud su primera derivada debe anularse. Imponiendo la condición  $\partial L(\beta_a) / \partial \beta = 0$ , se obtiene:

$$\beta_a = \beta + \left( -\frac{\partial^2 L(\beta_a)}{\partial \beta \partial \beta'} \right)^{-1} \left( \frac{\partial L(\beta)}{\partial \beta} \right) \quad (I.16)$$

que expresa como obtener el punto máxima  $\beta_a$ , a partir de un punto próximo cualquiera  $\beta$ . La ecuación depende de la matriz de segundas derivadas, que, en el

óptimo, es la inversa de la matriz de varianzas y covarianzas asintótica de los estimadores MV. Para obtener su expresión, derivando por segunda vez en (I.14), se obtiene:

$$\hat{M}^{-1} = \left( -\frac{\partial^2 L(\beta)}{\partial \beta \partial \beta'} \right) = \sum_{i=1}^n x_i x_i' \omega_i \quad (I.17)$$

donde los coeficientes  $\omega_i$  están dados por:

$$\omega_i = \frac{e^{x_i \beta}}{(1 + e^{x_i \beta})^2} = p_i (1 - p_i) \quad (I.18)$$

Sustituyendo en (I.16) las expresiones (I.17) y (I.14) y evaluando las derivadas en un estimador inicial  $\hat{\beta}$ , se obtiene el siguiente método para obtener un nuevo valor del estimador,  $\beta_a$ , a partir del  $\hat{\beta}$ .

$$\beta_a = \hat{\beta} + \left( \sum_{i=1}^n x_i x_i' \omega_i \right)^{-1} \left( \sum x_i (y_i - \hat{p}_i) \right) \quad (I.19)$$

donde  $\hat{p}_i$  y  $\omega_i$  se calculan con el valor  $\hat{\beta}$ . El algoritmo puede escribirse como:

$$\boxed{\beta_a = \hat{\beta} + (X' \hat{W} X)^{-1} X' (Y - \hat{Y})} \quad (I.20)$$

donde  $\hat{W}$  es una matriz diagonal con términos  $\hat{p}_i (1 - \hat{p}_i)$  y  $\hat{Y}$  es el vector de valores esperados de Y. La matriz de varianzas y covarianzas de los estimadores así obtenidos es aproximadamente, según (I.20),  $(X' \hat{W} X)^{-1}$ . Observemos que la ecuación (I.20) indica que debemos modificar el estimador si los residuos no son ortogonales a las variables explicativas, es decir, si  $X' (Y - \hat{Y}) \neq 0$ . La modificación del estimador depende de esta diferencia y se reparte entre los componentes de  $\hat{\beta}$  en función de su matriz de varianzas y covarianzas estimada.

La forma habitual de implementar este método es el siguiente algoritmo iterativo que proporciona, en convergencia, el estimador MV de  $\beta$ .

1. Fijar un valor arbitrario inicial,  $\hat{\beta}_1$ , para los parámetros y obtener el vector  $\hat{Y}_1$  para dicho valor en el modelo logit. Por ejemplo, si  $\hat{\beta}_1 = 0$ ,

$$\hat{y}_i = \hat{p}_i = \frac{1}{1 + e^{-0}} = \frac{1}{2}$$

y el vector  $\hat{Y}$  tiene todas sus componentes iguales a 1/2.

2. Definir una variable auxiliar  $z_i$  de residuos estandarizados por:

$$z_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i(1 - \hat{y}_i)}} = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}}$$

o vectorialmente:

$$Z = \hat{W}^{-\frac{1}{2}}(Y - \hat{Y})$$

donde  $\hat{W}$  es una matriz diagonal con términos  $\hat{y}_i(1 - \hat{y}_i)$ .

3. Estimar por mínimos cuadrados una regresión con variable dependiente Z y matriz de regresores  $T = \hat{W}^{\frac{1}{2}}X$ . Los parámetros estimados con esta regresión,  $\hat{b}_1$ , vendrán dados por:

$$\begin{aligned} \hat{b}_1 &= (T'T)^{-1}T'Z \\ &= (X'\hat{W}X)^{-1}X'(Y - \hat{Y}) \end{aligned}$$

y, comparando con (I.20), vemos que  $\hat{b}_1$  estima el incremento  $\beta_a - \hat{\beta}_1$  de los parámetros que nos acerca al máximo.

4. Obtener un nuevo estimador de los parámetros  $\hat{\beta}_2$  del modelo logístico mediante

$$\hat{\beta}_2 = \hat{\beta}_1 + \hat{b}_1$$

5. Tomar el valor estimado resultante de la etapa anterior, que en general llamaremos  $\hat{\beta}_h$  y sustituirlo en la ecuación del modelo logístico para obtener el vector de estimadores  $\hat{Y}(\hat{\beta}_h) = \hat{Y}_h$ . Utilizando este vector  $\hat{Y}_h$  construir la matriz  $\hat{W}_h$  y las nuevas variables  $Z_h$  y  $T_h$

$$Z_h = \hat{W}_h^{-\frac{1}{2}}(Y - \hat{Y}_h)$$

$$T_h = \hat{W}_h^{-\frac{1}{2}}X$$

y volver a la etapa 2. El proceso se repite hasta obtener la convergencia ( $\beta_{h+1} \approx \beta_h$ ).

- **Pruebas de Validación:**

**Prueba de Wald:**

En particular, este test puede aplicarse para comprobar si un parámetro es significativo. Sin embargo, es más habitual en estos casos comparar el parámetro estimado con su desviación típica. Los cocientes

$$\omega_j = \frac{\beta_j}{s(\beta_j)}$$

se denominan estadísticos de Wald y en muestras grandes se distribuyen, si el verdadero valor del parámetro es cero, como una normal estándar.

Una medida global del ajuste es

$$R^2 = 1 - \frac{D(\hat{\beta})}{D_o} = 1 - \frac{L(\hat{\beta})}{L(\beta_o)}$$

donde el numerador es la desviación (verosimilitud en el máximo) para el modelo con parámetros estimados  $\hat{\beta}$  y el denominador la desviación (verosimilitud) para el modelo que solo incluye la constante  $\beta_o$ . Observemos que, en este último caso, la estimación de la probabilidad  $\pi_i$  es constante para

todos los datos e igual a  $\frac{m}{n}$  siendo  $m$  el número de elementos en la muestra

con la variable  $y = 1$ . Entonces, sustituyendo en (I.11) la desviación máxima, que corresponde al modelo mas simple posible con solo  $\beta_0$  que asigna la misma probabilidad a todos los datos, es

$$D_o = -2L(\beta_0) = -2m \log m - 2(n - m) \log(n - m) + 2n \log n$$

Por otra lado, si el ajuste es perfecto, es decir, todas las observaciones con  $y = 1$  tienen  $p_i = 1$  y las de  $y = 0$  tienen  $p_i = 0$ , entonces, según (I.9) la desviación es cero y  $L(\hat{\beta}) = 0$  y  $R^2 = 1$ . Por el contrario, si las variables explicativas no influyen nada la desviación con las variables explicativas será igual que sin ellas,  $L(\hat{\beta}) = L(\beta_0)$  y  $R^2 = 0$ .

## CURVA ROC

Curva ROC (Receiver Operating characteristic curve – Curva de características operativas del receptor), su origen se dio en la estimación de errores en la transmisión de mensajes, actualmente es muy usada en estadística aplicada a la medicina. Es llamada también diagrama de Lorentz. En el eje de las abscisas se ubica la proporción de malos con score  $s$  y en el eje de las ordenadas la proporción de buenos con score  $s$ .

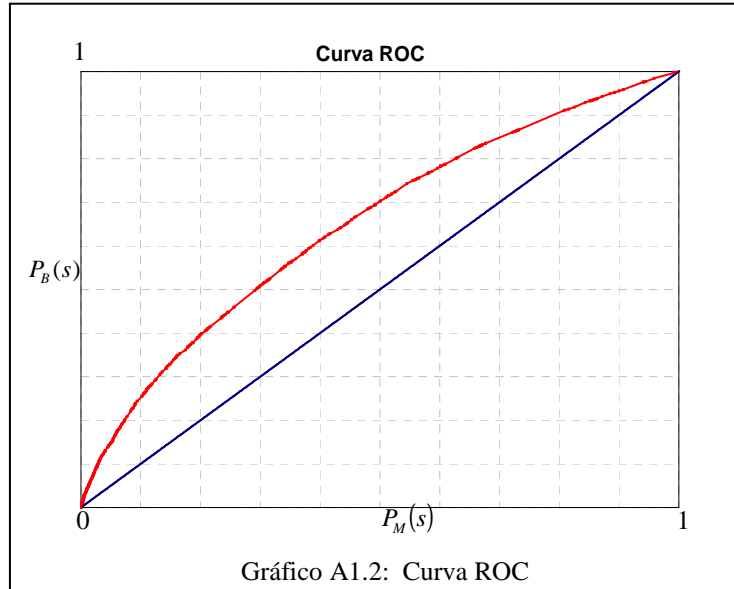
Otras definiciones relacionadas con la curva ROC, son la sensibilidad y la especificidad:

- Sensibilidad (Se): si los buenos son los casos que se desean detectar, es la proporción de buenos predichos como buenos.
- Especificidad (Sp): es la proporción de malos predichos como malos.

Para el modelo en estudio, donde los casos sobre el punto de corte (cutoff) son clasificados como malos, se tiene:  $Sp = 1 - P_M(s)$  y  $Se = P_B(s)$ , así la curva ROC sería un gráfico  $1 - Sp$  vs  $Se$ .



La curva ROC describe la propiedad de clasificación de la scorecard, cuando el punto de corte varía.



## INDICADOR DE INTERCORRELACIÓN KMO

Es una medida de la intercorrelación entre variables, se usa comúnmente para medir la adecuación del modelo factorial, es decir para verificar el supuesto de alta correlación de las variables que ingresan al análisis. El KMO se calcula por:

$$KMO = \frac{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} \sum_{i \neq j} a_{ij}^2}$$

Donde:

$r_{ij}$ , es el coeficiente de correlación de Pearson entre las variables  $i$  y  $j$ .

$a_{ij}$ , es el coeficiente de correlación parcial entre las variables  $i$  y  $j$

El índice KMO compara las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial. Si la suma de los coeficientes de correlación parcial al cuadrado es muy pequeña, el KMO será un índice muy próximo a la unidad y por tanto la correlación es

adecuada. En cambio valores pequeños en este índice nos indica la no conveniencia de aplicar el análisis factorial.

Una clasificación comúnmente aceptada para la interpretación del KMO es: (Kaiser 1974):

$1 \geq KMO > 0.90$  Excelente

$0.90 \geq KMO > 0.80$  Bueno

$0.80 \geq KMO > 0.70$  Aceptables

$0.70 \geq KMO > 0.60$  Mediocres o regulares

$0.60 \geq KMO > 0.50$  Malos

$KMO \leq 0.50$  Inacceptables o muy malos.

## **INDICE DE ESTABILIDAD DE LA POBLACIÓN PSI**

PSI, Population Stability Index, el índice de estabilidad de la población, evalúa el cambio en la población en el transcurso del tiempo. Teóricamente este índice se basa en la entropía relativa utilizada en teoría de la información, denominada información de Kullback simétrica. El PSI mide la separación (divergencia) entre las distribuciones.

Estadísticamente el PSI se calcula por la fórmula siguiente:

$$PSI = \sum_{i=1}^n (O_i - E_i) * \ln\left(\frac{O_i}{E_i}\right)$$

Donde  $O_i$  es la frecuencia observada para la característica  $i$  y  $E_i$  es la frecuencia esperada para la característica  $i$ .

- **Entropía Relativa**

En teoría de la información la entropía relativa o distancia de Kullback Leibler entre dos distribuciones de probabilidad (representadas por sus histogramas  $H_i$  y  $H_j$ ) cuantifica la ineficiencia de asumir que la distribución verdadera es la representada por  $H_i$  cuando en realidad lo es la representada por el histograma  $H_j$ .

$$KL_{ij} = \sum_k H_i(k) \log \left( \frac{H_i(k)}{H_j(k)} \right)$$

Donde:  $0 * \log \left( \frac{0}{H_j(k)} \right) = 0$  y  $H_i(k) \log \left( \frac{H_i(k)}{0} \right) = \infty$

Alternativamente  $KL_{ij} = \sum_k H_i(k) \{ \log(H_i(k)) - \log(H_j(k)) \}$

**Asimetría:** Esta medida no es una distancia en el sentido matemático del término ya que no es simétrica. Por ello cuando se necesita esta propiedad se suele usar  $KL_{ij} + KL_{ji}$ .

**Teorema:**  $KL_{ij} \geq 0$ . La igualdad se cumple si  $H_i(k) = H_j(k)$ .

## CADENAS DE MARKOV

Supongamos que  $M = \{X_n\}_{n=0}^{\infty}$  (o  $M = \{X_n\}_{n=-\infty}^{\infty}$ ) es una sucesión de variables aleatorias correlacionadas, donde cada (estado)  $X_n$  proviene de algún conjunto  $\Omega$  conocido como el espacio de estados. Se supone que los estados en  $\Omega$  pueden ser etiquetados mediante enteros, es decir,  $\Omega$  es discreto.

El proceso es una cadena de Markov si éste satisface la condición de Markov

$$\Pr(X_{n+1} = j | X_n = i, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \Pr(X_{n+1} = j | X_n = i)$$

Una distribución inicial  $\Pr(X_0 = i)$  para  $X_0$  y la probabilidad de transición para  $X_{n+1}$  dado  $X_n$ ,  $\Pr(X_{n+1} = j | X_n = i)$  determinan una cadena de Markov.

Si la probabilidad de transición no depende de  $n$ , es decir, si

$$\Pr(X_{n+m+1} = j | X_{n+m} = i) = \Pr(X_{n+1} = j | X_n = i); \quad \forall m \in \mathbb{Z}$$

decimos que la cadena de Markov  $M$  es homogénea y escribimos la probabilidad de transición como una matriz  $\underline{P}$  donde sus elementos están dados por

$$P_{ij} = \Pr(X_{n+1} = j | X_n = i)$$

Note que  $P_{ij}$  es la probabilidad condicional para pasar al estado  $j$  en el próximo paso dado que el estado actual sea  $i$ . Las probabilidades de transición satisfacen la condición de normalización

$$\sum_{j \in \Omega} P_{ij} = 1$$

ya que la cadena debe estar en algún estado en el próximo paso.

Una matriz de esta naturaleza, cuyas filas suman uno es conocida como matriz estocástica.

Ejemplo : Supongamos que  $\Omega = \{1, 2, 3\}$ , la matriz de transición es

$$\underline{P} = \begin{pmatrix} \frac{2}{5} & \frac{1}{2} & \frac{1}{10} \\ \frac{1}{5} & \frac{7}{10} & \frac{1}{10} \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \end{pmatrix}$$

y la distribución inicial es  $\Pr(X_0 = i) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ .

Se puede representar a la matriz de transición como un grafo, donde un vértice corresponde a un estado y un enlace orientado del vértice  $i$  al vértice  $j$  le corresponde una probabilidad de transición  $P_{ij}$  de  $i$  a  $j$  cuando ésta sea diferente de cero. Para el ejemplo tenemos su correspondiente representación en un grafo orientado en la Figura 1.



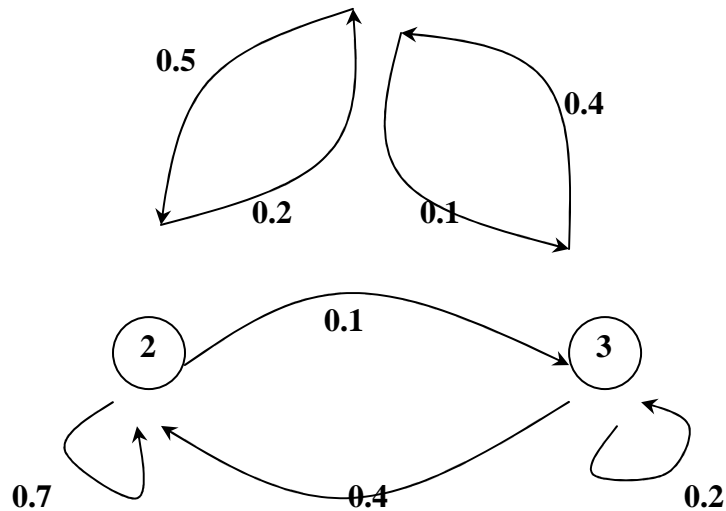


Figura1: Grafo orientado correspondiente al Ejemplo 1.

La distribución de  $X_n$

Considere  $\Pr(X_n = j)$ , la cual es la probabilidad que después de una simulación de una cadena de Markov para  $n$  pasos, el estado alcanzado sea  $X_n = j$ .

Estas probabilidades pueden ser arregladas en un vector fila  $\pi^{(n)}$  donde, por definición  $\pi_j^{(n)} = \Pr(X_n = j)$ . Cuando  $n=1$ , vemos que

$$\Pr(X_1 = j) = \sum_{i \in \Omega} \Pr(X_1 = j, X_0 = i) = \sum_{i \in \Omega} \Pr(X_1 = j | X_0 = i) \Pr(X_0 = i)$$

Esto puede ser escrito en forma matricial como

$$\pi_j^{(1)} = \sum_{i \in \Omega} \pi_i^{(0)} P_{ij} \quad \text{o} \quad \underline{\pi}^{(1)} = \underline{\pi}^{(0)} \underline{P} [0,1]$$

De manera similar

$$\underline{\pi}^{(n)} = \underline{\pi}^{(n-1)} \underline{P}$$

Supongamos que se cumple que

$$\underline{\pi} = \underline{\pi} \underline{P}$$

es decir,  $\underline{\pi}$  es un vector propio izquierdo normalizado de la matriz  $\underline{P}$ , ya que  $\sum_i \pi_i = 1$  y con valor propio 1. Entonces  $\underline{\pi}$  es una distribución estacionaria para  $\underline{P}$

ya que si  $\underline{\pi}^{(n)} = \underline{\pi}$  entonces  $\underline{\pi}^{(n+1)} = \underline{\pi}^{(n)} \underline{P} = \underline{\pi}$  también, es decir, una vez que la cadena está en la distribución  $\underline{\pi}$  ésta se queda en esta distribución.

Definición: Supongamos que  $\underline{\pi}^{(n)} \rightarrow \underline{\pi}$  cuando  $n \rightarrow \infty$  para cualquier  $\underline{\pi}^{(0)}$ .

Entonces  $\underline{\pi}$  es la distribución de equilibrio de la cadena  $M$  y la cadena se dice que es ergódica.

Para una cadena ergódica y para  $n$  suficientemente grande, los estados de  $M$  están distribuidos como  $\underline{\pi}$  y el sistema está “en equilibrio”.

Una cadena de Markov puede ser especificada por una matriz de transición, o dando la dinámica microscópica (es decir, un algoritmo que determina a  $X_{n+1}$  dado  $X_n$ ). El algoritmo de manera implícita fija la matriz de transición. Los problemas reales son usualmente demasiado complicados para poder dar la matriz de transición de manera explícita. Así que el problema de especificar una CM con una distribución de equilibrio deseada será sustituido por el problema de suministrar un algoritmo que se pueda probar que genera una CM con la distribución de equilibrio correcta. Se usa la idea de reversibilidad.

Ejemplo 11: Construir una cadena de Markov  $M$  ergódica y reversible en el espacio de estados  $\Omega = \{1, 2, 3, 4\}$  con distribución de equilibrio  $\pi_i = \frac{1}{4}$ , es decir uniforme en  $\Omega$ .

Ya que  $M$  es reversible, la matriz transición debe satisfacer

$$\pi_i P_{ij} = \pi_j P_{ji},$$

Y como  $\pi_i = \pi_j = \frac{1}{4}$  tenemos que  $P_{ij} = P_{ji}$ . Si  $\Omega$  es irreducible bajo  $\underline{P}$  (es decir, la cadena puede tomar cualquier estado en  $\Omega$ ) entonces se han satisfecho las condiciones del teorema de ergodicidad para cadenas reversibles. Así que cualquier matriz de transición  $\underline{P}$  simétrica e irreducible serviría.

Si por ejemplo,

$$\underline{P} = \begin{pmatrix} 3/4 & 1/4 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1/4 & 3/4 \end{pmatrix}$$

se satisface que  $\sum_j P_{ij} = 1$  y  $P_{ij} = P_{ji}$ . Ahora,  $\underline{\pi} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$  es un vector propio izquierdo de  $\underline{P}$ . Todas las condiciones de ergodicidad están satisfechas de manera que se espera que  $\underline{\pi}^{(n)} = \underline{\pi}^{(0)} \underline{P}^n$  tiende a  $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$  cuando  $n \rightarrow \infty$  desde cualquier estado inicial. De manera explícita, se puede hallar que

$$\underline{P}^2 = \begin{pmatrix} .625 & .3125 & .625 & 0 \\ .3125 & .375 & .25 & .625 \\ .625 & .25 & .375 & .3125 \\ 0 & .625 & .3125 & .625 \end{pmatrix}$$

$$\underline{P}^4 = \begin{pmatrix} .49219 & .32813 & .14063 & .03906 \\ .32813 & .30469 & .22656 & .14063 \\ .14063 & .22656 & .30469 & .32813 \\ .03906 & .14063 & .32813 & .49219 \end{pmatrix}$$

$$\underline{P}^{100} = \begin{pmatrix} .25000\ 00567 & .25000\ 00235 & .24999\ 99765 & .24999\ 99433 \\ .25000\ 00235 & .25000\ 00097 & .24999\ 99903 & .24999\ 99765 \\ .24999\ 99765 & .24999\ 99903 & .25000\ 00097 & .25000\ 00235 \\ .24999\ 99433 & .24999\ 99765 & .25000\ 00235 & .25000\ 00567 \end{pmatrix} \quad (0.1)$$

Note que este problema es trivial ya que se puede simplemente tomar  $P_{ij} = 1/4$  para todo  $i, j$  en  $\Omega$ . En cada actualización se muestrea uniformemente en  $\Omega$ . Entonces  $\underline{\pi}^{(n)} = \underline{\pi}$  para todo  $n$ .



## ANEXO II

### TABLAS Y GRÁFICOS

#### ÁRBOLES DE DECISIÓN

Node 0	Bueno	67,16	20879
	Malo	32,84	10208
scg_desciudad (Adj. P-value=0,0000, Chi-square=622,9995, df=2)			
<div style="border-left: 1px dashed black; padding-left: 10px;"> <p style="margin: 0;">QUITO;AMBATO;IBARRA;L ATACUNGA;MANTA;RIOB AMBA;PORTOVIEJO</p> </div>	<b>Node 5</b>	Bueno	75,27
		Malo	24,73
			9361
			3075
<div style="border-left: 1px dashed black; padding-left: 10px;"> <p style="margin: 0;">GUAYAQUIL;CUENCA</p> </div>	<b>Node 6</b>	Bueno	61,64
		Malo	38,36
			11216
			6981
<div style="border-left: 1px dashed black; padding-left: 10px;"> <p style="margin: 0;">LOJA</p> </div>	<b>Node 7</b>	Bueno	66,52
		Malo	33,48
			302
			152

Gráfico A2.1: Árbol de Decisión Variable Ciudad

Node 0	Bueno	67,16	20879
	Malo	32,84	10208
scg_indeudatitular (Adj. P-value=0,0000, Chi-square=45,8394, df=1)			
<div style="border-left: 1px dashed black; padding-left: 10px;"> <p style="margin: 0;">0</p> </div>	<b>Node 19</b>	Bueno	66,32
		Malo	33,68
			16945
			8604
<div style="border-left: 1px dashed black; padding-left: 10px;"> <p style="margin: 0;">1</p> </div>	<b>Node 20</b>	Bueno	71,04
		Malo	28,96
			3934
			1604
scg_indeudaconsumo6M (Adj. P-value=0,0000, Chi-square=103,6950, df=1)			
<div style="border-left: 1px dashed black; padding-left: 10px;"> <p style="margin: 0;">0</p> </div>	<b>Node 25</b>	Bueno	64,16
		Malo	35,84
			1595
			891
<div style="border-left: 1px dashed black; padding-left: 10px;"> <p style="margin: 0;">1</p> </div>	<b>Node 26</b>	Bueno	76,64
		Malo	23,36
			2339
			713

Gráfico A2.2: Árbol de Decisión Variable Deuda Titular, Deuda Consumo

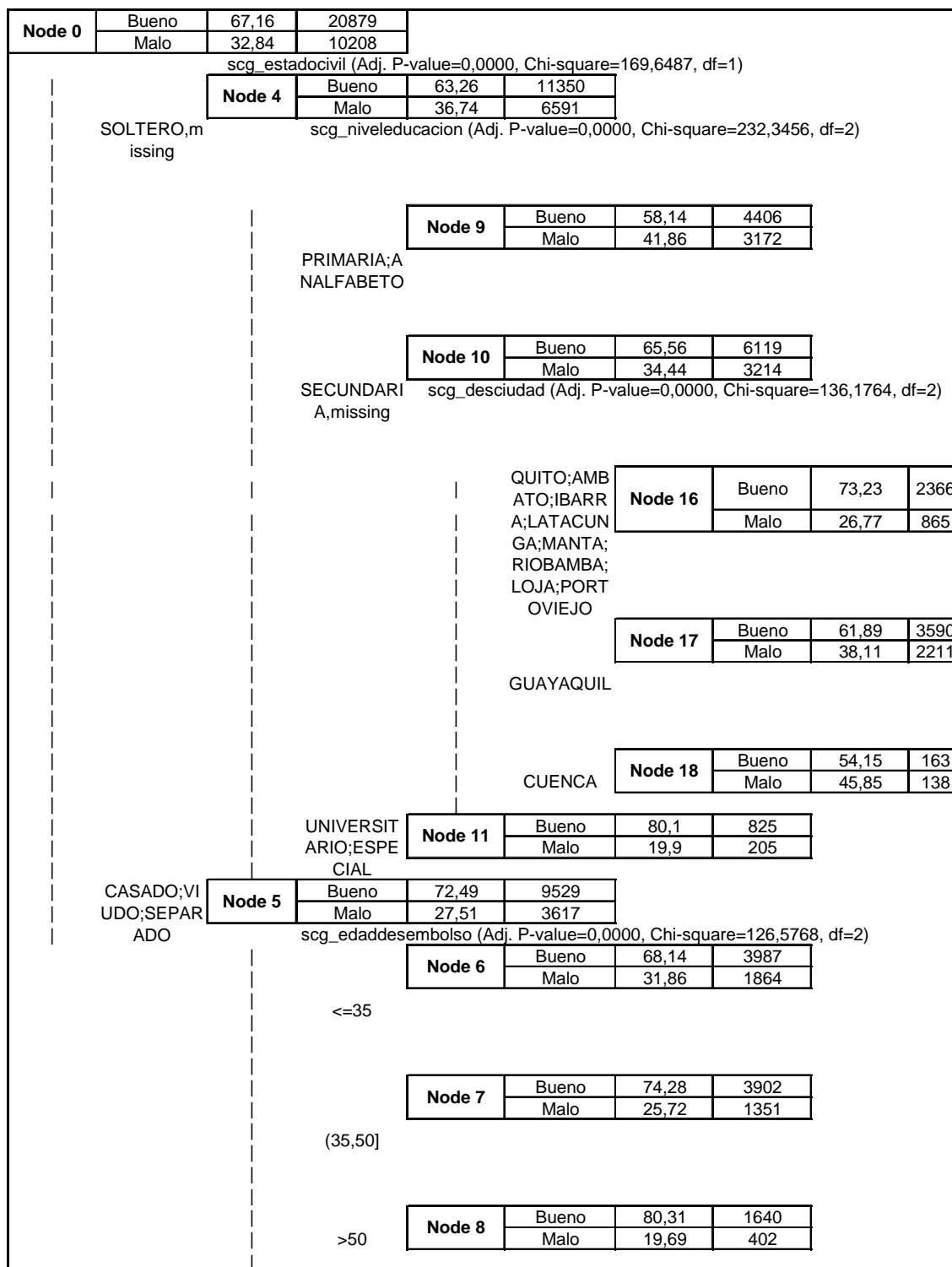


Gráfico A2.3: Árbol de Decisión Cruce Variables: Estado Civil, Edad, Nivel Educación

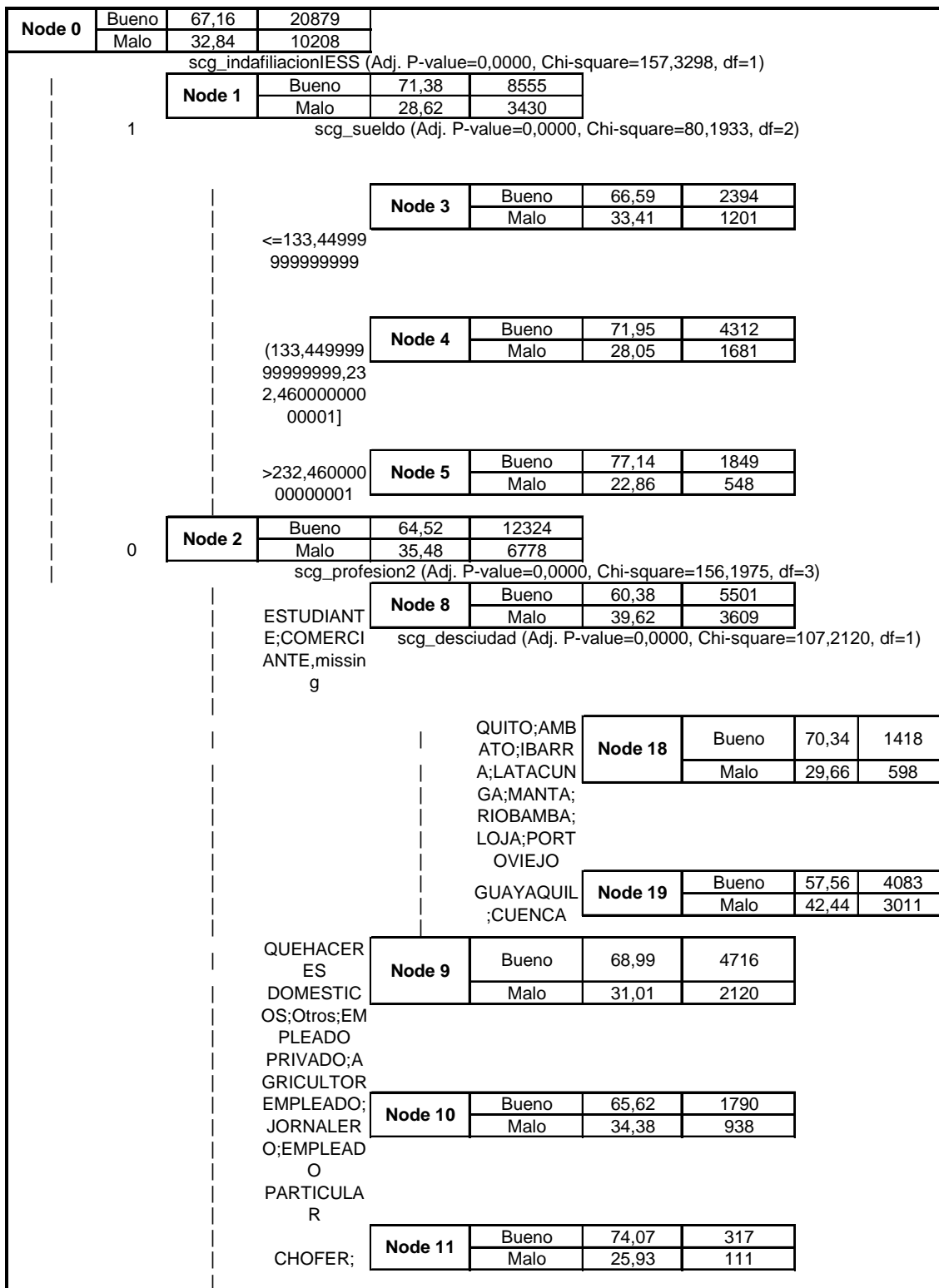


Gráfico A2.4: Árbol de Decisión Cruce Variables: IESS, Sueldo, Profesión, Ciudad

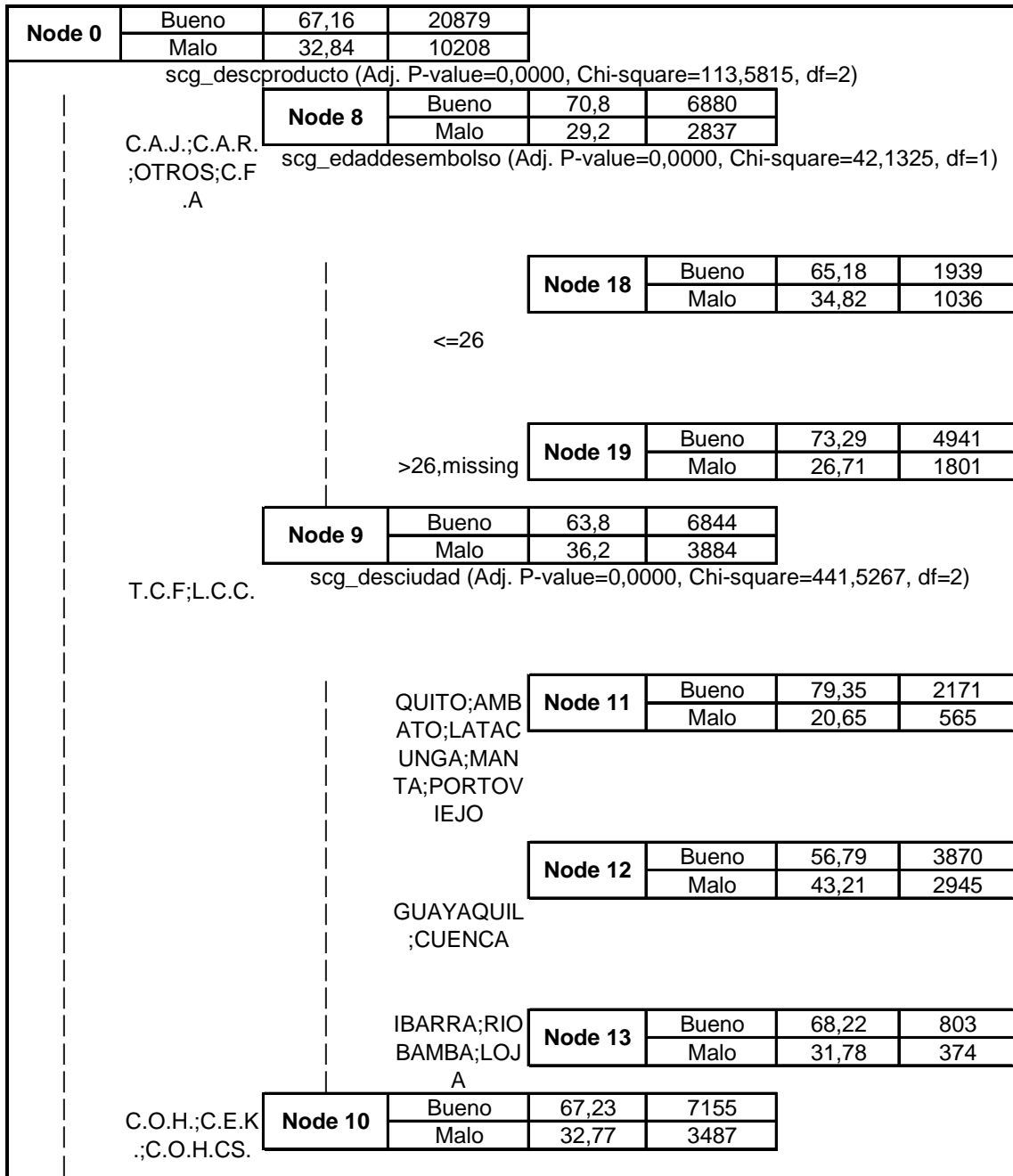


Gráfico A2.5: Árbol de Decisión Cruce Variables: Producto, Ciudad, Edad

Node 0	Bueno	67,16	20879	
	Malo	32,84	10208	
scg_desciudad (Adj. P-value=0,0000, Chi-square=630,7668, df=3)				
QUITO;AMBATO;IBARRA;LATACUNGA;MANTA;RIOBAMBA;PORTOVIEJO	Node 1	Bueno	75,27	9361
		Malo	24,73	3075
GUAYAQUIL	Node 2	Bueno	61,9	10465
		Malo	38,1	6440
CUENCA	Node 3	Bueno	58,13	751
		Malo	41,87	541
LOJA	Node 4	Bueno	66,52	302
		Malo	33,48	152

Gráfico A2.7: Árbol de Decisión Variable Ciudad

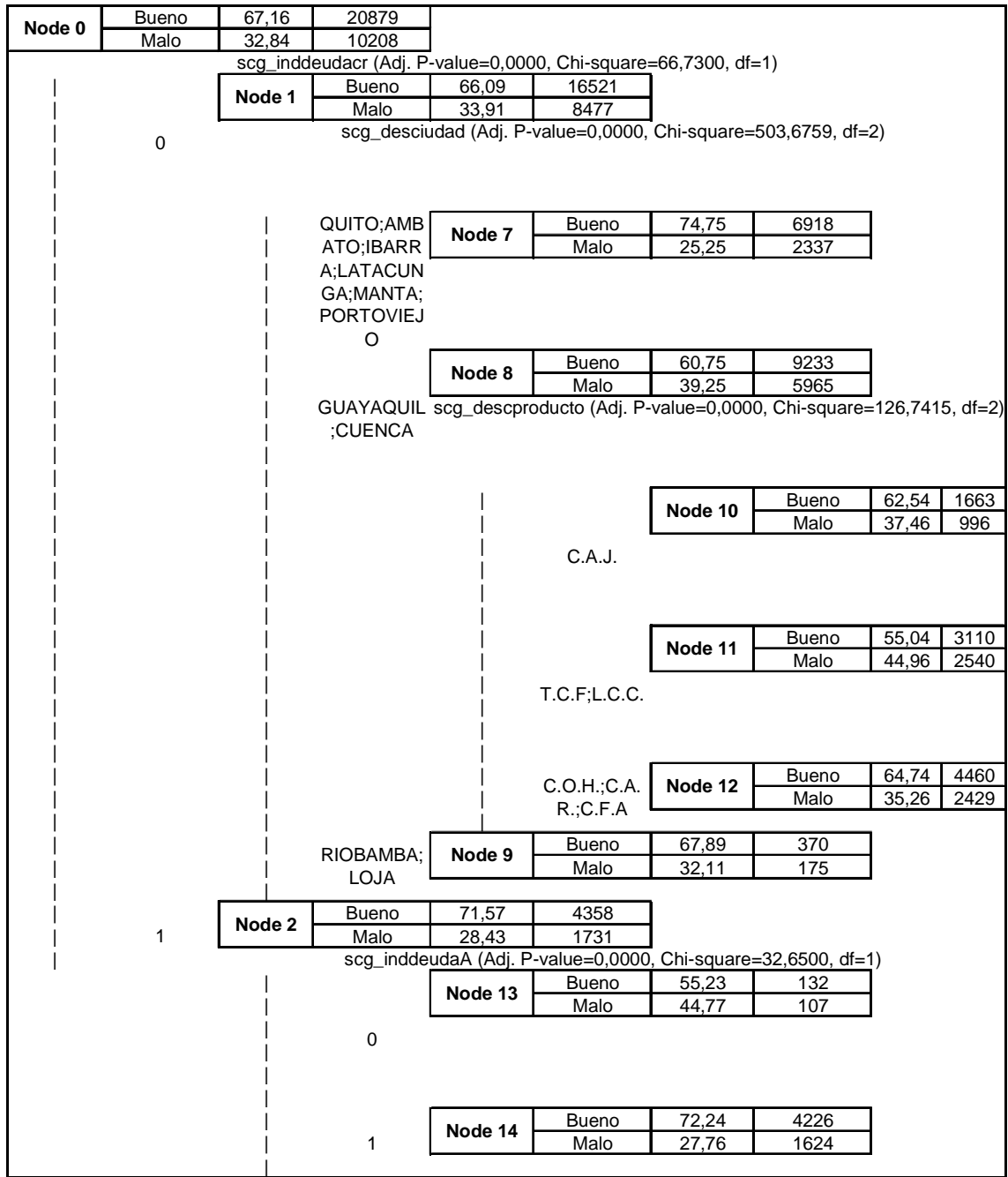


Gráfico A2.8: Árbol de Decisión Cruce de Variables: Deuda Central de Riesgos, Ciudad, Producto

Node 0	Bueno	67,16	20879	
	Malo		32,84	10208
scg_edadesembolso (Adj. P-value=0,0000, Chi-square=303,2204, df=4)				
<=26	Node 1	Bueno	62,39	6042
	Malo		37,61	3643
(26,35]	Node 2	Bueno	65,66	5976
	Malo		34,34	3126
(35,43]	Node 3	Bueno	69,36	4027
	Malo		30,64	1779
(43,50]	Node 4	Bueno	71,66	2440
	Malo		28,34	965
>50,missing	Node 5	Bueno	77,5	2394
	Malo		22,5	695

Gráfico A2.9: Árbol de Decisión Variable Edad

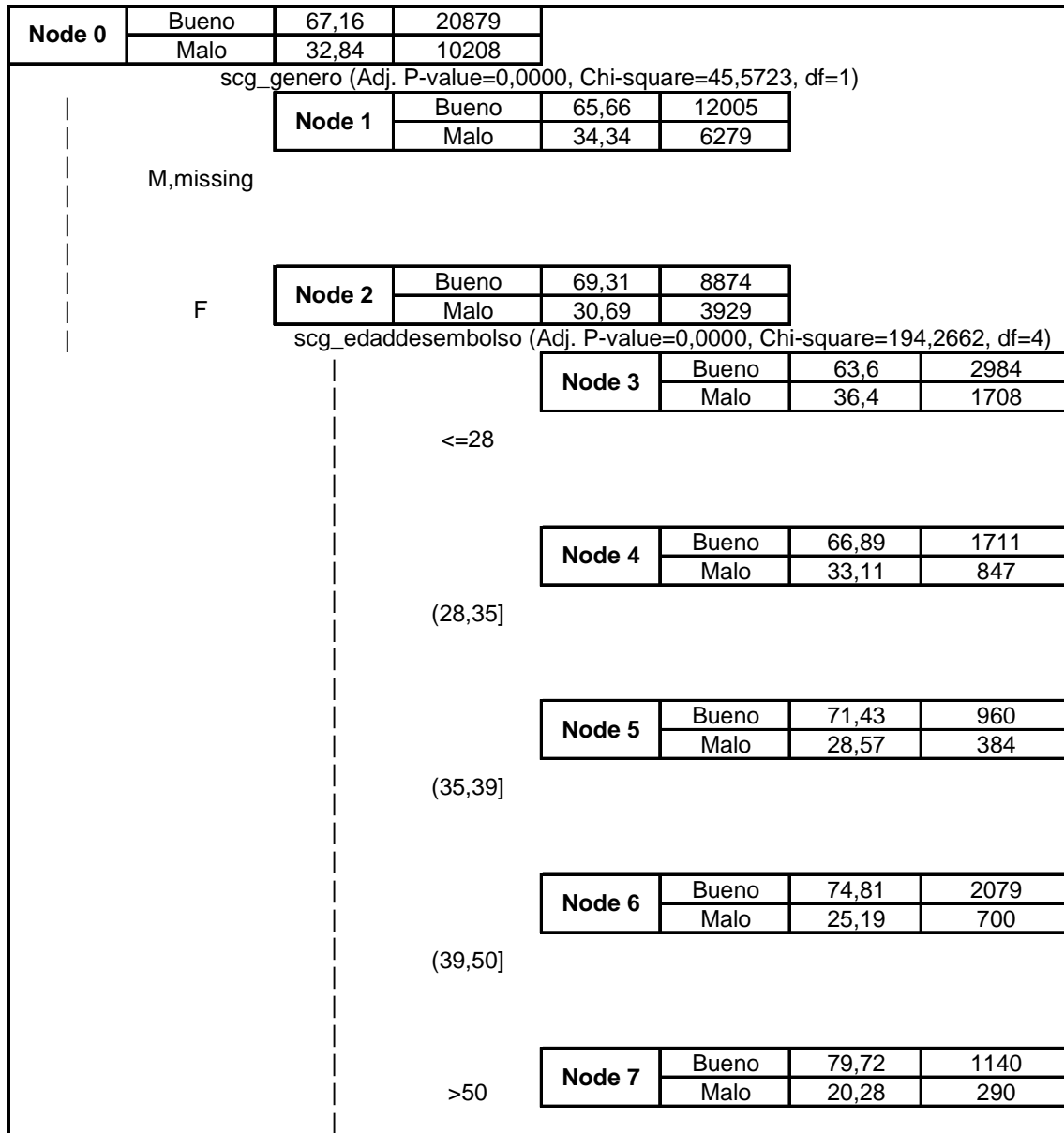


Gráfico A2.10: Árbol de Decisión Cruce de Variables: Género, Edad



<b>Node 0</b>	Bueno	67,16	20879	
	Malo	32,84	10208	
scg_indruc (Adj. P-value=0,0004, Chi-square=12,5428, df=1)				
0	<b>Node 27</b>	Bueno	66,68	16405
		Malo	33,32	8198
1	<b>Node 28</b>	Bueno	69	4474
		Malo	31	2010
scg_antiguedadRUC (Adj. P-value=0,0000, Chi-square=46,0180, df=2)				
	<b>Node 31</b>	Bueno	65,08	2000
		Malo	34,92	1073
		<=2		
	<b>Node 32</b>	Bueno	71,38	1606
		Malo	28,62	644
		(2,6]		
	<b>Node 33</b>	Bueno	74,76	868
		Malo	25,24	293
		>6		

Gráfico A2.11: Árbol de Decisión Cruce de Variables: RUC, Antigüedad, RUC

<b>Node 0</b>	Bueno	67,16	20879	
	Malo	32,84	10208	
scg_indcarros (Adj. P-value=0,0000, Chi-square=188,3627, df=1)				
0	<b>Node 20</b>	Bueno	66,19	19201
		Malo	33,81	9810
1	<b>Node 21</b>	Bueno	80,83	1678
		Malo	19,17	398

Gráfico A2.12: Árbol de Decisión Variable Vehículo

## ANÁLISIS DESCRIPTIVO DE VARIABLES EXPLICATIVAS

descproducto					
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado	
Válidos	C.A.J.	7.762	25,0	25,0	25,0
	C.A.R.	1.454	4,7	4,7	29,6
	C.E.K.	1	0,0	0,0	29,6
	C.F.A.	296	1,0	1,0	30,6
	C.O.H.	10.638	34,2	34,2	64,8
	C.O.H.CS.	3	0,0	0,0	64,8
	L.C.C.	431	1,4	1,4	66,2
	OTROS	205	0,7	0,7	66,9
	T.C.F	10.297	33,1	33,1	100,0
	Total	31.087	100,0	100,0	

Tabla A2.1: Análisis de Frecuencias Variable Producto

desciudad					
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado	
Válidos	AMBATO	1.090	3,5	3,5	3,5
	CUENCA	1.292	4,2	4,2	7,7
	GUAYAQUIL	16.905	54,4	54,4	62,0
	IBARRA	1.320	4,2	4,2	66,3
	LATACUNGA	183	0,6	0,6	66,9
	LOJA	454	1,5	1,5	68,3
	MANTA	1.621	5,2	5,2	73,6
	PORTOVIEJO	108	0,3	0,3	73,9
	QUITO	7.736	24,9	24,9	98,8
	RIOBAMBA	378	1,2	1,2	100,0
	Total	31.087	100,0	100,0	

Tabla A2.2: Análisis de Frecuencias Variable Ciudad

indcarros					
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado	
Válidos	0,00	29.011	93,3	93,3	93,3
	1,00	2.076	6,7	6,7	100,0
	Total	31.087	100,0	100,0	

Tabla A2.3: Análisis de Frecuencias Variable Carros

<b>totalcarros</b>				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0,00	29.011	93,3	93,3
	1,00	1.538	4,9	98,3
	2,00	390	1,3	99,5
	3,00	101	0,3	99,8
	4,00	31	0,1	99,9
	5,00	9	0,0	100,0
	6,00	5	0,0	100,0
	7,00	1	0,0	100,0
	8,00	1	0,0	100,0
	Total	31.087	100,0	100,0

Tabla A2.4: Análisis de Frecuencias Variable Número Carros

<b>estadocivil</b>				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	CASADO	12.020	38,7	39,1
	SEPARADO	285	0,9	40,0
	SOLTERO	17.584	56,6	97,3
	VIUDO	841	2,7	100,0
	Total	30.730	98,9	100,0
Perdidos	1	357	1,1	
	Total	31.087	100,0	

Tabla A2.5: Análisis de Frecuencias Variable Estado Civil

<b>niveleducacion</b>				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	ANALFABETO	298	1,0	1,0
	ESPECIAL	118	0,4	1,4
	PRIMARIA	12.677	40,8	42,6
	SECUNDARIA	15.430	49,6	92,8
	UNIVERSITARIO	2.207	7,1	100,0
	Total	30.730	98,9	100,0
Perdidos	1	357	1,1	
	Total	31.087	100,0	

Tabla A2.6: Análisis de Frecuencias Variable Nivel Educación

<b>genero</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	F	12.803	41,2	41,7	41,7
	M	17.927	57,7	58,3	100,0
	Total	30.730	98,9	100,0	
Perdidos	1	357	1,1		
Total		31.087	100,0		

Tabla A2.7: Análisis de Frecuencias Variable Género

<b>indafiliacion IESS</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0,00	19.102	61,4	61,4	61,4
	1,00	11.985	38,6	38,6	100,0
	Total	31.087	100,0	100,0	

Tabla A2.8: Análisis de Frecuencias Variable Indicador Afiliación IESS

<b>indruc</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0,00	24.603	79,1	79,1	79,1
	1,00	6.484	20,9	20,9	100,0
	Total	31.087	100,0	100,0	

Tabla A2.9: Análisis de Frecuencias Variable Indicador RUC

<b>vehículo año</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	31.087	100,0	100,0	100,0

Tabla A2.10: Análisis de Frecuencias Variable Vehículo Año

<b>numero propiedades</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	31.087	100,0	100,0	100,0

Tabla A2.11: Análisis de Frecuencias Variable Número Propiedades

<b>indcelulares</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	24.291	78,1	78,1	78,1
Válidos	1,00	6.796	21,9	21,9	100,0
	Total	31.087	100,0	100,0	

Tabla A2.12: Análisis de Frecuencias Variable Indicador Celulares

<b>indeudatitular</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	25.549	82,2	82,2	82,2
Válidos	1,00	5.538	17,8	17,8	100,0
	Total	31.087	100,0	100,0	

Tabla A2.13: Análisis de Frecuencias Variable Indicador Deuda Titular

<b>indeudasolidario</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	29.904	96,2	96,2	96,2
Válidos	1,00	1.183	3,8	3,8	100,0
	Total	31.087	100,0	100,0	

Tabla A2.14: Análisis de Frecuencias Variable Indicador Deuda Solidario

<b>indeudacodeudor</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	30.432	97,9	97,9	97,9
Válidos	1,00	655	2,1	2,1	100,0
	Total	31.087	100,0	100,0	

Tabla A2.15: Análisis de Frecuencias Variable Indicador Deuda Codeudor

<b>indmedidor</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	30.969	99,6	99,6	99,6
Válidos	1,00	118	0,4	0,4	100,0
	Total	31.087	100,0	100,0	

Tabla A2.16: Análisis de Frecuencias Variable Indicador Medidor

<b>tipoplancelular</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	PLAN	189	0,6	3,9	3,9
	PREPAGO	4.719	15,2	96,1	100,0
	Total	4.908	15,8	100,0	
Perdidos	1	26.179	84,2		
Total		31.087	100,0		

Tabla A2.17: Análisis de Frecuencias Variable Indicador Tipo Plan Celular

<b>indfonovivienda</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0,00	27.367	88,0	88,0	88,0
	1,00	3.720	12,0	12,0	100,0
	Total	31.087	100,0	100,0	

Tabla A2.18: Análisis de Frecuencias Variable Indicador Teléfono Vivienda

<b>categoriaempresa</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	GRANDE	21	0,1	0,5	0,5
	MEDIANA	2.441	7,9	58,4	58,9
	PEQUEÑA	1.114	3,6	26,7	85,6
	PUBLICA	602	1,9	14,4	100,0
	Total	4.178	13,4	100,0	
Perdidos	1	26.909	86,6		
Total		31.087	100,0		

Tabla A2.19: Análisis de Frecuencias Variable Categoría Empresa

<b>actividadempresa</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	COMERCIAL	994	3,2	22,5	22,5
	FINANCIERA	55	0,2	1,2	23,7
	INDUSTRIAL	448	1,4	10,1	33,8
	SERVICIOS	2.172	7,0	49,1	82,9
	SIN INFORMACION	758	2,4	17,1	100,0
	Total	4.427	14,2	100,0	
Perdidos	1	26.660	85,8		
Total		31.087	100,0		

Tabla A2.20: Análisis de Frecuencias Variable Actividad Empresa

subactividadempresa					
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado	
Válidos	ACTIVIDAD AGRICOLAS Y AGROPECUARIAS	79	0,3	14,0	14,0
	ARTICULOS COMPUTACIONALES	21	0,1	3,7	17,7
	ARTICULOS PLASTICOS	13	0,0	2,3	20,0
	BOUTIQUE	18	0,1	3,2	23,2
	ELECTRODOMESTICOS	20	0,1	3,5	26,8
	FARMACIAS	35	0,1	6,2	33,0
	FERRETERIAS	23	0,1	4,1	37,1
	GENERICOS	261	0,8	46,3	83,3
	HOTELES	10	0,0	1,8	85,1
	LICORERIAS	3	0,0	0,5	85,6
	MUEBLES	3	0,0	0,5	86,2
	PANADERIAS	10	0,0	1,8	87,9
	PAPELERIAS	5	0,0	0,9	88,8
	REPUESTOS AUTOMOTRICES	14	0,0	2,5	91,3
	RESTAURANTES	11	0,0	2,0	93,3
	TIENDAS Y MICROMERCADOS	38	0,1	6,7	100,0
	Total	564	1,8	100,0	
	Perdidos	1	30.523	98,2	
	Total	31.087	100,0		

Tabla A2.21: Análisis de Frecuencias Variable Subactividad Empresa

<b>scg_numentidades</b>				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	24.998	80,4	80,4
	1,00	4.826	15,5	95,9
	2,00	978	3,1	99,1
	3,00	214	0,7	99,8
Válidos	4,00	53	0,2	99,9
	5,00	14	0,0	100,0
	6,00	2	0,0	100,0
	7,00	2	0,0	100,0
	Total	31.087	100,0	100,0

Tabla A2.23: Análisis de Frecuencias Variable Número de Entidades

<b>indeudacr</b>				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	24.998	80,4	80,4
Válidos	1,00	6.089	19,6	100,0
	Total	31.087	100,0	100,0

Tabla A2.24: Análisis de Frecuencias Variable Indicador Deuda Central de Riesgos

<b>indeudaA</b>				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	25.237	81,2	81,2
Válidos	1,00	5.850	18,8	100,0
	Total	31.087	100,0	100,0

Tabla A2.25: Análisis de Frecuencias Variable Indicador Deuda Calificación A

<b>indeudaBC</b>				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	30.819	99,1	99,1
Válidos	1,00	268	0,9	100,0
	Total	31.087	100,0	100,0

Tabla A2.26: Análisis de Frecuencias Variable Indicador Deuda Calificación B y C



<b>indeudaDE</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	30.997	99,7	99,7	99,7
Válidos	1,00	90	0,3	0,3	100,0
	Total	31.087	100,0	100,0	

Tabla A2.27: Análisis de Frecuencias Variable Indicador Deuda Calificación D y E

<b>indeudacomercial</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	30.820	99,1	99,1	99,1
Válidos	1,00	267	0,9	0,9	100,0
	Total	31.087	100,0	100,0	

Tabla A2.28: Análisis de Frecuencias Variable Indicador Deuda Comercial

<b>scg_indeudaconsumo</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	26.925	86,6	86,6	86,6
Válidos	1,00	4.162	13,4	13,4	100,0
	Total	31.087	100,0	100,0	

Tabla A2.29: Análisis de Frecuencias Variable Indicador Deuda Consumo

<b>indeudavivienda</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	30.943	99,5	99,5	99,5
Válidos	1,00	144	0,5	0,5	100,0
	Total	31.087	100,0	100,0	

Tabla A2.30: Análisis de Frecuencias Variable Indicador Deuda Vivienda

<b>indeudaBC6M</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	30.852	99,2	99,2	99,2
Válidos	1,00	235	0,8	0,8	100,0
	Total	31.087	100,0	100,0	

Tabla A2.31: Análisis de Frecuencias Variable Indicador Deuda Calificación B o C hace seis meses

<b>inddeudacr6M</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	25.517	82,1	82,1	82,1
Válidos	1,00	5.570	17,9	17,9	100,0
	Total	31.087	100,0	100,0	

Tabla A2.32: Análisis de Frecuencias Variable Indicador  
Deuda Central de Riesgos hace seis meses

<b>indeudaA6M</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	25.827	83,1	83,1	83,1
Válidos	1,00	5.260	16,9	16,9	100,0
	Total	31.087	100,0	100,0	

Tabla A2.33: Análisis de Frecuencias Variable Indicador  
Deuda Calificación A hace seis meses

<b>indeudaDE6M</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	30.913	99,4	99,4	99,4
Válidos	1,00	174	0,6	0,6	100,0
	Total	31.087	100,0	100,0	

Tabla A2.34: Análisis de Frecuencias Variable Indicador  
Deuda Calificación D y E hace seis meses

<b>indeudacomercial6M</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	30.778	99,0	99,0	99,0
Válidos	1,00	309	1,0	1,0	100,0
	Total	31.087	100,0	100,0	

Tabla A2.35: Análisis de Frecuencias Variable Indicador  
Deuda Comercial hace seis meses

<b>indeudaconsumo6M</b>					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
	0,00	27.339	87,9	87,9	87,9
Válidos	1,00	3.748	12,1	12,1	100,0
	Total	31.087	100,0	100,0	

Tabla A2.36: Análisis de Frecuencias Variable Indicador  
Deuda Consumo hace seis meses

indeudavivienda6M				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	0,00	30.929	99,5	99,5
	1,00	158	0,5	100,0
Total		31.087	100,0	

Tabla A2.37: Análisis de Frecuencias Variable Indicador Deuda Vivienda hace seis meses

Ind_BM				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Bueno	20.879	67,2	67,2
	Malo	10.208	32,8	100,0
Total		31.087	100,0	

Tabla A2.38: Análisis de Frecuencias Variable Indicador Bueno Malo

## FUNCIONES LINEALES DE PARA LOS MESES DE ANTIGÜEDAD DEL 2 AL 11

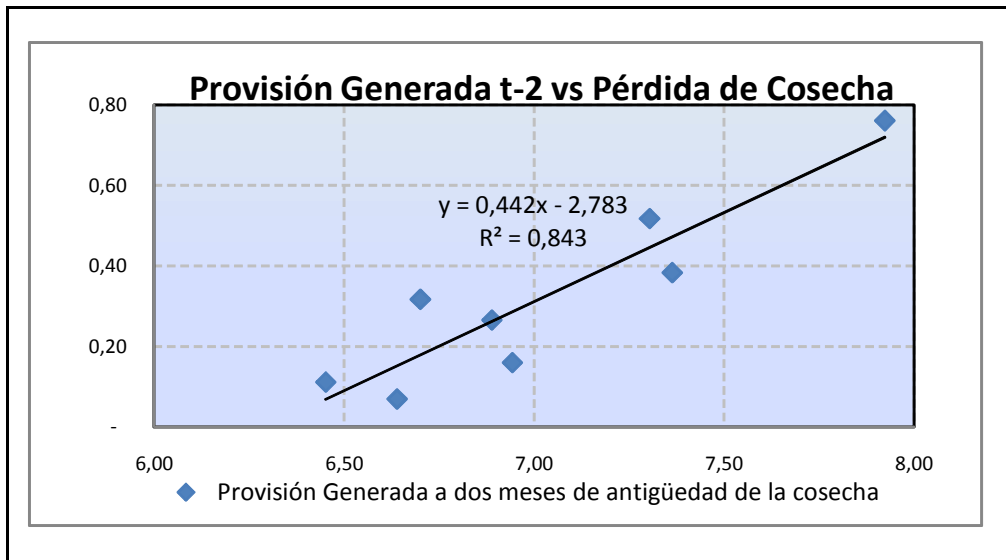


Gráfico A2.13: Regresión t-2

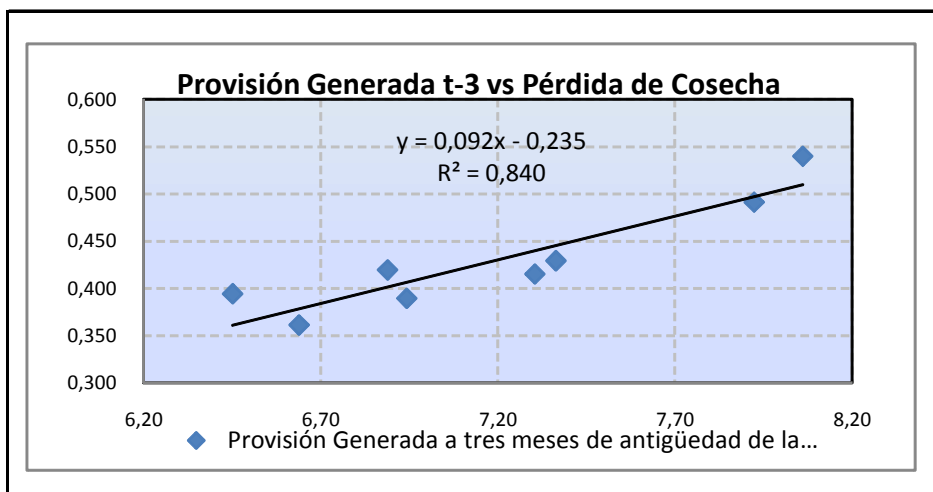


Gráfico A2.14: Regresión t-3

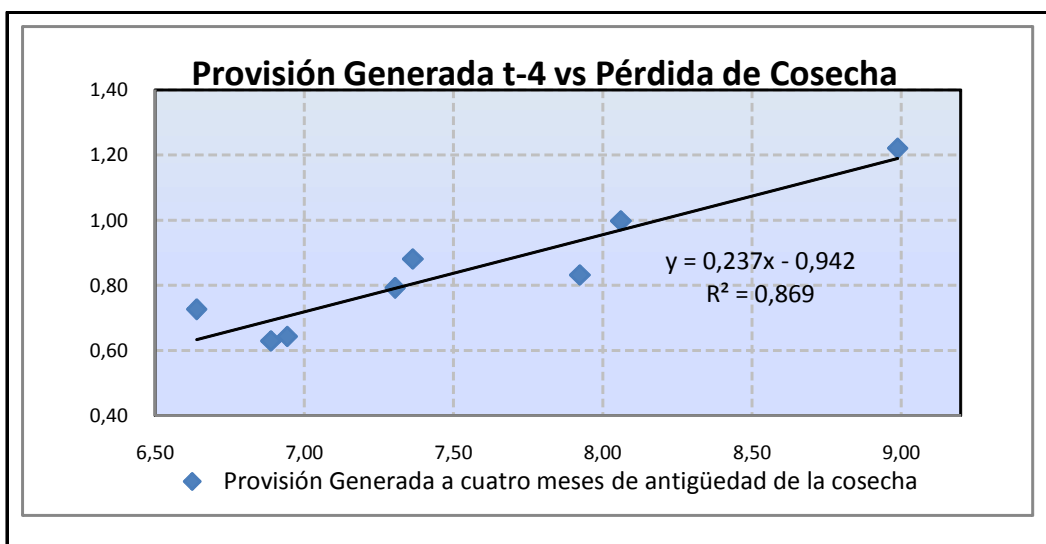


Gráfico A2.15: Regresión t-4

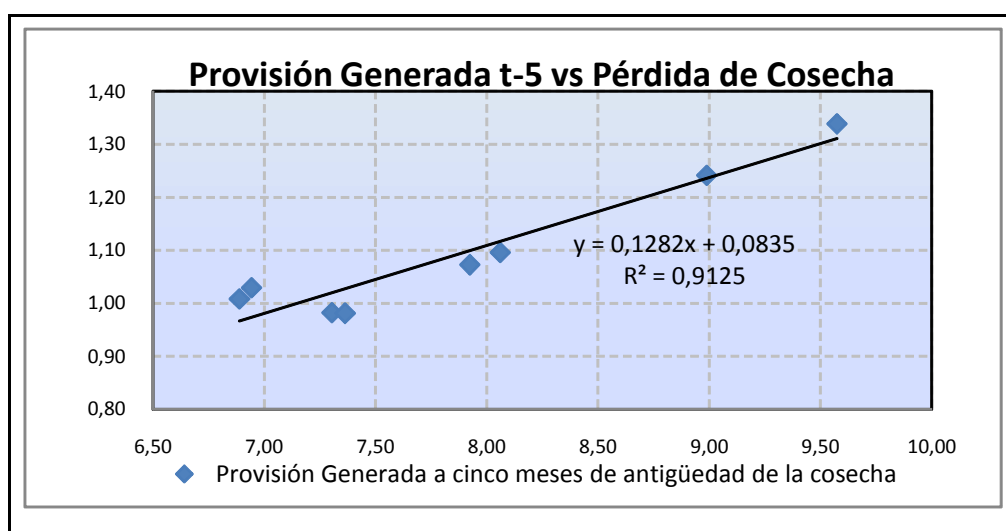


Gráfico A2.16 Regresión t-5

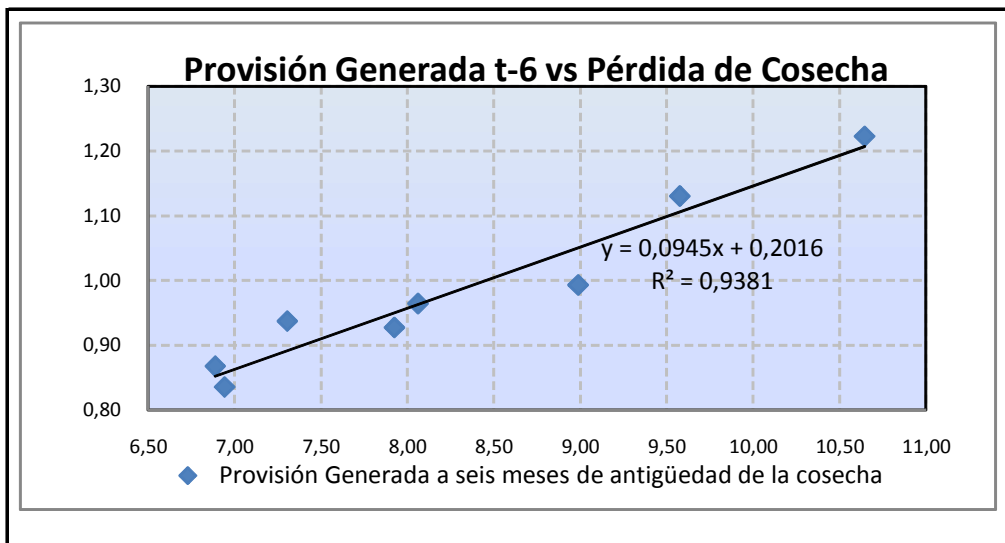


Gráfico A2.17 Regresión t-6

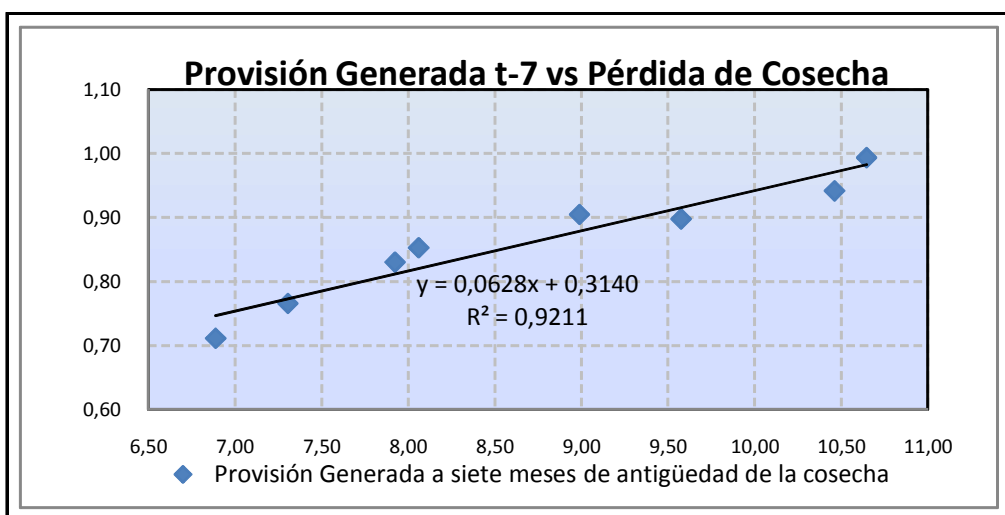


Gráfico A2.18 Regresión t-7

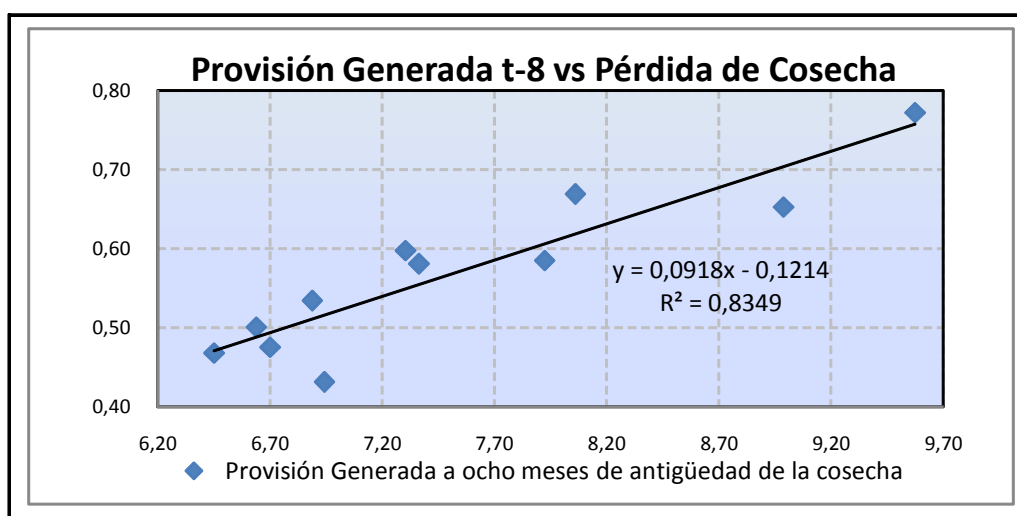


Gráfico A2.19 Regresión t-8

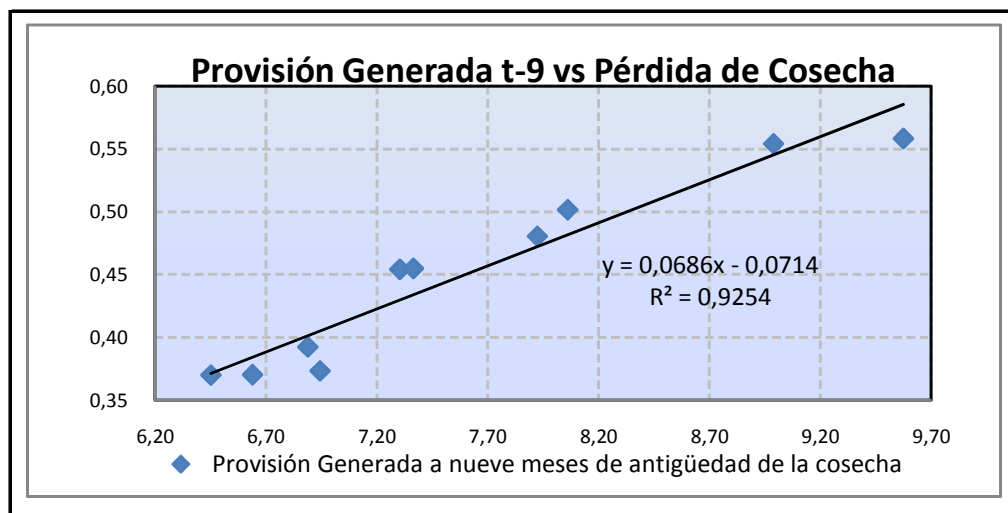


Gráfico A2.20 Regresión t-9

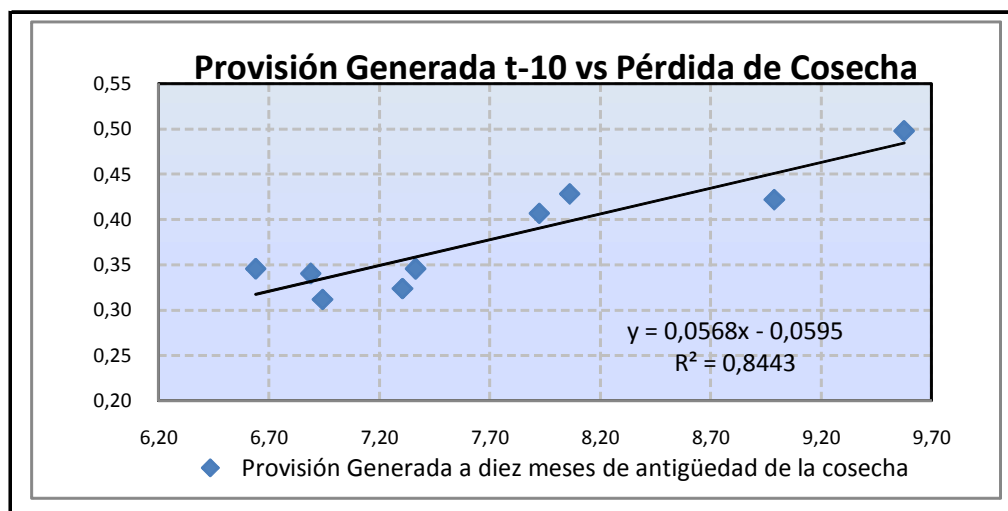


Gráfico A2.21 Regresión t-10

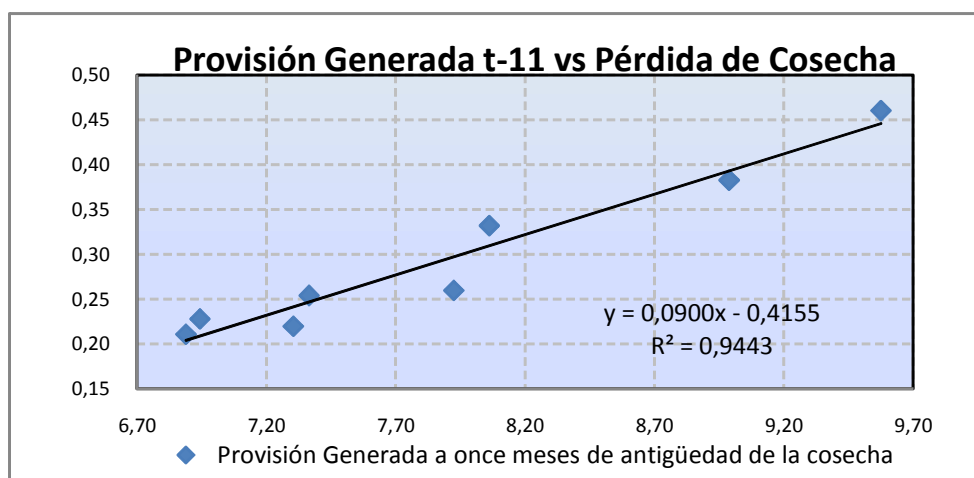


Gráfico A2.22 Regresión t-11

## ANEXO III

### CÓDIGO FUENTE PANEL DE MONITOREO

-----Este Módulo sirve para ingreso y salida de cada uno de los Reportes-----

-----

```
Sub Reportes_Financieros()  
Application.ScreenUpdating = False  
Sheets("Resultados_Pérdida").Visible = True  
Sheets("Reportes").Visible = False  
Sheets("Resultados_Pérdida").Select  
Cells(1, 1).Select  
Sheets("Resultados_Pérdida").ScrollArea = "A1:P38"  
End Sub
```

```
Sub Financieros_Reportes()  
Application.ScreenUpdating = False  
Sheets("Reportes").Visible = True  
Sheets("Resultados_Pérdida").Visible = False  
Sheets("Reportes").Select  
End Sub
```

```
Sub Reportes_Eficiencia()  
Application.ScreenUpdating = False  
Sheets("Eficiencia").Visible = True  
Sheets("Reportes").Visible = False  
Sheets("Eficiencia").Select  
Sheets("Eficiencia").ScrollArea = "A1:Q52"  
End Sub
```

```
Sub Eficiencia_reportes()  
Application.ScreenUpdating = False
```

```
Sheets("Reportes").Visible = True
Sheets("Eficiencia").Visible = False
Sheets("Reportes").Select
Range("a1").Select
End Sub
```

```
Sub Reportes_RR()
Application.ScreenUpdating = False
Sheets("RR").Visible = True
Sheets("Reportes").Visible = False
Sheets("RR").Select
Sheets("RR").ScrollArea = "A1:Q75"
End Sub
```

```
Sub RR_Reportes()
Application.ScreenUpdating = False
Sheets("Reportes").Visible = True
Sheets("RR").Visible = False
Sheets("Reportes").Select
Cells(1, 1).Select
End Sub
```

```
Sub Reportes_Corr()
Application.ScreenUpdating = False
Application.ScreenUpdating = False
Sheets("Resultados").Visible = True
Sheets("Reportes").Visible = False
Sheets("Resultados").Select
Sheets("Resultados").ScrollArea = "A1:U40"
```

```
End Sub
```

```
Sub Corr_Reporte()
Application.ScreenUpdating = False
Sheets("Reportes").Visible = True
```



```
Sheets("Resultados").Visible = False
Sheets("Reportes").Select
End Sub
```

```
Sub Reportes_Resultados2()
Application.ScreenUpdating = False
Application.ScreenUpdating = False
Sheets("Resultados (2)").Visible = True
Sheets("Reportes").Visible = False
Sheets("Resultados (2)").Select
Sheets("Resultados (2)").ScrollArea = "A1:U40"
```

```
End Sub
```

```
Sub Resultados2_Reportes()
Application.ScreenUpdating = False
Sheets("Reportes").Visible = True
Sheets("Resultados (2)").Visible = False
Sheets("Reportes").Select
```

```
End Sub
```

```
Sub Reportes_AC()
Application.ScreenUpdating = False
Application.ScreenUpdating = False
Sheets("AC").Visible = True
Sheets("Reportes").Visible = False
Sheets("AC").Select
Sheets("AC").ScrollArea = "A1:U65000"
```

```
End Sub
```

```
Sub AC_Reportes()
Application.ScreenUpdating = False
Sheets("Reportes").Visible = True
```

```

Sheets("AC").Visible = False
Sheets("Reportes").Select
End Sub

Sub Reportes_KMO()
    Sheets("KMO").Visible = True
    Sheets("KMO").Select
    Sheets("Reportes").Visible = False
End Sub

Sub KMO_Reportes()
    Sheets("Reportes").Visible = True
    Sheets("Reportes").Select
    Sheets("KMO").Visible = False
End Sub

Sub Reportes_PSI()
    Application.ScreenUpdating = False
    Sheets("PSI").Visible = True
    Sheets("Reportes").Visible = False
    Sheets("PSI").Select
End Sub

Sub Reportes_Corr2()
    Application.ScreenUpdating = False
    Sheets("Resultados (2)").Visible = True
    Sheets("Reportes").Visible = False
    Sheets("Resultados (2)").Select
End Sub

-----Módulo correspondiente a los Resultados por Perfil-----
-----

Sub Resultados_G()
    Application.ScreenUpdating = False
    Resultados_P
    Sheets("Resultados_Pérdida").Select

```

```
Range("d1").Select
End Sub
```

```
Sub Resultados2()
Sheets("Proceso Resultados").Visible = True
Sheets("Variables").Visible = True
Sheets("Variables").Select
Ciudad = Sheets("Variables").Range("Y2").Text
Sheets("Variables").Visible = False
Sheets("Proceso Resultados").Select
ActiveSheet.PivotTables("Resultados1").PivotFields("cre_ciudad_ub").CurrentPage =
Ciudad
Sheets("Proceso Resultados").Select
ActiveSheet.PivotTables("Resultados2").PivotFields("cre_ciudad_ub").CurrentPage =
Ciudad
End Sub
```

```
Sub Resultados_P()
Sheets("Variables").Visible = True
N = Sheets("Variables").Range("f20").Value
Sheets("Variables").Visible = False
Sheets("Proceso Resultados").Visible = True
Sheets("Proceso Resultados").Select
Range("AF2:AO2").Select
    Range(Selection, Selection.End(xlDown)).Select
    Range(Selection, Selection.End(xlDown)).Select
    Selection.ClearContents
a = Range("q1").Value
For i = 1 To N + 1
    If i <= a Then
        Range("af" & i + 1) = Range("n" & 24 + i).Text
        Range("ag" & i + 1) = Range("o" & 24 + i).Value
        Range("ah" & i + 1) = Range("p" & 24 + i).Value
    End If
End For
```

```

If i = 1 Then
    Range("ai" & 2) = Range("p" & 25)
Else
    If i <= a Then
        Range("ai" & i + 1) = Range("p" & 24 + i).Value + Range("ai" & i).Value
    Else
        Range("ai" & a + 1) = ""
    End If
End If

Range("aj" & i + 1) = Range("q" & 24 + i).Value
Range("ak" & i + 1) = Range("r" & 24 + i).Value
Range("al" & i + 1) = Range("s" & 24 + i).Value
Else
    If i > a And i < N + 1 Then
        Range("af" & i + 1) = "P" & i
        Range("ag" & i + 1) = ""
        Range("ah" & i + 1) = ""
        Range("ai" & i + 1) = 1
        Range("aj" & i + 1) = ""
        Range("ak" & i + 1) = ""
        Range("al" & i + 1) = ""
    Else
        If i = N + 1 Then
            Range("af" & i + 1) = "Total General"
            Range("ag" & i + 1) = Range("o" & 24 + i).Value
            Range("ah" & i + 1) = Range("p" & 24 + i).Value
            Range("aj" & i + 1) = Range("q" & 24 + i).Value
            Range("ak" & i + 1) = Range("r" & 24 + i).Value
            Range("al" & i + 1) = Range("s" & 24 + i).Value
        End If
    End If
End If
Next i

```

c = a - 1

```
Sheets("Resultados_Pérdida").Select
```

```
ActiveSheet.ChartObjects("Gráfico 13").Activate
```

```
ActiveChart.PlotArea.Select
```

```
ActiveChart.SeriesCollection(1).XValues = "=Resultados_Pérdida!R32C4:R" & 32 + c & "C4"
```

```
ActiveChart.SeriesCollection(1).Values = "=Resultados_Pérdida!R32C5:R" & 32 + c & "C5"
```

```
ActiveChart.SeriesCollection(2).XValues = "=Resultados_Pérdida!R32C4:R" & 32 + c & "C4"
```

```
ActiveChart.SeriesCollection(2).Values = "=Resultados_Pérdida!R32C6:R" & 32 + c & "C6"
```

```
ActiveChart.SeriesCollection(3).XValues = "=Resultados_Pérdida!R32C4:R" & 32 + c & "C4"
```

```
ActiveChart.SeriesCollection(3).Values = "=Resultados_Pérdida!R32C7:R" & 32 + c & "C7"
```

```
ActiveWindow.Visible = False
```

```
ActiveSheet.ChartObjects("Gráfico 18").Activate
```

```
ActiveChart.PlotArea.Select
```

```
ActiveChart.SeriesCollection(1).XValues = "=Resultados_Pérdida!R32C4:R" & 32 + c & "C4"
```

```
ActiveChart.SeriesCollection(1).Values = "=Resultados_Pérdida!R32C8:R" & 32 + c & "C8"
```

```
ActiveChart.SeriesCollection(2).XValues = "=Resultados_Pérdida!R32C4:R" & 32 + c & "C4"
```

```
ActiveChart.SeriesCollection(2).Values = "=Resultados_Pérdida!R32C9:R" & 32 + c & "C9"
```

```
ActiveChart.SeriesCollection(3).XValues = "=Resultados_Pérdida!R32C4:R" & 32 + c & "C4"
```

```
ActiveChart.SeriesCollection(3).Values = "=Resultados_Pérdida!R32C10:R" & 32 + c & "C10"
```

```
ActiveWindow.Visible = False
```

```

Sheets("Resultados_Pérdida").Range("k5").Value = Sheets("Proceso
Resultados").Range("u47").Value
Sheets("Resultados_Pérdida").Range("k7").Value = Sheets("Proceso
Resultados").Range("w47").Value
Sheets("Resultados_Pérdida").Range("m5").Value = Sheets("Proceso
Resultados").Range("w48").Value
Sheets("Resultados_Pérdida").Range("m7").Value = Sheets("Proceso
Resultados").Range("w49").Value
Sheets("Proceso Resultados").Visible = False
End Sub

```

-----Modulo Correspondiente al Reporte de la Razón de  
rechazo-----

```

Sub RR()
Application.ScreenUpdating = False
Sheets("RR").Select
'Proceso para generar Filtros sobre la BD
Ciudad = Sheets("Proceso_RR").Range("T3").Text

Sheets("Proceso_RR").PivotTables("Proceso_RR1").PivotFields("cre_ciudad_ub").Current
Page = Ciudad

Sheets("Proceso_RR").PivotTables("Proceso_RR2").PivotFields("cre_ciudad_ub").Current
Page = Ciudad
P = Sheets("Variables").Range("F20").Value
Sheets("Proceso_RR").Visible = True
Sheets("Proceso_RR").Select
Dim M_RR(1 To 10, 1 To 10) As Double

For j = 1 To 5
For i = 1 To P + 1 'Se agrega 1 por lo totales
'Primero Cargamos la tasa morosidad

```

```

If j = 1 Then
    Cells(5 + i, 7).Select
    M_RR(i, j) = Cells(5 + i, 7)
Else
    If j > 1 And j < 4 Then
        Cells(5 + i, 15 + j).Select
        M_RR(i, j) = Cells(5 + i, 14 + j)

    Else
        If j = 4 Then
            M_RR(i, j) = 1 - M_RR(i, 3)
        End If
    End If
End If
End If
Sheets("RR").Select
Cells(32 + i, 4 + j) = M_RR(i, j)
Sheets("Proceso_RR").Select
Next i
Next j
Sheets("Proceso_RR").Visible = False
Sheets("RR").Select
Cells(1, 1).Select

```

End

-----Modulo Correspondiente al Reporte de Eficiencia-----  
 -----

```

Sub Eficiencia_Percen()
    Application.ScreenUpdating = False
    Percentiles_Score
    Eficiencia
    Sheets("Eficiencia").Select

```

```
Range("a1").Select
End Sub
```

```
Sub Percentiles_Score()
    Sheets("ProcedimientoPercent").Visible = True
    Sheets("ProcedimientoPercent").Select
    h = Range("r1").Value
    a = Range("s1").Value
    Range("ce" & 2, "cp" & a).Select
        Selection.ClearContents
    Range("p" & 2, "q" & 22).Select
        Selection.ClearContents
    Range("A" & 1, "G" & a).AdvancedFilter Action:=xlFilterCopy, CriteriaRange:= _
        Range("al6:al7"), CopyToRange:=Range("CE1:CK1"), Unique:=False
    If h <= 0 Then
        MsgBox "Imposible realizar Partición"
    Else
        If h = 1 Then
            MsgBox "No esta realizando ¡Ninguna Partición!"
        Else
            k = (100 / h) / 100
            For i = 1 To h
                X = k * i
                Range("q" & i + 1).Value = X
            Next i
            Range("p2").Select
            b = "=+PERCENTILE(R2C83:R" & a & "C83,RC[1])"
            ActiveCell.FormulaR1C1 = b
            Range("p2").Select
            Selection.AutoFill Destination:=Range("p" & 2, "p" & h + 1),
            Type:=xlFillDefault
            c = Range("T1").Value
```



```

For k = 2 To c
    y = Range("CE" & k).Value
    j = 0
    Do While Range("CF" & k) = ""
        j = j + 1
        If y > Range("p" & j).Value And y <= Range("p" & j + 1).Value Then
Range("CF" & k).Value = j
        Loop
    Next k
For t = 2 To c
    v = 0
    Do While Range("cg" & t).Value = ""
        v = v + 1
        If Range("CF" & t).Value = v Then Range("cg" & t).Value = Range("p" &
v).Value
        If Range("cf" & t).Value = v Then Range("ch" & t).Value = Range("p" & v
+ 1).Value
        Loop
    Next t
Range("s14").Select
ActiveSheet.PivotTables("Percentiles").PivotCache.Refresh
ActiveWorkbook.ShowPivotTableFieldList = False
Selection.Sort Order1:=xlAscending, Type:=xlSortLabels, OrderCustom:=1, _
Orientation:=xlTopToBottom
End If
End If
Sheets("ProcedimientoPercent").Visible = False
End Sub

```

```

Sub Eficiencia()
Sheets("Eficiencia").Select
    Ciudad = Sheets("ProcedimientoPercent").Range("A14").Text

```

```
Sheets("ProcedimientoPercent").PivotTables("Percentiles2").PivotFields("cre_ciudad_ub")  
.CurrentPage = Ciudad
```

```
a = Sheets("ProcedimientoPercent").Range("r1").Value  
  Sheets("Eficiencia").Select  
  Range("a1").Select  
  Range("L31").Select  
  Range(Selection, Selection.End(xlToLeft)).Select  
  Range(Selection, Selection.End(xlDown)).Select  
  Selection.Delete Shift:=xlUp  
  For i = 1 To a  
  Range("b30:l30").Select  
  Selection.Copy  
  Range("b" & 29 + i).Select  
  Selection.PasteSpecial Paste:=xlPasteFormats, Operation:=xlNone, _  
    SkipBlanks:=False, Transpose:=False  
  Application.CutCopyMode = False  
  Selection.Copy  
  Application.CutCopyMode = False  
  Range("b" & 29 + i).Value = i  
  Range("c" & 29 + i).Select  
  ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-  
1],ProcedimientoPercent!C[16]:C[22],3,0)"  
  Range("d" & 29 + i).Select  
  ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-  
2],ProcedimientoPercent!C[15]:C[21],4,0)"  
  Range("e" & 29 + i).Select  
  ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-  
3],ProcedimientoPercent!C[14]:C[20],2,0)"  
  Range("f" & 29 + i).Select  
  ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-  
4],ProcedimientoPercent!C[13]:C[19],5,0)"
```

```

If i > 1 Then
    Range("g" & 29 + i).Value = Range("f" & 29 + i).Value + Range("g" & 28 +
i).Value
Else
    If i = 1 Then
        Range("g" & 29 + i).Value = Range("f" & 29 + i).Value
    End If
End If
Range("h" & 29 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-
6],ProcedimientoPercent!C[11]:C[17],6,0)"
If i > 1 Then
    Range("i" & 29 + i).Value = Range("h" & 29 + i).Value + Range("i" & 28 +
i).Value
Else
    If i = 1 Then
        Range("i" & 29 + i).Value = Range("h" & 29 + i).Value
    End If
End If
Range("j" & 29 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-
8],ProcedimientoPercent!C[9]:C[15],7,0)"
Range("k" & 29 + i).Value = Abs(Range("g" & 29 + i).Value - Range("i" & 29 +
i).Value)
If i > 1 Then
    Range("l" & 29 + i).Value = (Range("g" & 29 + i).Value - Range("g" & 28 +
i).Value) * (Range("i" & 29 + i).Value + Range("i" & 28 + i).Value)
Else
    If i = 1 Then
        Range("l" & 30).Value = Range("g" & 30).Value * Range("i" & 30).Value
    End If
End If
Next i

```

```

Range("k28").Select
ActiveCell.FormulaR1C1 = "=MAX(R[2]C:R[200]C)"
Selection.Copy
Range("k" & 30 + a).Select
Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
:=False, Transpose:=False
Application.CutCopyMode = False
Range("k" & 30 + a).Select
With Selection.Borders(xlEdgeLeft)
.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With
With Selection.Borders(xlEdgeTop)
.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With

With Selection.Borders(xlEdgeBottom)
.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With

With Selection.Borders(xlEdgeRight)
.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With
Selection.Font.ColorIndex = 11
Selection.Font.Bold = False
With Selection.Font
.Name = "Arial Black"

```

```

.Size = 12
.Strikethrough = False
.Superscript = False
.Subscript = False
.OutlineFont = False
.Shadow = False
.Underline = xlUnderlineStyleNone
.ColorIndex = 11
End With
With Selection.Font
.Name = "Arial Black"
.Size = 12
.Strikethrough = False
.Superscript = False
.Subscript = False
.OutlineFont = False
.Shadow = False
.Underline = xlUnderlineStyleNone
.ColorIndex = 11
End With
Range("I28").Select
ActiveCell.FormulaR1C1 = "=abs(1-sum(R[2]C:R" & 29 + a & "C))"
Selection.Copy
Range("I" & 30 + a).Select
Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
:=False, Transpose:=False
Application.CutCopyMode = False
Range("I" & 30 + a).Select
With Selection.Borders(xlEdgeLeft)
.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With

```

With Selection.Borders(xlEdgeTop)

.LineStyle = xlContinuous

.Weight = xlMedium

.ColorIndex = xlAutomatic

End With

With Selection.Borders(xlEdgeBottom)

.LineStyle = xlContinuous

.Weight = xlMedium

.ColorIndex = xlAutomatic

End With

With Selection.Borders(xlEdgeRight)

.LineStyle = xlContinuous

.Weight = xlMedium

.ColorIndex = xlAutomatic

End With

Selection.Font.ColorIndex = 11

Selection.Font.Bold = False

With Selection.Font

.Name = "Arial Black"

.Size = 12

.Strikethrough = False

.Superscript = False

.Subscript = False

.OutlineFont = False

.Shadow = False

.Underline = xlUnderlineStyleNone

.ColorIndex = 11

End With

With Selection.Font

.Name = "Arial Black"

.Size = 12

.Strikethrough = False

```

.Superscript = False
.Subscript = False
.OutlineFont = False
.Shadow = False
.Underline = xlUnderlineStyleNone
.ColorIndex = 11
End With
Selection.NumberFormat = "0.0% "
ActiveSheet.ChartObjects("Gráfico 2").Activate
ActiveChart.PlotArea.Select
ActiveChart.SeriesCollection(1).XValues = "=Eficiencia!R30C2:R" & 29 + a & "C2"
ActiveChart.SeriesCollection(1).Values = "=Eficiencia!R30C7:R" & 29 + a & "C7"
ActiveChart.SeriesCollection(2).XValues = "=Eficiencia!R30C2:R" & 29 + a & "C2"
ActiveChart.SeriesCollection(2).Values = "=Eficiencia!R30C9:R" & 29 + a & "C9"
ActiveSheet.ChartObjects("Gráfico 3").Activate
ActiveChart.PlotArea.Select
ActiveChart.SeriesCollection(1).Values = "=Eficiencia!R30C6:R" & 29 + a & "C6"
ActiveChart.SeriesCollection(2).Values = "=Eficiencia!R30C8:R" & 29 + a & "C8"
ActiveChart.SeriesCollection(3).XValues = "=Eficiencia!R30C2:R" & 29 + a & "C2"
ActiveChart.SeriesCollection(3).Values = "=Eficiencia!R30C10:R" & 29 + a & "C10"
Sheets("Eficiencia").Range("L6").Value =
Sheets("ProcedimientoPercent").Range("ah5").Value
Sheets("Eficiencia").Range("L8").Value =
Sheets("ProcedimientoPercent").Range("ah6").Value
Sheets("Eficiencia").Range("n6").Value =
Sheets("ProcedimientoPercent").Range("ah7").Value
Sheets("Eficiencia").Range("n8").Value =
Sheets("ProcedimientoPercent").Range("ah8").Value
End Sub

-----Modulo Correspondiente al Reporte de Correlaciones de VA incluidas-----
-----
Sub Proceso_General()

```

```

Tabla_Contingencia
CeldasVariables
Sheets("Resultados").Select
Range("a1").Select
End Sub

```

```

Sub Tabla_Contingencia()
    Sheets("Procedimiento").Visible = True
    Sheets("Variables").Visible = True
    Sheets("Procedimiento").Select
    Columns("Q").Select
    Selection.ClearContents
    Columns("s").Select
    Selection.ClearContents
    Columns("t").Select
    Selection.ClearContents
    Columns("u").Select
    Selection.ClearContents
    Columns("w").Select
    Selection.ClearContents
    If Range("d1").Value = "" And Range("c25").Text = "Malo" Then
    For i = 1 To Range("c1").Value
        With ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & i)
            .Orientation = xlRowField
            .Position = 1
        End With
    For j = 1 To 2
        d = Range("m" & 1).Value
        For k = 1 To d
            a = Range("g" & j)
            b = Range("j" & k).Value
            c = Range("n" & 1)

```



```

If j = 1 Then
    Range("f" & k + 25).Value = a * b * c
Else
    If j = 2 Then
        Range("g" & k - 1 + 26).Value = a * b * c
    End If
End If
Next k
Next j
b = "=+CHITEST(R26C2:R" & d + 25 & "C3, R26C6:R" & d + 25 & "C7)"
Range("j26").Value = b
Range("q" & i + 1) = Range("j26").Value
If Range("A27").Value = 1 Then
    Range("s" & i + 1).Value = Range("d27").Value /
Range("d28").Value
    Range("T" & i + 1).Value = Range("b27").Value /
Range("d27").Value
    Range("u" & i + 1).Value = Range("c27").Value /
Range("d27").Value
    If Range("c27").Value = 0 Then
        Range("w" & i + 1).Value = "----"
    Else
        If Range("c27").Value <> 0 Then
            Range("w" & i + 1).Value = Range("b27").Value /
Range("c27").Value
        End If
    End If
Else
    If Range("A27").Value <> 1 And Range("A26").Value = 1 Then
        Range("s" & i + 1).Value = 1
        Range("T" & i + 1).Value = Range("b27").Value /
Range("d27").Value

```

```

Range("u" & i + 1).Value = Range("c27").Value /
Range("d27").Value
If Range("c27").Value = 0 Then
    Range("w" & i + 1).Value = "----"
Else
    If Range("c27").Value <> 0 Then
        Range("w" & i + 1).Value = Range("b27").Value /
Range("c27").Value
    End If
End If
Else
    If Range("A27").Value <> 1 And Range("A26").Value = 0 Then
        Range("s" & i + 1).Value = 0
        Range("T" & i + 1).Value = "----"
        Range("u" & i + 1).Value = "----"
        Range("w" & i + 1).Value = "----"
    Else
        If Range("l1").Value > 4 Then
            h = Range("l1").Value
            If Range("a" & 25 + h - 2).Text <> "(en blanco)" Then
                Range("s" & i + 1).Value = 1
                Range("T" & i + 1).Value = Range("b" & 25 + h - 1).Value / Range("d" & 25
+ h - 1).Value
                Range("u" & i + 1).Value = Range("c" & 25 + h - 1).Value / Range("d" & 25
+ h - 1).Value
                If Range("c" & 25 + h - 1).Value = 0 Then
                    Range("w" & i + 1).Value = "----"
                Else
                    If Range("c" & 25 + h - 1).Value <> 0 Then
                        Range("w" & i + 1).Value = Range("b" & 25 + h - 1).Value / Range("c" & 25
+ h - 1).Value
                    End If
                End If
            End If
        End If
    End If
End If

```

```

        End If
        End If
    End If

    End If
End If
Range("q" & i + 1).Select
Range("f" & 26, "g" & d + 26 - 1).Select
Selection.ClearContents
Range("A25").Select
ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & i).Orientation = _
    xlHidden
Next i
Else
    If Range("d1").Value <> "" And Range("c25").Text = "Malo" Then
        z = Range("d1").Value
        Range("A25").Select
        ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & z).Orientation = _
            xlHidden
        For i = 1 To Range("c1").Value
            With ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & i)
                .Orientation = xlRowField
                .Position = 1
            End With
        Next i
    End If
End If

For j = 1 To 2
    d = Range("m" & 1).Value
    For k = 1 To d
        a = Range("g" & j)
        b = Range("j" & k).Value
        c = Range("n" & 1)
        If j = 1 Then
            Range("f" & k + 25).Value = a * b * c
        End If
    Next k
Next j

```

```

Else
    If j = 2 Then
        Range("g" & k - 1 + 26).Value = a * b * c
    End If
End If
Next k
Next j
b = "=+CHITEST(R26C2:R" & d + 25 & "C3, R26C6:R" & d + 25 & "C7)"
Range("j26").Value = b
Range("q" & i + 1) = Range("j26").Value
If Range("A27").Value = 1 Then
    Range("s" & i + 1).Value = Range("d27").Value /
Range("d28").Value
    Range("T" & i + 1).Value = Range("b27").Value /
Range("d27").Value
    Range("u" & i + 1).Value = Range("c27").Value /
Range("d27").Value
    If Range("c27").Value = 0 Then
        Range("w" & i + 1).Value = "----"
    Else
        If Range("c27").Value <> 0 Then
            Range("w" & i + 1).Value = Range("b27").Value /
Range("c27").Value
        End If
    End If
End If

Else
    If Range("A27").Value <> 1 And Range("A26").Value = 1 Then
        Range("s" & i + 1).Value = 1
        Range("T" & i + 1).Value = Range("b27").Value /
Range("d27").Value
        Range("u" & i + 1).Value = Range("c27").Value /
Range("d27").Value
    End If
End If

```

```

        If Range("c27").Value = 0 Then
            Range("w" & i + 1).Value = "----"
        Else
            If Range("c27").Value <> 0 Then
                Range("w" & i + 1).Value = Range("b27").Value /
Range("c27").Value
            End If
        End If
    Else
        If Range("A27").Value <> 1 And Range("A26").Value = 0 Then
            Range("s" & i + 1).Value = 0
            Range("T" & i + 1).Value = "----"
            Range("u" & i + 1).Value = "----"
            Range("w" & i + 1).Value = "----"
        Else
            If Range("l1").Value > 4 Then
                h = Range("l1").Value
                If Range("a" & 25 + h - 2).Text <> "(en blanco)" Then
                    Range("s" & i + 1).Value = 1
                    Range("T" & i + 1).Value = Range("b" & 25 + h - 1).Value / Range("d" & 25
+ h - 1).Value
                    Range("u" & i + 1).Value = Range("c" & 25 + h - 1).Value / Range("d" & 25
+ h - 1).Value
                If Range("c" & 25 + h - 1).Value = 0 Then
                    Range("w" & i + 1).Value = "----"
                Else
                    If Range("c" & 25 + h - 1).Value <> 0 Then
                        Range("w" & i + 1).Value = Range("b" & 25 + h - 1).Value / Range("c" & 25
+ h - 1).Value
                    End If
                End If
            End If
        End If
    End If
End If

```

```

        End If
    End If
End If
Range("q" & i + 1).Select
Range("f" & 26, "g" & d + 26 - 1).Select
Selection.ClearContents
Range("A25").Select
ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & i).Orientation = _
    xlHidden
Next i
End If
End If
If Range("c25").Text <> "Malo" Then
    MsgBox "No se puede ejecutar el proceso"
End If
Sheets("Resultados").Range("m4").Value = Sheets("Procedimiento").Range("b2").Value
Sheets("Resultados").Range("m6").Value = Sheets("Procedimiento").Range("b3").Value
Sheets("Resultados").Range("o4").Value = Sheets("Procedimiento").Range("b4").Value
Sheets("Resultados").Range("o6").Value = Sheets("Procedimiento").Range("b5").Value
Sheets("Procedimiento").Visible = False
Sheets("Variables").Visible = False
End Sub
Sub CeldasVariables()
    Sheets("Resultados").Select
    Selection.AutoFilter Field:=1
    Rows("11:11").Select
    Range(Selection, Selection.End(xlDown)).Select
    Selection.Delete Shift:=xlUp
    Range("a10").Value = "x1"
    Range("B10:q10").Select
    Selection.Copy
    w = Sheets("Procedimiento").Range("c1").Value - 1
    For i = 1 To w

```

```

Range("a" & 10 + i).Value = "x" & i + 1
Range("B" & i + 10, "p" & i + 10).Select
Selection.PasteSpecial Paste:=xlPasteFormats, Operation:=xlNone, _
SkipBlanks:=False, Transpose:=False
Application.CutCopyMode = False
Selection.Copy
Range("B10:D10").Select
Next i
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-1],Variables!C:C[1],2,0)"
Selection.AutoFill Destination:=Range("B10:D" & w + 10 & "")
Columns("Q:Q").Select
Selection.EntireColumn.Hidden = False
For j = 1 To w + 1
Range("E" & j + 9, "G" & j + 9).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-4],Procedimiento!C[11]:C[13],2,0)"
Range("h" & j + 9, "i" & j + 9).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-7],Procedimiento!C[8]:C[10],3,0)"
Range("i" & j + 9).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-8],Procedimiento!C[7]:C[10],4,0)"
Range("j" & j + 9).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-9],Procedimiento!C[6]:C[13],5,0)"
Range("k" & j + 9).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-10],Procedimiento!C[5]:C[12],6,0)"
Range("l" & j + 9).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-11],Procedimiento!C[4]:C[11],7,0)"
Range("m" & j + 9).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-12],Procedimiento!C[3]:C[10],8,0)"
If Range("m" & j + 9).Value <> "----" Then

If Range("l" & j + 9).Value > 0 And Range("m" & j + 9).Value <= Range("m6").Value
Then
Range("n" & j + 9).Value = "Alineado"
Else

```

```

If Range("l" & j + 9).Value < 0 And Range("m" & j + 9).Value >= Range("m6").Value
Then
    Range("n" & j + 9).Value = "Alineado"
    Else
    Range("n" & j + 9).Value = "Desalineado"
End If
End If
Else
    Range("n" & j + 9).Value = "----"
End If
If Range("h" & j + 9).Text = "Independiente" And Sheets("Resultados (2)").Range("h" & j
+ 9).Text = "Independiente" Then
    Range("p" & j + 9).Select
    ActiveCell.FormulaR1C1 = "1"
    Selection.Font.ColorIndex = 3
    With Selection.Interior
        .ColorIndex = 3
        .Pattern = xlSolid
    End With
Else
    If Range("m" & j + 9).Value <> "----" Then
    If Range("i" & j + 9).Value > 0.05 And Range("n" & j + 9).Text = "Desalineado"
    And Abs(Range("m" & j + 9).Value - Range("m" & 6).Value) > 3 Then
    Range("p" & j + 9).Select
    ActiveCell.FormulaR1C1 = "1"
    Selection.Font.ColorIndex = 3
    With Selection.Interior
        .ColorIndex = 3
        .Pattern = xlSolid
    End With
End With
Else

```



```

If Range("h" & j + 9).Text = "Independiente" And Sheets("Resultados (2)").Range("h" & j
+ 9).Text = "Correlacionado" And Abs(Range("i" & j + 9).Value - Sheets("Resultados
(2)").Range("i" & j + 9).Value) > 0.05 Then
Range("p" & j + 9).Select
ActiveCell.FormulaR1C1 = "2"
Selection.Font.ColorIndex = 6
With Selection.Interior
.ColorIndex = 6
.Pattern = xlSolid
End With
Else
If Range("h" & j + 9).Text = "Correlacionado" And Sheets("Resultados (2)").Range("h" &
j + 9).Text = "Independiente" And Abs(Range("i" & j + 9).Value - Sheets("Resultados
(2)").Range("i" & j + 9).Value) > 0.05 Then
Range("p" & j + 9).Select
ActiveCell.FormulaR1C1 = "2"
Selection.Font.ColorIndex = 6
With Selection.Interior
.ColorIndex = 6
.Pattern = xlSolid
End With
Else
If Range("h" & j + 9).Text = "Correlacionado" And Range("n" & j + 9).Text =
"Desalineado" Then
Range("p" & j + 9).Select
ActiveCell.FormulaR1C1 = "2"
Selection.Font.ColorIndex = 6
With Selection.Interior
.ColorIndex = 6
.Pattern = xlSolid
End With
Else
Range("p" & j + 9).Select

```

```

ActiveCell.FormulaR1C1 = "3"
Selection.Font.ColorIndex = 4
With Selection.Interior
.ColorIndex = 4
.Pattern = xlSolid
End With
End If
End If
End If
End If
End If
End If
End If
Range("q" & j + 9) = Range("p" & j + 9).Value
Next j
Columns("Q:Q").Select
    Selection.EntireColumn.Hidden = True
Range("o10").Select
Selection.Copy
    Range("o10:o" & w + 10 & "").Select
    Selection.PasteSpecial Paste:=xlPasteFormulas, Operation:=xlNone, _
        SkipBlanks:=False, Transpose:=False
Range("B1").Select
    Application.CutCopyMode = False
End Sub

```

-----Módulo Correspondiente al Reporte de Correlaciones de VA no Incluidas-----  
-----

```

Sub Proceso_General_2()
Tabla_Contingencia_2
CeldasVariables_2
Sheets("Resultados (2)").Select
Range("a1").Select

```

End Sub

Sub Tabla\_Contingencia\_2()

Sheets("Procedimiento (2)").Visible = True

Sheets("Variables").Visible = True

Sheets("Procedimiento (2)").Select

Ciudad = Sheets("Variables").Range("y2").Text

Columns("Q").Select

Selection.ClearContents

Columns("s").Select

Selection.ClearContents

Columns("t").Select

Selection.ClearContents

Columns("u").Select

Selection.ClearContents

Columns("w").Select

Selection.ClearContents

If Range("d1").Value = "" And Range("c25").Text = "Malo" Then

For i = 1 To Range("c1").Value

With ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & i)

.Orientation = xlRowField

.Position = 1

End With

For j = 1 To 2

d = Range("m" & j).Value

For k = 1 To d

a = Range("g" & j)

b = Range("j" & k).Value

c = Range("n" & j)

If j = 1 Then

Range("f" & k + 25).Value = a \* b \* c

Else

```

        If j = 2 Then
            Range("g" & k - 1 + 26).Value = a * b * c
        End If
    End If
Next k
Next j
b = "=+CHITEST(R26C2:R" & d + 25 & "C3, R26C6:R" & d + 25 & "C7)"
Range("j26").Value = b
Range("q" & i + 1) = Range("j26").Value
If Range("A27").Value = 1 Then
    Range("s" & i + 1).Value = Range("d27").Value / Range("d28").Value
    Range("T" & i + 1).Value = Range("b27").Value / Range("d27").Value
    Range("u" & i + 1).Value = Range("c27").Value / Range("d27").Value
    If Range("c27").Value = 0 Then
        Range("w" & i + 1).Value = "----"
    Else
        If Range("c27").Value <> 0 Then
            Range("w" & i + 1).Value = Range("b27").Value / Range("c27").Value
        End If
    End If
Else
    If Range("A27").Value <> 1 And Range("A26").Value = 1 Then
        Range("s" & i + 1).Value = 1
        Range("T" & i + 1).Value = Range("b27").Value / Range("d27").Value
        Range("u" & i + 1).Value = Range("c27").Value / Range("d27").Value
        If Range("c27").Value = 0 Then
            Range("w" & i + 1).Value = "----"
        Else
            If Range("c27").Value <> 0 Then
                Range("w" & i + 1).Value = Range("b27").Value / Range("c27").Value
            End If
        End If
    End If
Else

```

```

If Range("A27").Value <> 1 And Range("A26").Value = 0 Then
    Range("s" & i + 1).Value = 0
    Range("T" & i + 1).Value = "----"
    Range("u" & i + 1).Value = "----"
    Range("w" & i + 1).Value = "----"

Else
If Range("l1").Value > 4 Then
    h = Range("l1").Value
If Range("a" & 25 + h - 2).Text <> "(en blanco)" Then
    Range("s" & i + 1).Value = 1
    Range("T" & i + 1).Value = Range("b" & 25 + h - 1).Value / Range("d" & 25 + h -
1).Value
    Range("u" & i + 1).Value = Range("c" & 25 + h - 1).Value / Range("d" & 25 + h -
1).Value
If Range("c" & 25 + h - 1).Value = 0 Then
    Range("w" & i + 1).Value = "----"

Else
If Range("c" & 25 + h - 1).Value <> 0 Then
    Range("w" & i + 1).Value = Range("b" & 25 + h - 1).Value / Range("c" & 25 + h -
1).Value

End If
End If
End If
End If
End If
End If
End If

End If

Range("q" & i + 1).Select
Range("f" & 26, "g" & d + 26 - 1).Select
Selection.ClearContents
Range("A25").Select
ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & i).Orientation = _
xlHidden

```

```

Next i
Else
  If Range("d1").Value <> "" And Range("c25").Text = "Malo" Then
    z = Range("d1").Value
    Range("A25").Select
    ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & z).Orientation = _
      xlHidden
    For i = 1 To Range("c1").Value
      With ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & i)
        .Orientation = xlRowField
        .Position = 1
      End With
    End With

  For j = 1 To 2
    d = Range("m" & 1).Value
    For k = 1 To d
      a = Range("g" & j)
      b = Range("j" & k).Value
      c = Range("n" & 1)
      If j = 1 Then
        Range("f" & k + 25).Value = a * b * c
      Else
        If j = 2 Then
          Range("g" & k - 1 + 26).Value = a * b * c
        End If
      End If
    End If
  Next k
Next j
b = "=+CHITEST(R26C2:R" & d + 25 & "C3, R26C6:R" & d + 25 & "C7)"
Range("j26").Value = b
Range("q" & i + 1) = Range("j26").Value
If Range("A27").Value = 1 Then
  Range("s" & i + 1).Value = Range("d27").Value / Range("d28").Value

```

```

Range("T" & i + 1).Value = Range("b27").Value / Range("d27").Value
Range("u" & i + 1).Value = Range("c27").Value / Range("d27").Value
If Range("c27").Value = 0 Then
    Range("w" & i + 1).Value = "----"
Else
If Range("c27").Value <> 0 Then
    Range("w" & i + 1).Value = Range("b27").Value / Range("c27").Value
End If
End If
Else
    If Range("A27").Value <> 1 And Range("A26").Value = 1 Then
        Range("s" & i + 1).Value = 1
        Range("T" & i + 1).Value = Range("b27").Value / Range("d27").Value
        Range("u" & i + 1).Value = Range("c27").Value / Range("d27").Value
        If Range("c27").Value = 0 Then
            Range("w" & i + 1).Value = "----"
        Else
            If Range("c27").Value <> 0 Then
                Range("w" & i + 1).Value = Range("b27").Value / Range("c27").Value
                End If
            End If
        Else
            If Range("A27").Value <> 1 And Range("A26").Value = 0 Then
                Range("s" & i + 1).Value = 0
                Range("T" & i + 1).Value = "----"
                Range("u" & i + 1).Value = "----"
                Range("w" & i + 1).Value = "----"
            Else
                If Range("l1").Value > 4 Then
                    h = Range("l1").Value
                    If Range("a" & 25 + h - 2).Text <> "(en blanco)" Then
                        Range("s" & i + 1).Value = 1
                    End If
                End If
            End If
        End If
    End If
End If

```

```

Range("T" & i + 1).Value = Range("b" & 25 + h - 1).Value / Range("d" & 25 + h -
1).Value
Range("u" & i + 1).Value = Range("c" & 25 + h - 1).Value / Range("d" & 25 + h -
1).Value
If Range("c" & 25 + h - 1).Value = 0 Then
    Range("w" & i + 1).Value = "----"
Else
If Range("c" & 25 + h - 1).Value <> 0 Then
    Range("w" & i + 1).Value = Range("b" & 25 + h - 1).Value / Range("c" & 25 + h -
1).Value
End If
End If
End If
End If
End If
End If
End If
End If
End If
End If
Range("q" & i + 1).Select
Range("f" & 26, "g" & d + 26 - 1).Select
Selection.ClearContents
Range("A25").Select
ActiveSheet.PivotTables("Tabla Monitoreo").PivotFields("x" & i).Orientation = _
xlHidden
Next i
End If
End If
If Range("c25").Text <> "Malo" Then
    MsgBox "No se puede ejecutar el proceso"
End If
Sheets("Resultados (2)").Range("m4").Value = Sheets("Procedimiento
(2)").Range("b2").Value
Sheets("Resultados (2)").Range("m6").Value = Sheets("Procedimiento
(2)").Range("b3").Value

```



```

Sheets("Resultados (2)").Range("o4").Value = Sheets("Procedimiento
(2)").Range("b4").Value
Sheets("Resultados (2)").Range("o6").Value = Sheets("Procedimiento
(2)").Range("b5").Value
Sheets("Procedimiento (2)").Visible = False
Sheets("Variables").Visible = False
End Sub

```

```

Sub CeldasVariables_2()
    Sheets("Resultados (2)").Select
    Selection.AutoFilter Field:=1
    Rows("11:11").Select
    Range(Selection, Selection.End(xlDown)).Select
    Selection.Delete Shift:=xlUp
    Range("a10").Value = "x1"
    Range("B10:q10").Select
    Selection.Copy
    w = Sheets("Procedimiento (2)").Range("c1").Value - 1
    For i = 1 To w
        Range("a" & 10 + i).Value = "x" & i + 1
        Range("B" & i + 10, "r" & i + 10).Select
        Selection.PasteSpecial Paste:=xlPasteFormats, Operation:=xlNone, _
        SkipBlanks:=False, Transpose:=False
        Application.CutCopyMode = False
        Selection.Copy
        Range("B10:D10").Select
    Next i
    ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-1],Variables!C:C[1],2,0)"
    Selection.AutoFill Destination:=Range("B10:D" & w + 10 & "")
    Columns("Q:Q").Select
    Selection.EntireColumn.Hidden = False

```

```

For j = 1 To w + 1
    Range("E" & j + 9, "G" & j + 9).Select
    ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-4],'Procedimiento
(2)'!C[11]:C[13],2,0)"
    Range("h" & j + 9, "i" & j + 9).Select
    ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-7],'Procedimiento (2)'!C[8]:C[10],3,0)"
    Range("i" & j + 9).Select
    ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-8],'Procedimiento (2)'!C[7]:C[10],4,0)"
    Range("j" & j + 9).Select
    ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-9],'Procedimiento (2)'!C[6]:C[13],5,0)"
    Range("k" & j + 9).Select
    ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-10],'Procedimiento
(2)'!C[5]:C[12],6,0)"
    Range("l" & j + 9).Select
    ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-11],'Procedimiento
(2)'!C[4]:C[11],7,0)"
    Range("m" & j + 9).Select
    ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-12],'Procedimiento
(2)'!C[3]:C[10],8,0)"
    If Range("m" & j + 9).Value <> "----" Then
        If Range("l" & j + 9).Value < 0 And Range("m" & j + 9).Value <=
Range("m6").Value Then
            Range("n" & j + 9).Value = "Alineado"
        Else
            If Range("l" & j + 9).Value > 0 And Range("m" & j + 9).Value >=
Range("m6").Value Then
                Range("n" & j + 9).Value = "Alineado"
            Else
                Range("n" & j + 9).Value = "Desalineado"
            End If
        End If
    Else
        Range("n" & j + 9).Value = "----"
    End If
End For

```

```

End If
If Range("n" & j + 9).Text = "Desalineado" Then
    Range("p" & j + 9).Select
    ActiveCell.FormulaR1C1 = "1"
    Selection.Font.ColorIndex = 3
    With Selection.Interior
        .ColorIndex = 3
        .Pattern = xlSolid
    End With
Else
If Range("h" & j + 9).Text = "Independiente" Then
    Range("p" & j + 9).Select
    ActiveCell.FormulaR1C1 = "1"
    Selection.Font.ColorIndex = 3
    With Selection.Interior
        .ColorIndex = 3
        .Pattern = xlSolid
    End With

        Else: Range("p" & j + 9).Select
            ActiveCell.FormulaR1C1 = "3"
            Selection.Font.ColorIndex = 4
            With Selection.Interior
                .ColorIndex = 4
                .Pattern = xlSolid
            End With
        End If
End If
Range("q" & j + 9) = Range("p" & j + 9).Value
Next j
Columns("Q:Q").Select
    Selection.EntireColumn.Hidden = True
Range("o10").Select

```

```

Selection.Copy
    Range("o10:o" & w + 10 & "").Select
    Selection.PasteSpecial Paste:=xlPasteFormulas, Operation:=xlNone, _
        SkipBlanks:=False, Transpose:=False
Range("B1").Select
    Application.CutCopyMode = False
End Sub

```

----- Módulo Correspondiente al Reporte de Estabilidad Global -----  
 -----

```

Sub Percentiles_PSI()
    Sheets("ProcedimientoPSI").Visible = True
    Sheets("ProcedimientoPSI").Select
    h = Range("r1").Value
    a = Range("s1").Value
    d = Range("u1").Value
    Range("ce" & 2, "cp" & a).Select
        Selection.ClearContents
        Range("dk" & d, "dt" & d).Select
        Selection.ClearContents
    Range("p" & 2, "q" & 10000).Select
        Selection.ClearContents
    Range("A" & 1, "L" & a).AdvancedFilter Action:=xlFilterCopy, CriteriaRange:= _
        Range("BK1:BN2"), CopyToRange:=Range("CE1:CP1"), Unique:=False
    Range("DN1").Select
        Range("CV" & 1, "DD" & d).AdvancedFilter Action:=xlFilterCopy, CriteriaRange:= _
            Range("BT1:BT2"), CopyToRange:=Range("DN1:DV1"), Unique:=False
    If h <= 0 Then
        MsgBox "Imposible realizar Partición"
    Else
    If h = 1 Then
        MsgBox "No esta realizando ¡Ninguna partición!"

```

```

Else
    k = (100 / h) / 100
For i = 1 To h
    X = k * i
    Range("q" & i + 1).Value = X
Next i
Range("p2").Select
b = "=+PERCENTILE(R2C118:R" & d & "C118,RC[1])"
ActiveCell.FormulaR1C1 = b
Range("p2").Select
Selection.AutoFill Destination:=Range("p" & 2, "p" & h + 1), Type:=xlFillDefault
Range("p" & 2, "p" & h).Select
c = Range("T1").Value
For k = 2 To d
    y = Range("dn" & k).Value
    For j = 1 To h
        If y > Range("p" & j).Value And y <= Range("p" & j + 1).Value Then Range("do" &
k).Value = j
    Next j
Next k
For t = 2 To d
    For v = 1 To h
        If Range("do" & t).Value = v Then Range("dp" & t).Value = Range("p" & v).Value
        If Range("do" & t).Value = v Then Range("dq" & t).Value = Range("p" & v +
1).Value
    Next v
Next t
For k = 2 To c
    y = Range("CE" & k).Value
    For j = 1 To h
        If y > Range("p" & j).Value And y <= Range("p" & j + 1).Value Then Range("CF" &
k).Value = j
    Next j

```

```

Next k
For t = 2 To c
  For v = 1 To h
    If Range("CF" & t).Value = v Then Range("cg" & t).Value = Range("p" & v).Value
    If Range("cf" & t).Value = v Then Range("ch" & t).Value = Range("p" & v + 1).Value
  Next v
Next t

  Range("s14").Select
  ActiveSheet.PivotTables("PSI").PivotCache.Refresh
  ActiveWorkbook.ShowPivotTableFieldList = False
  Selection.Sort Order1:=xlAscending, Type:=xlSortLabels, OrderCustom:=1, _
  Orientation:=xlTopToBottom
  Range("ab31").Select
  ActiveSheet.PivotTables("PSI3").PivotCache.Refresh
  ActiveWorkbook.ShowPivotTableFieldList = False
  Selection.Sort Order1:=xlAscending, Type:=xlSortLabels, OrderCustom:=1, _
  Orientation:=xlTopToBottom
  Range("ab26").Select
  ActiveSheet.PivotTables("PSI4").PivotCache.Refresh
End If
End If
End Sub

Sub PSI()
Sheets("PSI").Select
a = Sheets("ProcedimientoPSI").Range("r1").Value
  Sheets("PSI").Select
  Range("a1").Select
  Rows("32:32").Select
  Range(Selection, Selection.End(xlDown)).Select
  Selection.Delete Shift:=xlUp
For i = 1 To a
Range("B31:O31").Select

```

```

Selection.Copy
Range("B" & 30 + i).Select
Selection.PasteSpecial Paste:=xlPasteFormats, Operation:=xlNone, _
    SkipBlanks:=False, Transpose:=False
    Application.CutCopyMode = False
    Selection.Copy
    Application.CutCopyMode = False
Range("B" & 30 + i).Value = i
Range("C" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-1],ProcedimientoPSI!C[25]:C[31],2,0)"
Range("d" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-2],ProcedimientoPSI!C[24]:C[30],3,0)"
Range("e" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-3],ProcedimientoPSI!C[23]:C[29],4,0)"
Range("f" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-4],ProcedimientoPSI!C[22]:C[28],5,0)"
Range("G" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-5],ProcedimientoPSI!C[21]:C[27],6,0)"
Range("H" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-6],ProcedimientoPSI!C[20]:C[26],7,0)"
Range("I" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-7],ProcedimientoPSI!C[10]:C[14],2,0)"
Range("J" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-8],ProcedimientoPSI!C[9]:C[13],3,0)"
Range("K" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-9],ProcedimientoPSI!C[8]:C[12],4,0)"
Range("L" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-10],ProcedimientoPSI!C[7]:C[11],5,0)"
Range("M" & 30 + i).Select
ActiveCell.FormulaR1C1 = "=(R" & 30 + i & "C9-R" & 30 + i & "C5)*LN(R" & 30 + i
& "C9/R" & 30 + i & "C5)"
Range("n" & 30 + i).Select

```

```
ActiveCell.FormulaR1C1 = "=(R" & 30 + i & "C10-R" & 30 + i & "C6)*LN(R" & 30 +  
i & "C10/R" & 30 + i & "C6)"
```

```
Range("o" & 30 + i).Select
```

```
ActiveCell.FormulaR1C1 = "=(R" & 30 + i & "C11-R" & 30 + i & "C7)*LN(R" & 30 +  
i & "C11/R" & 30 + i & "C7)"
```

```
Next i
```

```
Range("m28").Select
```

```
ActiveCell.FormulaR1C1 = "=sum(R[3]C:R[200]C)"
```

```
Selection.Copy
```

```
Range("m" & 31 + a).Select
```

```
Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
```

```
:=False, Transpose:=False
```

```
Application.CutCopyMode = False
```

```
' Range("m28").Select
```

```
' Selection.ClearContents
```

```
Range("m" & 31 + a).Select
```

```
With Selection.Borders(xlEdgeLeft)
```

```
.LineStyle = xlContinuous
```

```
.Weight = xlMedium
```

```
.ColorIndex = xlAutomatic
```

```
End With
```

```
With Selection.Borders(xlEdgeTop)
```

```
.LineStyle = xlContinuous
```

```
.Weight = xlMedium
```

```
.ColorIndex = xlAutomatic
```

```
End With
```

```
With Selection.Borders(xlEdgeBottom)
```

```
.LineStyle = xlContinuous
```

```
.Weight = xlMedium
```

```
.ColorIndex = xlAutomatic
```

```
End With
```

```
With Selection.Borders(xlEdgeRight)
```



```

.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With
Selection.Font.ColorIndex = 11
Selection.Font.Bold = False
With Selection.Font
.Name = "Arial Black"
.Size = 12
.Strikethrough = False
.Superscript = False
.Subscript = False
.OutlineFont = False
.Shadow = False
.Underline = xlUnderlineStyleNone
.ColorIndex = 11
End With
With Selection.Font
.Name = "Arial Black"
.Size = 12
.Strikethrough = False
.Superscript = False
.Subscript = False
.OutlineFont = False
.Shadow = False
.Underline = xlUnderlineStyleNone
.ColorIndex = 11
End With
Range("n28").Select
ActiveCell.FormulaR1C1 = "=sum(R[3]C:R[200]C)"
Selection.Copy
Range("n" & 31 + a).Select
Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _

```

```

:=False, Transpose:=False
Application.CutCopyMode = False
' Range("n28").Select
' Selection.ClearContents
Range("n" & 31 + a).Select
With Selection.Borders(xlEdgeLeft)
    .LineStyle = xlContinuous
    .Weight = xlMedium
    .ColorIndex = xlAutomatic
End With
With Selection.Borders(xlEdgeTop)
    .LineStyle = xlContinuous
    .Weight = xlMedium
    .ColorIndex = xlAutomatic
End With

With Selection.Borders(xlEdgeBottom)
    .LineStyle = xlContinuous
    .Weight = xlMedium
    .ColorIndex = xlAutomatic
End With
With Selection.Borders(xlEdgeRight)
    .LineStyle = xlContinuous
    .Weight = xlMedium
    .ColorIndex = xlAutomatic
End With
Selection.Font.ColorIndex = 11
Selection.Font.Bold = False
With Selection.Font
    .Name = "Arial Black"
    .Size = 12
    .Strikethrough = False
    .Superscript = False

```

```

.Subscript = False
.OutlineFont = False
.Shadow = False
.Underline = xlUnderlineStyleNone
.ColorIndex = 11
End With
With Selection.Font
.Name = "Arial Black"
.Size = 12
.Strikethrough = False
.Superscript = False
.Subscript = False
.OutlineFont = False
.Shadow = False
.Underline = xlUnderlineStyleNone
.ColorIndex = 11
End With
Range("o28").Select
ActiveCell.FormulaR1C1 = "=sum(R[3]C:R[200]C)"
Selection.Copy
Range("o" & 31 + a).Select
Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
:=False, Transpose:=False
Application.CutCopyMode = False
' Range("o28").Select
' Selection.ClearContents
Range("o" & 31 + a).Select
With Selection.Borders(xlEdgeLeft)
.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With
With Selection.Borders(xlEdgeTop)

```

```

.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With

With Selection.Borders(xlEdgeBottom)
.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With

With Selection.Borders(xlEdgeRight)
.LineStyle = xlContinuous
.Weight = xlMedium
.ColorIndex = xlAutomatic
End With

Selection.Font.ColorIndex = 11
Selection.Font.Bold = False
With Selection.Font
.Name = "Arial Black"
.Size = 12
.Strikethrough = False
.Superscript = False
.Subscript = False
.OutlineFont = False
.Shadow = False
.Underline = xlUnderlineStyleNone
.ColorIndex = 11
End With

With Selection.Font
.Name = "Arial Black"
.Size = 12
.Strikethrough = False
.Superscript = False

```

```

.Subscript = False
.OutlineFont = False
.Shadow = False
.Underline = xlUnderlineStyleNone
.ColorIndex = 11
End With
ActiveSheet.ChartObjects("Gráfico 9").Activate
ActiveChart.PlotArea.Select
ActiveChart.SeriesCollection(1).XValues = "=PSI!R31C2:R" & 30 + a & "C2"
ActiveChart.SeriesCollection(1).Values = "=PSI!R31C10:R" & 30 + a & "C10"
ActiveChart.SeriesCollection(2).XValues = "=PSI!R31C2:R" & 30 + a & "C2"
ActiveChart.SeriesCollection(2).Values = "=PSI!R31C6:R" & 30 + a & "C6"
ActiveSheet.ChartObjects("Gráfico 10").Activate
ActiveChart.PlotArea.Select
ActiveChart.SeriesCollection(1).XValues = "=PSI!R31C2:R" & 30 + a & "C2"
ActiveChart.SeriesCollection(1).Values = "=PSI!R31C11:R" & 30 + a & "C11"
ActiveChart.SeriesCollection(2).XValues = "=PSI!R31C2:R" & 30 + a & "C2"
ActiveChart.SeriesCollection(2).Values = "=PSI!R31C7:R" & 30 + a & "C7"
ActiveSheet.ChartObjects("Gráfico 11").Activate
ActiveChart.PlotArea.Select
ActiveChart.ChartArea.Select
ActiveChart.SeriesCollection(1).XValues = "=PSI!R31C2:R" & 30 + a & "C2"
ActiveChart.SeriesCollection(1).Values = "=PSI!R31C12:R" & 30 + a & "C12"
ActiveChart.SeriesCollection(2).XValues = "=PSI!R31C2:R" & 30 + a & "C2"
ActiveChart.SeriesCollection(2).Values = "=PSI!R31C8:R" & 30 + a & "C8"
Sheets("PSI").Range("K6").Value = Sheets("ProcedimientoPSI").Range("ah5").Value
Sheets("PSI").Range("K8").Value = Sheets("ProcedimientoPSI").Range("ah6").Value
Sheets("PSI").Range("M6").Value = Sheets("ProcedimientoPSI").Range("ah7").Value
Sheets("PSI").Range("M8").Value = Sheets("ProcedimientoPSI").Range("ah8").Value
Sheets("PSI").Range("O6").Value = Sheets("ProcedimientoPSI").Range("ah22").Value
Sheets("PSI").Range("O8").Value = Sheets("ProcedimientoPSI").Range("ah23").Value
Sheets("PSI").Range("Q6").Value = Sheets("ProcedimientoPSI").Range("ah24").Value
Sheets("PSI").Range("Q8").Value = Sheets("ProcedimientoPSI").Range("ah25").Value

```

```

End Sub
Sub PSI_Percen()
Sheets("ProcedimientoPSI").Visible = True
Percentiles_PSI
PSI
Sheets("ProcedimientoPSI").Visible = False
Sheets("PSI").Select
Range("a1").Select
End Sub

```

```

Sub PSI_REPORTES()
Sheets("Reportes").Visible = True
Sheets("REPORTES").Select
Sheets("PSI").Visible = False
End Sub

```

-----Modulo correspondiente al análisis de características-----

-

```

Sub AC()

Dim x(1 To 100) As String
Dim Etiqueta(1 To 100) As String
Dim V_C(1 To 100) As Double
Dim V_M(1 To 100) As Double
Dim y1(1 To 100) As Double
Dim y2(1 To 100) As Double
Dim z(1 To 100) As String

Sheets("AC_PROCESO").Visible = True
Sheets("Variables").Visible = True
Sheets("Variables").Select
n = Range("f23").Value

```

```

For i = 1 To n
    x(i) = Range("e" & 1 + i)
    y1(i) = Range("d" & 1 + i)
    y2(i) = Int(1 / (1 + Exp(y1(i))) * 1000)
    z(i) = Range("c" & 1 + i)
Next i
Sheets("AC_PROCESO").Select
For i = 1 To n
    If x(i) = "C" Then
        With ActiveSheet.PivotTables("Tabla AC C").PivotFields("x" & i)
            .Orientation = xlRowField
            .Position = 1
        End With
        k = Range("e6").Value
        With ActiveSheet.PivotTables("Tabla AC C").PivotFields("x" & i)
            .Orientation = xlPageField
            .Position = i
        End With
    Else
        k = 1
    End If
    Contador = Contador + k
Next i
Sheets("AC").Select
Rows("14:100").Select
Selection.Delete Shift:=xlUp
For i = 1 To Contador
    Range("C13:G13").Select
    Selection.Copy
    Range("C" & 13 + i, "G" & 13 + i).Select
    Selection.PasteSpecial Paste:=xlPasteFormats, Operation:=xlNone, _
        SkipBlanks:=False, Transpose:=False

```

```

Application.CutCopyMode = False
Range("C" & 13 + i).Select
Next i
Range("C13:G13").Select
Selection.Delete
Sheets("AC_PROCESO").Select
s = 0
Suma = 0
For i = 1 To n
    With ActiveSheet.PivotTables("Tabla AC C").PivotFields("x" & i)
        .Orientation = xlRowField
        .Position = 1
    End With
    With ActiveSheet.PivotTables("Tabla AC M").PivotFields("x" & i)
        .Orientation = xlRowField
        .Position = 1
    End With
    If x(i) = "C" Then
        h = Range("e6").Value
        For j = 1 To h
            Etiqueta(j) = z(i) & " " & Range("b" & 35 + j).Text
            V_C(j) = Range("C" & 35 + j).Value
            V_M(j) = Range("H" & 35 + j).Value
            s = s + 1
            Sheets("AC").Select
            Range("c" & 12 + s) = Etiqueta(j)
            Range("d" & 12 + s) = V_M(j)
            Range("e" & 12 + s) = V_C(j)
            Range("f" & 12 + s) = y2(i)
            Range("g" & 12 + s) = (V_M(j) - V_C(j)) * y2(i)
            Suma = Suma + Range("g" & 12 + s)
            Sheets("AC_PROCESO").Select
        Next j
    End If
Next i

```



```

Else
  For j = 1 To 2
  If Range("b" & 35 + j).Value = 1 Then
    Etiqueta(j) = z(i)
    V_C(j) = Range("C" & 35 + j).Value
    V_M(j) = Range("H" & 35 + j).Value
    s = s + 1
    Sheets("AC").Select
    Range("c" & 12 + s) = Etiqueta(j)
    Range("d" & 12 + s) = V_M(j)
    Range("e" & 12 + s) = V_C(j)
    Range("f" & 12 + s) = y2(i)
    Range("g" & 12 + s) = (V_M(j) - V_C(j)) * y2(i)
    Suma = Suma + Range("g" & 12 + s)
    Sheets("AC_PROCESO").Select
  End If
  Next j

End If

With ActiveSheet.PivotTables("Tabla AC C").PivotFields("x" & i)
  .Orientation = xlPageField
  .Position = 1
End With

With ActiveSheet.PivotTables("Tabla AC M").PivotFields("x" & i)
  .Orientation = xlPageField
  .Position = 1
End With

Next i

Sheets("AC").Select
Range("g" & Contador + 13) = Suma
End Sub

```

## REFERENCIAS BIBLIOGRÁFICAS

- Tomas Lyn, Edelman David and Crook Jonathan, “Credit Scoring and its Applications”, primera edición, SIAM Society for Industrial and Applied Mathematics, Filadelfia, 2002, págs 115 – 120.
- Lawrence David B., “Handbook of Consumer Lending”, primera edición, Prentice Hall Englewood Cliffs, New Jersey, 1992, págs. 98 – 102.
- G.S. Maddala, Introducción a la Econometría, segunda edición, Prentice-Hall Hispanoamericana, México, 1996, págs. 378 - 379

### Papers:

- Hsieh FY , “Sample size tables for logistic regression”, artículo para revista, Inglaterra, 1989.
- Robert B. Avery Gerald A. Hanweck , “A Dynamic Analysis of Bank Failures”, Board of Governors of the Federal Reserve System (U.S.) in its series Research Papers in Banking and Financial Economics with number 74, Estados Unidos, 1984.
- Scalar Consulting, “Aspectos Cuantitativos y Cualitativos de Riesgo Operativo: Prevención del Fraude, de Lavado de Activos y Prácticas de Buen Gobierno Corporativo vinculadas con riesgo operativo de las Instituciones Financieras”, Centro de Capacitación Scalar Consulting Cía. Ltda., Quito, 2007.
- Martha Galicia Romero, “Nuevos Enfoques de Riesgo de Crédito”, Instituto del Riesgo Financiero, México 2003.
- ELIZALDE Abel y REPULLO Rafael, “Capital Regulatorio y Capital Económico Un Análisis de sus Determinantes”, Banco de España, Revista Estabilidad Financiera No. 7, Noviembre 2004 .
- Reyes Samaniego Medina y Martín Marín José Luis, “El método IRB en el Acuerdo de Basilea”, Universidad Pablo de Olavide, Sevilla, 2005.

- BASEL COMMITTEE ON BANKING SUPERVISION, “Credit Risk Modelling: Current Practices and Applications”, Abril 1999.
- Latoli , “Requerimiento Mínimo de Capital: Replanteamientos para mejorar la exigencia de liquidez y solvencia de las instituciones de fianzas”.