

# **ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE CIENCIAS**

## **EL ANÁLISIS DE SUPERVIVENCIA EN LA GESTIÓN DEL RIESGO DE CRÉDITO: EL CASO DE UNA CARTERA HIPOTECARIA ECUATORIANA**

**TESIS DE GRADO PREVIO A LA OBTENCIÓN DEL GRADO DE MAGISTER  
EN ESTADÍSTICA APLICADA**

**PROYECTO DE INVESTIGACIÓN Y DESARROLLO**

**LUIS FELIPE GUEVARA URQUIZO**

`luis_felipe_guevara@yahoo.com`

**DIRECTOR: ING. MIGUEL FLORES SÁNCHEZ, MSc.**

`miguel.flores@epn.edu.ec`

**CODIRECTOR: MAT. LUIS ALCIDES HORNA HUARACA, PHD.**

`luis.horna@epn.edu.ec`

**Quito, enero 2018**

## DECLARACIÓN

Yo, Luis Felipe Guevara Urquiza, 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 Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



---

**Luis Felipe Guevara Urquiza**

## CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Luis Felipe Guevara Urquiza, bajo mi supervisión.

A handwritten signature in blue ink, appearing to read 'Miguel Flores', with a large, sweeping flourish extending to the right.

---

**Ing. Miguel Flores Sánchez, MSc.**

**DIRECTOR**

## CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Luis Felipe Guevara Urquiza, bajo mi supervisión.

A handwritten signature in blue ink, appearing to read 'Luis Horna Huaraca', is positioned above a horizontal line.

---

**Mat. Luis Horna Huaraca, PHD.**

**CODIRECTOR**

## **AGRADECIMIENTO**

A Dios por la vida y la oportunidad de estudiar.

A mi familia por su apoyo y comprensión.

A mi director de tesis Ing. Miguel Flores por la motivación y asesoría brindada.

A todos quienes contribuyeron en mi formación académica.

## **DEDICATORIA**

A la memoria de mi padre Luis  
A la vida de mi madre Clemencia y de mi hija Ivana  
Al amor incondicional de mi esposa Sofía  
A la entrega diaria de mi tía Carmelita  
Al corazón generoso de mi suegra María Elena

## ÍNDICE DE CONTENIDO

LISTA DE FIGURAS .....	i
LISTA DE TABLAS .....	ii
LISTA DE ANEXOS .....	iii
RESUMEN.....	iv
<i>ABSTRACT</i> .....	v
1. INTRODUCCIÓN .....	9
1.1. PREGUNTAS DE INVESTIGACIÓN .....	12
1.2. OBJETIVO GENERAL.....	12
1.3. OBJETIVOS ESPECÍFICOS .....	12
1.4. HIPÓTESIS .....	13
1.5. MARCO TEÓRICO .....	14
1.5.1 Análisis de Supervivencia y Modelo de Cox.....	14
1.5.2 El Análisis de Árboles de Decisión.....	23
2. METODOLOGÍA.....	25
2.1. PROCEDIMIENTO METODOLÓGICO .....	25
2.2. DETALLE DE ACTIVIDADES RELEVANTES .....	26
2.2.1. Definición de la variable de interés .....	26
2.2.2. Breve descripción de la base de datos a utilizar.....	26
2.2.3. Análisis descriptivo de las potenciales variables explicativas .....	27
2.2.4. Verificación de los supuestos del modelo de cox .....	30
2.2.5. Interpretación de los resultados del modelo de cox.....	34
2.2.6. Calibración del modelo de cox frente a un cambio en la variable dependiente .....	35
3. RESULTADOS Y DISCUSIÓN .....	36
3.1. EL SISTEMA FINANCIERO Y SU INFLUENCIA EN LA ECONOMÍA.....	36
3.1.1. La función de producción y la productividad marginal del capital.....	36
3.1.2. Relación entre el sistema financiero y el crecimiento económico.....	38
3.2. EL ANÁLISIS DE SUPERVIVENCIA EN EL SISTEMA FINANCIERO.....	39
3.2.1. Casos de Aplicación .....	39

3.3. EL SISTEMA FINANCIERO ECUATORIANO Y LA DINÁMICA DEL CRÉDITO HIPOTECARIO .....	46
3.3.1. Sistema Financiero Ecuatoriano.....	46
3.3.2. Desempeño general del sector de la construcción.....	52
3.3. APLICACIÓN DEL ANÁLISIS DE SUPERVIVENCIA: EL CASO DE UNA CARTERA HIPOTECARIA ECUATORIANA .....	56
3.3.1. Desarrollo del Modelo de Supervivencia.....	56
3.3.2. Segmentación de Clientes y Estrategias .....	108
3.4. AUTOMATIZACIÓN DEL ANÁLISIS DE SUPERVIVENCIA: EL CASO DE UNA CARTERA HIPOTECARIA ECUATORIANA .....	111
3.4.1. Programación en lenguaje “R” .....	111
4. CONCLUSIONES .....	122
REFERENCIAS BIBLIOGRÁFICAS .....	125
ANEXOS .....	128



## LISTA DE FIGURAS

Figura 1 – Gráfico función de supervivencia KM, ejemplo.....	28
Figura 2 - Saldo de captaciones - OSD .....	46
Figura 3 – Estructura anual de saldo de captaciones (%) .....	47
Figura 4 – Saldo de cartera anual (millones USD).....	48
Figura 5 – Saldo cartera de créditos 2015 y 2016 – Estructura por sectores (%)	49
Figura 6 – Morosidad de la cartera de crédito – Sistema Financiero Privado (%)	49
Figura 7 – Morosidad por segmento (%) .....	50
Figura 8 – Volumen de colocación crediticia – crédito vivienda (millones USD)....	51
Figura 9 – PIB Industria de la Construcción (base 2007), Tasa de variación anual – PIB Industria de la Construcción y (%) Participación - PIB Industria de la Construcción / PIB Total .....	53
Figura 10 - Índice de confianza empresarial del sector de la construcción y ciclo del índice .....	55
Figura 11 – Función supervivencia, método KM, var.: rango de avalúo (R_Avalúo2).....	59
Figura 12 – Función supervivencia, método KM, var.: género (GENERO).....	62
Figura 13 – Función supervivencia, método KM, var.: rango de plazo (R_Plazo) .	64
Figura 14 – Función supervivencia, método KM, var.: rango de ingreso (R_Ingp)	68
Figura 15 – Función supervivencia, método KM, var.: rango de monto financiado (R_Montf) .....	71
Figura 16 – Función supervivencia, método KM, var.: subregión geográfica (SubReg).....	74
Figura 17 – Función supervivencia, método KM, var.: rango de edad (R_Edad) ..	77
Figura 18 – Función supervivencia, método KM, var.: actividad económica (Acteco_r).....	82
Figura 19 – Función supervivencia, método KM, var.: ciclo económico al momento del desembolso (Ciclo) .....	86

Figura 20 – Función supervivencia, método KM, var.: cesante previo a la mora grave (Cesante) .....	89
Figura 21 – Árbol de decisión: Variable de segmentación: mora grave (Mayor90)91	
Figura 22 – Función de supervivencia resultante de la aplicación del modelo de cox (mora grave (> 90 días).....	100
Figura 23 – Residuos de Schonenfeld o dfbetas para cada uno de los coeficientes de las variables .....	103
Figura 24 – Residuos tipo deviance del modelo.....	103
Figura 25 – Residuos tipo score del modelo .....	105
Figura 26 – Residuos de martingala y variable tiempo de permanencia.....	105
Figura 27 – Función de supervivencia resultante de la aplicación del modelo de cox (mora grave > 90 días) vs Función de supervivencia resultante de la aplicación del modelo de cox (mora grave > 120 días).....	107
Figura 28 - Ejemplo Reporte análisis de supervivencia .....	121

## LISTA DE TABLAS

Tabla 1 – Ejemplo, cálculo estadístico KM para función de supervivencia.....	29
Tabla 2 – Ranking instituciones financieras – Volumen de crédito causado por entidades SB (millones de USD) - Segmento vivienda .....	51
Tabla 3 – Número de empleados por rama de actividad económica – septiembre 2015 vs septiembre 2016 .....	54
Tabla 4 – Tabla de contingencia - variable rango de avalúo (R_Avalúo2) vs variable presencia mora grave (Mayor90) .....	57
Tabla 5 – Función de supervivencia, método KM, variable rango de avalúo (R_Avalúo2).....	58
Tabla 6 – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de avalúo (R_Avalúo2).....	59
Tabla 7 – Tabla de contingencia - variable género (GENERO) vs variable presencia mora grave (Mayor90) .....	60
Tabla 8 – Función de supervivencia, método KM, variable género (GENERO) ....	61
Tabla 9 – Resultado del contraste de igualdad de funciones de supervivencia para la variable género (GENERO) .....	62
Tabla 10 – Tabla de contingencia - variable rango de plazo (R_Plazo) vs variable presencia mora grave (Mayor90) .....	63
Tabla 11 – Función de supervivencia, método KM, variable rango de plazo (R_Plazo) .....	64
Tabla 12 – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de plazo (R_Plazo) .....	65
Tabla 13 – Tabla de contingencia - variable rango de ingreso (R_Ingp) vs variable presencia mora grave (Mayor90) .....	65
Tabla 14 – Función de supervivencia, método KM, variable rango de ingreso (R_Ingp).....	67
Tabla 15 – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de ingreso (R_Ingp) .....	68
Tabla 16 – Tabla de contingencia - variable rango de monto financiado (R_Montf) vs variable presencia mora grave (Mayor90) .....	69

Tabla 17 – Función de supervivencia, método KM, variable rango de monto financiado (R_Montf) .....	71
Tabla 18 – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de monto financiado (R_Montf) .....	71
Tabla 19 – Tabla de contingencia - variable subregión geográfica (SubReg) vs variable presencia mora grave (Mayor90) .....	72
Tabla 20 – Función de supervivencia, método KM, variable subregión geográfica (SubReg).....	74
Tabla 21 – Resultado del contraste de igualdad de funciones de supervivencia para la variable subregión geográfica (SubReg) .....	75
Tabla 22 – Tabla de contingencia - variable rango de edad (R_Edad) vs variable presencia mora grave (Mayor90) .....	76
Tabla 23 – Función de supervivencia, método KM, variable rango de edad (R_Edad).....	77
Tabla 24 – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de edad (R_Edad).....	78
Tabla 25 – Tabla de contingencia - variable actividad económica (Acteco_r) vs variable presencia mora grave (Mayor90) .....	79
Tabla 26 – Función de supervivencia, método KM, variable actividad económica (Acteco_r).....	81
Tabla 27 – Resultado del contraste de igualdad de funciones de supervivencia para la variable actividad económica (Acteco_r) .....	82
Tabla 28 – Tabla de contingencia - variable ciclo económico al momento del desembolso (Ciclo_e) vs variable presencia mora grave (Mayor90) .....	84
Tabla 29 – Función de supervivencia, método KM, variable ciclo económico al momento del desembolso (Ciclo_e) .....	85
Tabla 30 – Resultado del contraste de igualdad de funciones de supervivencia para la variable ciclo económico al momento del desembolso (Ciclo_e) .....	86
Tabla 31 – Tabla de contingencia - variable cesante previo a la mora grave (Cesante) vs variable presencia mora grave (Mayor90) .....	87
Tabla 32 – Función de supervivencia, método KM, variable cesante previo a la mora grave (Cesante) .....	88

Tabla 33 – Resultado del contraste de igualdad de funciones de supervivencia para la variable cesante previo a la mora grave (Cesante) .....	89
Tabla 34 – Árbol de decisión: Variable de segmentación: mora grave (Mayor90)	90
Tabla 35 – Modelo de cox – Clientes de créditos hipotecarios que desempeñaron labores en "Empresa unipersonal / construcción / trabajo autónomo, rise y microempresa / sin información" – significancia individual de variables .....	95
Tabla 36 – Modelo de cox – Clientes de créditos hipotecarios que desempeñaron labores en "Empresa unipersonal / construcción / trabajo autónomo, rise y microempresa / sin información" – significancia global .....	96
Tabla 37 – Modelo de cox – Clientes de créditos hipotecarios que desempeñaron labores en "Empresa unipersonal / construcción / trabajo autónomo, rise y microempresa / sin información" – interpretación de coeficientes.....	97
Tabla 38 – Modelo de cox – Clientes de créditos hipotecarios que desempeñaron labores en "Empresa unipersonal / construcción / trabajo autónomo, rise y microempresa / sin información" – supuestos de riesgo proporcional.....	101
Tabla 39 – Función de supervivencia obtenida mediante la aplicación del modelo de cox propuesto .....	128
Tabla 40 – Modelo de cox 2 (mora mayor a 120 días).....	136
Tabla 41 – Función de supervivencia: Modelo de cox 2 (mora mayor a 120 días) .....	138
Tabla 42 – Supuesto de proporcionalidad, estadístico Rho: Modelo de cox 2 (mora mayor a 120 días).....	143

## LISTA DE ANEXOS

Anexo I – Detalle de la función de supervivencia obtenida mediante la aplicación del modelo de cox propuesto.....	128
Anexo II - Detalle del modelo de cox 2 (mora mayor a 120 días) y su función de supervivencia.....	136
Anexo III - Código de programación R - Etapa 1 - Preparación de la data.....	144
Anexo IV – Código de programación R – Etapa 2 - Análisis descriptivo de las potenciales variables explicativas .....	153
Anexo V – Código de programación R – Etapa 3 - Identificación de variables explicativas relacionadas en mayor medida con la variable dependiente.....	160
Anexo VI - Código de programación R – Etapa 4 - Formulación y validación del modelo de cox .....	160
Anexo VII - Código de programación R – Etapa 5 - Creación de reporte ejemplo .....	162

## RESUMEN

La presente tesis pretende analizar la aplicación del análisis de supervivencia, y específicamente del modelo de cox en un ámbito como el financiero, mismo que no ha sido tradicionalmente abordado en América Latina. Para ello, en primera instancia se ha descrito y analizado la bibliografía existente con respecto a los beneficios que la aplicación del análisis de supervivencia deriva en la dinámica del sistema financiero formal, posteriormente se ha propuesto un procedimiento metodológico que permita aplicar un análisis de supervivencia a una cartera de créditos hipotecario pertenecientes a una entidad financiera ecuatoriana, con el propósito de segmentarla e identificar situaciones internas y externas de cara a la definición de estrategias diferenciadas tanto para la concesión crediticia como para la gestión de cobranzas. Finalmente, se pretendió automatizar el análisis realizado a través de la programación en el software R y con ello facilitar la reproducción y parametrización del mismo de cara a su implementación como parte de las acciones de gestión y mitigación del riesgo de crédito asociado a la cartera hipotecaria analizada.

**Palabras clave:** Análisis de supervivencia, Modelo de Cox, Préstamos Hipotecarios.

## ***ABSTRACT***

This thesis aims to analyze the application of survival analysis, and specifically the cox model in an area such as finance, which has not traditionally been addressed in Latin America. In the first instance the existing bibliography has been described and analyzed with respect to the benefits that the application of the analysis of survival derives in the dynamics of the formal financial system, later a methodological procedure has been proposed that allows to apply an analysis of survival to a portfolio of mortgage loans belonging to an Ecuadorian financial institution, with the purpose of segmenting it and identifying internal and external situations in order to define differentiated strategies for both the credit granting and the collection management. Finally, it was intended to automate the analysis carried out through the programming in the R software and thereby facilitate the reproduction and parameterization of the same for its implementation as part of the management actions and mitigation of credit risk associated with the mortgage portfolio analyzed.

**Keywords:** Survival Analysis, Cox Model, Mortgage Loans.



## 1. INTRODUCCIÓN

El sistema financiero, en tanto se constituye en un esquema organizado que traslada liquidez desde unidades excedentarias hacia unidades deficitarias, aporta significativamente en la transformación del ahorro en inversión, y con ello al crecimiento económico de un país, dependiendo de las particularidades de cada una de las economías (González & López, 2008).

En esta línea, la sostenibilidad de las instituciones financieras, es un elemento relevante en el buen funcionamiento del engranaje denominado sistema financiero. Una de las aristas que contribuyen a este fin, es el adecuado manejo del riesgo de crédito, entendido como la administración de la posibilidad de pérdida suscitada por el no pago de obligaciones acordadas.

Entre los procesos planteados para la medición y gestión del mentado riesgo, se realizan actividades de revisión permanente del portafolio de créditos, las cuales buscan pre-identificar situaciones de mayor probabilidad de incumplimiento o deterioro de cartera a través del monitoreo de diferentes variables.

Para ello, a través de diferentes técnicas estadísticas las diferentes entidades buscan estimar la probabilidad de que un cliente pueda ser buen o mal pagador, de acuerdo a sus particularidades, en el mejor de los casos, a través de la implementación de mecanismos de calificación estadística (scoring) que utilizan modelos binarios logit o probit, y así mejorar los procesos de selección de clientes y posteriormente de gestión de cobranzas (Velandia, 2013).

Por otra parte, se puede mencionar que con el devenir del tiempo, el control de calidad, y específicamente el control estadístico de la calidad, se ha constituido en una práctica, inicialmente aplicada en el sector industrial, que se ha extendido hacia los diferentes sectores productores de bienes y servicios, públicos y privados.

El enfoque de control estadístico de calidad para monitorear el desempeño de un proceso cobra relevancia, ya que a través de la sistematización del mismo se puede tener una visión global y simultánea de los resultados del proceso, haciendo factible la retroalimentación del mencionado de forma oportuna.

En esta lógica se inserta el denominado “análisis de supervivencia”, cuyos inicios se produjeron a través de la elaboración de las tablas de mortalidad atribuidas al astrónomo Edmon Halley quien utilizó en primera instancia un registro de funerales y nacimientos en la ciudad de Breslau (Rebasa, 2005). Tal como se lo conoce hoy, el análisis de supervivencia se lo aplicó inicialmente en la ingeniería, encaminado a analizar la duración y fiabilidad de los elementos que forman una máquina.

Puntualmente, el análisis de supervivencia es un conjunto de procedimientos estadísticos que posibilitan el estudio de los datos y cuya variable de interés primordial es el tiempo de duración de un evento.

La problemática que pretende abordar la presente investigación aplicada, es develar los beneficios que la introducción del análisis de supervivencia, derivado del enfoque de control estadístico de calidad, puede detonar en el análisis y administración del riesgo de crédito en entidades financieras, y puntualmente en el caso de la cartera hipotecaria generada por una entidad financiera ecuatoriana.

Específicamente, mediante las técnicas estadísticas tradicionales se estima la probabilidad de pago y no pago de un cliente en una entidad financiera, pero no se identifica y monitorea el tiempo que transcurre desde la concesión del crédito hasta un posible evento de no pago y las razones que motivan ese comportamiento.

Es así, que a través de un ejercicio empírico, utilizando el análisis de supervivencia y particularmente el modelo de Cox (semi-paramétrico) se pretende destacar los beneficios que se suscitan de incorporar al tratamiento tradicional del riesgo de crédito el análisis de la temporalidad de no pago, ya que con ello se puede implementar con mayor certeza y de forma anticipada políticas de concesión y gestión crediticia que permitan mantener saludable la cartera de créditos de las diferentes entidades financieras. Cabe resaltar, que en la bibliografía existente, la aplicación de este tipo de modelos en el sector financiero aún es incipiente a nivel de América Latina y obviamente en el Ecuador, por lo cual se torna de interés la realización de investigación aplicada en este campo.

Pese a lo mencionado, existe bibliografía de planteamientos y aplicación del análisis de supervivencia vinculado al Sistema Financiero por fuera de América Latina, lo cual denota la factibilidad real de implementar este tipo de análisis a nivel Ecuador. Ejemplo de lo expuesto ha sido la aplicación del análisis de supervivencia con información de clientes de crédito de consumo en un banco Español (Cao, Vilar, & Devia, 2009), con información de créditos hipotecarios pertenecientes a minoristas en Estados Unidos (Zhang, 2009) y con información de créditos personales derivados de las instituciones que ofertan servicios financieros en Australia (Watkins, Vasnev, & Gerlach, 2009) entre algunos destacados.

Adicionalmente, cabe señalar que coyunturalmente, este tipo de análisis aplicado cobra relevancia a nivel país ya que dada la etapa de desaceleración económica que atraviesa el Ecuador, luego de una etapa de cíclica de recuperación y expansión económico positiva entre marzo 2010 y septiembre 2014 (BCE, 2016a) (Banco Central del Ecuador (a), 2016), en la cual se canalizó un volumen significativo de crédito hipotecario (Romero & Meléndez, 2016); es importante para las entidades financieras identificar anticipadamente el comportamiento de pago de sus clientes, y puntualmente identificar la temporalidad en la cual pueden suscitarse el evento de “no pago” y con ello tomar decisiones acertadas en pos de mantener la sostenibilidad institucional y con ello a nivel agregado, mantener el financiamiento hacia el sector de la construcción, que a través de su valor agregado aportó al PIB entre los años 2010 y 2015 en alrededor del 9,4% (Banco Central del Ecuador (b), 2016).

## **1.1. Preguntas de investigación**

¿De qué manera el análisis de supervivencia contribuye al buen funcionamiento del sistema financiero formal?

¿Empíricamente: en qué medida la utilización del análisis de supervivencia mediante el modelo de Cox permite identificar la probabilidad de que un crédito hipotecario puede caer en mora en un período de tiempo e identificar los factores que incidirían en el suceso de ese evento, y con ello detonar la identificación de grupos de clientes y situaciones externas que permitan la definición de estrategias diferenciadas tanto para la concesión crediticia como para la gestión de cobranzas?

¿Es factible automatizar la reproducción del análisis de supervivencia con la finalidad de facilitar la implementación práctica de este tipo de análisis en la gestión del riesgo de crédito en la entidad ecuatoriana generadora de la cartera hipotecaria analizada?

## **1.2. Objetivo general**

Implementar el análisis de supervivencia, a través del caso de estudio de una entidad financiera ecuatoriana que oferta crédito hipotecario, con la finalidad de develar los beneficios que su utilización implicaría en el tratamiento y gestión del riesgo de crédito.

## **1.3. Objetivos específicos**

Describir y analizar los beneficios que la aplicación del análisis de supervivencia puede generar al utilizarse dentro de la dinámica del Sistema Financiero Formal.

Aplicar a través del modelo de COX un análisis de supervivencia para los créditos del producto préstamo hipotecario de una entidad financiera ecuatoriana, con la finalidad de segmentar la base de clientes e identificar situaciones externas

de cara a la definición de estrategias diferenciadas tanto para la concesión crediticia como para la gestión de cobranzas.

Implementar el análisis de supervivencia en la entidad generadora de la cartera hipotecaria ecuatoriana analizada mediante la automatización de estimaciones del modelo de Cox a través del uso de software R.

#### **1.4. Hipótesis**

La aplicación e implementación de un modelo de supervivencia conllevaría beneficios de cara a la gestión del riesgo de crédito de una cartera hipotecaria ecuatoriana a través del análisis de la temporalidad en la que podrían suscitarse los “no pagos”, de la identificación de las variables que incidirían en el comportamiento referido; y de la implementación de políticas de crédito y gestión que permitan mitigar anticipadamente los riesgos identificados.

## 1.5. Marco Teórico

### 1.5.1 Análisis de Supervivencia y Modelo de Cox

#### Análisis de Supervivencia

##### *Introducción*

Según Klein & Moeschberger (1997), el análisis de supervivencia es una "Técnica que permite describir el comportamiento de los datos que corresponden a tiempo o duración desde un origen hasta la ocurrencia de un cambio".

Explícitamente, este análisis se utiliza cuando la variable dependiente representa un período de tiempo entre una situación inicial y el evento final. Es así que el análisis de supervivencia (*survival analysis*) permite identificar el efecto de una o más variables independientes cuando la variable dependiente puede expresarse en términos de tiempo hasta que se suscita el evento final.

Es decir, dada una variable cuyos valores corresponden al tiempo que transcurre hasta que se produce un determinado evento final, el análisis referido permite estimar en función del tiempo, la probabilidad de que ocurra dicho evento.

Consecuentemente, el análisis de supervivencia es un análisis de alerta temprana, derivado del control estadístico de calidad, que permite identificar el lapso (tiempo) que tarda en ocurrir un evento de interés y las variables que inciden en ello.

En terminología estadística puede entenderse como la probabilidad condicional de que ocurra un evento de interés en el período  $t$ , dado que este no había ocurrido en el período  $t-1$ .

En este análisis también es factible utilizar factores explicativos constantes y variables en diferentes períodos temporales.

## **Conceptos relevantes**

### Evento de interés y tiempo de falla

Un evento de interés es un suceso identificado por el investigador sobre el cual se requiere conocer el tiempo que tarda su realización desde un tiempo inicial definido (tiempo de falla).

### Censura

Una de las características distintivas del análisis de supervivencia es la consideración de datos censurados. La censura se suscita en aquellos datos en los cuales se dispone de alguna información acerca del tiempo de supervivencia individual, pero no se dispone exactamente del mismo (Meira-Machado, Introduction to Survival Analysis, 2016).

En otras palabras, en el análisis de supervivencia existen algunos individuos que experimentan el evento de interés o de falla durante el período de estudio y otros que no. En el caso de estos últimos, el tiempo de supervivencia es desconocido y con ello la información de estos casos se considera censurada (Rebasa, 2005).

El fenómeno de la censura se puede producir de diferentes formas, es así que los tipos identificados son los siguientes (Meira-Machado, Introduction to Survival Analysis, 2016):

*Censura por la derecha.*- Se presenta cuando el evento de interés o falla no se ha suscitado durante el período de estudio. Esta situación generalmente responde a los siguientes criterios y puede suscitarse de forma aleatoria: 1) Una pérdida de seguimiento, 2) Conclusión del estudio antes de tener lugar el evento y 3) El individuo abandona el estudio.

Adicionalmente, este tipo de censura se sub-clasifica en: censura tipo I (existe un tiempo predeterminado hasta el cual se debe suscitarse el evento) y censura tipo II (se da por terminado el estudio cuando un número  $r$  de  $n$  individuos que ingresaron al estudio incurren en el evento de interés).

*Censura por la izquierda.*- Se presenta cuando el evento de interés o de falla se ha suscitado de forma previa al período de estudio.

*Censura por intervalo.*- Se presenta cuando el evento de interés o de falla exclusivamente se suscita en un intervalo de tiempo específico.

### **Modelo de Supervivencia**

El modelo de supervivencia está caracterizado por variables aleatorias no negativas. Es así que parte de la definición de una variable aleatoria  $T$ , la cual representa el tiempo de falla o el tiempo de supervivencia. El comportamiento de la variable  $T$  puede ser caracterizado a partir de algunas funciones básicas dentro de las que destacan la función de supervivencia y la función de riesgo.

#### Función de supervivencia

Es aquella que describe la probabilidad de que un individuo sobreviva (no le ocurra el evento de interés) al menos hasta el período  $t$  (Godoy Aguilar, 2009).

Sea  $T$  una variable aleatoria no negativa cuya función de distribución es  $F(t)$  y cuya función de densidad es  $f(t)$ . Su correspondiente función de supervivencia  $S(t)$  sería:

$$S(t) = P(T > t) = 1 - F(t) \quad (1)$$

$$S(t) = P(\text{Un individuo sobreviva más allá de } t)$$

La función  $S(t)$  al ser no negativa y no creciente, implica que  $S(0) = 1$  y  $S(t) = 0$  cuando  $t$  tiende al infinito. Explícitamente, la probabilidad de sobrevivir al menos al tiempo 0 es 1 y la probabilidad de sobrevivir un tiempo infinito es 0.

Cuando la variable aleatoria  $T$  es continua, su correspondiente función de supervivencia ( $S(t)$ ) es la integral de la función de densidad de probabilidad.

$$S(t) = P(T > t) = \int_t^{\infty} f(t)dt \quad (2)$$



Cuando la variable aleatoria  $T$  es discreta, su correspondiente función de supervivencia ( $S(t)$ ) figura de la siguiente manera:

$$S(t) = P(T > t) = \sum_{t_j > t} f(t_j) \quad (3)$$

Función de riesgo

Es aquella que denota el riesgo instantáneo de que un evento de interés suceda en un intervalo de tiempo reducido ( $t, \Delta t$ ), dado que no ha ocurrido hasta el tiempo  $t$ . La expresión matemática correspondiente sería la siguiente (Barrera Rebellon, 2008):

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} \quad (4)$$

Trabajando la probabilidad condicional anterior el resultado sería:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{P(t \leq T < t + \Delta t)}{P(T > t)} \quad (5)$$

Desarrollando la expresión (5) en función de la expresión (1) se puede obtener lo siguiente:

$$h(t) = \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \quad (6)$$

Conforme a la definición de derivada, la expresión anterior se convierte en:

$$h(t) = f(t)/S(t) \quad (7)$$

Esta expresión también se conoce como función de razón de riesgos o tasa instantánea de fallas.

Adicionalmente, si se integra la función de riesgo (7) se obtiene la siguiente función de riesgo acumulado:

$$H(t) = \int_0^t h(s) ds = \int_0^t \frac{f(s)}{S(s)} ds = \int_0^t \frac{f(s)}{1 - F(s)} ds \quad (8)$$

Si en la expresión anterior se realiza la sustitución de  $u = 1 - F(s)$  y  $du = -f(s)ds$ , el resultado sería el siguiente:

$$H(t) = \int \frac{-du}{u} \quad (9)$$

A partir de esta expresión se pueden obtener las siguientes funciones relacionadas:

$$H(t) = -\ln S(t) \quad (10)$$

$$S(t) = e^{-H(t)} \quad (11)$$

$$h(t) = -\frac{d(\ln S(t))}{dt} \quad (12)$$

$$h(t) = -\frac{dS(t)}{dt} \quad (13)$$

### Estimador de Kaplan y Meier

El estimador Kaplan-Meier desarrollado en 1958 es un método no paramétrico que estima la probabilidad de supervivencia  $\hat{S}(t)$  en el momento en el que se ha suscitado el evento. Para ello, es importante tener en cuenta que para sobrevivir en un momento determinado  $t$ , se tiene que haber sobrevivido en todos los instantes anteriores a  $t$  (Meira-Machado, 2016).

El cálculo de este método parte de una probabilidad condicional compuesta, en la cual se multiplica la supervivencia en el instante anterior y la tasa de supervivencia en ese instante. El detalle explícito del cálculo de la probabilidad de supervivencia bajo este método sería el siguiente (Herranz Valera, 2015):

$$\hat{S}(t) = \prod_{j: t_j \leq t} \left[ 1 - \frac{d_j}{n_j} \right] \quad (14)$$

Donde  $d_j$  es el número de individuos a los cuales les suscitó el evento de interés en el tiempo  $t_j$ ,  $n_j$  es el número de individuos que se encontraban en riesgo de que les suscitara el evento de interés. Adicionalmente  $n_j$  es la muestra original de individuos menos aquellos que fueron censurados o aquellos a los que les suscitó el evento de interés antes del tiempo  $t_j$ .

### Modelo de Cox

### Definición y funciones características

Es una forma funcional utilizada en el análisis de supervivencia conocido también como modelo de riesgo proporcional (desarrollada en 1972), que hace uso de una propuesta semi-paramétrica para encontrar una función que describa el riesgo de ocurrencia de un evento en diferentes períodos de tiempo (Ayala, Borges, & Colmenares, 2007).

El modelo se asemeja a una regresión que describe la relación entre la ocurrencia de un evento de interés o falla y un conjunto de variables. Formalmente, cuando  $t \geq 0$  la referida función de riesgo luce de la siguiente manera (Obuda, 2014):

$$h(t|x) = h_0(t)e^{B'x} \quad (15)$$

Donde  $B'x$  es el vector de covariables, la parte paramétrica  $e^{B'x}$  se la conoce como “puntaje de riesgo” y la parte no paramétrica  $h_0(t)$  se la conoce como “función de riesgo base” o aquella que describe la probabilidad de ocurrencia del evento cuando todas las variables explicativas son 0.

Algunos de las características relevantes del modelo de Cox son:

- Es producto de una función en  $t$  y una función en  $x$ ;
- $x$  es independiente del tiempo;
- La función de riesgo base  $h_0(t)$  no depende de  $x$  sino exclusivamente del tiempo;
- El exponencial (e) considera las  $x$  pero no  $t$ ;
- Sigue los supuestos del modelo de riesgo proporcional;
- Las estimaciones de riesgo son no negativas.

La función de supervivencia relacionada al modelo de Cox es la siguiente:

$$S(t|x) = e^{-H_0(t)e^{(B'x)}} \quad (16)$$

La función de distribución del modelo de Cox es la siguiente:

$$F(t|x) = 1 - e^{-H_0(t)e^{(B'x)}} \quad (17)$$

Adicionalmente, sea  $Z$  una variable aleatoria con una función de distribución  $F$ , donde  $G = F(Z)$  sigue una distribución uniforme en el intervalo comprendido entre 0 y 1 ( $G \sim Uni[0,1]$ ). Conforme a lo mencionado  $1-G$  también seguiría una distribución uniforme ( $1 - G \sim Uni[0,1]$ ). Si en la ecuación de  $G$  se considera el tiempo de supervivencia del modelo de Cox, se tiene que:

$$G = e^{-H_0(t)e^{(B'x)}} \sim Uni[0,1] \quad (18)$$

Si  $h_0(t) > 0$  para todo  $t$ , y  $H_0(t)$  puede ser invertido, el tiempo de supervivencia  $T$  en el modelo Cox puede ser expresado de la siguiente manera:

$$T = H_0^{-1}[\log G]e^{(B'x)} \quad (19)$$

Esta ecuación es idónea para la generación de tiempos de supervivencia con números aleatorios que siguen una distribución uniforme.

### Supuestos del modelo de riesgo proporcional

Las variables explicativas actúan multiplicativamente sobre el riesgo en cualquier punto del tiempo. En virtud de lo anterior, el supuesto clave del modelo proporcional es que el riesgo de un evento en un grupo es un múltiplo constante del riesgo en cualquier otro grupo (Obuda, 2014).

Así también se puede mencionar que las tasas de riesgo son proporcionales cuando la relación entre el riesgo de dos individuos con variables explicativas  $Z$  y  $Z^*$  es constante en el tiempo, siendo el riesgo calculado para cada predictor una proporción del riesgo base  $h_0$  (Almeida Baroja, 2011).

Específicamente, los supuestos del modelo de riesgo proporcional son los siguientes:

- No existe influencia de los individuos en la estimación del modelo.
- No existe influencia de los individuos en la estimación de cada parámetro del modelo.
- Se supone inexistencia de heterogeneidad no observada.
- Se supone una adecuada forma funcional de cada variable continua.

### Método de estimación del modelo de Cox

Generalmente los modelos de regresión se estiman mediante el método de máxima verosimilitud, que radica en estimar los parámetros a partir de la maximización del logaritmo natural de la función de verosimilitud. En el caso del modelo de Cox, este procedimiento no es viable por cuanto se desconoce *a priori* la función de riesgo base considerada en el modelo (Barrera Rebellon, 2008).

En virtud de la anterior, Cox en el año 1972 desarrolló un método de estimación el cual no contempla una forma específica para la función de riesgo base y propone la utilización de una función denominada “de verosimilitud parcial” la cual exclusivamente depende de los parámetros de interés y se expresa de la siguiente manera (Barrera Rebellon, 2008):

$$l_p = \prod_{i=1}^k \frac{e^{x_i B}}{\sum_{j \in R(ti)} e^{x_j B}} \quad (20)$$

Donde la multiplicatoria se da en los distintos k tiempos de supervivencia y  $x_{(i)}$  hace referencia al valor de la covariable para el individuo en el tiempo  $ti$ , y  $R(ti)$  hace referencia a los individuos en riesgo en el tiempo  $ti$  o con tiempos de supervivencia o censura superiores a  $ti$ .

Si se calcula el logaritmo natural de la función de verosimilitud parcial se tiene la siguiente expresión:

$$L_p = \sum_{i=1}^k \{x_{(i)} B - \ln[\sum_{j \in R(ti)} e^{x_j B}]\} \quad (21)$$

La derivada de la expresión (21) con respecto a  $B$  resulta de la siguiente forma:

$$\frac{\partial L_p(B)}{\partial B} = \sum_{i=1}^k \left\{ x_{(i)} - \frac{\sum_{j \in R(ti)} x_j e^{x_j B}}{\sum_{j \in R(ti)} e^{x_j B}} \right\} \quad (22)$$

La expresión (22) se debe igualar a 0 y resolver la ecuación a través de un método numérico como el de Newton-Rapson con un criterio de parada dada la complejidad de la igualdad.

La segunda derivada de la expresión (21) es la estimación de la varianza del parámetro y luce de la siguiente forma:

$$\frac{\partial^2 L_p(B)}{\partial B^2} = - \sum_{i=1}^k \left\{ \frac{[\sum_{j \in R(ti)} e^{x_j B}][\sum_{j \in R(ti)} x_j^2 e^{x_j B}] - [\sum_{j \in R(ti)} x_j e^{x_j B}]^2}{[\sum_{j \in R(ti)} x_j e^{x_j B}]^2} \right\} \quad (23)$$

El lado derecho de la expresión (23) se denota como  $I(B)$  y se la conoce como “información observada”, en el caso de una covariable y de “matriz de información observada”, en el caso de más de una covariable.

El estimador de la varianza del coeficiente estimado es el inverso de la expresión (23) evaluado para  $\hat{B}$

$$\widehat{Var}(\hat{B}) = I^{-1}(\hat{B}) \quad (24)$$

Finalmente, la desviación estándar  $\widehat{SE}(\hat{B})$  del coeficiente estimado es la raíz cuadrada de la expresión (24).

## 1.5.2 El Análisis de Árboles de Decisión

Los árboles de clasificación o de decisión son un método utilizado frecuentemente en la minería de datos. Esta metodología se deriva de la denominada *automatic interaction detection*. Se constituye en una técnica interesante para la exploración inicial de datos sobre todo cuando existe un número importante de datos (Universidad Carlos III, 2008).

La construcción de un árbol de decisión, partiendo de una muestra de entrenamiento que incluye la información del grupo al que pertenece cada caso (criterio de clasificación), implica el cumplimiento de un algoritmo que hace lo siguiente:

- En un nodo inicial, se pretende dividir el conjunto de datos en dos partes lo más homogéneas posibles con la ayuda de una de las variables disponibles. Para ello se define un punto de corte en la data de la variable que permita la segmentación.
- El proceso se repite en adelante seleccionando variables y sus correspondientes puntos de corte hasta cuando se hayan clasificado la totalidad de las observaciones.

La decisión de la variable a seleccionar para la partición en cada nodo se la realiza a partir de la proporción de observaciones que pasan por el nodo para cada uno de los grupos.

Sea  $t$  cada uno de los nodos,  $p(g|t)$  la probabilidad de que las observaciones que llegan al nodo  $t$  pertenezcan a cada una de las clases e  $I(t)$  la impureza del nodo  $t$  (medida de entropía o diversidad) que se define de la siguiente manera:

$$I(t) = -\sum_{g=1}^G p(g|t) * \log p(g|t) \quad (25)$$

La variable que introduce en el nodo es la que minimiza la heterogeneidad o impureza que resulta de la división en el nodo.

La clasificación de las observaciones en los nodos terminales se realiza mediante la asignación de todas las observaciones del nodo al grupo más probable, es decir, al grupo con máxima  $p(g|t)$ . Si la impureza del nodo es cero, todas las observaciones pertenecerían al mismo nodo, o en su defecto podrían existir errores de clasificación.

Cuando el número de variables es grande, el árbol puede contener un número importante de nodos, circunstancia en la cual se hace necesario definir un procedimiento de poda o simplificación del mismo. Usualmente la poda se daría en el nodo en el cual el error de clasificación deje de disminuir (Universidad Carlos III, 2008).



## **2. METODOLOGÍA**

### **2.1. Procedimiento metodológico**

Metodológicamente para atender el primer objetivo específico, se realizará el desarrollo de tres temáticas relevantes. Un primero relacionado a develar desde la aportación de varios autores la relevancia del sistema financiero, sus principales características y su incidencia en el crecimiento económico de un país, un segundo relacionado con la descripción de las experiencias de implementación del análisis de supervivencia en el sistema financiero formal y un tercero relacionado a la caracterización del sistema financiero ecuatoriano, su coyuntura y la dinámica del crédito hipotecario en el país.

Posteriormente, para atender el segundo objetivo se elaborará un análisis de supervivencia, mediante el modelo no paramétrico de Kaplan y Meier y particularmente mediante el modelo de Cox para una base de créditos hipotecarios concedidos entre el año 2008 y 2014 perteneciente a una entidad financiera ecuatoriana. El evento de interés que se constituirá en la variable de interés será la mora mayor a 90 días. Es así que la función de supervivencia a estimar estará relacionada con el tiempo que existe desde el desembolso crediticio hasta la ocurrencia de una mora grave.

El procedimiento metodológico que se aplicará será el siguiente:

- a) Definición de la variable de interés.
- b) Breve descripción de la base de datos a utilizar.
- c) Análisis descriptivo de las variables consideradas inicialmente para formar parte del modelo de cox y construcción individual de funciones de supervivencia mediante el método de Kaplan y Meier.
- d) Aplicación del análisis de decisión (árbol de decisión) para identificar aquellas variables que se relacionan en mayor medida con el evento de interés.
- e) Formulación del modelo de Cox.
- f) Verificación de los supuestos del modelo de Cox.
- g) Interpretación de los resultados

- h) Calibración del modelo frente a un cambio de la variable dependiente.
- i) Segmentación de la base de clientes de la cartera de créditos hipotecarios, a partir del análisis de supervivencia desarrollado con la finalidad de plantear estrategias diferenciadas tanto de concesión como de gestión de cobranzas.

Finalmente, para atender el tercer objetivo se automatizará el análisis de supervivencia a través de la programación del análisis en el software R, con lo cual se podrá facilitar la reproducción y parametrización del análisis de cara a la implementación del análisis como parte de las acciones dentro de la mitigación y gestión de los riesgos de crédito asociados a la cartera hipotecaria analizada.

## **2.2. Detalle de actividades relevantes**

### **2.2.1. Definición de la variable de interés**

Se define el evento de interés a partir del cual se definirá el tiempo de supervivencia a estimar mediante las diferentes técnicas de supervivencia.

### **2.2.2. Breve descripción de la base de datos a utilizar**

Consiste en una descripción sucinta de las características de la base de datos a utilizar. Específicamente, de detalla el tipo de caso a utilizar, el alcance temporal de los mismos y el número de observaciones de la base de datos.

### 2.2.3. Análisis descriptivo de las potenciales variables explicativas

#### Análisis descriptivo y tabla de contingencia

Para cada una de las variables consideradas se realiza un análisis descriptivo de las mismas. Puntualmente, en cada variable se caracteriza su distribución y se la compara con la variable de interés (mora grave) con la finalidad de identificar las opciones de respuesta de la variable en la cual existe una mayor incidencia de la variable de interés (mora grave).

Este último análisis se lo realiza mediante la utilización de una tabla de contingencia a partir de la cual se analiza la distribución conjunta de las variables y también se calcula el estadístico ji-cuadrado ( $\hat{X}^2$ ) para validar o rechazar la hipótesis nula (0) “independencia de variables”.

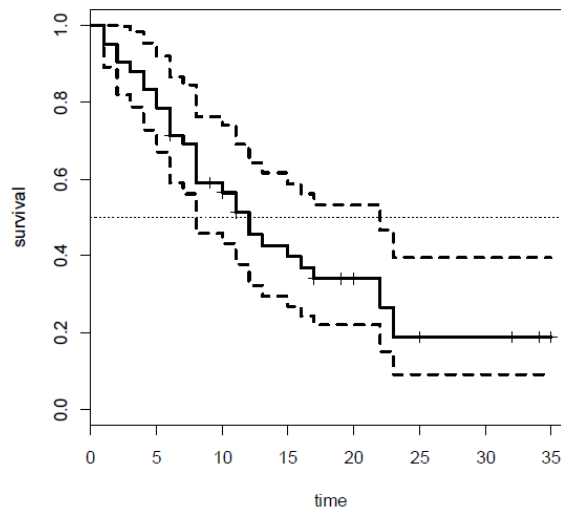
$$\hat{X}^2 = \frac{\sum_{i=1}^h \sum_{j=1}^k (n_{ij} - E_{ij})^2}{E_{ij}} \quad (26)$$

El valor ( $\hat{X}^2$ ) calculado se compara con el valor de la tabla ( $X^2$ ) para un nivel de confianza (n-1) (k-1) grados de libertad. Si el valor calculado es mayor que el valor de tablas, significará que las diferencias entre las frecuencias observadas y las frecuencias teóricas son elevadas, con lo cual con un nivel específico de confianza se afirmará que existe dependencia entre los factores o atributos analizados (Vicéns Otero & Medina Moral, 2005).

Específicamente, con un nivel de confianza del 95%, p-valores asociados al estadístico ji-cuadrado inferiores a 0,05 permiten rechazar la  $h_0$ , con lo cual se confirmaría una relación de dependencia estadísticamente significativa entre la variable independiente analizada y la variable de interés (mora grave); por el contrario, p-valores asociados al estadístico ji-cuadrado superiores a 0,05 permiten aceptar la  $h_0$ , con lo cual no existiría una relación de dependencia estadísticamente significativa entre las variables analizadas.

## Análisis de Kaplan y Meier para cada variable independiente

Conforme se lo detalló en el marco teórico el método de Kaplan-Meier (KM) es un método no paramétrico que estima la función de supervivencia, probabilidad acumulada de que no haya ocurrido el evento. Empieza en 1 y es una curva decreciente escalonada.



**Figura 1** – Gráfico función de supervivencia KM, ejemplo

El detalle de la función de supervivencia KM se presenta en una tabla con todos los tiempos existentes en el conjunto de datos, el número de casos en riesgo de incurrir en el evento de interés por tiempo, el número de eventos de interés que se suscitaron por tiempo, la probabilidad de supervivencia por tiempo, el error estándar, y el intervalo de confianza al 95% de confianza para la probabilidad de supervivencia.

El modelo de supervivencia KM que se aplica para cada variable es aquel construido a partir de la variable dependiente, tiempo de supervivencia hasta el evento interés (mora grave), acompañada de la variable evento de interés (mora grave), la cual indica la ocurrencia del evento (1) o la no ocurrencia del mismo (0) y censura por la derecha del caso. La variable dependiente planteada será explicada por cada una de las potenciales variables independientes.

La salida del modelo de supervivencia KM para cada variable presenta en las filas el detalle de las opciones de respuesta de la variable y en las columnas el número de casos por cada opción de respuesta, el número de eventos de interés

suscitados por cada opción de respuesta, la mediana de los tiempos de supervivencia por cada opción de respuesta y el correspondiente intervalo de confianza. En el programa estadístico R, la función `survfit()` es aquella que calcula el estimador KM de la función de supervivencia correspondiente.

**Tabla 1** – Ejemplo, cálculo estadístico KM para función de supervivencia

```
## Call: survfit(formula = Surv(tiempo supervivencia (t), presencia evento de interés(M)
~ potencial variable independiente (K)

##           n events median 0.95LCL 0.95UCL
## K=0  3681   3681   27.4    26.7    27.9
## K=1 75544  4073   20.0    19.8    20.5
```

Es importante señalar que la mediana y media de los tiempos de supervivencia no se puede calcular cuando existe la presencia de una importante cantidad de casos censurados (Meira-Machado, *Survival Analysis: Kaplan-Meier estimate*, 2016).

Finalmente, se realiza una comparación de las funciones de supervivencia que se generan para cada opción de respuesta de la variable en análisis, con la finalidad de identificar si los comportamientos son estadísticamente diferentes para cada opción de respuesta.

La citada comparación se la lleva a cabo mediante el estadístico log-rank (Q), el cual se calcula mediante el cociente entre la sumatoria de las diferencias entre los eventos de interés observados y los eventos esperados para la totalidad de los tiempos de supervivencia observados; y la estimación de varianza. El estadístico Q sigue una distribución ji-cuadrado con 1 grado de libertad

$$Q = \frac{[\sum_{i=1}^m (d_{1i} - \hat{e}_{1i})]^2}{\sum_{i=1}^m \hat{v}_{1i}} \quad (27)$$

La hipótesis nula a evaluar mediante el estadístico Q ( $h_0$ ) es la igualdad de las dos (2) curvas de supervivencia. Es así que a un nivel de confianza del 95%, si los p-valores asociados al estadístico Q son inferiores a 0,05, se rechaza la hipótesis nula ( $h_0$ ) de igualdad de funciones y se concluye que las funciones de supervivencia son estadísticamente diferentes. Por el contrario, a un nivel de confianza del 95%, si los p-valores asociados al estadístico Q son superiores a

0,05, se acepta la hipótesis nula ( $h_0$ ) de igualdad de funciones y se concluye que las funciones de supervivencia son estadísticamente iguales (Herranz Valera, 2015).

#### **2.2.4. Verificación de los supuestos del modelo de cox**

##### **Validación de la significancia individual de las variables**

La validación individual de la significancia estadística de las variables en el modelo se realiza mediante estadístico  $z$  que sigue una distribución normal.

La hipótesis nula ( $h_0$ ) evaluada mediante el estadístico  $z$  es: el coeficiente ( $\beta$ ) es igual a 0.

Es así, que a un nivel de confianza del 95% si el  $p$ -valor asociado al estadístico es inferior a 0,05 se rechaza la  $h_0$ , con lo cual se puede afirmar que el coeficiente analizado es estadísticamente diferente de 0 y con ello es significativo. Por el contrario, si el  $p$ -valor asociado al estadístico es superior a 0,05 no se rechaza la  $h_0$ , con lo cual se puede afirmar que el coeficiente analizado es igual a 0 y con ello no es estadísticamente significativo.

##### **Validación global del modelo**

La validación global del modelo de cox puede dar mediante tres (3) estadísticos: 1) prueba de razón de verosimilitud, 2) prueba de wald, y 3) prueba de puntajes. La hipótesis nula ( $h_0$ ) que evalúan los tres estadísticos es: los coeficientes del modelo en su conjunto son iguales a 0. De esta manera con un nivel de confianza del 95%, si el  $p$ -valor asociado a las diferentes pruebas es inferior a 0,05 se rechaza la  $h_0$ , con lo cual se puede afirmar los coeficientes estimados son estadísticamente diferentes de 0 y con ello los coeficientes en su conjunto son significativos. Por el contrario, si el  $p$ -valor asociado a las diferentes pruebas es superior a 0,05 no se rechaza la  $h_0$ , con lo cual se puede afirmar que los coeficientes en su conjunto son iguales a 0 y no son estadísticamente significativos.

### **Prueba de razón de verosimilitud (likelihood ratio test)**

Identificada como la prueba de mayor fiabilidad, puede definirse con la siguiente expresión:

$$X_{LR}^2 = 2[\log(L(\beta_0)) - \log(L(\hat{\beta}))] \quad (28)$$

Donde  $\beta_0$  son los valores iniciales de los coeficientes y  $\hat{\beta}$  son los valores ajustados de los coeficientes mediante el modelo de cox. Esta prueba sigue una distribución ji-cuadrado con n grados de libertad si  $h_0$  es cierta para muestras grandes (Pruenza García-Hinojosa, 2014).

### **Prueba de wald (wald test)**

Conocida como la prueba de mayor naturalidad, puede definirse con la siguiente expresión:

$$X_W^2 = (\hat{\beta} - \beta_0)^T I^{-1}(\hat{\beta})(\hat{\beta} - \beta_0) \quad (29)$$

Donde  $I(\beta)$  es la matriz de varianzas y covarianzas estimada. Esta prueba sigue una distribución ji-cuadrado con n grados de libertad si  $h_0$  es cierta para muestras grandes (Pruenza García-Hinojosa, 2014).

### **Prueba puntajes (score [logrank])**

La prueba de puntajes puede definirse con la siguiente expresión:

$$X_{SC}^2 = U(\beta_0)^T I^{-1}(\beta_0)U(\beta_0) \quad (30)$$

Donde U es el vector de derivadas de la función de log-verosimilitud parcial  $\log(L(\beta))$ . Esta prueba sigue una distribución ji-cuadrado con n grados de libertad si  $h_0$  es cierta para muestras grandes (Pruenza García-Hinojosa, 2014).

## **Validación de los supuestos de riesgo proporcional**

El supuesto de riesgo proporcional es uno de las más relevantes del modelo de cox. Este implica que las tasas de riesgo son proporcionales siempre que el cociente entre el riesgo de dos individuos con variables explicativas Z y Z\* se mantengan constantes en el tiempo, siendo el riesgo calculado para cada coeficiente una proporción del riesgo base h<sub>0</sub> (Almeida Baroja, 2011).

Conforme a lo expuesto, la razón de riesgo relativo se define como:

$$\frac{\lambda(t, Z_i(t))}{\lambda(t, Z_j^*(t))} = \frac{\lambda_0(t) e^{\beta' Z_i(t)}}{\lambda_0(t) e^{\beta' Z_j^*(t)}} = \frac{e^{\beta' Z_i(t)}}{e^{\beta' Z_j^*(t)}} = e^{\beta'(Z_i(t) - Z_j^*(t))} \quad (31)$$

La verificación del supuesto planteado se lo realiza a través de los residuos de Schonenfeld, mismos que se definen de la siguiente manera:

$$s_{ij}(\beta) = Z_{ij}(t_i) - \bar{Z}_j(\beta, t_i) \quad (32)$$

Donde i con los individuos, t<sub>i</sub> es el tiempo de ocurrencia del evento y j la definición de las variables independientes.

La distribución aleatoria de los residuos de Schonenfeld corrobora la hipótesis de riesgo proporcional.

Adicionalmente, la prueba estadística relacionada es la prueba (rho) misma que sigue una distribución ji-cuadrada y cuya hipótesis nula (h<sub>0</sub>) es el cumplimiento de los supuestos de riesgo proporcional.

De esta manera con un nivel de confianza del 95%, si el p-valor asociado al estadístico rho, es mayor a 0,05 se acepta (h<sub>0</sub>) y se puede afirmar que no existe evidencia estadística para rechazar el cumplimiento de los supuestos de proporcionalidad. Por su parte, si el p-valor asociado al estadístico rho, es menor a 0,05 se rechaza (h<sub>0</sub>) con lo cual se puede mencionar que el modelo no cumple con los supuestos de riesgo proporcional.



### **Influencia de los casos en la estimación del modelo**

Los residuos de los desvíos o desviante permiten la identificación de casos atípicos. El cálculo de este tipo de residuos se lo realiza mediante una transformación de normalización de los desvíos de martingala. Si todas las covariables son fijas en el tiempo los residuos toman la siguiente forma:

$$d_i = \text{signo}(\widehat{M}_i) * \sqrt{-\widehat{M}_i - N_i \log\left(\frac{N_i - \widehat{M}_i}{N_i}\right)} \quad (33)$$

La no presencia de casos atípicos refiere que estos no están influenciando al modelo de supervivencia desarrollado.

### **Influencia de los casos en la estimación de cada coeficiente**

Los residuos tipo puntaje o score permiten verificar la influencia individual de cada coeficiente y para la estimación robusta de la varianza. Puntualmente se definen de la siguiente manera (Hernández Domínguez, 2010):

$$U_{ij} = U_{ij}(\hat{\beta}, \infty) \quad (34)$$

Donde  $U_{ij}(\beta, t), j = 1 \dots p$  son las componentes del vector fila de longitud  $p$ , obtenido mediante el proceso de puntaje para el  $i$ -ésimo individuo:

$$U_i(\beta) = \int_0^t [Z_i(t) - \bar{Z}(\beta, t)] dN_i(t) \quad (35)$$

La no presencia de casos atípicos refiere en el caso de cada coeficiente que los anteriores no están influenciando su estimación.

## **Forma funcional de las variables continuas**

Los residuos de martingala permiten analizar la forma funcional de una variable continua que participa de un modelo de supervivencia. Específicamente se definen de la siguiente manera:

$$\widehat{M}_i(t) = N_i(t) - \widehat{E}_i(t) = N_i(t) - \int_0^t Y_i(s) e^{\beta' Z_i(s)} d\widehat{R}_0(\beta, s) \quad (36)$$

Donde  $\widehat{M}_i(t)$  se desprende de la martingala de un proceso de conteo para el  $i$  – ésimo individuo y se define como la diferencia entre el proceso de conteo  $N_i(t)$  y la integral de la función de intensidad (Pruenza García-Hinojosa, 2014). Dentro de la referida integral  $\widehat{R}_0(\beta, s)$  es el estimador de riesgo de Breslow o de Nelson y Aalen que se define de la siguiente manera:

$$\widehat{R}_0(\beta, s) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s) e^{\beta' Z_i(s)}} \quad (37)$$

La distribución de los residuos de martingala generalmente se presenta muy asimétrica y con una cola larga del lado derecho, especialmente para data de supervivencia que considera un solo evento.

El modelo de cox asume la forma funcional más sencilla de las variables continuas y el log hazard (log de la función de riesgo): lineal. Es así que una manera de comprobar lo expuesto es la gráfica de los residuos de martingala del modelo frente a los valores de la variable continua. Si la asunción de linealidad para la variable continua se cumple, un ajuste mediante una curva suavizada de la nube de puntos de los residuos de martingala se asemejará a una línea recta (Santabárbara, Rubio, & Martínez, 2016).

### **2.2.5. Interpretación de los resultados del modelo de cox**

En primera instancia es importante señalar que los signos de los coeficientes del modelo de cox, tienen una interpretación. Si el signo es positivo significa que la presencia de la variable cuyo coeficiente se analiza incrementa el riesgo de incurrir en evento de interés; si el signo es negativo, la interpretación es la contraria.

La interpretación de los coeficientes ( $\beta$ ) en el modelo de cox no se realiza de forma directa sino a partir del exponencial de los coeficientes  $\exp(\beta)$ , conocido como el ratio de Hazard que equivale al riesgo relativo. El detalle del ratio de Hazard es el siguiente:

$$\frac{\lambda(t, Z_i(t))}{\lambda(t, Z_j^*(t))} = e^{\beta_i} \quad (38)$$

El referido ratio se interpreta como cuantas veces más (o menos) cambia el riesgo frente al incremento en una unidad de la variable correspondiente mientras el resto de variables permanecen constantes (Almeida Baroja, 2011).

#### **2.2.6. Calibración del modelo de cox frente a un cambio en la variable dependiente**

En esta etapa se realiza un ejercicio adicional con respecto al modelo obtenido y se evalúa el modelo de cox desarrollado con la especificación de una variable de interés alternativa, es decir, un nivel de mora grave diferente a la marcación de 90 días, bajo la consigna de identificar su comportamiento y averiguar si una nueva marcación podría mejorar la gestión del riesgo de crédito para el caso de estudio analizado con las mismas variables del modelo obtenido con la marcación inicial.

### 3. RESULTADOS Y DISCUSIÓN

#### 3.1. El Sistema Financiero y su influencia en la Economía

La coexistencia de unidades excedentarias y deficitarias de liquidez dentro de una economía, justifican en principio la existencia de un sistema financiero, que entendido como un conjunto de instituciones, instrumentos y mercados; promueve la canalización del ahorro hacia la inversión (Altina & López, 2001).

La intermediación indirecta por sobre la intermediación directa, según Leland y Pyle (1977) se sustenta en el modelo de información asimétrica que presenta el mercado crediticio, donde para enfrentar el problema de selección adversa, la coalición de prestatarios (intermediario financiero) puede aglutinar y comunicar información de los diferentes implicados, a un menor costo comparado con el que debería incurrir de manera individual cada agente para revelar su perfil (Samartín, 2004).

##### 3.1.1. La función de producción y la productividad marginal del capital

Las publicaciones de Goldsmith (1969), Mckinnon (1973) y Shaw (1973), fueron el punto de partida de un debate matizado entre la relación desarrollo financiero y crecimiento económico. Para una mejor comprensión de la problemática se iniciará con la definición de una función de producción, definida por el stock de capital (De Gregorio & Guidotti, 1996):

$$y_t = f(k_t) \quad (39)$$

Sea ( $y$ ) la producción y ( $k$ ) el stock de capital, para un período ( $t$ ). Diferenciando (39), se obtiene que la tasa de crecimiento de la producción ( $\hat{y}_t$ ), es igual a la tasa de ahorro ( $dk / y_t = s$ ) multiplicada por la productividad marginal del capital ( $f'(k_t) \equiv \phi$ ).

$$\hat{y}_t = dk / y_t * f'(k_t) \equiv s_t \phi_t \quad (40)$$

Los modelos tradicionales de crecimiento económico como los de Solow o Cass-Koopman, consideran que la productividad marginal del capital ( $\phi$ ) es decreciente, con ello a medida que se incrementa ( $k$ ), el crecimiento de la producción alcanza un estado estacionario en el cual la productividad marginal del capital se iguala a cero. Sin embargo, la teoría del crecimiento endógeno, concibe que es posible evitar el estado estacionario si se incorpora al modelo componentes reproducibles, como el capital humano o nuevas tecnologías o como consecuencia de una actividad guiada por incentivos económicos individuales (inversión + desarrollo) (Conesa, 2006).

Relacionada con esta última, el desarrollo financiero incide en el crecimiento económico tanto desde el incremento en la eficiencia de la acumulación del capital (incremento en  $\phi$ ), como desde el incremento de la tasa de ahorro ( $s$ ), que eleva la tasa de inversión (De Gregorio & Guidotti, 1996).

Goldsmith (1969) fue quien inicialmente argumentó que existe una correlación positiva entre el desarrollo financiero y el nivel de PNB per cápita; gracias a que el mayor dinamismo en la intermediación financiera, estimula el uso eficiente del stock de capital. Así también concluyó que la evolución positiva del sistema financiero conlleva un círculo virtuoso que permite la retroalimentación del mismo y de esta manera su mejoramiento continuo (De Gregorio & Guidotti, 1996).

Por su parte, Mackinnon y Shaw (1973), complementan el argumento de Goldsmith, afirmando que adicionalmente al incremento de productividad del capital, el desarrollo financiero promueve una mayor tasa de ahorro y en consecuencia un mayor volumen de inversión. Adicionalmente, afirman que las tasas reales de interés marcadas por el proceso de liberalización financiera, inducen a las economías domésticas a incrementar el ahorro y en su defecto a mayores niveles de inversión y crecimiento. No obstante, esta hipótesis no dejó de ser cuestionada por autores como Díaz – Alejandro (1986), que haciendo referencia a la experiencia latinoamericana en la década de los ochenta, rebatió las bondades de la liberación financiera, citando que *“un proceso de sobreendeudamiento e inestabilidad macroeconómica termina convirtiendo la liberalización en pánico y colapso* (López, 2003).

### **3.1.2. Relación entre el sistema financiero y el crecimiento económico**

Resumiendo lo anteriormente expuesto, en el marco de la teoría del crecimiento endógeno, la relación sistema financiero-crecimiento económico mantiene un esquema bivalente por el cual el sistema financiero levanta y analiza información para posteriormente transferir fondos disponibles a las actividades económicas con mayores tasas de retorno, alterando así la tasa de innovación tecnológica y por ende el crecimiento económico de largo plazo (Barrios Pérez, 2004), mientras que en secuencia opuesta, la existencia de un crecimiento económico permite la masificación del sistema financiero (Greenwood y Jovanovic – 1990 / King y Levine - 1993) (Bench Colombia); y entre otros aspectos, que la intermediación financiera reduce el riesgo asociado a las necesidades de liquidez de los individuos, y en consecuencia evita la liquidación forzosa de inversiones (Bencivenga y Smith – 1991).

Finalmente, el trabajo empírico de De Gregorio y Guidotti publicado en el año 1996 en el N°29 de la revista “Pensamiento Iberoamericano” bajo el título “Desarrollo Financiero y Crecimiento Económico” concluyó que el desarrollo financiero efectivamente conduce a un mayor crecimiento económico; sin embargo, se puntualizó que la incidencia de la primera sobre la segunda varía según los países y el tiempo. Adicionalmente, se vislumbró que conforme a la experiencia latinoamericana de los años setenta y ochenta, pueden darse casos en los cuales la liberación financiera carente de un marco regulador apropiado y las expectativas de rescate gubernamental, revierten la conclusión inicial. La evidencia empírica recabada reveló también que el principal efecto que ejerce el desarrollo financiero sobre el crecimiento económico, se da a través de la mayor eficiencia en la inversión, más que en el volumen de la misma (De Gregorio & Guidotti, 1996).

## **3.2. El Análisis de Supervivencia en el Sistema Financiero**

### **3.2.1. Casos de Aplicación**

#### **Perú**

El análisis denominado “Análisis de supervivencia como alternativa metodológica para estimar probabilidades de incumplimiento de los deudores de créditos corporativos y grandes empresas en el Perú”, desarrollado por Cáceres Fátima y Palacios Ysea de la Facultad de Ingeniería de Industrial de la Universidad Nacional de San Martín, parte de que en la actualidad la mayoría de supervisores bancarios calculan las probabilidades de incumplimiento mediante el uso de matrices de transición, las cuales expresan el porcentaje del saldo de créditos de deudores que permanecen o migran de una categoría de riesgo crediticio a otra en un determinado periodo de tiempo. Las referidas probabilidades asumen tanto las características del deudor como la situación de la economía que se mantiene constante en el tiempo. Esta situación la analizan poco realista y pretenden utilizar metodologías alternativas que permitan el desarrollo de modelos más adecuados que utilicen de forma más eficiente la información disponible, facilitando la inclusión de variables asociadas al entorno económico y variables microeconómicas específicas de los deudores (Cáceres & Palacios, 2017).

Específicamente, propusieron la utilización del análisis de supervivencia para calcular probabilidades de incumplimiento de los deudores de créditos corporativos y a grandes empresas, resaltando que el cálculo de probabilidades de incumplimiento mediante este tipo de modelos representa un avance significativo con relación al uso de matrices de transición, debido a que posibilita obtener probabilidades de incumplimiento por deudor, permitiendo modelar, no solo si el deudor caerá en incumplimiento, sino también el tiempo que transcurrirá hasta que transite a ese estado.

Las conclusiones del análisis desarrollado fueron que las variables sector económico, endeudamiento, rentabilidad, tamaño de la empresa, saldo inicial, porcentaje en moneda extranjera, porcentaje de garantías, agrupación e índice de confianza empresarial fueron identificadas como factores asociados con el incumplimiento en el pago de créditos corporativos y a grandes empresas. Consecuentemente, sería importante monitorear las referidas variables al evaluar los créditos, con el fin de identificar potenciales incumplimientos para este tipo de deudores.

### **Holanda**

El trabajo denominado “Survival analysis in credit scoring: A framework for PD estimation”, desarrollado Ramon Man en el año 2014, pretendió desarrollar un modelo de supervivencia marco para la estimación de la probabilidad de *default* o de que un cliente no pague. Para ello utilizó, información de operaciones crediticias generadas en Rabobank Internacional, cuyas operaciones se desarrollan en Holanda, las observaciones utilizadas, organizadas en diferentes bases, superaron las 30.000 operaciones de crédito (Man, 2014).

Algunas consideraciones relevantes y conclusiones del análisis realizado fueron:

- La principal ventaja del análisis de supervivencia es la capacidad de incorporar datos censurados. Para estas observaciones no se observa incumplimiento durante el período de estudio.
- Enfoque tradicionales como el de regresión logística, no consideran observaciones censuradas y se eliminan del conjunto de datos. En virtud de lo anterior el análisis de supervivencia es capaz de incorporar estas observaciones, y puede proporcionar mejores resultados.

Adicionalmente, luego de desarrollar un modelo de riesgo proporcional de cox y de comparar su rendimiento con las técnicas de regresión logística concluye:



- La transformación de logrank (modelo de supervivencia de cox) supera la transformación logística y el enfoque estadístico óptimo, porque es más significativo en la predicción de la probabilidad de supervivencia basada en la estadística de prueba de Wald. Por otra parte, los resultados de la ROC / AUC, estadística de potencia y KS estadística mostró que hay poca diferencia en el rendimiento de los modelos de supervivencia y la regresión logística.
- Este modelo (refiriendo al modelo de cox y transformación logrank) requiere de una limpieza de datos significativamente menor debido a que el modelo estima la probabilidad de supervivencia en todo el conjunto de datos, en contraste con la regresión logística que sólo estima la probabilidad de supervivencia para un intervalo de tiempo fijo.

## **España**

El análisis denominado “Modelando el riesgo de crédito de los consumidores a través del análisis de supervivencia”, desarrollado por Cao Ricardo, Vilar Juan y Devia Andrés, pretendió identificar la probabilidad de caer en situación de no pago (default) utilizando el análisis de supervivencia. Para ello, utilizó tres métodos: modelo de regresión de Cox’s, modelos lineales generalizados bajo censura y estimación no paramétrica de Kernel.

La data utilizada para el ejercicio fue una base de 25.000 créditos de consumo pertenecientes a un banco español registrada entre julio 2004 y noviembre 2006. Adicionalmente, tomando en consideración algunas recomendaciones de Basilea para este tipo de créditos, se consideró únicamente aquellos créditos que presentaron una madurez de al menos un año. De manera general, la base estuvo compuesta por dos tipos de clientes, aquellos que no cayeron en *default* (92,8%) y aquellos que si cayeron en *default* (7,2%).

La variable dependiente fue el tiempo de maduración ( $y$ ) acompañada de la variable presencia de *default* (1 *default*, 0 censura), explicado por la variable continua score de crédito aplicado en el momento de la precalificación.

De manera general, la comparación realizada entre la aplicación de los diferentes métodos citados fue con respecto a su ajuste con los datos. En esta línea parecería que el modelo de Cox se ajusta de mejor manera a los datos, y adicionalmente en este modelo la variable score ( $x$ ) se mostró relevante. Finalmente, luego de una discusión generada por los autores, en la réplica correspondiente concluyeron que el modelo de cox y la estimación no paramétrica de Kernel permitieron discriminar de mejor manera entre clientes buenos y malos, en comparación a lo que se pudo realizar con el modelo lineal generalizado bajo censura.

No obstante, entre los elementos que se destacaron dentro del análisis se puede mencionar la distorsión que genera en este tipo de análisis la presencia de una importante proporción de datos censurados (Cao, Vilar, & Devia, 2009).

### **Reino Unido**

El análisis denominado “Métodos de análisis de supervivencia para datos de préstamos personales”, desarrollado por Stepanova Maria y Thomas Lyn, buscó aplicar técnicas de análisis de supervivencia a datos personales de una importante institución financiera del Reino Unido que disponía de 50.000 préstamos y un período de amortización de las mismas de 36 meses.

Para validar el modelo de cox ajustado utilizaron como herramientas de diagnósticos residuos de Cox-Snell, residuos de Martingala y residuos de desvío. Luego de la realización de las pruebas citadas los autores concluyeron que era necesario descartar aquellos préstamos que fueron pagados luego de un mes de concedidos ya que se constituyeron en casos atípicos. Con la depuración de casos realizados concluyeron que el modelo se ajustaba bien a los datos (Obuda, 2014).

## **Estados Unidos de Norte América**

El análisis notado como “Métodos estadísticos en el modelado del riesgo de crédito”, desarrollado por Zhang Aijun, entre otros elementos presentó un ejercicio de aplicación del análisis de supervivencia en tiempo dual en la cartera de hipotecas en la banca minorista. En este caso específico se aplicó el modelo de Cox, intentando identificar los determinantes de riesgo del incumplimiento y prepago de la hipoteca.

La información utilizada fue la correspondiente a hipotecas originadas entre el año 2001 y 2007, truncada a principios de 2005, censurada a la derecha en octubre de 2008 y censurada en 60 meses de permanencia.

El evento de interés propuesto para un préstamo hipotecario fue el causado por incumplimiento (default) y el de prepago de la operación de crédito (riesgos competitivos dependientes), el tiempo del evento observado corresponde al tiempo hasta el primer evento.

Las variables explicativas fueron: CLTV (índice de apalancamiento préstamo vs valor del bien), FICO (puntaje de crédito o score), NoteRate (tasa de la nota de hipoteca, se ajusta en el tiempo de acuerdo al precio en el mercado), Documentación completa, Propósito del préstamo (compra o refinanciamiento), Financiamiento exclusivo de intereses.

Los resultados obtenidos de la estimación fueron: 1) el riesgo de incumplimiento está dado por un alto CLVT, bajo FICO, un alto NoteRate, documentación incompleta, refinanciamiento y financiamiento exclusivo de intereses, y 2) el riesgo de pre-cancelación está dado por un bajo CLVT, alto FICO, alto NoteRate, documentación incompleta, compra y financiamiento exclusivo de intereses (Zhang, 2009).

## Australia

El análisis denominado “Análisis de supervivencia para credit scoring: incidencia y latencia”, desarrollado por Watkins John, Vasnev Andrey y Gerlach Richard, aplicó una la metodología que hace uso del modelo de tasa de curación a un conjunto de datos de préstamos personales originados entre el 1 de marzo de 2001 y el 31 de marzo de 2008 por parte de una entidad financiera de Australia.

Específicamente pretendió simular simultáneamente los eventos de prepago, cancelación y vencimiento de los préstamos. La estimación de máxima verosimilitud se la realizó mediante el método Nelder-Mead.

Los resultados de la estimación bajo la metodología propuesta proporcionaron la primera evidencia de coeficientes que actuaron en direcciones opuestas sobre la incidencia y la latencia de un evento. Puntualmente, mientras la variable LVR (ratio: monto de préstamo / activos totales) se incrementó, la incidencia en el evento cancelación se incrementó, la incidencia del prepago disminuyó y las latencias condicionales de cancelación y prepago progresaron de forma más lenta y de forma más rápida respectivamente. Así también, para la variable LT (pasivo total) se estimaron coeficientes positivo y negativo para la incidencia del prepago y los efectos de latencia respectivamente. Este comportamiento indicó que cuanto más alto es LT fue más alta la probabilidad de que se suscite el evento prepago, pero esto ocurriría más lentamente.

Así también, el coeficiente de la variable EmpYrs (número de meses que un solicitante ha estado trabajando en sus dos últimos trabajos) no fue significativo para explicar la incidencia en el evento de cancelación, mientras que fue significativo para explicar la latencia condicional de la cancelación. Sin embargo, presentó un significado económico marginal dado la baja magnitud del coeficiente positivo.

Según los autores, estos resultados hubieran sido inalcanzables utilizando otras metodologías. Este aspecto del modelo permitió un examen más profundo y más riguroso de los datos de crédito.

Finalmente, una observación que fue destacada dentro del análisis fue que la censura da lugar a más errores en las estimaciones de los parámetros para todas las metodologías (Watkins, Vasnev, & Gerlach, 2009).

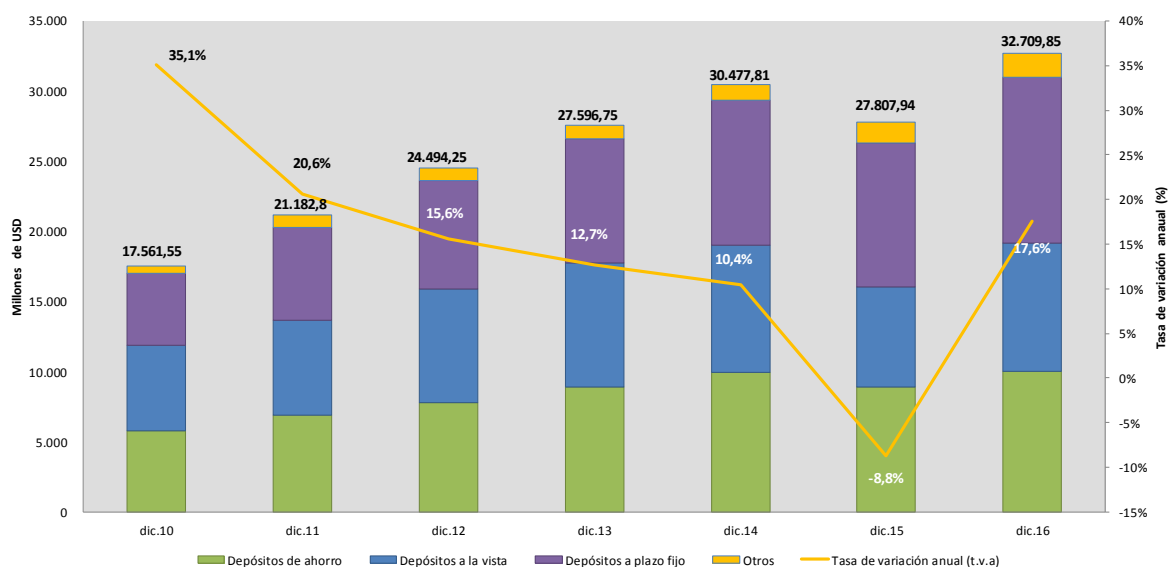
### 3.3. El Sistema Financiero Ecuatoriano y la Dinámica del Crédito Hipotecario

#### 3.3.1. Sistema Financiero Ecuatoriano

##### Captación de recursos

##### Evolución

El saldo de las captaciones realizadas por las denominadas Otras Sociedades de Depósito (OSD)<sup>1</sup> al mes de diciembre 2016 ascendió a 32.709,9 millones de USD, un 17,6% por sobre el saldo vigente al mes de diciembre 2015 (27.807,9 millones de USD). Durante los meses que conformaron el año 2016, el saldo de captaciones se incrementó en promedio en un 1,4% cada mes de la mano del incremento de los depósitos a la vista.

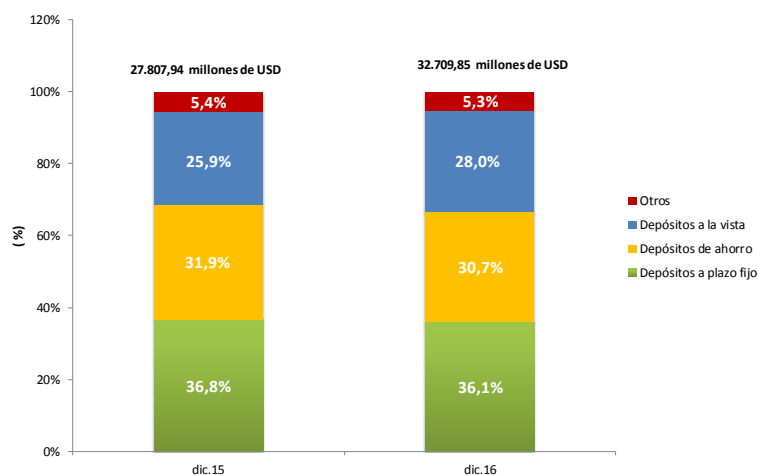


**Figura 2 - Saldo de captaciones - OSD**  
(Banco Central del Ecuador (c), 2017)

<sup>1</sup> N.B. Las OSD comprenden los bancos privados, sociedades financieras, mutualistas, cooperativas, tarjetas de crédito y BNF. Se excluye la información de la Banca Cerrada.

## Estructura de captaciones

El saldo de captaciones realizadas por las OSD del Ecuador al término del 2016 estuvo compuesto mayoritariamente por los depósitos a plazo fijo con un 36,1%, seguido de los depósitos de ahorro con un 30,7% y de los depósitos a la vista con un 28,0%. La estructura referida no varió significativamente en comparación a la vigente a diciembre 2015; sin embargo, cabe señalar que de la mano del incremento experimentado en los depósitos a la vista su participación frente al total de captaciones se incrementó en 2 puntos porcentuales (diciembre 2015 = 25,9% /diciembre 2016 = 28,0%) en detrimento de la participación de los depósitos de ahorro, la cual se contrajo en 1,2 puntos porcentuales (diciembre 2015 = 31,9% /diciembre 2016 = 30,7%).



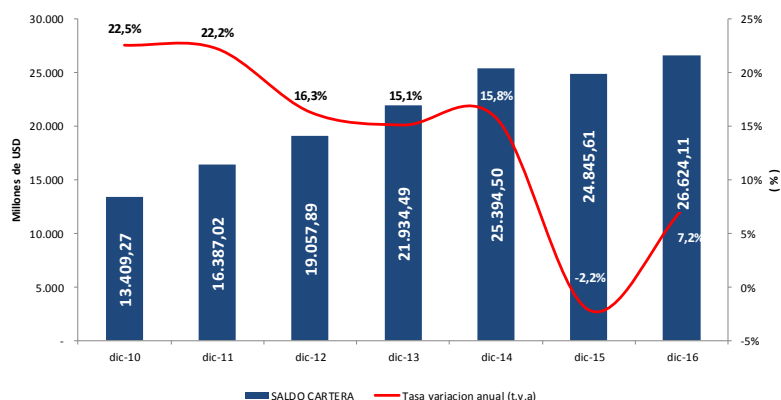
**Figura 3** – Estructura anual de saldo de captaciones (%)<sup>2</sup>  
(Banco Central del Ecuador (c), 2017)

<sup>2</sup> N.B. Dentro de la estructura de captaciones, la categoría Otros se compone de depósitos restringidos, operaciones de reporto y otros depósitos (no incluye depósitos por confirmar e incluye los fondos de tarjetahabientes).

## Concesión crediticia

### Cartera de créditos

El saldo de la cartera de las entidades del Sistema financiero formal ecuatoriano<sup>3</sup> en el período 2010-2014 presentó un crecimiento sostenido; sin embargo, a diciembre 2015 el saldo de cartera se situó en 24.845,6 millones de USD, un -2,2% por debajo del mostrado en diciembre 2014, y a diciembre 2016 en 26.624,11 millones de USD, un 7,2% por encima del presentado en diciembre 2015 recuperando de esta manera el descenso experimentado en diciembre 2015.



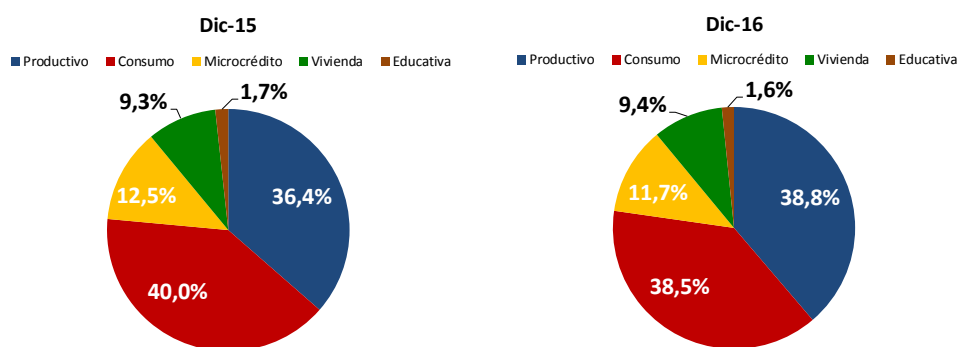
**Figura 4 – Saldo de cartera anual (millones USD)**  
(Banco Central del Ecuador (c), 2017)

La estructura de la cartera vigente a diciembre 2016 se concentró primordialmente en los segmentos productivo (ex - comercial) (10.318,1 millones de USD) y de consumo (10.252,9 millones de USD) con participaciones del 38,8% y 38,5% respectivamente, manteniendo de esta manera la estructura según segmentos de crédito vislumbrada históricamente.

Por su parte, la cartera del segmento de vivienda presentó a diciembre 2016 una participación del 9,4%, nivel ligeramente superior al 9,3% mostrado en diciembre 2015.

<sup>3</sup> N.B.: Para efectos del presente análisis se considera sistema financiero formal, a las entidades reguladas por la Superintendencia de Bancos del Ecuador, Banco del Instituto Ecuatoriano de Seguridad Social (BIESS) y reguladas por la Superintendencia de Economía Popular y Solidaria.

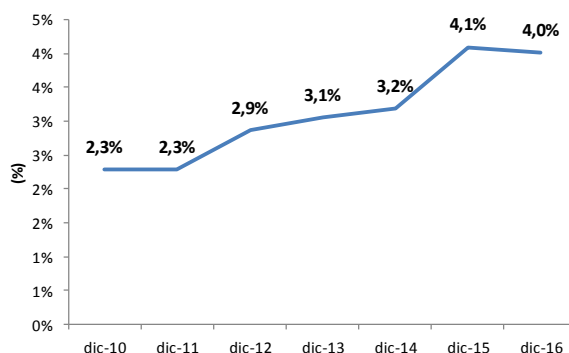




**Figura 5 – Saldo cartera de créditos 2015 y 2016 – Estructura por sectores (%)<sup>4</sup>**  
(Banco Central del Ecuador (c), 2017)

### Morosidad

La cartera de las entidades del sistema financiero formal ecuatoriano ha incrementado gradualmente su nivel de morosidad durante los últimos cinco años, pasando así del 2,3% en diciembre 2010 al 4,0% en diciembre 2016.

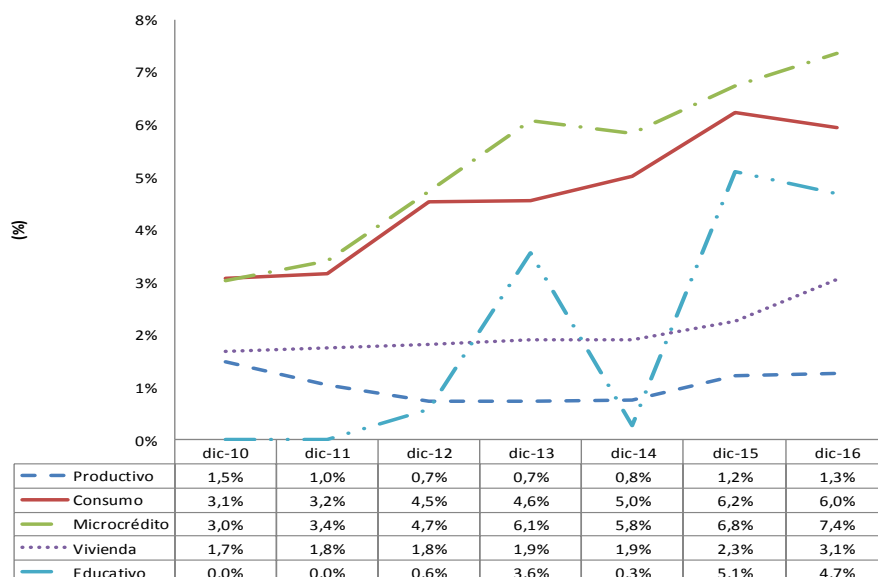


**Figura 6 – Morosidad de la cartera de crédito – Sistema Financiero Privado (%)**  
(Banco Central del Ecuador (c), 2017)

Si se analiza el nivel de morosidad según segmento de crédito se puede concluir que los mayores incrementos de morosidad durante el año 2016 se dieron en los segmentos de vivienda y microempresa que específicamente pasaron del 2,3% en diciembre 2015 al 3,1% en diciembre 2016 y del 6,8% en diciembre 2015 al 7,4% en diciembre 2016 respectivamente. La morosidad en el segmento productivo (ex - comercial) luego de haberse mostrado moderada hasta el año

<sup>4</sup> N.B. Dentro del segmento de consumo se agruparon los sub-segmentos de consumo prioritario y ordinario, dentro del segmento de vivienda se incluyó el sub-segmento vivienda de interés público y dentro del segmento comercial se agruparon los sub-segmentos comerciales ordinarios y prioritarios.

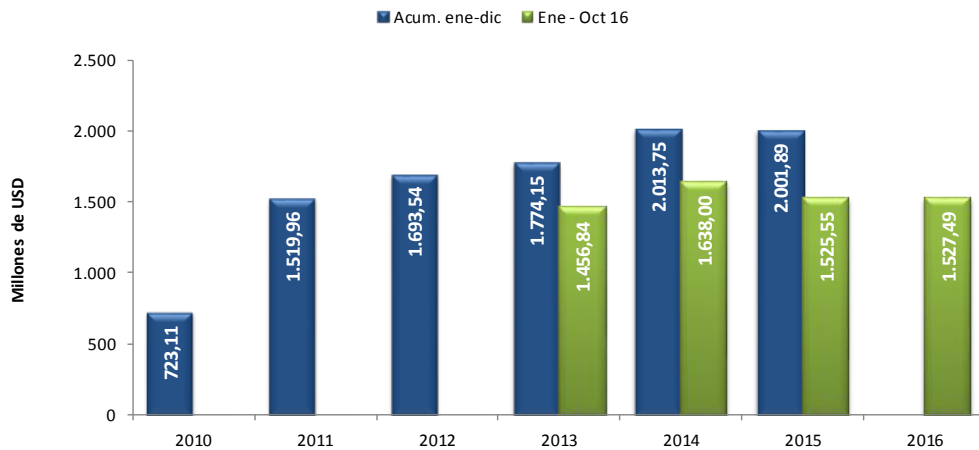
2014 se incrementó al 1,2% a diciembre 2015 y al 1,3% a diciembre 2016. Finalmente, en lo que respecta a la morosidad en el segmento de consumo, se puede señalar que luego de haber presentado un ratio del 6,2% en diciembre 2015, este se contrajo en 0,2 puntos porcentuales y se situó en el 6,0% a diciembre 2016.



**Figura 7 – Morosidad por segmento (%)**  
(Banco Central del Ecuador (c), 2017)

### **Volumen de Financiamiento – Crédito de vivienda - Sistema Financiero Ecuatoriano Regulado por la Superintendencia de Bancos (SB)**

El volumen de colocación de las entidades reguladas por la SB en el segmento de vivienda período enero – octubre 2016 se situó en 1.527,5 millones de USD, nivel superior al generado en el mismo periodo 2015 (1.525,6 millones de USD) en apenas un 0,1%.



**Figura 8** – Volumen de colocación crediticia – crédito vivienda (millones USD) (Superintendencia de Bancos (a))

En cuanto al aporte de las diferentes instituciones financieras, se puede afirmar que durante el periodo enero - octubre 2016, dos (2) instituciones financieras generaron el 79,8% del volumen de colocación total en el segmento de vivienda generado por las entidades reguladas por la SB.

Específicamente, las referidas fueron el Banco del Instituto Ecuatoriano de Seguridad Social (BIESS) con el 62,6% y el Banco del Pichincha con el 17,3%.

En lo que atañe a comportamientos particulares relevantes durante el periodo analizado 2016, se puede manifestar que a diferencia de lo vislumbrado con las restantes instituciones reguladas por la Superintendencia de Bancos del Ecuador, las cuales contrajeron sus volúmenes de concesión, el Banco del Pichincha y la Mutualista Pichincha incrementaron su volumen de colocación en el segmento de vivienda en un significativo 32,9% y 27,1 % respectivamente, en relación al generado en el periodo enero - octubre 2015.

**Tabla 2** – Ranking instituciones financieras – Volumen de crédito causado por entidades SB (millones de USD) - Segmento vivienda (Superintendencia de Bancos (a))

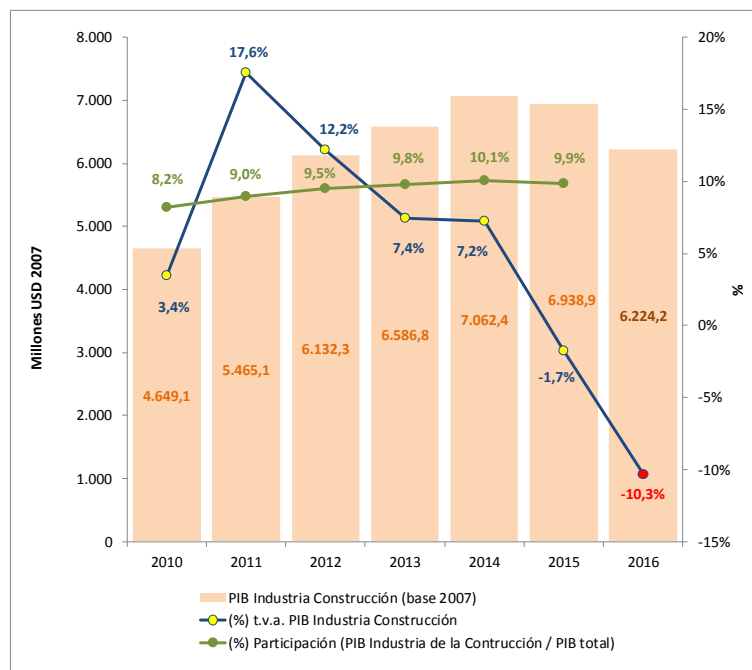
SEGMENTO DE VIVIENDA						
acum. ene - oct. 2015			acum. ene - oct. 2016			VARIACIÓN
SB	MONTO (Millones)	PARTICIPACIÓN (%)	SB	MONTO (Millones)	PARTICIPACIÓN (%)	
BIESS	1.047,79	64,3%	BIESS	904,92	62,6%	-13,6%
BP PICHINCHA	188,03	11,5%	BP PICHINCHA	249,88	17,3%	32,9%
MUT PICHINCHA	47,86	2,9%	MUT PICHINCHA	60,85	4,2%	27,1%
BP GUAYAQUIL	70,32	4,3%	BP GUAYAQUIL	48,72	3,4%	-30,7%
OTRAS ENTIDADES	276,26	16,9%	OTRAS ENTIDADES	182,35	12,6%	-34,0%
<b>TOTAL</b>	<b>1.630,25</b>	<b>100%</b>	<b>TOTAL</b>	<b>1.446,71</b>	<b>100%</b>	<b>-11,3%</b>

Finalmente, cabe mencionar que durante el periodo enero – octubre 2016 las instituciones financieras reguladas por la SB inmiscuidas en el segmento de vivienda, disminuyeron su número de operaciones de crédito en un – 12,5%.

### 3.3.2. Desempeño general del sector de la construcción

La industria de la construcción se mostró entre los años 2010 y 2014 como una industria dinámica que generó crecimientos de su PIB (base 2007) superiores al 7,0%, de la mano de un estado que a través de su política pública invirtió en proyectos de infraestructura, vivienda pública, desarrollo vial e impulsó la canalización de una mayor cantidad de recursos para el financiamiento de créditos hipotecarios, principalmente, a través del Banco del Instituto Ecuatoriano de Seguridad Social (BIESS); sin embargo, tras el anuncio del ingreso a la Asamblea Nacional de un proyecto de ley que modificaría tanto el impuesto a la plusvalía para los bienes inmuebles, como el impuesto a las herencias en el segundo semestre 2014 y la desaceleración económica que experimentó el país a partir del segundo semestre 2015; se generó una tasa de variación anual del PIB de la Construcción año 2015 (base 2007) negativa del -1,7% y tentativamente en una del -10,3% para el año 2016 según las previsiones realizadas por el Banco Central del Ecuador (Banco Central del Ecuador (d), 2016).

La incidencia de la industria de la construcción en la economía ecuatoriana ha sido importante durante los últimos 6 años, es así que el aporte del PIB de la Industria de la Construcción al PIB total del Ecuador entre los años 2010 y 2015 se mostró cercano al 9,4%, y específicamente se situó en el 9,9% en el año 2015 (Banco Central del Ecuador (b), 2016).



**Figura 9** – PIB Industria de la Construcción (base 2007), Tasa de variación anual – PIB Industria de la Construcción y (%) Participación - PIB Industria de la Construcción / PIB Total<sup>5</sup>  
(Banco Central del Ecuador (b), 2016)

Por otra parte, la desaceleración económica del país y de la industria de la construcción empezaron a incidir también en el sector real. De esta manera, el número de empleados en la rama de actividad “construcción” según nomenclatura del Instituto Nacional de Estadísticas y Censos (INEC) en el mes de septiembre 2016 se ubicó en aproximadamente 516.072, un -7,03% por debajo del número existente en septiembre 2015 (555.088). Es decir, entre septiembre 2015 y septiembre 2016 el número de empleados en la rama de actividad “construcción” se redujo en 39.017 empleados (Instituto Nacional de Estadísticas y Censos (a), 2016).

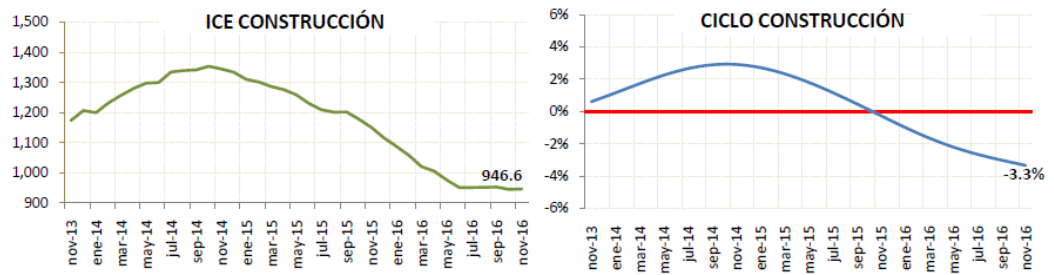
<sup>5</sup> Los valores de la gráfica correspondientes al año 2016, se derivan de las previsiones realizadas por el Banco Central del Ecuador en el mes de agosto 2016.

**Tabla 3 – Número de empleados por rama de actividad económica – septiembre 2015 vs septiembre 2016**  
(Instituto Nacional de Estadísticas y Censos (a), 2016)

Rama de Actividad	Nº de empleados sep 2015	Nº de empleados sep 2016	Diferencia Nº de empleados septiembre 2015 - septiembre 2016	Tasa de variación Nº de empleados septiembre 2015 - septiembre 2016 (%)
Agricultura, ganadería, caza y silvicultura y pesca	1.829.022	2.049.245	220.223	12,04%
Petróleo y minas	43.725	33.148	-10.576	-24,19%
Manufactura (Incluida refinación de petróleo)	801.201	826.232	25.031	3,12%
Suministro de electricidad y agua	40.657	46.379	5.722	14,07%
Construcción	555.088	516.072	-39.017	-7,03%
Comercio	1.369.317	1.424.746	55.429	4,05%
Alojamiento y servicios de comida	415.733	486.716	70.983	17,07%
Transporte	426.651	437.772	11.121	2,61%
Correo y Comunicaciones	78.691	78.687	-4	-0,01%
Actividades de servicios financieros	74.124	40.269	-33.855	-45,67%
Actividades profesionales, técnicas y administrativas	307.478	368.796	61.318	19,94%
Enseñanza y Servicios sociales y de salud	526.223	551.763	25.540	4,85%
Administración pública, defensa; planes de seguridad social obligatoria	339.597	285.321	-54.275	-15,98%
Servicio doméstico	179.734	196.348	16.614	9,24%
Otros Servicios	286.979	296.491	9.511	3,31%
<b>Total empleados:</b>	<b>7.274.221</b>	<b>7.637.986</b>	<b>363.765</b>	<b>5,00%</b>

Por otra parte, en lo que respecta a la percepción del empresariado vinculado al sector de la construcción se puede mencionar, según el Estudio Mensual de Opinión Empresarial generado por el Banco Central del Ecuador (BCE), que el índice de confianza de los empresarios (ICE) del sector de la construcción, que pretende medir las expectativas de los empresarios del sector con respecto a su actividad, ha descendido sostenidamente a partir de los primeros meses del año 2015, pasando así de un índice de 1.343,6 puntos en octubre 2014, a 1.310,1 puntos en enero 2015, a 1.116,2 puntos en diciembre 2015 y a 946,6 puntos en noviembre 2016 (Banco Central del Ecuador (e), 2016).

Obviamente, la situación expuesta implicó que a partir del mes de diciembre 2015 el ciclo del índice ICE se ubicó por debajo de la tendencia de crecimiento de largo plazo. Los factores que incidieron en el descenso referido fueron las expectativas de menores volúmenes de construcción y de menores niveles de ocupación de personal.



**Figura 10 - Índice de confianza empresarial del sector de la construcción y ciclo del índice**  
(Banco Central del Ecuador (e), 2016)

Es importante mencionar, que a partir del segundo semestre 2016, el índice se mantuvo relativamente estable sin presentar nuevamente descensos sustanciales.

### **3.3. Aplicación del Análisis de Supervivencia: El Caso de una Cartera Hipotecaria Ecuatoriana**

#### **3.3.1. Desarrollo del Modelo de Supervivencia**

##### **Definición de la variable de interés**

La variable de interés es el tiempo en meses que tarda un crédito hipotecario en caer en una mora grave, denominada “tiempoM90pv\_2mes” (tiempo de supervivencia). Se define como mora grave al hecho de permanecer en mora por más de 90 días consecutivos luego de la fecha en la cual debía hacerse el pago correspondiente. Explícitamente, la referida variable se condensa en una de tipo binaria denominada “Mayor90”, misma que toma el valor “1” si la operación de crédito hipotecario ha experimentado una mora grave y “0” si no lo ha experimentado durante el período de estudio (caso censurado por la derecha).

El vector conformado por la dos variables mencionadas en cada uno de los casos de los clientes del crédito hipotecario, se constituye en el tipo de variable dependiente requerido para la formulación de un análisis de supervivencia.

##### **Datos**

Los datos utilizados para la estimación fueron los créditos hipotecarios activos al 31 de diciembre de 2016 que denotaron una antigüedad de al menos 2 años y fueron destinados a la adquisición de vivienda terminada. Es decir que fueron concedidos entre los años 2008 y 2014. En virtud de lo anterior, el número de observaciones consideradas fue de 79.225 observaciones.

Complementariamente, el tiempo de estudio definido para el presente análisis fue el comprendido entre el 30 de julio del 2008 y el 31 de diciembre de 2016, es decir un período de 102,53 meses u 8,54 años.



## Caracterización de variables – modelo Kaplan - Meier

### Variable: rango de avalúo

La variable rango avalúo (R\_Avaluo2) es una variable nominal que condensa la variable avalúo de realización de una vivienda en tres (3) opciones de respuesta: “hasta 40.000 USD”, “entre 40.001 y 100.000 USD” y “mayor a 100.000 USD”.

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se distribuyeron según la variable rango de avalúo (R\_Avaluo2) en un 56,3% en el rango “entre 40.001 y 100.000 USD”, en un 33,7% en el rango “hasta 40.000 USD” y en un 9,9% en el rango “mayor a 100.000 USD”.

**Tabla 4** – Tabla de contingencia - variable rango de avalúo (R\_Avalúo2) vs variable presencia mora grave (Mayor90)

```
##
##      Cell Contents
## |-----|
## |              |
## |              |
## |              |
## |              |
## |              |
## |-----|
## Total Observations in Table:  79225
##
##      phcalf16_m2a$R_Avaluo2 | phcalf16_m2a$Mayor90
## |-----|-----|-----|
## |              | 0 | 1 | Row Total |
## |-----|-----|-----|
## | hasta 40.000 USD | 23601 | 3124 | 26725 |
## |              | 88.3% | 11.7% | 33.7% |
## |              | 33.0% | 40.3% |
## |-----|-----|-----|
## | entre 40.000,01 y 100.000 USD | 40685 | 3939 | 44624 |
## |              | 91.2% | 8.8% | 56.3% |
## |              | 56.9% | 50.8% |
## |-----|-----|-----|
## | mayor a 100.000 USD | 7185 | 691 | 7876 |
## |              | 91.2% | 8.8% | 9.9% |
## |              | 10.1% | 8.9% |
## |-----|-----|-----|
## |              | Column Total | 71471 | 7754 | 79225 |
## |              | 90.2% | 9.8% |
## |-----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----
## Chi^2 = 165.2799    d.f. = 2    p = 1.287997e-36
##
##
## Minimum expected frequency: 770.8489
```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable rango de avalúo (R\_Avaluo2) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en el rango de “hasta 40.000 USD” (11,7%), incluso superando levemente la estructura natural de la opción presencia de mora grave (1) de la variable mora grave (Mayor90) (9.8%).

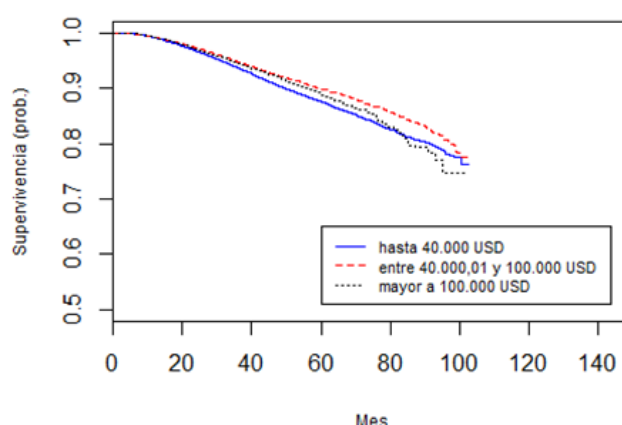
Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables rango de avalúo (R\_Avaluo2) y mora grave (Mayor90) se puede señalar que dado un p-valor asociado inferior a 0,05 ( $1,287997e-36$ ) se puede rechazar la hipótesis de independencia de variables ( $H_0$ ) y concluir que existe una relación de dependencia entre las dos variables analizadas.

Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM), para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable rango avalúo (R\_Avaluo2), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 20, y en mayor medida en aquellos clientes con un avalúo de hasta 40.000 USD.

Así también, cabe indicar que dada la alta proporción de casos censurados en cada uno de los rangos de avalúo no fue posible calcular la mediana de supervivencia en cada uno de los casos.

**Tabla 5** – Función de supervivencia, método KM, variable rango de avalúo (R\_Avalúo2)

```
## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$R_Avaluo2)
##
##                               n events median
## phcalf16_m2a$R_Avaluo2=hasta 40.000 USD      26725   3124   NA
## phcalf16_m2a$R_Avaluo2=entre 40.000,01 y 100.000 USD 44624   3939   NA
## phcalf16_m2a$R_Avaluo2=mayor a 100.000 USD      7876    691   NA
##                               0.95LCL 0.95UCL
## phcalf16_m2a$R_Avaluo2=hasta 40.000 USD           NA     NA
## phcalf16_m2a$R_Avaluo2=entre 40.000,01 y 100.000 USD NA     NA
## phcalf16_m2a$R_Avaluo2=mayor a 100.000 USD       NA     NA
```



**Figura 11** – Función supervivencia, método KM, var.: rango de avalúo (R\_Avalúo2)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor igual a 0, se rechaza la hipótesis nula ( $H_0$ ) de igualdad de funciones, y con ello se puede señalar que las funciones de supervivencia entre quienes integran los diferentes rangos de avalúo fueron diferentes.

**Tabla 6** – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de avalúo (R\_Avalúo2)

```
## Call:
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$R_Avaluo2)
##
##
##                               N Observed
## phcalf16_m2a$R_Avaluo2=hasta 40.000 USD      26725      3124
## phcalf16_m2a$R_Avaluo2=entre 40.000,01 y 100.000 USD 44624      3939
## phcalf16_m2a$R_Avaluo2=mayor a 100.000 USD      7876      691
##
##                               Expected (O-E)^2/E
## phcalf16_m2a$R_Avaluo2=hasta 40.000 USD      2774      44.282
## phcalf16_m2a$R_Avaluo2=entre 40.000,01 y 100.000 USD 4304      30.893
## phcalf16_m2a$R_Avaluo2=mayor a 100.000 USD      677      0.297
##
##                               (O-E)^2/V
## phcalf16_m2a$R_Avaluo2=hasta 40.000 USD      69.191
## phcalf16_m2a$R_Avaluo2=entre 40.000,01 y 100.000 USD 69.498
## phcalf16_m2a$R_Avaluo2=mayor a 100.000 USD      0.326
##
## Chisq= 75.6 on 2 degrees of freedom, p= 0
```

## Variable: género

La variable género (GENERO) es una variable nominal que contempla las opciones de respuesta: "Femenino" y "Masculino".

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se distribuyeron según la variable género (R\_Avaluo2) en un 59,4% en miembros del género masculino y en un 40,6% en miembros del género femenino.

**Tabla 7** – Tabla de contingencia - variable género (GENERO) vs variable presencia mora grave (Mayor90)

```
##
## Cell Contents
## |-----|
## |                Count |
## |                Row Percent |
## |                Column Percent |
## |-----|
##
## Total Observations in Table:  79225
##
##
##      phcalf16_m2a$Mayor90
## phcalf16_m2a$GENERO |      0      1 | Row Total |
## -----|-----|-----|
##      FEMENINO      | 29754   2423 | 32177     |
##                   | 92.5%   7.5% | 40.6%     |
##                   | 41.6%   31.2% |           |
## -----|-----|-----|
##      MASCULINO     | 41717   5331 | 47048     |
##                   | 88.7%   11.3% | 59.4%     |
##                   | 58.4%   68.8% |           |
## -----|-----|-----|
##      Column Total  | 71471   7754 | 79225     |
##                   | 90.2%   9.8% |           |
## -----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----|-----|-----|
## Chi^2 = 312.6322    d.f. = 1    p = 5.832179e-70
##
## Pearson's Chi-squared test with Yates' continuity correction
## -----|-----|-----|
## Chi^2 = 312.2019    d.f. = 1    p = 7.237197e-70
##
##
## Minimum expected frequency: 3149.264
```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable género (GENERO) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en el género masculino (11,3%), incluso superando levemente la estructura natural de la opción presencia de mora grave (1) de la variable mora grave (Mayor90) (9.8%).

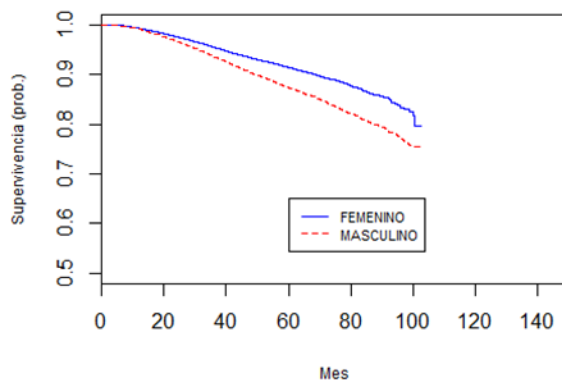
Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables género (GENERO) y mora grave (Mayor90) se puede señalar que dado un p-valor asociado inferior a 0,05 (5,832179e-70) se puede rechazar la hipótesis de independencia de variables (h0) y concluir que existe una relación de dependencia entre las dos variables analizadas.

Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM), para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable género (GENERO), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 20, y en mayor medida en aquellos clientes del género masculino.

Así también, cabe indicar que dada la alta proporción de casos censurados en cada uno de las opciones de género no fue posible calcular la mediana de supervivencia en cada uno de los casos.

**Tabla 8** – Función de supervivencia, método KM, variable género (GENERO)

```
## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$GENERO)
##
##               n events median 0.95LCL 0.95UCL
## phcalf16_m2a$GENERO=FEMENINO 32177 2423    NA     NA     NA
## phcalf16_m2a$GENERO=MASCULINO 47048 5331    NA     NA     NA
```



**Figura 12** – Función supervivencia, método KM, var.: género (GENERO)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor igual a 0, se rechaza la hipótesis nula ( $H_0$ ) de igualdad de funciones, con y con ello se puede manifestar que las funciones de supervivencia entre los miembros de los géneros masculino y femenino fueron diferentes.

**Tabla 9** – Resultado del contraste de igualdad de funciones de supervivencia para la variable género (GENERO)

```
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$GENERO)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## phcalf16_m2a$GENERO=FEMENINO 32177    2423    3112    152    255
## phcalf16_m2a$GENERO=MASCULINO 47048    5331    4642    102    255
##
## Chisq= 255 on 1 degrees of freedom, p= 0
```

### Variable: rango de plazo

La variable rango de plazo ( $R\_Plazo$ ) es una variable nominal que condensa la variable plazo de concesión de un crédito hipotecario en dos (2) opciones de respuesta: “menor a 20 años” y “mayor a 20 años”.

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se

distribuyeron según la variable rango de plazo (R\_Plazo) en un 50,8% en el rango “menor a 20 años” y en un 49,2% en el rango “mayor a 20 años”.

**Tabla 10** – Tabla de contingencia - variable rango de plazo (R\_Plazo) vs variable presencia mora grave (Mayor90)

```
##
## Cell Contents
## |-----|
## |                Count                |
## |                Row Percent          |
## |                Column Percent      |
## |-----|
##
## Total Observations in Table:  79225
##
##      phcalf16_m2a$Mayor90
## phcalf16_m2a$R_Plazo |      0      1 | Row Total |
## -----|-----|-----|
##      menor a 20 años | 37035   3177 | 40212    |
##                    | 92.1%    7.9% | 50.8%    |
##                    | 51.8%   41.0% |          |
## -----|-----|-----|
##      mayor a 20 años | 34436   4577 | 39013    |
##                    | 88.3%   11.7% | 49.2%    |
##                    | 48.2%   59.0% |          |
## -----|-----|-----|
##      Column Total   | 71471   7754 | 79225    |
##                    | 90.2%    9.8% |          |
## -----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----|-----|-----|
## Chi^2 = 329.2134    d.f. = 1    p = 1.425941e-73
##
## Pearson's Chi-squared test with Yates' continuity correction
## -----|-----|-----|
## Chi^2 = 328.7797    d.f. = 1    p = 1.772484e-73
##
##
##      Minimum expected frequency: 3818.325
```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable rango de plazo (R\_Plazo) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en el rango de “mayor a 20 años” (11,7%), incluso superando levemente la estructura natural de la opción presencia de mora grave (1) de la variable mora grave (Mayor90) (9.8%).

Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables rango de plazo (R\_Plazo) y mora grave (Mayor90) se puede señalar que dado un p-valor asociado inferior a 0,05

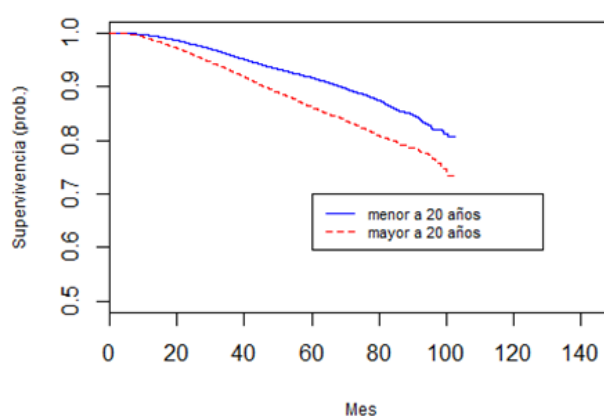
(1,425941e-73) se puede rechazar la hipótesis de independencia de variables ( $H_0$ ) y concluir que existe una relación de dependencia entre las dos variables analizadas.

Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM), para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable rango de plazo (R\_Plazo), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 10, y con mayor intensidad en aquellos clientes que optaron por un plazo para su crédito mayor a 20 años.

Así también, cabe indicar que dada la alta proporción de casos censurados en cada uno de los rangos de plazo no fue posible calcular la mediana de supervivencia en cada uno de los casos.

**Tabla 11** – Función de supervivencia, método KM, variable rango de plazo (R\_Plazo)

```
## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$R_Plazo)
##
##
##               n events median 0.95LCL 0.95UCL
## phcalf16_m2a$R_Plazo=menor a 20 años 40212  3177    NA      NA      NA
## phcalf16_m2a$R_Plazo=mayor a 20 años 39013  4577    NA      NA      NA
```



**Figura 13** – Función supervivencia, método KM, var.: rango de plazo (R\_Plazo)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor igual a 0, se rechaza la hipótesis nula ( $H_0$ ) de igualdad de



funciones, con lo cual se puede manifestar que las funciones de supervivencia entre quienes integran los diferentes rangos de plazo fueron diferentes.

**Tabla 12** – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de plazo (R\_Plazo)

```
## Call:
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$R_Plazo)
##
##
##           N Observed Expected (O-E)^2/E
## phcalf16_m2a$R_Plazo=menor a 20 años 40212      3177      4121       216
## phcalf16_m2a$R_Plazo=mayor a 20 años 39013      4577      3633       245
##
##           (O-E)^2/V
## phcalf16_m2a$R_Plazo=menor a 20 años      463
## phcalf16_m2a$R_Plazo=mayor a 20 años      463
##
## Chisq= 463 on 1 degrees of freedom, p= 0
```

### Variable: rango de ingreso

La variable rango de ingreso (R\_Ingp) es una variable nominal que condensa la variable nivel de ingreso del cliente de un crédito hipotecario al momento de la precalificación en cinco (5) opciones de respuesta: “hasta 770 USD (Q1)”, “entre 770,01 y 1.133 USD (Q2)”, “entre 1.133,01 y 1.603 USD (Q3)”, “entre 1.603,01 y 2.362 USD (Q4)” y “mayor a 2.362 USD (Q5)”.

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se distribuyeron según la variable rango de ingreso (R\_Ingp) por quintiles en un 22,2% en el rango “hasta 770 USD (Q1)”, en un 20,8% en el rango “entre 770,01 y 1.133 USD (Q2)”, en un 20,6% en el rango “entre 1.133,01 y 1.603 USD (Q3)”, en un 19,2% en el rango “entre 1.603,01 y 2.362 USD (Q4)” y en un 17,3% en el rango “mayor a 2.362 USD (Q5)”.

**Tabla 13** – Tabla de contingencia - variable rango de ingreso (R\_Ingp) vs variable presencia mora grave (Mayor90)

```

##
## Cell Contents
## |-----|
## |                Count |
## |            Row Percent |
## |            Column Percent |
## |-----|
##
## Total Observations in Table: 79225
##
##
##          phcalf16_m2a$R_Ingp | phcalf16_m2a$Mayor90
##          -----|-----|-----|
##          hasta 770 USD (Q1) | 15463 | 2104 | 17567
##          | 88.0% | 12.0% | 22.2%
##          | 21.6% | 27.1% |
##          -----|-----|-----|
##          entre 770,01 y 1.133 USD (Q2) | 14872 | 1585 | 16457
##          | 90.4% | 9.6% | 20.8%
##          | 20.8% | 20.4% |
##          -----|-----|-----|
##          entre 1.133,01 y 1.603 USD (Q3) | 14739 | 1551 | 16290
##          | 90.5% | 9.5% | 20.6%
##          | 20.6% | 20.0% |
##          -----|-----|-----|
##          entre 1.603,01 y 2.362 USD (Q4) | 13942 | 1268 | 15210
##          | 91.7% | 8.3% | 19.2%
##          | 19.5% | 16.4% |
##          -----|-----|-----|
##          mayor a 2.362 USD (Q5) | 12455 | 1246 | 13701
##          | 90.9% | 9.1% | 17.3%
##          | 17.4% | 16.1% |
##          -----|-----|-----|
##          Column Total | 71471 | 7754 | 79225
##          | 90.2% | 9.8% |
##          -----|-----|-----|
##
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----
## Chi^2 = 140.865    d.f. = 4    p = 1.84266e-29
##
##
##
## Minimum expected frequency: 1340.96

```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable rango de ingreso (R\_Ingp) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en el rango de “hasta 770 USD” (12,0%), incluso superando levemente la estructura natural de la opción presencia de mora grave (1) de la variable mora grave (Mayor90) (9.8%).

Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables rango de ingreso (R\_Ingp) y mora grave (Mayor90) se puede señalar que dado un p-valor asociado inferior a 0,05 (1,84266e-29) se puede rechazar la hipótesis de independencia de variables (h0)

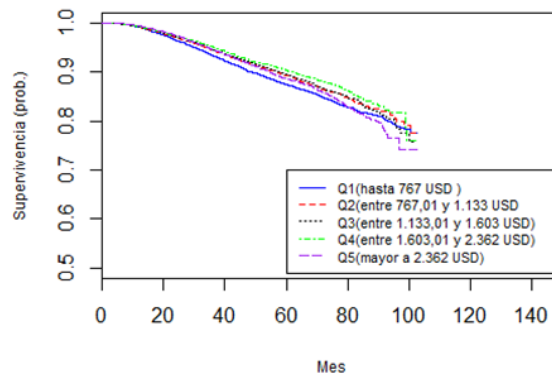
y concluir que existe una relación de dependencia entre las dos variables analizadas.

Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM) para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable rango de ingreso (R\_Ingp), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 20, y con mayor intensidad en aquellos clientes con un ingreso de hasta 767 USD. Adicionalmente, cabe mencionar que a partir del mes 80 la probabilidad de supervivencia descendió substancialmente en los clientes con ingresos superiores 2.362 USD, incluso en una magnitud inferior a la vislumbrada en los clientes de ingresos de hasta 767 USD.

Así también, cabe indicar que dada la alta proporción de casos censurados en cada uno de los rangos de ingreso no fue posible calcular la mediana de supervivencia en cada uno de los casos.

**Tabla 14** – Función de supervivencia, método KM, variable rango de ingreso (R\_Ingp)

```
## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$R_Ingp)
##
##
##               n events median
## phcalf16_m2a$R_Ingp=hasta 770 USD (Q1)      17567   2104   NA
## phcalf16_m2a$R_Ingp=entre 770,01 y 1.133 USD (Q2) 16457   1585   NA
## phcalf16_m2a$R_Ingp=entre 1.133,01 y 1.603 USD (Q3) 16290   1551   NA
## phcalf16_m2a$R_Ingp=entre 1.603,01 y 2.362 USD (Q4) 15210   1268   NA
## phcalf16_m2a$R_Ingp=mayor a 2.362 USD (Q5)      13701   1246   NA
##
##               0.95LCL 0.95UCL
## phcalf16_m2a$R_Ingp=hasta 770 USD (Q1)           NA     NA
## phcalf16_m2a$R_Ingp=entre 770,01 y 1.133 USD (Q2) NA     NA
## phcalf16_m2a$R_Ingp=entre 1.133,01 y 1.603 USD (Q3) NA     NA
## phcalf16_m2a$R_Ingp=entre 1.603,01 y 2.362 USD (Q4) NA     NA
## phcalf16_m2a$R_Ingp=mayor a 2.362 USD (Q5)       NA     NA
```



**Figura 14** – Función supervivencia, método KM, var.: rango de ingreso (R\_Ingp)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor cercano a 0 ( $1,51e-10$ ), se rechaza la hipótesis nula ( $h_0$ ) de igualdad de funciones, y con ello se puede señalar que las funciones de supervivencia entre quienes integran los diferentes rangos de ingreso fueron diferentes.

**Tabla 15** – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de ingreso (R\_Ingp)

```
## Call:
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$R_Ingp)
##
##
##                               N Observed
## phcalf16_m2a$R_Ingp=hasta 770 USD (Q1) 17567 2104
## phcalf16_m2a$R_Ingp=entre 770,01 y 1.133 USD (Q2) 16457 1585
## phcalf16_m2a$R_Ingp=entre 1.133,01 y 1.603 USD (Q3) 16290 1551
## phcalf16_m2a$R_Ingp=entre 1.603,01 y 2.362 USD (Q4) 15210 1268
## phcalf16_m2a$R_Ingp=mayor a 2.362 USD (Q5) 13701 1246
##
##                               Expected (O-E)^2/E
## phcalf16_m2a$R_Ingp=hasta 770 USD (Q1) 1873 28.534
## phcalf16_m2a$R_Ingp=entre 770,01 y 1.133 USD (Q2) 1617 0.639
## phcalf16_m2a$R_Ingp=entre 1.133,01 y 1.603 USD (Q3) 1609 2.117
## phcalf16_m2a$R_Ingp=entre 1.603,01 y 2.362 USD (Q4) 1436 19.729
## phcalf16_m2a$R_Ingp=mayor a 2.362 USD (Q5) 1218 0.629
##
##                               (O-E)^2/V
## phcalf16_m2a$R_Ingp=hasta 770 USD (Q1) 37.814
## phcalf16_m2a$R_Ingp=entre 770,01 y 1.133 USD (Q2) 0.808
## phcalf16_m2a$R_Ingp=entre 1.133,01 y 1.603 USD (Q3) 2.672
## phcalf16_m2a$R_Ingp=entre 1.603,01 y 2.362 USD (Q4) 24.234
## phcalf16_m2a$R_Ingp=mayor a 2.362 USD (Q5) 0.749
##
## Chisq= 51.8 on 4 degrees of freedom, p= 1.51e-10
```

## Variable: rango de monto financiado

La variable rango de monto financiado (R\_Montf) es una variable nominal que condensa la variable monto financiado por crédito hipotecario en cinco (5) opciones de respuesta: “hasta 40.000 USD”, “entre 40.000,01 y 70.000 USD”, “entre 70.000,01 y 100.000 USD”, “entre 100.000,01 y 150.000 USD” y “mayor a 150.000 USD”.

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se distribuyeron según la variable rango de monto financiado (R\_Montf) mayoritariamente en un 45,3% en el rango “hasta 40.000 USD”, en un 36,9% en el rango “entre 40.000,01 y 70.000 USD” y en un 13,1% en el rango “entre 70.000,01 y 100.000 USD”.

**Tabla 16** – Tabla de contingencia - variable rango de monto financiado (R\_Montf) vs variable presencia mora grave (Mayor90)

```
##
## Cell Contents
## |-----|
## |                Count |
## |            Row Percent |
## |            Column Percent |
## |-----|
##
## Total Observations in Table:  79225
##
##          phcalf16_m2a$R_Montf | phcalf16_m2a$Mayor90
##          -----|-----|-----|
##          hasta 40.000 USD      | 0 | 1 | Row Total |
##          -----|-----|-----|
##          | 32113 | 3740 | 35853 |
##          | 89.6% | 10.4% | 45.3% |
##          | 44.9% | 48.2% | -----|
##          -----|-----|-----|
##          entre 40.000,01 y 70.000 USD | 26560 | 2677 | 29237 |
##          | 90.8% | 9.2% | 36.9% |
##          | 37.2% | 34.5% | -----|
##          -----|-----|-----|
##          entre 70.000,01 y 100.000 USD | 9411 | 957 | 10368 |
##          | 90.8% | 9.2% | 13.1% |
##          | 13.2% | 12.3% | -----|
##          -----|-----|-----|
##          entre 100.000,01 y 150.000 USD | 2913 | 303 | 3216 |
##          | 90.6% | 9.4% | 4.1% |
##          | 4.1% | 3.9% | -----|
##          -----|-----|-----|
##          mayor a 150.000 USD | 474 | 77 | 551 |
##          | 86.0% | 14.0% | 0.7% |
##          | 0.7% | 1.0% | -----|
##          -----|-----|-----|
##          Column Total | 71471 | 7754 | 79225 |
##          | 90.2% | 9.8% | -----|
##          -----|-----|-----|
```

```

## -----|-----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----
## Chi^2 = 45.11045    d.f. = 4    p = 3.771178e-09
##
##
## Minimum expected frequency: 53.9281

```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable rango de monto financiado (R\_Montf) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en los rangos “mayor a 150.000 USD” (14,0%) y “hasta 40.000 USD” (10,4%), incluso superando levemente la estructura natural de la opción presencia de mora grave (1) de la variable mora grave (Mayor90) (9.8%).

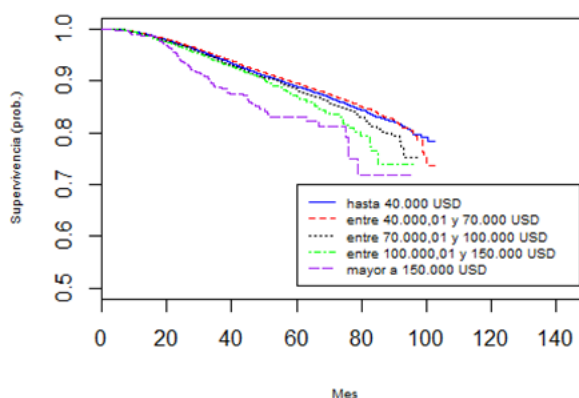
Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables rango de monto financiado (R\_Montf) y mora grave (Mayor90) se puede indicar que dado un p-valor asociado inferior a 0,05 (3,771178e-09) se puede rechazar la hipótesis de independencia de variables ( $H_0$ ) y concluir que existe una relación de dependencia entre las dos variables analizadas.

Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM) para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable rango de monto financiado (R\_Montf), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 20, y con mayor énfasis en aquellos clientes con un monto financiado superior a 150.000 USD. Adicionalmente, cabe mencionar que a partir del mes 46 la probabilidad de supervivencia descendió substancialmente en los clientes con montos financiados de entre 100.000,01 y 150.000 USD.

Así también, cabe indicar que dada la alta proporción de casos censurados en cada uno de los rangos de monto financiado no fue posible calcular la mediana de supervivencia en cada uno de los casos.

**Tabla 17** – Función de supervivencia, método KM, variable rango de monto financiado (R\_Montf)

```
## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$R_Montf)
##
##
##                               n events median
## phcalf16_m2a$R_Montf=hasta 40.000 USD      35853   3740   NA
## phcalf16_m2a$R_Montf=entre 40.000,01 y 70.000 USD 29237   2677   NA
## phcalf16_m2a$R_Montf=entre 70.000,01 y 100.000 USD 10368    957   NA
## phcalf16_m2a$R_Montf=entre 100.000,01 y 150.000 USD  3216    303   NA
## phcalf16_m2a$R_Montf=mayor a 150.000 USD      551     77   NA
##
##                               0.95LCL 0.95UCL
## phcalf16_m2a$R_Montf=hasta 40.000 USD      NA     NA
## phcalf16_m2a$R_Montf=entre 40.000,01 y 70.000 USD  NA     NA
## phcalf16_m2a$R_Montf=entre 70.000,01 y 100.000 USD  NA     NA
## phcalf16_m2a$R_Montf=entre 100.000,01 y 150.000 USD  NA     NA
## phcalf16_m2a$R_Montf=mayor a 150.000 USD      NA     NA
```



**Figura 15** – Función supervivencia, método KM, var.: rango de monto financiado (R\_Montf)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor cercano a 0 ( $8,91e-10$ ), se rechaza la hipótesis nula ( $h_0$ ) de igualdad de funciones, y con ello se puede manifestar que las funciones de supervivencia entre quienes integran los diferentes rangos de monto financiado fueron diferentes.

**Tabla 18** – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de monto financiado (R\_Montf)

```
## Call:
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$R_Montf)
##
##
##                               N Observed
## phcalf16_m2a$R_Montf=hasta 40.000 USD      35853   3740
## phcalf16_m2a$R_Montf=entre 40.000,01 y 70.000 USD 29237   2677
```

```

## phcalf16_m2a$R_Montf=entre 70.000,01 y 100.000 USD 10368 957
## phcalf16_m2a$R_Montf=entre 100.000,01 y 150.000 USD 3216 303
## phcalf16_m2a$R_Montf=mayor a 150.000 USD 551 77
## Expected (O-E)^2/E
## phcalf16_m2a$R_Montf=hasta 40.000 USD 3747 0.0145
## phcalf16_m2a$R_Montf=entre 40.000,01 y 70.000 USD 2818 7.0632
## phcalf16_m2a$R_Montf=entre 70.000,01 y 100.000 USD 891 4.9269
## phcalf16_m2a$R_Montf=entre 100.000,01 y 150.000 USD 255 9.1184
## phcalf16_m2a$R_Montf=mayor a 150.000 USD 43 26.8843
## (O-E)^2/V
## phcalf16_m2a$R_Montf=hasta 40.000 USD 0.0282
## phcalf16_m2a$R_Montf=entre 40.000,01 y 70.000 USD 11.1101
## phcalf16_m2a$R_Montf=entre 70.000,01 y 100.000 USD 5.5903
## phcalf16_m2a$R_Montf=entre 100.000,01 y 150.000 USD 9.4530
## phcalf16_m2a$R_Montf=mayor a 150.000 USD 27.0449
##
## Chisq= 48.1 on 4 degrees of freedom, p= 8.91e-10

```

### Variable: subregión geográfica

La variable subregión geográfica (SubReg) es una variable nominal que agrupa en subregiones geográficas las provincias en las cuales los clientes obtuvieron su crédito hipotecario en seis (6) opciones de respuesta: “SIERRA\_N (CAR., IMB., PIC.)”, “SIERRA\_C (COT., TUNG., BOL., CHI)”, “SIERRA\_S (CAÑ., AZU., LOJ.)”, “COSTA (incluye STO.DOM. y GAL.)” , “ORIENTE\_N (SUC., NAP., ORE.)” y “ORIENTE\_CS (PAS,MOR.,ZAM.)”.

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se distribuyeron según la variable subregión geográfica (SubReg) mayoritariamente en un 47,8% en la opción “COSTA (incluye STO.DOM. y GAL.)” y en un 38,6% en la opción “SIERRA\_N (CAR., IMB., PIC.)”.

**Tabla 19** – Tabla de contingencia - variable subregión geográfica (SubReg) vs variable presencia mora grave (Mayor90)

```

##
## Cell Contents
## |-----|
## |                Count |
## |                Row Percent |
## |                Column Percent |
## |-----|
##
## Total Observations in Table: 79225
##
##          phcalf16_m2a$SubReg | phcalf16_m2a$Mayor90
##          -----|-----|-----|
##          SIERRA_N (CAR., IMB., PIC.) | 28138 | 2442 | 30580 |
##          -----|-----|-----|
##          | 92.0% | 8.0% | 38.6% |

```



```

##          |          |          |          |
## -----|-----|-----|-----|
## SIERRA_C (COT., TUNG., BOL.,CHI) | 3368 | 214 | 3582 |
##          | 94.0% | 6.0% | 4.5% |
##          | 4.7% | 2.8% |      |
## -----|-----|-----|-----|
##      SIERRA_S (CAÑ., AZU., LOJ.) | 4816 | 265 | 5081 |
##          | 94.8% | 5.2% | 6.4% |
##          | 6.7% | 3.4% |      |
## -----|-----|-----|-----|
## COSTA (incluye STO.DOM. y GAL.) | 33370 | 4483 | 37853 |
##          | 88.2% | 11.8% | 47.8% |
##          | 46.7% | 57.8% |      |
## -----|-----|-----|-----|
##      ORIENTE_N (SUC., NAP., ORE.) | 1169 | 298 | 1467 |
##          | 79.7% | 20.3% | 1.9% |
##          | 1.6% | 3.8% |      |
## -----|-----|-----|-----|
##      ORIENTE_CS (PAS,MOR.,ZAM.) | 610 | 52 | 662 |
##          | 92.1% | 7.9% | 0.8% |
##          | 0.9% | 0.7% |      |
## -----|-----|-----|-----|
##          Column Total | 71471 | 7754 | 79225 |
##          | 90.2% | 9.8% |      |
## -----|-----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----
## Chi^2 = 659.7878    d.f. = 5    p = 2.425538e-140
##
##
##
##      Minimum expected frequency: 64.79202

```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable subregión geográfica (SubReg) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en las opciones “ORIENTE\_N (SUC., NAP., ORE.)” (20,3%) y “COSTA (incluye STO.DOM. y GAL.)” (11,8%), incluso superando levemente la estructura natural de la opción presencia de mora grave (1) de la variable mora grave (Mayor90) (9.8%).

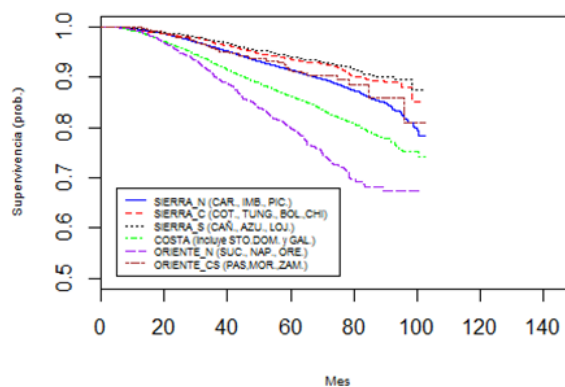
Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables subregión geográfica (SubReg) y mora grave (Mayor90) se puede señalar que dado un p-valor asociado inferior a 0,05 (2,425538e-140) se puede rechazar la hipótesis de independencia de variables (h0) y concluir que existe una relación de dependencia entre las dos variables analizadas.

Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM) para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable subregión geográfica (SubReg), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 20, y con mayor intensidad en aquellos clientes localizados en las subregiones “ORIENTE\_N (SUC., NAP., ORE.)” y “COSTA (incluye STO.DOM. y GAL.)”

Así también, cabe indicar que dada la alta proporción de casos censurados en cada una de las subregiones geográficas no fue posible calcular la mediana de supervivencia en cada uno de los casos.

**Tabla 20** – Función de supervivencia, método KM, variable subregión geográfica (SubReg)

```
## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$SubReg)
##
##
##                               n events median
## phcalf16_m2a$SubReg=SIERRA_N (CAR., IMB., PIC.)      30580      2442      NA
## phcalf16_m2a$SubReg=SIERRA_C (COT., TUNG., BOL.,CHI)   3582        214      NA
## phcalf16_m2a$SubReg=SIERRA_S (CAÑ., AZU., LOJ.)        5081        265      NA
## phcalf16_m2a$SubReg=COSTA (incluye STO.DOM. y GAL.)    37853      4483      NA
## phcalf16_m2a$SubReg=ORIENTE_N (SUC., NAP., ORE.)      1467         298      NA
## phcalf16_m2a$SubReg=ORIENTE_CS (PAS,MOR.,ZAM.)         662          52      NA
##
##                               0.95LCL 0.95UCL
## phcalf16_m2a$SubReg=SIERRA_N (CAR., IMB., PIC.)        NA        NA
## phcalf16_m2a$SubReg=SIERRA_C (COT., TUNG., BOL.,CHI)   NA        NA
## phcalf16_m2a$SubReg=SIERRA_S (CAÑ., AZU., LOJ.)        NA        NA
## phcalf16_m2a$SubReg=COSTA (incluye STO.DOM. y GAL.)    NA        NA
## phcalf16_m2a$SubReg=ORIENTE_N (SUC., NAP., ORE.)      NA        NA
## phcalf16_m2a$SubReg=ORIENTE_CS (PAS,MOR.,ZAM.)        NA        NA
```



**Figura 16** – Función supervivencia, método KM, var.: subregión geográfica (SubReg)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor cercano a 0, se rechaza la hipótesis nula ( $h_0$ ) de igualdad de funciones, y con ello se puede indicar que las funciones de supervivencia entre quienes integran las diferentes subregiones geográficas fueron diferentes.

**Tabla 21** – Resultado del contraste de igualdad de funciones de supervivencia para la variable subregión geográfica (SubReg)

```
## Call:
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$SubReg)
##
##
##                               N Observed
## phcalf16_m2a$SubReg=SIERRA_N (CAR., IMB., PIC.)      30580      2442
## phcalf16_m2a$SubReg=SIERRA_C (COT., TUNG., BOL.,CHI)  3582       214
## phcalf16_m2a$SubReg=SIERRA_S (CAÑ., AZU., LOJ.)       5081       265
## phcalf16_m2a$SubReg=COSTA (incluye STO.DOM. y GAL.)  37853     4483
## phcalf16_m2a$SubReg=ORIENTE_N (SUC., NAP., ORE.)     1467       298
## phcalf16_m2a$SubReg=ORIENTE_CS (PAS,MOR.,ZAM.)        662        52
##
##                               Expected (O-E)^2/E
## phcalf16_m2a$SubReg=SIERRA_N (CAR., IMB., PIC.)      3136     153.55
## phcalf16_m2a$SubReg=SIERRA_C (COT., TUNG., BOL.,CHI)  378       71.49
## phcalf16_m2a$SubReg=SIERRA_S (CAÑ., AZU., LOJ.)       519     124.37
## phcalf16_m2a$SubReg=COSTA (incluye STO.DOM. y GAL.)  3496     278.93
## phcalf16_m2a$SubReg=ORIENTE_N (SUC., NAP., ORE.)     154     134.96
## phcalf16_m2a$SubReg=ORIENTE_CS (PAS,MOR.,ZAM.)        71        5.11
##
##                               (O-E)^2/V
## phcalf16_m2a$SubReg=SIERRA_N (CAR., IMB., PIC.)      258.44
## phcalf16_m2a$SubReg=SIERRA_C (COT., TUNG., BOL.,CHI)  75.19
## phcalf16_m2a$SubReg=SIERRA_S (CAÑ., AZU., LOJ.)     133.33
## phcalf16_m2a$SubReg=COSTA (incluye STO.DOM. y GAL.)  509.97
## phcalf16_m2a$SubReg=ORIENTE_N (SUC., NAP., ORE.)    137.73
## phcalf16_m2a$SubReg=ORIENTE_CS (PAS,MOR.,ZAM.)       5.16
##
## Chisq= 771 on 5 degrees of freedom, p= 0
```

Es importante señalar, que con la finalidad de utilizar esta variable en la formulación de un modelo multivariado se decidió agrupar sus opciones de respuesta en dos (2): “COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE)” y “OTRAS SUBREGIONES”. El nuevo nombre de la variable con el ajuste citado sería SubReg2.

### Variable: rango de edad

La variable rango de edad (R\_Edad) es una variable nominal que condensa la variable edad de los clientes al momento de la precalificación de su crédito hipotecario en tres (3) opciones de respuesta: “hasta 35 años”, “entre 36 y 50 años” y “entre 51 y 75 años”.

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se distribuyeron según la variable rango de edad (R\_Edad) mayoritariamente en un 42,6% en el rango “entre 36 y 50 años” y en un 39,8% en el rango “hasta 35 años”.

**Tabla 22** – Tabla de contingencia - variable rango de edad (R\_Edad) vs variable presencia mora grave (Mayor90)

```
##
## Cell Contents
## |-----|
## |                Count |
## |            Row Percent |
## |            Column Percent |
## |-----|
##
## Total Observations in Table:  79225
##
##
##      phcalf16_m2a$Mayor90
## phcalf16_m2a$R_Edad |      0      1 | Row Total |
## -----|-----|-----|
##      hasta 35 años | 28042   3465 | 31507 |
##                   | 89.0%   11.0% | 39.8% |
##                   | 39.2%   44.7% |       |
## -----|-----|-----|
##      entre 36 y 50 años | 30356   3410 | 33766 |
##                   | 89.9%   10.1% | 42.6% |
##                   | 42.5%   44.0% |       |
## -----|-----|-----|
##      entre 51 y 75 años | 13073    879 | 13952 |
##                   | 93.7%    6.3% | 17.6% |
##                   | 18.3%   11.3% |       |
## -----|-----|-----|
##      Column Total | 71471   7754 | 79225 |
##                   | 90.2%    9.8% |       |
## -----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## -----|-----|-----|
## Chi^2 = 248.1312    d.f. = 2    p = 1.315213e-54
##
##
##
## Minimum expected frequency: 1365.526
```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable rango de edad (R\_Edad) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en los rangos “hasta 35 años” (11,0%) y “entre 36 y 50 años” (10,1%), incluso

superando levemente la estructura natural de la opción presencia de mora grave (1) de la variable mora grave (Mayor90) (9.8%).

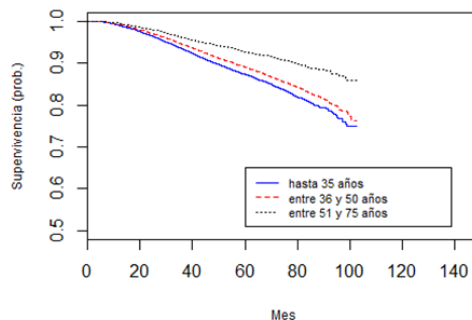
Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables rango de edad (R\_Edad) y mora grave (Mayor90) se puede señalar que dado un p-valor asociado inferior a 0,05 (1,315213e-54) se puede rechazar la hipótesis de independencia de variables ( $H_0$ ) y concluir que existe una relación de dependencia entre las dos variables analizadas.

Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM) para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable rango de edad (R\_Edad), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 20, y con mayor intensidad en aquellos clientes con edades de hasta 35 años.

Así también, cabe indicar que dada la alta proporción de casos censurados en cada uno de los rangos de edad no fue posible calcular la mediana de supervivencia en cada uno de los casos.

**Tabla 23** – Función de supervivencia, método KM, variable rango de edad (R\_Edad)

```
## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##      phcalf16_m2a$R_Edad)
##
##
##              n events median 0.95LCL 0.95UCL
## phcalf16_m2a$R_Edad=hasta 35 años      31507   3465    NA      NA      NA
## phcalf16_m2a$R_Edad=entre 36 y 50 años 33766   3410    NA      NA      NA
## phcalf16_m2a$R_Edad=entre 51 y 75 años 13952    879    NA      NA      NA
```



**Figura 17** – Función supervivencia, método KM, var.: rango de edad (R\_Edad)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor cercano a 0, se rechaza la hipótesis nula ( $H_0$ ) de igualdad de funciones, y con ello se puede señalar que las funciones de supervivencia entre quienes integran los diferentes rangos de edad fueron diferentes.

**Tabla 24** – Resultado del contraste de igualdad de funciones de supervivencia para la variable rango de edad (R\_Edad)

```
## Call:
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$R_Edad)
##
##
##              N Observed Expected (O-E)^2/E
## phcalf16_m2a$R_Edad=hasta 35 años      31507      3465      2964      84.7544
## phcalf16_m2a$R_Edad=entre 36 y 50 años  33766      3410      3416       0.0103
## phcalf16_m2a$R_Edad=entre 51 y 75 años  13952       879      1374     178.4789
##
##              (O-E)^2/V
## phcalf16_m2a$R_Edad=hasta 35 años      137.3088
## phcalf16_m2a$R_Edad=entre 36 y 50 años    0.0185
## phcalf16_m2a$R_Edad=entre 51 y 75 años   216.9781
##
## Chisq= 263 on 2 degrees of freedom, p= 0
```

### Variable: actividad económica del empleador

La variable actividad económica del empleador (Acteco\_r) es una variable nominal que agrupa las diferentes actividades económicas en las cuales se desempeñaban los clientes de créditos hipotecarios al momento de la precalificación en nueve (9) opciones de respuesta: “AGRÍCOLA”, “CONSTRUCCIÓN”, “EMPRESA PRIVADA, SOCIEDADES / COMPANIAS”, “EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA”, “ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS”, “ENTIDADES SECTOR PÚBLICO - SECCIONALES”, “OTROS”, “SIN INFORMACIÓN” y “TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA”.

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se distribuyeron según la variable actividad económica (Acteco\_r) mayoritariamente en un 49,9% en la opción “EMPRESA PRIVADA, SOCIEDADES / COMPANIAS”

y en un 34,9% en la opción “ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS”.

**Tabla 25** – Tabla de contingencia - variable actividad económica (Acteco\_r) vs variable presencia mora grave (Mayor90)

```
##
## Cell Contents
## |-----|
## |                Count |
## |            Row Percent |
## |            Column Percent |
## |-----|
##
## Total Observations in Table:  79225
##
##
##          phcalf16_m2a$Acteco_r | phcalf16_m2a$Mayor90
##          | 0 | 1 | Row Total |
##-----|-----|-----|-----|
##          AGRÍCOLA | 129 | 18 | 147 |
##          | 87.8% | 12.2% | 0.2% |
##          | 0.2% | 0.2% | |
##-----|-----|-----|-----|
##          CONSTRUCCIÓN | 292 | 85 | 377 |
##          | 77.5% | 22.5% | 0.5% |
##          | 0.4% | 1.1% | |
##-----|-----|-----|-----|
##          EMPRESA PRIVADA, SOCIEDADES / COMPANIAS | 34694 | 4817 | 39511 |
##          | 87.8% | 12.2% | 49.9% |
##          | 48.5% | 62.1% | |
##-----|-----|-----|-----|
##          EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA | 2428 | 566 | 2994 |
##          | 81.1% | 18.9% | 3.8% |
##          | 3.4% | 7.3% | |
##-----|-----|-----|-----|
##          ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS | 26182 | 1499 | 27681 |
##          | 94.6% | 5.4% | 34.9% |
##          | 36.6% | 19.3% | |
##-----|-----|-----|-----|
##          ENTIDADES SECTOR PÚBLICO - SECCIONALES | 4225 | 374 | 4599 |
##          | 91.9% | 8.1% | 5.8% |
##          | 5.9% | 4.8% | |
##-----|-----|-----|-----|
##          OTROS | 3366 | 351 | 3717 |
##          | 90.6% | 9.4% | 4.7% |
##          | 4.7% | 4.5% | |
##-----|-----|-----|-----|
##          SIN INFORMACIÓN | 50 | 23 | 73 |
##          | 68.5% | 31.5% | 0.1% |
##          | 0.1% | 0.3% | |
##-----|-----|-----|-----|
##          TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA | 105 | 21 | 126 |
##          | 83.3% | 16.7% | 0.2% |
##          | 0.1% | 0.3% | |
##-----|-----|-----|-----|
##          Column Total | 71471 | 7754 | 79225 |
##          | 90.2% | 9.8% | |
##-----|-----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
##-----|-----|-----|-----|
## Chi^2 = 1270.836    d.f. = 8    p = 4.726349e-269
##
```

```
##
##
##      Minimum expected frequency: 7.14474
```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable actividad económica (Acteco\_r) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en las opciones “SIN INFORMACIÓN” (31,5%), “CONSTRUCCIÓN” (22,5%), “EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA” (18,9%) y “TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA” (16,7%), incluso superando la estructura natural de la opción presencia de mora grave (1) de la variable mora grave (Mayor90) (9.8%).

Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables actividad económica (Acteco\_r) y mora grave (Mayor90) se puede manifestar que dado un p-valor asociado inferior a 0,05 (4,726349e-269) se puede rechazar la hipótesis de independencia de variables ( $H_0$ ) y concluir que existe una relación de dependencia entre las dos variables analizadas.

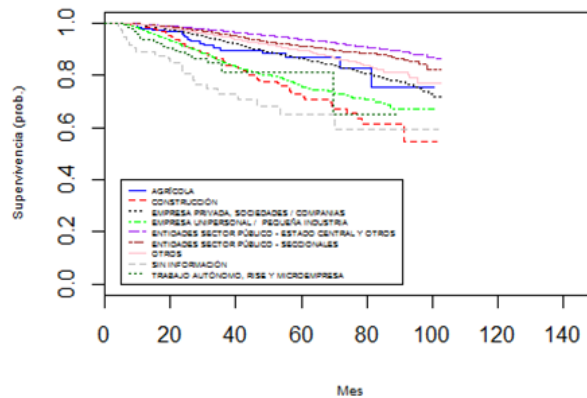
Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM) para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable actividad económica (Acteco\_r), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 5, y con mayor intensidad en aquellos clientes que desempeñaron actividades en las opciones: “SIN INFORMACIÓN”, “TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA”, “CONSTRUCCIÓN” y “EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA”.

Así también, cabe indicar que dada la alta proporción de casos censurados en cada una de las actividades económicas no fue posible calcular la mediana de supervivencia en cada uno de los casos.



**Tabla 26** – Función de supervivencia, método KM, variable actividad económica (Acteco\_r)

```
## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$Acteco_r)
##
##
##                                     n
## phcalf16_m2a$Acteco_r=AGRÍCOLA      147
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN  377
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPANIAS 39511
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA 2994
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS 27681
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES 4599
## phcalf16_m2a$Acteco_r=OTROS          3717
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN 73
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA 126
##                                     events
## phcalf16_m2a$Acteco_r=AGRÍCOLA      18
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN  85
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPANIAS 4817
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA 566
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS 1499
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES 374
## phcalf16_m2a$Acteco_r=OTROS          351
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN 23
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA 21
##                                     median
## phcalf16_m2a$Acteco_r=AGRÍCOLA      NA
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN  NA
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPANIAS NA
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA NA
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS NA
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES NA
## phcalf16_m2a$Acteco_r=OTROS          NA
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN NA
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA NA
##                                     0.95LCL
## phcalf16_m2a$Acteco_r=AGRÍCOLA      NA
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN  91.1
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPANIAS NA
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA NA
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS NA
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES NA
## phcalf16_m2a$Acteco_r=OTROS          NA
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN 70.3
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA 69.7
##                                     0.95UCL
## phcalf16_m2a$Acteco_r=AGRÍCOLA      NA
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN  NA
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPANIAS NA
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA NA
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS NA
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES NA
## phcalf16_m2a$Acteco_r=OTROS          NA
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN NA
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA NA
```



**Figura 18** – Función supervivencia, método KM, var.: actividad económica (Acteco\_r)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor cercano a 0, se rechaza la hipótesis nula ( $H_0$ ) de igualdad de funciones, y con ello se puede indicar que las funciones de supervivencia entre quienes integran las diferentes actividades económicas fueron diferentes.

**Tabla 27** – Resultado del contraste de igualdad de funciones de supervivencia para la variable actividad económica (Acteco\_r)

```
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$Acteco_r)
##
##
##                                     N
## phcalf16_m2a$Acteco_r=AGRÍCOLA      147
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN  377
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPAÑIAS 39511
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA 2994
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS 27681
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES 4599
## phcalf16_m2a$Acteco_r=OTROS         3717
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN 73
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA 126
##
## Observed
## phcalf16_m2a$Acteco_r=AGRÍCOLA      18
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN  85
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPAÑIAS 4817
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA 566
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS 1499
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES 374
## phcalf16_m2a$Acteco_r=OTROS         351
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN 23
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA 21
##
## Expected
## phcalf16_m2a$Acteco_r=AGRÍCOLA      13.42
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN  31.02
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPAÑIAS 3859.97
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA 234.32
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS 2747.44
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES 496.20
```

```

## phcalf16_m2a$Acteco_r=OTROS 358.90
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN 5.59
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA 7.15
## (0-E)^2/E
## phcalf16_m2a$Acteco_r=AGRÍCOLA 1.560
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN 93.951
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPANIAS 237.285
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA 469.514
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS 567.290
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES 30.093
## phcalf16_m2a$Acteco_r=OTROS 0.174
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN 54.199
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA 26.846
## (0-E)^2/V
## phcalf16_m2a$Acteco_r=AGRÍCOLA 1.563
## phcalf16_m2a$Acteco_r=CONSTRUCCIÓN 94.344
## phcalf16_m2a$Acteco_r=EMPRESA PRIVADA, SOCIEDADES / COMPANIAS 472.725
## phcalf16_m2a$Acteco_r=EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA 484.698
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS 879.017
## phcalf16_m2a$Acteco_r=ENTIDADES SECTOR PÚBLICO - SECCIONALES 32.171
## phcalf16_m2a$Acteco_r=OTROS 0.183
## phcalf16_m2a$Acteco_r=SIN INFORMACIÓN 54.243
## phcalf16_m2a$Acteco_r=TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA 26.879
##
## Chisq= 1482 on 8 degrees of freedom, p= 0

```

Es importante señalar, que con la finalidad de utilizar esta variable en la formulación de un modelo multivariado se decidió agrupar sus opciones de respuesta en dos (2): “OTRAS ACTIVIDADES” y “EMPRESA UNIPERSONAL/ CONSTRUCCIÓN/TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA/SIN INFORMACIÓN”. El nuevo nombre de la variable con el ajuste citado sería Acteco\_r2.

### **Variable: ciclo económico al momento del desembolso**

La variable ciclo económico al momento del desembolso (Ciclo) es una variable nominal que indica la fase del ciclo económico en el que se encuentra la economía con respecto al crecimiento promedio de largo plazo al momento de la precalificación del crédito hipotecario. La estimación de los períodos de los ciclos y la fase de cada ciclo lo desarrolla el Banco Central del Ecuador a través de su sistema de indicadores del ciclo económico (Banco Central del Ecuador (a), 2016). Las opciones de respuesta de esta variable son cuatro (4) y el detalle de las mismas es el siguiente: “Expansión”, “Desaceleración sobre la tendencia”, “Desaceleración bajo la tendencia” y “Recuperación”.

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se

distribuyeron según la variable ciclo económico al momento del desembolso (Ciclo) mayoritariamente en un 44,5% en el ciclo de “Expansión” y en un 33,8% en el ciclo “Desaceleración sobre la tendencia”.

**Tabla 28** – Tabla de contingencia - variable ciclo económico al momento del desembolso (Ciclo) vs variable presencia mora grave (Mayor90)

```
##
## Cell Contents
## |-----|
## |                Count |
## |                Row Percent |
## |                Column Percent |
## |-----|
##
## Total Observations in Table:  79225
##
##          phcalf16_m2a$Ciclo | phcalf16_m2a$Mayor90
##          -----|-----|-----|
##          Expansion          | 0          1          Row Total |
##          -----|-----|-----|
##          Expansion          | 32163     3094     35257 |
##          -----|-----|-----|
##          Expansion          | 91.2%     8.8%     44.5% |
##          -----|-----|-----|
##          Expansion          | 45.0%     39.9%     |
##          -----|-----|-----|
## Desaceleración sobre la tendencia | 24251     2505     26756 |
##          -----|-----|-----|
## Desaceleración sobre la tendencia | 90.6%     9.4%     33.8% |
##          -----|-----|-----|
## Desaceleración sobre la tendencia | 33.9%     32.3%     |
##          -----|-----|-----|
## Desaceleración bajo la tendencia | 6750      951      7701 |
##          -----|-----|-----|
## Desaceleración bajo la tendencia | 87.7%     12.3%     9.7% |
##          -----|-----|-----|
## Desaceleración bajo la tendencia | 9.4%      12.3%     |
##          -----|-----|-----|
## Recuperación                | 8307      1204     9511 |
##          -----|-----|-----|
## Recuperación                | 87.3%     12.7%     12.0% |
##          -----|-----|-----|
## Recuperación                | 11.6%     15.5%     |
##          -----|-----|-----|
## Column Total                | 71471     7754     79225 |
##          -----|-----|-----|
## Column Total                | 90.2%     9.8%     |
##          -----|-----|-----|
##
##
## Statistics for All Table Factors
##
## Pearson's Chi-squared test
## -----
## Chi^2 = 192.4184    d.f. = 3    p = 1.833022e-41
##
##
## Minimum expected frequency: 753.7211
```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable ciclo económico al momento del desembolso (Ciclo) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en las fases del ciclo “Recuperación” (12,7%) y “Desaceleración bajo la tendencia” (12,3%), incluso superando ligeramente la

estructura natural de la opción presencia de mora grave (1) de la variable mora grave (Mayor90) (9.8%).

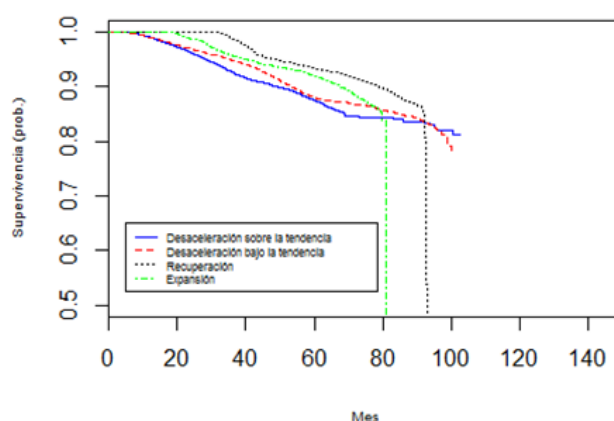
Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables ciclo económico al momento del desembolso (Ciclo) y mora grave (Mayor90) se puede señalar que dado un p-valor asociado inferior a 0,05 ( $1,833022e-41$ ) se puede rechazar la hipótesis de independencia de variables ( $H_0$ ) y concluir que existe una relación de dependencia entre las dos variables analizadas.

Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM) para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable ciclo económico al momento del desembolso (Ciclo), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 10, y con mayor intensidad en aquellos clientes que obtuvieron sus créditos en las fases del ciclo “Desaceleración sobre la tendencia” y “Desaceleración bajo la tendencia”.

Así también, cabe indicar que dada la alta proporción de casos censurados en cada uno de las fases del ciclo al momento del desembolso no fue posible calcular la mediana de supervivencia en cada uno de los casos.

**Tabla 29** – Función de supervivencia, método KM, variable ciclo económico al momento del desembolso (Ciclo)

```
## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$Ciclo)
##
##               n events median
## phcalf16_m2a$Ciclo=Expansión      35257      3094      NA
## phcalf16_m2a$Ciclo=Desaceleración sobre la tendencia 26756      2505      NA
## phcalf16_m2a$Ciclo=Desaceleración bajo la tendencia   7701       951     93.0
## phcalf16_m2a$Ciclo=Recuperación      9511     1204     81.1
##                               0.95LCL 0.95UCL
## phcalf16_m2a$Ciclo=Expansión              NA      NA
## phcalf16_m2a$Ciclo=Desaceleración sobre la tendencia  NA      NA
## phcalf16_m2a$Ciclo=Desaceleración bajo la tendencia   92.6     NA
## phcalf16_m2a$Ciclo=Recuperación           80.8     NA
```



**Figura 19** – Función supervivencia, método KM, var.: ciclo económico al momento del desembolso (Cicloe)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor cercano a 0, se rechaza la hipótesis nula ( $H_0$ ) de igualdad de funciones, y con ello se puede manifestar que las funciones de supervivencia entre quienes se aglutinaron en las diferentes fases del ciclo económico fueron diferentes.

**Tabla 30** – Resultado del contraste de igualdad de funciones de supervivencia para la variable ciclo económico al momento del desembolso (Cicloe)

```
## Call:
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$Cicloe)
##
##
##                               N Observed
## phcalf16_m2a$Cicloe=Expansión      35257      3094
## phcalf16_m2a$Cicloe=Desaceleración sobre la tendencia 26756      2505
## phcalf16_m2a$Cicloe=Desaceleración bajo la tendencia  7701       951
## phcalf16_m2a$Cicloe=Recuperación    9511      1204
##
##                               Expected (O-E)^2/E
## phcalf16_m2a$Cicloe=Expansión          2498      142.37
## phcalf16_m2a$Cicloe=Desaceleración sobre la tendencia 2486       0.14
## phcalf16_m2a$Cicloe=Desaceleración bajo la tendencia  1387     137.08
## phcalf16_m2a$Cicloe=Recuperación       1383      23.15
##
##                               (O-E)^2/V
## phcalf16_m2a$Cicloe=Expansión          228.128
## phcalf16_m2a$Cicloe=Desaceleración sobre la tendencia  0.217
## phcalf16_m2a$Cicloe=Desaceleración bajo la tendencia 191.551
## phcalf16_m2a$Cicloe=Recuperación        29.880
##
## Chisq= 359 on 3 degrees of freedom, p= 0
```

Es importante señalar, que con la finalidad de utilizar esta variable en la formulación de un modelo multivariado se decidió agrupar sus opciones de

respuesta en dos (2): “(0) Desempeño económico sobre el crecimiento promedio de largo plazo y “(1) Desempeño económico bajo crecimiento promedio de largo plazo”. El nuevo nombre de la variable con el ajuste citado sería Cicloe2.

**Variable: cesante previo a la mora grave**

La variable cesante previo a la mora grave (Cesante) es una variable nominal que indica la presencia o ausencia de la condición de cesante de forma previa al suceso de una mora grave. En caso de no darse el suceso de mora grave se registra la condición al término del estudio en cada uno de los casos. Las opciones de respuesta de la variable son dos (2): “(0) No cesante” y “(1) Cesante”.

El análisis de frecuencia de la variable analizada refirió en primera instancia no contar con casos perdidos, con lo cual contó con 79.225 casos. Así también develo que los clientes del crédito hipotecario considerados se distribuyeron según la variable cesante previo a la mora grave (Cesante) mayoritariamente en un 89,7% en la categoría “(0) No cesante” y en un 10,3% en la categoría “(1) Cesante”.

**Tabla 31** – Tabla de contingencia - variable cesante previo a la mora grave (Cesante) vs variable presencia mora grave (Mayor90)

```
##
## Cell Contents
## |-----|
## |                Count |
## |            Row Percent |
## |            Column Percent |
## |-----|
##
## Total Observations in Table:  79225
##
##
## phcalf16_m2a$cesante | phcalf16_m2a$Mayor90
## |-----|-----|-----|
## |                0 |                1 | Row Total |
## |-----|-----|-----|
## |                0 | 67497 | 3545 | 71042 |
## |                | 95.0% |  5.0% | 89.7% |
## |                | 94.4% | 45.7% |      |
## |-----|-----|-----|
## |                1 | 3974 | 4209 | 8183 |
## |                | 48.6% | 51.4% | 10.3% |
## |                |  5.6% | 54.3% |      |
## |-----|-----|-----|
## | Column Total | 71471 | 7754 | 79225 |
## |                | 90.2% |  9.8% |      |
## |-----|-----|-----|
##
##
## Statistics for All Table Factors
##
##
```

```

## Pearson's Chi-squared test
## -----
## Chi^2 = 17927.88    d.f. = 1    p = 0
##
## Pearson's Chi-squared test with Yates' continuity correction
## -----
## Chi^2 = 17922.62    d.f. = 1    p = 0
##
##
## Minimum expected frequency: 800.896

```

A partir del análisis de la variable mora grave (Mayor90) según las opciones de la variable cesante previo a la mora grave (Cesante) se puede señalar que la proporción de quienes incurrieron en una mora grave se suscitó en mayor medida en la categoría “(1) Cesante” (51,4%).

Adicionalmente, a partir la prueba ji-cuadrado de independencia realizada a la tabla de contingencia de las variables cesante previo a la mora grave (Cesante) y mora grave (Mayor90) se puede señalar que dado un p-valor asociado inferior a 0,05 (0) se puede rechazar la hipótesis de independencia de variables ( $H_0$ ) y concluir que existe una relación de dependencia entre las dos variables analizadas.

Por otra parte, la aplicación del análisis de supervivencia mediante el modelo de Kaplan – Meier (KM) para identificar la función de supervivencia frente a la posibilidad de incurrir en una mora grave (Mayor 90) según la variable cesante previo a la mora grave (Cesante), denotó que la probabilidad de supervivencia tendió a descender en mayor medida a partir del mes 5, y con mayor intensidad en aquellos clientes que cayeron en la condición de cesante.

Así también, cabe indicar que dada la alta proporción de casos censurados en cada una de los opciones de la variable cesante previo a la mora grave no fue posible calcular la mediana de supervivencia en cada uno de los casos.

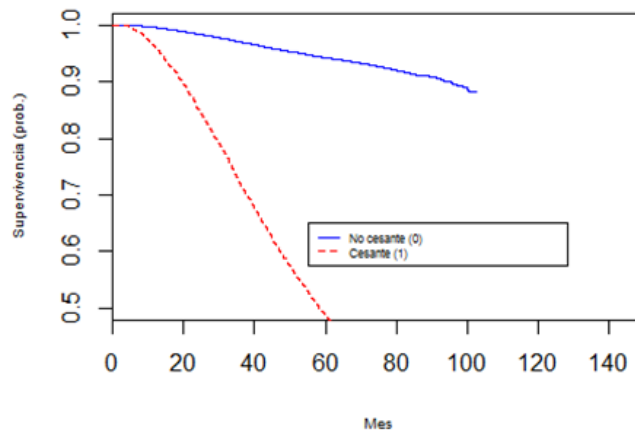
**Tabla 32** – Función de supervivencia, método KM, variable cesante previo a la mora grave (Cesante)

```

## Call: survfit(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$cesante)
##
##              n events median 0.95LCL 0.95UCL
## phcalf16_m2a$cesante=0 71042  3545    NA      NA      NA
## phcalf16_m2a$cesante=1  8183  4209  58.3     57     60

```





**Figura 20** – Función supervivencia, método KM, var.: cesante previo a la mora grave (Cesante)

Finalmente, luego de aplicar el estadístico de contraste log-rank el cual sigue una distribución ji-cuadrado para la evaluación de la hipótesis de igualdad de funciones de supervivencia a un nivel de confianzas del 95%, se puede afirmar que dado un p-valor cercano a 0, se rechaza la hipótesis nula ( $H_0$ ) de igualdad de funciones, y consecuentemente se puede indicar que las funciones de supervivencia entre no cesantes y cesantes.

**Tabla 33** – Resultado del contraste de igualdad de funciones de supervivencia para la variable cesante previo a la mora grave (Cesante)

```
## survdiff(formula = Surv(phcalf16_m2a$tiempoM90pv_2mes, phcalf16_m2a$Mayor90) ~
##   phcalf16_m2a$cesante)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## phcalf16_m2a$cesante=0 71042    3545    7080    1765    20338
## phcalf16_m2a$cesante=1  8183     4209     674    18551    20338
##
## Chisq= 20338 on 1 degrees of freedom, p= 0
```

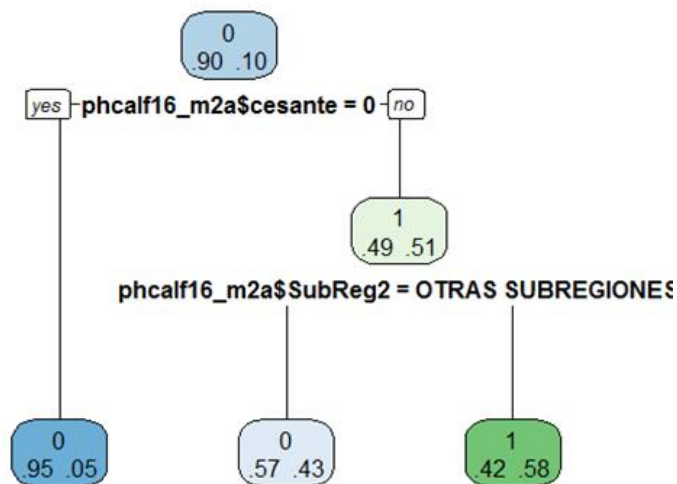
## Identificación de variables relacionadas con el evento de mora grave

Una vez que se caracterizaron alrededor de (11) variables y se identificaron las funciones de supervivencia con respecto al evento de incurrir en una mora grave relacionada con cada una de ellas, es pertinente identificar de forma previa a la formulación de un modelo de cox, aquellas variables que tienen mayor relación o pueden constituirse en variables explicativas de la variable mora grave (Mayor90).

Para ello, se utilizó el modelo no paramétrico denominado “árboles de decisión”. La variable de segmentación utilizada fue la variable mora grave (Mayor90) en la cual la opción de respuesta “0” indica la no ocurrencia de una mora grave y “1” indica la ocurrencia de una mora grave. Las variables consideradas inicialmente como explicativas fueron: rango de avalúo del bien inmueble (R\_Avaluo2), género (GENERO), rango de plazo (R\_Plazo), rango de ingreso (R\_Ingp), rango de monto financiado (R\_Montf), subregión geográfica resumida (Subreg2), rango de edad (R\_Edad), actividad económica del empleador resumida (Acteco\_r2), ciclo económico previo al desembolso (Ciclo) y cesante previo a la mora grave (Cesante).

**Tabla 34** – Árbol de decisión: Variable de segmentación: mora grave (Mayor90)

```
## Classification tree:
## rpart(formula = phcalf16_m2a$Mayor90 ~ phcalf16_m2a$R_Avaluo2 +
##   phcalf16_m2a$GENERO + phcalf16_m2a$R_Plazo + phcalf16_m2a$R_Ingp +
##   phcalf16_m2a$R_Montf + phcalf16_m2a$SubReg2 + phcalf16_m2a$R_Edad +
##   phcalf16_m2a$Acteco_r2 + phcalf16_m2a$Ciclo + phcalf16_m2a$cesante,
##   data = phcalf16_m2a, method = "class")
##
## Variables actually used in tree construction:
## [1] phcalf16_m2a$cesante phcalf16_m2a$SubReg2
##
## Root node error: 7754/79225 = 0.097873
##
## n= 79225
##
##          CP nsplit rel error  xerror   xstd
## 1 0.046428    0  1.00000  1.00000  0.010786
## 2 0.010000    2  0.90714  0.90714  0.010325
##
## n= 79225
##
## node), split, n, loss, yval, (yprob)
##   * denotes terminal node
##
## 1) root 79225 7754 0 (0.90212685 0.09787315)
##   2) phcalf16_m2a$cesante=0 71042 3545 0 (0.95009994 0.04990006) *
##   3) phcalf16_m2a$cesante=1 8183 3974 1 (0.48564096 0.51435904)
##     6) phcalf16_m2a$SubReg2=OTRAS SUBREGIONES 3573 1544 0 (0.56787014 0.43212986) *
##     7) phcalf16_m2a$SubReg2=COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 4610 1945 1
##        (0.42190889 0.57809111) *
```



**Figura 21** – Árbol de decisión: Variable de segmentación: mora grave (Mayor90)

Como resultado de la aplicación de la técnica de árbol de decisión se identificaron dos variables relevantes en la clasificación de clientes que incurrieron en una mora grave (1) y aquellos que no (0). Las referidas variables fueron: cesante previo a la mora grave (cesante) y subregiones geográficas resumidas (SubReg2).

El árbol de decisión resultante indicó, a través de su nodo inicial, que la probabilidad de no incurrir en una mora grave fue de 0,90, mientras que la de incurrir en una mora grave fue de apenas el 0,10.

La primera variable considerada para la segmentación fue la variable cesante previo a la mora grave (cesante), si el resultado de esta variable es “0” es decir el cliente no cae en la condición de cesante, la probabilidad de caer un una mora grave es de apenas 0,05 frente a la de no caer en esa situación (0,95) (nodo terminal). Por otra parte, si el resultado de la variable cesante previo a la mora grave es “1” es decir el cliente cayó en la condición de cesante, la probabilidad de caer un una mora grave se incrementa y se sitúa en 0,51 frente a la de no caer en esa situación (0,49).

A partir de lo anterior, si el cliente cayó en la condición de cesante y se localizó en la subregión denominada “Otras regiones”, la probabilidad de caer en una mora grave es del 0,43, frente a la de no caer en esta situación (0,57) (nodo terminal). Finalmente, si el cliente cayó en la condición de cesante y se localizó en

la subregión denominada “Costa (incl. Sto y Gal.) y Oriente N (Nap.,Suc.,Ore.)”, la probabilidad de caer en una mora grave es del 0,58, frente a la de no caer en esta situación (0,42)(nodo terminal).

En resumen, el perfil de cliente que mayor probabilidad tiene de incurrir en una mora grave fue el de quien cayó en la condición de cesante y se localizó geográficamente en las provincias de la costa ecuatoriana, incluidas las provincias de Santo Domingo de los Tsáchilas y Galápagos, así como en la provincias del oriente cercanas a la frontera norte con Colombia (0,58).

### **Formulación de un modelo de regresión de cox**

#### ***Consideraciones específicas y base de datos***

Conforme los análisis previos realizados con respecto a las variables a utilizar y sus correspondientes funciones de supervivencia construidas, se procedió a probar algunos modelos preliminares de COX utilizando la totalidad de la base de datos perfilada de manera previa (79.225 casos). En estos ejercicios se encontró variables significativas para el análisis pero bastantes débiles para la estimación.

En virtud de lo anterior, se decidió tomar una sub-base de datos para la formulación de un modelo de cox más robusto. Específicamente, para la construcción del modelo se utilizó una sub-base de datos, construida con los casos de los clientes de créditos hipotecarios que desempeñaron actividades en los sectores “EMPRESA UNIPERSONAL / CONSTRUCCIÓN / TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA / SIN INFORMACIÓN”, dado que en los análisis previos y específicamente en las curvas de supervivencia construidas de la variable “actividad económica del empleador” se vislumbró que quienes se encontraban asociados a las actividades mencionadas experimentaban, en términos relativos, el descenso más pronunciado en el tiempo de su probabilidad de sobrevivir al evento de una mora grave.

Con lo expuesto anteriormente, la base utilizada para la construcción del modelo de COX fue aquella constituida por 3.570 casos que cumplen con las consideraciones expuestas en el párrafo anterior.

### ***Variables utilizadas***

Como resultado de probar múltiples combinaciones de variables y de probar su validez y aporte, se definieron las siguientes como las idóneas entre las disponibles para integrar el modelo de cox propuesto.

#### *Variable dependiente*

- Variable continua tiempo en meses que tarda un crédito hipotecario en caer en una mora grave (tiempoM90pv\_2mes)
- Variable binaria mora grave (Mayor90).  
Opciones de respuesta:
  - "0".- No ha experimentado mora grave.
  - "1".- Experimentó una mora grave.

Las variables expuestas configuraron la tipología de variables dependientes requeridas para estimar un modelo de supervivencia.

#### *Variables explicativas*

En primera instancia, de las variables analizadas previamente se seleccionaron las siguientes:

- Variable nominal rango de ingreso (R\_Ingp).  
Opciones de respuesta:
  - "hasta 770 USD (Q1)" (opción base)
  - "entre 770,01 y 1.133 USD (Q2)"
  - "entre 1.133,01 y 1.603 USD (Q3)"
  - "entre 1.603,01 y 2.362 USD (Q4)"
  - "mayor a 2.362 USD (Q5)"
- Variable nominal rango de plazo (R\_Plazo).  
Opciones de respuesta:
  - "menor a 20 años" (opción base)
  - "mayor a 20 años"

- Variable nominal subregión geográfica resumida (SubReg2).
  - Opciones de respuesta:
    - "otras subregiones" (opción base)
    - "Costa (incluye Sto.Dom. y Gal.) y Oriente Norte (Nap., Suc., Ore.)"
  
- Variable nominal ciclo económico al momento del desembolso resumido (Cicloe2).
  - Opciones de respuesta:
    - "0".- desempeño económico sobre el crecimiento promedio de largo plazo (opción base)
    - "1".- desempeño económico bajo crecimiento promedio de largo plazo

Adicionalmente, con la finalidad de mejorar la capacidad predictiva del modelo se incorporaron las siguientes variables, algunas de las cuales son dependientes del tiempo:

- Variable discreta año de concesión (tyear).
  
- Variable nominal ciclo económico previo a la mora grave (Cicloef2).
  - Opciones de respuesta:
    - "0".- desempeño económico sobre el crecimiento promedio de largo plazo (opción base)
    - "1".- desempeño económico bajo crecimiento promedio de largo plazo
  
- Variable nominal cesante y ciclo económico por debajo del crecimiento de largo plazo previo a la mora grave (cesantexciclof).
  - Opciones de respuesta:
    - "0".- El cliente no se encontró simultáneamente en la condición de cesante y en un ciclo económico por debajo del crecimiento de largo plazo previo a la mora grave (opción base).
    - "1".- El cliente sí se encontró simultáneamente en la condición de cesante y en un ciclo económico por debajo del crecimiento de largo plazo previo a la mora grave

## Validación e interpretación de variables del modelo

### Validación individual de variables

Es importante señalar que en función de que el p-valor asociado al estadístico z en cada una de las variables explicativas fue inferior a 0,05, se rechazó la hipótesis nula de que cada uno de los coeficientes es igual a 0 (h0) y con ello se puede afirmar que todas las variables de manera individual resultaron estadísticamente significativas.

**Tabla 35** – Modelo de cox – Clientes de créditos hipotecarios que desempeñaron labores en "Empresa unipersonal / construcción / trabajo autónomo, rise y microempresa / sin información" – significancia individual de variables

```
## Call: coxph(formula = Surv(phcalf16$tiempoM90pv_2mes, phcalf16$Mayor90) ~
##      1)
##
## Null model
##   log likelihood= -95246.18
##     n= 110775

## Call:
## coxph(formula = Surv(phcalf16_actg$tiempoM90pv_2mes, phcalf16_actg$Mayor90) ~
##      phcalf16_actg$R_Ingp + phcalf16_actg$R_Plazo + phcalf16_actg$SubReg2 +
##      phcalf16_actg$Cicloef2 + phcalf16_actg$tyear + phcalf16_actg$Cicloef2 +
##      phcalf16_actg$cesantexciclof)
##
## n= 3570, number of events= 695
##
##      z
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)          3.390
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)       3.461
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4)       4.524
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5)                 6.641
## phcalf16_actg$R_Plazomayor a 20 años                       2.994
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 6.660
## phcalf16_actg$Cicloef21                                    5.433
## phcalf16_actg$tyear                                        21.255
## phcalf16_actg$Cicloef21                                   -42.941
## phcalf16_actg$cesantexciclof1                             19.137

##
##      Pr(>|z|)
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)         0.000700
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)       0.000537
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4)       6.07e-06
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5)                 3.11e-11
## phcalf16_actg$R_Plazomayor a 20 años                       0.002755
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 2.73e-11
## phcalf16_actg$Cicloef21                                    5.55e-08
## phcalf16_actg$tyear                                        < 2e-16
## phcalf16_actg$Cicloef21                                   < 2e-16
## phcalf16_actg$cesantexciclof1                             < 2e-16

##
##      ***
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)          ***
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)       ***
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4)       ***
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5)                 ***
## phcalf16_actg$R_Plazomayor a 20 años                       **
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) ***
## phcalf16_actg$Cicloef21                                    ***
```

```

## phcalf16_actg$tyear ***
## phcalf16_actg$Cicloef21 ***
## phcalf16_actg$cesantexciclof1 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

### Validación global del modelo

Por otra parte, los p-valor asociados a los estadísticos test de razón de verosimilitud (likelihood ratio test), test de wald (walt test) y test de puntajes (score [logrank]) al presentar magnitudes inferiores a 0,05, se puede rechazar la hipótesis de que los coeficientes del modelo en su conjunto son iguales a 0 ( $h_0$ ), y con ello se puede afirmar que el modelo en su conjunto es estadísticamente significativo.

Finalmente, el coeficiente de determinación r-cuadrado del modelo obtenido indicó que el modelo puede explicar el 0,488 de la varianza de la problemática analizada. Si bien la magnitud del coeficiente pudo lucir insuficiente ante la necesidad de estimar los tiempos de supervivencia a partir de las variables contempladas en el modelo, cabe señalar que esta situación puede mejorar substancialmente con la incorporación de dos variables que no estuvieron disponibles al momento de la construcción del presente modelo como son: 1) peor calificación crediticia en los 2 años anteriores a la concesión del crédito y 2) estado civil del cliente, las cuales han sido relevantes en la identificación de casos de mora grave en ejercicios empíricos realizados.

**Tabla 36** – Modelo de cox – Clientes de créditos hipotecarios que desempeñaron labores en "Empresa unipersonal / construcción / trabajo autónomo, rise y microempresa / sin información" – significancia global

```

## Call:
## coxph(formula = Surv(phcalf16_actg$tiempoM90pv_2mes, phcalf16_actg$Mayor90) ~
##   phcalf16_actg$R_Ingp + phcalf16_actg$R_Plazo + phcalf16_actg$SubReg2 +
##   phcalf16_actg$Cicloef2 + phcalf16_actg$tyear + phcalf16_actg$Cicloef2 +
##   phcalf16_actg$cesantexciclof)
##
## n= 3570, number of events= 695
##
## Concordance= 0.927 (se = 0.012 )
## Rsquare= 0.488 (max possible= 0.95 )
## Likelihood ratio test= 2387 on 10 df, p=0
## Wald test = 2037 on 10 df, p=0
## Score (logrank) test = 4066 on 10 df, p=0

```



### Interpretación de los resultados del modelo

En primera instancia, a partir de los signos de los coeficientes se puede señalar que las variables rango de edad con sus opciones de respuesta "entre 770,01 y 1.133 USD (Q2)", "entre 1.133,01 y 1.603 USD (Q3)", "entre 1.603,01 y 2.362 USD (Q4)" y "mayor a 2.362 USD (Q5)"; rango de plazo con su opción de respuesta "mayor a 20 años"; subregión geográfica con su opción de respuesta "Costa (incluye Sto. Dom. y Gal.) y Oriente Norte (Nap., Suc., Ore)"; ciclo económico previo al desembolso resumido con su opción de respuesta " (1) desempeño económico bajo crecimiento promedio de largo plazo"; año de concesión; y cesante y ciclo económico por debajo del crecimiento de largo plazo previo a la mora grave con su opción de respuesta "(1) el cliente sí se encontró simultáneamente en la condición de cesante y en un ciclo económico por debajo del crecimiento de largo plazo previo a la mora grave"; al presentar signo positivo denotaron que su presencia incrementa el riesgo de incurrir en una situación de mora grave.

Por su parte, la variable ciclo económico previo a la mora grave resumido con su opción de respuesta "(1) desempeño económico bajo crecimiento promedio de largo plazo"; al presentar signo negativo indicó que su presencia disminuye el riesgo de incurrir en una mora grave.

**Tabla 37** – Modelo de cox – Clientes de créditos hipotecarios que desempeñaron labores en "Empresa unipersonal / construcción / trabajo autónomo, rise y microempresa / sin información" – interpretación de coeficientes

```
## Call:
## coxph(formula = Surv(phcalf16_actg$tiempoM90pv_2mes, phcalf16_actg$Mayor90) ~
##   phcalf16_actg$R_Ingp + phcalf16_actg$R_Plazo + phcalf16_actg$SubReg2 +
##   phcalf16_actg$Cicloef2 + phcalf16_actg$tyear + phcalf16_actg$Cicloef2 +
##   phcalf16_actg$cesantexciclof)
##   n= 3570, number of events= 695
##
##                                     coef
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)      0.377662
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)    0.384910
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4)    0.535503
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5)             0.891376
## phcalf16_actg$R_Plazomayor a 20 años                   0.252111
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 0.544825
## phcalf16_actg$Cicloef21                                0.836868
## phcalf16_actg$tyear                                    1.178660
## phcalf16_actg$Cicloef21                                -5.699248
## phcalf16_actg$cesantexciclof1                          2.039368
##
##                                     exp(coef)
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)    1.458870
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)  1.469482
```

```

## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) 1.708307
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) 2.438484
## phcalf16_actg$R_Plazomayor a 20 años 1.286739
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 1.724306
## phcalf16_actg$Cicloef21 2.309123
## phcalf16_actg$tyear 3.250016
## phcalf16_actg$Cicloef21 0.003348
## phcalf16_actg$cesantexciclof1 7.685748

## se(coef)
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2) 0.111421
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3) 0.111200
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) 0.118367
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) 0.134217
## phcalf16_actg$R_Plazomayor a 20 años 0.084211
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 0.081800
## phcalf16_actg$Cicloef21 0.154043
## phcalf16_actg$tyear 0.055452
## phcalf16_actg$Cicloef21 0.132724
## phcalf16_actg$cesantexciclof1 0.106567

## exp(coef)
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2) 1.458870
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3) 1.469482
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) 1.708307
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) 2.438484
## phcalf16_actg$R_Plazomayor a 20 años 1.286739
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 1.724306
## phcalf16_actg$Cicloef21 2.309123
## phcalf16_actg$tyear 3.250016
## phcalf16_actg$Cicloef21 0.003348
## phcalf16_actg$cesantexciclof1 7.685748

## lower .95
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2) 1.172667
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3) 1.181710
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) 1.354602
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) 1.874452
## phcalf16_actg$R_Plazomayor a 20 años 1.090964
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 1.468879
## phcalf16_actg$Cicloef21 1.707360
## phcalf16_actg$tyear 2.915308
## phcalf16_actg$Cicloef21 0.002582
## phcalf16_actg$cesantexciclof1 6.237002

## upper .95
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2) 1.814923
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3) 1.827332
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) 2.154371
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) 3.172235
## phcalf16_actg$R_Plazomayor a 20 años 1.517646
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 2.024151
## phcalf16_actg$Cicloef21 3.122979
## phcalf16_actg$tyear 3.623152
## phcalf16_actg$Cicloef21 0.004343
## phcalf16_actg$cesantexciclof1 9.471013

```

En una segunda instancia, mediante la interpretación del exponencial de los coeficientes  $\beta$  (exp (coef)) se puede manifestar lo siguiente:

Aquellos clientes que se encontraron simultáneamente en la condición de cesante y en una fase del ciclo económico por debajo del crecimiento de largo plazo, fueron 7,69 veces más riesgosos de incurrir en una mora grave que quienes no presentaron estas características.

Así también aquellos clientes que en la variable ciclo económico previo al desembolso se mostraron en la opción "(1) desempeño económico bajo crecimiento promedio de largo plazo" fueron 2,31 veces más riesgosos de incurrir en una mora grave que quienes se mostraron en la opción "(0) desempeño económico sobre crecimiento promedio de largo plazo".

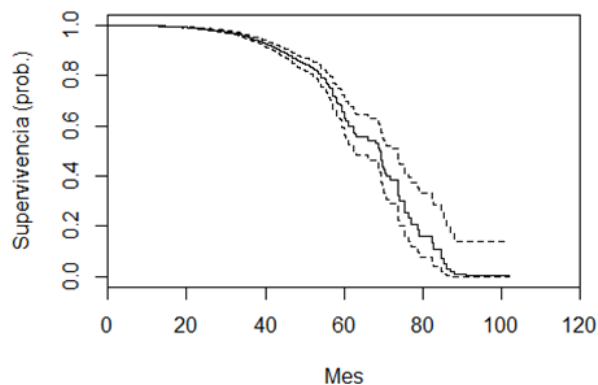
En lo que respecta al año de concesión del crédito se puede constatar que con cada año que pasó el riesgoso de incurrir en una mora grave se incrementó en 3,25 veces.

Por su parte, también se pudo constatar que a mayor ingreso del cliente existe mayor riesgo de incurrir en una mora grave. Específicamente, en comparación a quienes presentaron ingresos de "hasta 770 USD (Q1)" y se constituyen en la opción base, quienes mostraron ingresos "entre 770,01 y 1.133 USD (Q2)" fueron 1,46 veces más riesgosos de incurrir en una mora grave que aquellos clientes de la opción base. Y de esta manera consecutivamente quienes presentaron ingresos de "entre 1.133,01 y 1.603 USD (Q3)" fueron 1,47 veces más riesgosos, quienes mostraron ingresos de "entre 1.603,01 y 2.362 USD (Q4)" fueron 1,71 veces más riesgosos y finalmente, quienes presentaron ingresos "mayores a 2.362 USD (Q5)" fueron 2,44 veces más riesgosos de incurrir en una mora grave que quienes formaron parte de la opción base.

Adicionalmente, se avizó que aquellos clientes de crédito hipotecario y se localizaron geográficamente en las provincias de la opción "'Costa (incluye Sto.Dom. y Gal.) y Oriente Norte (Nap., Suc., Ore.)" fueron 1,72 veces más riesgosos de incurrir en una mora grave que quienes se dispusieron geográficamente en "otras subregiones".

Finalmente, entre los exponenciales de los coeficientes que resultaron relevantes se puede interpretar que quienes optaron por un plazo para su crédito hipotecario "mayor a 20 años" fueron 1,29 veces más riesgosos de incurrir en una mora grave que quienes optaron por un plazo "menor a 20 años".

Complementariamente, la función de supervivencia resultante del modelo propuesto denotó que la probabilidad de supervivencia a una mora grave por parte de los clientes del producto préstamo hipotecario que desempeñaron actividades económicas en los sectores "EMPRESA UNIPERSONAL / CONSTRUCCIÓN / TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA / SIN INFORMACIÓN" tendió a contraerse substancialmente a partir del mes 40 (3,33 años). La información detallada de la función de supervivencia obtenida se presenta en el anexo 1.



**Figura 22** – Función de supervivencia resultante de la aplicación del modelo de cox (mora grave (> 90 días))

#### Validación del supuesto de riesgo proporcional

De manera general, los p-valores asociados al estadístico Rho, que sigue una distribución ji-cuadrado, pertenecientes a casi la totalidad de las variables y al modelo de forma global presentaron una magnitud superior 0,05, con lo cual no existiría evidencia significativa al 95% de confianza de que exista una violación a los supuestos de riesgo proporcional, tanto a nivel de variables como del modelo de forma general. Consecuentemente, se acepta la hipótesis de cumplimiento de los supuestos de riesgo proporcional ( $h_0$ ).

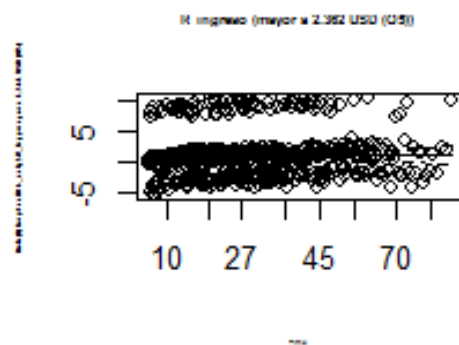
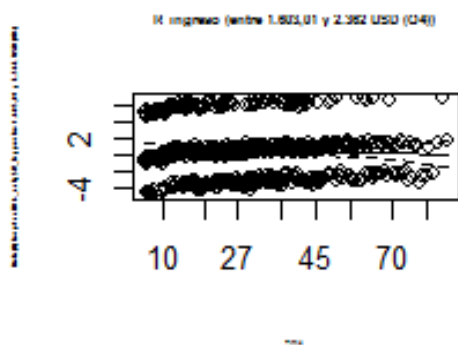
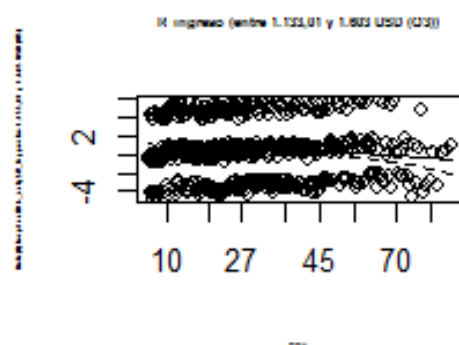
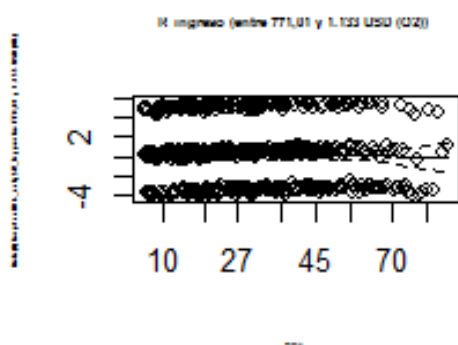
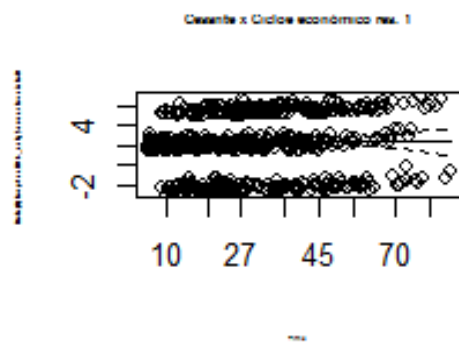
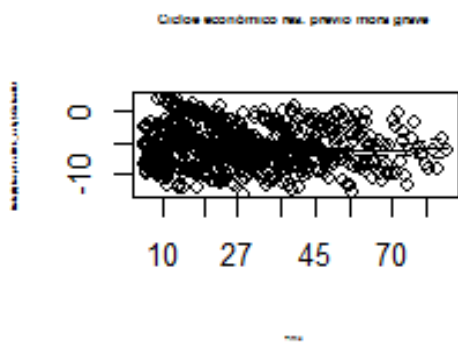
Sin embargo, es importante señalar que el p-valor asociado a la variable ciclo económico por debajo del crecimiento de largo plazo, opción (1) y la variable nominal combinada cesante y ciclo económico por debajo del crecimiento de largo plazo, opción (1); fue inferior a 0,05 con lo cual estas variables no cumpliría con el supuesto de riesgo proporcional por ser variables dependientes del tiempo. Sin embargo, su inclusión fue relevante al momento de mejorar la robustez del modelo

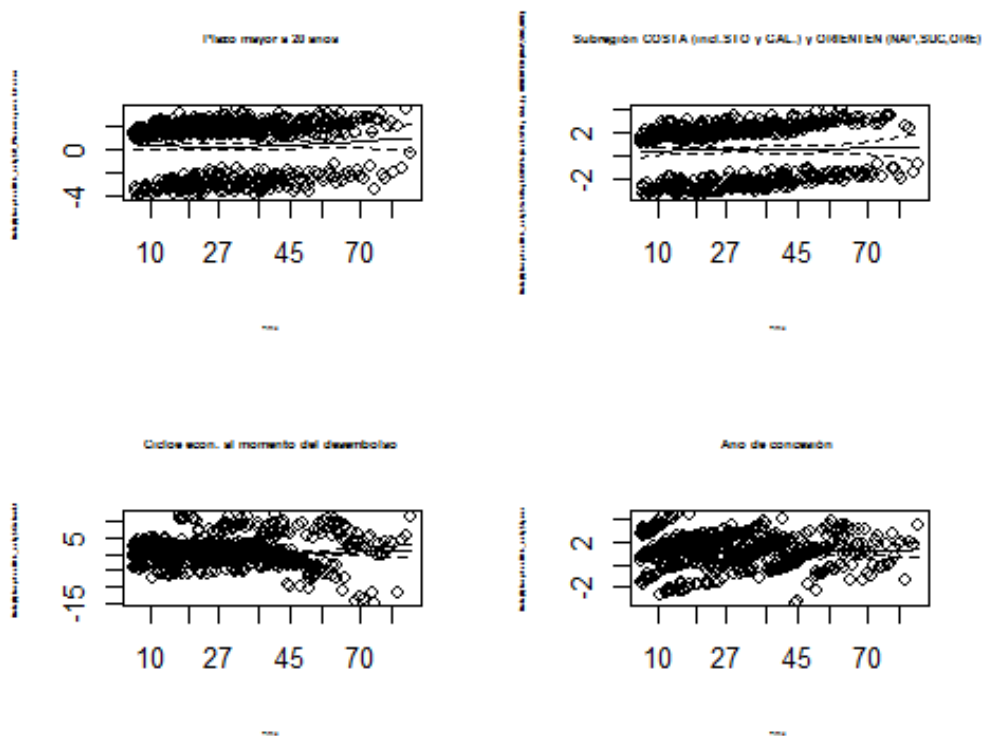
y adicionalmente su comportamiento se presentó acorde a la dinámica del negocio financiero analizado.

**Tabla 38** – Modelo de cox – Clientes de créditos hipotecarios que desempeñaron labores en "Empresa unipersonal / construcción / trabajo autónomo, rise y microempresa / sin información" – supuestos de riesgo proporcional

	rho
##	
## phcalf16_actg\$R_Ingpentre 770,01 y 1.133 USD (Q2)	-0.04096
## phcalf16_actg\$R_Ingpentre 1.133,01 y 1.603 USD (Q3)	-0.04176
## phcalf16_actg\$R_Ingpentre 1.603,01 y 2.362 USD (Q4)	-0.03911
## phcalf16_actg\$R_Ingpmayor a 2.362 USD (Q5)	0.03913
## phcalf16_actg\$R_Plazomayor a 20 años	0.02961
## phcalf16_actg\$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE)	0.03281
## phcalf16_actg\$Cicloef21	0.01316
## phcalf16_actg\$tyear	0.00723
## phcalf16_actg\$Cicloef21	-0.08527
## phcalf16_actg\$cesantexciclof1	0.09472
## GLOBAL	NA
##	chisq
## phcalf16_actg\$R_Ingpentre 770,01 y 1.133 USD (Q2)	1.1574
## phcalf16_actg\$R_Ingpentre 1.133,01 y 1.603 USD (Q3)	1.2158
## phcalf16_actg\$R_Ingpentre 1.603,01 y 2.362 USD (Q4)	1.0338
## phcalf16_actg\$R_Ingpmayor a 2.362 USD (Q5)	1.0680
## phcalf16_actg\$R_Plazomayor a 20 años	0.6224
## phcalf16_actg\$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE)	0.7607
## phcalf16_actg\$Cicloef21	0.1245
## phcalf16_actg\$tyear	0.0311
## phcalf16_actg\$Cicloef21	4.3134
## phcalf16_actg\$cesantexciclof1	6.2584
## GLOBAL	14.8211
##	p
## phcalf16_actg\$R_Ingpentre 770,01 y 1.133 USD (Q2)	0.2820
## phcalf16_actg\$R_Ingpentre 1.133,01 y 1.603 USD (Q3)	0.2702
## phcalf16_actg\$R_Ingpentre 1.603,01 y 2.362 USD (Q4)	0.3093
## phcalf16_actg\$R_Ingpmayor a 2.362 USD (Q5)	0.3014
## phcalf16_actg\$R_Plazomayor a 20 años	0.4301
## phcalf16_actg\$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE)	0.3831
## phcalf16_actg\$Cicloef21	0.7242
## phcalf16_actg\$tyear	0.8601
## phcalf16_actg\$Cicloef21	0.0378
## phcalf16_actg\$cesantexciclof1	0.0124
## GLOBAL	0.1387

Corroborando lo expuesto, también se pudo observar gráficamente una distribución relativamente aleatoria de los residuos de Schonenfeld o dfbetas para los coeficientes estimados, con lo cual se apuntaló desde otra perspectiva el cumplimiento de los supuestos de riesgo proporcional.

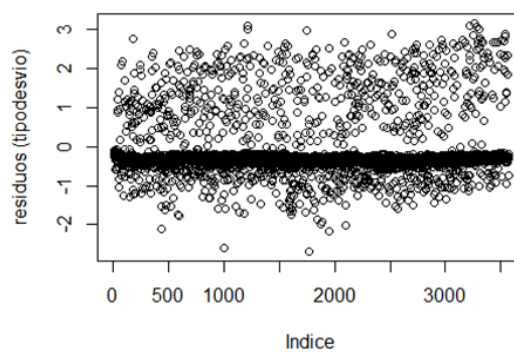




**Figura 23** – Residuos de Schoenfeld o dfbetas para cada uno de los coeficientes de las variables

*Influencia de los casos en la estimación del modelo*

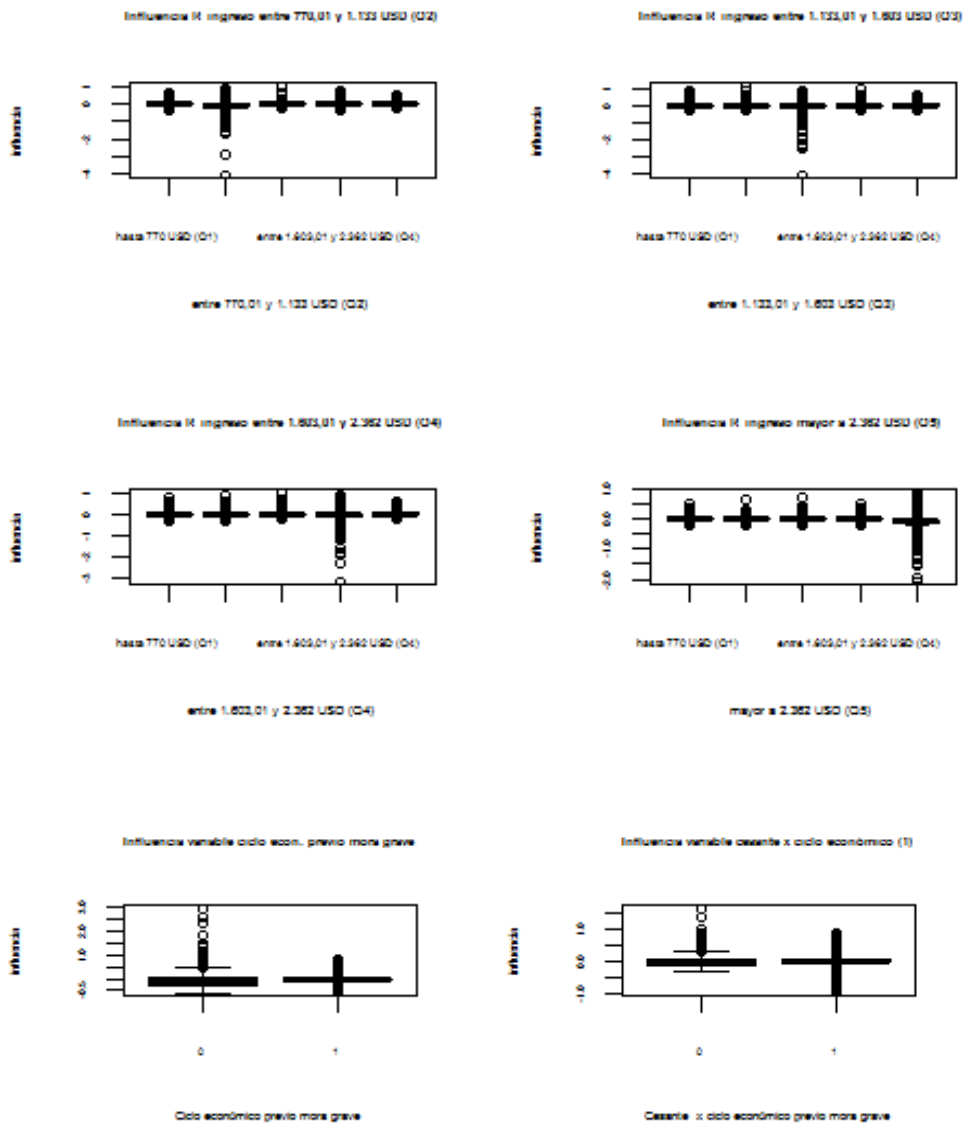
La observación de la distribución de los residuos tipo deviance permitió determinar que no se evidencian casos atípicos que puedan estar influenciando el ajuste del modelo de supervivencia desarrollado.



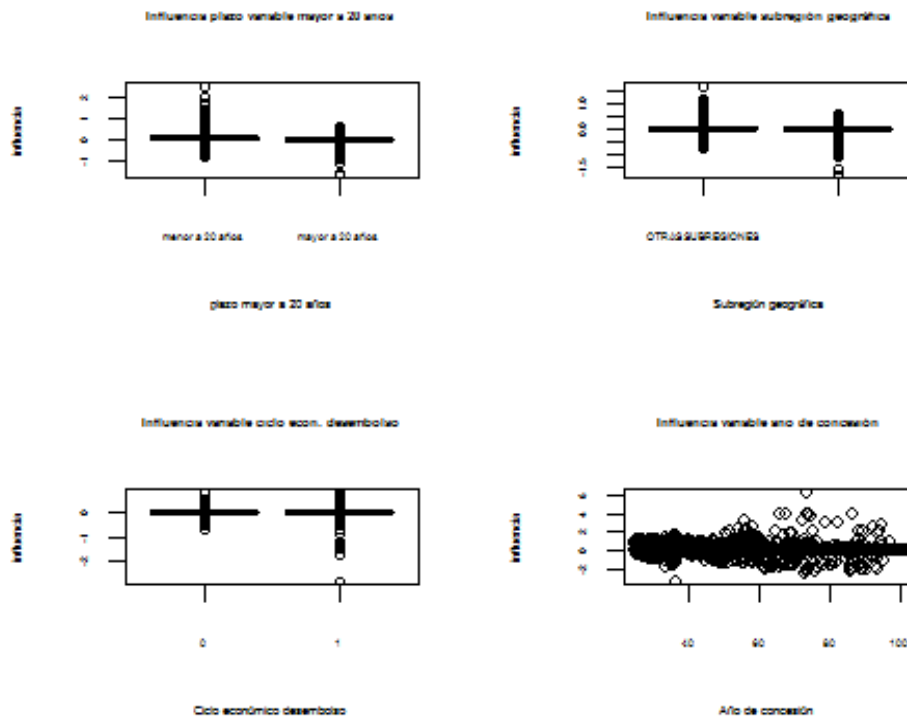
**Figura 24** – Residuos tipo deviance del modelo

Influencia de los casos en la estimación de cada coeficiente

La observación de la distribución de los residuos tipo score permitió determinar que no se evidencian casos atípicos que puedan estar influenciando la estimación de los diferentes coeficientes del modelo.



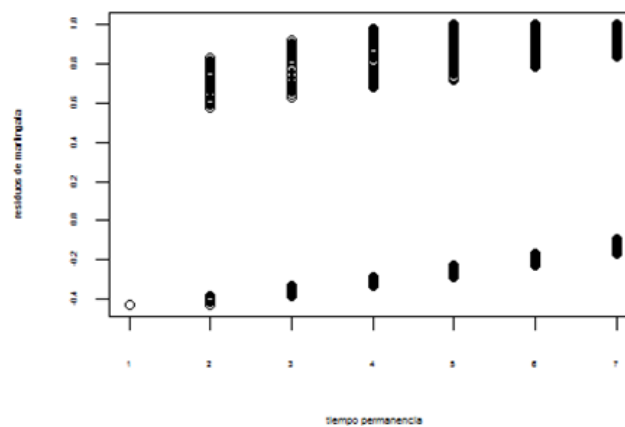




**Figura 25** – Residuos tipo score del modelo

Forma funcional de las variables continuas

Los residuos de martingala permitieron observar que la variable año de concesión, al ser una variable discreta y no propiamente continua mantuvo una distribución particular en la cual se dividió levemente el ajuste lineal de la misma frente al log hazard.



**Figura 26** – Residuos de martingala y variable tiempo de permanencia

### Calibración del modelo frente a un cambio de la variable dependiente

En esta instancia se realizó un pequeño ejercicio adicional, con la finalidad de constatar el cambio que experimenta el modelo de cox formulado frente a la modificación de su variable dependiente. Es decir, si la consideración de una mora grave, que implica una mora mayor a 90 días, se modificara a una marcación diferente, en la cual una mora grave implique una mora mayor a 120 días.

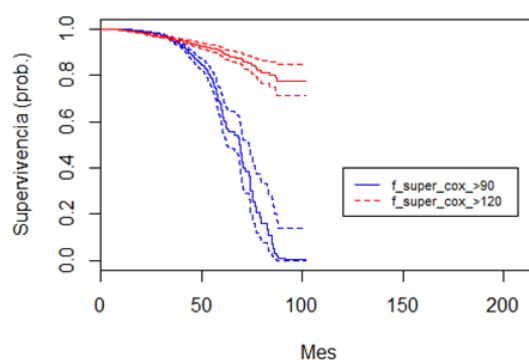
En primera instancia se puede indicar que el nivel de supervivencia en la función resultante del modelo de cox 2 (mora grave mayor a 120 días) es mayor que la constatada en el modelo de cox 1 (mora grave mayor a 90 días), principalmente a partir del mes 40. Específicamente, en el modelo de cox 2 el número de eventos (mora grave) fue de 455, mientras que en el modelo de cox 1 fue de 695. Con lo cual se puede concluir de manera obvia que en el modelo de cox 1 se identifica una mayor cantidad eventos (mora grave) en comparación de la cantidad constatada en el modelo de cox 2.

Adicionalmente, a diferencia del modelo de cox 1, el modelo de cox 2 no evidenció que la variable explicativa “ciclo económico al momento del desembolso resumido” (Ciclo2) fuera estadísticamente significativa de manera individual, no cumplió con los supuestos de riesgo proporcional, y su coeficiente de determinación r-cuadrado fue 0,269, inferior al 0,488 del modelo de cox 1, con lo cual el nivel de explicación de la varianza de los datos fue inferior en el modelo de cox 2.

En consecuencia, para efectos del caso analizado y con el mismo conjunto de variables explicativas en los modelos, el modelo de cox 1 brinda mayores posibilidades de gestionar la problemática de mora grave, principalmente de aplicar estrategias anticipadas que eviten una mayor maduración de cartera. En otras palabras el conjunto de variables explicativas utilizadas en el modelo cox fue relevante para un evento de mora grave mayor a 90 días, no así para un evento de mora grave mayor a 120 días.

Sin embargo, en función de este caso y dado que se requerirían otros elementos para determinarlo, no se podría afirmar que una marcación de mora grave de 90 es mejor que una marcación de mora grave mayor a 120 días, dado que en otro contexto y frente a mayores costos de seguimiento, una marcación de mora con mayor número de días podría ser una buena opción.

El detalle del modelo de cox 2 (mora mayor a 120 días) y su función de supervivencia se incluyeron en el Anexo 2.



**Figura 27** – Función de supervivencia resultante de la aplicación del modelo de cox (mora grave > 90 días) vs Función de supervivencia resultante de la aplicación del modelo de cox (mora grave > 120 días)

### **3.3.2. Segmentación de Clientes y Estrategias**

#### **Segmentación de clientes**

En virtud del análisis de supervivencia realizado para el caso de los clientes de créditos hipotecarios de una entidad financiera ecuatoriana se puede definir algunas tipologías de clientes sobre los cuales se pueden definir algunas estrategias específicas. Las tipologías identificadas fueron las siguientes:

1. Cliente general de crédito hipotecario
2. Cliente de crédito hipotecario, con ingresos mayores a 1,603 USD de ingreso mensual.
3. Cliente de crédito hipotecario, localizado geográficamente en las provincias de la Costa (incluye Santo Domingo y Galápagos) y del Oriente parte norte (Napo, Sucumbíos y Orellana).
4. Clientes de Crédito hipotecario que desempeñan labores en los sectores "EMPRESA UNIPERSONAL / CONSTRUCCIÓN / TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA / SIN INFORMACIÓN"

#### **Estrategias al momento de la concesión crediticia**

Las estrategias que se podrían aplicar al momento de la concesión crediticia en las tipologías de clientes identificadas serían las siguientes:

1. Cliente general de crédito hipotecario: Verificar que se cumplan a cabalidad los requisitos y políticas de crédito al momento de la concesión crediticia, especialmente cuando el entorno económico presenta un desempeño por debajo del crecimiento de largo plazo, dado que la necesidad de generar mayores volúmenes de colocación crediticia hipotecario conduce a otorgar créditos a quienes no necesariamente dan indicios de ser buenos clientes.
2. Cliente de crédito hipotecario, con ingresos mayores a 1,603 USD de ingreso mensual: Verificar con mayor intensidad la permanencia en el tiempo de sus ingresos. Identificación clara de la fuente de ingresos. Verificar que el porcentaje comprometido del ingreso mensual por efecto del endeudamiento hipotecario no supere el 40% del ingreso neto.

3. Cliente de crédito hipotecario, localizado geográficamente en las provincias de la Costa (incluye Santo Domingo y Galápagos) y del Oriente parte norte (Napó, Sucumbíos y Orellana): Verificar indicios del comportamiento de pago pasado. Peor calificación en los últimos 5 años. Aceptar preferiblemente clientes que hayan mantenido calificaciones de crédito A y B, ya que se identifica un perfil del cliente más riesgoso en estas localidades.
  
4. Clientes de Crédito hipotecario que desempeñan labores en los sectores "EMPRESA UNIPERSONAL / CONSTRUCCIÓN / TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA / SIN INFORMACIÓN": Verificar con mayor intensidad la permanencia en el tiempo de sus ingresos. Identificación clara de la fuente de ingresos. Verificar que el porcentaje comprometido del ingreso mensual por efecto del endeudamiento hipotecario no supere el 30% del ingreso neto.

### **Estrategias de gestión de cobranza**

Las estrategias que se podrían aplicar para optimizar la gestión de cobranza en las tipologías de clientes identificadas serían las siguientes:

1. Cliente general de crédito hipotecario: Verificar continuamente, si el cliente incurre en la condición de cesante, particularmente luego del mes 20 y con mayor intensidad a partir del mes 40 luego del otorgamiento del crédito. Recordar mediante mensajes de texto, correo electrónico y llamadas la importancia del repago del crédito. Intensificar la actividad si coyunturalmente la economía atraviesa un desempeño por debajo del crecimiento de largo plazo.
  
2. Cliente de crédito hipotecario, con ingresos mayores a 1,603 USD de ingreso mensual: Verificar continuamente, si el cliente incurre en la condición de cesante, particularmente luego del mes 20 y con mayor intensidad a partir del mes 80 luego del otorgamiento del crédito. Recordar mediante mensajes de texto, correo electrónico y llamadas la importancia del repago del crédito.

3. Cliente de crédito hipotecario, localizado geográficamente en las provincias de la Costa (incluye Santo Domingo y Galápagos) y del Oriente parte norte (Napó, Sucumbíos y Orellana): Verificar continuamente, si el cliente incurre en la condición de cesante, particularmente luego del mes 20 y si se localiza en las provincias orientales del norte. Recordar mediante mensajes de texto, correo electrónico y llamadas la importancia del repago del crédito.
4. Clientes de Crédito hipotecario que desempeñan labores en los sectores "EMPRESA UNIPERSONAL / CONSTRUCCIÓN / TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA / SIN INFORMACIÓN": Una vez que se incorpore al modelo de cox desarrollado variables que denoten el comportamiento de pago pasado, modelar específicamente la probabilidad de supervivencia de casos tipo. En virtud de lo anterior se desarrollaran acciones preventivas como recordatorios de pago y otros principalmente entre los meses 20 y 40 de vigencia de los créditos cuando no se ha producido una mora grave de forma previa. La priorización de los casos se da con el siguiente perfil: ingresos mayores a 1,603 USD, localizados en las provincias de la Costa (incluye Santo Domingo y Galápagos) y del Oriente parte norte (Napó, Sucumbíos y Orellana), con plazo de crédito mayor a 20 años, que se encuentren en la condición de cesante, que fueron concedidos durante un ciclo de la economía por debajo del crecimiento de largo plazo.

En virtud de que la variable cesante previo a la mora grave, y más aún, coincidente con la estancia en un ciclo económica por debajo del crecimiento de largo plazo, resultaron incidentes en la probabilidad de incurrir en una mora grave; sería importante desarrollar un modelo de supervivencia específico frente al evento de un mora grave luego de incurrir en el estado cesante, estratificado por la estancia en un ciclo económico por sobre o por debajo del crecimiento de largo plazo. Es decir, estimar el tiempo que tarda un individuo en incurrir en una mora grave luego de caer en el estado cesante e identificar las probabilidades de supervivencia asociadas. Esta actividad permitiría segmentar la base de clientes y aplicar una estrategia de propuesta anticipada de refinanciamientos y así evitar moras graves.

### **3.4. Automatización del Análisis de Supervivencia: El Caso de una Cartera Hipotecaria Ecuatoriana**

#### **3.4.1. Programación en lenguaje “R”**

El procedimiento metodológico implementado en el presente trabajo con la finalidad de verificar la utilidad de la aplicación del análisis de supervivencia en la cartera de créditos hipotecarios de una entidad ecuatoriana pudo llevarse a cabo mediante la utilización del lenguaje de programación R y la utilización específica de los *software* estadísticos R, R estudio y R markdown.

El tratamiento de la información realizado requirió de 5 etapas:

- Preparación de la data
- Análisis descriptivo de las potenciales variables explicativas
- Identificación de variables explicativas relacionadas en mayor medida con la variable dependiente (árbol de decisión)
- Formulación y validación del modelo de cox
- Creación de reportes

#### **Preparación de la data**

En esta etapa específicamente se realizó la carga inicial de la base datos desde un archivo plano, la definición de formatos adecuados para la información de las variables cargadas y la creación de todas las variables complementarias requeridas en el análisis.

Adicionalmente se utilizó el paquete base instalado en el programa y el paquete “dplyr”, siendo este último uno de aquellos que facilita el tratamiento de información en una base de datos.

#### ***Carga de información***

La carga de información desde un archivo plano con extensión tsv se lo realizó mediante la función “read.table ()”:

```
phcalf16_0 <- read.table("informacionCalificacionMora_21JUN.tsv",header=TRUE,sep = "\t",dec = ",")
```

En esta función, a más de especificar el archivo a leer con su respectiva extensión, se debió especificar mediante el argumento “header” si la primera fila del archivo contiene los nombres de las variables (TRUE) y finalmente, no menos importante, los argumentos “sep” y “dec” que indicaron el elemento mediante el cual se suscita la separación de la data en el archivo (“\t”, tabulaciones) y el signo que indica la presencia de decimales (“,”, coma).

### ***Formato adecuado de variables***

La definición del formato adecuado para las variables contenidas en el archivo leído fue relevante de cara a la utilización posterior de las mismas. Particularmente, cobró relevancia la definición del formato fecha para las variables que presentaron este tipo de información. La adecuación del formato en este caso se dio mediante la función “as.Date”.

```
phcalf16_0$FECHA_DESEMBOLSO<-as.Date(phcalf16_0$FECHA_DESEMBOLSO,format="%d/%m/%Y")
```

En ella, se especificó el nombre de la variable ligada a la base de datos perteneciente y el formato para la presentación de la fecha (“%d/%m/%Y”, día mes y año).

### ***Agrupación y filtro de data***

La agrupación de la data con la finalidad de especificar los casos idóneos para el análisis a realizar así como la aplicación de filtros para la obtención de la base de casos final se lo realizó principalmente mediante la aplicación de las funciones: “arrange”, “group by” y “subset”.

```
phcalf16_0<-arrange(phcalf16_0,NUT,TIPO_BENEF)
phcalf16_00 <- phcalf16_0%>%
group_by(NUT)%>%
summarise(last(NUMAFI),sum(SALDO),.....sum(AVALUOREALIZACION),.....,last(DIAS_MORA_GRAVE))
phcalf16<-subset(phcalf16_00,phcalf16_00$CODTIPSOLSER==35)
```



En este contexto, la función “arrange” permitió ordenar la base de datos según las variables de interés. En virtud de lo anterior, mediante la función “group by” se pretendió agrupar la base de datos por medio de una variable de interés, que para el caso puntual del ejemplo fue la variable “NUT” que hizo referencia al número de operación de crédito. Con la aplicación de esta última, se especificó a continuación, para cada una de las variables, la función de agrupación particular que podría ser suma (“sum()”), el último dato (“last()”), entre otras. Finalmente, la función “subset” permitió crear una sub-base con las observaciones que cumplieron con una característica específica. Específicamente en el ejemplo presentado se seleccionaron aquellos casos que cumplieron con el valor “35” en la variable “CODTIPSOLSER”, misma que hizo referencia a una operación de crédito destinada a la adquisición de vivienda terminada.

### ***Creación de variables complementarias***

La creación de variables complementarias a la base de datos inicial extraída a partir del archiv tsv, fue una necesidad relevante de cara a definir de mejor manera las variables que participaron en el análisis de supervivencia desarrollado. En esta línea, una de las actividades recurrentes fue la recodificación de variables continuas en variables tipo factor.

```
phcalf16$R_Avaluo[phcalf16$AVALUOREALIZACION<=40000] <- "hasta 40.000 USD"
phcalf16$R_Avaluo[phcalf16$AVALUOREALIZACION<=70000&phcalf16$AVALUOREALIZACION>40000] <- "entre 40.001 y 70.000 USD"

phcalf16$R_Avaluo <- as.factor(phcalf16$R_Avaluo)
phcalf16$R_Avaluo <- factor(phcalf16$R_Avaluo,levels = levels(phcalf16$R_Avaluo)[c(4,2,3,1,5)])
```

La implementación de la actividad referida se la hizo mediante la utilización de corchetes, que permitieron especificar la característica puntual de la variable que será sujeta de recodificación. Adicionalmente, mediante la función “as.factor” se definió el formato factor para las variables recodificadas. Finalmente, mediante la función “factor” se especificó las características del factor creado, como el orden de los niveles de la misma.

## **Consideraciones finales**

Es importante mencionar que las actividades desarrolladas durante esta etapa no se espera que posteriormente sean replicadas automáticamente en R ya que si bien en una primera instancia fueron programadas con la finalidad de identificar el procedimiento requerido, estas deberán realizarse a futuro mediante programas de procesamiento de bases de datos y ser entregadas ya incorporadas en un archivo plano para posterior realización del análisis de supervivencia. Obviamente, el código desarrollado en este trabajo será el insumo que de paso al proceso citado. De manera referencial, se adjuntó en el anexo III el código de programación de esta etapa.

## **Análisis descriptivo de las potenciales variables explicativas**

En esta etapa se implementó para cada potencial variable explicativa un análisis descriptivo mediante la utilización de una tabla de contingencia y un análisis de supervivencia individual de la variable mediante método de Kaplan y Meier.

Adicionalmente se utilizaron los paquetes “gmodels” y “survival”, los cuales permitieron la creación de tablas de contingencia y la realización del análisis de supervivencia de cada variable según el método de Kaplan y Meier.

### ***Tabla de contingencia***

La creación de tablas de contingencia permitió analizar la distribución de cada una de las potenciales variables explicativas y contrastarlas con la variable binaria que denotó la presencia o ausencia de la variable dependiente o de interés, que para el análisis planteado fue la presencia de una mora grave. Específicamente, para este fin se utilizó la función “CrossTable” del paquete “gmodels”.

```
table_R_Edad <- CrossTable(phcalf16_m2a$R_Edad,phcalf16_m2a$Mayor90, chisq = TRUE, prop.r = TRUE, prop.c = TRUE, format = "SPSS")
```

En esta función se definieron las variables a contrastar en la tabla de contingencia y adicionalmente se definieron argumentos como la realización de la prueba de independencia de variables ji-cuadrado (“chisq = TRUE”), y la presentación de la proporción con respecto a las columnas (“prop.c = TRUE”) y de la proporción con respecto a las filas (“prop.r = TRUE”).

### ***Análisis de supervivencia por variable – Kaplan y Meier***

La estimación de la función de supervivencia para cada una de las variables frente a un evento de interés utilizando el método de Kaplan y Meier se lo realizó mediante la función “survfit” perteneciente al paquete “survival”.

```
funsuper_R_Edad<-  
survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~phcalf16_m2a$R_Edad)  
print(funsuper_R_Edad)  
print(funsuper_R_Edad)
```

En esta función se especifica la variable dependiente en formato supervivencia mediante la función “surv” dentro de la cual se especificó la variable de tiempo hasta la consecución del evento y la variable de presencia o ausencia del evento de interés. Adicionalmente mediante el comando “~” se asignó la correspondiente variable explicativa. Finalmente, mediante la función “print” se imprimieron el resumen de los resultados de la función de supervivencia para cada opción de respuesta y mediante la función “plot” se graficó la función de supervivencia correspondiente.

```
compsuper_R_Edad<-  
survdif(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~phcalf16_m2a$R_Edad)  
print(compsuper_R_Edad)
```

Complementariamente, se realizó la evaluación de igualdad de funciones de supervivencia entre las opciones de respuesta de una variable mediante el estadístico logrank mediante la utilización de la función “survdif”.

## **Consideraciones finales**

Es importante mencionar que las actividades desarrolladas durante esta etapa si serán sujetas de replicarse automáticamente en R. Es por ello, que en el anexo IV se detalló el código R que puede ser utilizado una vez que se haya cargado y procesado la data del archivo plano inicial.

### **Identificación de variables explicativas relacionadas en mayor medida con la variable dependiente**

En esta etapa se buscó identificar preliminarmente aquellas potenciales variables explicativas que se relacionaron en mayor medida con la variable dependiente o de interés (mora grave). Para ello se hizo uso de la técnica de árbol de decisión.

Los paquetes adicionales a utilizar en esta etapa fueron “rpart” y “rpart.plot”.

```
arbol<- rpart(phcalf16_m2a$Mayor90~phcalf16_m2a$R_Avaluo2+phcalf16_m2a$GENERO+....  
phcalf16_m2a$Acteco_r2+phcalf16_m2a$cesante,method = "class",data = phcalf16_m2a)  
print(arbol)  
  
rpart.plot(arbol,extra = 4,cex = 0.75,cex.main=0.75)
```

Específicamente se utilizó la función “rpart” en la cual se definió la variable dependiente, y mediante el símbolo “~” se asignaron las potenciales variables explicativas. Con ello se creó el correspondiente árbol de decisión. El gráfico del correspondiente árbol de decisión se lo realizó mediante la función “rpart.plot”.

```
printcp(arbol)  
parbol <- prune(arbol,cp=0.010000)
```

Finalmente, mediante la función “printcp” se pudo constatar el error que se suscitó en cada uno de los nodos de clasificación. En el caso de disponer varios nodos de clasificación, mediante la función “prune”, se podría realizar la poda del

árbol en el nodo en el cual el error deja de disminuir, especificando este nodo mediante su correspondiente “cp”.

### ***Consideraciones finales***

Es importante mencionar que las actividades desarrolladas durante esta etapa si serán sujetas de replicarse automáticamente en R. Es por ello, que en el anexo V se detalló el código R que puede ser utilizado una vez que se haya cargado y procesado la data del archivo plano inicial.

### **Formulación y validación del modelo de cox**

En esta etapa se formuló un modelo de cox mediante aquellas variables que resultaron estadísticamente significativas tanto de manera individual como de manera global, se validaron los supuestos de riesgo proporcional y una adecuada distribución de los residuos del modelo.

El paquete “survival” fue el preponderante al momento de formular y validar el correspondiente modelo de cox.

### ***Formulación y validación de variables – modelo de cox***

La formulación del modelo de cox se la realizó mediante la función “coxph”, dentro de la cual se especificó la variable dependiente mediante la función “Surv” que consideró tanto la variable tiempo de supervivencia como la variable presencia del evento de interés o censura del mismo. Así también mediante el símbolo “~” se asignaron las diferentes variables explicativas. La aplicación de la función “summary” al modelo de cox estimado permitió obtener la especificación del modelo planteado, la validación de la significación estadística individual (prueba t) y global de las variables (pruebas: razón de verosimilitud, wald, puntaje).

```
coxsuper_actg<-  
coxph(Surv(phcalf16_actg$tiempoM90pv_2mes,phcalf16_actg$Mayor90)~phcalf16_actg$R_Ingp+ph  
calf16_actg$R_Plazo+.....+phcalf16_actg$cesantexciclof)  
summary(coxsuper_actg)
```

El detalle de la estimación de la función de supervivencia mediante el modelo de cox se dio mediante la función “survfit”. Y el gráfico de la función de supervivencia se dio mediante la aplicación de la función “plot” a la argumento especificado dentro de la función “survfit”.

```
plot(survfit(coxsuper_actg),xlab="Mes",ylab="Supervivencia(prob.)",main="Modelo de Cox(actg)",  
xlim=c(0,120), cex.main=1,ylim=c(0,1))
```

### ***Validación de supuestos de riesgo proporcional***

La validación del supuesto de riesgo proporcional se lo realizaron mediante el estadístico rho que valida el supuesto de riesgo proporcional. La función que evaluó el estadístico tanto a nivel de cada variable como a nivel global fue la denominada “cox.zph”.

```
cox.zph(coxsuper_actg)
```

Complementariamente, se graficó los residuos de Schonenfeld, cuya distribución aleatoria en la gráfica frente a las diferentes variables corroboran o no la hipótesis de riesgos proporcionales. La función “plot” aplicada a la función “cox.zph” del modelo construido y a cada una de las variables del modelo permitieron construir las gráficas señaladas.

```
plot(cox.zph(coxsuper_actg),var=5, main="Plazo mayor a 20 años",cex.main=0.45,cex.lab=0.25)  
plot(cox.zph(coxsuper_actg),var=6, main="Subregión COSTA (incl.STO y GAL.) y ORIENTEN  
(NAP,SUC,ORE)",cex.main=0.45,cex.lab=0.25)
```

### ***Influencia de los casos en la estimación del modelo***

La gráfica de los residuos de los desvíos o deviance permitieron identificar casos atípicos, cuya presencia podría evidencia influencia en la estimación del modelo. La utilización de la función “plot” aplicada a la función “resid” que contiene los residuos deviance del modelo de cox estimado, permitió obtener la gráfica citada.

```
plot(resid(coxsuper_actg,type="deviance"),xlab="Indice",ylab="residuos(tipodesvio)",main="Residuos  
(tipo deviance)_actg")
```

### ***Influencia de los casos en la estimación de cada coeficiente***

La gráfica de los residuos tipo puntaje o score permitieron verificar la presencia de casos atípicos que podrían evidencia influencia en la estimación de cada coeficiente. La utilización de la función “plot” aplicada a la variable cuyo coeficiente se analizó y a la función “resid” que contiene los residuos score del modelo de cox estimado, permitió obtener la gráfica citada.

```
residuo_actg <- resid(coxsuper_actg,type="score")
plot(phcalf16_actg$R_Plazo,residuo_actg[,5],xlab="plazo mayor a 20 años", ylab="influencia",
main="Influencia plazo variable mayor a 20 años",cex.axis=0.4,cex.main=0.45,cex.lab=0.45)
```

### ***Forma funcional de las variables continuas***

Las gráficas de los residuos de martingala permitieron verificar la forma funcional de la variable continua. La función “resid” permitió calcular los residuos de martingala a partir del modelo de cox general (sin variables dependientes). Posteriormente, el gráfico de la variable continua vs los residuos de martingala permitieron verificar la forma funcional de la variable analizada.

```
coxsuper_actg_0 <- coxph(Surv(phcalf16_actg$tiempoM90pv_2mes,phcalf16_actg$Mayor90)~1)
residmart_actg_0 <- resid(coxsuper_actg_0,type = "martingale")
plot(phcalf16_actg$tyear,residmart_actg_0,xlab=" tiempo ingreso",ylab="residuos de martingala",
main="Verificación forma funcional, variable tiempo de ingreso", cex.axis=0.4, cex.main=0.55,
cex.lab=0.45)
```

### ***Consideraciones finales***

Es importante mencionar que las actividades desarrolladas durante esta etapa si serán sujetas de replicarse automáticamente en R. Es por ello, que en el anexo VI se detalló el código R que puede ser utilizado una vez que se haya cargado y procesado la data del archivo plano inicial.

## **Creación de reportes**

En esta etapa se definió un tipo de reporte que pretende condensar información referente al análisis de supervivencia realizado y compartirlo para la toma de decisiones correspondiente. Específicamente, a manera de ejemplo se planteó reportar un resumen de los resultados de supervivencia del modelo de cox aplicado para aquellos clientes que desempeñaron actividades económicas en los sectores ""EMPRESA UNIPERSONAL / CONSTRUCCIÓN / TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA / SIN INFORMACIÓN" y adicionalmente visualizar la diferencia en las curvas de supervivencia entre quienes incurrieron en la condición de cesante en un ciclo económico por debajo del crecimiento de largo plazo, y quienes no.

La creación de este reporte implicó la utilización del programa "R Markdown", complemento de R estudio, mismo que utiliza la programación R y permite crear automáticamente reportes en formato html, Word o pdf.

Específicamente, el resumen del modelo de supervivencia se lo generó mediante la aplicación de la función "summary" al ajuste del modelo cox definido y contenido en la función "survfit". Adicionalmente, dentro de la función "summary" se pudo definir un vector con tiempos de interés sobre los cuales se deseaba visualizar los resultados de la función de supervivencia.

```
super_year_coxsuper_actg <- summary(survfit(coxsuper_actg), c(12,24,36,48,60,72,84,96))
print(super_year_coxsuper_actg)
```

Complementariamente, la graficación de las funciones de supervivencia se la realizó mediante la función "ggsurvplot" que formó parte de la librería "survminer". En ella particularmente se definió dentro de sus principales argumentos el ajuste del modelo de cox a graficar mediante la función "survfit" y la data que se utilizó para general el modelo de cox implicado.

```
ggsurvplot(survfit(coxsuper_actg),data = phcalf16_actg, palette = "#2E9FDF",ggtheme = theme_minimal())
```



De acuerdo a lo expuesto, a continuación se presenta el reporte ejemplo planteado conforme al enfoque detallado previamente e implementado mediante las fórmulas en R analizadas:

**Reporte 1 (ejemplo)**

**Tema:**

Supervivencia a una mora grave (mayor a 90 días) desde la concesión

**Producto:**

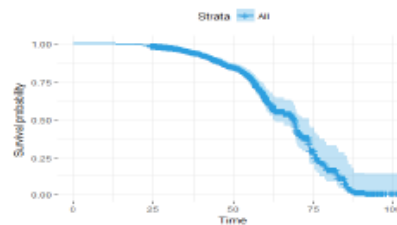
Préstamo Hipotecario

**Población objetivo:**

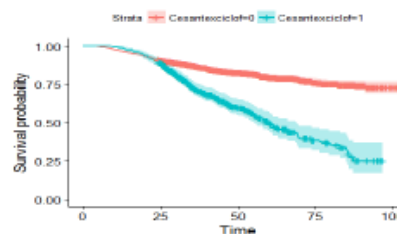
Clientes que desempeñaron actividades económica en los sectores ""EMPRESA UNIPERSONAL / CONSTRUCCIÓN / TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA / SIN INFORMACIÓN"

**Supervivencia general:**

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
12	3469	181	0.99793	0.008327	9.97e-01	0.999
24	3252	217	0.98637	0.001464	9.84e-01	0.989
36	2866	181	0.95849	0.004174	9.42e-01	0.959
48	1242	98	0.86494	0.011015	8.44e-01	0.887
60	704	53	0.64843	0.033411	5.86e-01	0.717
72	381	27	0.38436	0.054680	2.91e-01	0.508
84	181	12	0.10715	0.053689	4.01e-02	0.286
96	25	6	0.00241	0.004973	4.22e-05	0.138



**Supervivencia según clientes que incurrieron en la condición de cesante en un ciclo económico por debajo del crecimiento de largo plazo:**



**Figura 28 - Ejemplo Reporte análisis de supervivencia**

**Consideraciones finales**

Es importante mencionar que las actividades desarrolladas en esta etapa si serán sujetas de replicarse automáticamente en R. Sin embargo, cabe mencionar que exclusivamente es un ejemplo de reporte. En el el anexo VIII se detalló el código R que puede ser utilizado para la creación del reporte ejemplo.

## 4. CONCLUSIONES

La hipótesis planteada al inicio de la presente disertación fue la siguiente:

“La aplicación e implementación de un modelo de supervivencia conllevaría beneficios de cara a la gestión del riesgo de crédito de una cartera hipotecaria ecuatoriana a través del análisis de la temporalidad en la que podrían suscitarse los “no pagos”, de la identificación de las variables que incidirían en el comportamiento referido; y de la implementación de políticas de crédito y gestión que permitan mitigar anticipadamente los riesgos identificados”

En virtud del trabajo realizado se puede afirmar que la hipótesis planteada sí se cumpliría en virtud de las siguientes conclusiones

- Las experiencias de aplicación del análisis supervivencia en diferentes ámbitos y particularmente en el ámbito financiero denotaron que la incorporación de este análisis mejora el realizado con otras técnicas al momento de predecir la probabilidad de incumplimiento (default) como los modelos logísticos, en virtud de que tiene la capacidad de incorporar datos censurados (información perdida en otras técnicas). Adicionalmente, existen ejercicios de comparación de técnicas que denotaron que la transformación de logrank (modelo de supervivencia de cox) supera a la transformación logística y otros, en cuanto a la predicción de la probabilidad de supervivencia, principalmente basados en la prueba estadística de Wald (Man, 2014).
- Adicionalmente, las experiencias de aplicación relevaron que la identificación de la temporalidad en la cual se puede incurrir en incumplimiento (default) es un elemento importante de cara a mitigar el riesgo de crédito, para lo cual el típico análisis de matrices de transición es insuficiente. Así también se vislumbró la necesidad de incorporar a este tipo de modelos variables asociadas al entorno económico y variables microeconómicas específicas de los deudores (Cáceres & Palacios, 2017).
- Así también en la bibliografía empírica analizada se pudo constatar que la censura da lugar a errores en las estimaciones de los parámetros para todas las metodologías (Watkins, Vasnev, & Gerlach, 2009). Es por ello, que en

algunos casos se tomaron sub-bases de datos de la base general (créditos que no cumplen con una maduración mínima) con la finalidad de evitar casos censurados innecesarios (Obuda, 2014).

- Empíricamente, la aplicación del análisis de supervivencia a una base de datos de clientes del producto préstamo hipotecario en una entidad financiera ecuatoriana, permitió identificar los siguientes comportamientos con respecto a la incursión del cliente en una mora grave (mayor a 90 días):
  - La condición de cesante de un cliente, sumada a la estancia temporal en un ciclo económico que se ubica por debajo del crecimiento de largo plazo, se constituye en la condición más relevante identificada de cara a incurrir con mayor probabilidad en una mora grave.
  - Los clientes localizados en las provincias de la Costa (incluye Santo Domingo y Galápagos) y del Oriente parte norte (Napo, Sucumbíos y Orellana) son más propensos a caer en una mora grave que aquellos dispuestos en otros lugares.
  - A mayor ingreso mensual de los clientes, mayor probabilidad de incurrir en el evento de mora grave.
  - Los clientes que desempeñan labores en los sectores "EMPRESA UNIPERSONAL / CONSTRUCCIÓN / TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA / SIN INFORMACIÓN" son más propensos a caer en una mora grave que aquellos dispuestos en otros lugares.
- Así también, permitió identificar, de manera general que los clientes de crédito hipotecario de la entidad financiera analizada tienden a incrementar su probabilidad de caer en una mora grave a partir del mes 20 y con mayor intensidad a partir del mes 40.
- Los hallazgos especificados permitieron definir potenciales estrategias de concesión y gestión de cobranza para las tipologías de clientes identificados. Es por ello que uno de los aportes importantes del presente trabajo es haber definido y probado un procedimiento metodológico que permite mejorar la mitigación del riesgo de crédito. Específicamente, el procedimiento implementado constó de las siguientes actividades secuenciales: a) definición del evento de interés (mora grava) y con ello definir la variable de interés

correspondiente, b) breve descripción de la base de datos a utilizar, c) análisis descriptivo de las variables consideradas inicialmente relevantes y construcción individual de funciones de supervivencia mediante el método de Kaplan y Meier, d) aplicación del análisis de decisión (árbol de decisión) para identificar variables relacionadas con mayor intensidad al evento de interés, e) seleccionar la data y las variables que permitirán definir un modelo de cox representativo, f) formulación, validación e interpretación del modelo de cox, y g) segmentación de la base de clientes de la cartera de créditos hipotecarios, a partir del análisis de supervivencia desarrollado y plantear estrategias diferenciadas tanto de concesión como de gestión de cobranzas.

- Adicionalmente, el análisis de supervivencia realizado fue programado en lenguaje R con lo cual puede ser fácilmente replicado.

Adicionalmente, algunas de las principales recomendaciones serían las siguientes:

- Con la finalidad de mejorar los resultados predictivos del modelo de Cox desarrollado se recomienda la incorporación de variables que perfilen el comportamiento de crédito pasado del cliente en un horizonte de hasta 5 años.
- Medir periódicamente cada 3 o 6 meses la variable presencia de cesantía con la finalidad de construir un modelo de cox entendido, en el cual se incorporan variables que dependen del tiempo.
- En virtud de que la variable cesante previo a la mora grave, y más aún, coincidente con la estancia en un ciclo económica por debajo del crecimiento de largo plazo, resultaron incidentes en la probabilidad de incurrir en una mora grave; sería importante para la entidad oferente del crédito hipotecario desarrollar un modelo de supervivencia específico frente al evento de un mora grave luego de incurrir en el estado cesante, estratificado por la estancia en un ciclo económico por sobre o por debajo del crecimiento de largo plazo. Es decir, estimar el tiempo que tarda un individuo en incurrir en una mora grave luego de caer en el estado cesante e identificar las probabilidades de supervivencia asociadas. Esta actividad permitiría segmentar la base de clientes y aplicar una estrategia de propuesta anticipada de refinanciamientos y así evitar moras graves.

## REFERENCIAS BIBLIOGRÁFICAS

- Almeida Baroja, E. (2011). *Aplicación del Modelo de Supervivencia de Cox al caso de la Banca Ecuatoriana en el período 1996-2008*. Quito: Escuela Politécnica Nacional.
- Altina, S., & López, J. (2001). *Gestión Bancaria: Los nuevos retos en un entorno global*. McGraw-Hill / Interamericana de España. S.A.
- Ayala, M. A., Borges, R. E., & Colmenares, G. (junio de 2007). Análisis de Supervivencia aplicado a la Banca Comercial Venezolana, 1996 - 2004. (U. N. Colombia, Ed.) *Revista Colombiana de Estadística*, 30(1), 97-113.
- Banco Central del Ecuador (a). (2016). *Resultados del Ciclo Económico del Ecuador, Indicadores del Ciclo del PIB, coincidente y Adelantado, Información al II trimestre 2016*. Obtenido de <https://www.bce.fin.ec/index.php/component/k2/item/764>
- Banco Central del Ecuador (b). (2016). *Boletín de Cuentas Nacionales Trimestrales del Ecuador, Resultados 2000.1 – 2016.II, Precios Corrientes y Constantes, (96)*. Obtenido de <https://www.bce.fin.ec/index.php/component/k2/item/756>
- Banco Central del Ecuador (c). (1 de 2017). *Base - Monitoreo de los Principales Indicadores Monetarios y Financieros de la Economía Ecuatoriana*. Recuperado el 31 de 1 de 2017, de [https://contenido.bce.fin.ec/documentos/PublicacionesNotas/Presentacion\\_ene17.pdf](https://contenido.bce.fin.ec/documentos/PublicacionesNotas/Presentacion_ene17.pdf)
- Banco Central del Ecuador (d). (agosto de 2016). *Previsiones Macroeconómicas al año 2016, Resultados a precios corrientes y constantes, base 2007*.
- Banco Central del Ecuador (e). (noviembre de 2016). *Estudio Mensual de Opinión Empresarial*. Obtenido de <https://www.bce.fin.ec/index.php/component/k2/item/756>
- Barrera Rebellon, M. (2008). *Análisis de supervivencia aplicado al problema de la deserción estudiantil en la Universidad Tecnológica de Pereira*. Pereira: Universidad Tecnológica de Pereira.
- Barrios Pérez, V. (2004). ¿Por qué existen los bancos? *Boletín económico de ICE*, N° 2799, 34.

- Bench Colombia. (s.f.). *Una revisión de la literatura sobre el vínculo entre sistema financiero y crecimiento económico, Capítulo 1*. Obtenido de <http://bdigital.eafit.edu.co/bdigital/PROYECTO/P330.09861C957/capitulo1.pdf>
- Cáceres, F., & Palacios, Y. (2017). Análisis de supervivencia como alternativa metodológica para estimar probabilidades de incumplimiento de los deudores de créditos corporativos y a grandes empresas en el Perú. *Revista Industrial Data* 20(1): 07-16 (2017).
- Cao, R., Vilar, J., & Devia, A. (2009). *Modelling consumer credit risk via survival analysis*. Coruña, España: Universidade da Coruña.
- Conesa, J. C. (agosto de 2006). *Teoría Economía del Capital y la Renta*. Obtenido de <http://pareto.uab.es/jconesa/libro/?N=D>
- De Gregorio, J., & Guidotti, P. (1996). Desarrollo Financiero y Crecimiento Económico. *Revista de Economía Política: Pensamiento Iberoamericano*, 369.
- Godoy Aguilar, A. M. (2009). *Introducción al Análisis de Supervivencia con R*. Ciudad de México: Universidad Nacional Autónoma de México.
- González, A., & López, J. (2008). *Gestión Bancaria. Factores claves en un entorno competitivo*. Madrid, España: McGraw-Hill / Interamericana de España. S.A.U.
- Hernández Domínguez, A. M. (2010). *Análisis Estadísticos de Datos de Tiempos de Fallo en R*. Granada: Universidad de Granada.
- Herranz Valera, J. (2015). *Introducción al Análisis de Supervivencia con R*. Madrid: Instituto IMDEA Alimentación.
- Hofacker, A. (2008). *Rapid lean construction - quality rating model*. Manchester: s.n.
- Instituto Nacional de Estadísticas y Censos (a). (2016). *Tabulados Encuesta Nacional de Empleo, Subempleo y Desempleo (ENEMDU, (15 años y más), resultados septiembre 2016*.
- klein, J. P., & Moeschberger, M. L. (1997). *Survival Analysis Techniques for Censored and Truncated Data*. Springer-Verlag New York: primera edición.
- López, A. (2003). *Análisis de la Relación entre intermediación crediticia y crecimiento económico en Venezuela*. Banco Central de Venezuela.
- Man, R. (2014). *Survival analysis in credit scoring: A framework for PD estimation*. Rabobank International – Quantitative Risk Analytics & University of Twente.
- Meira-Machado, L. (2016). *Introduction to Survival Analysis*. Santiago de Compostella: Universidad Santiago de Compostela.

- Meira-Machado, L. (2016). *Survival Analysis: Kaplan-Meier estimate*. Santiago de Compostela: Universidad Santiago de Compostela.
- Obuda, F. Y. (2014). *Analysis of Credit Risk on Bank Loans using Cox's Proportional Hazard Model*. Nairobi: Universidad de Nairobi.
- Pruenza García-Hinojosa, C. (2014). *Estudio de Análisis de Supervivencia*. Madrid: Universidad Autónoma de Madrid, Escuela Politécnica Superior.
- Rebasa, P. (2005). Conceptos básicos del análisis de supervivencia. *Revista de cirugía española*(Artículo Especial, 78 (4), 36), 224,225.
- Romero, B., & Meléndez, A. (2016). El Sector de la Construcción no recobra el pulso. *Gestión*, 17.
- Samartín, M. (Mayo de 2004). *Algunos temas relevantes en la Teoría Bancaria*. Obtenido de <http://e-archivo.uc3m.es:8080/dspace/handle/10016/28>
- Santabárbara, J., Rubio, E., & Martínez, T. (2016). *Manual de Análisis de Supervivencia: Curvas de supervivencia y Regresión de Cox*. Zaragoza: Prensas de la Universidad de Zaragoza.
- Superintendencia de Bancos (a). (s.f.). *Volumen de concesión, datos enero – octubre de 2016*.
- Universidad Carlos III. (2008). *Análisis de Cluster y Árboles de Clasificación*. Obtenido de [halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema6dm.pdf](http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema6dm.pdf)
- Velandia, N. (2013). *Establecimiento de un Modelo Logit para la Medición del Riesgo de Incumplimiento en Créditos para una Entidad Financiera del Municipio de Arauca (tesis de postgrado)*. Arauca, Colombia: Universidad Nacional de Colombia.
- Vicéns Otero, J., & Medina Moral, E. (enero de 2005). *Análisis de Datos Cualitativos*. Obtenido de [https://www.google.com.ec/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&cad=rja&uact=8&ved=0ahUKEwjgouL319zXAhXEQ98KHVZyD4oQFghDMAU&url=https%3A%2F%2Fwww.uam.es%2Fpersonal\\_pdi%2Feconomicas%2Feva%2Fpdf%2Ftab\\_conting.pdf&usg=AOvVaw3omGoM-dQqWVyxIuFVYgKr](https://www.google.com.ec/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&cad=rja&uact=8&ved=0ahUKEwjgouL319zXAhXEQ98KHVZyD4oQFghDMAU&url=https%3A%2F%2Fwww.uam.es%2Fpersonal_pdi%2Feconomicas%2Feva%2Fpdf%2Ftab_conting.pdf&usg=AOvVaw3omGoM-dQqWVyxIuFVYgKr)
- Watkins, J., Vasnev, A., & Gerlach, R. (2009). *Survival Analysis for Credit Scoring: Incidence and Latency*. Madrid: OME Working Paper Series Madrid.
- Zhang, A. (2009). *Statistical Methods in Credit Risk Modeling (tesis de doctorado)*. Michigan, Estados Unidos de Norteamérica: The University of Michigan. .

## ANEXOS

**Anexo I** – Detalle de la función de supervivencia obtenida mediante la aplicación del modelo de cox propuesto

**Tabla 39** – Función de supervivencia obtenida mediante la aplicación del modelo de cox propuesto

```
## Call: survfit(formula = coxsuper_actg)
##
##   time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##   4.17  3570     1  0.99999  1.30e-05  1.00e+00  1.000
##   4.20  3569     2  0.99996  2.31e-05  1.00e+00  1.000
##   4.27  3567     1  0.99995  2.70e-05  1.00e+00  1.000
##   4.53  3566     1  0.99993  3.05e-05  1.00e+00  1.000
##   4.60  3565     1  0.99992  3.38e-05  1.00e+00  1.000
##   4.97  3564     1  0.99991  3.70e-05  1.00e+00  1.000
##   5.10  3563     1  0.99989  4.01e-05  1.00e+00  1.000
##   5.13  3562     2  0.99987  4.59e-05  1.00e+00  1.000
##   5.17  3560     1  0.99985  4.86e-05  1.00e+00  1.000
##   5.23  3559     2  0.99982  5.41e-05  1.00e+00  1.000
##   5.33  3557     1  0.99981  5.68e-05  1.00e+00  1.000
##   5.57  3556     1  0.99980  5.96e-05  1.00e+00  1.000
##   5.60  3555     1  0.99978  6.23e-05  1.00e+00  1.000
##   5.63  3554     1  0.99977  6.49e-05  1.00e+00  1.000
##   5.67  3553     1  0.99975  6.76e-05  1.00e+00  1.000
##   5.90  3552     1  0.99974  7.03e-05  1.00e+00  1.000
##   6.07  3551     1  0.99972  7.31e-05  1.00e+00  1.000
##   6.10  3550     1  0.99970  7.58e-05  1.00e+00  1.000
##   6.30  3549     1  0.99969  7.85e-05  1.00e+00  1.000
##   6.37  3548     1  0.99967  8.11e-05  1.00e+00  1.000
##   6.40  3547     1  0.99966  8.38e-05  9.99e-01  1.000
##   7.03  3546     2  0.99962  8.91e-05  9.99e-01  1.000
##   7.07  3544     1  0.99961  9.17e-05  9.99e-01  1.000
##   7.13  3543     1  0.99959  9.42e-05  9.99e-01  1.000
##   7.40  3542     1  0.99958  9.68e-05  9.99e-01  1.000
##   7.53  3541     1  0.99956  9.94e-05  9.99e-01  1.000
##   7.57  3540     1  0.99954  1.02e-04  9.99e-01  1.000
##   7.60  3539     1  0.99953  1.05e-04  9.99e-01  1.000
##   7.67  3538     1  0.99951  1.07e-04  9.99e-01  1.000
##   7.70  3537     1  0.99949  1.10e-04  9.99e-01  1.000
##   7.83  3536     2  0.99946  1.15e-04  9.99e-01  1.000
##   8.00  3534     1  0.99944  1.18e-04  9.99e-01  1.000
##   8.10  3533     1  0.99942  1.21e-04  9.99e-01  1.000
##   8.17  3532     1  0.99941  1.23e-04  9.99e-01  1.000
##   8.20  3531     1  0.99939  1.26e-04  9.99e-01  1.000
##   8.30  3530     1  0.99937  1.29e-04  9.99e-01  1.000
##   8.33  3529     1  0.99935  1.32e-04  9.99e-01  1.000
##   8.47  3528     1  0.99933  1.34e-04  9.99e-01  1.000
##   8.50  3527     1  0.99931  1.37e-04  9.99e-01  1.000
##   8.57  3526     1  0.99930  1.40e-04  9.99e-01  1.000
##   8.60  3525     1  0.99928  1.43e-04  9.99e-01  1.000
##   8.70  3524     1  0.99926  1.45e-04  9.99e-01  1.000
##   8.73  3523     1  0.99924  1.48e-04  9.99e-01  1.000
##   8.83  3522     1  0.99922  1.51e-04  9.99e-01  1.000
##   8.90  3521     1  0.99920  1.54e-04  9.99e-01  1.000
##   8.97  3520     1  0.99918  1.57e-04  9.99e-01  0.999
##   9.03  3519     1  0.99916  1.60e-04  9.99e-01  0.999
##   9.17  3518     1  0.99914  1.62e-04  9.99e-01  0.999
##   9.20  3517     3  0.99908  1.71e-04  9.99e-01  0.999
##   9.27  3514     1  0.99906  1.74e-04  9.99e-01  0.999
##   9.30  3513     1  0.99904  1.78e-04  9.99e-01  0.999
##   9.33  3512     1  0.99902  1.81e-04  9.99e-01  0.999
```



##	9.40	3511	1	0.99900	1.84e-04	9.99e-01	0.999
##	9.47	3510	1	0.99897	1.87e-04	9.99e-01	0.999
##	9.50	3509	1	0.99895	1.90e-04	9.99e-01	0.999
##	9.53	3508	3	0.99889	1.99e-04	9.98e-01	0.999
##	9.60	3505	1	0.99886	2.02e-04	9.98e-01	0.999
##	9.70	3504	2	0.99882	2.09e-04	9.98e-01	0.999
##	9.80	3502	1	0.99879	2.12e-04	9.98e-01	0.999
##	9.83	3501	1	0.99877	2.15e-04	9.98e-01	0.999
##	9.87	3500	1	0.99875	2.18e-04	9.98e-01	0.999
##	10.10	3499	1	0.99872	2.22e-04	9.98e-01	0.999
##	10.20	3498	1	0.99870	2.25e-04	9.98e-01	0.999
##	10.23	3497	1	0.99868	2.28e-04	9.98e-01	0.999
##	10.33	3496	1	0.99865	2.31e-04	9.98e-01	0.999
##	10.37	3495	2	0.99860	2.38e-04	9.98e-01	0.999
##	10.40	3493	3	0.99853	2.48e-04	9.98e-01	0.999
##	10.70	3490	1	0.99850	2.52e-04	9.98e-01	0.999
##	10.77	3489	1	0.99848	2.55e-04	9.98e-01	0.999
##	10.80	3488	1	0.99845	2.59e-04	9.98e-01	0.999
##	10.83	3487	1	0.99842	2.62e-04	9.98e-01	0.999
##	10.87	3486	2	0.99837	2.69e-04	9.98e-01	0.999
##	11.13	3484	1	0.99834	2.73e-04	9.98e-01	0.999
##	11.17	3483	1	0.99832	2.76e-04	9.98e-01	0.999
##	11.20	3482	1	0.99829	2.80e-04	9.98e-01	0.999
##	11.30	3481	3	0.99820	2.91e-04	9.98e-01	0.999
##	11.40	3478	1	0.99817	2.95e-04	9.98e-01	0.999
##	11.53	3477	2	0.99811	3.03e-04	9.98e-01	0.999
##	11.57	3475	1	0.99808	3.07e-04	9.97e-01	0.999
##	11.60	3474	2	0.99802	3.15e-04	9.97e-01	0.999
##	11.73	3472	1	0.99799	3.19e-04	9.97e-01	0.999
##	11.87	3471	1	0.99796	3.23e-04	9.97e-01	0.999
##	11.97	3470	1	0.99793	3.27e-04	9.97e-01	0.999
##	12.03	3469	2	0.99786	3.34e-04	9.97e-01	0.999
##	12.17	3467	1	0.99783	3.38e-04	9.97e-01	0.998
##	12.20	3466	1	0.99780	3.42e-04	9.97e-01	0.998
##	12.23	3465	1	0.99777	3.46e-04	9.97e-01	0.998
##	12.30	3464	1	0.99773	3.51e-04	9.97e-01	0.998
##	12.37	3463	1	0.99770	3.55e-04	9.97e-01	0.998
##	12.53	3462	1	0.99767	3.59e-04	9.97e-01	0.998
##	12.63	3461	2	0.99760	3.67e-04	9.97e-01	0.998
##	12.67	3459	1	0.99757	3.71e-04	9.97e-01	0.998
##	12.70	3458	1	0.99754	3.75e-04	9.97e-01	0.998
##	12.80	3457	1	0.99750	3.79e-04	9.97e-01	0.998
##	12.83	3456	1	0.99747	3.83e-04	9.97e-01	0.998
##	13.00	3455	2	0.99740	3.91e-04	9.97e-01	0.998
##	13.03	3453	1	0.99737	3.95e-04	9.97e-01	0.998
##	13.10	3452	1	0.99733	3.99e-04	9.97e-01	0.998
##	13.23	3451	1	0.99730	4.04e-04	9.97e-01	0.998
##	13.27	3450	1	0.99726	4.08e-04	9.96e-01	0.998
##	13.33	3449	2	0.99719	4.16e-04	9.96e-01	0.998
##	13.37	3447	2	0.99712	4.25e-04	9.96e-01	0.998
##	13.47	3445	1	0.99709	4.29e-04	9.96e-01	0.998
##	13.50	3444	1	0.99705	4.33e-04	9.96e-01	0.998
##	13.53	3443	2	0.99698	4.41e-04	9.96e-01	0.998
##	13.57	3441	2	0.99691	4.50e-04	9.96e-01	0.998
##	13.70	3439	2	0.99683	4.58e-04	9.96e-01	0.998
##	13.80	3437	1	0.99680	4.62e-04	9.96e-01	0.998
##	13.90	3436	3	0.99669	4.75e-04	9.96e-01	0.998
##	13.93	3433	1	0.99665	4.79e-04	9.96e-01	0.998
##	13.97	3432	1	0.99661	4.83e-04	9.96e-01	0.998
##	14.17	3431	1	0.99657	4.88e-04	9.96e-01	0.998
##	14.30	3430	1	0.99654	4.92e-04	9.96e-01	0.998
##	14.37	3429	2	0.99646	5.01e-04	9.95e-01	0.997
##	14.53	3427	1	0.99642	5.05e-04	9.95e-01	0.997
##	14.60	3426	1	0.99638	5.10e-04	9.95e-01	0.997
##	14.83	3425	1	0.99634	5.14e-04	9.95e-01	0.997
##	15.00	3424	1	0.99630	5.19e-04	9.95e-01	0.997
##	15.03	3423	1	0.99626	5.23e-04	9.95e-01	0.997
##	15.07	3422	2	0.99618	5.32e-04	9.95e-01	0.997
##	15.17	3420	1	0.99614	5.36e-04	9.95e-01	0.997
##	15.20	3419	1	0.99610	5.40e-04	9.95e-01	0.997
##	15.23	3418	1	0.99606	5.45e-04	9.95e-01	0.997
##	15.27	3417	1	0.99602	5.49e-04	9.95e-01	0.997

##	15.33	3416	1	0.99598	5.54e-04	9.95e-01	0.997
##	15.37	3415	1	0.99594	5.58e-04	9.95e-01	0.997
##	15.43	3414	1	0.99590	5.63e-04	9.95e-01	0.997
##	15.50	3413	1	0.99586	5.67e-04	9.95e-01	0.997
##	15.57	3412	1	0.99582	5.72e-04	9.95e-01	0.997
##	15.60	3411	2	0.99573	5.81e-04	9.95e-01	0.997
##	15.63	3409	2	0.99565	5.90e-04	9.94e-01	0.997
##	15.67	3407	2	0.99557	5.99e-04	9.94e-01	0.997
##	15.70	3405	1	0.99552	6.04e-04	9.94e-01	0.997
##	15.73	3404	2	0.99544	6.13e-04	9.94e-01	0.997
##	15.83	3402	1	0.99540	6.17e-04	9.94e-01	0.997
##	15.87	3401	1	0.99535	6.22e-04	9.94e-01	0.997
##	15.90	3400	1	0.99531	6.26e-04	9.94e-01	0.997
##	15.97	3399	1	0.99527	6.31e-04	9.94e-01	0.997
##	16.00	3398	1	0.99522	6.36e-04	9.94e-01	0.996
##	16.07	3397	1	0.99518	6.40e-04	9.94e-01	0.996
##	16.10	3396	2	0.99509	6.50e-04	9.94e-01	0.996
##	16.30	3394	2	0.99501	6.59e-04	9.94e-01	0.996
##	16.37	3392	1	0.99496	6.63e-04	9.94e-01	0.996
##	16.70	3391	1	0.99492	6.68e-04	9.94e-01	0.996
##	16.77	3390	1	0.99488	6.73e-04	9.94e-01	0.996
##	16.80	3389	1	0.99483	6.77e-04	9.94e-01	0.996
##	17.30	3388	1	0.99479	6.82e-04	9.93e-01	0.996
##	17.37	3387	4	0.99461	7.01e-04	9.93e-01	0.996
##	17.43	3383	1	0.99456	7.06e-04	9.93e-01	0.996
##	17.47	3382	3	0.99442	7.20e-04	9.93e-01	0.996
##	17.63	3379	1	0.99437	7.25e-04	9.93e-01	0.996
##	17.67	3378	1	0.99433	7.30e-04	9.93e-01	0.996
##	17.70	3377	1	0.99428	7.35e-04	9.93e-01	0.996
##	17.80	3376	1	0.99423	7.39e-04	9.93e-01	0.996
##	17.83	3375	2	0.99414	7.49e-04	9.93e-01	0.996
##	17.90	3373	1	0.99409	7.54e-04	9.93e-01	0.996
##	18.00	3372	1	0.99404	7.59e-04	9.93e-01	0.996
##	18.30	3371	2	0.99395	7.69e-04	9.92e-01	0.995
##	18.37	3369	1	0.99390	7.74e-04	9.92e-01	0.995
##	18.40	3368	1	0.99385	7.79e-04	9.92e-01	0.995
##	18.47	3367	1	0.99380	7.84e-04	9.92e-01	0.995
##	18.53	3366	1	0.99375	7.89e-04	9.92e-01	0.995
##	18.67	3365	2	0.99365	7.99e-04	9.92e-01	0.995
##	18.70	3363	2	0.99355	8.09e-04	9.92e-01	0.995
##	18.83	3361	1	0.99350	8.14e-04	9.92e-01	0.995
##	18.93	3360	1	0.99345	8.19e-04	9.92e-01	0.995
##	19.00	3359	1	0.99339	8.25e-04	9.92e-01	0.995
##	19.03	3358	2	0.99329	8.35e-04	9.92e-01	0.995
##	19.07	3356	1	0.99323	8.40e-04	9.92e-01	0.995
##	19.17	3355	1	0.99318	8.46e-04	9.92e-01	0.995
##	19.23	3354	1	0.99313	8.51e-04	9.91e-01	0.995
##	19.27	3353	1	0.99307	8.56e-04	9.91e-01	0.995
##	19.33	3352	1	0.99302	8.61e-04	9.91e-01	0.995
##	19.37	3351	1	0.99297	8.67e-04	9.91e-01	0.995
##	19.40	3350	1	0.99291	8.72e-04	9.91e-01	0.995
##	19.47	3349	2	0.99281	8.82e-04	9.91e-01	0.995
##	19.57	3347	3	0.99264	8.98e-04	9.91e-01	0.994
##	19.60	3344	2	0.99253	9.09e-04	9.91e-01	0.994
##	19.70	3342	1	0.99248	9.14e-04	9.91e-01	0.994
##	19.73	3341	1	0.99242	9.20e-04	9.91e-01	0.994
##	19.83	3340	1	0.99237	9.25e-04	9.91e-01	0.994
##	19.87	3339	1	0.99231	9.31e-04	9.90e-01	0.994
##	19.90	3338	3	0.99214	9.47e-04	9.90e-01	0.994
##	19.93	3335	1	0.99209	9.52e-04	9.90e-01	0.994
##	19.97	3334	3	0.99191	9.69e-04	9.90e-01	0.994
##	20.00	3331	1	0.99186	9.74e-04	9.90e-01	0.994
##	20.07	3330	1	0.99180	9.80e-04	9.90e-01	0.994
##	20.33	3329	1	0.99174	9.86e-04	9.90e-01	0.994
##	20.37	3328	2	0.99162	9.97e-04	9.90e-01	0.994
##	20.40	3326	1	0.99156	1.00e-03	9.90e-01	0.994
##	20.60	3325	1	0.99150	1.01e-03	9.90e-01	0.993
##	20.67	3324	2	0.99137	1.02e-03	9.89e-01	0.993
##	20.77	3322	2	0.99125	1.03e-03	9.89e-01	0.993
##	20.80	3320	4	0.99100	1.06e-03	9.89e-01	0.993
##	20.83	3316	1	0.99093	1.06e-03	9.89e-01	0.993
##	20.87	3315	1	0.99087	1.07e-03	9.89e-01	0.993

##	20.90	3314	1	0.99080	1.07e-03	9.89e-01	0.993
##	20.93	3313	1	0.99074	1.08e-03	9.89e-01	0.993
##	20.97	3312	1	0.99067	1.09e-03	9.89e-01	0.993
##	21.07	3311	2	0.99054	1.10e-03	9.88e-01	0.993
##	21.20	3309	1	0.99047	1.10e-03	9.88e-01	0.993
##	21.23	3308	1	0.99040	1.11e-03	9.88e-01	0.993
##	21.43	3307	1	0.99033	1.12e-03	9.88e-01	0.993
##	21.47	3306	1	0.99027	1.12e-03	9.88e-01	0.992
##	21.50	3305	1	0.99020	1.13e-03	9.88e-01	0.992
##	21.53	3304	1	0.99013	1.14e-03	9.88e-01	0.992
##	21.57	3303	1	0.99006	1.14e-03	9.88e-01	0.992
##	21.60	3302	2	0.98992	1.15e-03	9.88e-01	0.992
##	21.67	3300	1	0.98986	1.16e-03	9.88e-01	0.992
##	21.70	3299	1	0.98979	1.17e-03	9.88e-01	0.992
##	21.77	3298	1	0.98972	1.17e-03	9.87e-01	0.992
##	21.80	3297	2	0.98958	1.19e-03	9.87e-01	0.992
##	21.83	3295	1	0.98950	1.19e-03	9.87e-01	0.992
##	21.87	3294	1	0.98943	1.20e-03	9.87e-01	0.992
##	21.90	3293	1	0.98936	1.20e-03	9.87e-01	0.992
##	21.93	3292	1	0.98929	1.21e-03	9.87e-01	0.992
##	22.13	3291	1	0.98922	1.22e-03	9.87e-01	0.992
##	22.17	3290	1	0.98915	1.22e-03	9.87e-01	0.992
##	22.33	3289	2	0.98900	1.24e-03	9.87e-01	0.991
##	22.43	3287	1	0.98893	1.24e-03	9.86e-01	0.991
##	22.47	3286	1	0.98886	1.25e-03	9.86e-01	0.991
##	22.50	3285	1	0.98879	1.26e-03	9.86e-01	0.991
##	22.60	3284	1	0.98871	1.26e-03	9.86e-01	0.991
##	22.67	3283	1	0.98864	1.27e-03	9.86e-01	0.991
##	22.70	3282	1	0.98857	1.27e-03	9.86e-01	0.991
##	22.80	3281	2	0.98842	1.29e-03	9.86e-01	0.991
##	22.93	3279	1	0.98835	1.29e-03	9.86e-01	0.991
##	22.97	3278	2	0.98821	1.31e-03	9.86e-01	0.991
##	23.03	3276	1	0.98813	1.31e-03	9.86e-01	0.991
##	23.07	3275	1	0.98806	1.32e-03	9.85e-01	0.991
##	23.10	3274	2	0.98791	1.33e-03	9.85e-01	0.991
##	23.13	3272	1	0.98784	1.34e-03	9.85e-01	0.990
##	23.17	3271	1	0.98777	1.34e-03	9.85e-01	0.990
##	23.30	3270	1	0.98769	1.35e-03	9.85e-01	0.990
##	23.33	3269	1	0.98762	1.36e-03	9.85e-01	0.990
##	23.37	3268	2	0.98747	1.37e-03	9.85e-01	0.990
##	23.40	3266	3	0.98724	1.39e-03	9.85e-01	0.990
##	23.43	3263	1	0.98717	1.40e-03	9.84e-01	0.990
##	23.47	3262	2	0.98702	1.41e-03	9.84e-01	0.990
##	23.53	3260	1	0.98694	1.42e-03	9.84e-01	0.990
##	23.60	3259	1	0.98686	1.42e-03	9.84e-01	0.990
##	23.63	3258	1	0.98678	1.43e-03	9.84e-01	0.990
##	23.70	3257	1	0.98670	1.44e-03	9.84e-01	0.990
##	23.73	3256	1	0.98662	1.44e-03	9.84e-01	0.989
##	23.77	3255	1	0.98654	1.45e-03	9.84e-01	0.989
##	23.80	3254	1	0.98645	1.46e-03	9.84e-01	0.989
##	23.93	3253	1	0.98637	1.46e-03	9.84e-01	0.989
##	24.13	3252	1	0.98628	1.47e-03	9.83e-01	0.989
##	24.20	3251	1	0.98619	1.48e-03	9.83e-01	0.989
##	24.37	3250	1	0.98611	1.49e-03	9.83e-01	0.989
##	24.40	3249	2	0.98593	1.50e-03	9.83e-01	0.989
##	24.50	3238	1	0.98584	1.51e-03	9.83e-01	0.989
##	24.53	3237	2	0.98566	1.52e-03	9.83e-01	0.989
##	24.60	3235	1	0.98556	1.53e-03	9.83e-01	0.989
##	24.63	3231	1	0.98547	1.54e-03	9.82e-01	0.988
##	24.67	3220	1	0.98538	1.54e-03	9.82e-01	0.988
##	24.80	3203	1	0.98529	1.55e-03	9.82e-01	0.988
##	24.83	3191	1	0.98520	1.56e-03	9.82e-01	0.988
##	24.87	3183	2	0.98501	1.57e-03	9.82e-01	0.988
##	24.93	3167	2	0.98482	1.59e-03	9.82e-01	0.988
##	25.37	3117	1	0.98473	1.60e-03	9.82e-01	0.988
##	25.40	3116	2	0.98454	1.61e-03	9.81e-01	0.988
##	25.47	3114	1	0.98444	1.62e-03	9.81e-01	0.988
##	25.60	3093	1	0.98434	1.63e-03	9.81e-01	0.988
##	25.67	3082	1	0.98425	1.64e-03	9.81e-01	0.987
##	25.70	3081	1	0.98415	1.64e-03	9.81e-01	0.987
##	25.73	3076	1	0.98405	1.65e-03	9.81e-01	0.987
##	25.80	3058	1	0.98395	1.66e-03	9.81e-01	0.987

##	25.93	3042	1	0.98385	1.67e-03	9.81e-01	0.987
##	25.97	3037	1	0.98375	1.68e-03	9.80e-01	0.987
##	26.00	3032	2	0.98354	1.69e-03	9.80e-01	0.987
##	26.07	3019	1	0.98344	1.70e-03	9.80e-01	0.987
##	26.10	3016	2	0.98324	1.72e-03	9.80e-01	0.987
##	26.17	3014	1	0.98313	1.73e-03	9.80e-01	0.987
##	26.33	3005	1	0.98303	1.73e-03	9.80e-01	0.986
##	26.37	3004	2	0.98282	1.75e-03	9.79e-01	0.986
##	26.43	2987	2	0.98260	1.77e-03	9.79e-01	0.986
##	26.50	2968	1	0.98249	1.78e-03	9.79e-01	0.986
##	26.57	2959	1	0.98238	1.79e-03	9.79e-01	0.986
##	26.60	2958	1	0.98227	1.80e-03	9.79e-01	0.986
##	26.63	2957	2	0.98204	1.81e-03	9.78e-01	0.986
##	26.67	2950	1	0.98193	1.82e-03	9.78e-01	0.986
##	26.73	2936	2	0.98170	1.84e-03	9.78e-01	0.985
##	26.77	2928	1	0.98158	1.85e-03	9.78e-01	0.985
##	26.87	2924	2	0.98135	1.87e-03	9.78e-01	0.985
##	26.90	2911	1	0.98123	1.88e-03	9.78e-01	0.985
##	26.93	2901	2	0.98100	1.90e-03	9.77e-01	0.985
##	27.07	2882	1	0.98088	1.91e-03	9.77e-01	0.985
##	27.10	2881	1	0.98076	1.91e-03	9.77e-01	0.985
##	27.17	2875	1	0.98064	1.92e-03	9.77e-01	0.984
##	27.37	2853	1	0.98051	1.93e-03	9.77e-01	0.984
##	27.53	2834	1	0.98039	1.94e-03	9.77e-01	0.984
##	27.57	2833	2	0.98014	1.96e-03	9.76e-01	0.984
##	27.60	2824	1	0.98002	1.97e-03	9.76e-01	0.984
##	27.70	2807	2	0.97976	1.99e-03	9.76e-01	0.984
##	27.83	2788	1	0.97963	2.00e-03	9.76e-01	0.984
##	27.93	2775	1	0.97949	2.01e-03	9.76e-01	0.983
##	27.97	2765	1	0.97935	2.02e-03	9.75e-01	0.983
##	28.47	2738	2	0.97907	2.05e-03	9.75e-01	0.983
##	28.50	2736	1	0.97893	2.06e-03	9.75e-01	0.983
##	28.53	2718	1	0.97878	2.07e-03	9.75e-01	0.983
##	28.80	2674	1	0.97862	2.08e-03	9.75e-01	0.983
##	28.83	2669	2	0.97832	2.10e-03	9.74e-01	0.982
##	28.90	2665	1	0.97816	2.12e-03	9.74e-01	0.982
##	28.93	2664	2	0.97785	2.14e-03	9.74e-01	0.982
##	29.10	2649	1	0.97769	2.15e-03	9.73e-01	0.982
##	29.13	2635	3	0.97720	2.19e-03	9.73e-01	0.982
##	29.17	2632	1	0.97704	2.20e-03	9.73e-01	0.981
##	29.23	2628	1	0.97688	2.22e-03	9.73e-01	0.981
##	29.37	2621	1	0.97671	2.23e-03	9.72e-01	0.981
##	29.40	2620	1	0.97655	2.24e-03	9.72e-01	0.981
##	29.47	2619	3	0.97605	2.28e-03	9.72e-01	0.981
##	29.53	2600	3	0.97554	2.32e-03	9.71e-01	0.980
##	29.80	2582	2	0.97519	2.34e-03	9.71e-01	0.980
##	29.90	2569	2	0.97484	2.37e-03	9.70e-01	0.979
##	30.10	2549	1	0.97466	2.38e-03	9.70e-01	0.979
##	30.30	2535	1	0.97448	2.40e-03	9.70e-01	0.979
##	30.33	2534	2	0.97412	2.43e-03	9.69e-01	0.979
##	30.40	2529	1	0.97393	2.44e-03	9.69e-01	0.979
##	30.43	2526	1	0.97375	2.45e-03	9.69e-01	0.979
##	30.47	2521	2	0.97337	2.48e-03	9.69e-01	0.978
##	30.63	2506	1	0.97318	2.50e-03	9.68e-01	0.978
##	30.67	2496	1	0.97298	2.51e-03	9.68e-01	0.978
##	30.83	2477	2	0.97259	2.54e-03	9.68e-01	0.978
##	30.90	2472	1	0.97239	2.56e-03	9.67e-01	0.977
##	30.93	2467	1	0.97219	2.57e-03	9.67e-01	0.977
##	31.10	2449	1	0.97198	2.59e-03	9.67e-01	0.977
##	31.23	2433	1	0.97178	2.60e-03	9.67e-01	0.977
##	31.40	2423	1	0.97157	2.62e-03	9.66e-01	0.977
##	31.43	2420	2	0.97114	2.65e-03	9.66e-01	0.976
##	31.47	2418	2	0.97071	2.68e-03	9.65e-01	0.976
##	31.67	2398	1	0.97050	2.70e-03	9.65e-01	0.976
##	31.70	2391	2	0.97006	2.73e-03	9.65e-01	0.975
##	31.77	2389	1	0.96984	2.75e-03	9.64e-01	0.975
##	31.93	2376	1	0.96962	2.77e-03	9.64e-01	0.975
##	32.00	2375	3	0.96894	2.82e-03	9.63e-01	0.974
##	32.03	2367	1	0.96871	2.83e-03	9.63e-01	0.974
##	32.53	2335	1	0.96847	2.85e-03	9.63e-01	0.974
##	32.57	2328	1	0.96823	2.87e-03	9.63e-01	0.974
##	32.60	2323	1	0.96798	2.89e-03	9.62e-01	0.974

##	32.63	2317	1	0.96774	2.91e-03	9.62e-01	0.973
##	32.77	2311	1	0.96749	2.93e-03	9.62e-01	0.973
##	32.80	2306	1	0.96724	2.95e-03	9.61e-01	0.973
##	32.83	2303	1	0.96699	2.96e-03	9.61e-01	0.973
##	32.93	2300	1	0.96673	2.98e-03	9.61e-01	0.973
##	32.97	2299	1	0.96647	3.00e-03	9.61e-01	0.972
##	33.10	2282	1	0.96621	3.02e-03	9.60e-01	0.972
##	33.13	2281	1	0.96595	3.04e-03	9.60e-01	0.972
##	33.20	2276	1	0.96568	3.06e-03	9.60e-01	0.972
##	33.23	2273	1	0.96541	3.08e-03	9.59e-01	0.971
##	33.33	2264	1	0.96514	3.10e-03	9.59e-01	0.971
##	33.47	2255	1	0.96486	3.12e-03	9.59e-01	0.971
##	33.53	2253	2	0.96430	3.17e-03	9.58e-01	0.971
##	33.60	2241	1	0.96402	3.19e-03	9.58e-01	0.970
##	33.63	2240	1	0.96373	3.21e-03	9.57e-01	0.970
##	33.67	2231	2	0.96314	3.25e-03	9.57e-01	0.970
##	33.73	2226	1	0.96284	3.28e-03	9.56e-01	0.969
##	33.77	2224	1	0.96254	3.30e-03	9.56e-01	0.969
##	33.83	2222	1	0.96224	3.32e-03	9.56e-01	0.969
##	33.90	2216	1	0.96193	3.34e-03	9.55e-01	0.969
##	33.93	2212	1	0.96162	3.37e-03	9.55e-01	0.968
##	34.00	2205	2	0.96100	3.41e-03	9.54e-01	0.968
##	34.13	2196	2	0.96037	3.46e-03	9.54e-01	0.967
##	34.17	2191	2	0.95972	3.51e-03	9.53e-01	0.967
##	34.43	2181	1	0.95939	3.53e-03	9.52e-01	0.966
##	34.50	2180	1	0.95906	3.55e-03	9.52e-01	0.966
##	34.53	2179	1	0.95873	3.58e-03	9.52e-01	0.966
##	34.57	2178	1	0.95840	3.60e-03	9.51e-01	0.965
##	34.60	2172	1	0.95807	3.63e-03	9.51e-01	0.965
##	34.67	2166	1	0.95773	3.65e-03	9.51e-01	0.965
##	34.73	2160	1	0.95739	3.68e-03	9.50e-01	0.965
##	34.83	2159	1	0.95704	3.70e-03	9.50e-01	0.964
##	34.93	2152	1	0.95669	3.73e-03	9.49e-01	0.964
##	34.97	2151	2	0.95598	3.78e-03	9.49e-01	0.963
##	35.20	2129	1	0.95562	3.80e-03	9.48e-01	0.963
##	35.37	2123	1	0.95525	3.83e-03	9.48e-01	0.963
##	35.53	2119	3	0.95412	3.91e-03	9.46e-01	0.962
##	35.57	2115	1	0.95374	3.94e-03	9.46e-01	0.961
##	35.60	2110	1	0.95335	3.97e-03	9.46e-01	0.961
##	35.63	2104	1	0.95296	4.00e-03	9.45e-01	0.961
##	35.67	2098	1	0.95256	4.03e-03	9.45e-01	0.960
##	35.77	2087	1	0.95216	4.05e-03	9.44e-01	0.960
##	35.83	2080	1	0.95174	4.08e-03	9.44e-01	0.960
##	35.87	2077	1	0.95133	4.11e-03	9.43e-01	0.959
##	35.90	2072	1	0.95091	4.14e-03	9.43e-01	0.959
##	35.97	2071	1	0.95049	4.17e-03	9.42e-01	0.959
##	36.03	2063	3	0.94921	4.27e-03	9.41e-01	0.958
##	36.07	2054	1	0.94877	4.30e-03	9.40e-01	0.957
##	36.27	2027	1	0.94832	4.33e-03	9.40e-01	0.957
##	36.40	2020	1	0.94786	4.36e-03	9.39e-01	0.956
##	36.53	2017	1	0.94741	4.40e-03	9.39e-01	0.956
##	36.80	1994	2	0.94647	4.47e-03	9.38e-01	0.955
##	36.87	1983	1	0.94599	4.50e-03	9.37e-01	0.955
##	36.93	1981	2	0.94504	4.57e-03	9.36e-01	0.954
##	36.97	1978	1	0.94455	4.61e-03	9.36e-01	0.954
##	37.00	1974	1	0.94406	4.64e-03	9.35e-01	0.953
##	37.07	1970	1	0.94356	4.68e-03	9.34e-01	0.953
##	37.27	1962	1	0.94306	4.71e-03	9.34e-01	0.952
##	37.33	1961	1	0.94256	4.75e-03	9.33e-01	0.952
##	37.43	1960	1	0.94206	4.79e-03	9.33e-01	0.951
##	37.77	1940	2	0.94105	4.86e-03	9.32e-01	0.951
##	37.90	1929	1	0.94054	4.90e-03	9.31e-01	0.950
##	38.00	1916	1	0.94002	4.93e-03	9.30e-01	0.950
##	38.17	1906	2	0.93897	5.01e-03	9.29e-01	0.949
##	38.47	1880	1	0.93842	5.05e-03	9.29e-01	0.948
##	38.63	1863	1	0.93786	5.09e-03	9.28e-01	0.948
##	38.97	1833	1	0.93725	5.14e-03	9.27e-01	0.947
##	39.10	1825	1	0.93663	5.18e-03	9.27e-01	0.947
##	39.17	1819	2	0.93538	5.28e-03	9.25e-01	0.946
##	39.23	1817	1	0.93476	5.32e-03	9.24e-01	0.945
##	39.30	1810	1	0.93412	5.37e-03	9.24e-01	0.945
##	39.57	1790	2	0.93280	5.47e-03	9.22e-01	0.944

##	39.60	1788	1	0.93214	5.52e-03	9.21e-01	0.943
##	39.70	1782	1	0.93148	5.57e-03	9.21e-01	0.942
##	39.73	1777	2	0.93015	5.67e-03	9.19e-01	0.941
##	39.77	1770	2	0.92881	5.77e-03	9.18e-01	0.940
##	39.97	1758	1	0.92814	5.82e-03	9.17e-01	0.940
##	40.00	1751	1	0.92745	5.87e-03	9.16e-01	0.939
##	40.07	1743	1	0.92676	5.92e-03	9.15e-01	0.938
##	40.23	1731	1	0.92604	5.97e-03	9.14e-01	0.938
##	40.27	1728	1	0.92532	6.03e-03	9.14e-01	0.937
##	40.30	1723	1	0.92460	6.08e-03	9.13e-01	0.937
##	40.63	1713	1	0.92386	6.14e-03	9.12e-01	0.936
##	40.70	1706	1	0.92312	6.19e-03	9.11e-01	0.935
##	40.73	1701	2	0.92161	6.30e-03	9.09e-01	0.934
##	41.03	1673	1	0.92082	6.36e-03	9.08e-01	0.933
##	41.20	1659	1	0.92000	6.43e-03	9.07e-01	0.933
##	41.23	1651	1	0.91917	6.49e-03	9.07e-01	0.932
##	41.50	1637	1	0.91833	6.56e-03	9.06e-01	0.931
##	41.63	1636	2	0.91664	6.69e-03	9.04e-01	0.930
##	41.90	1613	2	0.91489	6.82e-03	9.02e-01	0.928
##	42.17	1606	1	0.91400	6.89e-03	9.01e-01	0.928
##	42.20	1605	1	0.91311	6.96e-03	9.00e-01	0.927
##	42.30	1601	1	0.91221	7.03e-03	8.99e-01	0.926
##	42.33	1598	1	0.91131	7.10e-03	8.97e-01	0.925
##	42.37	1596	2	0.90944	7.24e-03	8.95e-01	0.924
##	42.47	1587	1	0.90849	7.31e-03	8.94e-01	0.923
##	42.97	1553	1	0.90751	7.39e-03	8.93e-01	0.922
##	43.00	1548	1	0.90651	7.47e-03	8.92e-01	0.921
##	43.20	1539	2	0.90449	7.63e-03	8.90e-01	0.920
##	43.53	1515	1	0.90343	7.72e-03	8.88e-01	0.919
##	43.60	1513	2	0.90127	7.89e-03	8.86e-01	0.917
##	43.70	1503	1	0.90017	7.98e-03	8.85e-01	0.916
##	43.73	1500	2	0.89797	8.15e-03	8.82e-01	0.914
##	43.83	1486	1	0.89685	8.24e-03	8.81e-01	0.913
##	44.07	1474	1	0.89572	8.33e-03	8.80e-01	0.912
##	44.27	1456	2	0.89340	8.52e-03	8.77e-01	0.910
##	44.73	1437	1	0.89219	8.62e-03	8.75e-01	0.909
##	44.83	1431	1	0.89098	8.71e-03	8.74e-01	0.908
##	44.87	1426	1	0.88975	8.81e-03	8.73e-01	0.907
##	45.00	1414	1	0.88848	8.92e-03	8.71e-01	0.906
##	45.47	1384	1	0.88713	9.03e-03	8.70e-01	0.905
##	45.87	1373	1	0.88574	9.16e-03	8.68e-01	0.904
##	45.93	1369	2	0.88294	9.40e-03	8.65e-01	0.902
##	45.97	1367	2	0.88013	9.63e-03	8.61e-01	0.899
##	46.47	1341	1	0.87869	9.76e-03	8.60e-01	0.898
##	46.53	1337	1	0.87717	9.89e-03	8.58e-01	0.897
##	46.70	1332	2	0.87391	1.02e-02	8.54e-01	0.894
##	46.77	1325	1	0.87224	1.03e-02	8.52e-01	0.893
##	46.83	1307	1	0.87051	1.05e-02	8.50e-01	0.891
##	46.87	1303	1	0.86875	1.06e-02	8.48e-01	0.890
##	47.37	1280	1	0.86689	1.08e-02	8.46e-01	0.888
##	47.67	1272	1	0.86494	1.10e-02	8.44e-01	0.887
##	48.07	1238	3	0.85852	1.17e-02	8.36e-01	0.882
##	48.13	1229	1	0.85637	1.19e-02	8.33e-01	0.880
##	48.70	1216	1	0.85398	1.21e-02	8.31e-01	0.878
##	48.73	1215	1	0.85158	1.24e-02	8.28e-01	0.876
##	49.53	1168	1	0.84907	1.27e-02	8.25e-01	0.874
##	50.10	1131	1	0.84644	1.30e-02	8.21e-01	0.872
##	50.23	1121	1	0.84378	1.32e-02	8.18e-01	0.870
##	51.00	1080	1	0.84102	1.35e-02	8.15e-01	0.868
##	51.10	1069	1	0.83810	1.39e-02	8.11e-01	0.866
##	51.53	1050	1	0.83511	1.42e-02	8.08e-01	0.863
##	52.10	1025	1	0.83197	1.46e-02	8.04e-01	0.861
##	52.13	1024	1	0.82885	1.49e-02	8.00e-01	0.859
##	52.20	1020	1	0.82568	1.53e-02	7.96e-01	0.856
##	52.57	1006	1	0.82249	1.56e-02	7.92e-01	0.854
##	53.03	982	1	0.81919	1.60e-02	7.88e-01	0.851
##	53.17	975	1	0.81583	1.64e-02	7.84e-01	0.849
##	53.37	968	1	0.81246	1.67e-02	7.80e-01	0.846
##	53.40	966	2	0.80574	1.75e-02	7.72e-01	0.841
##	54.00	942	1	0.80224	1.78e-02	7.68e-01	0.838
##	54.17	936	1	0.79871	1.82e-02	7.64e-01	0.835
##	54.20	934	1	0.79520	1.85e-02	7.60e-01	0.832

##	54.27	929	1	0.79162	1.89e-02	7.55e-01	0.830
##	55.00	898	1	0.78792	1.93e-02	7.51e-01	0.827
##	55.07	897	1	0.78422	1.97e-02	7.47e-01	0.824
##	55.33	890	1	0.78045	2.00e-02	7.42e-01	0.821
##	55.40	886	1	0.77662	2.04e-02	7.38e-01	0.818
##	55.53	877	1	0.77256	2.08e-02	7.33e-01	0.815
##	55.70	870	1	0.76849	2.13e-02	7.28e-01	0.811
##	55.90	855	1	0.76432	2.17e-02	7.23e-01	0.808
##	56.17	849	1	0.76015	2.21e-02	7.18e-01	0.805
##	56.27	845	1	0.75592	2.25e-02	7.13e-01	0.801
##	56.30	844	1	0.75168	2.29e-02	7.08e-01	0.798
##	56.50	835	1	0.74737	2.33e-02	7.03e-01	0.795
##	57.00	816	1	0.74250	2.39e-02	6.97e-01	0.791
##	57.03	813	1	0.73737	2.44e-02	6.91e-01	0.787
##	57.07	810	1	0.73225	2.50e-02	6.85e-01	0.783
##	57.10	807	1	0.72701	2.55e-02	6.79e-01	0.779
##	57.17	804	1	0.72151	2.61e-02	6.72e-01	0.774
##	57.23	802	1	0.71605	2.67e-02	6.66e-01	0.770
##	57.77	783	1	0.71042	2.72e-02	6.59e-01	0.766
##	58.10	767	1	0.70465	2.78e-02	6.52e-01	0.761
##	58.17	765	1	0.69888	2.84e-02	6.45e-01	0.757
##	58.20	764	1	0.69316	2.90e-02	6.39e-01	0.752
##	58.43	757	1	0.68737	2.95e-02	6.32e-01	0.748
##	59.03	737	1	0.68114	3.02e-02	6.24e-01	0.743
##	59.30	730	1	0.67477	3.08e-02	6.17e-01	0.738
##	59.50	722	2	0.66177	3.21e-02	6.02e-01	0.728
##	59.60	718	1	0.65529	3.27e-02	5.94e-01	0.723
##	59.97	705	1	0.64843	3.34e-02	5.86e-01	0.717
##	60.07	700	1	0.64148	3.41e-02	5.78e-01	0.712
##	60.10	698	1	0.63458	3.47e-02	5.70e-01	0.706
##	60.23	691	1	0.62751	3.54e-02	5.62e-01	0.701
##	60.33	687	1	0.62035	3.60e-02	5.54e-01	0.695
##	60.93	673	1	0.61285	3.67e-02	5.45e-01	0.689
##	61.17	664	1	0.60532	3.74e-02	5.36e-01	0.683
##	61.23	662	1	0.59783	3.81e-02	5.28e-01	0.677
##	62.30	630	1	0.59018	3.88e-02	5.19e-01	0.671
##	62.37	627	1	0.58259	3.94e-02	5.10e-01	0.665
##	62.40	625	1	0.57490	4.00e-02	5.02e-01	0.659
##	62.63	621	1	0.56708	4.06e-02	4.93e-01	0.653
##	63.03	611	1	0.55913	4.13e-02	4.84e-01	0.646
##	66.03	520	1	0.54988	4.21e-02	4.73e-01	0.639
##	66.07	519	1	0.54059	4.29e-02	4.63e-01	0.632
##	68.10	469	1	0.53039	4.39e-02	4.51e-01	0.624
##	68.60	459	1	0.51966	4.49e-02	4.39e-01	0.616
##	68.87	456	1	0.50903	4.58e-02	4.27e-01	0.607
##	68.97	451	1	0.49791	4.68e-02	4.14e-01	0.599
##	69.30	444	1	0.48637	4.78e-02	4.01e-01	0.590
##	69.40	441	2	0.46310	4.96e-02	3.75e-01	0.571
##	69.73	427	2	0.43774	5.14e-02	3.48e-01	0.551
##	70.10	423	1	0.42538	5.22e-02	3.34e-01	0.541
##	70.27	419	1	0.41241	5.30e-02	3.21e-01	0.530
##	70.77	410	1	0.39940	5.37e-02	3.07e-01	0.520
##	71.67	394	1	0.38436	5.47e-02	2.91e-01	0.508
##	73.63	347	1	0.36257	5.72e-02	2.66e-01	0.494
##	73.80	344	1	0.34149	5.91e-02	2.43e-01	0.479
##	73.87	341	1	0.32038	6.05e-02	2.21e-01	0.464
##	74.13	340	1	0.30050	6.12e-02	2.02e-01	0.448
##	75.30	315	1	0.27973	6.18e-02	1.81e-01	0.431
##	75.47	311	1	0.25724	6.23e-02	1.60e-01	0.414
##	76.23	294	1	0.23359	6.26e-02	1.38e-01	0.395
##	76.90	280	1	0.21038	6.21e-02	1.18e-01	0.375
##	78.63	255	1	0.18690	6.12e-02	9.84e-02	0.355
##	79.20	244	1	0.16258	5.96e-02	7.93e-02	0.333
##	82.30	200	1	0.13255	5.77e-02	5.65e-02	0.311
##	82.77	195	1	0.10715	5.37e-02	4.01e-02	0.286
##	84.67	167	1	0.07499	4.82e-02	2.13e-02	0.265
##	85.40	159	1	0.05081	3.98e-02	1.10e-02	0.236
##	86.00	151	1	0.03223	3.03e-02	5.09e-03	0.204
##	87.17	131	1	0.01882	2.12e-02	2.06e-03	0.172
##	87.93	121	1	0.01019	1.36e-02	7.47e-04	0.139
##	91.13	69	1	0.00241	4.97e-03	4.22e-05	0.138

## Anexo II - Detalle del modelo de cox 2 (mora mayor a 120 días) y su función de supervivencia

**Tabla 40** – Modelo de cox 2 (mora mayor a 120 días)

```

## Call:
## coxph(formula = Surv(phcalf16_actg$tiempoM120pv_2mes, phcalf16_actg$Mayor120) ~
##   phcalf16_actg$R_Ingp + phcalf16_actg$R_Plazo + phcalf16_actg$SubReg2 +
##   phcalf16_actg$Cicloef2 + phcalf16_actg$tyear + phcalf16_actg$Cicloef2 +
##   phcalf16_actg$cesantexciclof)
##
## n= 3570, number of events= 455
##
##                                     coef
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)      0.42558
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)    0.46759
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4)    0.51048
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5)             0.82877
## phcalf16_actg$R_Plazomayor a 20 años                   0.29785
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 0.58730
## phcalf16_actg$Cicloef21                                0.16999
## phcalf16_actg$tyear                                     0.42954
## phcalf16_actg$Cicloef21                                -3.63624
## phcalf16_actg$cesantexciclof1                          1.93746
##                                                         exp(coef)
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)    1.53048
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)  1.59614
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4)  1.66610
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5)           2.29050
## phcalf16_actg$R_Plazomayor a 20 años                 1.34696
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 1.79912
## phcalf16_actg$Cicloef21                              1.18529
## phcalf16_actg$tyear                                   1.53656
## phcalf16_actg$Cicloef21                              0.02635
## phcalf16_actg$cesantexciclof1                        6.94110
##                                                         se(coef)
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)    0.14001
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)  0.14087
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4)  0.14882
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5)           0.16432
## phcalf16_actg$R_Plazomayor a 20 años                 0.10505
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 0.10308
## phcalf16_actg$Cicloef21                              0.18213
## phcalf16_actg$tyear                                   0.05176
## phcalf16_actg$Cicloef21                              0.12625
## phcalf16_actg$cesantexciclof1                        0.13723
##                                                         z
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)    3.040
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)  3.319
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4)  3.430
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5)           5.044
## phcalf16_actg$R_Plazomayor a 20 años                 2.835
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 5.698
## phcalf16_actg$Cicloef21                              0.933
## phcalf16_actg$tyear                                   8.299
## phcalf16_actg$Cicloef21                              -28.801
## phcalf16_actg$cesantexciclof1                       14.118
##                                                         Pr(>|z|)
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)    0.002369
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3)  0.000902
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4)  0.000603
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5)           4.57e-07
## phcalf16_actg$R_Plazomayor a 20 años                 0.004580
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 1.21e-08
## phcalf16_actg$Cicloef21                              0.350634
## phcalf16_actg$tyear                                   < 2e-16
## phcalf16_actg$Cicloef21                              < 2e-16
## phcalf16_actg$cesantexciclof1                        < 2e-16
##
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2)    **

```



```

## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3) ***
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) ***
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) ***
## phcalf16_actg$R_Plazomayor a 20 años **
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) ***
## phcalf16_actg$Cicloef21
## phcalf16_actg$tyear
## phcalf16_actg$Cicloef21 ***
## phcalf16_actg$cesantexciclof1 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## exp(coef)
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2) 1.53048
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3) 1.59614
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) 1.66610
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) 2.29050
## phcalf16_actg$R_Plazomayor a 20 años 1.34696
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 1.79912
## phcalf16_actg$Cicloef21 1.18529
## phcalf16_actg$tyear 1.53656
## phcalf16_actg$Cicloef21 0.02635
## phcalf16_actg$cesantexciclof1 6.94110
## exp(-coef)
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2) 0.6534
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3) 0.6265
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) 0.6002
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) 0.4366
## phcalf16_actg$R_Plazomayor a 20 años 0.7424
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 0.5558
## phcalf16_actg$Cicloef21 0.8437
## phcalf16_actg$tyear 0.6508
## phcalf16_actg$Cicloef21 37.9489
## phcalf16_actg$cesantexciclof1 0.1441
## lower .95
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2) 1.16318
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3) 1.21105
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) 1.24458
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) 1.65981
## phcalf16_actg$R_Plazomayor a 20 años 1.09631
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 1.47001
## phcalf16_actg$Cicloef21 0.82947
## phcalf16_actg$tyear 1.38832
## phcalf16_actg$Cicloef21 0.02057
## phcalf16_actg$cesantexciclof1 5.30414
## upper .95
## phcalf16_actg$R_Ingpentre 770,01 y 1.133 USD (Q2) 2.01376
## phcalf16_actg$R_Ingpentre 1.133,01 y 1.603 USD (Q3) 2.10367
## phcalf16_actg$R_Ingpentre 1.603,01 y 2.362 USD (Q4) 2.23037
## phcalf16_actg$R_Ingpmayor a 2.362 USD (Q5) 3.16083
## phcalf16_actg$R_Plazomayor a 20 años 1.65492
## phcalf16_actg$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE) 2.20190
## phcalf16_actg$Cicloef21 1.69376
## phcalf16_actg$tyear 1.70062
## phcalf16_actg$Cicloef21 0.03375
## phcalf16_actg$cesantexciclof1 9.08326
##
## Concordance= 0.911 (se = 0.014 )
## Rsquare= 0.269 (max possible= 0.861 )
## Likelihood ratio test= 1118 on 10 df, p=0
## Wald test = 1043 on 10 df, p=0
## Score (logrank) test = 1944 on 10 df, p=0

```

**Tabla 41 – Función de supervivencia: Modelo de cox 2 (mora mayor a 120 días)**

```
## Call: survfit(formula = coxsuper_actg_120)
##
##   time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##   5.17  3570     1    1.000  7.99e-05    1.000    1.000
##   5.20  3569     2    1.000  1.40e-04    0.999    1.000
##   5.27  3567     1    1.000  1.62e-04    0.999    1.000
##   5.53  3566     1    1.000  1.82e-04    0.999    1.000
##   5.60  3565     1    1.000  2.01e-04    0.999    1.000
##   5.97  3564     1    0.999  2.18e-04    0.999    1.000
##   6.13  3563     2    0.999  2.49e-04    0.999    1.000
##   6.17  3561     1    0.999  2.64e-04    0.999    1.000
##   6.23  3560     2    0.999  2.92e-04    0.998    1.000
##   6.57  3558     1    0.999  3.06e-04    0.998    1.000
##   6.60  3557     1    0.999  3.19e-04    0.998    0.999
##   6.63  3556     1    0.999  3.32e-04    0.998    0.999
##   6.67  3555     1    0.999  3.44e-04    0.998    0.999
##   6.90  3554     1    0.999  3.56e-04    0.998    0.999
##   7.30  3553     1    0.999  3.69e-04    0.998    0.999
##   7.37  3552     1    0.998  3.81e-04    0.998    0.999
##   7.40  3551     1    0.998  3.92e-04    0.998    0.999
##   8.03  3550     2    0.998  4.15e-04    0.997    0.999
##   8.07  3548     1    0.998  4.26e-04    0.997    0.999
##   8.13  3547     1    0.998  4.38e-04    0.997    0.999
##   8.40  3546     1    0.998  4.48e-04    0.997    0.999
##   8.53  3545     1    0.998  4.59e-04    0.997    0.999
##   8.57  3544     1    0.998  4.70e-04    0.997    0.999
##   8.60  3543     1    0.998  4.81e-04    0.997    0.999
##   8.70  3542     1    0.998  4.91e-04    0.997    0.999
##   8.83  3541     1    0.997  5.02e-04    0.996    0.998
##   9.00  3540     1    0.997  5.12e-04    0.996    0.998
##   9.10  3539     1    0.997  5.23e-04    0.996    0.998
##   9.17  3538     1    0.997  5.33e-04    0.996    0.998
##   9.20  3537     1    0.997  5.43e-04    0.996    0.998
##   9.30  3536     1    0.997  5.53e-04    0.996    0.998
##   9.33  3535     1    0.997  5.64e-04    0.996    0.998
##   9.47  3534     1    0.997  5.74e-04    0.996    0.998
##   9.60  3533     1    0.997  5.84e-04    0.996    0.998
##   9.73  3532     1    0.997  5.94e-04    0.996    0.998
##   9.83  3531     1    0.997  6.04e-04    0.995    0.998
##   9.90  3530     1    0.996  6.14e-04    0.995    0.998
##   9.97  3529     1    0.996  6.24e-04    0.995    0.998
##  10.17  3528     1    0.996  6.34e-04    0.995    0.998
##  10.20  3527     3    0.996  6.63e-04    0.995    0.997
##  10.27  3524     1    0.996  6.73e-04    0.995    0.997
##  10.30  3523     1    0.996  6.83e-04    0.994    0.997
##  10.33  3522     1    0.996  6.93e-04    0.994    0.997
##  10.47  3521     1    0.996  7.03e-04    0.994    0.997
##  10.50  3520     1    0.996  7.12e-04    0.994    0.997
##  10.53  3519     1    0.995  7.22e-04    0.994    0.997
##  10.60  3518     1    0.995  7.32e-04    0.994    0.997
##  10.70  3517     2    0.995  7.51e-04    0.994    0.997
##  10.80  3515     1    0.995  7.60e-04    0.994    0.997
##  10.83  3514     1    0.995  7.70e-04    0.993    0.996
##  10.87  3513     1    0.995  7.80e-04    0.993    0.996
##  11.10  3512     1    0.995  7.89e-04    0.993    0.996
##  11.20  3511     1    0.995  7.99e-04    0.993    0.996
##  11.33  3510     1    0.995  8.08e-04    0.993    0.996
##  11.37  3509     1    0.994  8.18e-04    0.993    0.996
##  11.40  3508     2    0.994  8.37e-04    0.993    0.996
##  11.77  3506     1    0.994  8.46e-04    0.993    0.996
##  11.80  3505     1    0.994  8.56e-04    0.992    0.996
##  11.87  3504     2    0.994  8.75e-04    0.992    0.996
##  12.13  3502     1    0.994  8.84e-04    0.992    0.996
##  12.17  3501     1    0.994  8.94e-04    0.992    0.995
##  12.30  3500     3    0.993  9.22e-04    0.992    0.995
##  12.40  3497     1    0.993  9.31e-04    0.991    0.995
##  12.53  3496     2    0.993  9.50e-04    0.991    0.995
##  12.57  3494     1    0.993  9.59e-04    0.991    0.995
##  12.60  3493     2    0.993  9.78e-04    0.991    0.995
```

##	12.73	3491	1	0.993	9.88e-04	0.991	0.995
##	12.87	3490	1	0.993	9.97e-04	0.991	0.995
##	13.03	3489	1	0.992	1.01e-03	0.990	0.994
##	13.23	3488	1	0.992	1.02e-03	0.990	0.994
##	13.30	3487	1	0.992	1.03e-03	0.990	0.994
##	13.53	3486	1	0.992	1.03e-03	0.990	0.994
##	13.63	3485	1	0.992	1.04e-03	0.990	0.994
##	13.67	3484	1	0.992	1.05e-03	0.990	0.994
##	13.70	3483	1	0.992	1.06e-03	0.990	0.994
##	13.80	3482	1	0.992	1.07e-03	0.990	0.994
##	13.83	3481	1	0.992	1.08e-03	0.990	0.994
##	14.10	3480	1	0.992	1.09e-03	0.989	0.994
##	14.23	3479	1	0.991	1.10e-03	0.989	0.994
##	14.27	3478	1	0.991	1.11e-03	0.989	0.993
##	14.37	3477	1	0.991	1.12e-03	0.989	0.993
##	14.47	3476	1	0.991	1.13e-03	0.989	0.993
##	14.50	3475	1	0.991	1.14e-03	0.989	0.993
##	14.53	3474	1	0.991	1.15e-03	0.989	0.993
##	14.57	3473	2	0.991	1.17e-03	0.988	0.993
##	14.70	3471	2	0.990	1.18e-03	0.988	0.993
##	14.90	3469	1	0.990	1.19e-03	0.988	0.993
##	14.93	3468	1	0.990	1.20e-03	0.988	0.993
##	14.97	3467	1	0.990	1.21e-03	0.988	0.992
##	15.17	3466	1	0.990	1.22e-03	0.988	0.992
##	15.30	3465	1	0.990	1.23e-03	0.987	0.992
##	15.37	3464	2	0.990	1.25e-03	0.987	0.992
##	15.53	3462	1	0.990	1.26e-03	0.987	0.992
##	15.60	3461	1	0.989	1.27e-03	0.987	0.992
##	15.83	3460	1	0.989	1.28e-03	0.987	0.992
##	16.00	3459	1	0.989	1.29e-03	0.987	0.992
##	16.03	3458	1	0.989	1.30e-03	0.987	0.992
##	16.07	3457	2	0.989	1.31e-03	0.986	0.991
##	16.17	3455	1	0.989	1.32e-03	0.986	0.991
##	16.20	3454	1	0.989	1.33e-03	0.986	0.991
##	16.23	3453	1	0.989	1.34e-03	0.986	0.991
##	16.33	3452	1	0.988	1.35e-03	0.986	0.991
##	16.50	3451	1	0.988	1.36e-03	0.986	0.991
##	16.60	3450	2	0.988	1.38e-03	0.985	0.991
##	16.63	3448	2	0.988	1.40e-03	0.985	0.991
##	16.67	3446	2	0.988	1.42e-03	0.985	0.990
##	16.70	3444	1	0.988	1.43e-03	0.985	0.990
##	16.73	3443	1	0.987	1.43e-03	0.985	0.990
##	16.90	3442	1	0.987	1.44e-03	0.985	0.990
##	16.97	3441	1	0.987	1.45e-03	0.984	0.990
##	17.00	3440	1	0.987	1.46e-03	0.984	0.990
##	17.07	3439	1	0.987	1.47e-03	0.984	0.990
##	17.10	3438	2	0.987	1.49e-03	0.984	0.990
##	17.30	3436	1	0.987	1.50e-03	0.984	0.990
##	17.37	3435	1	0.987	1.51e-03	0.984	0.989
##	17.80	3434	1	0.986	1.52e-03	0.983	0.989
##	18.37	3433	3	0.986	1.54e-03	0.983	0.989
##	18.43	3430	1	0.986	1.55e-03	0.983	0.989
##	18.47	3429	2	0.986	1.57e-03	0.983	0.989
##	18.67	3427	1	0.986	1.58e-03	0.983	0.989
##	18.70	3426	1	0.985	1.59e-03	0.982	0.989
##	18.83	3425	1	0.985	1.60e-03	0.982	0.989
##	18.90	3424	1	0.985	1.61e-03	0.982	0.988
##	19.00	3423	1	0.985	1.62e-03	0.982	0.988
##	19.30	3422	2	0.985	1.64e-03	0.982	0.988
##	19.37	3420	1	0.985	1.65e-03	0.982	0.988
##	19.47	3419	1	0.985	1.66e-03	0.981	0.988
##	19.53	3418	1	0.985	1.66e-03	0.981	0.988
##	19.67	3417	1	0.984	1.67e-03	0.981	0.988
##	19.70	3416	2	0.984	1.69e-03	0.981	0.988
##	19.83	3414	1	0.984	1.70e-03	0.981	0.987
##	19.93	3413	1	0.984	1.71e-03	0.981	0.987
##	20.03	3412	2	0.984	1.73e-03	0.980	0.987
##	20.17	3410	1	0.984	1.74e-03	0.980	0.987
##	20.27	3409	1	0.983	1.75e-03	0.980	0.987
##	20.33	3408	1	0.983	1.76e-03	0.980	0.987
##	20.47	3407	1	0.983	1.77e-03	0.980	0.987
##	20.57	3406	1	0.983	1.78e-03	0.980	0.987

##	20.60	3405	1	0.983	1.79e-03	0.980	0.987
##	20.70	3404	1	0.983	1.79e-03	0.979	0.986
##	20.73	3403	1	0.983	1.80e-03	0.979	0.986
##	20.87	3402	1	0.983	1.81e-03	0.979	0.986
##	20.90	3401	2	0.982	1.83e-03	0.979	0.986
##	20.93	3399	1	0.982	1.84e-03	0.979	0.986
##	20.97	3398	3	0.982	1.87e-03	0.978	0.986
##	21.33	3395	1	0.982	1.88e-03	0.978	0.985
##	21.40	3394	1	0.982	1.89e-03	0.978	0.985
##	21.60	3393	1	0.982	1.90e-03	0.978	0.985
##	21.67	3392	1	0.981	1.91e-03	0.978	0.985
##	21.77	3391	1	0.981	1.92e-03	0.978	0.985
##	21.80	3390	2	0.981	1.93e-03	0.977	0.985
##	21.87	3388	1	0.981	1.94e-03	0.977	0.985
##	21.90	3387	1	0.981	1.95e-03	0.977	0.985
##	21.93	3386	1	0.981	1.96e-03	0.977	0.985
##	22.20	3385	1	0.981	1.97e-03	0.977	0.984
##	22.23	3384	1	0.980	1.98e-03	0.977	0.984
##	22.43	3383	1	0.980	1.99e-03	0.976	0.984
##	22.47	3382	1	0.980	2.00e-03	0.976	0.984
##	22.50	3381	1	0.980	2.01e-03	0.976	0.984
##	22.53	3380	1	0.980	2.02e-03	0.976	0.984
##	22.57	3379	1	0.980	2.03e-03	0.976	0.984
##	22.60	3378	2	0.980	2.05e-03	0.976	0.984
##	22.67	3376	1	0.979	2.06e-03	0.975	0.983
##	22.70	3375	1	0.979	2.07e-03	0.975	0.983
##	22.77	3374	1	0.979	2.08e-03	0.975	0.983
##	22.80	3373	1	0.979	2.08e-03	0.975	0.983
##	22.83	3372	1	0.979	2.09e-03	0.975	0.983
##	22.87	3371	1	0.979	2.10e-03	0.975	0.983
##	23.13	3370	1	0.979	2.11e-03	0.975	0.983
##	23.33	3369	2	0.978	2.13e-03	0.974	0.983
##	23.50	3367	1	0.978	2.14e-03	0.974	0.982
##	23.60	3366	1	0.978	2.15e-03	0.974	0.982
##	23.67	3365	1	0.978	2.16e-03	0.974	0.982
##	23.70	3364	1	0.978	2.17e-03	0.974	0.982
##	23.80	3363	2	0.978	2.19e-03	0.973	0.982
##	23.93	3361	1	0.978	2.20e-03	0.973	0.982
##	23.97	3360	1	0.977	2.21e-03	0.973	0.982
##	24.03	3359	1	0.977	2.22e-03	0.973	0.982
##	24.07	3358	1	0.977	2.22e-03	0.973	0.981
##	24.10	3357	2	0.977	2.24e-03	0.972	0.981
##	24.17	3355	1	0.977	2.25e-03	0.972	0.981
##	24.30	3354	1	0.977	2.26e-03	0.972	0.981
##	24.33	3353	1	0.976	2.27e-03	0.972	0.981
##	24.37	3352	2	0.976	2.29e-03	0.972	0.981
##	24.40	3350	1	0.976	2.30e-03	0.972	0.981
##	24.43	3341	1	0.976	2.31e-03	0.971	0.980
##	24.47	3339	1	0.976	2.32e-03	0.971	0.980
##	24.70	3317	1	0.976	2.33e-03	0.971	0.980
##	24.73	3316	1	0.976	2.34e-03	0.971	0.980
##	25.13	3244	1	0.975	2.35e-03	0.971	0.980
##	25.40	3221	2	0.975	2.37e-03	0.971	0.980
##	25.50	3209	1	0.975	2.38e-03	0.970	0.980
##	25.53	3208	2	0.975	2.40e-03	0.970	0.979
##	25.60	3196	1	0.975	2.41e-03	0.970	0.979
##	25.67	3184	1	0.974	2.42e-03	0.970	0.979
##	25.83	3155	1	0.974	2.43e-03	0.970	0.979
##	25.87	3146	1	0.974	2.44e-03	0.969	0.979
##	26.40	3111	2	0.974	2.46e-03	0.969	0.979
##	26.47	3077	1	0.974	2.47e-03	0.969	0.979
##	26.70	3052	1	0.974	2.48e-03	0.969	0.978
##	26.73	3044	1	0.973	2.49e-03	0.969	0.978
##	26.80	3033	1	0.973	2.50e-03	0.968	0.978
##	26.97	2996	1	0.973	2.51e-03	0.968	0.978
##	27.00	2993	2	0.973	2.54e-03	0.968	0.978
##	27.10	2989	1	0.973	2.55e-03	0.968	0.978
##	27.17	2983	1	0.973	2.56e-03	0.968	0.978
##	27.33	2958	1	0.972	2.57e-03	0.967	0.977
##	27.37	2957	1	0.972	2.58e-03	0.967	0.977
##	27.43	2956	2	0.972	2.60e-03	0.967	0.977
##	27.67	2906	1	0.972	2.61e-03	0.967	0.977

##	27.73	2894	2	0.972	2.64e-03	0.966	0.977
##	27.77	2892	1	0.971	2.65e-03	0.966	0.977
##	27.87	2883	1	0.971	2.66e-03	0.966	0.976
##	27.93	2869	1	0.971	2.67e-03	0.966	0.976
##	28.10	2851	1	0.971	2.68e-03	0.966	0.976
##	28.17	2844	1	0.971	2.69e-03	0.965	0.976
##	28.57	2797	1	0.971	2.71e-03	0.965	0.976
##	28.70	2768	1	0.970	2.72e-03	0.965	0.976
##	28.93	2753	1	0.970	2.73e-03	0.965	0.976
##	28.97	2752	1	0.970	2.74e-03	0.965	0.975
##	29.50	2704	1	0.970	2.76e-03	0.965	0.975
##	29.80	2683	1	0.970	2.77e-03	0.964	0.975
##	29.83	2670	1	0.970	2.78e-03	0.964	0.975
##	29.93	2664	1	0.969	2.80e-03	0.964	0.975
##	30.13	2649	2	0.969	2.82e-03	0.964	0.975
##	30.23	2638	1	0.969	2.84e-03	0.963	0.974
##	30.37	2631	1	0.969	2.85e-03	0.963	0.974
##	30.40	2627	1	0.969	2.86e-03	0.963	0.974
##	30.47	2619	2	0.968	2.89e-03	0.963	0.974
##	30.53	2605	2	0.968	2.92e-03	0.962	0.974
##	30.80	2572	1	0.968	2.93e-03	0.962	0.973
##	30.90	2567	1	0.967	2.95e-03	0.962	0.973
##	31.10	2545	1	0.967	2.96e-03	0.961	0.973
##	31.33	2523	1	0.967	2.97e-03	0.961	0.973
##	31.43	2516	1	0.967	2.99e-03	0.961	0.973
##	31.47	2515	2	0.967	3.02e-03	0.961	0.972
##	31.63	2497	1	0.966	3.03e-03	0.960	0.972
##	31.67	2494	1	0.966	3.05e-03	0.960	0.972
##	31.83	2478	2	0.966	3.08e-03	0.960	0.972
##	31.90	2474	1	0.966	3.09e-03	0.960	0.972
##	31.93	2472	1	0.965	3.11e-03	0.959	0.971
##	32.10	2459	1	0.965	3.12e-03	0.959	0.971
##	32.23	2450	1	0.965	3.14e-03	0.959	0.971
##	32.40	2432	1	0.965	3.15e-03	0.959	0.971
##	32.43	2431	1	0.965	3.17e-03	0.958	0.971
##	32.47	2430	1	0.964	3.18e-03	0.958	0.971
##	32.67	2413	1	0.964	3.20e-03	0.958	0.970
##	32.70	2412	1	0.964	3.21e-03	0.958	0.970
##	32.77	2406	1	0.964	3.23e-03	0.957	0.970
##	33.00	2394	2	0.963	3.26e-03	0.957	0.970
##	33.60	2340	1	0.963	3.28e-03	0.957	0.970
##	33.63	2339	1	0.963	3.29e-03	0.957	0.969
##	33.77	2324	1	0.963	3.31e-03	0.956	0.969
##	33.80	2321	1	0.963	3.32e-03	0.956	0.969
##	33.83	2320	1	0.962	3.34e-03	0.956	0.969
##	33.93	2311	1	0.962	3.36e-03	0.956	0.969
##	34.13	2295	1	0.962	3.37e-03	0.955	0.969
##	34.23	2282	1	0.962	3.39e-03	0.955	0.968
##	34.33	2281	1	0.961	3.41e-03	0.955	0.968
##	34.47	2280	1	0.961	3.43e-03	0.955	0.968
##	34.67	2269	1	0.961	3.44e-03	0.954	0.968
##	34.73	2262	1	0.961	3.46e-03	0.954	0.968
##	34.83	2261	1	0.961	3.48e-03	0.954	0.967
##	35.00	2254	1	0.960	3.49e-03	0.954	0.967
##	35.13	2237	1	0.960	3.51e-03	0.953	0.967
##	35.17	2233	2	0.960	3.55e-03	0.953	0.967
##	35.43	2224	1	0.959	3.56e-03	0.952	0.966
##	35.50	2223	1	0.959	3.58e-03	0.952	0.966
##	35.53	2219	1	0.959	3.60e-03	0.952	0.966
##	35.60	2212	1	0.959	3.62e-03	0.952	0.966
##	35.67	2199	1	0.959	3.64e-03	0.951	0.966
##	35.73	2198	1	0.958	3.65e-03	0.951	0.965
##	35.83	2179	1	0.958	3.67e-03	0.951	0.965
##	35.93	2172	1	0.958	3.69e-03	0.951	0.965
##	35.97	2171	2	0.957	3.73e-03	0.950	0.965
##	36.53	2112	2	0.957	3.77e-03	0.949	0.964
##	36.57	2110	1	0.957	3.79e-03	0.949	0.964
##	36.60	2109	1	0.956	3.81e-03	0.949	0.964
##	36.63	2108	1	0.956	3.83e-03	0.949	0.964
##	36.83	2074	1	0.956	3.85e-03	0.948	0.963
##	36.87	2073	1	0.956	3.87e-03	0.948	0.963
##	36.90	2072	1	0.955	3.89e-03	0.948	0.963

##	37.03	2064	3	0.955	3.95e-03	0.947	0.962
##	37.40	2052	1	0.954	3.97e-03	0.947	0.962
##	37.53	2050	1	0.954	3.99e-03	0.946	0.962
##	37.80	2030	1	0.954	4.01e-03	0.946	0.962
##	37.93	2012	1	0.953	4.03e-03	0.946	0.961
##	37.97	2009	1	0.953	4.06e-03	0.945	0.961
##	38.00	2004	1	0.953	4.08e-03	0.945	0.961
##	38.27	1980	1	0.953	4.10e-03	0.945	0.961
##	38.33	1974	1	0.952	4.12e-03	0.944	0.960
##	38.77	1940	2	0.952	4.17e-03	0.944	0.960
##	38.90	1919	1	0.951	4.19e-03	0.943	0.960
##	39.63	1866	1	0.951	4.22e-03	0.943	0.959
##	40.17	1829	1	0.951	4.24e-03	0.943	0.959
##	40.23	1817	1	0.951	4.27e-03	0.942	0.959
##	40.57	1799	2	0.950	4.32e-03	0.941	0.958
##	40.70	1789	1	0.950	4.35e-03	0.941	0.958
##	40.73	1783	1	0.949	4.38e-03	0.941	0.958
##	40.77	1774	1	0.949	4.41e-03	0.940	0.958
##	41.07	1753	1	0.949	4.43e-03	0.940	0.957
##	41.27	1727	1	0.948	4.46e-03	0.939	0.957
##	41.30	1726	1	0.948	4.49e-03	0.939	0.957
##	41.73	1692	2	0.947	4.55e-03	0.938	0.956
##	42.03	1689	1	0.947	4.58e-03	0.938	0.956
##	42.20	1683	1	0.946	4.61e-03	0.937	0.955
##	42.23	1682	1	0.946	4.64e-03	0.937	0.955
##	42.50	1668	1	0.946	4.67e-03	0.937	0.955
##	42.90	1632	1	0.945	4.71e-03	0.936	0.955
##	43.17	1618	1	0.945	4.74e-03	0.936	0.954
##	43.20	1617	1	0.944	4.77e-03	0.935	0.954
##	43.30	1607	1	0.944	4.80e-03	0.935	0.954
##	43.33	1600	1	0.944	4.84e-03	0.934	0.953
##	43.37	1599	1	0.943	4.87e-03	0.934	0.953
##	43.47	1596	1	0.943	4.90e-03	0.933	0.953
##	43.97	1561	1	0.943	4.94e-03	0.933	0.952
##	44.00	1554	1	0.942	4.97e-03	0.932	0.952
##	44.20	1541	1	0.942	5.01e-03	0.932	0.952
##	44.53	1514	1	0.941	5.05e-03	0.931	0.951
##	44.60	1513	1	0.941	5.09e-03	0.931	0.951
##	44.70	1510	1	0.940	5.12e-03	0.930	0.950
##	44.73	1507	1	0.940	5.16e-03	0.930	0.950
##	44.83	1500	1	0.939	5.20e-03	0.929	0.950
##	45.27	1465	2	0.938	5.28e-03	0.928	0.949
##	45.73	1439	1	0.938	5.33e-03	0.928	0.948
##	45.87	1434	1	0.937	5.38e-03	0.927	0.948
##	46.00	1430	1	0.937	5.42e-03	0.926	0.948
##	46.47	1402	1	0.936	5.47e-03	0.926	0.947
##	46.87	1365	1	0.936	5.52e-03	0.925	0.947
##	46.93	1364	2	0.935	5.63e-03	0.924	0.946
##	46.97	1357	2	0.933	5.74e-03	0.922	0.945
##	47.47	1336	1	0.933	5.80e-03	0.922	0.944
##	47.77	1316	1	0.932	5.86e-03	0.921	0.944
##	47.87	1315	1	0.932	5.91e-03	0.920	0.943
##	48.37	1277	1	0.931	5.98e-03	0.919	0.943
##	49.07	1248	3	0.929	6.18e-03	0.917	0.941
##	49.70	1218	1	0.928	6.26e-03	0.916	0.940
##	49.73	1217	1	0.927	6.33e-03	0.915	0.940
##	50.53	1155	1	0.927	6.41e-03	0.914	0.939
##	51.10	1116	1	0.926	6.49e-03	0.913	0.939
##	51.23	1106	1	0.925	6.57e-03	0.912	0.938
##	52.00	1074	1	0.924	6.65e-03	0.911	0.937
##	52.10	1069	1	0.923	6.74e-03	0.910	0.937
##	52.53	1050	1	0.922	6.83e-03	0.909	0.936
##	53.10	1021	1	0.922	6.93e-03	0.908	0.935
##	53.20	1019	1	0.921	7.02e-03	0.907	0.934
##	53.57	1008	1	0.920	7.12e-03	0.906	0.934
##	54.03	987	1	0.919	7.22e-03	0.905	0.933
##	54.37	971	1	0.918	7.32e-03	0.904	0.932
##	55.00	941	1	0.917	7.42e-03	0.902	0.932
##	55.17	938	1	0.916	7.53e-03	0.901	0.931
##	56.53	878	1	0.915	7.65e-03	0.900	0.930
##	56.70	866	1	0.914	7.78e-03	0.899	0.929
##	57.27	844	1	0.913	7.91e-03	0.897	0.928

##	58.00	818	1	0.911	8.05e-03	0.896	0.927
##	58.03	814	1	0.910	8.20e-03	0.894	0.926
##	58.07	812	1	0.909	8.35e-03	0.893	0.925
##	59.10	780	1	0.907	8.52e-03	0.891	0.924
##	59.17	774	1	0.906	8.68e-03	0.889	0.923
##	59.20	773	1	0.905	8.85e-03	0.887	0.922
##	59.43	764	1	0.903	9.03e-03	0.886	0.921
##	60.03	745	1	0.902	9.20e-03	0.884	0.920
##	60.30	733	1	0.900	9.39e-03	0.882	0.919
##	60.50	727	1	0.899	9.57e-03	0.880	0.918
##	60.60	719	1	0.897	9.75e-03	0.878	0.917
##	60.97	715	1	0.896	9.93e-03	0.877	0.916
##	61.07	706	1	0.894	1.01e-02	0.875	0.914
##	61.10	705	1	0.893	1.03e-02	0.873	0.913
##	61.33	699	1	0.891	1.05e-02	0.871	0.912
##	62.17	674	1	0.890	1.07e-02	0.869	0.911
##	62.23	670	1	0.888	1.09e-02	0.867	0.910
##	63.30	640	1	0.886	1.11e-02	0.865	0.908
##	63.37	634	1	0.885	1.13e-02	0.863	0.907
##	63.40	631	1	0.883	1.15e-02	0.861	0.906
##	64.03	617	1	0.881	1.17e-02	0.859	0.905
##	67.03	526	1	0.879	1.20e-02	0.856	0.903
##	67.07	525	1	0.877	1.23e-02	0.853	0.902
##	69.10	479	1	0.875	1.27e-02	0.850	0.900
##	69.60	464	1	0.872	1.30e-02	0.847	0.898
##	69.97	454	1	0.869	1.35e-02	0.843	0.896
##	70.30	444	1	0.867	1.39e-02	0.840	0.894
##	70.40	439	1	0.863	1.44e-02	0.836	0.892
##	70.73	435	2	0.857	1.55e-02	0.827	0.888
##	72.67	390	1	0.853	1.62e-02	0.822	0.885
##	74.63	341	1	0.848	1.73e-02	0.814	0.882
##	74.87	336	1	0.842	1.84e-02	0.807	0.879
##	76.30	310	1	0.836	1.98e-02	0.798	0.876
##	77.23	293	1	0.829	2.12e-02	0.789	0.872
##	77.90	281	1	0.823	2.26e-02	0.779	0.868
##	79.63	250	1	0.814	2.47e-02	0.767	0.863
##	83.77	195	1	0.803	2.73e-02	0.751	0.858
##	86.40	157	1	0.789	3.10e-02	0.731	0.853
##	87.00	140	1	0.775	3.45e-02	0.711	0.846

**Tabla 42** – Supuesto de proporcionalidad, estadístico Rho: Modelo de cox 2 (mora mayor a 120 días)

##		rho
##	phcalf16_actg\$R_Ingpentre 770,01 y 1.133 USD (Q2)	0.00121
##	phcalf16_actg\$R_Ingpentre 1.133,01 y 1.603 USD (Q3)	0.01528
##	phcalf16_actg\$R_Ingpentre 1.603,01 y 2.362 USD (Q4)	-0.00734
##	phcalf16_actg\$R_Ingpmayor a 2.362 USD (Q5)	0.08242
##	phcalf16_actg\$R_Plazomayor a 20 años	0.05032
##	phcalf16_actg\$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE)	0.03831
##	phcalf16_actg\$Cicloef21	0.04685
##	phcalf16_actg\$tyear	-0.23118
##	phcalf16_actg\$Cicloef21	0.37273
##	phcalf16_actg\$cesantexciclof1	0.04100
##	GLOBAL	NA
##		chisq
##	phcalf16_actg\$R_Ingpentre 770,01 y 1.133 USD (Q2)	6.72e-04
##	phcalf16_actg\$R_Ingpentre 1.133,01 y 1.603 USD (Q3)	1.10e-01
##	phcalf16_actg\$R_Ingpentre 1.603,01 y 2.362 USD (Q4)	2.42e-02
##	phcalf16_actg\$R_Ingpmayor a 2.362 USD (Q5)	3.18e+00
##	phcalf16_actg\$R_Plazomayor a 20 años	1.21e+00
##	phcalf16_actg\$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE)	6.79e-01
##	phcalf16_actg\$Cicloef21	9.36e-01
##	phcalf16_actg\$tyear	2.03e+01
##	phcalf16_actg\$Cicloef21	5.61e+01
##	phcalf16_actg\$cesantexciclof1	7.57e-01
##	GLOBAL	1.15e+02
##		p

## phcalf16_actg\$R_Ingpentre 770,01 y 1.133 USD (Q2)	9.79e-01
## phcalf16_actg\$R_Ingpentre 1.133,01 y 1.603 USD (Q3)	7.41e-01
## phcalf16_actg\$R_Ingpentre 1.603,01 y 2.362 USD (Q4)	8.76e-01
## phcalf16_actg\$R_Ingpmayor a 2.362 USD (Q5)	7.44e-02
## phcalf16_actg\$R_Plazomayor a 20 años	2.72e-01
## phcalf16_actg\$SubReg2COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE)	4.10e-01
## phcalf16_actg\$Cicloef21	3.33e-01
## phcalf16_actg\$tyear	6.56e-06
## phcalf16_actg\$Cicloef21	6.86e-14
## phcalf16_actg\$cesantexciclof1	3.84e-01
## GLOBAL	0.00e+00

### Anexo III - Código de programación R - Etapa 1 - Preparación de la data

#librerías a cargar

library(ggplot2)

library(ggfortify)

library(dplyr)

library(survival)

library(rpart)

library(rpart.plot)

library(gmodels)

library(gtools)

#cargar base de datos

phcalf16\_0 <- read.table("informacionCalificacionMora\_21JUN.tsv",header=TRUE,sep = "\t",dec = ",")

str(phcalf16\_0)

#Cambiar de formato variables tipo fecha

phcalf16\_0\$FEC\_DESSS <- as.Date(phcalf16\_0\$FEC\_DESSS,format="%d/%m/%Y")

phcalf16\_0\$FECHA\_DESEMBOLSO<-as.Date(phcalf16\_0\$FECHA\_DESEMBOLSO, format="%d/%m/%Y")

phcalf16\_0\$FECHA\_CORTE\_REPORTE<-as.Date(phcalf16\_0\$FECHA\_CORTE\_REPORTE, format = "%d/%m/%Y")

phcalf16\_0\$FECHA\_FIN\_CREDITO <- as.Date(phcalf16\_0\$FECHA\_FIN\_CREDITO,format="%d/%m/%Y")

phcalf16\_0\$FEC\_MORA\_PRIMERA\_VEZ<- as.Date(phcalf16\_0\$FEC\_MORA\_PRIMERA\_VEZ, format="%d/%m/%Y")

phcalf16\_0\$FEC\_CAN\_MOR\_PRIMERA\_VEZ<-as.Date(phcalf16\_0\$FEC\_CAN\_MOR\_PRIMERA\_VEZ, format="%d/%m/%Y")

phcalf16\_0\$FEC\_NACIMIENTO <- as.Date(phcalf16\_0\$FEC\_NACIMIENTO,format="%d/%m/%Y")

phcalf16\_0\$FECHA\_MORA\_GRAVE <- as.Date(phcalf16\_0\$FECHA\_MORA\_GRAVE,format="%d/%m/%Y")

#Ordenar, agrupar casos por nut y filtrar aquellos referentes a vivienda terminada

phcalf16\_0<-arrange(phcalf16\_0,NUT,TIPO\_BENEF)

phcalf16\_00 <- phcalf16\_0%>%

group\_by(NUT)%>%



```

summarise(last(NUMAFI),sum(SALDO),last(NRO_OPERACION),last(ESTADO_CREDITO),last(FEC_DESSS
),last(ESTADO_AFILIADO),last(CODTIPSOLSER),last(TIPO_BENEF),sum(PORCENTAJE_PARTICIPACION
),last(FECHA_DESEMBOLSO),last(FECHA_CORTE_REPORTE),sum(MONTO_FINANCIADO),last(FECHA_
FIN_CREDITO),sum(AVALUOREALIZACION),last(PLAZO),last(TASA_NOMINAL_ACTUAL),sum(CUOTA_M
ENSUAL),last(ORIGEN),last(BIN_MOR_MAYOR_90),last(FEC_MORA_PRIMERA_VEZ),last(FEC_CAN_MO
R_PRIMERA_VEZ),last(DIAS_MORA_PRIMERA_VEZ),sum(INGRESO_PRECAL),last(ACT_ECO_EMP_PR
ECAL),last(GENERO),last(PROVINCIA),last(CANTON),last(FEC_NACIMIENTO),last(NOMBRES),last(FECH
A_MORA_GRAVE),last(DIAS_MORA_GRAVE))

```

```
colnames(phcalf16_00)<-
```

```

c("NUT","NUMAFI","SALDO","NRO_OPERACION","ESTADO_CREDITO","FEC_DESIESS","ESTADO_AFILI
ADO","CODTIPSOLSER","TIPO_BENEF","PORCENTAJE_PARTICIPACION","FECHA_DESEMBOLSO_BIE
SS","FECHA_CORTE_REPORTE","MONTO_FINANCIADO","FECHA_FIN_CREDITO","AVALUOREALIZACI
ON","PLAZO","TASA_NOMINAL_ACTUAL","CUOTA_MENSUAL","ORIGEN","BIN_MOR_MAYOR_90","FEC_
MORA_PRIMERA_VEZ","FEC_CAN_MOR_PRIMERA_VEZ","DIAS_MORA_PRIMERA_VEZ","INGRESO_PR
ECAL","ACT_ECO_EMP_PRECAL","GENERO","PROVINCIA","CANTON","FEC_NACIMIENTO","NOMBRES
","FECHA_MORA_GRAVE","DIAS_MORA_GRAVE")

```

```
phcalf16 <-subset(phcalf16_00,phcalf16_00$CODTIPSOLSER==35 | phcalf16_00$CODTIPSOLSER==36)
```

```
#recodificación de variables continuas a cualitativas.
```

```
# rango de avaluos de vivienda
```

```
phcalf16$R_Avaluo[phcalf16$AVALUOREALIZACION<=40000] <- "hasta 40.000 USD"
```

```
phcalf16$R_Avaluo[phcalf16$AVALUOREALIZACION<=70000 & phcalf16$AVALUOREALIZACION>40000]
<- "entre 40.001 y 70.000 USD"
```

```
phcalf16$R_Avaluo[phcalf16$AVALUOREALIZACION<=100000 & phcalf16$AVALUOREALIZACION>70000]
<- "entre 70.001 y 100.000 USD"
```

```
phcalf16$R_Avaluo[phcalf16$AVALUOREALIZACION<=150000 &
phcalf16$AVALUOREALIZACION>100000] <- "entre 100.001 y 150.000 USD"
```

```
phcalf16$R_Avaluo[phcalf16$AVALUOREALIZACION>150000] <- "mayor a 150.000 USD"
```

```
phcalf16$R_Avaluo <- as.factor(phcalf16$R_Avaluo)
```

```
levels(phcalf16$R_Avaluo)
```

```
phcalf16$R_Avaluo <- factor(phcalf16$R_Avaluo,levels = levels(phcalf16$R_Avaluo)[c(4,2,3,1,5)])
```

```
# rango de avaluos de vivienda resumido 2
```

```
phcalf16$R_Avaluo2[phcalf16$AVALUOREALIZACION<=40000] <- "hasta 40.000 USD"
```

```
phcalf16$R_Avaluo2[phcalf16$AVALUOREALIZACION<=100000 & phcalf16$AVALUOREALIZACION >
40000] <- "entre 40.000,01 y 100.000 USD"
```

```
phcalf16$R_Avaluo2[phcalf16$AVALUOREALIZACION>100000] <- "mayor a 100.000 USD"
```

```
phcalf16$R_Avaluo2 <- as.factor(phcalf16$R_Avaluo2)
```

```
levels(phcalf16$R_Avaluo2)
```

```
phcalf16$R_Avaluo2 <- factor(phcalf16$R_Avaluo2,levels = levels(phcalf16$R_Avaluo2)[c(2,1,3)])
```

```
levels(phcalf16$R_Avaluo2)
```

```
# rango plazo crédito
```

```

phcalf16$R_Plazo[phcalf16$PLAZO<=240] <- "menor a 20 años"
phcalf16$R_Plazo[phcalf16$PLAZO>240] <- "mayor a 20 años"
phcalf16$R_Plazo <- as.factor(phcalf16$R_Plazo)
levels(phcalf16$R_Plazo)
phcalf16$R_Plazo <- factor(phcalf16$R_Plazo,levels = levels(phcalf16$R_Plazo)[c(2,1)])

# rango ingresos al momento de la pre-calificación
summary(phcalf16$INGRESO_PRECAL)
quantile(phcalf16$INGRESO_PRECAL,prob=c(0.20,0.40,0.60,0.80,1))
phcalf16$INGRESO_PRECALaj <-
ifelse(phcalf16$INGRESO_PRECAL==0,phcalf16$CUOTA_MENSUAL/0.4,phcalf16$INGRESO_PRECAL)
quantile(phcalf16$INGRESO_PRECALaj,prob=c(0.20,0.40,0.60,0.80,1))

phcalf16$R_Ingp[phcalf16$INGRESO_PRECALaj<=770] <- "hasta 770 USD (Q1)"
phcalf16$R_Ingp[phcalf16$INGRESO_PRECALaj<=1133 & phcalf16$INGRESO_PRECALaj>770] <- "entre
770,01 y 1.133 USD (Q2)"
phcalf16$R_Ingp[phcalf16$INGRESO_PRECALaj<=1603 & phcalf16$INGRESO_PRECALaj>1133] <- "entre
1.133,01 y 1.603 USD (Q3)"
phcalf16$R_Ingp[phcalf16$INGRESO_PRECALaj<=2362 & phcalf16$INGRESO_PRECALaj>1603] <- "entre
1.603,01 y 2.362 USD (Q4)"
phcalf16$R_Ingp[phcalf16$INGRESO_PRECALaj>2362] <- "mayor a 2.362 USD (Q5)"
phcalf16$R_Ingp <- as.factor(phcalf16$R_Ingp)
levels(phcalf16$R_Ingp)
phcalf16$R_Ingp <- factor(phcalf16$R_Ingp,levels = levels(phcalf16$R_Ingp)[c(4,3,1,2,5)])

# rango monto financiado

summary(phcalf16$MONTO_FINANCIADO)
quantile(phcalf16$MONTO_FINANCIADO,prob=c(0.20,0.40,0.60,0.80,1))
phcalf16$R_Montf[phcalf16$MONTO_FINANCIADO<=40000] <- "hasta 40.000 USD"
phcalf16$R_Montf[phcalf16$MONTO_FINANCIADO<=70000 & phcalf16$MONTO_FINANCIADO>40000] <-
"entre 40.000,01 y 70.000 USD"
phcalf16$R_Montf[phcalf16$MONTO_FINANCIADO<=100000 & phcalf16$MONTO_FINANCIADO>70000] <-
"entre 70.000,01 y 100.000 USD"
phcalf16$R_Montf[phcalf16$MONTO_FINANCIADO<=150000 & phcalf16$MONTO_FINANCIADO>100000]
<- "entre 100.000,01 y 150.000 USD"
phcalf16$R_Montf[phcalf16$MONTO_FINANCIADO>150000] <- "mayor a 150.000 USD"
phcalf16$R_Montf <- as.factor(phcalf16$R_Montf)
levels(phcalf16$R_Montf)
phcalf16$R_Montf <- factor(phcalf16$R_Montf,levels = levels(phcalf16$R_Montf)[c(4,2,3,1,5)])

# rango monto financiado resumido 2

phcalf16$R_Montf2[phcalf16$MONTO_FINANCIADO<=150000] <- "hasta 150.000 USD"

```

```

phcalf16$R_Montf2[phcalf16$MONTO_FINANCIADO>150000] <- "mayor a 150.000 USD"
phcalf16$R_Montf2 <- as.factor(phcalf16$R_Montf2)
levels(phcalf16$R_Montf2)

# Sub_regiones geográficas
phcalf16$SubReg[phcalf16$PROVINCIA=="ESMERALDAS" | phcalf16$PROVINCIA=="MANABI" |
phcalf16$PROVINCIA=="SANTO DOMINGO DE LOS TSACHILAS" | phcalf16$PROVINCIA=="LOS RIOS" |
phcalf16$PROVINCIA=="SANTA ELENA" | phcalf16$PROVINCIA=="GALAPAGOS" |
phcalf16$PROVINCIA=="GUAYAS" | phcalf16$PROVINCIA=="EL ORO"] <- "COSTA (incluye STO.DOM. y
GAL.)"
phcalf16$SubReg[phcalf16$PROVINCIA=="CARCHI" | phcalf16$PROVINCIA=="IMBABURA" |
phcalf16$PROVINCIA=="PICHINCHA"] <- "SIERRA_N (CAR., IMB., PIC.)"
phcalf16$SubReg[phcalf16$PROVINCIA=="COTOPAXI" | phcalf16$PROVINCIA=="TUNGURAHUA" |
phcalf16$PROVINCIA=="BOLIVAR" | phcalf16$PROVINCIA=="CHIMBORAZO"] <- "SIERRA_C (COT.,
TUNG., BOL.,CHI)"
phcalf16$SubReg[phcalf16$PROVINCIA=="CAÑAR" | phcalf16$PROVINCIA=="AZUAY" |
phcalf16$PROVINCIA=="LOJA"] <- "SIERRA_S (CAÑ., AZU., LOJ.)"
phcalf16$SubReg[phcalf16$PROVINCIA=="SUCUMBOS" | phcalf16$PROVINCIA=="NAPO" |
phcalf16$PROVINCIA=="ORELLANA"] <- "ORIENTE_N (SUC., NAP., ORE.)"
phcalf16$SubReg[phcalf16$PROVINCIA=="PASTAZA" | phcalf16$PROVINCIA=="MORONA SANTIAGO" |
phcalf16$PROVINCIA=="ZAMORA CHINCHIPE"] <- "ORIENTE_CS (PAS,MOR.,ZAM.)"
phcalf16$SubReg <- as.factor(phcalf16$SubReg)
levels(phcalf16$SubReg)
phcalf16$SubReg <- factor(phcalf16$SubReg,levels = levels(phcalf16$SubReg)[c(5,4,6,1,3,2)])

# Sub_regiones geográficas resumidas 2
phcalf16$SubReg2[phcalf16$PROVINCIA=="ESMERALDAS" | phcalf16$PROVINCIA=="MANABI" |
phcalf16$PROVINCIA=="SANTO DOMINGO DE LOS TSACHILAS" | phcalf16$PROVINCIA=="LOS RIOS" |
phcalf16$PROVINCIA=="SANTA ELENA" | phcalf16$PROVINCIA=="GALAPAGOS" |
phcalf16$PROVINCIA=="GUAYAS" | phcalf16$PROVINCIA=="EL ORO" |
phcalf16$PROVINCIA=="SUCUMBOS" | phcalf16$PROVINCIA=="NAPO" |
phcalf16$PROVINCIA=="ORELLANA"] <- "COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE)"
phcalf16$SubReg2[phcalf16$PROVINCIA=="CARCHI" | phcalf16$PROVINCIA=="IMBABURA" |
phcalf16$PROVINCIA=="PICHINCHA" | phcalf16$PROVINCIA=="COTOPAXI" |
phcalf16$PROVINCIA=="TUNGURAHUA" | phcalf16$PROVINCIA=="BOLIVAR" |
phcalf16$PROVINCIA=="CHIMBORAZO" | phcalf16$PROVINCIA=="CAÑAR" |
phcalf16$PROVINCIA=="AZUAY" | phcalf16$PROVINCIA=="LOJA" | phcalf16$PROVINCIA=="PASTAZA" |
phcalf16$PROVINCIA=="MORONA SANTIAGO" | phcalf16$PROVINCIA=="ZAMORA CHINCHIPE"] <-
"OTRAS SUBREGIONES"

phcalf16$SubReg2 <- as.factor(phcalf16$SubReg2)
levels(phcalf16$SubReg2)
phcalf16$SubReg2 <- factor(phcalf16$SubReg2,levels = levels(phcalf16$SubReg2)[c(2,1)])

```

```

# rangos de edad al momento del desembolso
#calcular edad
phcalf16$Edad<-ifelse(is.na(phcalf16$FECHA_DESEMBOLSO_BIESS),as.numeric(floor((
phcalf16$FEC_DESSSPhcalf16$FEC_NACIMIENTO)/365)),as.numeric(floor((phcalf16$FECHA_DESEMBOL
SO-phcalf16$FEC_NACIMIENTO)/365)))
summary(phcalf16$Edad)
quantile(phcalf16$Edad,prob=c(0.20,0.40,0.60,0.80,1))

#calcular rangos de edad
phcalf16$R_Edad[phcalf16$Edad<=35] <- "hasta 35 años"
phcalf16$R_Edad[phcalf16$Edad<=50 & phcalf16$Edad>35] <- "entre 36 y 50 años"
phcalf16$R_Edad[phcalf16$Edad>50] <- "entre 51 y 75 años"
phcalf16$R_Edad <- as.factor(phcalf16$R_Edad)
levels(phcalf16$R_Edad)
phcalf16$R_Edad <- factor(phcalf16$R_Edad,levels = levels(phcalf16$R_Edad)[c(3,1,2)])

# actividad económica recodificada
str(phcalf16$ACT_ECO_EMP_PRECAL)
summary(phcalf16$ACT_ECO_EMP_PRECAL)

phcalf16$Acteco_r[phcalf16$ACT_ECO_EMP_PRECAL=="2-EMPRESA PRIVADA -SOCIEDADES /
COMPANIAS"] <- "EMPRESA PRIVADA, SOCIEDADES / COMPANIAS"
phcalf16$Acteco_r[phcalf16$ACT_ECO_EMP_PRECAL=="3-EMPRESA UNIPERSONAL / PEQUEÑA
INDUSTRIA"] <- "EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA"
phcalf16$Acteco_r[phcalf16$ACT_ECO_EMP_PRECAL=="1- INSTITUCION FINANCIERA PRIVADA,
BANCOS Y CIA DE SEGUROS22.50%" | phcalf16$ACT_ECO_EMP_PRECAL=="13-ENTIDADES
FINANCIERAS" | phcalf16$ACT_ECO_EMP_PRECAL=="26-COOP AHORRO CRED. VIVIENDA
ETC.CONTR SUPER DE BANCOS 20.50%" | phcalf16$ACT_ECO_EMP_PRECAL=="28-OTRAS FINANC.
PRIVAD, MUTUALISTAS ETC.20.50%" | phcalf16$ACT_ECO_EMP_PRECAL=="8-ARTESANAL" |
phcalf16$ACT_ECO_EMP_PRECAL=="25-EMPLEADOR DOMESTICO" |
phcalf16$ACT_ECO_EMP_PRECAL=="17-ORGANIZACION CAMPESINA" |
phcalf16$ACT_ECO_EMP_PRECAL=="24-MIEMBROS DE SINDICATOS Y COOP DE TRANSPORTE
TERRESTRE" | phcalf16$ACT_ECO_EMP_PRECAL=="27-COOPERA NO CONTROLADAS POR LA
SUPERINTENDENCIA DE BANCOS" | phcalf16$ACT_ECO_EMP_PRECAL=="29-EMBAJADAS, MISIONES
DIPL., CONSULARES, ORG. INTERNACIONALES" | phcalf16$ACT_ECO_EMP_PRECAL=="5-CURIA
DIOSESANA" | phcalf16$ACT_ECO_EMP_PRECAL=="18-CONTINUACION VOLUNTARIA"] <- "OTROS"
phcalf16$Acteco_r[phcalf16$ACT_ECO_EMP_PRECAL=="10-ORGANISMOS DEL REGIMEN SECCIONAL"]
<- "ENTIDADES SECTOR PÚBLICO - SECCIONALES"
phcalf16$Acteco_r[phcalf16$ACT_ECO_EMP_PRECAL=="14-FUNCION EJECUTIVA, LEGISL, JUDICIAL Y
ORGANISMOS DEL ESTADO" | phcalf16$ACT_ECO_EMP_PRECAL=="9-NOTARIAS, (Los Registradores
son Públicos) " | phcalf16$ACT_ECO_EMP_PRECAL=="12-ENTIDADES PUBLICAS CON AUTONOMIA" |
phcalf16$ACT_ECO_EMP_PRECAL=="12-ENTIDADES PUBLICAS CON AUTONOMIA " |
phcalf16$ACT_ECO_EMP_PRECAL=="16-ENTIDADES PÚBLICAS DE EDUCACIÓN SUPERIOR"] <-
"ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS"

```

```

phcalf16$Acteco_r[phcalf16$ACT_ECO_EMP_PRECAL=="6-CONSTRUCCION"] <- "CONSTRUCCIÓN"
phcalf16$Acteco_r[is.na(phcalf16$Acteco_r)] <- "SIN INFORMACIÓN"
phcalf16$Acteco_r[phcalf16$ACT_ECO_EMP_PRECAL=="4-AGRICOLA SEGURO GENERAL" |
phcalf16$ACT_ECO_EMP_PRECAL=="7-AGRICOLA LABORES DE CAMPO"] <- "AGRÍCOLA"
phcalf16$Acteco_r[phcalf16$ACT_ECO_EMP_PRECAL=="30-TRABAJADORES AUTONOMOS Y SIN
RELACIÓN DE DEPENDENCIA " | phcalf16$ACT_ECO_EMP_PRECAL=="31- MICRO EMPRESA (NO
ARTESANAL)" | phcalf16$ACT_ECO_EMP_PRECAL=="32-AFILIACION VOLUNTARIA(TIPEM-32)" |
phcalf16$ACT_ECO_EMP_PRECAL=="34-RISE-RÉGIMEN IMPOSITIVO SIMPLIFICADO"] <- "TRABAJO
AUTÓNOMO, RISE Y MICROEMPRESA"
phcalf16$Acteco_r<- as.factor(phcalf16$Acteco_r)

```

```

# actividad económica recodificada resumida 2

```

```

phcalf16$Acteco_r2[phcalf16$ACT_ECO_EMP_PRECAL=="2-EMPRESA PRIVADA -SOCIEDADES /
COMPANIAS" | phcalf16$ACT_ECO_EMP_PRECAL=="1- INSTITUCION FINANCIERA PRIVADA, BANCOS
Y CIA DE SEGUROS22.50%" | phcalf16$ACT_ECO_EMP_PRECAL=="13-ENTIDADES FINANCIERAS" |
phcalf16$ACT_ECO_EMP_PRECAL=="26-COOP AHORRO CRED. VIVIENDA ETC.CONTR SUPER DE
BANCOS 20.50%" | phcalf16$ACT_ECO_EMP_PRECAL=="28-OTRAS FINANC. PRIVAD, MUTUALISTAS
ETC.20.50%" | phcalf16$ACT_ECO_EMP_PRECAL=="8-ARTESANAL" |
phcalf16$ACT_ECO_EMP_PRECAL=="25-EMPLEADOR DOMESTICO" |
phcalf16$ACT_ECO_EMP_PRECAL=="17-ORGANIZACION CAMPESINA" |
phcalf16$ACT_ECO_EMP_PRECAL=="24-MIEMBROS DE SINDICATOS Y COOP DE TRANSPORTE
TERRESTRE" | phcalf16$ACT_ECO_EMP_PRECAL=="27-COOPERA NO CONTROLADAS POR LA
SUPERINTENDENCIA DE BANCOS" | phcalf16$ACT_ECO_EMP_PRECAL=="29-EMBAJADAS, MISIONES
DIPL., CONSULARES, ORG. INTERNACIONALES" | phcalf16$ACT_ECO_EMP_PRECAL=="5-CURIA
DIOSESANA" | phcalf16$ACT_ECO_EMP_PRECAL=="18-CONTINUACION VOLUNTARIA" |
phcalf16$ACT_ECO_EMP_PRECAL=="10-ORGANISMOS DEL REGIMEN SECCIONAL" |
phcalf16$ACT_ECO_EMP_PRECAL=="14-FUNCION EJECUTIVA, LEGISL, JUDICIAL Y ORGANISMOS
DEL ESTADO" | phcalf16$ACT_ECO_EMP_PRECAL=="9-NOTARIAS, (Los Regitradores son Públicos) " |
phcalf16$ACT_ECO_EMP_PRECAL=="12-ENTIDADES PUBLICAS CON AUTONOMIA" |
phcalf16$ACT_ECO_EMP_PRECAL=="12-ENTIDADES PUBLICAS CON AUTONOMIA " |
phcalf16$ACT_ECO_EMP_PRECAL=="16-ENTIDADES PÚBLICAS DE EDUCACIÓN SUPERIOR" |
phcalf16$ACT_ECO_EMP_PRECAL=="4-AGRICOLA SEGURO GENERAL" |
phcalf16$ACT_ECO_EMP_PRECAL=="7-AGRICOLA LABORES DE CAMPO"] <- "OTRAS ACTIVIDADES"
phcalf16$Acteco_r2[phcalf16$ACT_ECO_EMP_PRECAL=="3-EMPRESA UNIPERSONAL / PEQUEÑA
INDUSTRIA" | phcalf16$ACT_ECO_EMP_PRECAL=="6-CONSTRUCCION" |
phcalf16$ACT_ECO_EMP_PRECAL=="30-TRABAJADORES AUTONOMOS Y SIN RELACIÓN DE
DEPENDENCIA " | phcalf16$ACT_ECO_EMP_PRECAL=="31- MICRO EMPRESA (NO ARTESANAL)" |
phcalf16$ACT_ECO_EMP_PRECAL=="32-AFILIACION VOLUNTARIA(TIPEM-32)" |
phcalf16$ACT_ECO_EMP_PRECAL=="34-RISE-RÉGIMEN IMPOSITIVO SIMPLIFICADO" |
is.na(phcalf16$Acteco_r2)] <- "EMPRESA UNIPERSONAL/ CONSTRUCCIÓN/TRABAJO AUTÓNOMO, RISE
Y MICROEMPRESA/SIN INFORMACIÓN"
phcalf16$Acteco_r2<- as.factor(phcalf16$Acteco_r2)
levels(phcalf16$Acteco_r2)
phcalf16$Acteco_r2 <- factor(phcalf16$Acteco_r2,levels = levels(phcalf16$Acteco_r2)[c(2,1)])

```

```

# ciclo económico al momento del desembolso
phcalf16$Cicloe[phcalf16$FECHA_DESEMBOLSO <="2011-04-30" &
phcalf16$FECHA_DESEMBOLSO_BIESS >"2010-05-31"] <- "Recuperación"
phcalf16$Cicloe[phcalf16$FECHA_DESEMBOLSO <="2011-12-31" &
phcalf16$FECHA_DESEMBOLSO_BIESS >"2011-04-30"] <- "Expansión"
phcalf16$Cicloe[phcalf16$FECHA_DESEMBOLSO <="2013-05-31" &
phcalf16$FECHA_DESEMBOLSO_BIESS >"2011-12-31"] <- "Desaceleración sobre la tendencia"
phcalf16$Cicloe[phcalf16$FECHA_DESEMBOLSO <="2014-10-31" &
phcalf16$FECHA_DESEMBOLSO_BIESS >"2013-05-31"] <- "Expansión"
phcalf16$Cicloe[phcalf16$FECHA_DESEMBOLSO <="2015-07-31" &
phcalf16$FECHA_DESEMBOLSO_BIESS >"2014-10-31"] <- "Desaceleración sobre la tendencia"
phcalf16$Cicloe[phcalf16$FECHA_DESEMBOLSO <="2016-05-31" & phcalf16$FECHA_DESEMBOLSO
>"2015-07-31"] <- "Desaceleración bajo la tendencia"
phcalf16$Cicloe[phcalf16$FECHA_DESEMBOLSO <="2017-02-28" &
phcalf16$FECHA_DESEMBOLSO_BIESS >"2016-05-31"] <- "Recuperación"

phcalf16$Cicloe[phcalf16$FEC_DESSS <="2008-09-30" & phcalf16$FEC_DESSS >"2008-01-31"] <-
"Expansión"
phcalf16$Cicloe[phcalf16$FEC_DESSS <="2009-05-31" & phcalf16$FEC_DESSS >"2008-09-30"] <-
"Desaceleración sobre la tendencia"
phcalf16$Cicloe[phcalf16$FEC_DESSS <="2010-05-31" & phcalf16$FEC_DESSS >"2009-05-31"] <-
"Desaceleración bajo la tendencia"
phcalf16$Cicloe[phcalf16$FEC_DESSS <="2011-04-30" & phcalf16$FEC_DESSS >"2010-05-31"] <-
"Recuperación"
phcalf16$Cicloe<- as.factor(phcalf16$Cicloe)
levels(phcalf16$Cicloe)
phcalf16$Cicloe <- factor(phcalf16$Cicloe,levels = levels(phcalf16$Cicloe)[c(3,2,1,4)])

# ciclo económico al momento del desembolso resumido 2
phcalf16$Cicloe2[phcalf16$Cicloe=="Recuperación" | phcalf16$Cicloe=="Desaceleración bajo la tendencia"]
<- 1
phcalf16$Cicloe2[phcalf16$Cicloe=="Expansión" | phcalf16$Cicloe=="Desaceleración sobre la tendencia"] <-
0
phcalf16$Cicloe2<- as.factor(phcalf16$Cicloe2)
levels(phcalf16$Cicloe2)

# Fecha de mora grave completada Na
phcalf16$FECHA_MORA_GRAVEaj <-phcalf16$FECHA_MORA_GRAVE
phcalf16$FECHA_MORA_GRAVEaj[is.na(phcalf16$FECHA_MORA_GRAVEaj)] <- "2016-12-31"
str(phcalf16$FECHA_MORA_GRAVEaj)

# Fecha de inicio estudio
phcalf16$FEC_INICIO <- phcalf16$FECHA_CORTE_REPORTE

```

```
phcalf16$FEC_INICIO[phcalf16$FEC_INICIO=="2016-12-31"] <- "2008-07-30"
str(phcalf16$FEC_INICIO)
```

```
# ciclo económico al momento de la muerte
```

```
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2008-09-30" &
phcalf16$FECHA_MORA_GRAVEaj >"2008-01-31"] <- "Expansión"
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2009-05-31" &
phcalf16$FECHA_MORA_GRAVEaj >"2008-09-30"] <- "Desaceleración sobre la tendencia"
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2010-05-31" &
phcalf16$FECHA_MORA_GRAVEaj >"2009-05-31"] <- "Desaceleración bajo la tendencia"
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2011-04-30" &
phcalf16$FECHA_MORA_GRAVEaj >"2010-05-31"] <- "Recuperación"
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2011-12-31" &
phcalf16$FECHA_MORA_GRAVEaj >"2011-04-30"] <- "Expansión"
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2013-05-31" &
phcalf16$FECHA_MORA_GRAVEaj >"2011-12-31"] <- "Desaceleración sobre la tendencia"
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2014-10-31" &
phcalf16$FECHA_MORA_GRAVEaj >"2013-05-31"] <- "Expansión"
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2015-07-31" &
phcalf16$FECHA_MORA_GRAVEaj >"2014-10-31"] <- "Desaceleración sobre la tendencia"
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2016-05-31" &
phcalf16$FECHA_MORA_GRAVEaj >"2015-07-31"] <- "Desaceleración bajo la tendencia"
phcalf16$Cicloef1[phcalf16$FECHA_MORA_GRAVEaj <="2017-02-28" &
phcalf16$FECHA_MORA_GRAVEaj >"2016-05-31"] <- "Recuperación"
phcalf16$Cicloef1 <- as.factor(phcalf16$Cicloef1)
levels(phcalf16$Cicloef1)
phcalf16$Cicloef1 <- factor(phcalf16$Cicloef1,levels = levels(phcalf16$Cicloef1)[c(3,2,1,4)])
```

```
# ciclo económico al momento de la muerte resumido 2
```

```
phcalf16$Cicloef2[phcalf16$Cicloef1=="Recuperación" | phcalf16$Cicloef1=="Desaceleración bajo la
tendencia"] <- 1
phcalf16$Cicloef2[phcalf16$Cicloef1=="Expansión" | phcalf16$Cicloef1=="Desaceleración sobre la
tendencia"] <- 0
```

```
# variable cesante al momento de la muerte
```

```
phcalf16$cesante <- ifelse(phcalf16$ESTADO_AFILIADO=="CESANTE",1,0)
```

```
# variable año de desembolso
```

```
phcalf16$year <- ifelse(is.na(phcalf16$FECHA_DESEMBOLSO_BIESS),year(phcalf16$FEC_DESIESS),
year(phcalf16$FECHA_DESEMBOLSO_BIESS))
phcalf16$year <- as.factor(phcalf16$year)
```

```

# variable tiempo en función del año
phcalf16$tyear[phcalf16$year==2008] <- 1
phcalf16$tyear[phcalf16$year==2009] <- 2
phcalf16$tyear[phcalf16$year==2010] <- 3
phcalf16$tyear[phcalf16$year==2011] <- 4
phcalf16$tyear[phcalf16$year==2012] <- 5
phcalf16$tyear[phcalf16$year==2013] <- 6
phcalf16$tyear[phcalf16$year==2014] <- 7
phcalf16$tyear[phcalf16$year==2015] <- 8
phcalf16$tyear[phcalf16$year==2016] <- 9

# variable cesante x ciclo econonómico al final de la vida
phcalf16$cesantexciclof<- phcalf16$cesante*phcalf16$Cicloef2
phcalf16$Cicloef2 <- as.factor(phcalf16$Cicloef2)
phcalf16$cesantexciclof <- as.factor(phcalf16$cesantexciclof)
phcalf16$cesante <- as.factor(phcalf16$cesante)

#variable para separar créditos con al menos dos años de vida
phcalf16$mas2a <- ifelse(phcalf16$FECHA_DESEMBOLSO <="2014-12-31" |
is.na(phcalf16$FECHA_DESEMBOLSO),1,0)

#Crearnueva variable binaria para supervivencia, evento o censurado (evento mora mayor a 90 días)
phcalf16$Mayor90<-ifelse(is.na(phcalf16$FECHA_MORA_GRAVE),0,1)

#Crear nueva variable con el tiempo de supervivencia para casos no censurados (evento mora mayor a 90
días).
phcalf16$tiempoM90pv<-ifelse(phcalf16$Mayor90==1 & is.na(phcalf16$FECHA_DESEMBOLSO),
as.numeric(phcalf16$FECHA_MORA_GRAVE-phcalf16$FEC_DESSS),
as.numeric(phcalf16$FECHA_MORA_GRAVE-phcalf16$FECHA_DESEMBOLSO))

#Crear nueva variable con el tiempo de supervivencia para todos los casos en días (evento mora mayor a 90
días).
phcalf16$tiempoM90pv_2<-
ifelse(phcalf16$Mayor90==1,phcalf16$tiempoM90pv,ifelse(is.na(phcalf16$tiempoM90pv) &
is.na(phcalf16$FECHA_DESEMBOLSO), as.numeric(phcalf16$FECHA_CORTE_REPORTE-
phcalf16$FEC_DESSS),as.numeric(phcalf16$FECHA_CORTE_REPORTE-
phcalf16$FECHA_DESEMBOLSO)))

#Crear nueva variable con el tiempo de supervivencia para todos los casos en meses (evento mora mayor a
90 días).
phcalf16$tiempoM90pv_2mes <-as.numeric(phcalf16$tiempoM90pv_2/30)

#variables adicionales para calibración. Evento (mora mayor a 120 días)

```



```

phcalf16$DIAS_MORA_GRAVE_AJUST<-ifelse(phcalf16$Mayor90==1
is.na(phcalf16$DIAS_MORA_GRAVE),as.numeric(phcalf16$FECHA_CORTE_REPORTE-
phcalf16$FECHA_MORA_GRAVE),phcalf16$DIAS_MORA_GRAVE)
phcalf16$Mayor120 <- ifelse(phcalf16$DIAS_MORA_GRAVE_AJUST<=120
is.na(phcalf16$DIAS_MORA_GRAVE_AJUST),0,1)
phcalf16$tiempoM120pv_2<-ifelse(phcalf16$Mayor120==1,
as.numeric(phcalf16$tiempoM90pv_2+30),ifelse(is.na(phcalf16$FECHA_DESEMBOLSO),
as.numeric(phcalf16$FECHA_CORTE_REPORTE-
phcalf16$FEC_DESSS),as.numeric(phcalf16$FECHA_CORTE_REPORTE-
phcalf16$FECHA_DESEMBOLSO)))
phcalf16$tiempoM120pv_2mes <-as.numeric(phcalf16$tiempoM120pv_2/30)

```

\*\*\* Análisis de Supervivencia (créditos con al menos 2 años de vida)

#Sub-base exclusivo créditos con mas de 2 años de vida

```
phcalf16_m2a <- subset(phcalf16,phcalf16$mas2a==1)
```

#### **Anexo IV – Código de programación R – Etapa 2 - Análisis descriptivo de las potenciales variables explicativas**

\*\*\* Análisis de Supervivencia (créditos con al menos 2 años de vida)

#Sub-base exclusivo créditos con mas de 2 años de vida

```
phcalf16_m2a <- subset(phcalf16,phcalf16$mas2a==1)
```

#calcular función de supervivencia global por el método Kaplan y Meier, se utiliza la función Surv(Objeto con variable tipo análisis de supervivencia)

```
funsuper<- survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~1)
```

```
summary(funsuper)
```

```
plot(funsuper,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia general, método Kaplan y Meier",xlim=c(0,216),cex.main=1)
```

#calcular función de supervivencia K y M según rango de avaluo 2

#tabla R avaluo 2

```
tableR_avaluo2 <- CrossTable(phcalf16_m2a$R_Avaluo2,phcalf16_m2a$Mayor90,expected = FALSE,digits = 1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
```

#distribución diferente entre opciones de respuesta variable rango de avaluo2.

```
funsuper_R_avaluo2 <- survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~phcalf16_m2a$R_Avaluo2)
```

```
print(funsuper_R_avaluo2)
```

```
plot(funsuper_R_avaluo2,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método Kaplan y Meier, variable: rango avalúo 2",lty=c(1,2,3),mark.time=FALSE,col=c("blue","red","black"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.75,cex.lab=0.75)
```

```

legend(60,0.65,legend=c("hasta 40.000 USD","entre 40.000,01 y 100.000 USD","mayor a 100.000
USD"),lty=c(1,2,3),col=c("blue","red","black"),text.width = 65,cex = 0.7)
#comparar funciones K y M entre categorías rango de avalúo 2
compsuper_R_avaluo2<-          survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$R_Avaluo2)
print(compsuper_R_avaluo2)
#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia
(significancia del 5%)

#calcular función de supervivencia K y M según género
#tabla género
table_genero <- CrossTable(phcalf16_m2a$GENERO,phcalf16_m2a$Mayor90,expected = FALSE,digits =
1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
#distribución diferente entre opciones de respuesta variable género
funsuper_genero<-          survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$GENERO)
print(funsuper_genero)
plot(funsuper_genero,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método
Kaplan y Meier, variable: género",lty=c(1,2),mark.time=FALSE,col =
c("blue","red"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.75,cex.lab=0.75)
legend(60,0.65,legend=c("FEMENINO","MASCULINO"),lty=c(1,2),col = c("blue","red"),text.width = 25,cex =
0.7)
#comparar funciones KM global entre opciones de género
compsuper_genero<-          survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$GENERO)
print(compsuper_genero)
#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia
(significancia del 5%)

#calcular función de supervivencia K y M según rango de plazo
#tabla rango de plazo
table_R_plazo <- CrossTable(phcalf16_m2a$R_Plazo,phcalf16_m2a$Mayor90,expected = FALSE,digits =
1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
#distribución diferente entre opciones de respuesta variable rango de plazo
funsuper_R_plazo<-          survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$R_Plazo)
print(funsuper_R_plazo)
plot(funsuper_R_plazo,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método
Kaplan y Meier, variable: rango plazo",lty=c(1,2),mark.time=FALSE,col=c("blue","red"),xlim=c(0,150),
ylim=c(0.5,1),cex.main=0.75,cex.lab=0.75)
legend(60,0.70,legend=c("menor a 20 años","mayor a 20 años"),lty=c(1,2),col=c("blue","red"),text.width =
50,cex = 0.7)
#comparar funciones K y M global entre categorías rango de plazo

```

```

compsuper_R_plazo<-          survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$R_Plazo)
print(compsuper_R_plazo)
#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia
(significancia del 5%)

#calcular función de supervivencia K y M según rango de ingreso
#tabla rango de ingreso
table_R_Ingp <- CrossTable(phcalf16_m2a$R_Ingp,phcalf16_m2a$Mayor90,expected = FALSE,digits =
1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
#distribución diferente entre opciones de respuesta variable rango de ingreso
funsuper_R_Ingp<-          survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$R_Ingp)
print(funsuper_R_Ingp)
plot(funsuper_R_Ingp,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método
Kaplan y Meier, variable: rango ingreso",lty=c(1,2,3,4,5),mark.time=FALSE,
col=c("blue","red","black","green","purple"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.75,cex.lab=0.75)
legend(60,0.70,legend=c("Q1(hasta 767 USD )","Q2(entre 767,01 y 1.133 USD)","Q3(entre 1.133,01 y 1.603
USD)","Q4(entre 1.603,01 y 2.362 USD)","Q5(mayor a 2.362
USD)"),lty=c(1,2,3,4,5),col=c("blue","red","black","green","purple"),text.width = 70,cex=0.70)
#comparar funciones K y M global entre categorías rango de ingresos
compsuper_R_Ingp<-          survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$R_Ingp)
print(compsuper_R_Ingp)
#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia
(significancia del 5%)

#calcular función de supervivencia K y M según rango de monto financiado
#tabla rango de monto financiado
table_R_Montf <- CrossTable(phcalf16_m2a$R_Montf,phcalf16_m2a$Mayor90,expected = FALSE,digits =
1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
#distribución diferente entre opciones de respuesta variable rango de monto financiado
funsuper_R_Montf<-          survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$R_Montf)
print(funsuper_R_Montf)
plot(funsuper_R_Montf,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método
Kaplan y Meier, variable: rango monto financiado",lty=c(1,2,3,4,5),mark.time=FALSE,
col=c("blue","red","black","green","purple"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.65,cex.lab=0.65)
legend(60,0.70,legend=c("hasta 40.000 USD","entre 40.000,01 y 70.000 USD","entre 70.000,01 y 100.000
USD","entre 100.000,01 y 150.000
USD","mayor a 150.000
USD"),lty=c(1,2,3,4,5),col=c("blue","red","black","green","purple"),text.width = 70,cex = 0.65)
#comparar funciones K y M entre categorías rango de monto financiado
compsuper_R_Montf<-          survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$R_Montf)

```

```

print(compsuper_R_Montf)
#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia
(significancia del 5%)

#calcular función de supervivencia K y M según subregiones geográficas
#tabla subregiones geográficas
table_SubReg <- CrossTable(phcalf16_m2a$SubReg,phcalf16_m2a$Mayor90,expected = FALSE,digits =
1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
#distribución diferente entre opciones de respuesta variable subregiones geográficas
funsuper_SubReg<-          survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$SubReg)
print(funsuper_SubReg)
plot(funsuper_SubReg,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método
Kaplan y Meier, variable: sub-región geográfica",lty=c(1,2,3,4,5,6),mark.time=FALSE,
col=c("blue","red","black","green","purple","brown"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.65,cex.lab=0.65)
legend(5,0.68,legend=c("SIERRA_N (CAR., IMB., PIC.)","SIERRA_C (COT., TUNG., BOL.,CHI)","SIERRA_S
(CAÑ., AZU., LOJ.)","COSTA (incluye STO.DOM. y GAL.)","ORIENTE_N (SUC., NAP.,
ORE.)","ORIENTE_CS
(PAS,MOR.,ZAM.)"),lty=c(1,2,3,4,5,6),col=c("blue","red","black","green","purple","brown"),text.width = 57,cex =
0.5)
#comparar funciones K y M entre subregiones geográficas
compsuper_SubReg<-          survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$SubReg)
print(compsuper_SubReg)
#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia
(significancia del 5%)

#calcular función de supervivencia K y M según rango de edad
#tabla rangos de edad
table_R_Edad <- CrossTable(phcalf16_m2a$R_Edad,phcalf16_m2a$Mayor90,expected = FALSE,digits =
1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
#distribución diferente entre opciones de respuesta variable rango de edad
funsuper_R_Edad<-
survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~phcalf16_m2a$R_Edad)
print(funsuper_R_Edad)
plot(funsuper_R_Edad,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método
Kaplan y Meier, variable: rango de edad",lty=c(1,2,3),mark.time=FALSE,col=c("blue","red","black"),
xlim=c(0,150),ylim=c(0.5,1),cex.main=0.75,cex.lab=0.75)
legend(60,0.65,legend=c("hasta 35 años","entre 36 y 50 años","entre 51 y 75
años"),lty=c(1,2,3),col=c("blue","red","black"),text.width = 60,cex = 0.7)
#comparar funciones K y M entre categorías rango de edad
compsuper_R_Edad<-          survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$R_Edad)
print(compsuper_R_Edad)

```

#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia (significancia del 5%)

#calcular función de supervivencia K y M según actividad económica del empleador

#tabla actividad económica del empleador reducida

```
table_Acteco_r<- CrossTable(phcalf16_m2a$Acteco_r,phcalf16_m2a$Mayor90,expected = FALSE,digits = 1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
```

#distribución diferente entre opciones de respuesta variable actividad económica

```
funsuper_Acteco_r<- survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~ phcalf16_m2a$Acteco_r)
```

```
print(funsuper_Acteco_r)
```

```
plot(funsuper_Acteco_r,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método Kaplan y Meier, variable: actividad económica emp.",lty=c(1,2,3,4,5,6,7,8,9),mark.time=FALSE,col=c("blue","red","black","green","purple","brown","pink","gray","darkgreen"),xlim=c(0,150),ylim=c(0,1),cex.main=0.65,cex.lab=0.65)
```

```
legend(5,0.40,legend=c("AGRÍCOLA","CONSTRUCCIÓN","EMPRESA PRIVADA, SOCIEDADES / COMPANIAS","EMPRESA UNIPERSONAL / PEQUEÑA INDUSTRIA","ENTIDADES SECTOR PÚBLICO - ESTADO CENTRAL Y OTROS","ENTIDADES SECTOR PÚBLICO - SECCIONALES","OTROS","SIN INFORMACIÓN","TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA"),lty=c(1,2,3,4,5,6,7,8,9),col=c("blue","red","black","green","purple","brown","pink","gray","darkgreen"),text.width = 75,cex = 0.4)
```

#comparar funciones K y M entre opciones de actividad económica del empleador

```
compsuper_Acteco_r<- survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~ phcalf16_m2a$Acteco_r)
```

```
print(compsuper_Acteco_r)
```

#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia (significancia del 5%)

#calcular función de supervivencia K y M según ciclo de la economía al momento de la precalificación

#tabla ciclo de la economía al momento de la precalificación

```
table_Ciclo_e<- CrossTable(phcalf16_m2a$Ciclo_e,phcalf16_m2a$Mayor90,expected = FALSE,digits = 1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
```

#distribución diferente entre opciones de respuesta variable ciclo económico al momento de la precalificación

```
funsuper_Ciclo_e<- survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~ phcalf16_m2a$Ciclo_e)
```

```
print(funsuper_Ciclo_e)
```

```
plot(funsuper_Ciclo_e,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método Kaplan y Meier, variable: ciclo de la economía inicio",lty=c(1,2,3,4),mark.time=FALSE,col = c("blue","red","black","green"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.65,cex.lab=0.65)
```

```
legend(5,0.65,legend=c("Desaceleración sobre la tendencia", "Desaceleración bajo la tendencia", "Recuperación", "Expansión"),lty=c(1,2,3,4),col = c("blue","red","black","green"),text.width = 60,cex=0.5)
```

#comparar funciones KM entre categorías ciclo económico al momento de la precalificación

```
compsuper_Ciclo_e<- survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~ phcalf16_m2a$Ciclo_e)
```

```
print(compsuper_Ciclo_e)
```

#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia (significancia del 5%)

#calcular función de supervivencia K y M según ciclo de la economía al momento de la precalificación resumida

#tabla ciclo de la economía al momento de la precalificación resumida

```
table_Ciclo2<- CrossTable(phcalf16_m2a$Ciclo2,phcalf16_m2a$Mayor90,expected = FALSE,digits = 1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
```

#distribución diferente entre opciones de respuesta variable ciclo económico al momento de la precalificación resumido

```
funsuper_Ciclo2<- survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~ phcalf16_m2a$Ciclo2)
```

```
print(funsuper_Ciclo2)
```

```
plot(funsuper_Ciclo2,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método Kaplan y Meier, variable: ciclo de la economía inicio resum.",lty=c(1,2),mark.time=FALSE,col = c("blue","red"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.65,cex.lab=0.65)
```

```
legend(1,0.65,legend=c("Desempeño sobre crecimiento promedio de largo plazo","Desempeño bajo crecimiento promedio de largo plazo"),lty=c(1,2),col = c("blue","red"),text.width = 70,cex=0.5)
```

#comparar funciones KM entre categorías ciclo económico al momento de la precalificación resumida

```
compsuper_Ciclo<- survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~ phcalf16_m2a$Ciclo)
```

```
print(compsuper_Ciclo)
```

#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia (significancia del 5%)

#calcular función de supervivencia K y M según ciclo de la economía previo a la mora grave

#tabla ciclo de la economía previo a la mora grave

```
table_Cicloef1<- CrossTable(phcalf16_m2a$Cicloef1,phcalf16_m2a$Mayor90,expected = FALSE,digits = 1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
```

#distribución diferente entre opciones de respuesta variable ciclo económico al momento de la precalificación resumido

```
funsuper_Cicloef1<- survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~ phcalf16_m2a$Cicloef1)
```

```
print(funsuper_Cicloef1)
```

```
plot(funsuper_Cicloef1,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método Kaplan y Meier, variable: ciclo de la economía previo mora grave",lty=c(1,2,3,4),mark.time=FALSE,col = c("blue","red","black","green"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.65,cex.lab=0.65)
```

```
legend(60,0.65,legend=c("Expansión","Desaceleración sobre la tendencia","Desaceleración bajo la tendencia","Recuperación"),lty=c(1,2,3,4),col = c("blue","red","black","green"),text.width = 50,cex=0.5)
```

#comparar funciones KM entre categorías ciclo económico previo a la mora grave

```
compsuper_Cicloef1<- survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~ phcalf16_m2a$Cicloef1)
```

```
print(compsuper_Cicloef1)
```

#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia (significancia del 5%)

```
#calcular función de supervivencia K y M según ciclo de la economía previo a la mora grave resumido
#tabla ciclo de la economía previo a la mora grave resumido
table_Cicloef2<- CrossTable(phcalf16_m2a$Cicloef2,phcalf16_m2a$Mayor90,expected = FALSE,digits =
1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
#distribución diferente entre opciones de respuesta variable ciclo económico previo a la mora grave resumido
funsuper_Cicloef2<-          survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$Cicloef2)
print(funsuper_Cicloef2)
plot(funsuper_Cicloef2,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método
Kaplan y Meier, variable: ciclo de la economía previo mora grave resumido",lty=c(1,2),mark.time=FALSE,col =
c("blue","red"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.65,cex.lab=0.65)
legend(55,0.65,legend=c("Desempeño sobre crecimiento promedio de largo plazo","Desempeño bajo
crecimiento promedio de largo plazo"),lty=c(1,2),col = c("blue","red"),text.width = 77,cex=0.5)
#comparar funciones KM entre categorías ciclo económico previo a la mora grave resumido
compsuper_Cicloef2<-          survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$Cicloef2)
print(compsuper_Cicloef2)
#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia
(significancia del 5%)
```

```
#calcular función de supervivencia K y M según condición de cesante previo a la mora grave
#tabla cesante previo a la mora grave
table_cesante<- CrossTable(phcalf16_m2a$cesante,phcalf16_m2a$Mayor90,expected = FALSE,digits =
1,chisq = TRUE,prop.chisq = FALSE,prop.c = TRUE,prop.t = FALSE,format = "SPSS")
#distribución diferente entre opciones de respuesta variable cesante previo a la mora grave
funsuper_cesante<-          survfit(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$cesante)
print(funsuper_cesante)
plot(funsuper_cesante,xlab="Mes",ylab="Supervivencia (prob.)",main="Función de supervivencia, método
Kaplan y Meier, variable: cesante previo mora grave",lty=c(1,2),mark.time=FALSE,col =
c("blue","red"),xlim=c(0,150),ylim=c(0.5,1),cex.main=0.65,cex.lab=0.65)
legend(55,0.65,legend=c("No cesante (0)","Cesante (1)"),lty=c(1,2),col = c("blue","red"),text.width =
60,cex=0.5)
#comparar funciones KM entre categorías No cesante y cesante previo a la mora grave
compsuper_cesante<-          survdiff(Surv(phcalf16_m2a$tiempoM90pv_2mes,phcalf16_m2a$Mayor90)~
phcalf16_m2a$cesante)
print(compsuper_cesante)
#dado que p valor es menor a 0,05 se rechaza la hipótesis nula de igualdad de funciones de supervivencia
(significancia del 5%)
```

## Anexo V – Código de programación R – Etapa 3 - Identificación de variables explicativas relacionadas en mayor medida con la variable dependiente

```
#árbol de decisión
set.seed(2)
arbol <- rpart(phcalf16_m2a$Mayor90~phcalf16_m2a$R_Avaluo2+phcalf16_m2a$GENERO+
phcalf16_m2a$R_Plazo+phcalf16_m2a$R_Ingp+phcalf16_m2a$R_Montf+phcalf16_m2a$SubReg2+phcalf16
_m2a$R_Edad+phcalf16_m2a$Acteco_r2+phcalf16_m2a$Cicloef+phcalf16_m2a$cesante,method = "class",
data = phcalf16_m2a)
print(arbol)
par(mfrow=c(1,1))
rpart.plot(arbol,extra = 4,cex = 0.75,cex.main=0.75)
printcp(arbol)
plotcp(arbol)
parbol <- prune(arbol,cp=0.010000)
printcp(parbol)
rpart.plot(parbol,extra = 4)
#El modelo con dos ramas considera las variables: cesante previo a la mora grave y subregión geográfica. No
fue necesario podar.
```

## Anexo VI - Código de programación R – Etapa 4 - Formulación y validación del modelo de cox

```
#Modelo de Cox con base actividad económica y al menos 2 años de vida.
#Sub-base créditos actividad económica y al menos 2 años de vida
phcalf16_actg <- subset(phcalf16_m2a,phcalf16_m2a$Acteco_r2=="EMPRESA UNIPERSONAL/
CONSTRUCCIÓN/TRABAJO AUTÓNOMO, RISE Y MICROEMPRESA/SIN INFORMACIÓN")

#Modelo de cox
coxsuper_actg <- coxph(Surv(phcalf16_actg$tiempoM90pv_2mes,phcalf16_actg$Mayor90)~
phcalf16_actg$R_Ingp+phcalf16_actg$R_Plazo+phcalf16_actg$SubReg2+phcalf16_actg$Cicloef2+phcalf16_
actg$tyear+phcalf16_actg$Cicloef2+phcalf16_actg$cesantexciclof)
summary(coxsuper_actg)

plot(survfit(coxsuper_actg),xlab="Mes",ylab="Supervivencia (prob.)",main="Modelo Cox (actg)",xlim=c(0,120),
cex.main=1,ylim=c(0,1))
summary(survfit(coxsuper_actg))

#Verificación de cumplimiento de los supuestos de riesgo proporcional
#Prueba estadística de proporcionalidad de RHo
cox.zph(coxsuper_actg)
#Se concluye que no existe evidencia significativa al 5% de que se viole los supuestos de riesgo proporcional,
de forma particular o global (p > 0.05)
```



```

#Residuos de Schonenfeld o dfbetas (corrobora hipótesis de riesgos proporcionales)
par(mfrow=c(2,2))
plot(cox.zph(coxsuper_actg),var=1, main="R_ingreso (entre 771,01 y 1.133 USD (Q2))",cex.main=0.45,
cex.lab=0.25)
plot(cox.zph(coxsuper_actg),var=2, main="R_ingreso (entre 1.133,01 y 1.603 USD (Q3))",cex.main=0.45,
cex.lab=0.25)
plot(cox.zph(coxsuper_actg),var=3, main="R_ingreso (entre 1.603,01 y 2.362 USD (Q4))",cex.main=0.45,
cex.lab=0.25)
plot(cox.zph(coxsuper_actg),var=4, main="R_ingreso (mayor a 2.362 USD (Q5))",cex.main=0.45,
cex.lab=0.25)
par(mfrow=c(2,2))
plot(cox.zph(coxsuper_actg),var=5, main="Plazo mayor a 20 años",cex.main=0.45,cex.lab=0.25)
plot(cox.zph(coxsuper_actg),var=6, main="Subregión COSTA (incl.STO y GAL.) y ORIENTEN
(NAP,SUC,ORE)",cex.main=0.45,cex.lab=0.25)
plot(cox.zph(coxsuper_actg),var=7, main="Ciclo econ. al momento del
desembolso",cex.main=0.45,cex.lab=0.25)
plot(cox.zph(coxsuper_actg),var=8, main="Año de concesión",cex.main=0.45,cex.lab=0.25)
par(mfrow=c(2,2))
plot(cox.zph(coxsuper_actg),var=9, main="Ciclo económico res. previo mora
grave",cex.main=0.45,cex.lab=0.25)
plot(cox.zph(coxsuper_actg),var=10, main="Cesante x Ciclo económico res. 1",cex.main=0.45,cex.lab=0.25)
#Se observa que existe una distribución aleatoria de los residuos dfbetas para todos los coeficientes
estimados

#Verificación de la influencia de las observaciones en la estimación del modelo, residuos deviance
par(mfrow=c(1,1))
plot(resid(coxsuper_actg,type="deviance"),xlab="Indice",ylab="residuos (tipodesvio)", main="Residuos (tipo
deviance)_actg")
#se evidencia que no existen observaciones atípicas que estén influenciando el ajuste del modelo

#influencia de los casos sobre la estimación de cada coeficiente
residuo_actg <- resid(coxsuper_actg,type="score")
par(mfrow=c(2,2))
plot(phcalf16_actg$R_Ingp,residuo_actg[,1],xlab="entre 770,01 y 1.133 USD (Q2)",ylab="influencia",
main="Influencia R_ingreso entre 770,01 y 1.133 USD (Q2)",cex.axis=0.40,cex.main=0.45,cex.lab=0.45)
plot(phcalf16_actg$R_Ingp,residuo_actg[,2],xlab="entre 1.133,01 y 1.603 USD (Q3)",ylab="influencia",
main="Influencia R_ingreso entre 1.133,01 y 1.603 USD (Q3)",cex.axis=0.40,cex.main=0.45,cex.lab=0.45)
plot(phcalf16_actg$R_Ingp,residuo_actg[,3],xlab="entre 1.603,01 y 2.362 USD (Q4)",ylab="influencia",
main="Influencia R_ingreso entre 1.603,01 y 2.362 USD (Q4)",cex.axis=0.40,cex.main=0.45,cex.lab=0.45)
plot(phcalf16_actg$R_Ingp,residuo_actg[,4],xlab="mayor a 2.362 USD (Q5)",ylab="influencia",
main="Influencia R_ingreso mayor a 2.362 USD (Q5)",cex.axis=0.40,cex.main=0.45,cex.lab=0.45)
par(mfrow=c(2,2))
plot(phcalf16_actg$R_Plazo,residuo_actg[,5],xlab="plazo mayor a 20 años",ylab="influencia",main="Influencia
plazo variable mayor a 20 años",cex.axis=0.4,cex.main=0.45,cex.lab=0.45)

```

```

plot(phcalf16_actg$SubReg2,residuo_actg[,6],xlab="Subregión geográfica",ylab="influencia",main="Influencia
variable subregión geográfica",cex.axis=0.4,cex.main=0.45,cex.lab=0.45)
plot(phcalf16_actg$Cicloef2,residuo_actg[,7],xlab="Ciclo económico desembolso",ylab="influencia",
main="Influencia variable ciclo econ. desembolso",cex.axis=0.4,cex.main=0.45,cex.lab=0.45)
plot(phcalf16_actg$tyear,residuo_actg[,8],xlab="Año de concesión",ylab="influencia",main="Influencia variable
tiempo de permanencia",cex.axis=0.4,cex.main=0.45,cex.lab=0.45)
par(mfrow=c(2,2))
plot(phcalf16_actg$Cicloef2,residuo_actg[,9],xlab="Ciclo económico previo mora grave",ylab="influencia",
main="Influencia variable ciclo econ. previo mora grave",cex.axis=0.4,cex.main=0.45,cex.lab=0.45)
plot(phcalf16_actg$Cicloef2,residuo_actg[,10],xlab="Cesante x ciclo económico previo mora grave",
ylab="influencia",main="Influencia variable cesante x ciclo económico (1)",cex.axis=0.4,
cex.main=0.45,cex.lab=0.45)
# no se evidencias casos atípicos extremos, consecuentemente los casos particularmente no están
influyendo la estimación de los coeficientes del modelo

#verificación de la forma funcional de la variable año de concesión
coxsuper_actg_0 <- coxph(Surv(phcalf16_actg$tiempoM90pv_2mes,phcalf16_actg$Mayor90)~1)
residmart_actg_0 <- resid(coxsuper_actg_0,type = "martingale")
par(mfrow=c(1,1))
plot(phcalf16_actg$tyear,residmart_actg_0,xlab=" tiempo permanencia",ylab="residuos de
martingala",main="Verificación forma funcional, variable año de concesión",cex.axis=0.4,
cex.main=0.55,cex.lab=0.45)
#los residuos de la variable año de concesión al ser discreta muestra una distribución particular sin embargo
tiene levemente a ajustarse a una recta

```

## Anexo VII - Código de programación R – Etapa 5 - Creación de reporte ejemplo

```

#Reporte modelo de supervivencia
#resumen función de supervivencia modelo, evaluado para cada año
super_year_coxsuper_actg <- summary(survfit(coxsuper_actg),c(12,24,36,48,60,72,84,96))
print(super_year_coxsuper_actg)
#gráfico modelo de cox actg
ggsurvplot(survfit(coxsuper_actg),data = phcalf16_actg, palette = "#2E9FDF",ggtheme = theme_minimal())

#gráfica funciones de supervivencia cesante x ciclo
caso1 <- with(phcalf16_actg,data.frame(R_Ingp=c("entre 1.133,01 y 1.603 USD (Q3)","entre 1.133,01 y 1.603
USD (Q3)"),R_Plazo=c("mayor a 20 años","mayor a 20 años"),SubReg2=c("COSTA (incl.STO y GAL.) y
ORIENTEN (NAP,SUC,ORE)","COSTA (incl.STO y GAL.) y ORIENTEN (NAP,SUC,ORE)"),
Cicloef2=c("1","1"),tyear=c(3,3),Cicloef2=c("1","1"),cesantexciclof=c("1","0")))
scoxc1 <- survfit(coxsuper_actg,newdata = caso1)
super_year_coxsuper_actg_c1 <- summary(scoxc1)
ggsurvplot(scoxc1,data = phcalf16_actg,conf.int = TRUE, legend.labs=c("Cesantexciclof=0",
"Cesantexciclof=1"))

```