

**ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE CIENCIAS**

**EVALUACIÓN DEL RIESGO CREDITICIO EN UNA  
CARTERA DE CONSUMO BAJO DOS SISTEMAS DE  
CLASIFICACIÓN: PROCÍCLICO Y ACICLICO**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERA  
MATEMÁTICA**

**PROYECTO DE INVESTIGACIÓN**

**CAROLINA ELIZABETH YÉPEZ CASTILLO**

**caritoelizabeth\_20@hotmail.com**

**DIRECTORA: ADRIANA UQUILLAS ANDRADE, PHD.**

**adriana.uquillas@epn.edu.ec**

**QUITO, MARZO DE 2019**

## Declaración

Yo, Carolina Elizabeth Yépez Castillo, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado 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 de la Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

.....

Carolina Elizabeth Yépez Castillo

## **Certificación**

Certifico que el presente trabajo fue desarrollado por Carolina Elizabeth Yépez Castillo, bajo mi supervisión.

.....

Adriana Uquillas Andrade, PhD.

Directora

## Agradecimientos

Agradezco infinitamente a Dios por permitirme culminar mi tesis, por darme la sabiduría necesaria para haber terminado mi carrera.

A mis padres Jesús y Jacqueline quienes han sido mi pilar, mi soporte, mi apoyo incondicional en todo este arduo camino. Gracias por sus palabras de apoyo y aliento en los momentos más duros, me ayudaron a no darme por vencida. Quisiera que nunca me falten y sean eternos.

A mi hermana Arelyta que es la felicidad de mi hogar y la que siempre nos une, gracias por todo tu amor y alegría.

A Junior, la persona que ha caminado de mi mano en todo este largo proceso. Gracias por tu paciencia y amor.

De igual manera quisiera agradecer a mi directora de titulación la Dra. Adriana Uquillas quien fue mi guía para realizar esta tesis.

## **Dedicatoria**

El presente trabajo está dedicado a:

Mis queridos padres, quienes se merecen esto y mucho más.

## Índice General

<b>Índice de figuras</b>	<b>X</b>
<b>Índice de tablas</b>	<b>XI</b>
<b>Resumen</b>	<b>XIII</b>
<b>Abstract</b>	<b>XIV</b>
<b>1 Introducción</b>	<b>1</b>
1.1 Justificación del Problema . . . . .	1
1.2 Planteamiento del Problema . . . . .	3
1.3 Objetivos . . . . .	9
1.3.1 Objetivo General . . . . .	9
1.3.2 Objetivos Específicos . . . . .	9
<b>2 Marco Teórico</b>	<b>10</b>
2.1 Definiciones y conceptos básicos del sistema financiero . . . . .	10
2.1.1 ¿Qué es un banco? . . . . .	10
2.1.2 ¿Para qué sirve un banco? . . . . .	10
2.2 Crédito . . . . .	11
2.2.1 Las 5 C's del Crédito . . . . .	11
2.2.2 Tipos de Crédito . . . . .	12
2.3 Riesgos Financieros . . . . .	13

2.3.1	Riesgo de Crédito . . . . .	13
2.3.2	Riesgo Operativo . . . . .	14
2.3.3	Riesgo de Liquidez . . . . .	14
2.3.4	Riesgo de Mercado . . . . .	14
2.4	Gestión de Riesgos en el Sistema Financiero . . . . .	14
2.4.1	Comité de Basilea . . . . .	16
2.5	Regulaciones de Basilea . . . . .	16
2.5.1	Análisis de Basilea I . . . . .	17
2.5.2	Análisis de Basilea II . . . . .	19
2.6	Filosofías de clasificación de riesgos . . . . .	23
2.6.1	Definición y Comportamientos Esperados . . . . .	24
2.6.2	Filosofía de Clasificación TTC . . . . .	25
2.6.3	Filosofía de Clasificación PIT . . . . .	26
2.6.4	Diferencias entre PIT y TTC . . . . .	26
2.6.5	Evaluación de Desempeño PIT y TTC . . . . .	26
2.7	Dinámica de los sistemas de clasificación . . . . .	27
2.7.1	Grupos de Riesgo y Estabilidad . . . . .	28
2.7.2	Dinámica de las PD agrupadas . . . . .	29
2.7.3	Cuantificación de las PD agrupadas . . . . .	30
2.8	Modelo Credit Scoring . . . . .	31

2.8.1	Ventajas y Desventajas de las distintas Metodologías del Credit Scoring . . . . .	31
2.9	Metodología Estadística . . . . .	33
2.9.1	Metodología para el Desarrollo del Modelo de Credit Scoring . . . . .	33
2.9.2	Metodología para la Creación de los Grupos de Riesgo . . . . .	41
<b>3</b>	<b>Desarrollo del Modelo Estadístico</b>	<b>46</b>
3.1	Descripción de la base de datos . . . . .	46
3.1.1	Estructura de la base de datos . . . . .	46
3.1.2	Adición de Variables . . . . .	47
3.1.3	Madurez de la Cartera . . . . .	47
3.2	Definición del Default . . . . .	48
3.2.1	Status de los Clientes en los Pagos. . . . .	48
3.2.2	Definición de buenos y malos clientes . . . . .	49
3.2.3	Análisis Roll Rate . . . . .	50
3.3	Definición de la Variable Dependiente . . . . .	53
3.3.1	Comportamiento de la variable dependiente (PD) . . . . .	54
3.4	Tratamiento de la base de datos . . . . .	54
3.5	Categorización de las Variables . . . . .	55
3.5.1	Variables Cualitativas . . . . .	55
3.5.2	Variables Cuantitativas . . . . .	56
3.6	Selección de las Variables . . . . .	56



3.6.1	Information Value . . . . .	56
3.6.2	Estadística KS . . . . .	57
3.6.3	Multicolinealidad . . . . .	57
3.7	Descripción de variables seleccionadas . . . . .	58
3.7.1	Variables Independientes . . . . .	58
3.8	Creación del Modelo Scoring . . . . .	59
3.8.1	Modelo PIT sin informaciones sistémicas puras . . . . .	59
3.8.2	Interpretación del Modelo PIT sin informaciones sistémicas puras . . . . .	63
3.8.3	Modelo TTC . . . . .	64
3.8.4	Interpretación del Modelo TTC . . . . .	66
3.8.5	Modelo PIT con informaciones sistémicas puras . . . . .	66
3.8.6	Interpretación del Modelo PIT con informaciones sistémicas puras . . . . .	70
3.9	Resultados de los Modelos . . . . .	71
3.10	Validación del Modelo . . . . .	72
3.10.1	Poder de Discriminación de los Modelos . . . . .	72
3.10.2	Precisión y Error de los Modelos . . . . .	73
<b>4</b>	<b>Creación de Grupos de Riesgo</b> . . . . .	<b>76</b>
4.1	Generación de los grupos homogéneos de riesgo . . . . .	76
4.1.1	Grupos de Riesgo: Modelo PIT . . . . .	76
4.1.2	Grupos de Riesgo: Modelo TTC . . . . .	78
4.2	Validación de los grupos de riesgo . . . . .	80

4.2.1	Pruebas de Comparaciones Múltiples de Diferencias de Medias .	80
4.3	Estabilidad de los grupos de riesgo . . . . .	83
4.3.1	Modelo PIT . . . . .	83
4.3.2	Modelo TTC . . . . .	84
<b>5</b>	<b>Conclusiones y recomendaciones</b>	<b>86</b>
5.1	Conclusiones . . . . .	86
5.2	Recomendaciones . . . . .	89
<b>6</b>	<b>Referencias</b>	<b>115</b>

**Índice de figuras**

1	Filosofías de Clasificación . . . . .	25
2	Estructura de la base de datos . . . . .	46
3	Estado de los Clientes . . . . .	50
4	Porcentaje de Clientes Buenos, Malos e Indeterminados . . . . .	53
5	Tasa de Mora observada a lo largo del tiempo . . . . .	54
6	Tasa de Mora en los distintos modelos . . . . .	72
7	Grupos de Riesgo del Modelo PIT . . . . .	78
8	Grupos de Riesgo del Modelo TTC . . . . .	80
9	Porcentaje de clientes entre los grupos de riesgo: PIT . . . . .	84
10	Porcentaje de clientes entre los grupos de riesgo: TTC . . . . .	85

## Índice de tablas

1	Diferencias del sistema PIT y TTC . . . . .	26
2	Correlación de PDs agrupados con las fluctuaciones de la economía. . .	30
3	Regla para el Information Value . . . . .	38
4	Tipos de categorías . . . . .	49
5	Tabla de Probabilidades . . . . .	51
6	Tabla Roll Rate . . . . .	52
7	Resultados del Modelo PIT sin informaciones sistémicas puras . . . . .	62
8	VIF - Modelo PIT sin informaciones sistémicas puras . . . . .	63
9	Resultados del Modelo TTC . . . . .	65
10	VIF - Modelo TTC . . . . .	66
11	Resultados del Modelo PIT con informaciones sistémicas puras . . . . .	69
12	VIF - Modelo PIT con informaciones sistémicas puras . . . . .	70
13	Estadística KS en el conjunto de Prueba y de Validación. . . . .	72
14	AUC en el conjunto de Prueba, y de Validación . . . . .	73
15	Resultados de los Modelos - SIN INDETERMINADOS . . . . .	74
16	Resultados de los Modelos - CON INDETERMINADOS . . . . .	75
17	Resumen de los grupos del modelo PIT . . . . .	77
18	Resumen de los grupos del modelo TTC . . . . .	79
19	Prueba de Tukey para los grupos del modelo PIT . . . . .	81
20	Prueba de Tukey para los grupos del modelo TTC . . . . .	82

21	Pueba de Dunnet T3 para los grupos del modelo PIT . . . . .	82
22	Pueba de Dunnet T3 para los grupos del modelo TTC . . . . .	83

## Resumen

Debido a las crisis financieras que se han presentado en la banca y considerando la prociclicidad que existe en el sistema financiero, el Comité de Supervisión Bancaria de Basilea ha planteado distintos sistemas de clasificación con el fin de evaluar el riesgo de crédito en carteras de créditos ya concedidos, y observar el comportamiento de pago de los prestatarios.

El presente trabajo trata sobre la modelización del riesgo de crédito en una cartera de consumo considerando dos enfoques de clasificación en términos de prociclicidad: PIT (Punto en el tiempo) y TTC (A través del ciclo).

Con este fin, como primer paso se estimó un modelo de elección binaria logit para obtener la probabilidad de incumplimiento (PD) de una cartera de créditos concedidos de consumo, la cual fue proporcionada por una entidad financiera de Guatemala. Para este análisis se han considerado 32585 individuos y se han tratado alrededor de 219 variables explicativas en los periodos desde Julio/2015 hasta Noviembre/2015; lo cual implicó una gran dificultad para modelar la PD con tan pocos periodos de tiempo. Luego para abordar la estabilidad del sistema de clasificación se crearon grupos homogéneos de riesgo, con el fin de observar cambios o migraciones de los clientes en cada sistema de clasificación durante las crisis económicas.

Los resultados de la investigación muestran que el sistema clasificación PIT tiene mejor poder predictivo en comparación con el sistema de clasificación TTC. Los cuales pueden ser usados como una herramienta para tener una adecuada gestión del riesgo crediticio en futuras renovaciones de crédito. Además, al crear los grupos de riesgo y observar la estabilidad de los sistemas de clasificación, se concluye que la filosofía TTC es más estable que la filosofía PIT, es decir, existe menor migración de los deudores en los grupos de riesgo a través del tiempo.

**Palabras clave:** prociclicidad, Comité de Supervisión Bancaria de Basilea, cartera de consumo, PIT, TTC, logit, PD, grupos de riesgo.

## Abstract

Due to financial crisis that have occurred in banking, and considering the procyclicality that exists in the financial system, the Basel Committee on Banking Supervision has proposed different rating systems in order to assess credit risk for the loan portfolio and observe the payment behavior of the obligors.

This work deals with the credit risk modeling of in a personal lending, considering two rating approaches in terms of procyclicality: PIT (Point in time) and TTC (Through the cycle).

For this purpose, as a first step a logit model was estimated to obtain the default probability (PD) of a personal loan portfolio, and the data was provided by a financial institution from Guatemala. For this analysis 32585 individuals have been considered and 219 explanatory variables have been treated, between the analysis periods from July/2015 to November/2015; which implied a great difficulty to model the probability of default (PD) having so few periods of time. Then in order to address the issue of rating stability, homogeneous risk buckets were created to observe changes or migrations of obligors in each rating system during economic crisis.

The research results show that the PIT rating system has greater robustness than TTC rating system, and has better predictive power than TTC rating system. These results can be used as a tool to have an adequate credit risk management in future credit renewals in the financial institution. In addition, when creating the risk buckets and observe the stability of the rating systems, it can be concluded that the TTC rating is more stable than the PIT rating, that is, there is less migration of the obligors in risk buckets over time.

**Palabras clave:** procyclicality, Basel Committee on Banking Supervision, loan portfolio, PIT, TTC, logit, PD, risk buckets.

# Capítulo 1

## Introducción

### 1.1 Justificación del Problema

Durante las décadas de 1980 y 1990 del siglo pasado se produjeron diversas crisis en las instituciones financieras, tanto de países desarrollados como en desarrollo; pero, los efectos fueron más devastadores en los del segundo grupo, afectando tanto el PIB como la inversión estatal en educación y salud. Guatemala, al pertenecer al segundo grupo, vio su economía severamente afectada (Mendoza, 2010).

Ante la situación antes descrita, con un escenario mundial cuyas finanzas podrían provocar agudas crisis nacionales y regionales, el Comité Basilea <sup>1</sup> se reunió en 2004 para emitir las recomendaciones generales para los organismos de control financiero en los diferentes países, buscando una gestión financiera eficaz. (Comité de Basilea, 2006)

Por lo tanto, se introdujeron nuevas directrices de cumplimiento, como los acuerdos de Basilea y de solvencia propuestos por la Supervisión Bancaria de Basilea (Dirick, Baesensand y Claeskens, 2016). Entre los acuerdos introducidos junto con la metodología que los acompaña, están los sistemas de clasificación de riesgo PIT (Point in time) y TTC (Through the cycle).

Los Principios Básicos de Basilea <sup>2</sup> constan de 25 principios fundamentales para la supervisión y regulación bancaria adecuada de las instituciones financieras dentro de cualquier país.

Este Comité trabaja indirectamente para reactivar la confianza de los clientes bancarios, por lo que, su primera propuesta es proteger los depósitos con el capital de las

---

<sup>1</sup>El Comité de Supervisión Bancaria de Basilea, creado en 1975 por los Gobernadores de los bancos centrales del G-10, está compuesto por altos representantes de autoridades de supervisión bancaria y de bancos centrales de: Alemania, Bélgica, Canadá, España, Estados Unidos, Francia, Italia, Japón, Luxemburgo, Suecia, Suiza, los Países Bajos y el Reino Unido.

<sup>2</sup>Los Principios Básicos se entienden como un marco voluntario de normas mínimas sobre mejores prácticas de supervisión bancaria.



instituciones financieras. También plantea que los supervisores bancarios propongan un mínimo de capital propio para afrontar riesgos y pérdidas. (Comité de Basilea, 2006)

Además, exige que el capital sea calculado para cubrir los riesgos crediticios de la institución financiera, y que podrían manejarse dos estrategias:

1. Impartiendo una normativa general, y,
2. Que el supervisor apruebe la utilización de sistemas internos de calificación (Comité de Basilea, 2006)

De esta forma, las instituciones financieras pueden construir modelos de clasificación interna que permitan evaluar el riesgo de crédito de un cliente, calificándolo y manteniendo ciertos niveles de aprobación, con la mayor precisión posible.

Un estudio erróneo del cliente puede implicar tanto el sobreendeudamiento del mismo, o también adquirir una provisión superior que prevenga las posibles fluctuaciones de la economía del país. Estos motivos mencionados y entre otros, son los que pueden conllevar a una cartera morosa, impidiendo seguir con el giro normal de la institución financiera (L. Bonilla, 2008).

El crecimiento de las operaciones y los errores en los análisis de los clientes, puede conllevar a tener pérdidas y malas provisiones en la cartera de una institución financiera. Además, con el fin de tecnificar el análisis al momento de la otorgación de un crédito, se trabajará juntamente con una institución financiera a fin de establecer mejores prácticas, disminuyendo el riesgo de crédito y cumpliendo con los requerimientos de Basilea II en cuanto al cálculo mínimo de capital regulatorio.

Teniendo en cuenta este requerimiento, se propone crear un modelo de scoring de calificación de crédito, con el fin de determinar la probabilidad de incumplimiento PD de los clientes. Cuando se hace la comparación del comportamiento de la tasa de mora observada y la PD estimada durante un auge o una recesión económica, se pueden observar distintas propiedades. Es decir, durante una recesión económica, algo común que puede ocurrir, es que la banca tenga un alto número de incumplimientos por parte de los clientes, mientras que en una expansión económica los incumplimientos se redu-

cirán. Es por eso que se menciona que un cambio en las condiciones macroeconómicas de un país afectará a la PD.

Se propone desarrollar dos sistemas de clasificación de riesgo PIT y TTC, para analizar cómo es el comportamiento de la PD, y observar el cambio en la predicción de PD bajo los dos sistemas de clasificación.

El sistema de clasificación PIT (Punto en el tiempo) es un sistema procíclico, donde la clasificación de riesgo de los clientes cambia de acuerdo a épocas de expansión y recesión de la economía, y de cambios en la política de crédito.

El sistema de clasificación TTC (A través del ciclo) es un sistema donde la clasificación de los clientes en su mayoría no se ve afectado por las condiciones macroeconómicas del país.

Estos dos sistemas de clasificación de riesgo PIT y TTC usan características idiosincráticas del cliente, características sistémicas e inclusive características macroeconómicas del país.

En la práctica no existen tantos estudios acerca de este tema y, por ende, no existe una definición fija o una regla que describa las características de la filosofía PIT o TTC.

Sin embargo, basándose en Basilea (Basel Committee on Banking Supervision, 2005) se puede decir que el sistema PIT usa toda la información disponible de un cliente y, además, información externa. Aquí la PD puede cambiar a medida que la economía del país mejore; mientras que el sistema TTC usa solo información disponible del cliente y la PD se mantiene constante a lo largo de las fluctuaciones económicas.

Este proyecto se centrará en comparar estos dos sistemas de clasificación de riesgo, determinando si la PD sigue una filosofía PIT o TTC, para luego asignarlas a grupos de riesgo.

## **1.2 Planteamiento del Problema**

El Comité de Supervisión Bancaria de Basilea es una organización responsable de mantener la solidez en el sistema financiero y también se encarga de plantear los requisitos mínimos de capital para bancos internacionales. Pero cabe mencionar que las reco-

mendaciones planteadas por este comité también son adoptadas por países de todo el mundo, aunque no pertenezcan a los 13 países <sup>3</sup> que constituyen el comité (Allen, DeLong y Saunders, 2004). Este comité propuso acuerdos de capital también conocido como Basilea II, que propone determinar requisitos de capital para afrontar el riesgo de crédito en una entidad bancaria.

Es por eso que con la introducción de Basilea II, bajo el enfoque de calificaciones internas, las instituciones financieras tienen la facultad de calificar el riesgo de sus clientes mediante modelos desarrollados internamente cumpliendo con su capital mínimo regulatorio, es decir, proponer una metodología para validar los sistemas de clasificación internos y así determinar la probabilidad de incumplimiento (PD) de un cliente.

Es fundamental realizar un análisis de los requerimientos estadísticos del comité de Basilea para poder desarrollar de forma adecuada el enfoque de este proyecto, y para entender las prácticas y normativas de supervisión bancaria.

Por otro lado, los componentes de riesgo que son: *PD (probabilidad de incumplimiento)*, *LGD (pérdida dada la morosidad)*, *EAD (exposición por al riesgo)* son considerados factores de vital importancia para el cálculo del capital regulatorio en una institución financiera. (Basel Committee on Banking Supervision, 2005)

Basilea II considera las diversas probabilidades de incumplimiento de los distintos clientes que dependen tanto del tipo de cliente, de las propiedades dinámicas de la metodología que se use, así como de la cantidad de datos que disponga la entidad bancaria, entre otros. Para ello, Basilea II se basa en un conjunto de requerimientos, apoyadas en tres pilares: (Trías y March, 2016)

- **Pilar I.** Cálculo de los requisitos mínimos de capital;
- **Pilar II.** Proceso de supervisión de la gestión de los fondos propios; y,
- **Pilar III.** Disciplina de mercado.

Para cumplir con los 3 pilares, la institución financiera debe evaluar el riesgo de crédito teniendo en cuenta la calidad crediticia de los clientes (*utilizando ratings externos o*

---

<sup>3</sup>Alemania, Bélgica, Canadá, España, Estados Unidos, Francia, Italia, Japón, Luxemburgo, Suecia, Suiza, los Países Bajos y el Reino Unido

*internos*) para la posterior estimación de los componentes de riesgo PD, LGD, y EAD. Además, se añaden verificaciones de requisitos de capital. Los modelos de clasificación de una institución financiera buscan identificar y combinar los factores (variables) que permitan establecer lo óptimo en términos de riesgo. Para ello, se deben estimar los componentes de riesgo PD, LGD y EAD, desarrollando un sistema de clasificación crediticio.

Esto es necesario para el cálculo posterior del mínimo capital regulatorio, el cual busca guardar una determinada porción del capital social de una empresa como garantía de solvencia frente a pérdidas inesperadas. Es decir, cuando los clientes no pagan los créditos dentro de un periodo de tiempo. Esto se denomina default o incumplimiento. Este tipo de riesgo se denomina riesgo de crédito (Ramon, 2014).

En la actualidad, para observar el comportamiento de pago del cliente existen varios modelos que nos permiten estimar la PD, uno de los más utilizados es la regresión logística, el cual es un proceso binomial es decir que puede tener solo dos posibles resultados: “éxito: 1” o “fracaso: 0”, tal que la probabilidad de cada uno de ellos es constante en una serie de repeticiones (Lemeshow y Hosmer, 2005).

Este proyecto aplicará este enfoque, con el propósito de estimar la probabilidad de que un individuo incumpla en el pago de su cuota, dado un conjunto de variables  $X$ , las cuales son características de los clientes que indican la solvencia de los mismos, información propia del cliente, información macroeconómica del país, o información de políticas de crédito. Por tal razón, se debe hacer una cuidadosa selección previa de las variables pertinentes.

Para estimar esa probabilidad de incumplimiento (PD), se lo realiza estableciendo una variable binaria; esto es, se observa el comportamiento de pago del individuo y se lo clasifica con una marca de default que puede tomar valor 0, si no hay incumplimiento, o 1 en caso de incumplimiento. En general, dichos modelos son calibrados en un momento, y se realizan seguimientos periódicos.

Ahora, surge la duda de: *¿cómo se comportará la probabilidad de incumplimiento PD durante un auge económico, o en épocas de recesión, o en cuando existan cambios en las políticas de crédito?*

Bessis (1998) menciona que, durante un auge económico o un mejoramiento en las condiciones sistémicas, es de esperarse un mejoramiento generalizado de la capacidad de pago de los individuos. Mientras que en épocas de recesión o de empeoramiento de las condiciones del sistema financiero, se espera que esa capacidad se deteriore.

Por esa razón surge la pregunta:

*¿Cómo contribuirán las condiciones sistémicas con modificaciones en las probabilidades de incumplimiento estimadas, durante una recesión o un auge económico?*

Para responder a las preguntas mencionadas, se analizarán las distintas filosofías de clasificación de riesgo las cuales consideran tanto las características sistémicas (*variables sistémicas*)<sup>4</sup>, como también las características propias del cliente (*variables idiosincráticas*)<sup>5</sup>

Según Basilea, para calificar a los clientes y posteriormente asignarlos a los grupos de riesgo, la institución financiera puede basarse en una de las siguientes filosofías o clasificaciones de riesgo: PIT (Point in the time) o TTC (Through the cycle); su diferencia radica en como los clientes son clasificados y además en el tipo de información que usa para clasificar a los clientes (Bassel Commmite on Banking Supervision, 2005), o sea, en las características específicas de las PDs que posee cada cliente.

El sistema de clasificación **PIT** ("Punto en el tiempo") usa toda la información disponible del cliente, es decir toda la información agregada y específica del cliente para asignarlos a los grupos de riesgo (Bassel Commmite on Banking Supervision, 2004). Por tanto, el PIT utiliza información sistémica e idiosincrática, y aquí el comportamiento de la PD sube en recesión y cae en periodos de expansión, en otras palabras, acompañan las fluctuaciones macroeconómicas, con lo cual, las calificaciones PIT tenderán a ajustarse rápidamente a un entorno económico cambiante (Uquillas, 2018).

El sistema de clasificación **TTC** ("A través del ciclo") utiliza características estáticas y dinámicas del cliente, es decir usa principalmente información idiosincrática del clien-

---

<sup>4</sup>Las variables sistémicas son características de la macroeconomía del país, y características de las políticas de crédito que establece una institución financiera

<sup>5</sup>Las variables idiosincráticas son características reales y que son únicas para cada individuo.

te. Además, el comportamiento de la PD no necesariamente acompaña a las fluctuaciones de la economía, ni de políticas de crédito.

El sistema de clasificación TTC tiende a mantenerse más o menos constante aun cuando las condiciones macroeconómicas del país cambien con el tiempo. En cambio, el sistema de clasificación PIT tiende a adaptarse a un entorno económico variable. Entre estos dos casos extremos se encuentran los sistemas de clasificación llamados **Híbridos**, que poseen características de las filosofías de clasificación PIT y TTC (Bassel Commmite on Banking Supervision, 2005).

Con el fin de comparar las dos filosofías antes mencionadas, tanto en ordenamiento como en predicción de la PD, se propone la formulación de varios modelos considerando el tipo de información que usa cada sistema de clasificación.

Para poder evaluar con precisión las PD hay que considerar el tipo de sistema de clasificación en el que se basa la institución financiera.

Y además se debe tomar en cuenta que los modelos de clasificación interna de una institución financiera deben ser capaces de:

- Ordenar el riesgo.
- Cuantificar el riesgo.

Por otro lado, según Basilea II, una vez obtenidas las PD estimadas con las distintas filosofías de clasificación, la institución financiera deberá asignar a los clientes a grupos de riesgo homogéneos <sup>6</sup>, donde los clientes que se encuentran en un mismo grupo deben poseer la misma calidad de crédito. Luego que cada cliente este asignado a un grupo, se debe calcular una PD del grupo (PD agrupada) para cada distinto grupo de riesgo (Bassel Commmite on Banking Supervision, 2005). Además, se debe demostrar que los grupos de riesgo son válidos mediante pruebas estadísticas de diferencia de medias.

Por esta razón, después de la creación del modelo de clasificación, se generarán grupos homogéneos de riesgo ordenados con relación al riesgo de crédito, los cuales están

---

<sup>6</sup>Un grupo de riesgo homogéneo es un grupo caracterizado por estar dotado de la misma probabilidad de incumplimiento (Vilariño, 2010).

destinados a distinguir entre clientes de diferente calidad crediticia atribuyendo una PD distinta a cada grupo.

Se usará un enfoque histórico para cuantificar las PD agrupadas, es decir, se hará uso de datos históricos, considerando la experiencia de incumplimiento de los clientes en períodos pasados. Así finalmente la PD agrupada de cada grupo se obtendrá calculando el promedio de la tasa de mora de los clientes, esto es, la frecuencia de incumplimientos observados en los clientes pertenecientes a un determinado grupo de riesgo.

El objetivo de asignar a los clientes a estos grupos de riesgo es observar y predecir su comportamiento en ocasiones futuras. Dependiendo de la filosofía de clasificación que se use, se podrá determinar si durante una desaceleración económica, o una expansión económica, o cambios en las políticas de crédito, el cliente tenderá a hacer una transición hacia otros grupos de riesgo, o tenderá a permanecer estable en su grupo ya establecido.

Por esto, la información del cliente que se use para crear cada grupo homogéneo es un factor clave para construir un sistema de clasificación de riesgo.

En este proyecto se usará una base de datos de clientes de una institución financiera de Guatemala, que es una cartera de créditos ya concedidos de consumo. El período considerado para el modelado fue entre julio del 2015 y noviembre del 2015, las cuales son las fechas de referencia.

Esta base proporcionada por la institución financiera consta de aproximadamente 219 variables, entre las cuales hay variables que describen información específica de cada cliente, es decir, que es propia de un cliente como la edad, género, salario entre otros. A su vez existen variables propias de la institución financiera como la tasa de interés, número de cuotas en contrato, cantidad de productos de crédito, etc. Todas estas variables son informaciones históricas de las variables explicativas que se las usará para pronosticar.

Para la marcación de mora del cliente, se tiene dos variables que son: los días en atraso en los próximos 3 meses y en los próximos 6 meses a partir de la fecha de referencia.

También hay información agregada de cada cliente, por lo cual se incluyó variables

macroeconómicas del país de Guatemala como el IMAE <sup>7</sup>, el IPC <sup>8</sup> y la inflación de Guatemala en los periodos en consideración.

### 1.3 Objetivos

#### 1.3.1 Objetivo General

Establecer las diferencias entre los sistemas de clasificación PIT y TTC en la evaluación de riesgo de crédito en el producto de consumo en una entidad financiera, mediante un modelo de score.

#### 1.3.2 Objetivos Específicos

- Determinar el componente de riesgo PD (probabilidad de incumplimiento) de la institución financiera realizando una validación del mismo.
- Analizar la viabilidad de la metodología de regresión logística en la evaluación de riesgo de crédito.
- Sustentar la creación de grupos homogéneos de riesgo tras la obtención de predicciones de cada modelo.
- Analizar los requerimientos estadísticos de Basilea que están relacionados con la atribución de la PD y con los sistemas de calificación de riesgos.

---

<sup>7</sup>El Índice Mensual de Actividad Económica (IMAE) es un indicador que permite seguir el pulso a la actividad económica del año mientras se concreta la cifra de la producción anual (PIB).

<sup>8</sup>El Índice de Precios al Consumidor (IPC) es una variable que tiene por objeto evaluar a través del tiempo la variación de los precios de los artículos de consumo y servicio en Guatemala.



## Capítulo 2

### Marco Teórico

#### 2.1 Definiciones y conceptos básicos del sistema financiero

##### 2.1.1 ¿Qué es un banco?

Un banco según la Corporación Financiera Nacional (2016):

“Es una institución financiera que administra los recursos de sus accionistas y el de sus clientes, y que utiliza esos recursos para prestar a otras personas o empresas cobrándoles un interés.”

En otras palabras, un banco se considera una entidad de carácter privado o público cuyas operaciones consisten en conceder créditos y captar recursos mediante depósitos, lo cual contribuye al buen funcionamiento y desarrollo de la economía de un país (Gobat, 2012).

En base a lo descrito, los bancos son instituciones financieras, de ahorro y préstamos, que ofrecen un lugar seguro para manejar el dinero. Sin embargo, la minuciosa tarea de otorgar créditos requiere de un sistema de análisis para evitar el incumplimiento de los créditos concedidos.

##### 2.1.2 ¿Para qué sirve un banco?

La importancia de un banco parte de su papel en la intermediación de los fondos prestables. Según Merton (1993):

“Un sistema financiero bien desarrollado y saludable facilita la *alocación*<sup>9</sup> eficiente de consumo de los individuos y del capital físico de las empresas en usos más productivos.”

La teoría actual de la banca, según Fabozzi, Modigliani y Ferri (1996), clasifica las funciones bajo cuatro categorías:

---

<sup>9</sup>Alocación: Distribución o colocación.

1. Intermediación bancaria.
2. Facilitar sistemas de pago.
3. Gestionar riesgos.
4. Financiar proyectos de inversión.

Es necesario destacar que no todos los bancos cumplen con estas cuatro funciones, existen bancos especializados y más pequeños que tienen funciones determinadas.

Por otro lado, gracias al desarrollo del mercado de créditos, un país podría evolucionar mucho más rápido. Es decir, el hecho de ofertar créditos por parte de la institución financiera contribuirá al desarrollo económico del cliente y por ende del país (Beckman, 1949).

## 2.2 Crédito

Según la Superintendencia de Bancos Guatemala (2017) los créditos son aquellos activos crediticios, en el cual el acreditante (entidad financiera) se obliga a poner una suma de dinero a disposición del acreditado (cliente), quien deberá restituir las sumas de que disponga o a proveer las cantidades pagaderas por su cuenta, y a pagar los gastos, comisiones e intereses que resulten a su cargo.

La concesión de créditos a una persona natural mediante la entidad bancaria es lo que genera un crecimiento en la actividad económica, ya que se le otorga recursos financieros a un determinado cliente para que pueda cubrir sus necesidades ya sean eventuales o permanentes; cobrando una tasa de interés por el monto (Gobat, 2012).

Es por eso que esta actividad es considerada de bastante riesgo, pues la institución financiera está expuesta a que el cliente no cumpla con sus obligaciones de pago total o parcial en la devolución del crédito; a este riesgo se lo denomina *riesgo de crédito*.

### 2.2.1 Las 5 C's del Crédito

Para evaluar el incumplimiento de los pagos, los bancos, en términos muy generales,

pueden adoptar el criterio de las cinco C's del crédito. En estas variables se encuentra la información básica de un individuo que permitirá estimar la insolvencia de ese posible cliente.

1. **Carácter:** disposición a cumplir con las obligaciones crediticias contraídas.
2. **Capacidad:** nivel de solvencia del cliente para cubrir sus obligaciones de crédito por medio de sus flujos operacionales.
3. **Capital:** nivel de reservas monetarias del cliente.
4. **Colateral:** hipotecas entregadas en caso de incumplimiento.
5. **Condiciones:** ambiente económico en el negocio del cliente.

Para la concesión de un crédito, un banco podría realizar un análisis previo de las cinco C's. Por ejemplo, podría usar la información disponible de cada cliente y asignarle un valor del 1 al 10 a cada una de las C's y luego sumar los resultados, en caso de que alcance una puntuación de 30 se concede el crédito (Ross, Jaffe y Westerfield, 2000).

### 2.2.2 Tipos de Crédito

Según la Superintendencia de Bancos Guatemala (2017), existen varios tipos de créditos entre los cuales están los siguientes:

- **Empresarial Mayor:** Son créditos que pueden ser concedidos a una persona natural o jurídica, que sirven para financiar actividades productivas y de comercialización ya sea de grandes empresas o de instituciones públicas.
- **Empresarial Menor:** Son créditos que pueden ser concedidos a una persona natural o jurídica. Sirven para financiar actividades productivas y de comercialización, los cuales son concedidos a pequeñas o medianas empresas.
- **Consumo:** Son créditos que son concedidos a una persona natural, para que financien la adquisición de bienes de consumo o para realizar algún pago de servicios. No tienen que estar relacionados con ninguna actividad empresarial.

- **Microcrédito:** Estos créditos son concedidos en su mayoría a personas que poseen un negocio propio o familiar. Sirven para financiar actividades productivas y de comercialización de bienes y/o servicios. En el país de estudio *Guatemala*, estos créditos no superan USD\$20,800.00.
- **Hipotecario para vivienda:** Son créditos que están destinados a financiar la compra, construcción, remodelación, o reparación de inmuebles para personas individuales teniendo como garantía un bien inmueble hipotecado a nombre de la institución financiera (Superintendencia de Bancos Guatemala, 2017).

## 2.3 Riesgos Financieros

Los riesgos financieros según Lizarzaburu, Berggrun y Quispe, (pág. 97):

“...son contingencias que están asociados a los productos y servicios que se pueden adquirir en los mercados financieros, de donde las empresas obtienen los recursos o los colocan.”

Además, el riesgo financiero se puede entender como la incertidumbre vinculada a una posición financiera futura. En un escenario sin contingencia las decisiones financieras resultan triviales, ya que la probabilidad de ocurrencia de los eventos sería del 100%. Sin embargo, en la realidad existen muchos sucesos que promueven un escenario de contingencia, en estos casos, es útil construir una lista de eventos y asignarles una probabilidad de ocurrencia (Arias, Rave y Castaño, 2006).

Los riesgos financieros en el sistema bancario más estudiados son: el riesgo de crédito y el riesgo de mercado.

A continuación se mencionan y definen los tipos de riesgos más relevantes en el sistema financiero (De Lara Haro, 2005):

### 2.3.1 Riesgo de Crédito

Es la contingencia de que una institución incurra en pérdidas como consecuencia de que un deudor o contraparte incumpla sus obligaciones en los términos acordados.

Incluye, entre otros, el riesgo cambiario crediticio (Superintendencia de Bancos Guatemala, 2017).

El riesgo de crédito se refiere a lo relacionado con los créditos concedidos por la institución financiera como el no pago o el cambio en la calificación de un cliente.

### **2.3.2 Riesgo Operativo**

El riesgo operativo se relaciona con las fallas en los procesos y sistemas de la organización financiera, así como en la falta de capacitación del personal que pudieran afectar la rentabilidad esperada de la institución financiera.

Es la contingencia de pérdida que una institución tenga como consecuencia de daños, interrupción, alteración, o fallas derivados del uso o dependencia en la infraestructura de la plataforma informática (Superintendencia de Bancos Guatemala, 2017).

### **2.3.3 Riesgo de Liquidez**

Es la contingencia de que una institución no tenga capacidad para fondar incrementos en sus activos o cumplir con sus obligaciones oportunamente, sin incurrir en costos financieros fuera del mercado (Superintendencia de Bancos Guatemala, 2017).

### **2.3.4 Riesgo de Mercado**

Es la contingencia de que una institución incurra en pérdidas como consecuencia de movimientos adversos en precios en los mercados financieros. Incluye, entre otros, los riesgos de tasa de interés y cambiario o de tipo de cambio (Superintendencia de Bancos Guatemala, 2017).

## **2.4 Gestión de Riesgos en el Sistema Financiero**

Una vez que se ha definido el concepto de riesgo financiero es necesario mencionar las características de la gestión de riesgos en el sistema financiero. Según De Lara Haro (2005), la gestión de riesgos:

“...es en esencia un método racional y sistemático para entender los riesgos, medirlos y controlarlos en un entorno en el que prevalecen los instrumentos financieros sofisti-

cados, mercados financieros que se mueven con gran rapidez y avances tecnológicos en los sistemas de información que marcan nuestra era.” (pág 11)

Además, la gestión de riesgos consiste en controlar los riesgos a los que se enfrentan las instituciones financieras.

Y se puede decir que la única manera de mitigar por completo un riesgo es que no exista, por lo tanto, es necesaria la gestión de riesgos financieros para identificar los factores de riesgo que puedan afectar al nivel de rendimiento esperado (Lizarzaburu y cols., 2012).

Por lo que, para una correcta gestión de riesgos financieros, es útil considerar el origen y el nivel de exposición que se tiene ante los diferentes tipos de riesgos y así poder realizar acciones con el fin de mitigarlos (Arias y cols., 2006).

El proceso de gestión de riesgos implica:

1. Identificar los riesgos.
2. Cuantificar y controlar estableciendo niveles de tolerancia.
3. Modificar o mitigar la exposición a los riesgos o planificar coberturas.

En primer lugar, al momento de identificar los riesgos es necesario conocer su naturaleza ya que, como se mencionó, existen muchos tipos de riesgos, como el riesgo de crédito, de mercado, de liquidez, legal, operativo y de reputación. Cada uno de estos riesgos financieros provoca distintos efectos y, además, se correlacionan entre sí.

En segundo lugar, la cuantificación de los riesgos financieros se ha desarrollado sobre todo para los riesgos de mercado, para lo cual se han desarrollado herramientas como el valor en riesgo (VaR), análisis de sensibilidad, beta, delta, etc. La cuantificación del riesgo de crédito consiste en estimar la probabilidad de impago de un cliente, de ahí que surgen muchos modelos de score para clasificar a los clientes (De Lara Haro, 2005).

Con el fin de lograr una adecuada gestión de riesgos y, por ende, estabilidad en el sistema financiero se creó un fórum internacional dedicado a la promoción de contactos, discusiones e intercambio entre los bancos centrales llamado *Comité de Supervisión Bancaria de Basilea* (o simplemente Comité de Basilea).

### 2.4.1 Comité de Basilea

Este Comité de Basilea es el ente regulador de la banca internacional (Bassel Commite on Banking Supervision, 2005), el cual se encarga de formular estándares internacionales sobre regulación y supervisión bancaria.

Es la principal contraparte de los bancos centrales para transacciones financieras.

**2.4.1.1 Objetivo:** Fortalecimiento de la estabilidad del sistema financiero mundial, mejorando las prácticas de gestión y gobernanca de lo riesgos.

## 2.5 Regulaciones de Basilea

Para evitar confusiones, en lo que sigue se dirá que el movimiento de un sistema de clasificación es "procíclico" si tiende a amplificar las fluctuaciones de la economía. De acuerdo con esta definición, por ejemplo, los sistemas de clasificación se comportan de forma procíclica si aumentan en recesión y disminuyen en un auge económico (Borio, Furfine y Lowe, 2001).

La evolución del sector financiero en las últimas décadas ha jugado un papel importante en la conformación de los resultados macroeconómicos de varios países. Por lo que esta evolución ha reforzado el impulso de las fluctuaciones económicas subyacentes y, en algunos casos, también ha provocado oscilaciones extremas en la actividad económica.

Todo lo mencionado ha conllevado a la preocupación de que el sistema financiero sea excesivamente procíclico. Y, por ende, todo esto ha provocado que se establezcan cambios tanto en la regulación prudencial como en las prácticas de medición de riesgos, esto con el objetivo de mejorar el sistema financiero y la estabilidad macroeconómica (Bassel Commite on Banking Supervision, 2001).

Ahora, las políticas orientadas a la adecuación de capital y al aprovisionamiento son herramientas importantes que sirven para abordar la prociclicidad excesiva en el sistema financiero. A su vez, estas políticas de adecuación de capital dependen de la evaluación de riesgo realizada por los propios bancos, como lo ha propuesto el Comité de Supervisión Bancaria de Basilea.

Es por eso que en la evolución de la banca el Comité de Basilea, el Banco de Pagos Internacionales y el Fondo Monetario Internacional se han convertido en entes reguladores y de supervisión en el sistema bancario, brindando soluciones eficientes a todos estos problemas del sistema financiero (Reinhart y Rogoff, 2008).

En estos temas se ha designado al Banco de Pagos Internacionales (BIS), el cual ha promulgado diversos acuerdos conocidos como los acuerdos de Basilea. Estos acuerdos buscan mejorar la seguridad del sistema bancario con la imposición de requisitos mínimos de capital, seguro para los depósitos, estructura de los balances homogéneos y la constante supervisión de los resultados (Sotelsek y Pavón, 2012), basándose en un buen control y administración de los riesgos que pueden tomar las entidades financieras.

Para llevar a cabo una buena regulación bancaria, garantizar estabilidad financiera y, por ende, que cada institución tenga solidez financiera se establecieron los *Acuerdos sobre Capital del Comité de Basilea* (Bassel Commmite on Banking Supervision, 2004), entre los cuales se encuentran el Acuerdo de Basilea I, Basilea II y Basilea III, los cuales tienen por objetivo principal establecer controles de los posibles riesgos en las instituciones financieras.

A continuación, se llevará a cabo una descripción de los Acuerdo de Basilea I y Basilea II, prestando especial atención al acuerdo de Basilea II ya que se enfoca en el tema del presente estudio, que trata de modelos internos para el control de riesgo de crédito en una institución financiera.

### **2.5.1 Análisis de Basilea I**

El primer Acuerdo de Basilea celebrado en 1988 surgió de la necesidad de regular el capital mínimo requerido para solventar el riesgo crediticio del sistema bancario internacional. Para ello, se propuso un sistema de valoración de activos homogéneos que considera incluso a los activos que no figuran en los balances. La norma indica que los bancos deben considerar un mínimo del 8% de capital sobre sus activos ponderados por riesgo (Sotelsek y Pavón, 2012).

Para el año 1999, el Acuerdo de Basilea era considerado un importante régimen de



disposiciones cautelares, así lo reconocían el Fondo Monetario Internacional y el Banco Mundial, por lo que, llegó a ser aplicado en más de 130 países a nivel mundial (Pancorbo, 2002).

En este Acuerdo de Basilea se agregaron requerimientos de capital con el objetivo de que las instituciones financieras se manejaran con un capital mínimo que estuviera acorde a los riesgos que pudieran asumir, salvaguardando la solvencia de la institución financiera.

### *2.5.1.1 Componentes del Capital*

En el Acuerdo de Basilea (Bassel Commmite on Banking Supervision, 1998) se establecen los siguientes componentes del capital:

- **Capital Esencial (patrimonio básico):** Este es el mayor componente del capital analizado en las instituciones financieras. Por su visibilidad en las cuentas publicadas, el capital es útil para realizar juicios de adecuación del capital y analizar los márgenes de rentabilidad y la competitividad de las instituciones financieras. El Comité pretende asegurar calidad y buenos niveles de capital mantenidos por los bancos (Bassel Commmite on Banking Supervision, 1998).
- **Capital Complementario:** El Comité de Basilea señala que es necesario considerar otras partes legítimas del capital que deben ser consideradas para estimar la cobertura del riesgo de crédito. Dentro del capital complementario se pueden encontrar las reservas no divulgadas, las reservas de revaluación, las provisiones generales, reservas generales para préstamos, los instrumentos híbridos de deuda de capital y la deuda subordinada a término (Bassel Commmite on Banking Supervision, 1998).

### *2.5.1.2 Las Ponderaciones de Riesgo*

El Comité de Basilea recomienda que los bancos construyan un coeficiente, el cual tome en cuenta el capital que se encuentra dentro o fuera de los balances, y que este coeficiente se relacione con los riesgos relativos de crédito para de esa forma poder mantener un nivel adecuado de capital. Las ponderaciones por riesgo iban desde el

0% hasta el 100%, donde el 100% representa activos que deben ser cubiertos en la totalidad de su valor (Bassel Commmite on Banking Supervision, 1998).

Debido a la simpleza del proceso de medición de riesgos en Basilea I, y al no tomar en consideración otros riesgos adicionales que tienen las instituciones bancarias, se desarrolló un nuevo aporte conocido como Basilea II, el cual permite a los supervisores obtener indicadores más precisos para estimar el capital óptimo en la cobertura de riesgos. (Pancorbo, 2002).

## **2.5.2 Análisis de Basilea II**

El acuerdo de Basilea II nace como una mejora al acuerdo de Basilea I argumentando que el capital no es suficiente para garantizar la estabilidad del sistema financiero, pues también se requiere de una adecuada gestión y control de riesgos. Además, se propone la regulación interna por medio de modelos internos y control del supervisor encargado. Se espera que la modelización interna ayude a controlar los riesgos de mercado, de crédito y operacional, aunque su implementación sea compleja y se requiera de una base de datos muy amplia (Gutiérrez y Fernández, 2006).

El objetivo principal del Acuerdo de Basilea II es fortalecer la solidez y mantener una estabilidad en el sistema financiero, afinando y mejorando el Acuerdo de Basilea I mediante el uso de buenas prácticas en el manejo de riesgos de las instituciones financieras.

Basilea II busca determinar el capital mínimo regulatorio, asignando responsabilidades a supervisores bancarios con criterios para emitir juicios objetivos acerca del riesgo que asume la institución. Con este acuerdo se busca mejorar tanto la eficiencia como el rendimiento de las transacciones y procesos bancarios.

Para alcanzar los objetivos planteados, el Acuerdo de Basilea II se basa en tres pilares fundamentales, que se describen a continuación.

### ***2.5.2.1 Primer Pilar: Requerimientos Mínimos de Capital***

Este pilar se centra en los fondos propios mínimos que tienen que tener las entidades bancarias para conseguir una correcta gestión de los riesgos, mediante el uso de modelos que midan el nivel de estos. Se sigue manteniendo el 8% del capital con respecto a

los riesgos asumidos. En la nueva definición de capital propio requerido se incorpora en el análisis el riesgo operacional como sigue:

$$\frac{\text{Patrimonio Neto}}{RWA + 12,5 * (RCRM + RCRO)} \geq 8\% \quad (2.1)$$

Donde:

*RWA* : Activos ponderados por riesgo de crédito.

*RCRM* : Requerimiento de capital por riesgo de mercado.

*RCRO* : Requerimiento de capital por riesgo operativo.

La especificación del requerimiento mínimo de capital muestra las diferencias en cómo se valoran los riesgos. Para el riesgo de crédito se determinan directamente los activos ponderados, y, por otra parte, para los riesgos de mercado y operacional se determinan montos de capital.

Los requerimientos de capital sirven para cubrirse de pérdidas no esperadas y para estimar provisiones para las pérdidas esperadas. Si existe déficit en la cobertura de pérdidas esperadas se recomienda usar el capital de TIER 1<sup>10</sup> y TIER 2<sup>11</sup> en partes iguales, por otra parte, si existe exceso en la cobertura de pérdidas esperadas se recomienda usar el TIER 2.

En Basilea II se presentaron varios cambios para el caso de riesgo de crédito, estableciendo el *método estándar (ME)* y los *métodos basados en sistemas de clasificación interna (IRB)* para estimar el capital mínimo requerido (Gutiérrez y Fernández, 2006).

### *Método Estándar*

Este método es considerado como una extensión del acuerdo de Basilea I que asigna diferentes ponderaciones a los activos y otras operaciones que no estén consideradas en los balances, con el fin de obtener un total de activos ponderados por riesgo. Las ponderaciones se basan en el tipo de cliente y en el criterio de evaluación de crédito realizada por una institución externa.

El hecho de que este método fuera desarrollado por agencias calificadoras externas

<sup>10</sup>TIER 1: Acciones ordinarias y reservas reveladas

<sup>11</sup>TIER 2: Instrumentos de capital de deuda híbrida

conllevó a que reabriera diversas críticas que estaban basadas en el sesgo que podría beneficiar a ciertos países pertenecientes al Comité de Basilea. Por lo cual el Comité, propuso otra posibilidad para medir el nivel de riesgos en las entidades bancarias, mediante la implementación de los modelos internos.

### *Modelos internos IRB*

Este tipo de modelos buscan estimar la calidad crediticia de los clientes y las pérdidas futuras potenciales, basados en datos propios de la institución financiera. Es importante mencionar que las calificaciones de los clientes mediante este método van a depender no solo de la solvencia del cliente, sino también de las condiciones macroeconómicas en las que se encuentre el país.

Para esto se utilizan 3 variables: la probabilidad de incumplimiento *PD*, Pérdida por incumplimiento *LGD* y la exposición al riesgo *EAD*.

Existen dos variantes de estos modelos internos, el primero es un método básico donde un supervisor fija las magnitudes de la *LGD* y la *EAD*, de modo que la institución financiera solo debe estimar la *PD*, y hacer un mapeo de los clientes mediante un sistema de clasificación PIT o TTC. El segundo es un método avanzado donde el propio banco realiza las estimaciones de las 3 variables, aunque de igual forma se debe someter al criterio del supervisor para su aprobación.

Para utilizar sus propios parámetros de riesgo para asignación de capital los modelos internos deben ser capaces de:

1. **Ordenar el riesgo:** Clasificar a los individuos en grupos homogéneos en relación al riesgo de crédito y de acuerdo a un sistema de clasificación PIT o TTC.
2. **Cuantificar el riesgo:** Atribuir parámetros de riesgo (*PD*, *LGD*, *EAD*) a cada uno de los grupos.

Y, luego, se tiene que mostrar que esos grupos de riesgo creados son válidos (Gutiérrez y Fernández, 2006), verificando la diferencia de medias entre los grupos y, analizando la migración de los clientes en cada grupo de riesgo.

#### *2.5.2.2 Segundo Pilar: El proceso de Revisión del Supervisor*

Este pilar parte de la necesidad de incorporar otros elementos que afectan a los niveles de riesgo de los bancos y que no han sido considerados en el pilar 1. De esta manera, los supervisores deben preocuparse de que los bancos elaboren modelos internos que se adapten a sus características internas. Además, deben evaluar la forma en que estos se realizan.

En otras palabras, el supervisor debe encargarse de controlar que la posición de capital del banco esté de acuerdo a sus características y, de darse el caso de que la posición de capital no sea buena, debe realizarse una pronta intervención para mantener un nivel óptimo de capital frente al riesgo.

El proceso de supervisión se basa en los siguientes principios (Bassel Commite on Banking Supervision, 2004):

- Los bancos deben establecer procesos para evaluar la suficiencia de su capital total y para mantener niveles óptimos de cobertura según su perfil de riesgo.
- Los supervisores <sup>12</sup> deben monitorear los procesos internos de evaluación de la suficiencia de capital y verificar si los bancos son capaces de mantener un capital regulador que permita cubrir los riesgos. En caso de ser necesario, los supervisores deben intervenir.
- Los supervisores esperan que los bancos operen por encima del monto mínimo del capital regulador.
- En el caso de que los niveles de capital estén por debajo de los coeficientes del capital regulador el supervisor podrá exigir inmediatamente el uso de medidas correctoras para mantener un capital regulador por encima de los niveles mínimos requeridos.

### 2.5.2.3 Tercer Pilar: *Disciplina de Mercado*

La disciplina de mercado se refiere a la acción de los supervisores para solucionar la ineficiencia en el mercado de depósitos por la falta de información acerca de los bancos. Por otra parte, el análisis de la disciplina de mercado sirve para evitar la quiebra de las

---

<sup>12</sup>Los supervisores son ejecutivos que tienen a su disposición un rango adecuado de herramientas de supervisión para realizar acciones correctivas oportunas.

instituciones financieras ineficientes, ya que esto podría tener efectos negativos en el ámbito socioeconómico.

La información completa sobre el capital que los bancos tienen como colchón ante posibles pérdidas y sus exposiciones al riesgo, permite a la contraparte realizar una valoración efectiva. Esta información deberá ser presentada semestralmente de forma general y, trimestralmente, para verificar la suficiencia del capital. Además, la información debe ser presentada de forma homogénea, de acuerdo a los estándares establecidos que permitan su comparación a nivel nación e internacional (Gutiérrez y Fernández, 2006).

## **2.6 Filosofías de clasificación de riesgos**

En general, la evidencia empírica es coherente con la opinión de que la prociclicidad del sistema financiero puede estar en la raíz de la inestabilidad financiera y que las medidas de riesgo se comportan como si el riesgo disminuyera durante un auge económico y se elevara durante una recesión, es decir, el riesgo se verá afectado por las condiciones macroeconómicas generales que afecten al país.

Es por esa razón que el Acuerdo de Basilea II introdujo la posibilidad de que los bancos usen las filosofías de clasificación PIT o TTC, y posterior a esto, que creen grupos homogéneos de riesgo tomando en consideración las condiciones económicas durante una expansión o recesión económica.

Los sistemas de clasificación crediticia asignan los clientes a diferentes grupos de riesgo, con el fin de distinguir entre clientes de diferente calidad crediticia, es decir, obtener una diferencia significativa de las PD.

Usualmente, los sistemas de clasificación hacen uso tanto de modelos estadísticos (análisis objetivo) como de juicios expertos (análisis subjetivo) para asignar los clientes a las calificaciones.

En la literatura se menciona al sistema de clasificación PIT y al sistema de clasificación TTC, como modelos de clasificación que describen las características dinámicas de los clientes, en los cuales se hace la inclusión de efectos macroeconómicos. Uno de estos enfoques se basa en el estado real de la economía y del estado de políticas de crédito de la institución financiera, el cual se conoce como PIT; otro enfoque es el TTC el cual

analiza características demográficas de los clientes.

No existe una definición concisa de TTC, pero generalmente hacen referencia a las estimaciones de PD estresadas <sup>13</sup>. De hecho, se manifiesta que una PD estresada se destaca por tener un enfoque TTC, mientras que una PD no estresada pertenece al enfoque PIT (Bassel Commite on Banking Supervision, 2005).

Heitfield (2005) manifiesta que el PD sin estrés de un grupo de riesgo PIT y el PD estresado de un grupo de riesgo TTC se mantienen constantes en el tiempo. Sin embargo, un PD estresado de un grupo de riesgo PIT y un PD sin estrés de un grupo de riesgo TTC si varían con el tiempo.

### 2.6.1 Definición y Comportamientos Esperados

Para estimar la PD el banco puede seguir una filosofía de clasificación de riesgo PIT (Point in the time) o TTC (through the cycle).

Valles (2006) manifiesta que un sistema PIT hace referencia a la información utilizada por el sistema (sin tomar en consideración las condiciones de estrés), y el sistema TTC utiliza información del cliente, sin tomar en consideración factores sistémicos, ni macroeconómicos.

Estas filosofías son maneras como clasificamos a los clientes en grupos de riesgo de acuerdo a la información que se tiene disponible. Una de las maneras de clasificar a los clientes es a través de los modelos de score o también conocidos como modelos de clasificación, los cuales son usados para automatizar el proceso de gestión de créditos. En ese contexto la diferencia entre PIT y TTC está en las informaciones que el score de cada sistema usa (Bassel Commite on Banking Supervision, 2005).

Esta información disponible está dividida en categorías:

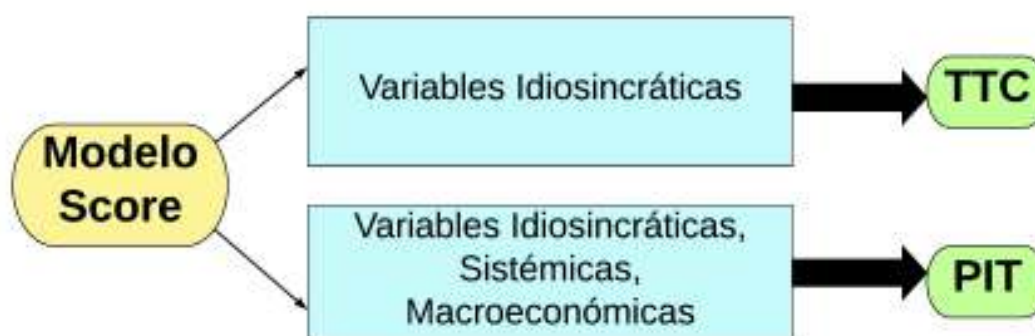
- *Información específica del cliente*, que es aquella información que es propia de un cliente en particular como por ejemplo el estado civil, edad, ingresos, ocupación, etc. Este tipo de información son las denominadas variables idiosincráticas.

---

<sup>13</sup>Las PD pueden evidenciar ciertos escenarios de estrés. En donde un escenario de estrés es aquel en donde se supone que las condiciones económicas futuras tienen muy poca probabilidad de que ocurran, pero si ocurrieran conllevaría a tener altas pérdidas crediticias.

- *Información agregada*, que es aquella información que es observable en el momento que se calcula la PD estimada del cliente, es decir información acerca de la situación macroeconómica del país como el PIB, la inflación, etc. Este tipo de información son las denominadas variables sistémicas. (Bassel Commmite on Banking Supervision, 2005)

En la figura 1 se muestra el tipo de información que usa cada filosofía de clasificación:



**Fuente:** Elaboración Propia

Figura 1. Filosofías de Clasificación

### 2.6.2 Filosofía de Clasificación TTC

Los sistemas TTC usan información específica del cliente y no usan información sistémica, es decir, se basan en características estáticas y dinámicas propias de los clientes.

El score ordena a los clientes en relación a su calidad de crédito, y esa ordenación es válida en cualquier instante durante las fluctuaciones en épocas de recesión, o expansión, o de políticas de crédito; ya que, estas calificaciones TTC no tienden a adaptarse a los posibles cambios que ocurran en las variables macroeconómicas del país.

El sistema TTC se basa en el rendimiento probable que tendrá un cliente durante un auge o recesión económica, bajo condiciones macroeconómicas adversas, cuando sus calificaciones tienden a no ajustarse a los posibles cambios, manteniéndose en el mismo grupo sin migrar a otro, a menos que cambie alguna característica idiosincrática (Cesaroni, 2015).

Por lo que el sistema de clasificación TTC es llamado acíclico (Bassel Commmite on Banking Supervision, 2005).



### 2.6.3 Filosofía de Clasificación PIT

El sistema PIT (punto en el tiempo) se basa en la información actual de un cliente y, también, puede basarse en el estado futuro que tendrá probablemente el cliente dentro de un determinado tiempo (BIS, 2000).

Por otra parte, este sistema utiliza toda la información agregada y específica de los clientes que se encuentra disponible, es decir tanto información idiosincrática, sistémica y macroeconómica. En este sistema de clasificación PIT se espera que un cliente responda a los cambios a medida que las condiciones macroeconómicas o de políticas de crédito, cambien. En general, las calificaciones PIT tenderán a caer durante expansiones económicas y aumentar durante las recesiones económicas.

Por lo que el sistema de clasificación PIT es llamado procíclico (Bassel Commmite on Banking Supervision, 2005).

### 2.6.4 Diferencias entre PIT y TTC

	PIT	TTC
<b>i) Tipo de Información</b>	Idiosincráticas y sistémicas	Principalmente Idiosincráticas
<b>ii) Comportamiento de la PD atribuida al individuo</b>	Sube en recesión y cae en períodos de expansión	No necesariamente acompañan las fluctuaciones durante una expansión o recesión económica.
<b>iii) Tasa de Mora observada del grupo</b>	Menos volátiles: más próximas de la PD atribuida al grupo en cualquier momento.	Más volátiles: con el pasar del tiempo la tasa de mora observada media tiende a ser la PD media del grupo.

**Fuente:** (Bassel Commmite on Banking Supervision, 2005)

Tabla 1

*Diferencias del sistema PIT y TTC*

### 2.6.5 Evaluación de Desempeño PIT y TTC

Como ya se mencionó, los factores que determinan el nivel de riesgo son: factor idiosincrático, factor macroeconómico y factor sistémico. El tipo de información usada antes de crear cada grupo homogéneo es el factor clave para construir un sistema de clasificación (score).

La tasa de mora atribuida a cada grupo de clientes <sup>14</sup>, su dinámica y su capital regulatorio dependerán fuertemente del tipo de filosofía adoptada (Bassel Commite on Banking Supervision, 2005).

El objetivo es obtener un sistema de clasificación confiable de tal forma que prediga lo más acertado posible la probabilidad de incumplimiento de un cliente, y que no exista variabilidad o migración excesiva en los grupos de riesgo.

Pero, existe un limitante cuando se hace uso de la filosofía de clasificación PIT, ya que este sistema, aunque posee un potente poder predictivo puede conllevar a la migración de los clientes entre grupos de riesgo a través del tiempo. Lo contrario ocurre con un sistema de clasificación TTC, el cual tiene un menor poder predictivo, sin embargo, puede presentar más estabilidad dentro de los grupos de riesgo.

Para determinar la estabilidad de los clientes en los grupos de riesgo, es válido tomar en cuenta la distancia de los grupos, ya que los grupos más cercanos tendrán más alta la probabilidad de migración de los clientes a lo largo del tiempo.

## **2.7 Dinámica de los sistemas de clasificación**

Una probabilidad de incumplimiento es una predicción prospectiva de la probabilidad de que un cliente cualquiera incumpla con su pago en un horizonte fijo de evaluación. En este trabajo se considera que la PD del grupo (PD agrupada) de un grupo de riesgo, es la tasa de mora promedio de los clientes que han sido asignados a determinado grupo.

Es válido centrarnos en las características de las probabilidades de incumplimiento que estén asociados con los clientes individuales. Una PD específica del cliente puede poseer información que se considera de gran importancia al momento de evaluar la capacidad y disposición del cliente para pagar sus deudas con la institución financiera. Además, puede considerar información acerca del entorno económico en el que se encuentra el cliente.

Algunos bancos pueden preferir usar cierto tipo de información del cliente para pre-

---

<sup>14</sup>La tasa de mora de cada grupo es la cantidad de clientes morosos dentro del grupo dividido para el total de clientes del grupo específico.

decir el incumplimiento, por ejemplo, no poner atención a las variables macroeconómicas. Pero, cabe mencionar que estas predicciones de incumplimiento al igual que los pronósticos económicos, deben considerar implícita o explícitamente supuestos sobre las condiciones económicas futuras (Bassel Commite on Banking Supervision, 2005).

Estos supuestos pueden evidenciar ciertos escenarios de estrés. Por lo que las características de las PD específicas de los clientes son sensibles tanto al tipo de información que usan, como a los supuestos de escenarios de estrés.

- **PD sin estrés:** Mide la probabilidad de que un cliente incumpla, usando toda la información que se encuentra disponible actualmente, considerando también características estáticas y dinámicas del cliente y a la vez datos agregados. Como esta PD hace uso de condiciones macroeconómicas, su comportamiento tendrá una variación, es decir, cuando estas condiciones mejoren tenderán a disminuir la PD, y a medida que las condiciones se deterioren el comportamiento de la PD aumentará.
- **PD estresado:** Es la probabilidad de que un cliente incumpla, dada toda la información disponible del cliente; pero, en este caso se asume que las condiciones macroeconómicas son adversas del escenario de estrés. Como esta PD hace uso de características dinámicas del cliente, tenderá a cambiar a medida que cambien las características individuales del cliente, sin embargo, no estará correlacionada con las fluctuaciones durante un auge, o recesión económica, o cambios de políticas de crédito.

### 2.7.1 Grupos de Riesgo y Estabilidad

Se debe realizar una selección óptima de los grupos de riesgo para cada sistema de clasificación, buscando obtener un equilibrio entre las migraciones de los clientes entre grupos a través del tiempo y la precisión de cada sistema de tal forma que, se discrimine adecuadamente la capacidad de pago y la solvencia de cada cliente.

Con el fin de reducir la prociclicidad se puede modificar la longitud de los grupos de riesgo, de tal forma que no se tenga concentrados gran cantidad de clientes en un solo

grupo de riesgo, sino más bien que se encuentren en distintos grupos tal que compartan PD similares en el grupo asignado y sea evidente cuando exista un cambio en la probabilidad de incumplimiento.

Ahora analizaremos la dinámica que tienen las PD agrupadas para cada grupo de riesgo.

### **2.7.2 Dinámica de las PD agrupadas**

Se asumirá que la PD agrupada de un grupo de riesgo será igual a la tasa de mora promedio de los clientes que están asignados a determinado grupo.

Según Cesaroni (2015), para un sistema de calificación PIT, todos los clientes deberán compartir aproximadamente la misma PD sin estrés, lo cual implicaría que la PD agrupada sin estrés para un grupo de riesgo PIT deberá permanecer usualmente fija durante el transcurso de fluctuaciones económicas o de políticas de crédito. Y cuando las condiciones macroeconómicas mejoren y las PD sin estrés de los clientes tiendan a subir, los clientes de mayor calidad crediticia se retirarán de un grupo de riesgo PIT (migrando a un grupo de mayor calificación crediticia), mientras que los clientes que tienen una menor calidad crediticia serán transferidos a un grupo tal que la PD agrupada sin estrés para ese grupo permanezca fija. Sucederá todo lo contrario cuando las condiciones macroeconómicas se deterioren.

En cambio, un PD agrupado con estrés de un grupo de riesgo PIT no tiene la misma dinámica, ya que este PD se correlacionará positivamente con las fluctuaciones económicas durante un auge o recesión económica, esto quiere decir que a medida que las condiciones macroeconómicas mejoren la PD aumentará, y cuando las condiciones macroeconómicas empeoren la PD disminuirá.

En la literatura se manifiesta que un sistema de clasificación TTC está diseñado de tal forma que todos los clientes que pertenezcan a un grupo de riesgo TTC tengan la misma PD con estrés, lo cual implica que la PD agrupada con estrés para un grupo de riesgo TTC se mantiene estable a lo largo de las fluctuaciones económicas o de políticas de crédito.

Cuando cambian las condiciones o circunstancias particulares de un cliente específico, este cliente puede migrar dentro y fuera de los grupos de riesgo TTC.

En los grupos de riesgo TTC se espera que las PD agrupadas sin estrés se correlacionen negativamente con las fluctuaciones económicas, es decir, las PD tenderán a disminuir durante un crecimiento económico, y tenderán a aumentar durante una recesión económica.

En la siguiente tabla se resume las características dinámicas de las PD agrupadas:

<b>Propiedades de las PDs de los clientes</b>	<b>Filosofías de Clasificación</b>	
	<b>PIT</b>	<b>TTC</b>
No Estresada	Estable	Negativo
Estresada	Positivo	Estable

**Fuente:** (Bassel Commmite on Banking Supervision, 2005)

Tabla 2

*Correlación de PDs agrupados con las fluctuaciones de la economía.*

### 2.7.3 Cuantificación de las PD agrupadas

En la práctica, es indispensable estimar las PD agrupadas a partir de los datos que se tienen disponibles. En el Marco revisado se describe tres enfoques amplios para cuantificar las PD agrupadas.

- **Enfoque histórico:** Es aquel que se basa en la experiencia de incumplimiento, es decir el PD agrupado para un grupo de riesgo se lo estima a partir de datos históricos tomando en consideración la frecuencia de incumplimientos observados en los clientes que pertenecen a ese grupo de riesgo.
- **Enfoque de Modelo Estadístico:** Primero se hace uso de los modelos estadísticos predictivos para estimar la PD específica de cada cliente que está asignado a un grupo, entonces la PD agrupada del grupo de riesgo se la calcula como el promedio de las PD específicas del cliente.
- **Enfoque de Mapeo Externo:** Se establece un mapeo que vincula cada grupo de riesgo interno de una institución financiera con calificaciones externas.

En este proyecto nos basaremos en un enfoque histórico para cuantificar las PD agrupadas.

## 2.8 Modelo Credit Scoring

Los modelos de score o credit scoring son considerados un buen punto de partida para comenzar el análisis formal de las filosofías de clasificación; básicamente sirven para estimar el riesgo de que un cliente no pague, basados en fórmulas estadísticas para identificar buenos o malos clientes.

La estimación se realiza por medio de la discriminación de los nuevos clientes con el uso de datos sobre el cumplimiento de clientes antiguos (Rodríguez, Becerra y Cardona, 2017), con el fin de determinar el comportamiento de pago de esos nuevos clientes.

### 2.8.1 Ventajas y Desventajas de las distintas Metodologías del Credit Scoring

En la literatura se mencionan distintas metodologías para modelizar el riesgo de crédito, entre los métodos más comunes y más usados se encuentran la regresión logística, redes neuronales, análisis discriminante, support vector machine y probit.

Todas estas metodologías se basan en analizar la relación entre ciertas variables consideradas predictoras de riesgo con la variable dependiente PD (probabilidad de incumplimiento) mediante una función.

En este subcapítulo se describirá algunos trabajos de credit scoring que usan distintas metodologías, y se mostrarán sus ventajas, desventajas y las características que identifican los diferentes casos.

M. Bonilla, Puertas y Olmeda (2003) propusieron varios modelos con el fin de predecir la calificación crediticia de un cliente, y al realizar las comparaciones respectivas, concluyeron que las redes neuronales dan mejores resultados. Sin embargo, esta metodología no paramétrica realiza un proceso interno de aprendizaje por lo que les resulta complicado la interpretación de los resultados del modelo.

SunHoHa, Ramayya y Krishnan (2012) proponen un modelo con el cual se pueda pre-

decir el posible comportamiento de crédito de los clientes, específicamente, en el producto de tarjetas de crédito. El modelo agrupa a los clientes por segmentos y obtiene una predicción de pago para cada segmento de clientes a través del modelo de riesgo proporcional de Cox.

Aguas y Castillo (2002) proponen un modelo para la modelización de riesgo de crédito usando la regresión logística con el fin de reducir las pérdidas potenciales de la institución financiera cuando otorgan un crédito al cliente. Este trabajo es bastante interesante, pues considera información de la economía del país. Es decir, cuando el país está pasando por una recesión o expansión, y añade al modelo el análisis llamado “hipótesis de la inestabilidad”.

Cardoso y Díaz (2015) implementan un modelo de supervivencia mixto, en el cual se establecen distintos grupos de clientes donde cada grupo posee un comportamiento crediticio diferente. Mediante este modelo es posible notar el comportamiento heterogéneo no observado de los clientes, gracias a las combinaciones de distribuciones log normal.

Un estudio que fue realizado en la industria financiera de Camerún por Abdouab, Dongmo, Ntim y Baker (2016), implementó distintas técnicas como las redes neuronales, la regresión logística, y también árboles de clasificación y regresión en el desarrollo del modelo de predicción. Las conclusiones a las que llegaron los autores es que con la técnica de redes neuronales se obtuvieron mejores predicciones a comparación con las otras.

Finalmente se puede llegar a la conclusión de que la técnica más usada para modelar el riesgo de crédito es la regresión logística (Cox y Snell, 1989). Haciendo una comparación de la técnica de redes neuronales o support vector machine con la regresión logística, las primeras tienen una mejor capacidad de predicción, pero la desventaja es que no poseen robustez ni una adecuada interpretabilidad de los resultados. Es por esa razón que no son usadas en el ámbito bancario. Cabe mencionar que también se podría modelar usando el análisis de supervivencia, pero debido a la limitación de los datos con respecto a los meses de estudio se dificulta la implementación de esta técnica.

Por todo lo dicho anteriormente la regresión logística es la metodología elegida para

aplicarla en este trabajo.

## 2.9 Metodología Estadística

Cada institución financiera emplea una metodología diferente para determinar sus pérdidas potenciales con respecto al riesgo de crédito, basadas en las políticas de crédito, estrategias financieras, niveles de riesgo que pueda asumir cada entidad, entre otros. A continuación, se explicará la teoría de toda la metodología, métodos y pruebas estadísticas que serán usados para el desarrollo de este proyecto.

### 2.9.1 Metodología para el Desarrollo del Modelo de Credit Scoring

#### 2.9.1.1 Modelo de Incumplimiento del cliente

El autor (Heitfield, 2005) propone un modelo para presentar las dos filosofías de clasificación PIT y TTC, basándose en las características de las PD estresadas y no estresadas.

Este modelo muestra la relación entre las características dinámicas de un sistema de clasificación crediticia y las características de las PD agrupadas asociadas con ese sistema de clasificación.

Sea  $Z_{it}$  una v.a que está normalmente distribuida, la cual es única para el cliente  $i$  y para cada fecha  $t$ , además se considera que esta variable  $Z_{it}$  depende de ciertos factores de riesgo que son observables e inobservables. El modelo es de la forma:

$$Z_{i,t+1} = \alpha + \beta_W W_i + \beta_X X_{it} + \beta_Y Y_t + U_{i,t+1} \quad (2.2)$$

$W_i$  es un factor de riesgo fijo que considera las características del cliente  $i$  que no varían en el tiempo,  $Y_t$  es un factor de riesgo el cual afecta directamente la calidad crediticia de todos los clientes en la cartera del banco. Se asume que el entorno macroeconómico puede ser observado por un banco en la fecha  $t$ .

El factor de riesgo  $X_{it}$  considera las características dinámicas del cliente  $i$  que son observables en la fecha  $t$ , y además que  $X_{it}$  es independiente de la variable  $Y_t$ .

En conjunto,  $W_i$ ,  $Y_t$  y  $X_{it}$  representan el conjunto completo de información disponible



para evaluar la calidad crediticia del cliente  $i$  en la fecha  $t$ .

La variable  $U_{i,t+1}$  representa aquella información que afecta al estado de incumplimiento <sup>15</sup> de un cliente en el tiempo  $t + 1$ , y es información que no ha podido ser observada por el banco en la fecha  $t$ .

$$U_{i,t+1} = \omega V_{t+1} + \sqrt{1 - \omega^2} E_{i,t+1} \quad (2.3)$$

Donde  $V_{t+1}$  es un factor de riesgo sistemático que comparten todos los clientes, y  $E_{i,t+1}$  es un factor de riesgo idiosincrático que es exclusivo para el cliente  $i$ . El parámetro  $\omega$  es aquel que determina la sensibilidad del incumplimiento al factor de riesgo sistemático inobservable.

Una vez visto el modelo de incumplimiento del cliente, se procede a adaptar este modelo a un modelo de regresión logística el cual se presenta a continuación.

### 2.9.1.2 Regresión Logística

Los modelos de elección cualitativa o binaria son aquellos que se usan cuando la variable que se desea modelar es dicotómica (binaria), y sirven para predecir la probabilidad de que ocurra un fenómeno dadas ciertas variables explicativas. Por ejemplo, en el presente estudio, nos interesa determinar cuál es la probabilidad de que un cliente cumpla o no cumpla con los pagos de sus préstamos considerando un conjunto de características del cliente que pueden afectar a la variable a predecir.

Por ende la variable dependiente solo puede tomar los valores de 0 o 1.

---

<sup>15</sup>El incumplimiento se refiere cuando un cliente entra en default, es decir no cumple con su pago en el banco.

El modelo estará definido por la siguiente función de distribución logística:

$$P\left(\sum_{i=1}^n Y_i = 1 \mid \sum_{i=1}^n \sum_{j=0}^k X_{ij}\right) = \pi(X_{ij}) = \frac{\exp\left(\sum_{j=0}^k \beta_j X_{ij}\right)}{1 + \exp\left(\sum_{j=0}^k \beta_j X_{ij}\right)} \quad (2.4)$$

Donde:

$Y_i$  es 1 si el cliente entra en mora, y 0 caso contrario para cada individuo  $i$ , donde  $i = 1, 2, \dots, n$ .

$X_{ij}$  son las  $j$  variables independientes del individuo  $i$ , donde  $i = 1, 2, \dots, n$  y  $j = 0, 1, \dots, k$ .

$\beta_j$  son los parámetros estimados asociados a las  $X_{ij}$ .

La ecuación (2.4) también puede expresarse de la siguiente forma:

$$\frac{\pi(X_{ij})}{1 - \pi(X_{ij})} = \exp\left(\sum_{j=0}^k \beta_j X_{ij}\right) \quad (2.5)$$

Es necesario realizar una transformación a la ecuación (2.5) para expresar el modelo de regresión logística de forma logit, con lo cual tomando logaritmos neperianos ( $\ln$ ) nos quedaría la función de la siguiente manera:

$$\ln\left(\frac{\pi(X_{ij})}{1 - \pi(X_{ij})}\right) = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} \quad (2.6)$$

Después de aplicar esta transformación a la ecuación 2.5, se convierte en lineal, por lo tanto, es más sencilla su interpretación y análisis.

Se hará uso del método de máxima verosimilitud para estimar los coeficientes de la regresión, de tal forma que esos estimadores contribuyan a maximizar la función logística y sean valores próximos a los observados.

Gujarati y Porter (2010) muestran una forma de como hallar esos estimadores en el modelo de regresión logística, para esto definen la función:  $f(Y_i)$ , la cual indica la proba-

bilidad de que el individuo  $i$  incumpla con su pago ( $Y_i = 1$ ) o caso contrario ( $Y_i = 0$ ), donde  $i = 1, 2, \dots, n$ . Luego al observar la probabilidad conjunta de los  $n$  individuos, la función de verosimilitud (FV) quedaría expresada de la siguiente forma:

$$f(Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n f_i(Y_i) = \prod_{i=1}^n \pi(X_{ij})^{Y_i} (1 - \pi(X_{ij}))^{1-Y_i} \quad (2.7)$$

para cada  $j = 0, 1, \dots, k$ .

Ahora al aplicar el logaritmo natural a la FV, se obtiene la función logaritmo de verosimilitud (FLV), suponiendo que cada  $Y_i$  es independiente y además que tienen igual función de densidad:

$$\ln f(Y_1, Y_2, \dots, Y_n) = \sum_{i=1}^n \left[ Y_i \ln \left( \frac{\pi(X_{ij})}{1 + \pi(X_{ij})} \right) \right] + \sum_{i=1}^n \ln (1 - \pi(X_{ij})) \quad (2.8)$$

Luego a partir de la ecuación (2.6) se puede expresar de otra forma la ecuación (2.8), obteniendo los siguientes resultados:

$$\ln f(Y_1, Y_2, \dots, Y_n) = \sum_{i=1}^n Y_i \left[ \beta_0 + \sum_{j=1}^k \beta_j X_{ij} \right] - \sum_{i=1}^n \ln \left[ 1 + \exp \left( \beta_0 + \sum_{j=1}^k \beta_j X_{ij} \right) \right] \quad (2.9)$$

El objetivo ahora es maximizar la función logaritmo de verosimilitud FLV establecida en la ecuación previa (2.9), pero existe el problema de que las soluciones que devuelve no son explícitas y se requieren de otros métodos para obtener soluciones numéricas. Para esto se pueden usar métodos de estimación no lineal, y así se garantiza que las estimaciones de los  $\beta$  sean numéricos, y por ende se pueda estimar fácilmente la ecuación (2.4).

### 2.9.1.3 Metodología para selección de variables

#### *Odds Ratio (OR)*

Los OR determinan la influencia que poseen las variables explicativas con respecto a la variable dependiente, mediante una razón de probabilidades en el caso de que la variable dependiente sea éxito. Se halla el resultado mediante la siguiente ecuación:

$$\frac{\pi(X_{ij})}{1 - \pi(X_{ij})} = \exp \left( \sum_{j=0}^k \beta_j X_{ij} \right) \quad (2.10)$$

### *Information Value*

En la calificación crediticia, Fair Isaac (FI) <sup>16</sup> introdujo una medida que denomino Information Value, que es usada para medir el poder predictivo de una característica. Y se la conoce técnicamente como la medida de divergencia de Kullback, y se utiliza para medir la diferencia entre dos distribuciones.

Es por eso que, durante la etapa del análisis preliminar de los datos, se deben elegir las variables predictoras adecuadas, eligiendo aquellas variables o atributos que posean las mejores propiedades de predicción para pronosticar la probabilidad de incumplimiento de un cliente, es decir, distinguir entre clientes que van a incumplir o no.

Se hace uso de la medida del Information Value (IV), para clasificar a las variables predictoras según su poder clasificador (Yap, Ong y Mohamed, 2011), y poder separar los altos riesgos de los bajos riesgos.

Este método ayudará para hacer la selección de las variables y luego hacer su inclusión en cada modelo. El Information Value IV se calcula de la siguiente forma:

$$IV = \sum_{i=1}^k \left[ \left( \frac{N_i}{\sum N} - \frac{P_i}{\sum P} \right) * \ln \left( \frac{\frac{N_i}{\sum N}}{\frac{P_i}{\sum P}} \right) \right] \quad (2.11)$$

Donde:

$P$ : es la ocurrencia (positiva), es decir, el número de malos clientes.

$N$ : es la no ocurrencia (negativa), es decir, el número de buenos clientes.

$i$ : es el índice del atributo que se está evaluando.

<sup>16</sup>Los pioneros más conocidos de la calificación crediticia son el ingeniero Bill Fair y el matemático Earl Isaac, quienes fundaron su consultora, Fair Isaac (FI), en San Francisco en 1956 (Anderson, 2007).

$k$ : es el número total de atributos.

El resultado para cada atributo o característica se denomina "contribución".

Los valores del IV siempre serán positivos y pueden estar por encima de 3 cuando se evalúen los puntajes proporcionados por tablas de puntuación de comportamiento altamente predictivas. Las características con valores de menos de 0.10 generalmente se consideran débiles, mientras que las características con valores de más de 0.30 son las que se buscan y es probable que aparezcan en los modelos de clasificación (Anderson, 2007).

Para poder interpretar los resultados del Information Value, se tiene la siguiente regla:

<b>Information Value</b>	<b>Poder de Predicción</b>
<0.02	Predictor Inútil
0.02 to 0.1	Predictor Débil
0.1 to 0.3	Predictor Medio
0.3 to 0.5	Predictor Fuerte
>0.5	Predictor Sospechoso

**Fuente:** (SAS, 2009)

Tabla 3

*Regla para el Information Value*

Lo ideal sería elegir aquellas variables que tengan un poder predictivo medio y fuerte, con lo cual basándonos en la Tabla (3) el information value debe ser mayor que 0.1 para que se considere la inclusión de una variable en el modelo.

#### **2.9.1.4 Pruebas de Especificación y Desempeño del Modelo**

##### ***Multicolinealidad***

Para analizar la relación existente entre la variable dependiente  $Y$  y las variables explicativas  $X_1, X_2, \dots, X_k$ , se hace uso de la multicolinealidad, es decir, se verifica que no haya correlación entre las variables explicativas incluidas en el modelo de regresión.

Este análisis se debe realizar para evitar efectos adversos en la predicción del modelo ya que repercute, de manera directa en el error estándar de los estimadores de Mínimos cuadrados ordinarios (MCO) de los parámetros relacionados con dichas variables; ya que se reflejan incrementados, lo que conlleva a que la estimación de los coeficientes no sea precisa y se obtenga coeficientes inestables y con alta varianza (Herrera, Raúl y Guevara, 2010).

Para detectar la existencia de multicolinealidad se procede, al análisis de la matriz de correlaciones de todas las variables predictoras conjuntas.

Sea  $C$  la matriz de correlaciones para las  $k$  variables regresoras, y sea  $l_j$  la  $j$ -ésima raíz latente de  $C$  con el correspondiente vector latente  $v_j$ . Denotemos  $L$  como la matriz diagonal con sus elementos  $l_j$  colocados de acuerdo a su magnitud, y sea la matriz de vectores latentes la siguiente:

$$V = [v_1, v_2, \dots, v_k]$$

Para calcular el grado de multicolinealidad entre las variables regresoras se usa frecuentemente la siguiente medida llamada *Factores de Inflación de Varianza Grandes VIF* (Mansfield y Helms, (1982)), que relaciona las raíces latentes con los vectores latentes de la matriz de correlación, donde  $VIF(j)$  corresponde a los elementos de la diagonal de la inversa de la matriz de correlación  $C$ , tal que:

$$VIF(j) = C_{jj}^{-1}$$

Donde:

$$C^{-1} = VL^{-1}V'$$

Este factor de inflación de la varianza (VIF), lo que hace es cuantificar la intensidad de la multicolinealidad en un análisis de regresión de mínimos cuadrados. El VIF proporciona un índice de medición, el cual mide hasta qué punto la varianza de un coeficiente de regresión estimado se incrementa a causa de la colinealidad.

Los valores de VIF mayores que 10 se consideran indicativos de multicolinealidad (Gujarati y Porter, 2010).

### *Poder Discriminatorio - Curva ROC*

Para poder evaluar el poder discriminatorio de un sistema de clasificación, Basilea propone varias metodologías estadísticas que pueden ser usadas en la industria financiera como medidas de discriminación. En este trabajo se hará uso de la curva ROC.

Cabe recalcar que este poder discriminatorio debe probarse tanto para un conjunto de desarrollo (train) como también para un conjunto de validación (test).

La Curva ROC (Receiver Operating Characteristic) se puede construir a partir de dos muestras que se encuentren disponibles de los puntajes de clientes malos (morosos) y buenos (no morosos). Estas dos muestras serían la PD observada y la PD obtenida de la predicción. El área bajo la curva ROC, llamada comúnmente AUC (Área Bajo la Curva), es una transformación lineal del ratio de precisión, donde en el eje horizontal se encuentra el porcentaje de falsos positivos, es decir, clientes malos que el modelo clasificó como clientes buenos (no morosos), y en el eje vertical se encuentra el porcentaje de verdaderos positivos que son los clientes buenos (no morosos) que fueron clasificados como buenos en el modelo.

El área bajo la curva AUC puede tomar valores entre 0.5 y 1. Un porcentaje del AUC del 50 % implica que el modelo no es mejor que hacer una suposición aleatoria; y un valor de 100 % indicaría la improbable aparición de predicciones perfectamente correctas. Sin embargo, para el caso de sistemas de clasificación lo ideal es tomar valores del AUC que sean mayores a 0.7, y entonces se dirá que esos modelos tienen una buena capacidad de clasificación (Anderson, 2007).

### *Kolmogorov – Smirnov (KS)*

Una de las estadísticas comúnmente utilizadas en la calificación crediticia, es la estadística KS (Anderson, 2007). La cual se basa en un análisis de la función de distribución empírica acumulativa (ECDF).

Es una medida no paramétrica que se usa para evaluar el error o "bondad de ajuste" en el ajuste de curvas.

Segun Mays (2004) la estadística KS es la estadística más utilizada para medir el poder predictivo de los sistemas de clasificación.

Para calcular el estadístico de Kolmogorov-Smirnov, sea  $D$  la función de distribución de las muestras  $F_{n_1}(x)$  y  $F_{n_2}(x)$  hay que calcular su máxima diferencia absoluta (siendo la diferencia absoluta máxima entre las dos curvas: la de clientes malos acumulados y clientes buenos acumulados), sobre todos los valores  $x$  tal que:

$$D = \max_x |F_{n_1}(x) - F_{n_2}(x)| \quad (2.12)$$

Mays (2004) también menciona que, como medida de la capacidad de clasificación, los valores de KS deben oscilar entre el 20 % (por debajo del cual debe cuestionarse el valor del modelo), y el 70 % (por encima del cual "probablemente sea demasiado bueno para ser cierto").

### 2.9.2 Metodología para la Creación de los Grupos de Riesgo

Luego de estimar las PD específicas de cada cliente mediante un modelo logístico, se hace uso de esta información para asignar a los clientes a los grupos de riesgo, tal que en el mismo grupo todos los clientes compartan la misma PD o valores similares.

El sistema de clasificación PIT, asignará a los clientes a los grupos de riesgos de tal forma que compartan PD similares sin estrés, mientras que el sistema de clasificación TTC asignará a los clientes a grupos de riesgo tal que los clientes en cada grupo compartan PD similares con estrés.

Heitfield se basó en el modelo de incumplimiento del cliente, y se definió el grupo de riesgo PIT de la siguiente forma:

$$\Gamma_t^{PIT} = \{i, |\alpha + \beta_W W_i + \beta_X X_{it} + \beta_Y Y_t = -\gamma_{PIT}\} \quad (2.13)$$

Donde  $-\gamma$  es el valor esperado del indicador de distancia al incumplimiento.

Y, entonces, todos los clientes en ese grupo de riesgo comparten la misma PD sin estrés,



la cual está definida como:

$$PD_{PIT}^{SE} | i \in \Gamma_t^{PIT} = \Phi(\gamma_{PIT}) \quad (2.14)$$

$\Phi$  denota la función de distribución normal acumulada.

Y la PD agrupada sin estrés para el grupo de riesgo PIT se define:

$$PPD_{PIT}^{SE} = E \left[ PD_{PIT}^{SE} | i \in \Gamma_t^{PIT}, Y_t = y_t \right] = \Phi(\gamma_{PIT})$$

De igual manera, basándose en el modelo de incumplimiento del cliente Heitfield ha definido el grupo de riesgo TTC de la siguiente forma:

$$\Gamma_t^{TTC} = \left\{ i, \left| \frac{\alpha + \beta_W W_i + \beta_X X_{it} + \psi}{\sqrt{1 - \omega^2}} \right| = -\gamma_{TTC} \right\} \quad (2.15)$$

donde  $\psi = \beta_Y Y_t + \omega V_{t+1}$ .

Por lo tanto, todos los clientes en ese grupo de riesgo comparten la misma PD estresada, la cual está definida como:

$$PD_{TTC}^E | i \in \Gamma_t^{TTC} = \Phi(\gamma_{TTC}) \quad (2.16)$$

Y, la PD agrupada con estrés para el grupo de riesgo TTC se define como:

$$PPD_{TTC}^E = E \left[ PD_{TTC}^E | i \in \Gamma_t^{TTC}, Y_t = y_t \right] = \Phi(\gamma_{TTC}) \quad (2.17)$$

### 2.9.2.1 Pruebas de Comparaciones Múltiples de Diferencias de Medias

Una vez obtenidos los grupos homogéneos de riesgo ordenados con relación al riesgo de crédito, se debe verificar que estos grupos son válidos. Para esto se debe realizar un análisis de comparación múltiple de diferencias de medias utilizando distintas técnicas como las pruebas de Duncan, Tukey, Scheffé, Games and Howell, Dunnet T3, entre otras.

Estas pruebas de comparación múltiple permiten probar igualdad de medias con un

coeficiente de confianza conjunto para todas las combinaciones. El propósito de estas pruebas es verificar si la PD de un determinado grupo homogéneo es estadísticamente diferente de la PD de otro grupo. Esta comparación se la conoce como comparación 2 a 2.

Las pruebas adecuadas que se usarán en este proyecto son la prueba de Tukey y la prueba Dunnett T3.

### *Prueba de Tukey*

Esta prueba sirve para comparar todas las diferencias entre medias posibles; es decir, compara las medias de cada grupo con los medias de cualquier otro grupo, y se basa en una distribución *studentized range*( $q(\alpha, k, v)$ ).

La prueba estadística es la siguiente:

$$\frac{n_i n_j (\bar{Y}_i - \bar{Y}_j)^2}{(n_i + n_j) MS_w} \sim Student\ range\ q(\alpha, k, v) \quad (2.18)$$

Donde:

$\bar{Y}_i$  y  $\bar{Y}_j$  son las medias muestrales del i-ésimo y j-ésimo grupo.

$n_i$  y  $n_j$  son los tamaños de muestra respectivos de los grupos i y j.

$MS_w$  es la varianza media de todos los grupos.

$\alpha$  es el nivel de significancia.

$k$  es el número de grupos.

$v$  son los grados de libertad.

El método de Tukey identifica cualquier diferencia entre dos medias que sea mayor que el error estándar esperado; y es conservador cuando hay tamaños de muestra desiguales.

Las suposiciones de la prueba Tukey son las siguientes:

- a) Normalidad
- b) Homocedasticidad
- c) Independencia entre la observaciones

d) Muestra aleatoria.

Y las hipótesis nula y alternativa son:

$$H_0: u_1 = u_2 = \dots = u_i$$

$$H_1: u_1 \neq u_2 \neq \dots \neq u_i$$

Donde  $u_i$  es la media del grupo  $i$ .

### Procedimiento:

1. Se calcula el valor crítico de todas las comparaciones por pares.
2. Se obtiene el error estándar de cada media.
3. Se establece  $\alpha$ .
4. Se realiza las comparaciones de las medias y del valor crítico.

La ventaja de usar esta prueba de Tukey es que es considerada el mejor método para todas las comparaciones de tratamiento por pares.

### *Prueba de Dunnet T3*

Esta prueba de comparación por pares se basa en el módulo máximo de *studentized*.

Esta prueba es apropiada cuando los tamaños de las muestras son desiguales.

Además se usa cuando los tamaños de muestra son pequeños, es decir,  $n < 50$  (Shingala y Rajyaguru, 2015); y se usa también si el grado de libertad es pequeño.

La prueba estadística es la siguiente:

$$\frac{(\bar{Y}_i - \bar{Y}_j)^2}{\frac{S_i^2}{n_i} + \frac{S_j^2}{n_j}} \sim t - Student \quad (2.19)$$

Donde:

$\bar{Y}_i$  y  $\bar{Y}_j$  son las medias muestrales del  $i$ -ésimo y  $j$ -ésimo grupo.

$S_i^2$  y  $S_j^2$  son las varianzas muestrales del i-ésimo y j-ésimo grupo.  
 $n_i$  y  $n_j$  son los tamaños de muestra respectivos de los grupos i y j.

Las suposiciones de la prueba Dunnett T3 son:

- a) Normalidad
- b) Homocedasticidad

Y las hipótesis nula y alternativa de esta prueba son:

$$H_0: u_1 = u_2 = \dots = u_i$$

$$H_1: u_1 \neq u_2 \neq \dots \neq u_i$$

Donde  $u_i$  es la media del grupo  $i$ .

Las ventajas de usar este método es que mantiene bajo control la tasa de error cuando se dispone de un grupo de tamaño desigual y, además, este método es robusto cuando no hay normalidad (Day y Quinn, 1989).

## Capítulo 3

### Desarrollo del Modelo Estadístico

#### 3.1 Descripción de la base de datos

En el presente estudio se usará una base de datos de corte transversal (la cual contiene observaciones sobre múltiples fenómenos en un momento determinado y, en donde, el orden de las observaciones es irrelevante) de una cartera de créditos concedidos de consumo proporcionada por una institución financiera de Guatemala, la cual consta de un total de 32585 individuos y de 219 variables explicativas que describen toda la información disponible de los clientes.

El período considerado para el modelado fue entre julio del 2015 y noviembre del 2015. A estos meses los llamaremos fechas de referencia.

##### 3.1.1 Estructura de la base de datos

La base de clientes proporcionada tiene la siguiente estructura:

Clientes	Fecha referencia: $t_0$	VARIABLES EXPLICATIVAS HISTÓRICAS					MARCACIÓN DE MORA		
		$X1_{(t-6)}$	$X1_{(t-3)}$	$X2_{(t-6)}$	$X2_{(t-3)}$	$X2_{(t_0)}$	$X3_{(t_0)}$	$D_{(t_1 t_2)}$	$D_{(t_1 t_4)}$
Cliente 1	jul-15								
Cliente 2	ago-15								
Cliente 3	nov-15								
.	.								
.	.								
Cliente n	oct-15								

**Fuente:** Elaboración Propia

Figura 2. Estructura de la base de datos

Donde:

$t_0$  se refiere a la fecha de referencia en la que se encuentra determinado cliente. Por ejemplo, para el cliente 1, la fecha de referencia es Jul-2015.

$t_{-3}$ : se refiere a la información de determinada variable o atributo, de 3 meses anteriores a la fecha de referencia.

$t_{-6}$ : se refiere a la información de determinada variable o atributo, de 6 meses anteriores a la fecha de referencia.

$t_1 : t_3$  se refiere a información futura a partir de la fecha de referencia, es decir, la información de los próximos 3 meses.

$t_1 : t_6$  se refiere a información futura a partir de la fecha de referencia, es decir, la información de los próximos 6 meses.

Es importante mencionar que en el momento de observación del cliente (fecha de referencia), se observaron 3 meses anteriores y hasta 6 meses anteriores para cada fecha de referencia; estas informaciones históricas de las variables explicativas se las usará para pronosticar. Y, para la marcación de mora del cliente, se observará el futuro, es decir, los días en atraso en los próximos 3 meses y en los próximos 6 meses a partir de la fecha de referencia.

Por ejemplo, para el cliente 1 que se encuentra en la figura 2, su fecha de referencia es Jul-2015. Para este cliente se tiene información histórica de hasta 6 meses atrás (Ene-2015) con respecto a las variables explicativas; y, para marcar la mora se observa el futuro, es decir, el número máximo de días en atraso hasta los próximos 3 meses (Oct-2015), y hasta los próximos 6 meses (Ene-2016).

### 3.1.2 Adición de Variables

Además de las 219 variables de la base, se agregaron 3 variables macroeconómicas como información agregada para tomar en cuenta la situación macroeconómica del país; estas variables agregadas son el *IMAE*<sup>17</sup>, el *IPC*<sup>18</sup> y la Inflación de Guatemala<sup>19</sup>, que se obtuvieron de la página oficial del Banco Central de Guatemala<sup>20</sup>.

### 3.1.3 Madurez de la Cartera

Se estableció que la madurez mínima en la cartera es de 3 meses. Esto con el fin de

<sup>17</sup>El Índice Mensual de Actividad Económica (IMAE) es un indicador que permite seguir el pulso a la actividad económica del año mientras se concreta la cifra de la producción anual (PIB).

<sup>18</sup>El Índice de Precios al Consumidor (IPC) es una variable que tiene por objeto evaluar a través del tiempo la variación de los precios de los artículos de consumo y servicio en Guatemala.

<sup>19</sup>El concepto de inflación se define técnicamente como el alza generalizada y persistente en el nivel de los precios internos de la economía de un país. Ver (Superintendencia de Bancos Guatemala, 2017).

<sup>20</sup><https://www.banguat.gob.gt/inc/main.asp?id=111348&aud=1&lang=1>

garantizar que el tiempo transcurrido luego de la fecha de otorgado el crédito, sea el tiempo suficiente para que se puedan observar características de comportamiento de pago de un cliente.

Esto quiere decir que se tomó en consideración los clientes que ya tienen al menos 3 meses con el crédito concedido.

### **3.2 Definición del Default**

Para definir el default o incumplimiento se considera el número máximo de días de atrasos que tiene un cliente en el pago de sus cuotas, ya sea en los próximos 3 meses o en los próximos 6 meses a partir de la fecha de referencia. Definir bien esta variable es un factor muy importante ya que es la variable que se va a predecir mediante el modelo de scoring.

#### **3.2.1 Status de los Clientes en los Pagos.**

Cuando la institución financiera concede el crédito a un cliente, se establecen periodos y fechas de pago para que el cliente cancele su cuota establecida sin cobrarle intereses. El problema aparece cuando el cliente no dispone del capital suficiente para cubrir su cuota mensual y cae en default. Es decir, el cliente migra a otro status en donde se toma en cuenta los días que tiene vencido. Es por esa razón que cada institución financiera tiene la facultad de decidir cuál será el nivel máximo que asuma de riesgos.

Por lo tanto, la cartera de crédito de consumo de la institución financiera con la que se está trabajando se la clasificó en las siguientes categorías, según el número de días vencidos.

Tipo de Categoría	Días de Atraso
Categoría A	0 – 30
Categoría B	31 – 60
Categoría C	61 – 90
Categoría D	91 – 120
Categoría E	121 – 150
Categoría F	151 – 180
Categoría G	> 181 días

Fuente: Elaboración Propia

Tabla 4

*Tipos de categorías*

### 3.2.2 Definición de buenos y malos clientes

Para poder establecer si un cliente es bueno o malo, se considera los acuerdos de Basilea, la propia experiencia de la entidad bancaria, la experiencia de otras entidades, la experiencia de los oficiales de crédito. Para ellos se pondrá mayor atención en el comportamiento de pago del cliente a lo largo del tiempo, para luego asignarlo como cliente bueno, malo, o indeterminado.

Se dice que un cliente es *bueno* si (Nieto, 2010):

- Cumple con el pago de su cuota mensual a tiempo, es decir aquel cliente que cancele dentro de sus límites de pago.
- Se encuentra dentro del intervalo de la categoría A y la categoría B, es decir que tenga como máximo 2 meses de atraso en sus cuotas.

Un cliente se denomina *indeterminado* si no es posible clasificarlo ni como buen ni como mal cliente. Esto puede ocurrir cuando un individuo adeuda cuotas tanto de capital como de interés acumulados, lo cual implica un mayor análisis del cliente para ver su comportamiento en los próximos meses.

Basados en Basilea (Bassel Commmite on Banking Supervision, 2005) se dice que un cliente es *malo* si el individuo no cancela sus deudas después de 90 días de atraso y,



por ende, causa pérdidas económicas a la institución financiera.

La figura 3 muestra la clasificación que se realizó con los clientes de la cartera de consumo considerando el máximo de los días de atrasos en los que no se canceló a tiempo el pago.



**Fuente:** Elaboración Propia  
 Figura 3. Estado de los Clientes

### 3.2.3 Análisis Roll Rate

Mediante el análisis Roll Rate se podrá definir la variable dependiente que tendrá el modelo de scoring. Además, se podrá identificar la categoría actual en la que se encuentre el cliente con respecto a la categoría siguiente, de tal forma que se obtenga una tasa de cambio de una categoría a otra.

La limitación de este estudio es que no se posee la información de atraso de clientes mayor a 6 meses, por lo tanto, es imprescindible realizar un análisis roll rate de la variable de marcación de mora (número máximo de días en atraso del cliente, ya sea, en los próximos 3 meses o en los próximos 6 meses a partir de la fecha de referencia), considerando 4 posibilidades para marcar a un determinado cliente como mal pagador:

- más de 30 días de atraso en los próximos 3 meses
- más de 30 días de atraso en los próximos 6 meses
- más de 60 días de atraso en los próximos 3 meses
- más de 60 días de atraso en los próximos 6 meses

El Roll Rate representa una matriz de transición de los clientes que proporciona una tabla de probabilidades condicionales, en la que se observó cómo cambiaban los días de atraso de los clientes en los próximos 6 meses con respecto a los días de atraso en los próximos 3 meses.

A continuación, se muestra la tabla roll rate con las correspondientes probabilidades de los clientes en cada categoría, en donde las filas representan los días de atraso en los próximos 3 meses a partir de la fecha de referencia y, en las columnas, se representa los días de atraso en los próximos 6 meses a partir de la fecha de referencia:

6 meses 3 meses	Categoría A	Categoría B	Categoría C	Categoría D	Categoría E	Categoría F	Categoría G
Categoría A	91.41 %	5.30 %	2.16 %	1.07 %	0.05 %	0.00 %	0.00 %
Categoría B	0.00 %	42.49 %	23.85 %	17.40 %	15.75 %	0.51 %	0.00 %
Categoría C	0.00 %	0.00 %	31.48 %	22.36 %	12.95 %	29.67 %	3.54 %
Categoría D	0.00 %	0.00 %	0.00 %	18.67 %	11.58 %	8.39 %	61.36 %
Categoría E	0.00 %	0.00 %	0.00 %	0.00 %	6.67 %	4.44 %	88.89 %

**Fuente:** Elaboración propia  
**Tabla 5**  
*Tabla de Probabilidades*

Lo que se está comparando es cómo cambian los días de atraso cuando los clientes van de 3 meses a 6 meses.

Y, una vez obtenida esta tabla, se obtuvo el porcentaje de clientes que se mantuvieron en la misma categoría (es decir aquellos que no migraron a otra categoría en los próximos 6 meses) y, el porcentaje de clientes que avanzaron a otra categoría (es decir aquellos que empeoraron como clientes ya que no están cumpliendo con su cuota de pago y por ende tienen más días de mora al pasar de los 3 meses a los 6 meses).

CATEGORÍAS	% de Clientes AVANZAN	% Clientes NO AVANZAN
Categoría A	8.6 %	91.4 %
Categoría B	57.5 %	42.5 %
Categoría C	68.5 %	31.5 %
Categoría D	81.3 %	18.7 %
Categoría E	93.3 %	6.7 %

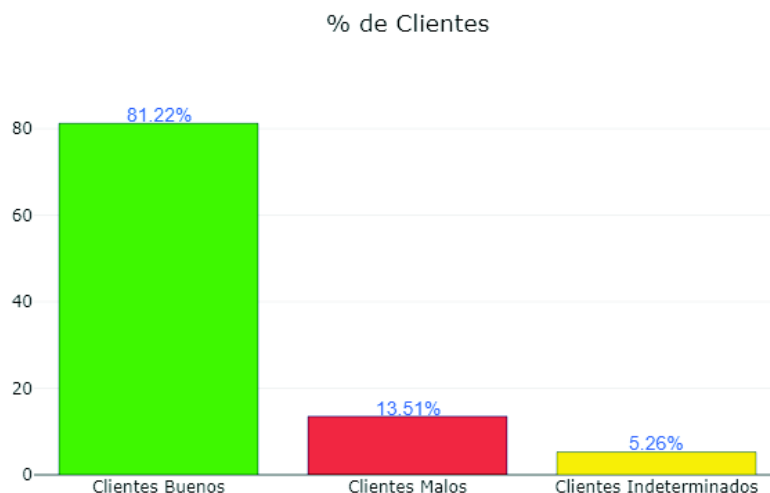
Fuente: Elaboración propia

Tabla 6

*Tabla Roll Rate*

Es por eso que en este trabajo se definió:

- Como clientes *malos* a aquellos que superan o igualan el 60 % de la tasa de avance (es decir los que se encuentran en la zona de color rojo). En otras palabras, aquellos clientes que se encuentran a partir de la Categoría C (más de 60 días de atraso) serán considerados como *malos*.
- Como clientes *indeterminados* a aquellos que su tasa de avance se encuentra entre 40 % y 59 % (es decir en la zona de color amarillo). Dicho de otra forma, aquellos clientes que se encuentran en la categoría B (entre 31 y 60 días de atraso) serán considerados como *indeterminados*.
- Como clientes *buenos* a aquellos que están fuera de los rangos antes mencionados (en la zona de color verde con un porcentaje de avance suficientemente pequeño). Esto quiere decir que los clientes *buenos* serán aquellos que se encuentran dentro de la categoría A (entre 0 y 30 días de atraso).



**Fuente:** Elaboración Propia

*Figura 4.* Porcentaje de Clientes Buenos, Malos e Indeterminados

### 3.3 Definición de la Variable Dependiente

Basándonos en el análisis roll rate de la sección anterior, donde se determinó como clientes *malos* a aquellos donde la tasa de cambio es mayor o igual al 60%, es decir, aquellos clientes que se encuentran a partir de la Categoría C. Por lo tanto, la variable de incumplimiento o "probabilidad de default PD" estará definida de la siguiente forma:

$$Y = PD = \begin{cases} 1 & \text{si el cliente tiene más de 60 días en atraso en los próximos 6 meses} \\ 0 & \text{caso contrario} \end{cases} \quad (3.1)$$

Esta variable  $Y$  que se desea predecir nos devolverá un resultado binario aplicando la regresión logística, dado que se conoce las características propias del individuo y variables macroeconómicas de Guatemala, las cuales serán tratadas como las variables independientes del modelo.

La variable dependiente  $Y$  se relaciona con las demás variables explicativas  $X_1, X_2, X_3, \dots, X_k$  mediante la ecuación 2.4 descrita en el capítulo anterior.

### 3.3.1 Comportamiento de la variable dependiente (PD)

La figura que se encuentra a continuación nos muestra la tasa de mora <sup>21</sup> durante el periodo de estudio:



**Fuente:** Elaboración Propia

*Figura 5.* Tasa de Mora observada a lo largo del tiempo

En la figura 5 se observa que el porcentaje de la tasa de mora, es mayor al inicio del estudio en el periodo 2015-07, mientras que en los siguientes meses se mantuvo una estabilidad de la tasa de mora en el tiempo observado.

La figura 5 nos muestra un gráfico de tasa la mora bastante difícil de modelar, pues solo se dispone de 5 periodos de estudio.

### 3.4 Tratamiento de la base de datos

En la sección anterior ya se definió la variable dependiente y, en base a esto, se obtuvo la tasa de mora mes a mes y, a su vez, se identificó en la base de datos cuáles serían los clientes buenos, malos e indeterminados.

Es importante mencionar que para realizar la modelización se descartó del análisis (no se los eliminó) a aquellos individuos que se consideran como indeterminados, por

<sup>21</sup>Es la cantidad de clientes malos dividido para el total de clientes.

lo que, no sabemos a ciencia cierta qué comportamiento tendrían, si fueran buenos o malos clientes. Este proceso se realiza para luego, en la validación del modelo, incluir estos clientes que son indeterminados y ver su comportamiento haciendo otro análisis de los resultados, con el fin de clasificar a los individuos indeterminados como buenos o malos.

La base de datos proporcionada fue dividida aleatoriamente en un 80 % para realizar la modelización, y un 20 % para realizar la validación del modelo.

Como ya se mencionó, la base de datos pertenece a una cartera de consumo, y que dispone de distintos tipos de información que hacen referencia a: información de comportamiento en atrasos, información del contrato, información de renovaciones, información de otros productos de crédito, información de productos de crédito en otros bancos, información de comportamiento crediticio, información idiosincrática, información socioeconómica, entre otras.

Además de las variables proporcionadas en la base, se tomaron en consideración algunas **variables macroeconómicas**, que son:

- Inflación Guatemala
- IMAE Guatemala
- IPC Guatemala

Por otro lado, aquellas variables que tenían más del 25 % de valores faltantes o inclusive nulos, se las descartó del análisis, pues no proporcionaban información relevante dentro del mismo.

### **3.5 Categorización de las Variables**

#### **3.5.1 Variables Cualitativas**

Para realizar la modelización, se optó por transformar las variables cualitativas en va-

riables *dummies* <sup>22</sup>, las cuales toman el valor de 0 o de 1 indicando la ausencia o presencia de alguna cualidad del individuo.

Además, a ciertas variables cualitativas se aplicó la técnica de recategorización. La cual se basa en agrupar las categorías o estados de una variable con el fin de optimizar y reducir el número de categorías, evaluando si una reducción en el número de categorías de la variable independiente es significativa o no, tomando en cuenta la variable dependiente (PD).

También, se hizo uso del análisis de clúster y la creación de *dummies* para las variables: Sexo, Tipo de vivienda, Escolaridad, y Estado Civil.

La variable Edad se dividió en 3 intervalos: jóvenes, adultos, y mayores y, después, se realizó las *dummies* de estas variables. El mismo procedimiento se aplicó a la variable Renta, subdividiendo en Renta Baja, Renta Media, y Renta Alta.

### 3.5.2 Variables Cuantitativas

Con respecto a las variables cuantitativas, se analizó la existencia de variabilidad en estas variables, con lo cual, a las variables cuantitativas que no tenían mucha variabilidad se decidió usar la técnica de recategorización, ya que, era más apropiado tratarlas como cualitativas estableciendo rangos de acuerdo a la variable dependiente.

Además, se verificó la presencia de valores perdidos o atípicos y, también, se trató la no linealidad realizando transformaciones cuando fue necesario.

## 3.6 Selección de las Variables

Para realizar la selección de variables que entrarán a los modelos, se aplicó las metodologías mencionadas en el marco teórico:

### 3.6.1 Information Value

Esta técnica del Information Value (IV) mencionado en el capítulo 2 del marco teórico, ayuda a determinar qué variables explicativas de la base de datos tienen poder

---

<sup>22</sup>Es una variable que toma el valor 0 o 1, con lo cual indica la categoría o estado al que pertenece cada individuo.

predictivo o si tienen influencia en el valor de la variable dependiente  $Y$ .

Como paso preliminar para la aplicación del modelo de regresión logística, se aplica este método con el fin de reducir lo mayor posible el número de variables explicativas que se introducirán en el modelo, ya que se dispone de 122 variables y nos interesa conocer cuál de ellas es potencialmente buena con respecto a la variable dependiente. Basándonos en la tabla 3 del capítulo Marco teórico, para la aplicación del modelo lo ideal sería elegir aquellas variables que tengan un poder predictivo bien fuerte, fuerte, y medio.

Por lo tanto, luego de hacer uso del software estadístico R, y de los paquetes respectivos, se obtuvieron los resultados del Information Value aplicado a las variables categóricas de la base de datos, y se tomaron en consideración las primeras 41 variables con mayor Information Value para realizar el análisis y la modelización, ya que estas variables poseían un poder predictivo bien fuerte, fuerte y medio. Las otras variables fueran descartadas ya que no poseían poder predictivo dentro del modelo.

### **3.6.2 Estadística KS**

Esta técnica se usó de igual manera que el Information Value, para hacer una correcta selección de las variables que entrarán al modelo. Con este método se calculó la estadística KS (Kolmogorov-Smirnov), con la cual se determina la eficacia que tienen los modelos. Mientras la estadística KS sea más grande, por ende, la variable analizada tendrá mayor poder predictivo y entonces el modelo será más eficaz (Prabhakaran, 2016).

De igual forma se utilizó el software estadístico R, y se calculó la estadística KS para las variables continuas con respecto a la variable dependiente. Se descartó aquellas variables con un estadístico KS muy bajo, pues no aportan dentro del modelo de regresión logística. Se eligieron aquellas variables continuas, con un valor de KS alto para que sean consideradas dentro del modelo, en este caso fueron seleccionadas 6 variables continuas, y las demás fueron descartadas.

### **3.6.3 Multicolinealidad**



Para evitar la multicolinealidad en el modelo, es decir, que no exista una alta correlación entre las variables explicativas del modelo se obtuvo la matriz de correlaciones de la base de datos nueva (solo con las 47 variables). Esto se lo realiza para medir los coeficientes de correlación entre pares de variables.

Con lo cual se observó que existen varias variables explicativas que están altamente correlacionadas entre sí, por lo que, se colocó un umbral del 70 %, para que en aquellas variables que exista una correlación superior al 70 % sea descartada una de las 2 variables en el estudio, tomando en cuenta el orden del Information Value de cada variable para su exclusión.

Luego de analizar la matriz de correlaciones y los coeficientes de cada par de variables, se llegó a la conclusión que deben ser excluidas 16 variables ya que poseen una alta correlación con respecto a otras variables, con lo cual nos quedarían 31 variables con las cuales finalmente se trabajará en la aplicación del modelo estadístico.

### **3.7 Descripción de variables seleccionadas**

#### **3.7.1 Variables Independientes**

Quedaron un total de 31 variables, las cuales tienen más influencia en la mora de un cliente, es decir, son las que mejor miden las características del cliente que provocan que los clientes incumplan con su pago.

Las variables seleccionadas son aquellas que están relacionadas con el comportamiento de pago de las cuotas del producto u otros productos de crédito, el tiempo de relacionamiento con el banco, los días de atraso y el saldo en atraso, el tamaño de la deuda, las características del producto, todas estas siendo variables idiosincráticas y sistémicas. Además, están variables sistémicas puras como el IPC, IMAE, Inflación de Guatemala y la tasa de interés; finalmente se encuentran variables idiosincráticas puras como género, escolaridad, edad, etc.

### 3.8 Creación del Modelo Scoring

Para hacer una correcta selección de las variables que deben ingresar al modelo, tomando en consideración que se tienen alrededor de 31 variables preseleccionadas, se hace uso del método de selección automática **Stepwise** para elegir las variables que mejor predicen la probabilidad de incumplimiento (Robert y Bendell, 2012). Luego de que se haya elegido las mejores variables que pueden ingresar al modelo mediante el stepwise, se corre el modelo y se vuelven a analizar los parámetros y el p-valor, considerando que las variables son significativas si el  $p\_valor \leq 0.05$ , por ende, se eliminarán aquellas variables con un  $p\_valor > 0.05$ . Este proceso se repite hasta obtener un modelo final con las variables adecuadas para cada modelo.

Se propone la creación de los modelos PIT, TTC e Híbrido para luego hacer una comparación entre estos modelos que usan distinto tipo de información y de esa forma entender las consecuencias de cada filosofía de clasificación en la práctica.

El modelo PIT será con información idiosincrática y sistémica de los clientes, pero sin considerar las variables sistémicas puras (variables macroeconómicas y políticas de crédito del banco). A este modelo PIT lo llamaremos **modelo PIT sin informaciones sistémicas puras**.

El modelo TTC será solamente con información idiosincrática de los clientes.

El modelo híbrido tendrá tanto información idiosincrática y sistémica, como también información sistémica pura. A este modelo Híbrido lo llamaremos **modelo PIT con informaciones sistémicas puras**.

#### 3.8.1 Modelo PIT sin informaciones sistémicas puras

Para la implementación de este sistema de clasificación, se usó información propia del cliente. También se incluyó información acerca del comportamiento del cliente en el contrato y en los demás productos de crédito, pero se descartó la información sistémica pura de los clientes y la información macroeconómica de Guatemala.

Se obtuvo los siguientes resultados mostrados en la tabla 7:

Variable	Categoría	Parámetro	p-valor	Std. Error
Intercepto		-2.413	<2e-16	0.095688
Disminución de cuotas pendientes (cantidad de de- crementos consecutivos)		0.07938	<2e-16	0.001667
Total de días en atraso		0.6164	<2e-16	0.057242
Saldo en atrasos cortos	<i>D4 &gt; 38.91</i>	-0.2647	5.14e-05	0.065371
	<i>D5 &gt; 112.91</i>	0.3023	1.43e-05	0.069661
Saldo en atrasos medios		0.4785	1.54e-13	0.064812
Tipo de residencia	arrendada	-	-	-
	propia	-	-	-
	prestada	-	-	-
	familiares	0.5524	<2e-16	0.0465
Deuda mensual activa	<i>D3 &gt; 162.528</i>	1.1745	<2e-16	0.073387
	<i>D4 &gt; 444.059</i>	1.3838	<2e-16	0.062852
	<i>D5 &gt; 754.146</i>	1.131	<2e-16	0.078253
Cantidad de productos de crédito cancelados		-0.3661	8.45e-11	0.056384
Tiempo de relacionamiento con el banco		-0.0084	1.22e-15	0.001045

Variable	Categoría	Parámetro	p-valor	Std. Error
Indicador de retanqueo (renovación)		-0.3275	6.68e-09	0.056484
% medio de cuotas pagadas en otros productos	D3 > 60.00	0.4475	4.65e-10	0.071826
	D6 > 864.91	0.3043	0.000106	0.078527
% de cuotas ya pagadas en el producto		-0.6617	2.98e-09	0.111537
Edad	Joven	0.1144	0.006782	0.04227
	Adulto	-	-	-
	Tercera Edad	-	-	-
Estado civil	soltero	0.2481	2.48e-09	0.041603
	casado	-	-	-
	desconocido	-	-	-
Grado de estudio	básica	-0.1767	3.72e-05	0.042855
	media	-	-	-
	superior	-	-	-
Saldo de crédito del cliente		0.6631	<2e-16	0.063221

Variable	Categoría	Parámetro	p-valor	Std. Error
Género	Masculino	-0.1094	0.006948	0.040546
	Femenino	-	-	-

Tabla 7  
*Resultados del Modelo PIT sin informaciones sistémicas puras*

Se observa en la tabla 7, que todas las variables del modelo PIT sin informaciones sistémicas puras son estadísticamente significativas, y en la siguiente sección se procede a la interpretación de los resultados.

### **Cálculo del VIF**

Se procedió al cálculo del factor de inflación de la varianza (VIF) para cada variable del modelo PIT sin informaciones sistémicas puras y se obtuvieron los siguientes resultados:

Variables	VIF
Disminución de cuotas pendientes	1.027
Total de días en atraso	1.268
Saldo en atrasos cortos	1.817
Saldo en atrasos medios	1.594
Tipo de residencia	1.1462
Deuda mensual activa	1.786
Cantidad de productos de crédito cancelados	1.4032
Tiempo de relacionamiento con el banco	1.324
Deuda mensual activa	1.286
Indicador de retanqueo	1.1139
% medio de cuotas pagadas en otros productos	1.596
% de cuotas ya pagadas en el producto	1.091
Edad	1.072
Estado civil	1.0352
Grado de estudio	1.012
Saldo de crédito del cliente	1.040
Género	1.019

Tabla 8

*VIF - Modelo PIT sin informaciones sistémicas puras*

Con los resultados previos, mostrados en la tabla 8, se verifica que no existe multicolinealidad entre las variables explicativas del modelo PIT sin informaciones sistémicas puras. Esto se concluye ya que los valores del VIF de las variables son bajos, el valor máximo de VIF fue apenas de 1.817.

### **3.8.2 Interpretación del Modelo PIT sin informaciones sistémicas puras**

Mediante la aplicación de la regresión logística, los parámetros de cada variable muestran el cambio en el logaritmo de la razón de probabilidades, es decir, cada parámetro (positivo o negativo) indica si la variable explicativa influye positiva o negativamente en la probabilidad de incumplimiento de un cliente haciendo que esta PD aumente o disminuya respectivamente.

- Un cliente que tiene una deuda mensual activa entre ]444.06, 754.15] unidades monetarias tiene 3.99 más veces de ser mal pagador que cualquier otro individuo con una deuda mensual activa entre ]162.53, 444.06] unidades monetarias.
- Si la variable Saldo de crédito del cliente se incrementa en una unidad, entonces en promedio la probabilidad de incumplimiento se incrementará en 66.31 %, con las demás variables constantes.
- Si el % de cuotas ya pagadas por parte del cliente se incrementa en una unidad, entonces eso impacta a que la probabilidad de incumplimiento disminuya en un 66.17 % en promedio, si se mantienen las otras variables constantes, indicando una relación negativa entre ambos.
- Considerando que el total de días de atraso que tiene un cliente en el pago de la cuota se incrementará en una unidad, esto provocará que la probabilidad de incumplimiento se incremente en 61.64 % en promedio, si las demás variables permanecen constantes; lo cual implica un riesgo bastante alto para la institución financiera.
- Un cliente con un tipo de residencia "vive con familiares" tiene 1.74 más veces de ser mal pagador que quienes tienen un tipo de residencia propia.

### 3.8.3 Modelo TTC

Para la implementación de este sistema de clasificación, se usó solo información propia del cliente, es decir solo información idiosincrática.

Los resultados se presentan en la tabla 9:

Variable	Categoría	Parámetro	p-valor	Std. Error
Intercepto		-2.307	<2e-16	0.05229

Variable	Categoría	Parámetro	p-valor	Std. Error
Tipo de residencia	arrendada	-	-	-
	propia	-	-	-
	prestada	-	-	-
	familiares	0.8407	<2e-16	0.04097
Edad	Joven	0.2923	1.45e-14	0.0380
	Adulto	-	-	-
	Tercera Edad	-	-	-
Estado civil	soltero	0.2257	2.74e-09	0.03796
	casado	-	-	-
	desconocido	-	-	-
Grado de estudio	básica	-0.1943	7.05e-07	0.03917
	media	-	-	-
	superior	-	-	-
Género	Masculino	-0.1840	6.82e-07	0.03705
	Femenino	-	-	-

Tabla 9  
Resultados del Modelo TTC

Se observa en la tabla 9, que todas las variables del modelo TTC son estadísticamente significativas, y en la siguiente sección se procede a la interpretación de los resultados.

### Cálculo del VIF

Se realizó el cálculo del factor de inflación de la varianza (VIF) para cada variable del



modelo TTC y se obtuvieron los siguientes resultados:

VARIABLES	VIF
Tipo de residencia	1.015
Edad	1.031
Estado Civil	1.028
Grado de estudio	1.008
Género	1.007

Tabla 10  
VIF - Modelo TTC

Con lo cual basados en la tabla 10, se verifica que no existe multicolinealidad entre las variables explicativas del modelo descrito. Esto se concluye ya que los valores del VIF de las variables son bajos, el valor máximo de VIF fue apenas de 1.031.

#### 3.8.4 Interpretación del Modelo TTC

La interpretación de los resultados se los plantea a continuación:

- Un cliente con un tipo de residencia "vive con familiares" tiene 2.32 más veces de ser mal pagador que quienes tienen un tipo de residencia propia.
- Un cliente con un rango de edad "joven" tiene 1.34 más veces de ser mal pagador que quienes tienen un rango de edad "adulta".
- Un cliente con estado civil soltero tiene 1.25 más veces de ser mal pagador que quienes tienen un estado civil casado.

#### 3.8.5 Modelo PIT con informaciones sistémicas puras

Para la implementación de este sistema de clasificación, se considero toda la información disponible, información agregada, e información macroeconómica de Guatemala.

Se obtuvieron los siguientes resultados presentados en la tabla: 11:

Variable	Categoría	Parámetro	p-valor	Std. Error
Intercepto		-2.507	<2e-16	9.82E-02

Variable	Categoría	Parámetro	p-valor	Std. Error
Disminución de cuotas pendientes (cantidad de decrementos consecutivos)		0.01948	<2e-16	1.67E-03
Total de días en atraso		0.6198	<2e-16	5.72E-02
Saldo en atrasos cortos	D4 > 38.91	-0.2734	2.92e-05	6.54E-02
	D5 > 112.91	0.2965	2.04e-05	6.96E-02
Saldo en atrasos medios		0.4775	1.71e-13	6.48E-02
Tipo de residencia	arrendada	-	-	-
	propia	-	-	-
	prestada	-	-	-
	familiares	0.5534	<2e-16	4.66E-02
Deuda mensual activa	D3 > 162.528	1.2040	<2e-16	7.39E-02
	D4 > 444.059	1.4140	<2e-16	6.34E-02
	D5 > 754.146	1.142	<2e-16	7.82E-02
Cantidad de productos de crédito cancelados		-0.3557	4.11e-10	5.69E-02
Tiempo de relacionamiento con el banco		-0.0078	7.58e-13	1.09E-03

Variable	Categoría	Parámetro	p-valor	Std. Error
Indicador de retanqueo (renovación)		-0.2927	2.68e-07	1.09E-03
% medio de cuotas pagadas en otros productos	D3 > 60.00	0.4783	3.85e-11	7.24E-02
	D6 > 864.91	0.2975	0.000151	7.85E-02
Monto bruto		0.0000	5.09e-05	8.78E-07
% de cuotas ya pagadas en el producto		-0.5929	1.31e-07	1.12E-01
Tasa de interés nominal		-0.1455	0.005232	5.21E-02
Edad	Joven	0.1141	0.007015	4.23E-02
	Adulto	-	-	-
	Tercera Edad	-	-	-
Estado civil	soltero	0.2494	2.21e-09	4.17E-02
	casado	-	-	-
	desconocido	-	-	-
Plazo (Número de cuotas en contrato)		-0.2318	0.002936	7.79E-02
Grado de estudio	básica	-0.1748	4.56e-05	4.29E-02
	media	-	-	-
	superior	-	-	-

Variable	Categoría	Parámetro	p-valor	Std. Error
Saldo de crédito del cliente		0.6790	<2e-16	6.35E-02
Género	Masculino	-0.1238	0.002357	4.07E-02
	Femenino	-	-	-

Tabla 11  
*Resultados del Modelo PIT con informaciones sistémicas puras*

Se observa en la tabla 11, que todas las variables del modelo PIT con informaciones sistémicas puras son estadísticamente significativas, y en la siguiente sección se procede a la interpretación de los resultados.

#### **Cálculo del VIF**

Se realizó el cálculo del factor de inflación de la varianza (VIF) para cada variable del modelo PIT con informaciones sistémicas puras y se obtuvieron los siguientes resultados:

Variables	VIF
Disminución de cuotas pendientes	1.028
Total de días en atraso	1.269
Saldo en atrasos cortos	1.819
Saldo en atrasos medios	1.594
Tipo de residencia	1.1467
Deuda mensual activa	1.800
Cantidad de productos de crédito cancelados	1.424
Tiempo de relacionamiento con el banco	1.422
Deuda mensual activa	1.808
Indicador de retanqueo	1.136
% medio de cuotas pagadas en otros productos	1.608
Monto bruto	1.108
% de cuotas ya pagadas en el producto	1.105
% medio de cuotas pagadas en otros productos	1.661
Tasa de interés nominal	1.165
Edad	1.073
Estado civil	1.0377
Plazo	1.038
Grado de Estudio	1.013
Saldo de crédito del cliente	1.044
Género	1.028

Tabla 12  
VIF - Modelo PIT con informaciones sistémicas puras

Con los resultados previos, mostrados en la tabla 12, se verifica que no existe multicolinealidad entre las variables explicativas del modelo PIT con informaciones sistémicas puras. Esto se concluye ya que los valores del VIF de las variables son bajos, el valor máximo de VIF fue apenas de 1.819.

### 3.8.6 Interpretación del Modelo PIT con informaciones sistémicas puras

Como ya se mencionó los parámetros de cada variable del modelo se interpretan como el cambio porcentual en el logaritmo de la razón de probabilidades.

- Un cliente que tiene una deuda mensual activa entre ]444.06, 754.15] unidades monetarias tiene 4.11 más veces de ser mal pagador que cualquier otro individuo con una deuda mensual activa entre ]162.53, 444.06] unidades monetarias.

- Ahora si la variable Saldo de crédito del cliente se incrementará en una unidad, entonces en promedio el logit estimado se incrementará en 67.90 %, con las demás variables constantes.

- Existe una relación positiva entre la variable total de días de atraso y la probabilidad de mora, pues si la variable mencionada si incrementa en una unidad, en promedio la probabilidad de mora aumentará en 61.98 %, siempre y cuando las demás variables se mantengan constantes.

- La relación entre la tasa de interés y la probabilidad de mora en este caso no era la esperada, pues se obtiene una relación negativa entre ambas. Esto se puede deber a la posible existencia de endogeneidad <sup>23</sup> de las variables, puede ser que este coeficiente salió negativo (-0.1455) por que no se hayan tomado en cuenta ciertas políticas de la institución financiera, etc.

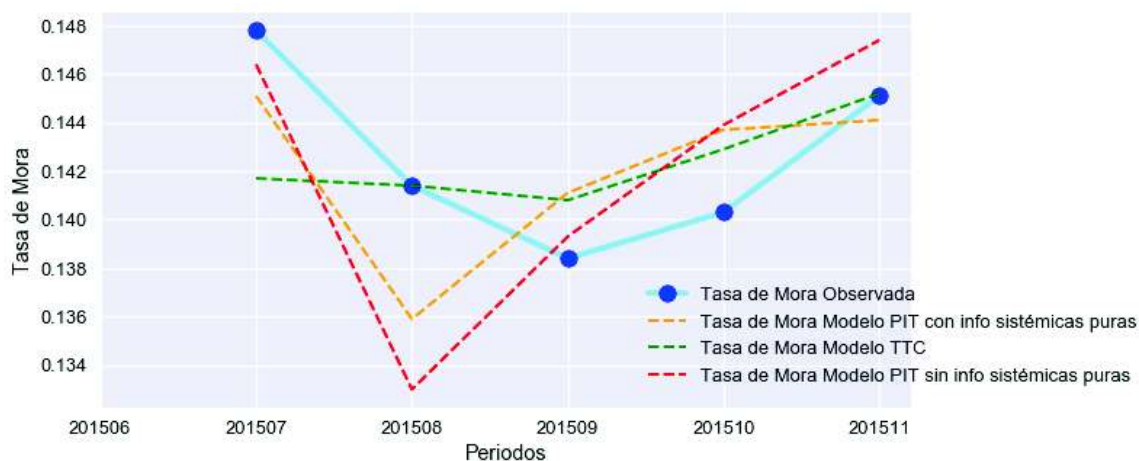
### 3.9 Resultados de los Modelos

En esta sección se presentan los resultados de modelar la probabilidad de mora con las distintas filosofías de clasificación. Se puede observar en el gráfico 6 la tasa de mora de observada, la tasa de mora para el modelo PIT sin informaciones sistémicas puras, TTC, y PIT con informaciones sistémicas puras a través del tiempo.

Con lo cual se puede decir que el modelo que mejor estima la tasa de mora observada, con solo observar el gráfico es el modelo PIT sin informaciones sistémicas puras y el modelo PIT con informaciones sistémicas puras, ya que tienen una tendencia parecida.

---

<sup>23</sup>Endogeneidad: Se refiere a que la relación entre una variable explicativa  $x$  y otra que queremos explicar  $y$  viene determinada por otras variables que no se han tenido en cuenta  $u$  y que covarían con  $x$ . De este modo,  $Cov(x, u) \neq 0$ .



Fuente: Elaboración Propia

Figura 6. Tasa de Mora en los distintos modelos

### 3.10 Validación del Modelo

#### 3.10.1 Poder de Discriminación de los Modelos

##### Estadística KS

Como ya se mencionó en el capítulo del Marco Teórico, se hace uso de la estadística KS para medir el poder predictivo de los sistemas de clasificación creados. Obteniéndose los siguientes resultados:

Modelos	Prueba	Validación
Modelo PIT sin info sistémicas puras	0.462	0.421
Modelo TTC	0.242	0.205
Modelo PIT con info sistémicas puras	0.4601	0.4199

Tabla 13

Estadística KS en el conjunto de Prueba y de Validación.

Basándonos en Mays (2004), quien menciona que un modelo de regresión logística es considerado un buen modelo, al tener un buen poder predictivo, si se obtiene una estadística KS entre el 20 % y el 70 %.

Por lo tanto, se concluye que el modelo PIT sin info sistémicas puras, al tener un KS del 46.2 % en el conjunto de prueba, y un KS del 42.1 % en el conjunto de validación, es considerado un buen modelo, que tiene una buena capacidad de clasificación.

El modelo TTC, tiene una estadística KS de 24.2 % en el conjunto de prueba, y una estadística KS de 20.52 % en el conjunto de validación, por lo tanto, se puede mencionar

que el KS para este modelo es bajo, pero, se mantiene dentro del rango de poder de discriminación.

Y, el modelo PIT con info sistémicas puras, al tener un KS del 46.01 % en el conjunto de prueba, y un KS de 41.99 % en el conjunto de validación; con lo cual, también es considerado un modelo con un buen poder predictivo.

### Área bajo la curva ROC

Se hace uso de la curva ROC y se calculó el Área bajo la curva (AUC) como medida de validación para diferenciar los buenos de los malos clientes, obteniéndose los siguientes resultados para cada modelo, tanto como para el conjunto de prueba (test), como para el conjunto de validación (con indeterminados):

Modelos	Prueba	Validación
Modelo PIT sin info sistémicas puras	0.792	0.7735
Modelo TTC	0.655	0.6322
Modelo PIT con info sistémicas puras	0.793	0.7754

Tabla 14  
*AUC en el conjunto de Prueba, y de Validación*

Para el conjunto de datos de prueba, es decir con los individuos que no fueron entrenados en los modelos; se observó que, para el modelo PIT sin informaciones sistémicas puras, la probabilidad de que un individuo seleccionado aleatoriamente sea clasificado correctamente por el modelo es del 79.2 %. En el modelo TTC, esa probabilidad es más baja, alrededor del 65.5 %. Y, en el modelo PIT con informaciones sistémicas puras, la probabilidad de que el modelo clasifique bien a un individuo elegido aleatoriamente es del 79.3 %.

De igual manera, para el conjunto de validación (clientes indeterminados), se concluye que los modelos propuestos son adecuados, pues los resultados tienen un comportamiento similar al conjunto de prueba. Por lo tanto, los modelos propuestos presentan un buen poder predictivo basándonos en el cálculo del área bajo la curva AUC.

### 3.10.2 Precisión y Error de los Modelos

Se calculó la precisión de los modelos antes descritos haciendo uso del estadístico  $R^2$ ,



el cual presenta un significado similar al coeficiente de determinación de la regresión lineal. Es decir, indica la proporción de clientes morosos que el modelo predice correctamente tanto en el conjunto de prueba test (sin indeterminados), como en el conjunto de validación (con los clientes indeterminados).

De igual forma se calculó la raíz del error cuadrático medio (RMSE), el cual mide el porcentaje de error que existe entre dos conjuntos de datos, es decir compara el valor de la predicción con el valor observado para cada modelo. Se presentan los resultados sin tomar en consideración los clientes indeterminados en la tabla 15:

Modelos	$R^2$ Ajustado	RMSE
Modelo PIT sin info sistémicas puras	86.83 %	0.3152577
Modelo TTC	80.98 %	0.3439934
Modelo PIT con info sistémicas puras	86.74 %	0.3155832

Tabla 15

*Resultados de los Modelos - SIN INDETERMINADOS*

El  $R^2$  de la tabla 15 nos indica que en el modelo PIT sin informaciones sistémicas puras, las variables independientes explican el 86.83 % de la variable dependiente. En el modelo TTC, indica que la variable dependiente es explicada por el 85.09 % por las variables independientes. Y, por último, en el modelo PIT con informaciones sistémicas puras muestra que el 86.74 % de las variables independientes explican la variable dependiente PD.

El RMSE muestra la proporción de clientes que son morosos y el modelo clasificó como clientes buenos, es decir, el porcentaje de error en el que incurrió el modelo al clasificar a los clientes. En el modelo PIT sin informaciones sistémicas puras, se presentó un RMSE de alrededor de 0.315; en el modelo TTC, se obtuvo un RMSE de 0.344; y, en el modelo PIT con informaciones sistémicas puras, 0.316. Con lo cual, se puede concluir que existe un mayor error en las predicciones con el modelo TTC.

Ahora se presentan los resultados considerando a los clientes indeterminados en la tabla 16:

El  $R^2$  muestra la precisión que posee cada modelo. En el modelo PIT sin informacio-

<b>Modelos</b>	<b>R<sup>2</sup> Ajustado</b>	<b>RMSE</b>
Modelo PIT sin info sistémicas puras	87.19 %	0.313791
Modelo TTC	81.76 %	0.3374658
Modelo PIT con info sistémicas puras	87.17 %	0.3138389

Tabla 16

*Resultados de los Modelos - CON INDETERMINADOS*

nes sistémicas puras, las variables independientes explican el 87.19 % de la variable dependiente. En el modelo TTC, indica que la variable dependiente es explicada por el 81.76 % por las variables independientes. Y, por último, en el modelo PIT con informaciones sistémicas puras muestra que el 87.17 % de las variables independientes explican la variable dependiente PD.

El RMSE devuelve el porcentaje de error en el que incurrió el modelo al clasificar a los clientes. En el modelo PIT sin informaciones sistémicas puras se presentó un RMSE de alrededor de 0.3137, en el modelo TTC se obtuvo un RMSE de 0.3374 y en el modelo PIT con informaciones sistémicas puras 0.3138. Con lo cual se puede concluir que, aunque se añada a los clientes INDETERMINADOS, se mantiene un mayor error en las predicciones con el modelo TTC.

## Capítulo 4

### Creación de Grupos de Riesgo

#### 4.1 Generación de los grupos homogéneos de riesgo

Para la creación de los grupos homogéneos de riesgo sólo se considerará el modelo PIT (PIT sin informaciones sistémicas puras) y el TTC.

Se desconsidera el modelo híbrido (PIT con informaciones sistémicas puras) por que tiene un desempeño parecido al modelo PIT, y, además, ya que el coeficiente de la variable tasa de interés no presentó el signo esperado en los resultados debido a la posible presencia de endogeneidad.

Para generar estos segmentos se tomó en consideración la PD estimada específica de cada cliente y de cada modelo respectivamente, y se hizo una partición de las PD en cuartiles para obtener los grupos de riesgo adecuados, de tal forma que los individuos que se encuentren en un mismo grupo compartan PD similares, y no exista una alta migración de clientes entre los grupos.

Como primer paso, se creó una nueva variable llamada "Período", la cual indica la fecha en la que se encuentra el cliente, considerando el intervalo de tiempo de estudio de la base de datos. Posteriormente, se dividió para cada período la PD estimada en cuartiles, luego se seleccionó los datos correspondientes de cada período estableciendo los puntos de corte para la división y, finalmente, se obtuvieron los grupos de riesgo. Estos grupos de riesgo creados tienen que cumplir ciertas características, deben ser homogéneos en riesgo, es decir, que todos los individuos de un mismo grupo deben tener PD similares; deben ordenar en relación al riesgo es decir que las PD agrupadas deben ser estadísticamente diferentes; y, además, se deben evitar concentraciones tanto en riesgo como en exposición, es decir, que hay que diversificar.

##### 4.1.1 Grupos de Riesgo: Modelo PIT

En la tabla 17 se muestra un resumen de cómo están distribuidos los grupos de riesgo

para el modelo PIT:

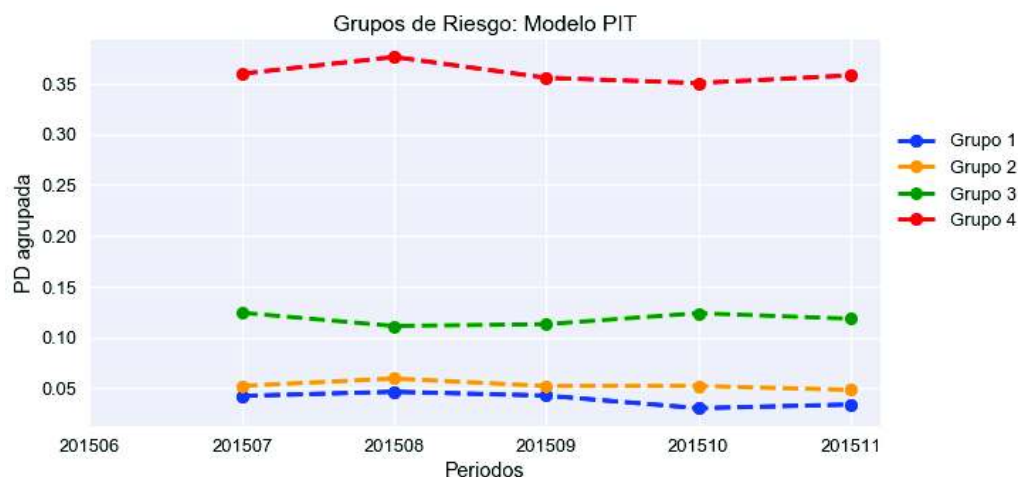
<b>Grupo 1</b>	
0	1
7419	302
<b>Grupo 2</b>	
0	1
7302	414
<b>Grupo 3</b>	
0	1
6802	914
<b>Grupo 4</b>	
0	1
4944	2773

Tabla 17

*Resumen de los grupos del modelo PIT*

Se obtuvo que la cantidad adecuada de grupos de riesgo para este modelo son 4. Posterior a esto, se asignó a cada grupo de riesgo las PD específicas de cada cliente, de acuerdo con una escala de calificación, la cual fue calculada mediante la partición de las PD en cuartiles. Esta escala está dada por los siguientes intervalos de probabilidad:  $(0.0086 - 0.0467)$ ,  $(.0467 - 0.0808)$ ,  $(0.0808 - 0.1755)$ ,  $(0.1755 - 0.9144)$ . Luego se calculó la tasa de mora promedio de cada grupo, para esto, se observó la frecuencia de incumplimientos de los clientes de cada grupo.

En el siguiente gráfico se presentan las PD agrupadas para los periodos de muestra, en los 4 grupos de riesgos creados bajo la filosofía PIT:



**Fuente:** Elaboración Propia

*Figura 7.* Grupos de Riesgo del Modelo PIT

En el gráfico 7 se observa la variación de las PD agrupadas para el modelo PIT a lo largo de los períodos observados. En general, los resultados muestran una sensibilidad<sup>24</sup> significativa de las migraciones de clientes entre los grupos de riesgo con el tiempo.

Los Grupos 1 y 2 son considerados de menor riesgo, el Grupo 3 es considerado un grupo de riesgo nivel medio, y el Grupo 4 es un grupo de riesgo alto.

#### 4.1.2 Grupos de Riesgo: Modelo TTC

A continuación, se presentan los resultados de cómo están distribuidos los grupos de riesgo para el modelo TTC:

<sup>24</sup>Se entiende por sensibilidad de una prueba a su capacidad para detectar diferencias o efectos allí donde los haya (Cárdenas y Arancibia, 2014).

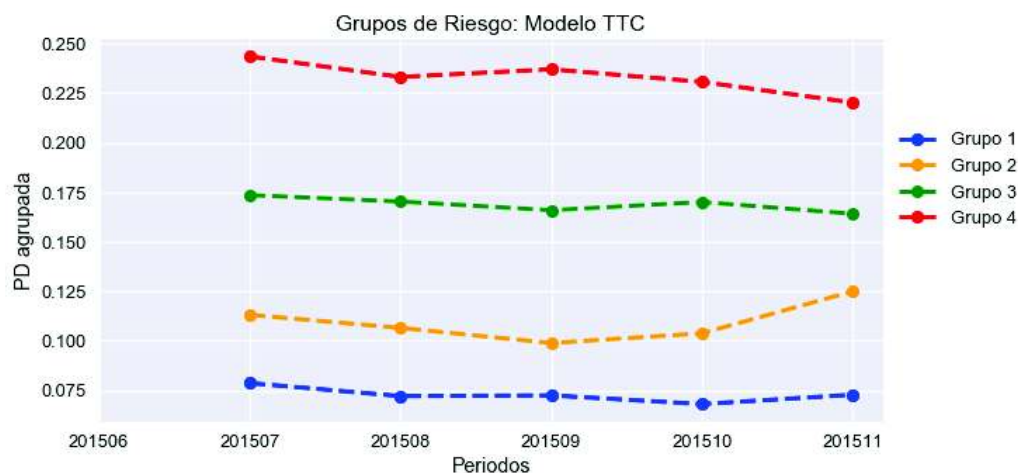
<b>Grupo 1</b>	
0	1
7588	601
<b>Grupo 2</b>	
0	1
7294	940
<b>Grupo 3</b>	
0	1
6374	1296
<b>Grupo 4</b>	
0	1
5211	1566

Tabla 18

*Resumen de los grupos del modelo TTC*

De igual forma se obtuvo que la cantidad de grupos de riesgo adecuados es 4. Luego se realizó la partición de las PD en cuartiles considerando los siguientes intervalos de probabilidad:  $(0.064 - 0.083)$ ,  $(0.083 - 0.136)$ ,  $(0.136 - 0.192)$ ,  $(0.192 - 0.279)$ . Posteriormente se realizó el cálculo de la tasa de mora promedio de cada grupo, tomando en consideración la frecuencia de incumplimientos observados que posee cada cliente en el grupo.

A continuación, se presenta el gráfico de las PD agrupadas en cada periodo de tiempo en los 4 grupos de riesgos creados:



Fuente: Elaboración Propia

Figura 8. Grupos de Riesgo del Modelo TTC

En el gráfico 8 se observa la variación de las PD agrupadas para el modelo TTC a lo largo de los periodos observados, y se observa un cambio en el gráfico con respecto al modelo PIT presentado previamente.

Se observa que los grupos de riesgo para el modelo TTC se encuentran más distantes, evitando las concentraciones en un solo grupo.

## 4.2 Validación de los grupos de riesgo

### 4.2.1 Pruebas de Comparaciones Múltiples de Diferencias de Medias

Luego de la generación de los grupos homogéneos de riesgo (creados en la sección anterior), se debe comprobar que esos grupos son válidos. Para eso, se realiza un análisis de comparación múltiple de diferencias de medias utilizando las pruebas de Tukey y de Dunnet T3.

#### 4.2.1.1 Prueba de Tukey

##### *Modelo PIT*

En la siguiente tabla 19 se presentan los resultados de aplicar la prueba de Tukey a los grupos de riesgo del modelo PIT:

Tukey multiple comparisons of means  
95% family-wise confidence level

Fit: `aov(formula = x ~ f)`

Grupos	diff	lwr	upr	p adj
2-1	0.02885431	0.02567868	0.03202994	0
3-1	0.08655359	0.08337796	0.08972923	0
4-1	0.31952357	0.31634804	0.32269910	0
3-2	0.05769929	0.05452314	0.06087543	0
4-2	0.29066926	0.28749322	0.29384531	0
4-3	0.23296998	0.22979394	0.23614602	0

Tabla 19

*Prueba de Tukey para los grupos del modelo PIT*

Con lo cual se concluye que, al hacer la comparación de diferencia de medias por pares, (haciendo uso de la Prueba de Tukey) se obtiene un  $p - adj < 0.05$  y, por lo tanto, los resultados son significativos y, por ende, los grupos de riesgo para el modelo PIT son válidos. Y también se concluye que las medias son diferentes en cada grupo de riesgo.

### *Modelo TTC*

Ahora se presentan los resultados de la Prueba de Tukey aplicados a los grupos de riesgo del modelo TTC:

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: `aov(formula = x ~ f)`



Grupos	diff	lwr	upr	p adj
2-1	0.04329467	0.04260329	0.04398604	0
3-1	0.09735886	0.09665492	0.09806280	0
4-1	0.14933355	0.14860606	0.15006104	0
3-2	0.05406419	0.05336119	0.05476720	0
4-2	0.10603888	0.10531229	0.10676547	0
4-3	0.05197469	0.05123614	0.05271324	0

Tabla 20

*Prueba de Tukey para los grupos del modelo TTC*

Los resultados muestran que los grupos de riesgo son válidos, con un  $p - adj < 0.05$ . Y por tanto se rechaza la hipótesis nula y se concluye que las medias de los grupos de riesgo son distintas para el modelo TTC.

#### 4.2.1.2 Prueba de Dunnet T3

##### *Modelo PIT*

En la tabla 21 se presentan los resultados de aplicar la prueba de Dunnet T3 a los grupos de riesgo del modelo PIT:

Dunnett's test for comparing several treatments with a control:

95% family-wise confidence level

Grupos	diff	lwr.ci	upr.ci	pval
1-4	-0.3195236	-0.3224274	-0.3166198	<2e-16
2-4	-0.2906693	-0.2935735	-0.2877650	<2e-16
3-4	-0.2329700	-0.2358743	-0.2300657	<2e-16

Tabla 21

*Prueba de Dunnet T3 para los grupos del modelo PIT*

Se puede concluir que los grupos de riesgo para el modelo PIT son válidos, pues al observar los resultados de la prueba de comparación de medias Dunnet T3 se obtiene un valor  $pval < 0.05$ . Entonces se rechaza la hipótesis nula, con lo cual, se puede concluir que las medias de los grupos son diferentes en el modelo PIT.

### *Modelo TTC*

Ahora se presentan los resultados de la Prueba de Dunnett T3 aplicados a los grupos de riesgo del modelo TTC:

Dunnett's test for comparing several treatments with a control:

95% family-wise confidence level

Grupos	diff	lwr.ci	upr.ci	pval
1-4	-0.14933355	-0.14999650	-0.14867059	<2e-16
2-4	-0.10603888	-0.10670102	-0.10537675	<2e-16
3-4	-0.05197469	-0.05264772	-0.05130165	<2e-16

Tabla 22

*Prueba de Dunnett T3 para los grupos del modelo TTC*

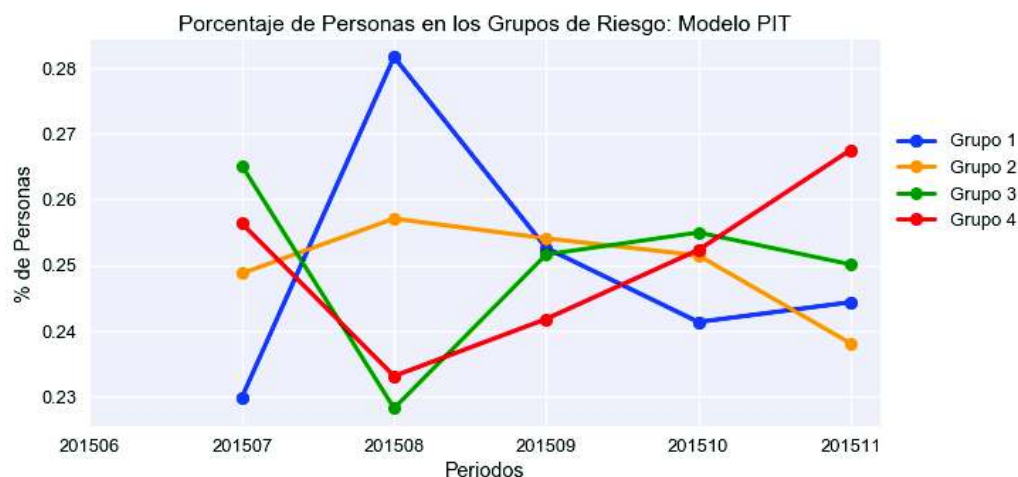
Los resultados al aplicar la prueba de Dunnett T3 muestran que los grupos de riesgo son válidos, ya que el  $pval < 0.05$  y, por ende, se rechaza la hipótesis nula, y se puede concluir que las medias de los grupos son diferentes.

### **4.3 Estabilidad de los grupos de riesgo**

Una vez que ya se comprobó que los grupos de riesgo creados son válidos para cada modelo, se procede a analizar la estabilidad que tendrán los clientes en los distintos grupos de riesgo y en cada modelo. Esto con el fin de tener una visión más clara de la estabilidad que posee cada sistema de clasificación durante un auge o recesión económica.

#### **4.3.1 Modelo PIT**

Para observar la estabilidad que posee este sistema de clasificación durante fluctuaciones económicas, se presenta el siguiente gráfico que muestra la variación porcentual de los clientes entre los grupos de riesgo bajo la filosofía PIT:



**Fuente:** Elaboración Propia

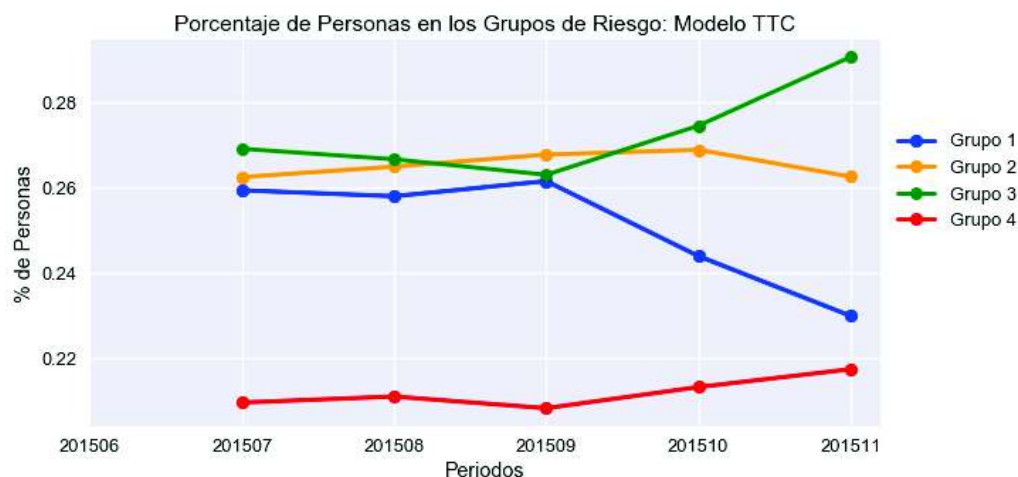
*Figura 9.* Porcentaje de clientes entre los grupos de riesgo: PIT

La Figura 9 muestra una gran variación de los clientes en los distintos grupos de riesgo, es decir, se percibe una alta migración de clientes entre grupos.

En el periodo 201508 se observa que los clientes han migrado considerablemente a los grupos de menor riesgo (Grupo 1 y Grupo 2). En los periodos 201509 y 201510 los clientes tienen una baja migración entre los grupos. Y en el periodo 201511 se observa que los clientes han migrado a los grupos de mayor riesgo (Grupo 4). Finalmente, gracias al gráfico, se observa que el porcentaje de clientes entre grupos para el modelo PIT es muy variable lo cual conlleva a que los clientes estén migrando entre estos grupos a través del tiempo.

#### 4.3.2 Modelo TTC

En la Figura 10 se muestra la variación porcentual de los clientes entre los grupos de riesgo bajo la filosofía TTC:



**Fuente:** Elaboración Propia

*Figura 10.* Porcentaje de clientes entre los grupos de riesgo: TTC

La Figura 10 muestra que con este sistema de clasificación los clientes permanecen más estables a lo largo del tiempo de estudio, es decir, que la migración entre grupos bajo esta filosofía es mínima.

En los períodos 201507, 201508, y 201519 se observa que los clientes permanecen estables, sin presentar una migración entre los grupos de riesgo. En los períodos 201510 y 201511 se observa una pequeña migración de los clientes de menor riesgo (Grupo 1) hacia un grupo de mayor riesgo (Grupo 3). En general, se observa que el porcentaje de clientes entre grupos para el modelo TTC es estable, con lo cual, los clientes van a permanecer en el mismo grupo, sin migrar a otros de mayor riesgo.

## Capítulo 5

### Conclusiones y recomendaciones

#### 5.1 Conclusiones

- Basados en Anderson (2007) y Mays (2004), los modelos PIT obtenidos son considerados modelos con un buen poder predictivo, ya que, el KS y el % de AUC, están dentro de los rangos establecidos (en el capítulo *Marco Teórico* página 40) para indicar que los modelos son buenos.
  - El modelo PIT sin informaciones sistémicas puras, tiene un KS de 0.462 y un % de AUC de 79.2 %.
  - El modelo TTC, tiene un KS de 0.242 y un % de AUC de 65.5 %.
  - El modelo PIT con informaciones sistémicas puras, tiene un KS de 0.4601 y un % de AUC de 79.3 %.

Aunque el KS del modelo TTC es bajo, está todavía dentro del rango de poder de discriminación. Sin embargo, para ser un modelo de comportamiento la verdad es un KS bastante bajo, pero es esperado ya que el modelo TTC se basa en características idiosincráticas de los clientes (las cuales no van a variar mucho en el tiempo).

Por lo tanto, para este tipo de cartera y con el histórico que se tenía disponible sería mejor usar el modelo PIT sin informaciones sistémicas puras; aunque, después la institución financiera tendrá que hacer una validación, es decir, decidir si quiere un capital procíclico o acíclico.

- La principal diferencia de los sistemas de clasificación propuestos PIT y TTC, es cómo los clientes son clasificados de acuerdo al tipo de información que usa cada sistema. Además, se puede concluir a partir de los resultados vistos en la página 71 del documento, basados en la figura 6, en la tabla 13 (que muestra los resultados del KS) y en tabla 14 (que muestra los resultados del % de AUC) que:

- El sistema de clasificación PIT tiene un buen poder predictivo, según Anderson (2007) y Mays (2004). Y, también, se observa que bajo este sistema de clasificación se tiene una alta migración de los clientes entre los grupos, a través del tiempo. Lo cual se puede observar en la página 83 del documento en la figura 9.
- Por el contrario, se observó que el sistema de clasificación TTC tiene una menor capacidad predictiva que el PIT, lo que es esperado, porque el sistema TTC busca solo ordenar los clientes y no predecir la PD. Al analizar la estabilidad en este sistema de clasificación, se obtuvo que los clientes permanecen más estables en los grupos de riesgo a través del tiempo. Lo cual se puede observar en la página 84 del documento en la figura 10.
- Se observó que existen dos tipos de PD: la PD del cliente, la cual fue obtenida a través del modelo de regresión logística y, la PD del grupo (PD agrupada), obtenida a través de la creación de los grupos de riesgo. Analizamos los resultados obtenidos para estos dos tipos de PD.

### **Comportamiento de la PD de los clientes**

- Sistema de clasificación PIT: En este sistema si la economía mejora, entonces, la PD disminuirá, ya que el PIT si es susceptible a las fluctuaciones de la economía o de políticas de crédito, debido a que posee información sistémica y macroeconómica.
- Sistema de clasificación TTC: Este sistema no se ve afectado por las fluctuaciones de la economía, ya que, solo usa información idiosincrática de los clientes. Por lo tanto, la PD se mantiene constante, bajo este sistema de clasificación.

### **Comportamiento de la PD del grupo (PD agrupada)**

- Sistema de clasificación PIT: Se puede concluir que las PD de los grupos tienden a acompañar las fluctuaciones económicas o de políticas de crédito y, por tanto, los clientes de un mismo grupo migrarán a un grupo de mayor o menor riesgo según el caso. Los resultados se muestran en la figura 9 en el capítulo *Creación de grupos*.

- Sistema de clasificación TTC: Este sistema, al no verse afectado por los cambios en la economía ni de políticas de crédito, los clientes no migrarán de grupos y se mantendrán en el mismo grupo de riesgo establecido, por tanto, la PD del grupo se mantiene constante. Los resultados de esta conclusión se encuentran en la figura 10.
- Cada sistema de clasificación proporciona diferentes impactos en precificación y asignación de capital. Al comparar entre PIT y TTC se pudo observar que en un sistema PIT los clientes migran y, por ende, el capital es procíclico, mientras que en el sistema TTC, los clientes se mantienen estables y, por lo tanto, el capital es estable, es decir acíclico.
- Al validar los modelos creados con el grupo de "clientes indeterminados", se concluye que los modelos predicen muy bien a este segmento de clientes, obteniéndose un KS y un % de AUC bastante parecidos a los que se obtuvieron con el grupo de clientes sin indeterminados.  
Para el sistema de clasificación PIT sin informaciones sistémicas puras, se obtuvo un KS de 0.421 y un % de AUC de 77.35 %.  
Para el sistema de clasificación TTC, se obtuvo un KS de 0.205 y un % de AUC de 63.22 %.  
Para el sistema de clasificación PIT con informaciones sistémicas puras, se obtuvo un KS de 0.4199 y un % de AUC de 77.54 %.
- Basilea sugiere que se creen grupos homogéneos de riesgo y, basados en los resultados obtenidos en el capítulo de *Creación de grupos riesgo*, se evidenció la existencia de grupos de riesgo homogéneos en PD y heterogéneos entre ellos en PD, tanto para el sistema de clasificación PIT y para el sistema TTC. Lo que sustenta el hecho de que clientes que pertenecen a un mismo grupo de riesgo, deben tener la misma PD (o similar), y diferente a la de los otros grupos de riesgo, de acuerdo a lo que establece Basilea.  
Conforme se muestra en la figura 7 y figura 8, y mediante las pruebas realizadas para la validación de los grupos de riesgo que se muestran en las tablas 19, 20, 21 y 22; se comprobó que los grupos de riesgo creados son válidos.

## 5.2 Recomendaciones

- Se recomienda para futuros estudios conseguir información de la institución financiera con un intervalo de tiempo de, por lo menos, un año. Además, incentivar a los bancos para que guarden su información y así obtener modelos robustos; donde, quizá la metodología no cambie, pero, sí podría ayudar a obtener coeficientes robustos, al tener mayor cantidad de datos.
- Se recomienda hacer evaluaciones comparativas (benchmarking) como complemento de la evaluación estadística formal de los sistemas de calificación interna planteados. Esto con el fin de que el banco determine la aceptación o rechazo de los sistemas propuestos, considerando sus prácticas internas de gestión de riesgos. Esta evaluación comparativa se puede realizar comparando las estimaciones internas de la PD de un conjunto común de clientes entre distintos bancos.
- Se recomienda, como siguiente paso, estimar los otros componentes de riesgo: LGD (pérdida dada por incumplimiento), y EAD (exposición en el incumplimiento) para determinar la alocaión de capital de la institución financiera, dada una determinada exposición. Todo esto con el fin de medir la pérdida esperada y la pérdida inesperada que tendría la institución.
- Para gestionar de forma correcta la prociclicidad se puede trabajar sobre la cantidad de grupos homogéneos que se crean para cada modelo, evitando que un solo grupo de riesgo pueda concentrar todas las exposiciones de la cartera tanto en riesgo como en exposición.



# Anexos

## Anexo 1:

### Categorización de Variables - Código R

```

# Lectura de la base de clientes
# (sin considerar los interminados)
basesin <- read.xlsx("datos_sin.xlsx")

# Lectura de la base donde están
# los deciles que tiene cada variable:
b2s <- read.xlsx('deciles.xlsx')

# Nuevo data frame donde se guardará
# las variables categorizadas.
b1s <- data.frame()

#####
# Recategorización, y variables dummies:
#####

##### Var07
var07.uniq <- unique(
c(0, b2s$Var07, max(b2s$Var07)))

# Dummies para la var07
b1s[, "var07.1"] <- as.numeric(
basesin$Var07_NumPer>=var07.uniq[1]
& basesin$Var07_NumPer<=var07.uniq[2])
b1s[, "var07.2"] <- as.numeric(
basesin$Var07_NumPer>var07.uniq[2]
& basesin$Var07_NumPer<=var07.uniq[3])
b1s[, "var07.3"] <- as.numeric(
basesin$Var07_NumPer>var07.uniq[3]
& basesin$Var07_NumPer<=var07.uniq[4])
b1s[, "var07.4"] <- as.numeric(
basesin$Var07_NumPer>var07.uniq[4]
& basesin$Var07_NumPer<=var07.uniq[5])

##### Var09
var09.uniq <- unique(
c(0, b2s$Var09, max(b2s$Var09)))

# Dummies para la var09
b1s[, "var09.1"] <- as.numeric(
basesin$Var09_TasNom>=var09.uniq[1]
& basesin$Var09_TasNom<=var09.uniq[2])
b1s[, "var09.2"] <- as.numeric(
basesin$Var09_TasNom>var09.uniq[2]
& basesin$Var09_TasNom<=var09.uniq[3])
b1s[, "var09.3"] <- as.numeric(
basesin$Var09_TasNom>var09.uniq[3]
& basesin$Var09_TasNom<=var09.uniq[4])
b1s[, "var09.4"] <- as.numeric(
basesin$Var09_TasNom>var09.uniq[4]
& basesin$Var09_TasNom<=var09.uniq[5])
b1s[, "var09.5"] <- as.numeric(
basesin$Var09_TasNom>var09.uniq[5]
& basesin$Var09_TasNom<=var09.uniq[6])
b1s[, "var09.6"] <- as.numeric(
basesin$Var09_TasNom>var09.uniq[6]
& basesin$Var09_TasNom<=var09.uniq[7])
b1s[, "var09.7"] <- as.numeric(
basesin$Var09_TasNom>var09.uniq[7]
& basesin$Var09_TasNom<=var09.uniq[8])
b1s[, "var09.8"] <- as.numeric(
basesin$Var09_TasNom>var09.uniq[8]
& basesin$Var09_TasNom<=var09.uniq[9])
b1s[, "var09.9"] <- as.numeric(
basesin$Var09_TasNom>var09.uniq[9]
& basesin$Var09_TasNom<=var09.uniq[10])

##### Var10
var10.uniq <- unique(
c(-1, b2s$Var10, max(b2s$Var10)))

# Dummies para la var10
b1s[, "var10.1"] <- as.numeric(
basesin$Var10_SalTot>=var10.uniq[1]
& basesin$Var10_SalTot<=var10.uniq[2])
b1s[, "var10.2"] <- as.numeric(
basesin$Var10_SalTot>var10.uniq[2]
& basesin$Var10_SalTot<=var10.uniq[3])
b1s[, "var10.3"] <- as.numeric(
basesin$Var10_SalTot>var10.uniq[3]

```

```

& basesin$Var10_SalTot<=var10.uniq[4])

##### Var36
var36.uniq <- unique(
c(-1, b2s$Var36, max(b2s$Var36)))

# Dummies para la var36
b1s[, "var36.1"] <- as.numeric(
basesin$Var36_NumCuoVen>=
var36.uniq[1] &
basesin$Var36_NumCuoVen<=var36.uniq[2])
b1s[, "var36.2"] <- as.numeric(
basesin$Var36_NumCuoVen>
var36.uniq[2] &
basesin$Var36_NumCuoVen<=var36.uniq[3])

##### Var37
var37.uniq <- unique(
c(-1, b2s$Var37, max(b2s$Var37)))

# Dummies para la var37
b1s[, "var37.1"] <- as.numeric(
basesin$Var37_SalVen>=var37.uniq[1]
& basesin$Var37_SalVen<=var37.uniq[2])
b1s[, "var37.2"] <- as.numeric(
basesin$Var37_SalVen>var37.uniq[2]
& basesin$Var37_SalVen<=var37.uniq[3])

##### Var111
var111.uniq <- unique(
c(-1, b2s$Var111, max(b2s$Var111)))

# Dummies para la var111
b1s[, "var111.1"] <- as.numeric(
basesin$Var111_NumCuo>=
var111.uniq[1] &
basesin$Var111_NumCuo<=var111.uniq[2])
b1s[, "var111.2"] <- as.numeric(
basesin$Var111_NumCuo>
var111.uniq[2] &
basesin$Var111_NumCuo<=var111.uniq[3])
b1s[, "var111.3"] <- as.numeric(
basesin$Var111_NumCuo>
var111.uniq[3] &
basesin$Var111_NumCuo<=var111.uniq[4])

##### Var38_Atr3
var38_Atr3.uniq <- unique(
c(-1, b2s$Var38_Atr3,
max(b2s$Var38_Atr3)))

# Dummies para la Var38_Atr3
b1s[, "var38_Atr3.1"] <- as.numeric(
basesin$Var38_Atr3Max>=
var38_Atr3.uniq[1] &
basesin$Var38_Atr3Max<=
var38_Atr3.uniq[2])
b1s[, "var38_Atr3.2"] <- as.numeric(
basesin$Var38_Atr3Max>
var38_Atr3.uniq[2] &
basesin$Var38_Atr3Max<=
var38_Atr3.uniq[3])
b1s[, "var38_Atr3.3"] <- as.numeric(
basesin$Var38_Atr3Max>
var38_Atr3.uniq[3] &
basesin$Var38_Atr3Max<=
var38_Atr3.uniq[4])
b1s[, "var38_Atr3.4"] <- as.numeric(
basesin$Var38_Atr3Max>
var38_Atr3.uniq[4] &
basesin$Var38_Atr3Max<=
var38_Atr3.uniq[5])
b1s[, "var38_Atr3.5"] <- as.numeric(
basesin$Var38_Atr3Max>
var38_Atr3.uniq[5] &
basesin$Var38_Atr3Max<=
var38_Atr3.uniq[6])
b1s[, "var38_Atr3.6"] <- as.numeric(
basesin$Var38_Atr3Max>
var38_Atr3.uniq[6] &
basesin$Var38_Atr3Max<=
var38_Atr3.uniq[7])

##### Var38_Atr6
var38_Atr6.uniq <- unique(
c(-1, b2s$Var38_Atr6, max(b2s$Var38_Atr6)))

# Dummies para la Var38_Atr6
b1s[, "var38_Atr6.1"] <- as.numeric(
basesin$Var38_Atr6Max>=
var38_Atr6.uniq[1] &
basesin$Var38_Atr6Max<=
var38_Atr6.uniq[2])
b1s[, "var38_Atr6.2"] <- as.numeric(
basesin$Var38_Atr6Max>
var38_Atr6.uniq[2] &
basesin$Var38_Atr6Max<=

```

```

var38_Atr6.uniq[3])
b1s[, "var38_Atr6.3"] <- as.numeric(
basesin$Var38_Atr6Max>
var38_Atr6.uniq[3] &
basesin$Var38_Atr6Max<=
var38_Atr6.uniq[4])
b1s[, "var38_Atr6.4"] <- as.numeric(
basesin$Var38_Atr6Max>
var38_Atr6.uniq[4] &
basesin$Var38_Atr6Max<=
var38_Atr6.uniq[5])
b1s[, "var38_Atr6.5"] <- as.numeric(
basesin$Var38_Atr6Max>
var38_Atr6.uniq[5] &
basesin$Var38_Atr6Max<=
var38_Atr6.uniq[6])
b1s[, "var38_Atr6.6"] <- as.numeric(
basesin$Var38_Atr6Max>
var38_Atr6.uniq[6] &
basesin$Var38_Atr6Max<=
var38_Atr6.uniq[7])
b1s[, "var38_Atr6.7"] <- as.numeric(
basesin$Var38_Atr6Max>
var38_Atr6.uniq[7] &
basesin$Var38_Atr6Max<=
var38_Atr6.uniq[8])

##### Var40_Atr3
###Variable con missings:
# si el %de misssing es mayor al 25%
# eliminar esas variables, caso contrario
# se categorizó, donde una categoría es
# la variable con valores y otra categoría
# serían los missings.
b1s[, "Var40_Atr3.1"] <- as.numeric(is.na(
basesin$Var40_Atr3Ant)==FALSE)
b1s[, "Var40_Atr3.2"] <- as.numeric(is.na(
basesin$Var40_Atr3Ant)==TRUE)

##### Var41_Atr3
b1s[, "Var41_Atr3.1"] <- as.numeric(is.na(
basesin$Var41_Atr3Mes)==FALSE)
b1s[, "Var41_Atr3.2"] <- as.numeric(is.na(
basesin$Var41_Atr3Mes)==TRUE)

#####Var41_Atr6
basesin$Var41_Atr6Mes[ is.na(
basesin$Var41_Atr6Mes)] <- 0

var41_Atr6.uniq <- unique(
c(-1, b2s$Var41_Atr6,
max(b2s$Var41_Atr6)))

# Dummies para la Var38_Atr3
b1s[, "var41_Atr6.1"] <- as.numeric(
basesin$Var41_Atr6Mes>=
var41_Atr6.uniq[1] &
basesin$Var41_Atr6Mes<=
var41_Atr6.uniq[2])
b1s[, "var41_Atr6.2"] <- as.numeric(
basesin$Var41_Atr6Mes>
var41_Atr6.uniq[2] &
basesin$Var41_Atr6Mes<=
var41_Atr6.uniq[3])

##### Var44_Atr3
var44_Atr3.uniq <- unique(
c(-1, b2s$Var44_Atr3,
max(b2s$Var44_Atr3)))

# Dummies para la Var44_Atr3
b1s[, "var44_Atr3.1"] <- as.numeric(
basesin$Var44_Atr3SalPro>=
var44_Atr3.uniq[1] &
basesin$Var44_Atr3SalPro<=
var44_Atr3.uniq[2])
b1s[, "var44_Atr3.2"] <- as.numeric(
basesin$Var44_Atr3SalPro>
var44_Atr3.uniq[2] &
basesin$Var44_Atr3SalPro<=
var44_Atr3.uniq[3])
b1s[, "var44_Atr3.3"] <- as.numeric(
basesin$Var44_Atr3SalPro>
var44_Atr3.uniq[3] &
basesin$Var44_Atr3SalPro<=var44_Atr3.uniq[4])
b1s[, "var44_Atr3.4"] <- as.numeric(
basesin$Var44_Atr3SalPro>
var44_Atr3.uniq[4] &
basesin$Var44_Atr3SalPro<=
var44_Atr3.uniq[5])
b1s[, "var44_Atr3.5"] <- as.numeric(
basesin$Var44_Atr3SalPro>
var44_Atr3.uniq[5] &
basesin$Var44_Atr3SalPro<=
var44_Atr3.uniq[6])
b1s[, "var44_Atr3.6"] <- as.numeric(

```

```

basesin$Var44_Atr3SalPro>
var44_Atr3.uniq[6] &
basesin$Var44_Atr3SalPro<=
var44_Atr3.uniq[7])

##### Var44_Atr6
var44_Atr6.uniq <- unique(
c(-1, b2s$Var44_Atr6, max(b2s$Var44_Atr6)))

# Dummies para la Var44_Atr6
b1s[, "var44_Atr6.1"] <- as.numeric(
basesin$Var44_Atr6SalPro>=
var44_Atr6.uniq[1] &
basesin$Var44_Atr6SalPro<=
var44_Atr6.uniq[2])
b1s[, "var44_Atr6.2"] <- as.numeric(
basesin$Var44_Atr6SalPro>
var44_Atr6.uniq[2] &
basesin$Var44_Atr6SalPro<=
var44_Atr6.uniq[3])
b1s[, "var44_Atr6.3"] <- as.numeric(
basesin$Var44_Atr6SalPro>
var44_Atr6.uniq[3] &
basesin$Var44_Atr6SalPro<=
var44_Atr6.uniq[4])
b1s[, "var44_Atr6.4"] <- as.numeric(
basesin$Var44_Atr6SalPro>
var44_Atr6.uniq[4] &
basesin$Var44_Atr6SalPro<=
var44_Atr6.uniq[5])
b1s[, "var44_Atr6.5"] <- as.numeric(
basesin$Var44_Atr6SalPro>
var44_Atr6.uniq[5] &
basesin$Var44_Atr6SalPro<=
var44_Atr6.uniq[6])
b1s[, "var44_Atr6.6"] <- as.numeric(
basesin$Var44_Atr6SalPro>
var44_Atr6.uniq[6] &
basesin$Var44_Atr6SalPro<=
var44_Atr6.uniq[7])
b1s[, "var44_Atr6.7"] <- as.numeric(
basesin$Var44_Atr6SalPro>
var44_Atr6.uniq[7] &
basesin$Var44_Atr6SalPro<=
var44_Atr6.uniq[8])

##### Var45_Atr3
var45_Atr3.uniq <- unique(
c(-1, b2s$Var45_Atr3,

```

```

max(b2s$Var45_Atr3)))

# Dummies para la Var45_Atr3
b1s[, "var45_Atr3.1"] <- as.numeric(
basesin$Var45_Atr3SalMax>=
var45_Atr3.uniq[1] &
basesin$Var45_Atr3SalMax<=
var45_Atr3.uniq[2])
b1s[, "var45_Atr3.2"] <- as.numeric(
basesin$Var45_Atr3SalMax>
var45_Atr3.uniq[2] &
basesin$Var45_Atr3SalMax<=
var45_Atr3.uniq[3])
b1s[, "var45_Atr3.3"] <- as.numeric(
basesin$Var45_Atr3SalMax>
var45_Atr3.uniq[3] &
basesin$Var45_Atr3SalMax<=
var45_Atr3.uniq[4])
b1s[, "var45_Atr3.4"] <- as.numeric(
basesin$Var45_Atr3SalMax>
var45_Atr3.uniq[4] &
basesin$Var45_Atr3SalMax<=
var45_Atr3.uniq[5])
b1s[, "var45_Atr3.5"] <- as.numeric(
basesin$Var45_Atr3SalMax>
var45_Atr3.uniq[5] &
basesin$Var45_Atr3SalMax<=
var45_Atr3.uniq[6])
b1s[, "var45_Atr3.6"] <- as.numeric(
basesin$Var45_Atr3SalMax>
var45_Atr3.uniq[6] &
basesin$Var45_Atr3SalMax<=
var45_Atr3.uniq[7])

##### Var45_Atr6
var45_Atr6.uniq <- unique(
c(-1, b2s$Var45_Atr6,
max(b2s$Var45_Atr6)))

# Dummies para la Var45_Atr6
b1s[, "var45_Atr6.1"] <- as.numeric(
basesin$Var45_Atr6SalMax>=
var45_Atr6.uniq[1] &
basesin$Var45_Atr6SalMax<=
var45_Atr6.uniq[2])
b1s[, "var45_Atr6.2"] <- as.numeric(
basesin$Var45_Atr6SalMax>
var45_Atr6.uniq[2] &
basesin$Var45_Atr6SalMax<=

```

```

var45_Atr6.uniq[3])
b1s[, "var45_Atr6.3"] <- as.numeric(
basesin$Var45_Atr6SalMax>
var45_Atr6.uniq[3] &
basesin$Var45_Atr6SalMax<=
var45_Atr6.uniq[4])
b1s[, "var45_Atr6.4"] <- as.numeric(
basesin$Var45_Atr6SalMax>
var45_Atr6.uniq[4] &
basesin$Var45_Atr6SalMax<=
var45_Atr6.uniq[5])
b1s[, "var45_Atr6.5"] <- as.numeric(
basesin$Var45_Atr6SalMax>
var45_Atr6.uniq[5] &
basesin$Var45_Atr6SalMax<=
var45_Atr6.uniq[6])
b1s[, "var45_Atr6.6"] <- as.numeric(
basesin$Var45_Atr6SalMax>
var45_Atr6.uniq[6] &
basesin$Var45_Atr6SalMax<=
var45_Atr6.uniq[7])
b1s[, "var45_Atr6.7"] <- as.numeric(
basesin$Var45_Atr6SalMax>
var45_Atr6.uniq[7] &
basesin$Var45_Atr6SalMax<=
var45_Atr6.uniq[8])

##### Var46
var46.uniq <- unique(
c(-1, b2s$Var46, max(b2s$Var46)))

# Dummies para la Var46
b1s[, "var46.1"] <- as.numeric(
basesin$Var46_CanProCre>=var46.uniq[1] &
basesin$Var46_CanProCre<=var46.uniq[2])
b1s[, "var46.2"] <- as.numeric(
basesin$Var46_CanProCre>var46.uniq[2] &
basesin$Var46_CanProCre<=var46.uniq[3])
b1s[, "var46.3"] <- as.numeric(
basesin$Var46_CanProCre>var46.uniq[3] &
basesin$Var46_CanProCre<=var46.uniq[4])

##### Var48
var48.uniq <- unique(
c(-1, b2s$Var48, max(b2s$Var48)))

# Dummies para la Var48
b1s[, "var48.1"] <- as.numeric(
basesin$Var48_SalAtr0ProCre>=var48.uniq[1] &
basesin$Var48_SalAtr0ProCre<=var48.uniq[2])
b1s[, "var48.2"] <- as.numeric(
basesin$Var48_SalAtr0ProCre>var48.uniq[2] &
basesin$Var48_SalAtr0ProCre<=var48.uniq[3])

##### Var49
var49.uniq <- unique(
c(-1, b2s$Var49, max(b2s$Var49)))

# Dummies para la Var49
b1s[, "var49.1"] <- as.numeric(
basesin$Var49_SalAtr15ProCre>=var49.uniq[1] &
basesin$Var49_SalAtr15ProCre<=var49.uniq[2])
b1s[, "var49.2"] <- as.numeric(
basesin$Var49_SalAtr15ProCre>var49.uniq[2] &
basesin$Var49_SalAtr15ProCre<=var49.uniq[3])

##### Var50
var50.uniq <- unique(
c(-1, b2s$Var50, max(b2s$Var50)))

# Dummies para la Var50
b1s[, "var50.1"] <- as.numeric(
basesin$Var50_SalAtr30ProCre>=var50.uniq[1] &
basesin$Var50_SalAtr30ProCre<=var50.uniq[2])
b1s[, "var50.2"] <- as.numeric(
basesin$Var50_SalAtr30ProCre>var50.uniq[2] &
basesin$Var50_SalAtr30ProCre<=var50.uniq[3])

##### Var56_Can3
var56_Can3.uniq <- unique(
c(-1, b2s$Var56_Can3, max(b2s$Var56_Can3)))

# Dummies para la Var56_Can3
b1s[, "var56_Can3.1"] <- as.numeric(
basesin$Var56_Can3IncSal>=var56_Can3.uniq[1] &
basesin$Var56_Can3IncSal<=var56_Can3.uniq[2])
b1s[, "var56_Can3.2"] <- as.numeric(
basesin$Var56_Can3IncSal>var56_Can3.uniq[2] &
basesin$Var56_Can3IncSal<=var56_Can3.uniq[3])

##### Var56_Can6
var56_Can6.uniq <- unique(
c(-1, b2s$Var56_Can6, max(b2s$Var56_Can6)))

# Dummies para la Var56_Can6
b1s[, "var56_Can6.1"] <- as.numeric(
basesin$Var56_Can6IncSal>=var56_Can6.uniq[1] &
basesin$Var56_Can6IncSal<=var56_Can6.uniq[2])

```

```

b1s[, "var56_Can6.2"] <- as.numeric(
basesin$Var56_Can6IncSal>var56_Can6.uniq[2] &
basesin$Var56_Can6IncSal<=var56_Can6.uniq[3])
b1s[, "var56_Can6.3"] <- as.numeric(
basesin$Var56_Can6IncSal>var56_Can6.uniq[3] &
basesin$Var56_Can6IncSal<=var56_Can6.uniq[4])

##### Var57_Can3
var57_Can3.uniq <- unique(
c(-1, b2s$Var57_Can3, max(b2s$Var57_Can3)))

# Dummies para la Var57_Can3
b1s[, "var57_Can3.1"] <- as.numeric(
basesin$Var57_Can3DecSal>=var57_Can3.uniq[1] &
basesin$Var57_Can3DecSal<=var57_Can3.uniq[2])
b1s[, "var57_Can3.2"] <- as.numeric(
basesin$Var57_Can3DecSal>var57_Can3.uniq[2] &
basesin$Var57_Can3DecSal<=var57_Can3.uniq[3])
b1s[, "var57_Can3.3"] <- as.numeric(
basesin$Var57_Can3DecSal>var57_Can3.uniq[3] &
basesin$Var57_Can3DecSal<=var57_Can3.uniq[4])

##### Var57_Can6
var57_Can6.uniq <- unique(
c(-1, b2s$Var57_Can6, max(b2s$Var57_Can6)))

# Dummies para la Var57_Can6
b1s[, "var57_Can6.1"] <- as.numeric(
basesin$Var57_Can6DecSal>=var57_Can6.uniq[1] &
basesin$Var57_Can6DecSal<=var57_Can6.uniq[2])
b1s[, "var57_Can6.2"] <- as.numeric(
basesin$Var57_Can6DecSal>var57_Can6.uniq[2] &
basesin$Var57_Can6DecSal<=var57_Can6.uniq[3])
b1s[, "var57_Can6.3"] <- as.numeric(
basesin$Var57_Can6DecSal>var57_Can6.uniq[3] &
basesin$Var57_Can6DecSal<=var57_Can6.uniq[4])

##### Var58_Can3
var58_Can3.uniq <- unique(
c(-1, b2s$Var58_Can3, max(b2s$Var58_Can3)))

# Dummies para la Var58_Can3
b1s[, "var58_Can3.1"] <- as.numeric(
basesin$Var58_Can3IncCas>=var58_Can3.uniq[1] &
basesin$Var58_Can3IncCas<=var58_Can3.uniq[2])
b1s[, "var58_Can3.2"] <- as.numeric(
basesin$Var58_Can3IncCas>var58_Can3.uniq[2] &
basesin$Var58_Can3IncCas<=var58_Can3.uniq[3])

##### Var58_Can6
var58_Can6.uniq <- unique(
c(-1, b2s$Var58_Can6, max(b2s$Var58_Can6)))

# Dummies para la Var58_Can6
b1s[, "var58_Can6.1"] <- as.numeric(
basesin$Var58_Can6IncCas>=var58_Can6.uniq[1] &
basesin$Var58_Can6IncCas<=var58_Can6.uniq[2])
b1s[, "var58_Can6.2"] <- as.numeric(
basesin$Var58_Can6IncCas>var58_Can6.uniq[2] &
basesin$Var58_Can6IncCas<=var58_Can6.uniq[3])

##### Var59_Can3
var59_Can3.uniq <- unique(
c(-1, b2s$Var59_Can3, max(b2s$Var59_Can3)))

# Dummies para la Var59_Can3
b1s[, "var59_Can3.1"] <- as.numeric(
basesin$Var59_Can3DecCas>=var59_Can3.uniq[1] &
basesin$Var59_Can3DecCas<=var59_Can3.uniq[2])
b1s[, "var59_Can3.2"] <- as.numeric(
basesin$Var59_Can3DecCas>var59_Can3.uniq[2] &
basesin$Var59_Can3DecCas<=var59_Can3.uniq[3])

##### Var59_Can6
var59_Can6.uniq <- unique(
c(-1, b2s$Var59_Can6, max(b2s$Var59_Can6)))

# Dummies para la Var59_Can6
b1s[, "var59_Can6.1"] <- as.numeric(
basesin$Var59_Can6DecCas>=var59_Can6.uniq[1] &
basesin$Var59_Can6DecCas<=var59_Can6.uniq[2])
b1s[, "var59_Can6.2"] <- as.numeric(
basesin$Var59_Can6DecCas>var59_Can6.uniq[2] &
basesin$Var59_Can6DecCas<=var59_Can6.uniq[3])

##### Var60_Can3
b1s[, "Var60_Can3.1"] <- as.numeric(is.na(
basesin$Var60_Can3ProCre)==FALSE)
b1s[, "Var60_Can3.2"] <- as.numeric(is.na(
basesin$Var60_Can3ProCre)==TRUE)

##### Var60_Can6
b1s[, "Var60_Can6.1"] <- as.numeric(is.na(
basesin$Var60_Can6ProCre)==FALSE)
b1s[, "Var60_Can6.2"] <- as.numeric(is.na(
basesin$Var60_Can6ProCre)==TRUE)

```

```

##### Var65
var65.uniq <- unique(
c(-1, b2s$Var65, max(b2s$Var65)))

# Dummies para la var65
b1s[, "var65.1"] <- as.numeric(
basesin$Var65_CanProCreRef>=var65.uniq[1] &
basesin$Var65_CanProCreRef<=var65.uniq[2])
b1s[, "var65.2"] <- as.numeric(
basesin$Var65_CanProCreRef>var65.uniq[2] &
basesin$Var65_CanProCreRef<=var65.uniq[3])

##### Var69_Alia
var69_Alia.uniq <- unique(
c(-1, b2s$Var69_Alia,
max(b2s$Var69_Alia)))

# Dummies para la var69_Alia
b1s[, "var69_Alia.1"] <- as.numeric(
basesin$Var69_Alianza>=var69_Alia.uniq[1] &
basesin$Var69_Alianza<=var69_Alia.uniq[2])
b1s[, "var69_Alia.2"] <- as.numeric(
basesin$Var69_Alianza>var69_Alia.uniq[2] &
basesin$Var69_Alianza<=var69_Alia.uniq[3])

##### Var69_Cons
var69_Cons.uniq <- unique(
c(-1, b2s$Var69_Cons,
max(b2s$Var69_Cons)))

# Dummies para la var69_Cons
b1s[, "var69_Cons.1"] <- as.numeric(
basesin$Var69_Consumo>=var69_Cons.uniq[1] &
basesin$Var69_Consumo<=var69_Cons.uniq[2])
b1s[, "var69_Cons.2"] <- as.numeric(
basesin$Var69_Consumo>var69_Cons.uniq[2] &
basesin$Var69_Consumo<=var69_Cons.uniq[3])

##### Var69_IncoJapon
var69_IncoJapon.uniq <- unique(
c(-1, b2s$Var69_Inco,
max(b2s$Var69_Inco)))

# Dummies para la Var69_IncoJapon
b1s[, "var69_IncoJapon.1"] <- as.numeric(
basesin$Var69_IncomelJapon>=
var69_IncoJapon.uniq[1] &
basesin$Var69_IncomelJapon<=

var69_IncoJapon.uniq[2])
b1s[, "var69_IncoJapon.2"] <- as.numeric(
basesin$Var69_IncomelJapon>
var69_IncoJapon.uniq[2] &
basesin$Var69_IncomelJapon<=
var69_IncoJapon.uniq[3])

##### Var69_Micr
var69_Micr.uniq <- unique(
c(-1, b2s$Var69_Micr,
max(b2s$Var69_Micr)))

# Dummies para la Var69_Micr
b1s[, "var69_Micr.1"] <- as.numeric(
basesin$Var69_MicroCredito>=var69_Micr.uniq[1] &
basesin$Var69_MicroCredito<=var69_Micr.uniq[2])
b1s[, "var69_Micr.2"] <- as.numeric(
basesin$Var69_MicroCredito>var69_Micr.uniq[2] &
basesin$Var69_MicroCredito<=var69_Micr.uniq[3])

##### Var70_Alia
var70_Alia.uniq <- unique(
c(-1, b2s$Var70_Alia, max(b2s$Var70_Alia)))

# Dummies para la var70_Alia
b1s[, "var70_Alia.1"] <- as.numeric(
basesin$Var70_Alianza>=var70_Alia.uniq[1] &
basesin$Var70_Alianza<=var70_Alia.uniq[2])
b1s[, "var70_Alia.2"] <- as.numeric(
basesin$Var70_Alianza>var70_Alia.uniq[2] &
basesin$Var70_Alianza<=var70_Alia.uniq[3])

##### Var70_Cons
var70_Cons.uniq <- unique(
c(-1, b2s$Var70_Cons, max(b2s$Var70_Cons)))

# Dummies para la var70_Cons
b1s[, "var70_Cons.1"] <- as.numeric(
basesin$Var70_Consumo>=var70_Cons.uniq[1] &
basesin$Var70_Consumo<=var70_Cons.uniq[2])
b1s[, "var70_Cons.2"] <- as.numeric(
basesin$Var70_Consumo>var70_Cons.uniq[2] &
basesin$Var70_Consumo<=var70_Cons.uniq[3])

##### Var70_IncoJapon
var70_IncoJapon.uniq <- unique(
c(-1, b2s$Var70_IncomelJa,
max(b2s$Var70_IncomelJa)))

```



```

# Dummies para la Var70_IncoJapon
b1s[, "var70_IncoJapon.1"] <- as.numeric(
basesin$Var70_IncomelJapon>=
var70_IncoJapon.uniq[1] &
basesin$Var70_IncomelJapon<=
var70_IncoJapon.uniq[2])
b1s[, "var70_IncoJapon.2"] <- as.numeric(
basesin$Var70_IncomelJapon>
var70_IncoJapon.uniq[2] &
basesin$Var70_IncomelJapon<=
var70_IncoJapon.uniq[3])

##### Var70_IncomelCC
var70_IncomelCC.uniq <- unique(
c(-1, b2s$Var70_IncomelCC,
max(b2s$Var70_IncomelCC)))

# Dummies para la Var70_IncomelCC
b1s[, "var70_IncomelCC.1"] <- as.numeric(
basesin$Var70_IncomelCCo>=
var70_IncomelCC.uniq[1] &
basesin$Var70_IncomelCCo<=
var70_IncomelCC.uniq[2])
b1s[, "var70_IncomelCC.2"] <- as.numeric(
basesin$Var70_IncomelCCo>
var70_IncomelCC.uniq[2] &
basesin$Var70_IncomelCCo<=
var70_IncomelCC.uniq[3])

##### Var70_Micr
var70_Micr.uniq <- unique(
c(-1, b2s$Var70_Micr, max(b2s$Var70_Micr)))

# Dummies para la Var70_Micr
b1s[, "var70_Micr.1"] <- as.numeric(
basesin$Var70_MicroCredito>=
var70_Micr.uniq[1] &
basesin$Var70_MicroCredito<=
var70_Micr.uniq[2])
b1s[, "var70_Micr.2"] <- as.numeric(
basesin$Var70_MicroCredito>
var70_Micr.uniq[2] &
basesin$Var70_MicroCredito<=
var70_Micr.uniq[3])

##### Var70_Tarj
var70_Tarj.uniq <- unique(
c(-1, b2s$Var70_Tarj, max(b2s$Var70_Tarj)))

# Dummies para la Var70_Tarj
b1s[, "var70_Tarj.1"] <- as.numeric(
basesin$Var70_TarjetaCredito>=
var70_Tarj.uniq[1] &
basesin$Var70_TarjetaCredito<=
var70_Tarj.uniq[2])
b1s[, "var70_Tarj.2"] <- as.numeric(
basesin$Var70_TarjetaCredito>
var70_Tarj.uniq[2] &
basesin$Var70_TarjetaCredito<=
var70_Tarj.uniq[3])

##### Var71_Alia
var71_Alia.uniq <- unique(
c(-1, b2s$Var71_Alia, max(b2s$Var71_Alia)))

# Dummies para la var71_Alia
b1s[, "var71_Alia.1"] <- as.numeric(
basesin$Var71_Alianza>=
var71_Alia.uniq[1] &
basesin$Var71_Alianza<=
var71_Alia.uniq[2])
b1s[, "var71_Alia.2"] <- as.numeric(
basesin$Var71_Alianza>
var71_Alia.uniq[2] &
basesin$Var71_Alianza<=
var71_Alia.uniq[3])

##### Var71_Cons
var71_Cons.uniq <- unique(
c(-1, b2s$Var71_Cons, max(b2s$Var71_Cons)))

# Dummies para la var71_Cons
b1s[, "var71_Cons.1"] <- as.numeric(
basesin$Var71_Consumo>=
var71_Cons.uniq[1] &
basesin$Var71_Consumo<=
var71_Cons.uniq[2])
b1s[, "var71_Cons.2"] <- as.numeric(
basesin$Var71_Consumo>
var71_Cons.uniq[2] &
basesin$Var71_Consumo<=
var71_Cons.uniq[3])

```



```

##### Var71_IncoJapon
var71_IncoJapon.uniq <- unique(
c(-1, b2s$Var71_IncomelJa ,
max(b2s$Var71_IncomelJa)))

# Dummies para la Var71_IncoJapon
b1s["var71_IncoJapon.1"] <- as.numeric(
basesin$Var71_IncomelJapon>=
var71_IncoJapon.uniq[1] &
basesin$Var71_IncomelJapon<=
var71_IncoJapon.uniq[2])
b1s["var71_IncoJapon.2"] <- as.numeric(
basesin$Var71_IncomelJapon>
var71_IncoJapon.uniq[2] &
basesin$Var71_IncomelJapon<=
var71_IncoJapon.uniq[3])

##### Var71_IncomelCC
var71_IncomelCC.uniq <- unique(
c(-1, b2s$Var71_IncomelCC,
max(b2s$Var71_IncomelCC)))

# Dummies para la Var71_IncomelCC
b1s["var71_IncomelCC.1"] <- as.numeric(
basesin$Var71_IncomelCCo>=
var71_IncomelCC.uniq[1] &
basesin$Var71_IncomelCCo<=
var71_IncomelCC.uniq[2])
b1s["var71_IncomelCC.2"] <- as.numeric(
basesin$Var71_IncomelCCo>
var71_IncomelCC.uniq[2] &
basesin$Var71_IncomelCCo<=
var71_IncomelCC.uniq[3])

##### Var71_Micr
var71_Micr.uniq <- unique(
c(-1, b2s$Var71_Micr, max(b2s$Var71_Micr)))

# Dummies para la Var71_Micr
b1s["var71_Micr.1"] <- as.numeric(
basesin$Var71_MicroCredito>=
var71_Micr.uniq[1] &
basesin$Var71_MicroCredito<=
var71_Micr.uniq[2])
b1s["var71_Micr.2"] <- as.numeric(
basesin$Var71_MicroCredito>
var71_Micr.uniq[2] &
basesin$Var71_MicroCredito<=
var71_Micr.uniq[3])

##### Var71_Tarj
var71_Tarj.uniq <- unique(
c(-1, b2s$Var71_Tarj, max(b2s$Var71_Tarj)))

# Dummies para la Var71_Tarj
b1s["var71_Tarj.1"] <- as.numeric(
basesin$Var71_TarjetaCredito>=
var71_Tarj.uniq[1] &
basesin$Var71_TarjetaCredito<=
var71_Tarj.uniq[2])
b1s["var71_Tarj.2"] <- as.numeric(
basesin$Var71_TarjetaCredito>
var71_Tarj.uniq[2] &
basesin$Var71_TarjetaCredito<=
var71_Tarj.uniq[3])

##### Var115_Alia
var115_Alia.uniq <- unique(
c(-1, b2s$Var115_Alia,
max(b2s$Var115_Alia)))

# Dummies para la var115_Alia
b1s["var115_Alia.1"] <- as.numeric(
basesin$Var115_Alianza>=
var115_Alia.uniq[1] &
basesin$Var115_Alianza<=
var115_Alia.uniq[2])
b1s["var115_Alia.2"] <- as.numeric(
basesin$Var115_Alianza>
var115_Alia.uniq[2] &
basesin$Var115_Alianza<=
var115_Alia.uniq[3])

##### Var115_Cons
var115_Cons.uniq <- unique(c(-1, b2s$Var115_Cons,
max(b2s$Var115_Cons)))

# Dummies para la var115_Cons
b1s["var115_Cons.1"] <- as.numeric(
basesin$Var115_Consumo>=
var115_Cons.uniq[1] &
basesin$Var115_Consumo<=

```

```

var115_Cons.uniq[2])
b1s[, "var115_Cons.2"] <- as.numeric(
basesin$Var115_Consumo>
var115_Cons.uniq[2] &
basesin$Var115_Consumo<=
var115_Cons.uniq[3])

##### Var115_IncoJapon
var115_IncoJapon.uniq <- unique(
c(-1, b2s$Var115_IncomelJa ,
max(b2s$Var115_IncomelJa )))
var115_IncoJapon.uniq

# Dummies para la Var115_IncoJapon
b1s[, "var115_IncoJapon.1"] <- as.numeric(
basesin$Var115_IncomelJapon>=
var115_IncoJapon.uniq[1] &
basesin$Var115_IncomelJapon<=
var115_IncoJapon.uniq[2])
b1s[, "var115_IncoJapon.2"] <- as.numeric(
basesin$Var115_IncomelJapon>
var115_IncoJapon.uniq[2] &
basesin$Var115_IncomelJapon<=
var115_IncoJapon.uniq[3])

##### Var115_IncomelCC
var115_IncomelCC.uniq <- unique(
c(-1, b2s$Var115_IncomelCC ,
max(b2s$Var115_IncomelCC)))

# Dummies para la Var115_IncomelCC
b1s[, "var115_IncomelCC.1"] <- as.numeric(
basesin$Var115_IncomelCCo>=
var115_IncomelCC.uniq[1] &
basesin$Var115_IncomelCCo<=
var115_IncomelCC.uniq[2])

b1s[, "var115_IncomelCC.2"] <- as.numeric(
basesin$Var115_IncomelCCo>
var115_IncomelCC.uniq[2] &
basesin$Var115_IncomelCCo<=
var115_IncomelCC.uniq[3])

##### Var115_Micr
var115_Micr.uniq <- unique(
c(-1, b2s$Var115_Micr ,
max(b2s$Var115_Micr)))

# Dummies para la Var115_Micr
b1s[, "var115_Micr.1"] <- as.numeric(
basesin$Var115_MicroCredito>=
var115_Micr.uniq[1] &
basesin$Var115_MicroCredito<=
var115_Micr.uniq[2])
b1s[, "var115_Micr.2"] <- as.numeric(
basesin$Var115_MicroCredito>
var115_Micr.uniq[2] &
basesin$Var115_MicroCredito<=
var115_Micr.uniq[3])

##### Var115_Tarj
var115_Tarj.uniq <- unique(
c(-1, b2s$Var115_Tarj ,
max(b2s$Var115_Tarj)))

# Dummies para la Var115_Tarj
b1s[, "var115_Tarj.1"] <- as.numeric(
basesin$Var115_TarjetaCredito>=
var115_Tarj.uniq[1] &
basesin$Var115_TarjetaCredito<=
var115_Tarj.uniq[2])
b1s[, "var115_Tarj.2"] <- as.numeric(
basesin$Var115_TarjetaCredito>
var115_Tarj.uniq[2] &
basesin$Var115_TarjetaCredito<=
var115_Tarj.uniq[3])

##### Var72_Alia
var72_Alia.uniq <- unique(
c(-1, b2s$Var72_Alia ,
max(b2s$Var72_Alia)))

# Dummies para la var72_Alia
b1s[, "var72_Alia.1"] <- as.numeric(
basesin$Var72_Alianza>=
var72_Alia.uniq[1] &
basesin$Var72_Alianza<=var72_Alia.uniq[2])
b1s[, "var72_Alia.2"] <- as.numeric(
basesin$Var72_Alianza>
var72_Alia.uniq[2] &
basesin$Var72_Alianza<=var72_Alia.uniq[3])

```

```

##### Var72_Cons
var72_Cons.uniq <- unique(
c(-1, b2s$Var72_Cons,
max(b2s$Var72_Cons)))

# Dummies para la var72_Cons
b1s[, "var72_Cons.1"] <- as.numeric(
basesin$Var72_Consumo>=
var72_Cons.uniq[1] &
basesin$Var72_Consumo<=var72_Cons.uniq[2])
b1s[, "var72_Cons.2"] <- as.numeric(
basesin$Var72_Consumo>
var72_Cons.uniq[2] &
basesin$Var72_Consumo<=var72_Cons.uniq[3])

##### Var72_IncoJapon
var72_IncoJapon.uniq <- unique(
c(-1, b2s$Var72_IncomelJa,
max(b2s$Var72_IncomelJa)))

# Dummies para la Var72_IncoJapon
b1s[, "var72_IncoJapon.1"] <- as.numeric(
basesin$Var72_IncomelJapon>=
var72_IncoJapon.uniq[1] &
basesin$Var72_IncomelJapon<=
var72_IncoJapon.uniq[2])

b1s[, "var72_IncoJapon.2"] <- as.numeric(
basesin$Var72_IncomelJapon>
var72_IncoJapon.uniq[2] &
basesin$Var72_IncomelJapon<=
var72_IncoJapon.uniq[3])

##### Var72_IncomelCC
var72_IncomelCC.uniq <- unique(
c(-1, b2s$Var72_IncomelCC,
max(b2s$Var72_IncomelCC)))

# Dummies para la Var72_IncomelCC
b1s[, "var72_IncomelCC.1"] <- as.numeric(
basesin$Var72_IncomelCCo>=
var72_IncomelCC.uniq[1] &
basesin$Var72_IncomelCCo<=
var72_IncomelCC.uniq[2])

b1s[, "var72_IncomelCC.2"] <- as.numeric(
basesin$Var72_IncomelCCo>

var72_IncomelCC.uniq[2] &
basesin$Var72_IncomelCCo<=
var72_IncomelCC.uniq[3])

##### Var72_Micr
var72_Micr.uniq <- unique(
c(-1, b2s$Var72_Micr,
max(b2s$Var72_Micr)))

# Dummies para la Var72_Micr
b1s[, "var72_Micr.1"] <- as.numeric(
basesin$Var72_MicroCredito>=
var72_Micr.uniq[1] &
basesin$Var72_MicroCredito<=
var72_Micr.uniq[2])
b1s[, "var72_Micr.2"] <- as.numeric(
basesin$Var72_MicroCredito>
var72_Micr.uniq[2] &
basesin$Var72_MicroCredito<=
var72_Micr.uniq[3])

##### Var72_Tarj
var72_Tarj.uniq <- unique(
c(-1, b2s$Var72_Tarj,
max(b2s$Var72_Tarj)))

# Dummies para la Var72_Tarj
b1s[, "var72_Tarj.1"] <- as.numeric(
basesin$Var72_TarjetaCredito>=
var72_Tarj.uniq[1] &
basesin$Var72_TarjetaCredito<=
var72_Tarj.uniq[2])
b1s[, "var72_Tarj.2"] <- as.numeric(
basesin$Var72_TarjetaCredito>
var72_Tarj.uniq[2] &
basesin$Var72_TarjetaCredito<=
var72_Tarj.uniq[3])

##### Var73_Can3IncSalAlia
var73_Can3IncSalAlia.uniq <- unique(
c(-1, b2s$Var73_Can3IncSalAlia,
max(b2s$Var73_Can3IncSalAlia)))

# Dummies para la var73_Can3IncSalAlia
b1s[, "var73_Can3IncSalAlia.1"] <- as.numeric(
basesin$Var73_Can3IncSalAlianza>=

```

```

var73_Can3IncSalAlia . uniq[1] &
basesin$Var73_Can3IncSalAlianza <=
var73_Can3IncSalAlia . uniq[2])

b1s[,"var73_Can3IncSalAlia .2"] <- as.numeric(
basesin$Var73_Can3IncSalAlianza >
var73_Can3IncSalAlia . uniq[2] &
basesin$Var73_Can3IncSalAlianza <=
var73_Can3IncSalAlia . uniq[3])

##### Var73_Can3IncSalCons
var73_Can3IncSalCons . uniq <- unique(
c(-1, b2s$Var73_Can3IncSalCons ,
max(b2s$Var73_Can3IncSalCons)))

# Dummies para la var73_Can3IncSalCons
b1s[,"var73_Can3IncSalCons .1"] <- as.numeric(
basesin$Var73_Can3IncSalConsumo>=
var73_Can3IncSalCons . uniq[1] &
basesin$Var73_Can3IncSalConsumo<=
var73_Can3IncSalCons . uniq[2])

b1s[,"var73_Can3IncSalCons .2"] <- as.numeric(
basesin$Var73_Can3IncSalConsumo >
var73_Can3IncSalCons . uniq[2] &
basesin$Var73_Can3IncSalConsumo<=
var73_Can3IncSalCons . uniq[3])

##### Var73_Can3IncSalIncoJap
var73_Can3IncSalIncoJap . uniq <- unique(
c(-1, b2s$Var73_Can3IncSalIncoJap ,
max(b2s$Var73_Can3IncSalIncoJap )))

# Dummies para la Var73_Can3IncSalIncoJap
b1s[,"var73_Can3IncSalIncoJap .1"]<-as.numeric(
basesin$Var73_Can3IncSalIncomelJapon>=
var73_Can3IncSalIncoJap . uniq[1] &
basesin$Var73_Can3IncSalIncomelJapon<=
var73_Can3IncSalIncoJap . uniq[2])

b1s[,"var73_Can3IncSalIncoJap .2"]<-as.numeric(
basesin$Var73_Can3IncSalIncomelJapon >
var73_Can3IncSalIncoJap . uniq[2] &
basesin$Var73_Can3IncSalIncomelJapon<=
var73_Can3IncSalIncoJap . uniq[3])

##### Var73_Can3IncSalIncoCCO
var73_Can3IncSalIncoCCO . uniq <- unique(
c(-1, b2s$Var73_Can3IncSalIncoCCO ,
max(b2s$Var73_Can3IncSalIncoCCO)))

# Dummies para la Var73_Can3IncSalIncoCCO
b1s[,"var73_Can3IncSalIncoCCO .1"] <- as.numeric(
basesin$Var73_Can3IncSalIncomelCCo>=
var73_Can3IncSalIncoCCO . uniq[1] &
basesin$Var73_Can3IncSalIncomelCCo<=
var73_Can3IncSalIncoCCO . uniq[2])

b1s[,"var73_Can3IncSalIncoCCO .2"] <- as.numeric(
basesin$Var73_Can3IncSalIncomelCCo >
var73_Can3IncSalIncoCCO . uniq[2] &
basesin$Var73_Can3IncSalIncomelCCo<=
var73_Can3IncSalIncoCCO . uniq[3])

##### Var73_Can3IncSalMicr
var73_Can3IncSalMicr . uniq <- unique(
c(-1, b2s$Var73_Can3IncSalMicr ,
max(b2s$Var73_Can3IncSalMicr)))

# Dummies para la Var73_Can3IncSalMicr
b1s[,"var73_Can3IncSalMicr .1"] <- as.numeric(
basesin$Var73_Can3IncSalMicroCredito>=
var73_Can3IncSalMicr . uniq[1] &
basesin$Var73_Can3IncSalMicroCredito<=
var73_Can3IncSalMicr . uniq[2])

b1s[,"var73_Can3IncSalMicr .2"] <- as.numeric(
basesin$Var73_Can3IncSalMicroCredito >
var73_Can3IncSalMicr . uniq[2] &
basesin$Var73_Can3IncSalMicroCredito<=
var73_Can3IncSalMicr . uniq[3])

##### Var74_Can3DecSalAlia
var74_Can3DecSalAlia . uniq <- unique(
c(-1, b2s$Var74_Can3DecSalAlia ,
max(b2s$Var74_Can3DecSalAlia )))

# Dummies para la var74_Can3DecSalAlia
b1s[,"var74_Can3DecSalAlia .1"] <- as.numeric(
basesin$Var74_Can3DecSalAlianza >=
var74_Can3DecSalAlia . uniq[1] &
basesin$Var74_Can3DecSalAlianza <=

```

```

var74_Can3DecSalAlia . uniq[2])

b1s[, "var74_Can3DecSalAlia . 2"] <- as.numeric(
basesin$Var74_Can3DecSalAlianza >
var74_Can3DecSalAlia . uniq[2] &
basesin$Var74_Can3DecSalAlianza <=
var74_Can3DecSalAlia . uniq[3])

##### Var74_Can3DecSalCons
var74_Can3DecSalCons . uniq <- unique(
c(-1, b2s$Var74_Can3DecSalCons ,
max(b2s$Var74_Can3DecSalCons)))

# Dummies para la var74_Can3DecSalCons
b1s[, "var74_Can3DecSalCons . 1"] <- as.numeric(
basesin$Var74_Can3DecSalConsumo >=
var74_Can3DecSalCons . uniq[1] &
basesin$Var74_Can3DecSalConsumo <=
var74_Can3DecSalCons . uniq[2])

b1s[, "var74_Can3DecSalCons . 2"] <- as.numeric(
basesin$Var74_Can3DecSalConsumo >
var74_Can3DecSalCons . uniq[2] &
basesin$Var74_Can3DecSalConsumo <=
var74_Can3DecSalCons . uniq[3])

##### Var74_Can3DecSalIncomelJ
var74_Can3DecSalIncomelJ . uniq <- unique(
c(-1, b2s$Var74_Can3DecSalIncomelJ ,
max(b2s$Var74_Can3DecSalIncomelJ)))

# Dummies para la Var74_Can3DecSalIncomelJ
b1s[, "var74_Can3DecSalIncomelJ . 1"] <- as.numeric(
basesin$Var74_Can3DecSalIncomelJapon >=
var74_Can3DecSalIncomelJ . uniq[1] &
basesin$Var74_Can3DecSalIncomelJapon <=
var74_Can3DecSalIncomelJ . uniq[2])

b1s[, "var74_Can3DecSalIncomelJ . 2"] <- as.numeric(
basesin$Var74_Can3DecSalIncomelJapon >
var74_Can3DecSalIncomelJ . uniq[2] &
basesin$Var74_Can3DecSalIncomelJapon <=
var74_Can3DecSalIncomelJ . uniq[3])

##### Var74_Can3DecSalIncomelC
var74_Can3DecSalIncomelC . uniq <- unique(
c(-1, b2s$Var74_Can3DecSalIncomelC ,
max(b2s$Var74_Can3DecSalIncomelC)))

# Dummies para la Var74_Can3DecSalIncomelC
b1s[, "var74_Can3DecSalIncomelC . 1"] <- as.numeric(
basesin$Var74_Can3DecSalIncomelCCo >=
var74_Can3DecSalIncomelC . uniq[1] &
basesin$Var74_Can3DecSalIncomelCCo <=
var74_Can3DecSalIncomelC . uniq[2])

b1s[, "var74_Can3DecSalIncomelC . 2"] <- as.numeric(
basesin$Var74_Can3DecSalIncomelCCo >
var74_Can3DecSalIncomelC . uniq[2] &
basesin$Var74_Can3DecSalIncomelCCo <=
var74_Can3DecSalIncomelC . uniq[3])

##### Var74_Can3DecSalMicr
var74_Can3DecSalMicr . uniq <- unique(
c(-1, b2s$Var74_Can3DecSalMicr ,
max(b2s$Var74_Can3DecSalMicr)))

# Dummies para la Var74_Can3DecSalMicr
b1s[, "var74_Can3DecSalMicr . 1"] <- as.numeric(
basesin$Var74_Can3DecSalMicroCredito >=
var74_Can3DecSalMicr . uniq[1] &
basesin$Var74_Can3DecSalMicroCredito <=
var74_Can3DecSalMicr . uniq[2])

b1s[, "var74_Can3DecSalMicr . 2"] <- as.numeric(
basesin$Var74_Can3DecSalMicroCredito >
var74_Can3DecSalMicr . uniq[2] &
basesin$Var74_Can3DecSalMicroCredito <=
var74_Can3DecSalMicr . uniq[3])

##### Var75_Can3IncomeJa
var75_Can3IncomeJa . uniq <- unique(
c(-1, b2s$Var75_Can3IncomeJa ,
max(b2s$Var75_Can3IncomeJa)))

# Dummies para la Var75_Can3IncomeJa
b1s[, "var75_Can3IncomeJa . 1"] <- as.numeric(
basesin$Var75_Can3IncCasIncomelJapon >=
var75_Can3IncomeJa . uniq[1] &
basesin$Var75_Can3IncCasIncomelJapon <=
var75_Can3IncomeJa . uniq[2])

```

```

b1s[, "var75_Can3IncomeJa.2"] <- as.numeric(
basesin$Var75_Can3IncCasIncomelJapon >
var75_Can3IncomeJa.uniq[2] &
basesin$Var75_Can3IncCasIncomelJapon <=
var75_Can3IncomeJa.uniq[3])

##### Var76_Can3DecCasAlia
var76_Can3DecCasAlia.uniq <- unique(
c(-1, b2s$Var76_Can3DecCasAlia,
max(b2s$Var76_Can3DecCasAlia)))

# Dummies para la var76_Can3DecCasAlia
b1s[, "var76_Can3DecCasAlia.1"] <- as.numeric(
basesin$Var76_Can3DecCasAlianza >=
var76_Can3DecCasAlia.uniq[1] &
basesin$Var76_Can3DecCasAlianza <=
var76_Can3DecCasAlia.uniq[2])

b1s[, "var76_Can3DecCasAlia.2"] <- as.numeric(
basesin$Var76_Can3DecCasAlianza >
var76_Can3DecCasAlia.uniq[2] &
basesin$Var76_Can3DecCasAlianza <=
var76_Can3DecCasAlia.uniq[3])

##### Var76_Can3DecCasCons
var76_Can3DecCasCons.uniq <- unique(
c(-1, b2s$Var76_Can3DecCasCons,
max(b2s$Var76_Can3DecCasCons)))

# Dummies para la var76_Can3DecCasCons
b1s[, "var76_Can3DecCasCons.1"] <- as.numeric(
basesin$Var76_Can3DecCasConsumo >=
var76_Can3DecCasCons.uniq[1] &
basesin$Var76_Can3DecCasConsumo <=
var76_Can3DecCasCons.uniq[2])

b1s[, "var76_Can3DecCasCons.2"] <- as.numeric(
basesin$Var76_Can3DecCasConsumo >
var76_Can3DecCasCons.uniq[2] &
basesin$Var76_Can3DecCasConsumo <=
var76_Can3DecCasCons.uniq[3])

##### Var76_Can3DecCasIncomelJ
var76_Can3DecCasIncomelJ.uniq <- unique(
c(-1, b2s$Var76_Can3DecCasIncomelJ,
max(b2s$Var76_Can3DecCasIncomelJ)))

# Dummies para la Var76_Can3DecCasIncomelJ
b1s[, "var76_Can3DecCasIncomelJ.1"] <- as.numeric(
basesin$Var76_Can3DecCasIncomelJapon >=
var76_Can3DecCasIncomelJ.uniq[1] &
basesin$Var76_Can3DecCasIncomelJapon <=
var76_Can3DecCasIncomelJ.uniq[2])

b1s[, "var76_Can3DecCasIncomelJ.2"] <- as.numeric(
basesin$Var76_Can3DecCasIncomelJapon >
var76_Can3DecCasIncomelJ.uniq[2] &
basesin$Var76_Can3DecCasIncomelJapon <=
var76_Can3DecCasIncomelJ.uniq[3])

##### Var76_Can3DecCasIncomelC
var76_Can3DecCasIncomelC.uniq <- unique(
c(-1, b2s$Var76_Can3DecCasIncomelC,
max(b2s$Var76_Can3DecCasIncomelC)))

# Dummies para la Var76_Can3DecCasIncomelC
b1s[, "var76_Can3DecCasIncomelC.1"] <- as.numeric(
basesin$Var76_Can3DecCasIncomelCCo >=
var76_Can3DecCasIncomelC.uniq[1] &
basesin$Var76_Can3DecCasIncomelCCo <=
var76_Can3DecCasIncomelC.uniq[2])

b1s[, "var76_Can3DecCasIncomelC.2"] <- as.numeric(
basesin$Var76_Can3DecCasIncomelCCo >
var76_Can3DecCasIncomelC.uniq[2] &
basesin$Var76_Can3DecCasIncomelCCo <=
var76_Can3DecCasIncomelC.uniq[3])

##### Var76_Can3DecCasMicr
var76_Can3DecCasMicr.uniq <- unique(
c(-1, b2s$Var76_Can3DecCasMicr,
max(b2s$Var76_Can3DecCasMicr)))

# Dummies para la Var76_Can3DecCasMicr
b1s[, "var76_Can3DecCasMicr.1"] <- as.numeric(
basesin$Var76_Can3DecCasMicro >=
var76_Can3DecCasMicr.uniq[1] &
basesin$Var76_Can3DecCasMicro <=
var76_Can3DecCasMicr.uniq[2])

b1s[, "var76_Can3DecCasMicr.2"] <- as.numeric(
basesin$Var76_Can3DecCasMicro >

```

```

var76_Can3DecCasMicr . uniq [2] &
basesin$Var76_Can3DecCasMicro<=
var76_Can3DecCasMicr . uniq [3])

##### Var73_Can6IncSalAlia
var73_Can6IncSalAlia . uniq <- unique(
c(-1, b2s$Var73_Can6IncSalAlia ,
max(b2s$Var73_Can6IncSalAlia )))

# Dummies para la var73_Can6IncSalAlia
b1s[, "var73_Can6IncSalAlia . 1" ] <- as.numeric(
basesin$Var73_Can6IncSalAlianza >=
var73_Can6IncSalAlia . uniq [1] &
basesin$Var73_Can6IncSalAlianza <=
var73_Can6IncSalAlia . uniq [2])

b1s[, "var73_Can6IncSalAlia . 2" ] <- as.numeric(
basesin$Var73_Can6IncSalAlianza >
var73_Can6IncSalAlia . uniq [2] &
basesin$Var73_Can6IncSalAlianza <=
var73_Can6IncSalAlia . uniq [3])

##### Var73_Can6IncSalCons
var73_Can6IncSalCons . uniq <- unique(
c(-1, b2s$Var73_Can6IncSalCons ,
max(b2s$Var73_Can6IncSalCons )))

# Dummies para la var73_Can6IncSalCons
b1s[, "var73_Can6IncSalCons . 1" ] <- as.numeric(
basesin$Var73_Can6IncSalConsumo>=
var73_Can6IncSalCons . uniq [1] &
basesin$Var73_Can6IncSalConsumo<=
var73_Can6IncSalCons . uniq [2])

b1s[, "var73_Can6IncSalCons . 2" ] <- as.numeric(
basesin$Var73_Can6IncSalConsumo>
var73_Can6IncSalCons . uniq [2] &
basesin$Var73_Can6IncSalConsumo<=
var73_Can6IncSalCons . uniq [3])

##### Var73_Can6IncSalIncoJap
var73_Can6IncSalIncoJap . uniq <- unique(
c(-1, b2s$Var73_Can6IncSalIncomelJ ,
max(b2s$Var73_Can6IncSalIncomelJ )))

# Dummies para la Var73_Can6IncSalIncoJap
b1s[, "var73_Can6IncSalIncoJap . 1" ] <- as.numeric(
basesin$Var73_Can6IncSalIncomelJapon >=
var73_Can6IncSalIncoJap . uniq [1] &
basesin$Var73_Can6IncSalIncomelJapon <=
var73_Can6IncSalIncoJap . uniq [2])

b1s[, "var73_Can6IncSalIncoJap . 2" ] <- as.numeric(
basesin$Var73_Can6IncSalIncomelJapon >
var73_Can6IncSalIncoJap . uniq [2] &
basesin$Var73_Can6IncSalIncomelJapon <=
var73_Can6IncSalIncoJap . uniq [3])

##### Var73_Can6IncSalIncoCCO
var73_Can6IncSalIncoCCO . uniq <- unique(
c(-1, b2s$Var73_Can6IncSalIncomelC ,
max(b2s$Var73_Can6IncSalIncomelC )))

# Dummies para la Var73_Can6IncSalIncoCCO
b1s[, "var73_Can6IncSalIncoCCO . 1" ] <- as.numeric(
basesin$Var73_Can6IncSalIncomelCCo >=
var73_Can6IncSalIncoCCO . uniq [1] &
basesin$Var73_Can6IncSalIncomelCCo <=
var73_Can6IncSalIncoCCO . uniq [2])

b1s[, "var73_Can6IncSalIncoCCO . 2" ] <- as.numeric(
basesin$Var73_Can6IncSalIncomelCCo >
var73_Can6IncSalIncoCCO . uniq [2] &
basesin$Var73_Can6IncSalIncomelCCo <=
var73_Can6IncSalIncoCCO . uniq [3])

##### Var73_Can6IncSalMicr
var73_Can6IncSalMicr . uniq <- unique(
c(-1, b2s$Var73_Can6IncSalMicr ,
max(b2s$Var73_Can6IncSalMicr )))

# Dummies para la Var73_Can6IncSalMicr
b1s[, "var73_Can6IncSalMicr . 1" ] <- as.numeric(
basesin$Var73_Can6IncSalMicroCredito >=
var73_Can6IncSalMicr . uniq [1] &
basesin$Var73_Can6IncSalMicroCredito <=
var73_Can6IncSalMicr . uniq [2])

b1s[, "var73_Can6IncSalMicr . 2" ] <- as.numeric(
basesin$Var73_Can6IncSalMicroCredito >
var73_Can6IncSalMicr . uniq [2] &
basesin$Var73_Can6IncSalMicroCredito <=
var73_Can6IncSalMicr . uniq [3])

```



```

##### Var74_Can6DecSalAlia
var74_Can6DecSalAlia.uniql <- unique(
c(-1, b2s$Var74_Can6DecSalAlia,
max(b2s$Var74_Can6DecSalAlia)))

# Dummies para la var74_Can6DecSalAlia
b1s[, "var74_Can6DecSalAlia.1"] <- as.numeric(
basesin$Var74_Can6DecSalAlianza >=
var74_Can6DecSalAlia.uniql[1] &
basesin$Var74_Can6DecSalAlianza <=
var74_Can6DecSalAlia.uniql[2])

b1s[, "var74_Can6DecSalAlia.2"] <- as.numeric(
basesin$Var74_Can6DecSalAlianza >
var74_Can6DecSalAlia.uniql[2] &
basesin$Var74_Can6DecSalAlianza <=
var74_Can6DecSalAlia.uniql[3])

##### Var74_Can6DecSalCons
var74_Can6DecSalCons.uniql <- unique(
c(-1, b2s$Var74_Can6DecSalCons,
max(b2s$Var74_Can6DecSalCons)))

# Dummies para la var74_Can6DecSalCons
b1s[, "var74_Can6DecSalCons.1"] <- as.numeric(
basesin$Var74_Can6DecSalConsumo >=
var74_Can6DecSalCons.uniql[1] &
basesin$Var74_Can6DecSalConsumo <=
var74_Can6DecSalCons.uniql[2])

b1s[, "var74_Can6DecSalCons.2"] <- as.numeric(
basesin$Var74_Can6DecSalConsumo >
var74_Can6DecSalCons.uniql[2] &
basesin$Var74_Can6DecSalConsumo <=
var74_Can6DecSalCons.uniql[3])

##### Var74_Can6DecSalIncomelJ
var74_Can6DecSalIncomelJ.uniql <- unique(
c(-1, b2s$Var74_Can6DecSalIncomelJ,
max(b2s$Var74_Can6DecSalIncomelJ)))

# Dummies para la Var74_Can6DecSalIncomelJ
b1s[, "var74_Can6DecSalIncomelJ.1"] <- as.numeric(
basesin$Var74_Can6DecSalIncomelJapon >=
var74_Can6DecSalIncomelJ.uniql[1] &
basesin$Var74_Can6DecSalIncomelJapon <=
var74_Can6DecSalIncomelJ.uniql[2])

b1s[, "var74_Can6DecSalIncomelJ.2"] <- as.numeric(
basesin$Var74_Can6DecSalIncomelJapon >
var74_Can6DecSalIncomelJ.uniql[2] &
basesin$Var74_Can6DecSalIncomelJapon <=
var74_Can6DecSalIncomelJ.uniql[3])

##### Var74_Can6DecSalIncomelC
var74_Can6DecSalIncomelC.uniql <- unique(
c(-1, b2s$Var74_Can6DecSalIncomelC,
max(b2s$Var74_Can6DecSalIncomelC)))

# Dummies para la Var74_Can6DecSalIncomelC
b1s[, "var74_Can6DecSalIncomelC.1"] <- as.numeric(
basesin$Var74_Can6DecSalIncomelCCo >=
var74_Can6DecSalIncomelC.uniql[1] &
basesin$Var74_Can6DecSalIncomelCCo <=
var74_Can6DecSalIncomelC.uniql[2])

b1s[, "var74_Can6DecSalIncomelC.2"] <- as.numeric(
basesin$Var74_Can6DecSalIncomelCCo >
var74_Can6DecSalIncomelC.uniql[2] &
basesin$Var74_Can6DecSalIncomelCCo <=
var74_Can6DecSalIncomelC.uniql[3])

##### Var74_Can6DecSalMicr
var74_Can6DecSalMicr.uniql <- unique(
c(-1, b2s$Var74_Can6DecSalMicr,
max(b2s$Var74_Can6DecSalMicr)))

# Dummies para la Var74_Can6DecSalMicr
b1s[, "var74_Can6DecSalMicr.1"] <- as.numeric(
basesin$Var74_Can6DecSalMicroCredito >=
var74_Can6DecSalMicr.uniql[1] &
basesin$Var74_Can6DecSalMicroCredito <=
var74_Can6DecSalMicr.uniql[2])

b1s[, "var74_Can6DecSalMicr.2"] <- as.numeric(
basesin$Var74_Can6DecSalMicroCredito >
var74_Can6DecSalMicr.uniql[2] &
basesin$Var74_Can6DecSalMicroCredito <=
var74_Can6DecSalMicr.uniql[3])

```



```

##### Var75_Can6IncCasAlia
var75_Can6IncCasAlia .uniq <- unique(
c(-1, b2s$Var75_Can6IncCasAlia ,
max(b2s$Var75_Can6IncCasAlia )))

# Dummies para la Var75_Can6IncCasAlia
b1s[,"var75_Can6IncCasAlia .1"] <- as.numeric(
basesin$Var75_Can6IncCasAlianza>=
var75_Can6IncCasAlia .uniq[1] &
basesin$Var75_Can6IncCasAlianza<=
var75_Can6IncCasAlia .uniq[2])

b1s[,"var75_Can6IncCasAlia .2"] <- as.numeric(
basesin$Var75_Can6IncCasAlianza>
var75_Can6IncCasAlia .uniq[2] &
basesin$Var75_Can6IncCasAlianza<=
var75_Can6IncCasAlia .uniq[3])

##### Var75_Can6IncCasIncomelJ
var75_Can6IncCasIncomelJ .uniq<-unique(
c(-1, b2s$Var75_Can6IncCasIncomelJ ,
max(b2s$Var75_Can6IncCasIncomelJ )))

# Dummies para la Var75_Can6IncCasIncomelJ
b1s[,"var75_Can6IncCasIncomelJ .1"]<-as.numeric(
basesin$Var75_Can6IncCasIncomelJapon>=
var75_Can6IncCasIncomelJ .uniq[1] &
basesin$Var75_Can6IncCasIncomelJapon<=
var75_Can6IncCasIncomelJ .uniq[2])

b1s[,"var75_Can6IncCasIncomelJ .2"]<-as.numeric(
basesin$Var75_Can6IncCasIncomelJapon>
var75_Can6IncCasIncomelJ .uniq[2] &
basesin$Var75_Can6IncCasIncomelJapon<=
var75_Can6IncCasIncomelJ .uniq[3])

##### Var76_Can6DecCasAlia
var76_Can6DecCasAlia .uniq <- unique(
c(-1, b2s$Var76_Can6DecCasAlia ,
max(b2s$Var76_Can6DecCasAlia )))

# Dummies para la var76_Can6DecCasAlia
b1s[,"var76_Can6DecCasAlia .1"] <- as.numeric(
basesin$Var76_Can6DecCasAlianza>=
var76_Can6DecCasAlia .uniq[1] &
basesin$Var76_Can6DecCasAlianza<=
var76_Can6DecCasAlia .uniq[2])

b1s[,"var76_Can6DecCasAlia .2"] <- as.numeric(
basesin$Var76_Can6DecCasAlianza>
var76_Can6DecCasAlia .uniq[2] &
basesin$Var76_Can6DecCasAlianza<=
var76_Can6DecCasAlia .uniq[3])

##### Var76_Can6DecCasCons
var76_Can6DecCasCons .uniq <- unique(
c(-1, b2s$Var76_Can6DecCasCons ,
max(b2s$Var76_Can6DecCasCons )))

# Dummies para la var76_Can6DecCasCons
b1s[,"var76_Can6DecCasCons .1"] <- as.numeric(
basesin$Var76_Can6DecCasConsumo>=
var76_Can6DecCasCons .uniq[1] &
basesin$Var76_Can6DecCasConsumo<=
var76_Can6DecCasCons .uniq[2])

b1s[,"var76_Can6DecCasCons .2"] <- as.numeric(
basesin$Var76_Can6DecCasConsumo>
var76_Can6DecCasCons .uniq[2] &
basesin$Var76_Can6DecCasConsumo<=
var76_Can6DecCasCons .uniq[3])

##### Var76_Can6DecCasIncomelJ
var76_Can6DecCasIncomelJ .uniq <- unique(
c(-1, b2s$Var76_Can6DecCasIncomelJ ,
max(b2s$Var76_Can6DecCasIncomelJ )))

# Dummies para la Var76_Can6DecCasIncomelJ
b1s[,"var76_Can6DecCasIncomelJ .1"] <- as.numeric(
basesin$Var76_Can6DecCasIncomelJapon>=
var76_Can6DecCasIncomelJ .uniq[1] &
basesin$Var76_Can6DecCasIncomelJapon<=
var76_Can6DecCasIncomelJ .uniq[2])

b1s[,"var76_Can6DecCasIncomelJ .2"] <- as.numeric(
basesin$Var76_Can6DecCasIncomelJapon>
var76_Can6DecCasIncomelJ .uniq[2] &
basesin$Var76_Can6DecCasIncomelJapon<=
var76_Can6DecCasIncomelJ .uniq[3])

##### Var76_Can6DecCasIncomelC
var76_Can6DecCasIncomelC .uniq <- unique(
c(-1, b2s$Var76_Can6DecCasIncomelC ,

```

```

max(b2s$Var76_Can6DecCasIncomelC))

# Dummies para la Var76_Can6DecCasIncomelC
b1s[, "var76_Can6DecCasIncomelC.1"] <- as.numeric(
basesin$Var76_Can6DecCasIncomelCCo>=
var76_Can6DecCasIncomelC.uniq[1] &
basesin$Var76_Can6DecCasIncomelCCo<=
var76_Can6DecCasIncomelC.uniq[2])

b1s[, "var76_Can6DecCasIncomelC.2"] <- as.numeric(
basesin$Var76_Can6DecCasIncomelCCo>
var76_Can6DecCasIncomelC.uniq[2] &
basesin$Var76_Can6DecCasIncomelCCo<=
var76_Can6DecCasIncomelC.uniq[3])

##### Var76_Can6DecCasMicr
var76_Can6DecCasMicr.uniq <- unique(
c(-1, b2s$Var76_Can6DecCasMicr,
max(b2s$Var76_Can6DecCasMicr)))

# Dummies para la Var76_Can6DecCasMicr
b1s[, "var76_Can6DecCasMicr.1"] <- as.numeric(
basesin$Var76_Can6DecCasMicro>=
var76_Can6DecCasMicr.uniq[1] &
basesin$Var76_Can6DecCasMicro<=
var76_Can6DecCasMicr.uniq[2])

b1s[, "var76_Can6DecCasMicr.2"] <- as.numeric(
basesin$Var76_Can6DecCasMicro>
var76_Can6DecCasMicr.uniq[2] &
basesin$Var76_Can6DecCasMicro<=
var76_Can6DecCasMicr.uniq[3])

##### Var80_Can3ProCreAlia
b1s[, "Var80_Can3ProCreAlia.1"] <- as.numeric(
is.na(basesin$Var80_Can3ProCreAlianza)==FALSE)

b1s[, "Var80_Can3ProCreAlia.2"] <- as.numeric(
is.na(basesin$Var80_Can3ProCreAlianza)==TRUE)

##### Var80_Can3ProCreCons
b1s[, "Var80_Can3ProCreCons.1"] <- as.numeric(
is.na(basesin$Var80_Can3ProCreConsumo)==FALSE)

b1s[, "Var80_Can3ProCreCons.2"] <- as.numeric(
is.na(basesin$Var80_Can3ProCreConsumo)==TRUE)

##### Var80_Can3ProCreIncomelJ
b1s[, "Var80_Can3ProCreIncomelJ.1"] <- as.numeric(
is.na(basesin$Var80_Can3ProCreIncomelJapon)==FALSE)

b1s[, "Var80_Can3ProCreIncomelJ.2"] <- as.numeric(
is.na(basesin$Var80_Can3ProCreIncomelJapon)==TRUE)

##### Var80_Can3ProCreMicr
b1s[, "Var80_Can3ProCreMicr.1"] <- as.numeric(
is.na(basesin$Var80_Can3ProCreMicroCredito)==FALSE)

b1s[, "Var80_Can3ProCreMicr.2"] <- as.numeric(
is.na(basesin$Var80_Can3ProCreMicroCredito)==TRUE)

##### Var80_Can6ProCreAlia
b1s[, "Var80_Can6ProCreAlia.1"] <- as.numeric(
is.na(basesin$Var80_Can6ProCreAlianza)==FALSE)

b1s[, "Var80_Can6ProCreAlia.2"] <- as.numeric(
is.na(basesin$Var80_Can6ProCreAlianza)==TRUE)

##### Var80_Can6ProCreCons
b1s[, "Var80_Can6ProCreCons.1"] <- as.numeric(
is.na(basesin$Var80_Can6ProCreConsumo)==FALSE)

b1s[, "Var80_Can6ProCreCons.2"] <- as.numeric(
is.na(basesin$Var80_Can6ProCreConsumo)==TRUE)

##### Var80_Can6ProCreIncomelJ
b1s[, "Var80_Can6ProCreIncomelJ.1"] <- as.numeric(
is.na(basesin$Var80_Can6ProCreIncomelJapon)==FALSE)

b1s[, "Var80_Can6ProCreIncomelJ.2"] <- as.numeric(
is.na(basesin$Var80_Can6ProCreIncomelJapon)==TRUE)

##### Var80_Can6ProCreMicr
b1s[, "Var80_Can6ProCreMicr.1"] <- as.numeric(
is.na(basesin$Var80_Can6ProCreMicroCredito)==FALSE)

b1s[, "Var80_Can6ProCreMicr.2"] <- as.numeric(
is.na(basesin$Var80_Can6ProCreMicroCredito)==TRUE)

##### Var84
b1s[, "Var84.1"] <- as.numeric(is.na(
basesin$Var84_TotalSaldoPendienteSinBDA)==FALSE)

b1s[, "Var84.2"] <- as.numeric(is.na(
basesin$Var84_TotalSaldoPendienteSinBDA)==TRUE)

```

```

##### Var85
b1s[, "Var85.1"] <- as.numeric(is.na(
basesin$Var85_CantidadOperacionesCuotasS)==FALSE)
b1s[, "Var85.2"] <- as.numeric(is.na(
basesin$Var85_CantidadOperacionesCuotasS)==TRUE)

##### Var86
b1s[, "Var86.1"] <- as.numeric(is.na(
basesin$Var86_TotalSaldoPendienteSinBDAS)==FALSE)
b1s[, "Var86.2"] <- as.numeric(is.na(
basesin$Var86_TotalSaldoPendienteSinBDAS)==TRUE)

##### Var87
b1s[, "Var87.1"] <- as.numeric(is.na(
basesin$Var87_ValorCuotaParceladosSinBDA)==FALSE)
b1s[, "Var87.2"] <- as.numeric(is.na(
basesin$Var87_ValorCuotaParceladosSinBDA)==TRUE)

##### Var88
b1s[, "Var88.1"] <- as.numeric(is.na(
basesin$Var88_CantidadDiasAtrasoSinBDA)==FALSE)
b1s[, "Var88.2"] <- as.numeric(is.na(
basesin$Var88_CantidadDiasAtrasoSinBDA)==TRUE)

##### Var89
b1s[, "Var89.1"] <- as.numeric(is.na(
basesin$Var89_SaldoAtrasoCorteSinBDA)==FALSE)
b1s[, "Var89.2"] <- as.numeric(is.na(
basesin$Var89_SaldoAtrasoCorteSinBDA)==TRUE)

##### Var90
b1s[, "Var90.1"] <- as.numeric(is.na(
basesin$Var90_CantidadRotativosSinBDA)==FALSE)
b1s[, "Var90.2"] <- as.numeric(is.na(
basesin$Var90_CantidadRotativosSinBDA)==TRUE)

##### Var91
b1s[, "Var91.1"] <- as.numeric(is.na(
basesin$Var91_SaldoPendienteRotativosSin) == FALSE)
b1s[, "Var91.2"] <- as.numeric(is.na(
basesin$Var91_SaldoPendienteRotativosSin) == TRUE)

##### Var92
b1s[, "Var92.1"] <- as.numeric(is.na(
basesin$Var92_SaldoTotalDemandaSinBDA) == FALSE)
b1s[, "Var92.2"] <- as.numeric(is.na(
basesin$Var92_SaldoTotalDemandaSinBDA) == TRUE)

##### Var93_RenNa
# Se asigna el valor de 0 a los que tienen NA
basesin$Var93_RenNat[is.na(
basesin$Var93_RenNat)] <- 0
# var93_RenNat.uniq <- unique(
# c(-1, b2s$Var93_RenNa, max(b2s$Var93_RenNa)))
var93_renta <- c(0, 2500, 15000, 300000)

##### Var93_RenMi
var93_RenMi.uniq <- unique(
c(-1, b2s$Var93_RenMi,
max(b2s$Var93_RenMi)))
# Dummies para la Var93_RenMi
b1s[, "var93_RenMi.1"] <- as.numeric(
basesin$Var93_RenMic>=var93_RenMi.uniq[1] &
basesin$Var93_RenMic<=var93_RenMi.uniq[2])
b1s[, "var93_RenMi.2"] <- as.numeric(
basesin$Var93_RenMic>var93_RenMi.uniq[2] &
basesin$Var93_RenMic<=var93_RenMi.uniq[3])
b1s[, "var93_RenMi.3"] <- as.numeric(
basesin$Var93_RenMic>var93_RenMi.uniq[3] &
basesin$Var93_RenMic<=var93_RenMi.uniq[4])

##### Var_Sexo
# El grupo 1 será si el cliente es Masculino.
# El grupo 2 será si el cliente es Femenino.

```

```

# Dummies para la Var_Sexo
b1s[, "var_Sexo.1"] <- as.numeric(
basesin$Var95_Sexo=="MASCULINO")
b1s[, "var_Sexo.2"] <- as.numeric(
basesin$Var95_Sexo=="FEMENINO")

##### Var99_TipViv
# Clúster de la variable Tipo de Vivienda,
# hay 4 grupos:
# ARRENDADA: Grupo 1
# PROPIA HIPOTECADA, PROPIA NO
# HIPOTECADA : Grupo 2
# PRESTADA : Grupo 3
# SI , VIVE CON FAMILIARES : Grupo 4

# Dummies para tipo de vivienda
b1s[, "viv_arrendada"] <- as.numeric(
basesin$Var99_TipViv=="ARRENDADA")
b1s[, "viv_propia"] <- as.numeric(
basesin$Var99_TipViv=="PROPIA HIPOTECADA")
+ as.numeric(
basesin$Var99_TipViv=="PROPIA NO HIPOTECADA")
b1s[, "viv_prestada"] <- as.numeric(
basesin$Var99_TipViv=="PRESTADA") +
as.numeric(
basesin$Var99_TipViv=="VIVE CON FAMILIARES")

##### Var102_Esc
#Clúster de la variable Escolaridad,
# hay 3 grupos:
# Formación intermedia , Primaria ,
# Sin estudios: Grupo 1
# Secundaria–Universidad: Grupo 2
# Postgrado : Grupo 3

basesin$Var102_Esc[is.na(
basesin$Var102_Esc)==TRUE] <- "DESCONOCIDO"

# Dummies para Escolaridad Var102_Esc
b1s[, "esc_basica"] <- as.numeric(
basesin$Var102_Esc=="FORMACION INTERMEDIA")
+ as.numeric(
basesin$Var102_Esc=="PRIMARIA")
+ as.numeric(
basesin$Var102_Esc=="SIN ESTUDIOS")

+ as.numeric(
basesin$Var102_Esc=="DESCONOCIDO")
b1s[, "esc_media"] <- as.numeric(
basesin$Var102_Esc=="SECUNDARIA")
+ as.numeric(
basesin$Var102_Esc=="UNIVERSIDAD")
b1s[, "esc_superior"] <- as.numeric(
basesin$Var102_Esc=="POSTGRADO")

##### Var103_EstCiv
# Clúster de la variables Estado Civil ,
# hay 3 grupos:
#SOLTERO, DIVORCIADO,
# VIUDO : Grupo 1
# CASADO, UNIN LIBRE: Grupo 2
# SIN INFORMACION : Grupo 3

# Dummies para Estado Civil Var103_EstCiv
b1s[, "estciv_soltero"] <- as.numeric(
basesin$Var103_EstCiv=="SOLTERO")
+ as.numeric(
basesin$Var103_EstCiv=="DIVORCIADO")
+ as.numeric(
basesin$Var103_EstCiv=="VIUDO")
b1s[, "estciv_casado"] <- as.numeric(
basesin$Var103_EstCiv=="CASADO")
+ as.numeric(
basesin$Var103_EstCiv=="UNIN LIBRE")
b1s[, "estciv_desconocido"] <- as.numeric(
basesin$Var103_EstCiv=="SI")

##### Var47
var47.uniq <- unique(
c(-1, b2s$Var47, max(b2s$Var47)))

# Dummies para la Var47
b1s[, "var47.1"] <- as.numeric(
basesin$Var47_SalTotProCre >=var47.uniq[1] &
basesin$Var47_SalTotProCre <=var47.uniq[2])
b1s[, "var47.2"] <- as.numeric(
basesin$Var47_SalTotProCre >var47.uniq[2] &
basesin$Var47_SalTotProCre <=var47.uniq[3])
b1s[, "var47.3"] <- as.numeric(
basesin$Var47_SalTotProCre >var47.uniq[3] &
basesin$Var47_SalTotProCre <=var47.uniq[4])

```

```

##### Var114
var114.uniq <- unique(
c(-1, b2s$Var114, max(b2s$Var114)))

# Dummies para la Var114
b1s[, "var114.1"] <- as.numeric(
basesin$Var114_AtrMaxProCre>=var114.uniq[1] &
basesin$Var114_AtrMaxProCre<=var114.uniq[2])
b1s[, "var114.2"] <- as.numeric(
basesin$Var114_AtrMaxProCre>var114.uniq[2] &
basesin$Var114_AtrMaxProCre<=var114.uniq[3])

##### Var51
var51.uniq <- unique(
c(-1, b2s$Var51, max(b2s$Var51)))

# Dummies para la Var51
b1s[, "var51.1"] <- as.numeric(
basesin$Var51_SalValCuoCre>=var51.uniq[1] &
basesin$Var51_SalValCuoCre<=var51.uniq[2])
b1s[, "var51.2"] <- as.numeric(
basesin$Var51_SalValCuoCre>var51.uniq[2] &
basesin$Var51_SalValCuoCre<=var51.uniq[3])
b1s[, "var51.3"] <- as.numeric(
basesin$Var51_SalValCuoCre>var51.uniq[3] &
basesin$Var51_SalValCuoCre<=var51.uniq[4])

##### Var52
var52.uniq <- unique(
c(-1, b2s$Var52, max(b2s$Var52)))

# Dummies para la Var52
b1s[, "var52.1"] <- as.numeric(
basesin$Var52_PromCuoPagProCre>=var52.uniq[1] &
basesin$Var52_PromCuoPagProCre<=var52.uniq[2])
b1s[, "var52.2"] <- as.numeric(
basesin$Var52_PromCuoPagProCre>var52.uniq[2] &
basesin$Var52_PromCuoPagProCre<=var52.uniq[3])
b1s[, "var52.3"] <- as.numeric(
basesin$Var52_PromCuoPagProCre>var52.uniq[3] &
basesin$Var52_PromCuoPagProCre<=var52.uniq[4])

##### var61_cre3salpag
b1s[, "Var61_cre3salpag.1"] <- as.numeric(
is.na(basesin$Var61_Cre3SalPag)==FALSE)

b1s[, "Var61_cre3salpag.2"] <- as.numeric(
is.na(basesin$Var61_Cre3SalPag)==TRUE)

##### var61_cre6salpag
b1s[, "Var61_cre6salpag.1"] <- as.numeric(
is.na(basesin$Var61_Cre6SalPag)==FALSE)
b1s[, "Var61_cre6salpag.2"] <- as.numeric(
is.na(basesin$Var61_Cre6SalPag)==TRUE)

##### var62_pro3salatr
var62_pro3salatr.uniq <- unique(
c(-1, b2s$var62_pro3salatr,
max(b2s$var62_pro3salatr)))

# Dummies para la Var62_pro3salatr
b1s[, "var62_pro3salatr.1"] <- as.numeric(
basesin$Var62_Pro3SalAtr>=var62_pro3salatr.uniq[1] &
basesin$Var62_Pro3SalAtr<=var62_pro3salatr.uniq[2])
b1s[, "var62_pro3salatr.2"] <- as.numeric(
basesin$Var62_Pro3SalAtr>var62_pro3salatr.uniq[2] &
basesin$Var62_Pro3SalAtr<=var62_pro3salatr.uniq[3])

##### var62_pro6salatr
var62_pro6salatr.uniq <- unique(
c(-1, b2s$var62_pro6salatr,
max(b2s$var62_pro6salatr)))

# Dummies para la Var62_pro6salatr
b1s[, "var62_pro6salatr.1"] <- as.numeric(
basesin$Var62_Pro6SalAtr>=var62_pro6salatr.uniq[1] &
basesin$Var62_Pro6SalAtr<=var62_pro6salatr.uniq[2])
b1s[, "var62_pro6salatr.2"] <- as.numeric(
basesin$Var62_Pro6SalAtr>var62_pro6salatr.uniq[2] &
basesin$Var62_Pro6SalAtr<=var62_pro6salatr.uniq[3])

##### var63_sal3atrmax
var63_sal3atrmax.uniq <- unique(
c(-1, b2s$var63_sal3atrmax,
max(b2s$var63_sal3atrmax)))

# Dummies para la Var63_sal3atrmax
b1s[, "var63_sal3atrmax.1"] <- as.numeric(
basesin$Var63_Sal3AtrMax>=
var63_sal3atrmax.uniq[1] &
basesin$Var63_Sal3AtrMax<=

```

```

var63_sal3atrmax.uniq[2])
b1s[, "var63_sal3atrmax.2"] <- as.numeric(
basesin$Var63_Sal3AtrMax>
var63_sal3atrmax.uniq[2] &
basesin$Var63_Sal3AtrMax<=
var63_sal3atrmax.uniq[3])

##### var63_sal6atrmax
var63_sal6atrmax.uniq <- unique(
c(-1, b2s$var63_sal6atrmax,
max(b2s$var63_sal6atrmax)))

# Dummies para la Var63_sal6atrmax
b1s[, "var63_sal6atrmax.1"] <- as.numeric(
basesin$Var63_Sal6AtrMax>=
var63_sal6atrmax.uniq[1] &
basesin$Var63_Sal6AtrMax<=
var63_sal6atrmax.uniq[2])
b1s[, "var63_sal6atrmax.2"] <- as.numeric(
basesin$Var63_Sal6AtrMax>
var63_sal6atrmax.uniq[2] &
basesin$Var63_Sal6AtrMax<=
var63_sal6atrmax.uniq[3])

##### Var68
b1s[, "Var68.1"] <- as.numeric(
is.na(basesin$Var68_TieUltProCreRef)==FALSE)
b1s[, "Var68.2"] <- as.numeric(
is.na(basesin$Var68_TieUltProCreRef)==TRUE)

##### Var64_SalTotProCre
b1s[, "Var64_SalTotProCre.1"] <- as.numeric(
is.na(basesin$Var64_SalTotProCreRef)==FALSE)
b1s[, "Var64_SalTotProCre.2"] <- as.numeric(
is.na(basesin$Var64_SalTotProCreRef)==TRUE)

##### Var66_AtrMaxProCre
b1s[, "Var66_AtrMaxProCre.1"] <- as.numeric(
is.na(basesin$Var66_AtrMaxProCreRef)==FALSE)
b1s[, "Var66_AtrMaxProCre.2"] <- as.numeric(
is.na(basesin$Var66_AtrMaxProCreRef)==TRUE)

##### Var67_SalAtrProCre
b1s[, "Var67_SalAtrProCre.1"] <- as.numeric(

```

```

is.na(basesin$Var67_SalAtrProCreRef)==FALSE)
b1s[, "Var67_SalAtrProCre.2"] <- as.numeric(
is.na(basesin$Var67_SalAtrProCreRef)==TRUE)

##### Var01_30dias
var01_30dias.uniq <- unique(
c(-1, b2s$Var01_30dias, max(b2s$Var01_30dias)))

# Dummies para la Var01_30dias
b1s[, "var01_30dias.1"] <- as.numeric(
basesin$Var01_30dias_origem>=var01_30dias.uniq[1] &
basesin$Var01_30dias_origem<=var01_30dias.uniq[2])
b1s[, "var01_30dias.2"] <- as.numeric(
basesin$Var01_30dias_origem>var01_30dias.uniq[2] &
basesin$Var01_30dias_origem<=var01_30dias.uniq[3])

##### var40_30dias
basesin$var40_30dias[is.na(
basesin$var40_30dias)] <- 0

var40_30dias.uniq <- unique(
c(-1, b2s$var40_30dias, max(b2s$var40_30dias)))

# Dummies para la Var40_30dias
b1s[, "var40_30dias.1"] <- as.numeric(
basesin$var40_30dias>=var40_30dias.uniq[1] &
basesin$var40_30dias<=var40_30dias.uniq[2])
b1s[, "var40_30dias.2"] <- as.numeric(
basesin$var40_30dias>var40_30dias.uniq[2] &
basesin$var40_30dias<=var40_30dias.uniq[3])

##### var40_15dias
basesin$var40_15dias[is.na(
basesin$var40_15dias)] <- 0

var40_15dias.uniq <- unique(
c(-1, b2s$var40_15dias, max(b2s$var40_15dias)))

# Dummies para la Var40_15dias
b1s[, "var40_15dias.1"] <- as.numeric(
basesin$var40_15dias>=var40_15dias.uniq[1] &
basesin$var40_15dias<=var40_15dias.uniq[2])
b1s[, "var40_15dias.2"] <- as.numeric(
basesin$var40_15dias>var40_15dias.uniq[2] &
basesin$var40_15dias<=var40_15dias.uniq[3])

```

```

##### Var24_IndRen
var24_IndRen.uniq <- unique(
c(-1, b2s$Var24_IndRen, max(b2s$Var24_IndRen)))

# Dummies para la Var24_IndRen
b1s[, "var24_IndRen.1"] <- as.numeric(
basesin$Var24_IndRen>=var24_IndRen.uniq[1] &
basesin$Var24_IndRen<=var24_IndRen.uniq[2])
b1s[, "var24_IndRen.2"] <- as.numeric(
basesin$Var24_IndRen>var24_IndRen.uniq[2] &
basesin$Var24_IndRen<=var24_IndRen.uniq[3])
b1s[, "var24_IndRen.3"] <- as.numeric(
basesin$Var24_IndRen>var24_IndRen.uniq[3] &
basesin$Var24_IndRen<=var24_IndRen.uniq[4])

##### Var39_Atr3
var39_Atr3.uniq <- unique(
c(-1, b2s$Var39_Atr3, max(b2s$Var39_Atr3)))

# Dummies para la Var39_Atr3
b1s[, "var39_Atr3.1"] <- as.numeric(
basesin$Var39_Atr3MaxMes>=var39_Atr3.uniq[1] &
basesin$Var39_Atr3MaxMes<=var39_Atr3.uniq[2])
b1s[, "var39_Atr3.2"] <- as.numeric(
basesin$Var39_Atr3MaxMes>var39_Atr3.uniq[2] &
basesin$Var39_Atr3MaxMes<=var39_Atr3.uniq[3])

##### Var39_Atr6
var39_Atr6.uniq <- unique(
c(-1, b2s$Var39_Atr6, max(b2s$Var39_Atr6)))

# Dummies para la Var39_Atr6
b1s[, "var39_Atr6.1"] <- as.numeric(
basesin$Var39_Atr6MaxMes>=var39_Atr6.uniq[1] &
basesin$Var39_Atr6MaxMes<=var39_Atr6.uniq[2])
b1s[, "var39_Atr6.2"] <- as.numeric(
basesin$Var39_Atr6MaxMes>var39_Atr6.uniq[2] &
basesin$Var39_Atr6MaxMes<=var39_Atr6.uniq[3])

##### Var Segmento

# Dummies para la Var Segmento
b1s[, "var_Segmento.1"] <- as.numeric(
basesin$Segmento=="ALIANZAS")
b1s[, "var_Segmento.2"] <- as.numeric(
basesin$Segmento=="CONSUMO")
b1s[, "var_Segmento.3"] <- as.numeric(
basesin$Segmento=="INCOMEL - CCO")
b1s[, "var_Segmento.4"] <- as.numeric(
basesin$Segmento=="INCOMEL - JAPON CARIBE")
b1s[, "var_Segmento.5"] <- as.numeric(
basesin$Segmento=="MICROCREDITO")

##### Var116
basesin$Var116_SldPendBDASinHip[is.na(
basesin$Var116_SldPendBDASinHip)] <- 0

var116.uniq <- unique(
c(-1, b2s$Var116, max(b2s$Var116)))

# Dummies para la Var116
b1s[, "var116.1"] <- as.numeric(
basesin$Var116_SldPendBDASinHip>=var116.uniq[1] &
basesin$Var116_SldPendBDASinHip<=var116.uniq[2])
b1s[, "var116.2"] <- as.numeric(
basesin$Var116_SldPendBDASinHip>var116.uniq[2] &
basesin$Var116_SldPendBDASinHip<=var116.uniq[3])
b1s[, "var116.3"] <- as.numeric(
basesin$Var116_SldPendBDASinHip>var116.uniq[3] &
basesin$Var116_SldPendBDASinHip<=var116.uniq[4])
b1s[, "var116.4"] <- as.numeric(
basesin$Var116_SldPendBDASinHip>var116.uniq[4] &
basesin$Var116_SldPendBDASinHip<=var116.uniq[5])
b1s[, "var116.5"] <- as.numeric(
basesin$Var116_SldPendBDASinHip>var116.uniq[5] &
basesin$Var116_SldPendBDASinHip<=var116.uniq[6])
b1s[, "var116.6"] <- as.numeric(
basesin$Var116_SldPendBDASinHip>var116.uniq[6] &
basesin$Var116_SldPendBDASinHip<=var116.uniq[7])
b1s[, "var116.7"] <- as.numeric(
basesin$Var116_SldPendBDASinHip>var116.uniq[7] &
basesin$Var116_SldPendBDASinHip<=var116.uniq[8])

##### Var117
basesin$Var117_NInstitBDASinHip[is.na(
basesin$Var117_NInstitBDASinHip)] <- 0

var117.uniq <- unique(
c(-1, b2s$Var117, max(b2s$Var117)))

# Dummies para la Var117

```



```

b1s[, "var117.1"] <- as.numeric(
basesin$Var117_NInstitBDASinHip>=
var117.uniq[1] &
basesin$Var117_NInstitBDASinHip<=
var117.uniq[2])
b1s[, "var117.2"] <- as.numeric(
basesin$Var117_NInstitBDASinHip>
var117.uniq[2] &
basesin$Var117_NInstitBDASinHip<=
var117.uniq[3])
b1s[, "var117.3"] <- as.numeric(
basesin$Var117_NInstitBDASinHip>
var117.uniq[3] &
basesin$Var117_NInstitBDASinHip<=
var117.uniq[4])
b1s[, "var117.4"] <- as.numeric(
basesin$Var117_NInstitBDASinHip>
var117.uniq[4] &
basesin$Var117_NInstitBDASinHip<=
var117.uniq[5])
b1s[, "var117.5"] <- as.numeric(
basesin$Var117_NInstitBDASinHip>
var117.uniq[5] &
basesin$Var117_NInstitBDASinHip<=
var117.uniq[6])

##### Var118
basesin$Var118_NOperBDASinHip[is.na(
basesin$Var118_NOperBDASinHip)] <- 0

var118.uniq <- unique(
c(-1, b2s$Var118, max(b2s$Var118)))

# Dummies para la Var118
b1s[, "var118.1"] <- as.numeric(
basesin$Var118_NOperBDASinHip>=
var118.uniq[1] &
basesin$Var118_NOperBDASinHip<=
var118.uniq[2])
b1s[, "var118.2"] <- as.numeric(
basesin$Var118_NOperBDASinHip>
var118.uniq[2] &
basesin$Var118_NOperBDASinHip<=
var118.uniq[3])
b1s[, "var118.3"] <- as.numeric(
basesin$Var118_NOperBDASinHip>
var118.uniq[3] &
basesin$Var118_NOperBDASinHip<=

var118.uniq[4])
b1s[, "var118.4"] <- as.numeric(
basesin$Var118_NOperBDASinHip>
var118.uniq[4] &
basesin$Var118_NOperBDASinHip<=
var118.uniq[5])
b1s[, "var118.5"] <- as.numeric(
basesin$Var118_NOperBDASinHip>
var118.uniq[5] &
basesin$Var118_NOperBDASinHip<=
var118.uniq[6])
b1s[, "var118.6"] <- as.numeric(
basesin$Var118_NOperBDASinHip>
var118.uniq[6] &
basesin$Var118_NOperBDASinHip<=
var118.uniq[7])

##### Var24_DIndRen
var24_DIndRen.uniq <- unique(
c(-1, b2s$Var24_DIndRen, max(b2s$Var24_DIndRen)))

# Dummies para la Var24_DIndRen
b1s[, "var24_DIndRen.1"] <- as.numeric(
basesin$Var24_DIndRen>=var24_DIndRen.uniq[1] &
basesin$Var24_DIndRen<=var24_DIndRen.uniq[2])
b1s[, "var24_DIndRen.2"] <- as.numeric(
basesin$Var24_DIndRen>var24_DIndRen.uniq[2] &
basesin$Var24_DIndRen<=var24_DIndRen.uniq[3])

##### Var46_DCanProCre
var46_DCanProCre.uniq <- unique(
c(-1, b2s$Var46_DCanProCre,
max(b2s$Var46_DCanProCre)))

# Dummies para la Var46_DCanProCre
b1s[, "var46_DCanProCre.1"] <- as.numeric(
basesin$Var46_DCanProCre>=var46_DCanProCre.uniq[1] &
basesin$Var46_DCanProCre<=var46_DCanProCre.uniq[2])
b1s[, "var46_DCanProCre.2"] <- as.numeric(
basesin$Var46_DCanProCre>var46_DCanProCre.uniq[2] &
basesin$Var46_DCanProCre<=var46_DCanProCre.uniq[3])

##### Var96_Edad
# Jovenes, adultos, y mayores.
var96_Edad.uniq <- c(19.99, 38, 49, 73)

```



```

# Dummies para la Var96_Edad
b1s[, "var96_Edad.1"] <- as.numeric(
basesin$Var96_Edad>=var96_Edad.uniq[1] &
basesin$Var96_Edad<=var96_Edad.uniq[2])
b1s[, "var96_Edad.2"] <- as.numeric(
basesin$Var96_Edad>var96_Edad.uniq[2] &
basesin$Var96_Edad<=var96_Edad.uniq[3])
b1s[, "var96_Edad.3"] <- as.numeric(
basesin$Var96_Edad>var96_Edad.uniq[3] &
basesin$Var96_Edad<=var96_Edad.uniq[4])

#####
# Variables cuantitativas #
#####

##### Var08
b1s[, "Var08"] <- basesin$Var08_MonBru

##### Var13
b1s[, "Var13"] <- basesin$Var13_TimIniCon

##### Var104
b1s[, "Var104"] <- basesin$Var104_NumCuoFal

##### Var105
b1s[, "Var105"] <- basesin$Var105_MonBru

##### Var106
b1s[, "Var106"] <- basesin$Var106_Sal

##### Var107
b1s[, "Var107"] <- basesin$Var107_TieIniCon

##### Var108
b1s[, "Var108"] <- basesin$Var108_NumCuoPag

##### Var109
b1s[, "Var109"] <- basesin$Var109_NumCuoDebSerPag

##### Var110
b1s[, "Var110"] <- basesin$Var110_NumCuoCub

##### Var40_Atr6
b1s[, "Var40_Atr6"] <- basesin$Var40_Atr6Ant

b1s$Var40_Atr6[is.na(b1s$Var40_Atr6)] <- 0

##### Var100
b1s[, "Var100"] <- basesin$Var100_TimRelBcoMM

##### Inflación
b1s[, "Inflación"] <- basesin$Inflación.Guatemala

##### IMAE Guat
b1s[, "IMAE.Guatemala"] <- basesin$IMAE.Guatemala

##### IPC
b1s[, "IPC.Guatemala"] <- basesin$IPC.Guatemala

```

## 6 Referencias

- Abdouab, H., Dongmo, M., Ntim, C. y Baker, R. (2016). Predicting creditworthiness in retail banking with limited scoring data. *Knowledge-Based Systems*, PP. 89–103.
- Aguas, D. y Castillo, M. (2002). Modelo de administración del riesgo crediticio para la cartera comercial de una entidad financiera colombiana. *Apuntes de Banca y Finanzas*, PP. 1-8.
- Allen, L., DeLong, G. y Saunders, A. (2004). Issues in the credit risk modeling of retail. *Journal of Banking and Finance*, PP. 727–752.
- Anderson, R. (2007). *The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation*. Oxford University Press.
- Arias, L., Rave, S. y Castaño, J. (2006). Metodologías para la medición del riesgo financiero en inversiones. *Scientia et technica*.
- Bassel Commite on Banking Supervision. (1998). *Convergencia internacional sobre ponderación de capital y niveles de capital. (sin enmendar)*. Banco de Pagos Internacionales (BIS).
- Bassel Commite on Banking Supervision. (2001). *Marrying the macro and microprudential dimensions of financial stability*.
- Bassel Commite on Banking Supervision. (2004). *International convergence of capital measurement and capital standards: A revised framework*. Bank for International Settlements.
- Bassel Commite on Banking Supervision. (2005). *Studies on the validation of internal rating systems working paper n 14*. Bank for International Settlements.
- Beckman, T. (1949). Cases in credits and collections. *New York: McGraw-Hill*, PP. 145.
- Bessis, J. (1998). *Risk management in banking* (Second Edition. ed.). Other Wiley Editorial Offices.
- BIS. (2000). Range of practice in banks' internal ratings systems. *BIS Discussion Paper*.
- Bonilla, L. (2008). *Diseño y validación del scoring de aprobación de riesgo crediticio para la coop. de ahorro y crédito progreso ltda. para la cartera de vivienda y de consumo con modelos estadísticos y econométricos. (tesis de pregrado)*. Escuela Politécnica Nacional.
- Bonilla, M., Puertas, R. y Olmeda, I. (2003). Modelo paramétricos y no paramétricos en

- problemas de credit scoring. *Revista Española de Financiación y Contabilidad*, 27, PP. 833 - 869.
- Borio, C., Furfine, C. y Lowe, P. (2001). Procyclicality of the financial system and financial stability: Issues and policy options. *BIS Paper No. 1*, PP. 1-57.
- Cardoso, B. y Díaz, J. (2015). Survival mixture models in behavioral scoring. *Expert Systems with Applications.*, PP. 3902-3910.
- Cesaroni, T. (2015). Procyclicality of credit rating systems: How to manage it. *Journal of Economics and Business*.
- Comité de Basilea. (2006). Principios básicos para una supervisión bancaria eficaz. *Banco de Pagos Internacionales*.
- Corporación Financiera Nacional. (2016). *Glosario de términos financieros*. Obtenido de: <http://www.cfn.fin.ec/wp-content/uploads/2016/05/Glosario-de-Términos-Financieros-2016.pdf>.
- Cox, D. y Snell, E. (1989). *Analysis of binary data* (Second Edition ed.). Chapman & Hall, London.
- Cárdenas, M. y Arancibia, H. (2014). Potencia estadística y cálculo del tamaño del efecto en g\*power: Complementos a las pruebas de significación estadística y su aplicación en psicología. *Revista Salud Sociedad*, PP. 210-224.
- Day, R. y Quinn, G. (1989). Comparisons of treatments after an analysis of variance. *Ecological Society of America*, 59(4), PP. 433-463.
- De Lara Haro, A. (2005). *Medición y control de riesgos financieros*. Editorial Limusa.
- Dirick, L., Baesensand, B. y Claeskens, G. (2016). *Using survival analysis to model time to default*. Obtenido de: <http://www.dataminingapps.com/2016/01/using-survival-analysis-to-model-time-to-default/>.
- Fabozzi, F., Modigliani, F. y Ferri, M. (1996). *Mercados e instituciones financieras*. Pearson Educación.
- Gobat, J. (2012). ¿qué es un banco? *Finanzas & Desarrollo*, PP. 38-39.
- Gujarati, D. y Porter, D. (2010). *Econometría* (5th edition ed.). McGRAW-HILL.
- Gutiérrez, C. y Fernández, J. (2006). Evolución del proceso de regulación bancaria hasta basilea 2: origen, características y posibles efectos. *Pecunia*, (2), PP. 23-63.
- Heitfield, E. (2005). Dynamics of rating systems. *BIS [2005b]*.

- Herrera, L., Raúl, H. y Guevara, R. (2010). Aplicación del análisis factorial como una alternativa de solución al problema de multicolinealidad. *Revista Multidisciplinaria del Consejo de Investigación de la Universidad de Oriente*, PP. 181-184.
- Lemeshow, S. y Hosmer, D. (2005). *Applied logistic regression* (Second Edition ed.). John Wiley & Sons, Inc.
- Lizarzaburu, E., Berggrun, L. y Quispe, J. (2012). Gestión de riesgos financieros. experiencia en un banco latinoamericano. *Estudios Gerenciales*, PP. 96-103.
- Mansfield, E. y Helms, B. ((1982)). Detecting multicollinearity. *The American Statistician*, 36:3a, PP. 158-160.
- Mays, E. (2004). *Credit scoring for risk managers: The handbook for lenders*. South-Western Publishing: Mason, OH.
- Mendoza, E. (2010). *Derecho bancario: de la crisis financiera ecuatoriana, causas, consecuencias, soluciones*. Revista Financiera.
- Merton, R. C. (1993). *Operation and regulation in financial intermediation: A functional perspective*.
- Nieto, S. (2010). *Crédito al consumo: La estadística aplicada a un problema de riesgo crediticio*. (proyecto de tesis). Universidad Autónoma Metropolitana.
- Pancorbo, A. (2002). Basilea ii: Las nuevas normas internacionales de solvencia. *Información Comercial Española, ICE: Revista de economía*, (801), PP. 35-50.
- Prabhakaran, S. (2016). *r-statistics.co*. Obtenido de: <http://r-statistics.co/Information-Value-With-R.html>.
- Ramon, M. (2014). *Survival analysis in credit scoring: A framework for pd estimation*. (tesis de pregrado). University of Twente.
- Reinhart, C. y Rogoff, K. (2008). This time is different: A panoramic view of eight centuries of financial crises. nber working paper no. 13882. *Financial Times*.
- Robert, B. y Bendell, A. (2012). Comparison of stopping rules in forward “stepwise” regression. *Journal of the American Statistical Association.*, Volume 72.
- Rodríguez, D., Becerra, J. y Cardona, D. (2017). Modelos y metodologías de credit score para personas naturales: una revisión literaria. *Revista CEA*, 3(5), PP. 13-28.
- Ross, S., Jaffe, J. y Westerfield, R. (2000). *Finanzas corporativas* (no. 658.15/. 16). McGraw-Hill Interamericana.

- SAS, I. I. (2009). *Building credit scorecards using credit scoring for sas enterprise miner: A sas best practices paper. a sas white paper.* (Cary ed.). NC: SAS Institute, Inc.
- Shingala, M. y Rajyaguru, A. (2015). Comparison of post hoc tests for unequal variance. *International Journal of New Technologies in Science and Engineering, Vol. 2,,* PP. 22-33.
- Sotelsek, D. y Pavón, L. (2012). Evolución de los acuerdos de basilea: diagnóstico de los estándares de regulación bancaria internacional. *Economía UNAM, 9(25),* PP. 29-50.
- SunHoHa, Ramayya y Krishnan. (2012). Predicting repayment of the credit card debt. *Computers and Operations Research, 39,* PP. 765-773.
- Superintendencia de Bancos Guatemala. (2017). *Educacion financiera glosario.* Obtenido de: <https://www.sib.gob.gt>.
- Trías, R. y March, A. (2016). *La influencia de basilea 2 en la gestión del riesgo hipotecario.* Obtenido de: <http://www.iberfinanzas.com/index.php/Articulos-especiales/La-influencia-de-Basilea-II-en-la-gestion-del-riesgo-hipotecario.html>.
- Uquillas, A. (2018). Determinantes del riesgo comportamental en préstamos de consumo y microcrédito: Un estudio de caso en centro américa. *Journal of Research in Accounting and Management Science., Vol. 3(Num. 1).*
- Valles, V. (2006). Stability of a “through-the-cycle” rating system during a financial crisis. *Financial Stability Institute, Award 2006 Winning Paper.*
- Vilariño, A. (2010). La regulación financiera ante la crisis de crédito y liquidez. *Jornadas de Economía Crítica..*
- Yap, B., Ong, S. y Mohamed, N. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications, 38,* PP. 13274 - 13283.