



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

REGRESIÓN LOGÍSTICA DEL INCUMPLIMIENTO FINANCIERO

TESIS PARA OBTENER EL TÍTULO DE
LICENCIADO EN ACTUARÍA

PRESENTA:

DAVID ADRIAN ZAMORA MORALES

DIRECTOR DE TESIS

DR. FERNANDO VELASCO LUNA

PUEBLA, PUE. 22 DE SEPTIEMBRE 2020



A mis padres y hermana por su apoyo en todo momento,
siendo los pilares de mi vida, mi ejemplo de superación y motivación.

AGRADECIMIENTOS.

Primero a mis padres quienes a lo largo de mi vida me han apoyado y motivado mi formación académica, su tenacidad y lucha interminable han hecho de ellos un gran ejemplo a seguir, por sus ánimos y atenciones, por siempre desear y anhelar lo mejor para la vida de mi hermana y mía, gracias por cada consejo y por cada una de sus palabras que me guiaron durante mi vida.

A mi hermana por impulsarme a buscar nuevos retos y a tener más paciencia al realizar proyectos, su sinceridad, generosidad y tiempo compartido al ser un ejemplo de valentía, capacidad y superación; al igual que me mostró que una sobrina se ha vuelto alguien tan importante para mí ay que con su sincero cariño ocupa un importante lugar en mi corazón.

Extender mi agradecimiento a mi director de tesis, el Dr. Velasco Luna Fernando por su paciencia y generosidad para compartir su experiencia y amplio conocimiento en la elaboración de lo estudiado en esta tesis que se vio reflejado en muchos de los resultados obtenidos, al igual que a sus rápidas respuestas en mis diversos cuestionamientos.

A los miembros de mi jurado, el Dr. Tajonar Sanabria Francisco Solano, M.C. Zavala López Brenda y M.C. Hernández Asunción José; por su disponibilidad y ser grandes maestros esenciales en mi desarrollo académico por su forma en la que transmite sus conocimientos y motivar a seguir aprendiendo. Por aceptar el examinar y emitir un dictamen de este trabajo.

Mis familiares abuelos, tíos y primos que han cambiado mi vida sin intentarlo y lo mucho que significan para mí, no puedo imaginar cómo serían las cosas si no los hubiera conocido.

Por último a mis compañeros en la universidad, parte fundamental de esta etapa de mi vida con los que compartí largas horas de estudio y agradables momentos a lo largo de la licenciatura.

Índice.

Introducción.....	1
Capítulo 1. El crédito y las Instituciones Financieras.....	5
1.1. Antecedentes del Crédito.....	5
1.2. Ventajas y Desventajas del Crédito.....	8
1.3. Clasificación del Crédito.....	9
1.4. Instituciones Financieras.....	10
1.5. El Otorgamiento del Crédito.....	11
1.6. Investigación de Crédito.....	13
1.7. Riesgo de Crédito.....	16
1.8. El Riesgo de Crédito: El Enfoque Actuarial.....	18
1.9. Metodologías Estadísticas.....	19
1.10. Morosidad.....	23
Capítulo 2. Análisis de Regresión.....	27
2.1 Modelos de Regresión.....	27
2.2. Regresión Logística Simple, Binaria y Multinomial.....	30
2.3. Pruebas estadísticas para el modelo logístico.....	34
2.3.1. Desviación.....	34
2.3.2. Estadístico de Wald y el Test de Hosmer-Lemeshow.....	34
2.4. Eligiendo un Mejor Modelo.....	37
Capítulo 3. Caso Aplicado de la Regresión Logística.....	38
3.1. Diseño Estadístico.....	38
3.1.1. Aspectos Generales.....	38
3.1.2. Criterios de Inclusión.....	39
3.1.3. Criterios de Exclusión.....	39
3.2. Análisis Estadístico.....	41
Capítulo 4. Resultados.....	42
4.1. Análisis Preliminar.....	42
4.1.1. Análisis Univariado.....	42
4.1.2. Análisis Bivariado.....	46
4.2. Regresión Logística.....	51
4.2.1. Comparando Resultados.....	58
4.2.2. Comportamiento de la Regresión.....	62
Conclusiones.....	64
Bibliografía.....	66

INTRODUCCIÓN.

Es incuestionable la importancia de las finanzas en nuestra sociedad no sólo como una herramienta para la manipulación del dinero o la gestión del riesgo, sino como la administración de todos los activos de la sociedad. Los posibles riesgos que puede asumir una entidad financiera en el mercado son muy diversos como lo son el riesgo de liquidez, riesgo operacional que puede provocar pérdidas debido a errores humanos, el riesgo de tipo de cambio caracterizado por la potencial pérdida como consecuencia de las fluctuaciones de las divisas de acuerdo a la volatilidad y posición de ésta en un momento determinado, entre otros.

Por su naturaleza, las instituciones de crédito se ven comúnmente expuestas a contingencias y una de ellas es el riesgo de contraparte, el cual existe

Cuando se da la posibilidad de que una de las partes de un contrato financiero sea incapaz de cumplir con las obligaciones financieras contraídas, haciendo que la otra parte del contrato incurra en una pérdida. El riesgo de crédito es el caso particular cuando el contrato es uno de crédito, y el deudor no puede pagar su deuda. (Definiciones básicas de riesgos, 2005, p.7)

Ante el riesgo crediticio las instituciones deben contar con la solvencia para hacer frente a las obligaciones adquiridas y al mismo tiempo obtener un rendimiento sobre su capital. Con la finalidad de lograr esto se recurre a la aplicación de algunas técnicas estadísticas para desarrollar modelos que den información para una acertada toma de decisiones.

Dado el continuo crecimiento del mercado de crédito al consumo, el tomar eficientemente decisiones en aspectos sociales y de rentabilidad exige un gran interés en modelar con métodos estadísticos de una manera oportuna para evitar riesgos por incumplimiento de pago.

Es a partir de la Gran Depresión de 1929 que los banqueros norteamericanos empezaron a utilizar los análisis financieros mediante ratios, como técnica de gestión ante las contingencias de sus clientes; actualmente en la concesión de crédito a los particulares las entidades financieras utilizan el scoring, como herramienta básica de análisis de un cliente. El cual consiste, en un algoritmo que proporciona una puntuación relacionada con la calidad crediticia de una operación de crédito y limitando la decisión de ser o no aprobada.

Este **rating** permite conocer la calidad crediticia, es decir, la capacidad de un cliente para hacer frente al compromiso solicitado. De igual forma podemos entenderlo como la forma de calificar la asignación de la probabilidad de incumplimiento en el pago del evento solicitado, dependiendo de esta se concederá o no el respaldo a este compromiso, igual nos da una idea del costo de la operación al ser aprobada.

El **scoring** puede aplicarse a empresas o particulares, esta técnica se aplica para el crédito al consumo, compras de electrodomésticos, vehículos, construcción, etc e incluso para préstamos personales e hipotecarios.

Esta técnica consiste en una puntuación (**score**) de las características personales y objetivas de un cliente, como los ingresos de nómina, intereses de depósitos a plazo fijo, endeudamiento previo al crédito solicitado, saldos medios en sus cuentas de depósito, etc. En base a la puntuación obtenida por el solicitante de financiación se le concede o no el crédito. La mayoría de entidades crediticias disponen de un sistema de scoring propio. (Picanyol, 2013, p.111)

Por otro lado, el préstamo, proveniente del latín *praestarium*, es una actividad que los humanos han realizado desde épocas muy antiguas, incluso antes que existiera la figura del dinero, como instrumento para facilitar el intercambio de bienes y servicios.

Con el transcurrir del tiempo esta actividad sigue vigente hasta nuestros días; según el Banco Mundial, poco más del 30% de adultos en el mundo mencionó haber pedido dinero prestado. Sin embargo, las costumbres, tradiciones, hábitos, características económicas y geográficas de cada región y país, han sido determinantes para que no todos recurran o lo utilicen de la misma manera. (Proteja su dinero, (s.f.), p.32)

Según la CONDUSEF, en México, el 17.7% de adultos utiliza, al menos, un mecanismo de préstamo formal, a través de instituciones financieras; y el 23.9% recurre al crédito informal como la familia, amigos, casas de empeño, cajas de ahorro entre otros; 9.8% utiliza ambos tipos de créditos y 48.6% ninguno.

En nuestro país, el préstamo informal se usa primordialmente para solventar gastos personales, mientras que el formal no dirigido (nómina, personal y grupal), se destina a la remodelación o compra de vivienda. Otro dato importante a tomar en cuenta es que en México las mujeres utilizan más el crédito formal, mientras que los hombres recurren mayormente al préstamo informal. Estos datos se han obtenido por medio del portal de la Comisión Nacional para la Protección y Defensa de los Usuarios de Servicios Financieros (CONDUSEF).

En el país, de acuerdo con la Encuesta Nacional de Inclusión Financiera (ENIF) 2019, publicado por el gobierno de México, de 76 millones de adultos de entre 18 y 70 años, 71% no tiene ningún tipo de crédito formal: ni bancario, ni tarjeta departamental, ni de tienda de autoservicio, ni con otras instituciones públicas.

Son muchos los motivos por los que los mexicanos (sin contar a las empresas) no acceden a opciones de crédito formales. El (33%) porque no cumplen con los requisitos que las entidades les piden para tenerlo; es decir, no tienen trabajo o sus ingresos son insuficientes.

Juárez (2016) afirma, “Uno de los principales motivos por los que a los mexicanos les rechazan una solicitud de crédito formal, según la ENIF, es precisamente que no pueden comprobar ingresos o les son insuficientes”.

(Recuperado de <https://www.eleconomista.com.mx/sectorfinanciero/En-Mexico-54-millones-de-adultos-no-tienen-credito-formal-20160619-0077.html>)

Objetivos.

Determinar si las variables estado civil, tipo de vivienda, nivel máximo de estudios, tipo de economía y préstamo o valor del proyecto, tienen influencia en la probabilidad de incumplimiento del pago de una obligación.

Objetivos Particulares.

- Determinar si la variable estado civil influye en el incumplimiento del pago.
- Determinar si la variable tipo de vivienda influye en el incumplimiento del pago.
- Determinar si la variable nivel máximo de estudios influye en el incumplimiento del pago.
- Determinar si la variable tipo de economía influye en el incumplimiento del pago.
- Determinar si la variable valor del proyecto influye en el incumplimiento del pago.

Capítulo 1

El crédito y las Instituciones Financieras

“El que maneja el crédito maneja más la moneda que el que la emite. El que maneja el crédito maneja más el comercio de exportación e importación que el que compra y el que vende. El que maneja el crédito estimula determinadas formas de producción y debilita otras; el que maneja el crédito establece qué es lo que se ha de producir y que es lo que no; determina lo que puede y lo que no puede llegar al mercado con facilidades de venta, y maneja por consecuencia el consumo.”

Arturo Jauretche. (s.f.)

1.1. Antecedentes del Crédito

Para abordar el tema de incumplimiento el cual es el objetivo general de este trabajo, debemos empezar con el tema de los créditos, un aspecto importante en ellos es identificar el comportamiento de los clientes y su morosidad. En la actualidad se han desarrollado investigaciones sobre este fenómeno económico en todo el mundo; ya que afecta en gran medida la situación bancaria mundial, ya que cada año ha ido creciendo la tasa de morosidad en los créditos. Por esta razón se han implementado técnicas como el credit scoring que es una herramienta numérica basada en un análisis de nivel de los archivos de crédito de una persona, para representar la capacidad de pago en un individuo, sirven para evaluar el riesgo potencial que

representa el prestar dinero a consumidores y aminorar las pérdidas provocadas por la cartera vencida. Un ejemplo de esto son los estudios hechos en la Universidad de Birmingham en su departamento de Economía, uno de sus trabajos relevantes es el de la “Implementación de un sistema de gestión de riesgos de crédito basado en técnicas innovadoras de puntuación”, donde presentan un sistema de calificación crediticia que tiene como objetivo controlar el riesgo crediticio. Al igual desde el punto de vista contable se ha estudiado la morosidad y sus efectos en las instituciones financieras como ejemplo tenemos la Universidad Nacional de Trujillo donde la Facultad de Ciencias Económicas en el área de Contabilidad y Finanzas, estudiaron este fenómeno en la ciudad de Chepén en Perú.

Estas y otras investigaciones han servido como antecedentes para impulsar este trabajo y generar la idea de estudiar algunas variables y como afectan al incumplimiento de pago en estados de la República Mexicana haciendo uso de herramientas matemáticas; con este fin empezaremos por adentrarnos en los antecedentes del crédito.

La reglamentación del préstamo con interés es un elemento central de los primeros corpus jurídicos de la antigüedad, basados en la posibilidad física de disponer materialmente de alguna cosa, por esa razón algunos historiadores suponen que el crédito podría tener un origen antiguo y remontarse a la sedentarización agrícola del Neolítico.

Entenderemos como crédito al compromiso pactado entre una persona o institución que otorga capacidad de compra por adelantado al deudor, que también puede ser una persona física o moral. El crédito permite realizar ventas y satisfacer la necesidad de compra de los consumidores, conforme a su capacidad de pago. Las condiciones del convenio que permiten el acuerdo comercial

pueden ser flexibles y negociables en cuanto a plazos, montos, tipos de interés, etc.; con la finalidad de concluir en buenos términos el compromiso de crédito adquirido. (Nieto, 2010, p. 5)

Los créditos se han otorgados a los diferentes individuos de la sociedad para adquirir productos y servicios: a los campesinos para la adquisición de herramientas de labranza y pagar deudas; a los manufactureros para incrementar su volumen de fabricación; a los comerciantes para la adquisición de productos y equipos de distribución; a los gobiernos quienes necesitan saldar sus deudas. Se ha usado a lo largo de la historia como una herramienta diversa.

En el período alejandrino, el crédito lo ocupaban los comerciantes y los negociantes para acrecentar sus empresas; durante el dominio de los romanos, los usureros se apoderaban de las tierras de los campesinos cuando no pagaban sus deudas; en los últimos dos siglos se usó el crédito para reactivar la economía.

El primer pueblo en condenar el préstamo con intereses es el hebreo y lo expresa en su éxodo; el mal manejo del intercambio en los bienes y el alto interés en los pagos dieron paso los primeros Montes de Piedad, en Perusa en 1462 y en Gubbio en 1463; se difundieron por toda Italia gracias a los franciscanos, estas oficinas de empeños ofrecían un interés reducido para ayudar al pobre y evitar que los cristianos recurrieran a la usura. Por desgracia resultaron ser muy frágiles, muchos de ellos desaparecen por falta de liquidez, apropiación de fondos por empleados, fraudes de los clientes, saqueos de las ciudades, malas tomas de decisión a la hora de dar el préstamo, etcétera.

El comienzo de que un comerciante venda mercancías y reciba pagos parciales en periodos fijos, encuentra sus orígenes en Estados Unidos de América, alrededor de 1850, la Compañía de Máquinas de Coser Singer, emprendió vendiendo sus productos basándose en un sistema de pagos

mensuales y semanales. Aún con el estigma social de ese periodo, el plan de venta en abonos aumento en popularidad.

En el México independiente, Morales (2014) afirma “se fundó en 1864 el primer banco de características modernas, El Banco de Londres, México y Sudamérica. Asimismo, se fundaron otros bancos entre los cuales sobresalen: Banco Nacional Mexicano, en 1882, y Banco Mercantil, con la publicación de sus estatutos en 1881” (p.16).

1.2. Ventajas y Desventajas del Crédito

Actualmente el crédito es de vital importancia para la economía, sin embargo como todo proceso tiene sus ventajas y desventajas dependiendo de cómo se aplique. Si se utiliza adecuadamente produce los siguientes beneficios:

- Aumento de los volúmenes de venta
- Incremento en la producción de bienes y servicios, y como consecuencia, disminución de los costos unitarios.
- Da flexibilidad a la oferta y demanda.
- Elevación del consumo, ya que permite a determinados sectores socioeconómicos adquirir bienes y servicios que no podrán pagar de contado.
- Hace más productivo el Capital.
- Acelera la producción y la distribución.
- Creación de fuentes de empleo, mediante nuevas empresas y la ampliación de las ya existentes.
- Aumenta el volumen de los negocios.

- Desarrollo tecnológico, favorecido indirectamente por el incremento de los volúmenes de venta.
- Ampliación y apertura de nuevos mercados.
- Facilita la transferencia de dinero. (Del valle, 2014, p.7)

Las desventajas radican en el mal uso de finanzas personales, el incremento de intereses y usar dinero que aún no tenemos podría hacer que caigan los solicitantes en el incumplimiento; la principal desventaja reside en que al aumentar el volumen de créditos la oferta monetaria crece, lo que provoca el alza de los precios produciendo inflación.

1.3. Clasificación del Crédito

Existen diversos criterios para clasificar el crédito, por eso pueden existir diferentes clasificaciones, dado esto presentaremos una clasificación general:

- **Créditos de Consumo o Créditos Comerciales:** Son otorgados por empresas para la adquisición de bienes o servicios de uso personal en plazos determinados.
- **Créditos Empresariales:** Si una empresa pretende adquirir materia prima, insumos, servicios, etc.; solicita el servicio o bien a otras empresas a crédito para continuar con sus labores, realizan convenios para cubrir el adeudo.
- **Créditos Bancarios:** Estos son los otorgados por entidades bancarias o empresas del sistema financiero a personas físicas o morales que necesitan recursos para financiar sus actividades, pueden abarcar la adquisición de bienes, servicios, pagar deudas, etc.

1.4. Instituciones Financieras

El papel principal de las instituciones financieras es proporcionar liquidez a la economía y permitir un mayor nivel de actividad económica. Según el Brookings Institute, los bancos

consiguen esto de tres maneras principales: ofreciendo crédito, administrando mercados y agrupando el riesgo entre los consumidores.

La fuente más obvia de la liquidez proporcionada por las instituciones financieras se presenta en forma de crédito disponible. El crédito emitido por el banco ayuda a impulsar la expansión económica al permitir que las empresas emprendan nuevas empresas sin ahorrar el dinero necesario. Esto permite a los actores económicos aprovechar las oportunidades a medida que surgen, en lugar de perder las ganancias potenciales por falta de efectivo disponible.

Las instituciones financieras tienen otras herramientas a su disposición para proporcionar liquidez. Una de ellas, también descrita en el informe del Instituto Brookings, es la compra y venta de valores. Al "hacer un mercado" de esta manera, los bancos se colocan en el corazón de los mercados financieros. Otra forma más en que las instituciones financieras sirven al mercado es actuando como depósitos de riesgo.

Al agrupar el riesgo de, digamos, fluctuaciones de las materias primas, las instituciones que comercian con derivados pueden actuar para estabilizar lo que de otro modo serían mercados volátiles, estas transacciones de derivados ayudan a proteger a los inversores más pequeños de la exposición a las fluctuaciones de precios. (What Is the Role of Financial Institutions?, (s.f.), Reference*, recuperado de <https://www.reference.com/business-finance/role-financial-institutions-6172cf8df7b8095e>).

1.5. El Otorgamiento del Crédito

Las **instituciones de crédito** son establecimientos que realizan transacciones financieras, tales como inversiones, préstamos y depósitos. Casi todos los entes se relacionan con las

instituciones de crédito de manera habitual. Para entender el funcionamiento entenderemos como institución crediticia al otorgante del crédito y al deudor como sujeto de crédito o cliente.

La disposición de otorgar o negar un crédito depende del resultado de la averiguación que se hace acerca del interesado para conocer si lo pagará; y de la percepción que el analista tenga respecto del solicitante en su comportamiento probable de pagar el crédito.

Una investigación de crédito se lleva a cabo para ayudar a tomar una sana decisión respecto al crédito; este propósito debiera tenerse constantemente en mente cuando se emprende una investigación de crédito [...]. La investigación no busca información por simple capricho, busca información pertinente a la decisión específica de crédito, suficiente para asegurar que la decisión recaerá dentro del límite de la probabilidad de recuperación de acuerdo con las políticas de operación, (Cole, 1997, p.195).

Los analistas de crédito deben determinar los elementos que tienen que conocer para tomar una decisión de aceptación o rechazo en la solicitud de crédito, de manera que estos, en su mayoría, sean recuperados.

Por lo general cada empresa definen de alguna manera sus condiciones para otorgar y administrar los créditos a clientes, habrá empresas que tengan algún departamento dedicado particularmente al otorgamiento de créditos y otro para la cobranza, en otras empresas quizás el departamento de finanzas se encargue a la par de esas funciones . Para el otorgamiento de créditos generalmente se especifican los criterios para la investigación sobre el cliente, normas para fijar los límites de crédito, el monto de crédito, los pagos periódicos y la asignación de responsabilidades para gestionar todo el proceso, el cual generalmente está compuesto de:

- **Investigación:** La concesión de crédito a nuevos clientes, así como a clientes ya establecidos, requieren el conocimiento de la capacidad financiera del cliente para contraer endeudamiento. Antes de expedir el primer pedido se llevara a cabo una investigación sobre los antecedentes financieros del cliente.
- **Análisis:** Es el estudio de los datos que permitirán decidir si se otorga o niega la solicitud de crédito, lo que preferentemente se analiza son los hábitos de pago del cliente y de su capacidad para atender sus compromisos.
- **Aceptación del Cliente:** Al terminar la investigación y el análisis del perfil del cliente, si es que se decide otorgarle la línea de crédito, se le calcula un límite de crédito en función de su potencial de compras y su capacidad para pagar en tiempo y forma el monto de crédito otorgado.

Si la investigación indicara la conveniencia de denegar el crédito y según el grado de calificación negativa, se podrían considerar algunas alternativas como garantías adicionales para el buen fin de las operaciones.

1.6. Investigación de Crédito

“La investigación de crédito reúne la información histórica por medio del análisis que se usa como base para predecir la conducta futura. La predicción que se busca es de conducta de crédito, esto es, una predicción del tipo e importe de las obligaciones por concepto de crédito que puede esperarse que paguen y una estimación del importe máximo del crédito que debiera aceptarse del cliente individual”, (Cole, 1997, p.192).

El objetivo de la investigación es la creencia que al saber su comportamiento pasado se puede predecir el comportamiento futuro.

Las instituciones que otorgan créditos inician la investigación del solicitante con elementos fundamentales para conocer el comportamiento del individuo. Algunas empresas deben hacer un análisis complejo del interesado, porque la cantidad de los recursos que se prestaran son grandes en relación con sus los ingresos y probablemente el plazo sea de varios años; un crédito hipotecario podría ser un ejemplo. Hay otros créditos en los que no necesita hacerse una investigación profunda, como es el caso del dueño de una tienda de abarrotes que da crédito a sus clientes, en ese caso los elementos a considerar son más simples que los del anterior ejemplo.

Entonces los elementos que deben investigarse dependen del tipo de crédito que se otorgue. Pero entre los elementos fundamentales que se investigan están los siguientes conocidos como la técnica de las 5c

- Carácter.
- Capacidad.
- Capital.
- Colateral.
- Condiciones.

A continuación describiremos cada una

Carácter: Para conocer el carácter del solicitante de crédito es necesario diseñar con precisión los indicadores que permitirán conocer la conducta al liquidar sus deudas.

La actitud hacia las obligaciones se revela mejor por medio de una pregunta que inquiriere si la persona hace reclamaciones injustificadas o actúa de acuerdo con el contrato, que por preguntas generales acerca de la actitud respecto a obligaciones. La situación de la familia se revela de un modo más significativo por las peculiaridades del matrimonio, el número de hijos, lugar y

duración de la residencia, la propiedad de la casa, los principales activos propiedad de la familia, el seguro sobre la vida, los ahorros principales y las inversiones, que por vagas preguntas relativas a la situación personal y de la familia. El carácter es el registro del cumplimiento de las obligaciones pasadas del solicitante de crédito, (Morales, 2014, p.109).

Capacidad: Morales (2014) afirma “La capacidad es una cualidad de crédito [...] simplemente puede querer decir la capacidad de pagar una obligación precisa en dinero cuando se adeuda [...] a veces se resume bajo los encabezados de ingreso y empleo” (p.109). Para conocer la capacidad de un solicitante de crédito, además de conocer los ingresos que posee, es necesario conocer como los distribuye en sus diferentes gastos, y cuanto es lo que dispone de ingreso libre para el pago de los creditos que solicita actualmente. La capacidad es la posibilidad del solicitante para reembolsar el crédito requerido.

Capital: Morales (2014) concluye “Es la solidez financiera del solicitante y se puede medir por su posición como propietario de activos; aquí se recomienda un análisis de la deuda con el capital para conocer la solidez del capital del solicitante”, (p.109).

Colateral: La garantía colateral es la cantidad de activos que el solicitante tiene disponible como garantía de que pagara el crédito; cuanto mayor sea la cantidad de activos disponibles mayor será la probabilidad de que el prestatario recupere el monto prestado.

Condiciones: Es el ambiente económico y de negocios y la situación particular que influya en alguna de las partes de la transacción del crédito, por ejemplo, en el caso de que los productos que se desee vender sean pasados de moda, la empresa deseara venderlos lo más rápido posible, en ese sentido las características del crédito probablemente sean más laxas, (Morales, 2014, p.109).

Algunas cualidades personales del individuo que solicita el crédito que se verifican con frecuencia en cualquier análisis son:

- Registro de pagos
- Empleo
- Estado civil
- Edad
- Referencias
- Activos de reserva o propiedad que tenga a su nombre
- Valor del artículo que desea adquirir

Toda esta información es el mínimo esencial que se toma en cuenta para tomar una decisión acertada a la hora de conceder o negar el crédito, dado que considerar de forma oportuna y certera ayuda mucho tanto a la institución como al solicitante, se han creado diversos métodos para optimizar el proceso.

1.7. Riesgo de Crédito

Antes de tomar una decisión es conveniente hacer un estudio del riesgo que se puede adquirir al otorgar un crédito; dentro de las principales definiciones de riesgo podemos recordar las siguientes:

“El riesgo de crédito surge cuando las contrapartes están indispuestas o son totalmente incapaces de cumplir sus obligaciones contractuales. El riesgo de crédito abarca tanto el riesgo de incumplimiento, que es la valuación objetiva de la probabilidad de que una contraparte incumpla, como el riesgo del mercado que mide la pérdida financiera que será experimentada si el cliente incumple”.

Philippe Jorion. (1999)

“Se determina que existen dos tipos de riesgo de crédito: el riesgo de incumplimiento, que se refiere a la pérdida potencial derivada de que la contraparte no pueda cumplir con sus obligaciones financieras en las condiciones definidas contractualmente; y el riesgo de mercado, que se define como la pérdida potencial que podría sufrir un tenedor de un portafolio de préstamos, instrumentos financieros o derivados, como consecuencia de que el valor de mercado de estos disminuya. La segunda definición, plantea exposición al riesgo de crédito aún en el caso de que la contraparte no sufra quebranto alguno”.

Comité de Supervisión Bancaria de Basilea. (s.f.)

El riesgo de crédito puede definirse sencillamente como la pérdida potencial ocasionada por el hecho de que un deudor o contraparte incumpla con sus obligaciones de acuerdo con los términos establecidos.

Elizondo (2004) nos dice “El incumplimiento es un acontecimiento que tiene asignada una probabilidad, la cual puede analizarse a nivel de acreditado, o deudor, es decir, a nivel individual; Este evento se conoce como riesgo individual donde se consideran los siguientes elementos:

- La tasa de incumplimiento
- La tasa de recuperación
- La migración del crédito

El riesgo de portafolio considera el análisis desde el punto de vista del agregado de créditos y este dependerá de la composición y naturaleza de cada cartera. La estimación de la pérdida agregada en este caso deberá considerar tanto la composición como la concentración de los créditos

que compone cada cartera, así como tomar en cuenta las correlaciones que puedan existir entre las fuentes de riesgo de los mismos, (Elizondo, 2004, p.).

Los aspectos más importantes a considerar son:

- La correlación entre la probabilidad de incumplimiento y la calidad del crédito
- La concentración de riesgo
- El riesgo de incumplimiento hace énfasis a la incertidumbre asociada a la habilidad de una institución

La estimación de correlaciones entre los deudores es un instrumento básico para el análisis del portafolio.

Cuando se lleva a cabo el análisis de riesgo de crédito de una institución financiera, se espera obtener como resultado el nivel de pérdidas de capital que dicha institución puede llegar a tener como resultado del incumplimiento de sus acreditados. El incumplimiento a su vez está asociado al deterioro gradual que puede observarse en la calidad de los activos de la institución, lo cual se traduce en lo que se conoce como la pérdida esperada.

1.8. El Riesgo de Crédito: El Enfoque Actuarial

La incertidumbre es una de las características principales con las cuales debe vivir una institución financiera. Una amplia serie de fenómenos, cuyo comportamiento es impredecible, tiene un impacto directo en el desempeño de dichas instituciones.

El problema de estimar pérdidas por riesgo de crédito de una cartera presenta ciertas similitudes con los portafolios de asegurados. El incumplimiento de crédito es un evento incierto al igual que el siniestro de un asegurado. El análisis de las variaciones de factores cuyo

comportamiento es impredecible puede ser realizado por medio de diversas herramientas estadísticas, lo cual, en el caso particular de las compañías de seguros, ha dado lugar a la teoría de riesgo. Una de las aplicaciones tradicionales de la teoría de riesgo es encontrar la distribución de probabilidad de pérdida originada por los instrumentos financieros adquiridos por un conjunto de individuos, (Gutiérrez, (s.f.), Riesgo de crédito: el enfoque actuarial. Recuperado de http://www.actuaires.org/EVENTS/congresses/Cancun/ica2002_subject/credit_risk/credit_71_gutierrez_elizondo.pdf).

Ambos fenómenos pueden ser modelados por una distribución de pérdida para poder ponderar la situación, como ya mencionamos la teoría de riesgo ha resultado de gran utilidad en el ámbito bancario y de seguros.

La distribución de pérdida derivada del análisis de un portafolio muestra las posibles pérdidas en un periodo determinado y nos permite tomar en cuenta diversas particularidades colectivas del grupo de individuos que lo conforman. Este tipo de distribuciones da origen a medidas que se conocen en el ámbito de las finanzas con el nombre de Valor en Riesgo (VaR), mismas que han sido promovidas y avaladas por organismos internacionales regulatorios, para medir el riesgo de mercado.

Recientemente el uso de la teoría de riesgo ha sido explotada exitosamente en el ámbito bancario para construir la distribución de pérdidas de portafolios crediticios. En 1997 Credit Suisse Financial Products (CSFP) publicó el modelo CreditRisk+ para la medición del riesgo de crédito. El modelo CreditRisk+ resuelve muchas de las críticas que se han hecho a los esquemas regulatorios vigentes para cuantificar las pérdidas por riesgo de crédito, (Gutiérrez, (s.f.), Riesgo de crédito: el enfoque actuarial. Recuperado de

http://www.actuaires.org/EVENTS/congresses/Cancun/ica2002_subject/credit_risk/credit_71_gutierrez_elizondo.pdf).

1.9. Metodologías Estadísticas

Se denominan ciclos económicos a las oscilaciones recurrentes de la economía en las que una fase de expansión va seguida de otra contracción. Se le conoce como ciclo debido a que una vez terminado comienza de nuevo desde el inicio formando una rueda continua. Sin embargo, debido a su imprevisibilidad no se puede tomar esto como una regla formal.

Suele dividirse al ciclo económico en diferentes partes:

- **Recuperación:** Etapa del ciclo en que la economía está estancada o crece ligeramente. Se incrementa demanda de crédito con instituciones financieras.
- **Expansión:** Fase donde se presenta el mayor crecimiento económico. La captación bancaria crece más rápido que el PIB y el crédito bancario se incrementa en consecuencia y con una tendencia irracional a aceptar contratos con tasas de interés fijas o topadas.
- **Auge:** Periodo en que el crecimiento económico comienza a mostrar señales de retroceso. Ante los temores de escasez o controles crediticios, se negocian líneas de crédito
- **Recesión:** La actividad económica se comprime provocando una disminución del consumo, inversión y producción de bienes y servicios, iniciando el despido de trabajadores aumentando la tasa de desempleo excediendo el 10%. Por estas razones se presenta la restricción de la liquidez y del crédito.

- Depresión: Si nos encontramos en una fase de recesión constante y sin perspectiva de mejora. Dependen los individuos en mayor medida del financiamiento informal y menos de los créditos bancarios.

La duración del ciclo económico es variada ya que raramente han tenido la misma durabilidad, se han presentado las cuatro fases en dos años y en otras ocasiones han pasado más de 10 años.

A lo largo de la historia se han usado diferentes métodos para el análisis de estos fenómenos, a principio de los 80 la metodología más usada para el estudio de información financiera era el análisis discriminante, con el pasar de los años se han desarrollado diversos métodos así que nombraremos los más importantes:

- Modelo Z-Score: El análisis discriminante en su forma más simple busca obtener una combinación lineal de las características que mejor discriminan entre los grupos, es decir que maximice la varianza entre grupos y minimice la varianza dentro de cada grupo, en otras palabras identificar variables que permitan homogeneidad de varianzas dentro de cada grupo y heterogeneidad de varianzas entre grupos distintos.

El Modelo Z-Score es el nombre que se le da al resultado de aplicar el análisis discriminante a un conjunto de indicadores financieros, que tienen como propósito clasificar a las empresas en dos grupos: bancarota y no bancarota.

El modelo original Z de Altman fue desarrollado para predecir las quiebras de las empresas

- Modelo Z: En 1977, Altman, Haldeman y Narayanan construyen un modelo en el que introducen algunas modificaciones al modelo Z-Score original. El propósito de este modelo

conocido como Modelo Z es clasificar a las empresas en bancarrota incluyendo lo siguiente:

1. Empresas medianas y grandes en el análisis.
2. Empresas del sector no-manufacturero.
3. Los cambios en los estándares de cálculo de las principales razones financieras y nuevas prácticas contables.
4. Técnicas más recientes del análisis estadístico para la estimación del análisis discriminante.

El resultado de la estimación del modelo Z resulta ser superior al Z-Score ya que permite predecir la bancarrota de las empresas con anticipación de 5 años, con un nivel de confianza de 70%, y predice con un año de anticipación con una confianza de 96%.

- Modelo EMS: Este modelo muestra el análisis de clasificación de riesgo financiero de una empresa y puede ser utilizado para calcular el valor relativo entre los diversos créditos, combinando rigurosamente los análisis estadísticos del desempeño financiero de las empresas emisoras de bonos, con factores cualitativos críticos, con el objeto de asignar una calificación de bonos equivalente para todas las empresas emisoras.
- Modelo Credit Monitor de KMV: Las empresas KMV desarrollo un modelo de probabilidad de incumplimiento, Credit Monitor (CM), que permite estimar la Frecuencia de Incumplimiento Esperada (EDF). Este modelo está diseñado para transformar la información contenida en el precio de una acción en medida de riesgo de incumplimiento, Sellers, Vasisek y Levinson (2000).

- **Modelo RPA:** El modelo Recursive Partitioning Algorithm (RPA), presentado originalmente por Fryedman H., Altman Edward, Duen-Likao, es considerado como un procedimiento de calificación Bayesiana para el análisis financiero, ya que es una técnica de clasificación no paramétrica, basada en patrones de reconocimiento que tiene los atributos tanto del enfoque de clasificación invariado clásico como el de los procedimientos multivariados.
- **Matrices de Transición:** La matriz de transición es la principal herramienta para determinar la probabilidad de que un crédito con calificación determinada cambie de calificación crediticia durante un periodo específico. A esta probabilidad se le conoce como “probabilidad de migración en la calidad de un crédito”. Las matrices pueden utilizarse para otros fines alternativos, en particular permiten resumir la información inter temporal de la calidad de la cartera de créditos de los bancos.
- **Modelos de Elección Cualitativa:** existen situaciones en las que se desea explicar, a través de un modelo, la relación que existe entre un conjunto de variables o atributos y una variable dicotómica que representa la elección entre dos opciones cualitativas denominadas como éxito o fracaso, las cuales se denotan con los valores 0 y 1, cuyo principio es la probabilidad de que un evento ocurra depende de ciertos atributos que caracterizan al individuo que realiza la elección. Ejemplo son los siguientes modelos:
 1. **Modelo Probit:** Un modelo que garantiza que las probabilidades estimadas se encuentren en el intervalo (0,1) y cuya relación guardan estas con el vector de atributos es no lineal, es el modelo probit, que se basa en la función de distribución acumulada normal.

2. Modelo Logit: Una alternativa para la estimación de modelos de elección binaria es el modelo Logit, en este caso, se supone que la función de distribución acumulada en la que se basa el modelo es la función Logística.

Estos últimos modelos serán de relevancia para nuestro tema de estudio.

1.10. Morosidad

"Morosidad" es un término general usado para abarcar varios comportamientos diferentes entre comprador y vendedor, pero en general se usa para referirse a una situación en la que un comprador no paga la factura de bienes o servicios prestados en la fecha de vencimiento acordada según los términos establecidos con el vendedor.

La morosidad bancaria se mide como la relación entre los moros y el total de operaciones realizadas. En su esencia se trata de la relación entre riesgos dudosos, esto es, casos en los que existen dudas relevantes de que la obligación no se cumplirá, y los riesgos totales (todas las operaciones realizadas que implicaron cierto riesgo)", (Roldán Paula, (s.f.) Morosidad bancaria. Recuperado de <https://economipedia.com/definiciones/morosidad-bancaria.html>).

Como sucede a menudo con los términos generales, los pagos atrasados no son situaciones únicas que son el resultado de asociaciones o culturas comerciales defectuosas, sino que a menudo son el producto final lógico de las estructuras y normas de la industria, los ciclos comerciales estándar, el desequilibrio en jerarquías de mercado, infraestructuras financieras en entornos empresariales y la relativa fortaleza y debilidad de los sistemas judiciales, entre otros factores.

El pago tardío es una consecuencia natural del crédito comercial, esto también nos deja observar que las causas fundamentales que afectan el retraso en el pago no son aleatorias, sino que surgen de la interacción entre ciertas circunstancias, un ejemplo es facilidad de acceso a crédito

a corto plazo por parte de intermediarios, como bancos y otras fuentes finales formales. La disminución continuada de la cartera crediticia de las entidades financieras, provoca que el porcentaje de morosidad es más elevado sin necesidad de que los retrasos aumenten, en caso de que aumentara la proporción sería aún mayor; que es el caso del sector financiero mundial.

Según la economista Asli Demirgüç-Kunt del Banco Mundial, “Ops. Cit. (1998). Determinants of Banking Crisis in Developing and Developed Countries”. La morosidad puede ocasionar en una entidad financiera graves problemas de solvencia si alcanza porcentajes elevados, que pueden desembocar en una quiebra.

La morosidad y las notas negativas en su historial de crédito es una constante para casi 6 de cada 10 mexicanos; quienes registran saldos o deudas atrasadas equivalentes a 2.9 veces su salario o ingreso mensual, el nivel de endeudamiento activo o atrasado de dicho sector promedia 35 mil 499 pesos y su ingreso mensual apenas alcanza 12 mil 250 pesos.

La Comisión Nacional Bancaria y de Valores (CNBV) reportó que al cierre de abril pasado el índice de Morosidad (IMOR) de la cartera de consumo –donde se ubican los créditos personales, de nómina y tarjetas de crédito– se ubicó en 4.38% a nivel nacional. Al considerar el Índice de Morosidad Ajustada (IMORA), que incluye la cartera vencida o en impago por más de tres meses, quitas (reducción del adeudo) y castigos, el nivel de incumplimiento de los clientes se dispara hasta 12.87%.

El uso de la tecnología para analizar y dar seguimiento a quién le prestan y, ser más rigurosas al momento de otorgar créditos ha ayudado a los bancos a bajar la morosidad. En el primer trimestre de 2019, la cartera vencida o morosidad se ubicó en 2.1%, su nivel más bajo desde

2010, cuando los informes de la Comisión Nacional Bancaria y de Valores (CNBV) comenzaron a ser mensuales.

Dado que la morosidad en el pago es un resultado problemático de este proceso comercial, en lugar de una casualidad aleatoria, una mejor comprensión de los factores y razones que influyen en mayores o menores casos de demora en los pagos nos ayudarían a enmarcar mejores métodos para la toma de decisiones. Bancos como HSBC han dicho que en el primer trimestre lograron disminuir su morosidad ante la venta de cartera con problemas, una mejora en los perfiles crediticios y el manejo activo de los portafolios, al igual que un buen análisis en su cartera de clientes.

A pesar de la gran variedad de trabajos relacionados con el riesgo crediticio y técnicas para abordarlo como son la creación de una Scorecard, la cual nos da un puntaje basado en un cuestionario que ofrecen una gran cantidad de información personal para poder tomar una decisión; no se han hecho trabajos específicos sobre el incumplimiento de un pago.

De lo establecido antes una financiera empezó un programa político en el cuál otorgó créditos a personas en diversas partes del país tuvieran o no manera de comprobar sus ingresos, se les pidió contestaran un breve cuestionario para conocer factores que eran de interés para el programa, esto con el fin de poder observar el comportamiento de estas personas. El tiempo que dio seguimiento en su forma de pagar va desde Julio del 2016 hasta Diciembre del 2017, el número de personas en esa base de datos fue de 1182.

La institución, que por privacidad prefiere omitir su nombre, otorgó la investigación para poder obtener información de ella. Se pretende que a partir de los resultados estadísticos la empresa

tenga conocimiento de cuáles son las variables que influyen en el incumplimiento de pago para poder tener una mejor toma de decisiones.

La decisión está basada en observar que la técnica del credit scoring da un puntaje general de todos los factores dentro de los cuestionarios pero en pocas ocasiones se ve si cada variables afecta directamente al cumplimiento del pago.

Capítulo 2

Análisis de Regresión

“Las preguntas más importantes de la vida, de hecho, no son en su mayoría más que problemas de probabilidad. “

Pierre Simon Laplace. (s.f.)

2.1 Modelos de Regresión

Como se mencionó en el anterior capítulo, una mayor comprensión de los factores que intervienen en el incumplimiento de pago ayudaría a una mejor toma de una decisión y la razón es porque nos brinda una mejor capacidad analítica que nos permite elegir el mejor camino posible, para esta investigación haremos uso de la regresión logística por ser una herramienta versátil y que con variables dicotómicas es la más usada. Para esto debemos preguntarnos, ¿Qué es la regresión logística?, el cuestionamiento anterior nos ayudará a comprender por qué se hará uso de esta herramienta en el análisis; por esta razón debemos tener algunos elementos de fondo que nos ayudaran a comprender con mayor claridad.

En estadística, el análisis de regresión trata del estudio de la dependencia de una variable de interés con respecto de una o más variables explicativas mediante técnicas para su análisis y modelación. Nos ayuda a estimar la esperanza condicional de la variable dependiente dada las

variables independientes. El objetivo de la estimación es una función de las variables independientes llamada la función de regresión y describir la variación de la variable de interés basándonos en una distribución de probabilidad.

Muchos métodos se han desarrollado para llevar a cabo el análisis de regresión y su ocupación en la práctica depende de la forma del proceso de generación de datos el cual depende de hacer suposiciones acerca de este proceso.

El método de regresión más conocido es la regresión lineal, este utiliza el MCO, también conocido como análisis de regresión convencional; el método es limitado ya que solo es aplicable si la variable dependiente es continua, independiente e idénticamente distribuida, cuando la variable dependiente es categórica, el análisis de regresión convencional no es el más apropiado esto porque incumple que debe ser continua, puede tomar valores negativos, debe distribuirse normalmente y en términos de error debe ser independiente e idénticamente distribuida, estos no se cumplen en los casos en que la variable dependiente es dicotómica. Los anteriores incumplimientos impulsan la evolución hacia los modelos lineales generalizados, dentro de los cuales pueden incluirse los denominados log-lineales y los denominados LOGIT.

La regresión logística, como la regresión de mínimos cuadrados, es una técnica estadística que se utiliza para explorar la relación entre una variable dependiente y al menos una variable independiente. La diferencia es que la regresión lineal se usa cuando la variable dependiente es continua, mientras que las técnicas de regresión logística se usan con variables dependientes categóricas.

La regresión logística, como cualquier otra técnica de construcción de modelos en estadística, tiene como objetivo encontrar el modelo más adecuado, más económico y sensible para evaluar la relación entre las variables de respuesta y al menos una variable independiente. Se

diferencia de la regresión lineal en que puede aplicarse cuando la variable dependiente es categórica y no requiere suposiciones rigurosas.

Se considera preciso establecer las diferencias entre la regresión lineal y logística antes de proseguir con el tema, para esto haremos uso de la tabla 2.1

Tabla 2.1
Comparación modelo lineal y logístico

	Lineal simple	Logística binaria
DIFERENCIAS.		
Escala de medición de la variable respuesta.	Cuantitativa continua	Dicotómica o binaria
Distribución de probabilidad $y = f(x)$	$Normal : N(\mu = 0, \sigma^2 = 1)$	Bernoulli (p)
Parámetros desconocidos.	μ, σ^2	p
Parámetros de interés: Media de y/x	$E(y/x) = \mu$	$E(y/x) = p$
Rango de valores del parámetro de interés.	$\mu \in (-\infty, +\infty)$	$p \in (0, 1)$
Modelo propuesto.	Línea (recta de regresión)	Por determinar
SEMEJANZAS.	* Estimar la asociación entre una o más variables independientes y una variable de respuesta o dependiente. * Estimación del valor del parámetro de interés en diversos escenarios.	

Adaptación de Hernández, M. (2007).

Dada las semejanzas del modelo es conveniente entender el funcionamiento y limitaciones del modelo lineal, esto para comprender porque optamos por un modelo logístico. Para entender usamos el modelo más simple:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

El modelo de la ecuación formula a Y como una función lineal de las variables explicativas (X_i). La expresión $E(Y_i | X_i)$ se interpreta como la probabilidad condicional de que el evento ocurra cuando se de X_i y donde Y_i es nuestra variable explicativa; es decir, $\Pr(Y_i = 1 | X_i)$.

Si suponemos $E(\varepsilon_i) = 0$, como es normal para obtener estimadores insesgados, se obtiene la ecuación:

$$E(Y_i | X_i) = \beta_0 + \beta_1 X_i$$

Definiendo P_i como la probabilidad de que $Y_i = 1$, el evento ocurra y $1 - P_i$ la probabilidad de que $Y_i = 0$, el evento no ocurra; por la definición de valor esperado obtenemos la ecuación:

$$E(Y_i) = 0(1 - P_i) + 1(P_i) = P_i$$

Igualando las dos ecuaciones previas conseguimos:

$$E(Y_i | X_i) = \beta_0 + \beta_1 X_i = P_i \quad (2.1)$$

De esta forma la expectativa condicional del modelo definido puede interpretarse como la probabilidad condicional de Y_i . Como el rango de valores de la probabilidad que el evento ocurra va de 0 a 1, se asume la siguiente desigualdad, $0 \leq E(Y_i | X_i) \leq 1$.

2.2. Regresión Logística Simple, Binaria y Multinomial

La aceptación de esta regresión se atribuye a que se basa en la función logaritmo natural el cual tiene la propiedad de ser una función monótona creciente y aplica valores en el intervalo $(0, \infty)$, pero al igual que una recta, se obtienen números reales.

Si tomamos la función logarítmica y la modelamos como una función lineal obtendremos la función logística, la expresaremos en la ecuación siguiente:

$$f(Y) = \frac{1}{1+e^{-y}}$$

Cualquier cantidad la pasa a un valor entre los límites 0 y 1 ($0 \leq f(Y) \leq 1$), como observamos a continuación.

- $Y = -\infty$ $f(-\infty) = \frac{1}{1+e^{-(-\infty)}} = \frac{1}{1+\infty} = 0.$
- $Y = 0$ $f(0) = \frac{1}{1+e^{-(0)}} = \frac{1}{1+1} = 0.5.$
- $Y = \infty$ $f(\infty) = \frac{1}{1+e^{-(\infty)}} = \frac{1}{1+0} = 1.$

La función logística tiene forma de “S” como podemos observar en las siguientes figuras.

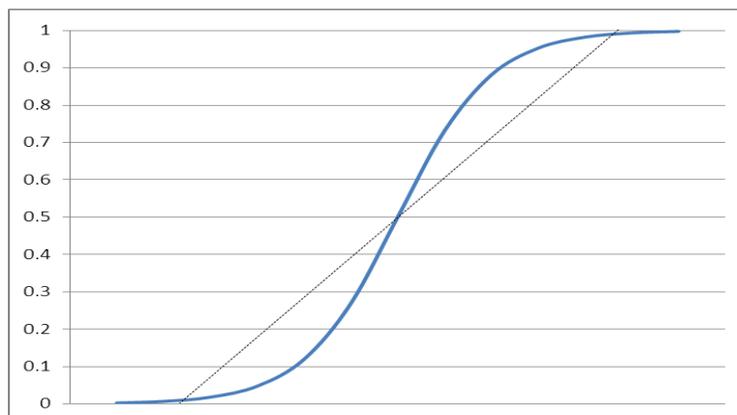


Figura 2.1 Función logística.

Como ya se ha mencionado la función logística indica la probabilidad de que ocurra la variable dependiente y sus valores siempre oscilarán entre el 0 y 1, para cualquier valor de X.

Para la regresión logística binaria simple con una sola variable independiente, el modelo se define de la siguiente manera:

$$f(Y) = \log \left[\frac{P(y=1)}{1-P(y=1)} \right] = \beta_0 + \beta_x \quad (2.2)$$

Los parámetros del modelo logístico sirven para calcular un parámetro de cuantificación de riesgo conocido como Odds ratio o razón de probabilidades. Para ilustrarlo tomemos en un ejemplo, cuando $P(y=1)=0.8$, la Odds equivale a $0.8/0.2=4$, esto nos indica que la probabilidad de ocurrencia es cuatro veces más que la probabilidad de no ocurrencia. Entonces el Odds asociado es el cociente entre la probabilidad de que ocurra un evento con la probabilidad de que no ocurra.

$$Odds = \frac{p_i}{1-p_i}$$

Con esta información podemos observar que en la función 2,2 se usa el logaritmo con la Odds; este proceso es nombrado transformación logística o logit para simplificar, el modelo se abrevia de la siguiente manera:

$$f(Y) = p(Y) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_i)}}$$

Esta es la función de ocurrencia. Ahora si lo que deseamos es la probabilidad de la no ocurrencia se abrevia de la forma:

$$1-P(y=1) = \frac{e^{-(\beta_0+\beta_1 X_i)}}{1+e^{-(\beta_0+\beta_1 X_i)}}$$

De acuerdo con las anteriores ecuaciones tenemos:

$$f(Y) = \log \left[\frac{P(y=1)}{1-P(y=1)} \right] = \left(\frac{1}{\frac{1+e^{-(\beta_0+\beta_1 X_i)}}{e^{-(\beta_0+\beta_1 X_i)}}} \right) = \log \left(\frac{1}{e^{-(\beta_0+\beta_1 X_i)}} \right) = \beta_0 + \beta_1 X_i$$

Con la información que hemos visto podemos asegurar que el parámetro β_1 es igual al log (OR) donde OR será Odds ratio que sabemos ya denota a la razón de probabilidad, y lo calculamos al exponenciar β_1 de la siguiente forma $e^{\beta_1} = e^{\log(OR)}$. En el caso de tener una tabla de 2x2, el OD estimada sin ajuste y el parámetro de regresión β_1 mantienen esta misma relación.

El parámetro β_0 es el valor de la constante y el parámetro β_1 nos indica el ascenso o descenso de la curva a medida que X incrementa su valor, podríamos compararla con la pendiente en la regresión lineal.

La ecuación 2.2 que se encuentra en escala logarítmica, al ser una función lineal de la variable explicativa, nos facilita estimar y hacer una interpretación del modelo.

De los sucesos de interés para un estudio, la mayoría suelen ser bastante complicados y es necesario usar más de una variable dependiente con la finalidad de plantear un solución a los fenómenos cuestionados; para estos casos, el modelo de regresión binaria multinomial nos permite incorporar un mayor número de variables independientes y comprender las variaciones en las respuestas obtenidas de los individuos. Al ser una extensión del binario es definido de la siguiente manera:

$$f(Y) = p(Y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}}$$

Podemos notar que $y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$.

2.3. Pruebas estadísticas para el modelo logístico

Al utilizar modelos se espera que los estimadores tengan una facultad discriminatoria, para cuantificar esa capacidad se hace uso de diferentes técnicas de prueba las cuales veremos.

2.3.1. Desviación

La desviación tiene la función $D(\beta) = -2L(\beta)$

$$D(\beta) = 2 \sum_{i=1}^n \left[\log \left(1 + e^{x_i^T \beta} \right) \right]$$

Y en términos de las probabilidades

$$D(\beta) = 2 \sum_{i=1}^n \left[y_i \log p_i + (1 - y_i) \log (1 - p_i) \right]$$

Esta nos da una medida de la desviación máxima del modelo.

2.3.2. Estadístico de Wald y el test de Hosmer-Lemeshow

Para saber si una variable tiene un peso significativo y debe ser incluida en el modelo, se hace uso de la prueba de estadístico de Wald. Esta consiste en contrastar la hipótesis nula contra la alternativa:

$$H_0 = \beta_i = 0$$

$$H_1 = \beta_i \neq 0$$

Con el estadístico de prueba:

$$w_j = \frac{\hat{\beta}_i}{s(\hat{\beta}_i)}$$

Con el supuesto de que H_0 es cierto, seguimos una distribución t con $n - p - 1$ grados de libertad, se distribuye como una normal para muestras grandes. Aquí lo que buscamos probar es que w_i tiene un valor alejado de cero, será evidencia de que H_0 es falsa, entonces la región crítica es de la forma $|w_j| > t_{\alpha/2}$ con un nivel de significancia apropiado. Al igual podemos expresar la región crítica con el p-value usado en la mayoría de los paquetes estadísticos, $p = P(t > |w_j|)$, su región crítica es de la forma $p < \alpha$.

La prueba de Wald también permite comprobar si cada nueva variable independiente mejora la capacidad predictiva del modelo. En muestras grandes tanto máxima verosimilitud y la prueba de Wald ofrecen resultados similares, aunque para muestras pequeñas la razón de máxima verosimilitud nos da una mayor precisión de la estimación.

El Test de Hosmer y Lemeshow es un test muy utilizado en Regresión logística. Se trata de un test de bondad de ajuste al modelo propuesto. Un Test de bondad de ajuste lo que hace es comprobar si el modelo propuesto puede explicar lo que se observa. Es un Test donde se evalúa la distancia entre un observado y un esperado.

El Test básicamente consiste en dividir el recorrido de valores de la variable dependiente (0,1) en una serie de intervalos. Intervalos que contengan un número de observaciones suficientemente grande (5 ó más). Se trata, entonces, de contar intervalo por intervalo el esperado y el observado para cada uno de los dos resultados posibles de la variable dependiente dicotómica (tiene la enfermedad o no la tiene, es hombre o mujer, etc). El observado es lo que se tiene y el esperado es el valor esperado teórico calculado mediante el modelo construido. El estadístico es un estadístico de la ji-cuadrado, como el visto en el tema dedicado a la relación entre variables cualitativas. De hecho, buena parte de los test de bondad de ajuste a un modelo parten de esta idea de comparar lo observado con lo esperado. (Jaume Llopis Pérez, (diciembre 2013), Test de Hosmer y Lemeshow.

Recuperado de <https://estadisticaorquestainstrumento.wordpress.com/2013/12/19/test-de-hosmer-y-lemeshow/>).

Construye una tabla de contingencia y aplica una prueba de contraste Chi- cuadrada, con grados de libertad igual a $n-2$

$$X^2 = \sum_{i=1}^g \frac{(O_{s,i} - E_{s,i})}{E_{s,i}} - \frac{(O_{f,i} - E_{f,i})}{E_{f,i}}$$

Donde $O_{s,i}$, $O_{f,i}$ los dos representan fracasos observados y $E_{s,i}$, $E_{f,i}$ el número esperado de éxitos y fracasos, usualmente el valor más usado por paquetes estadísticos son los deciles. El modelo no explica obligatoriamente una gran parte de la varianza de la variable dependiente, lo que hace es ver que la cantidad de varianza explicada resulte significativa.

A medida que el tamaño de la muestra aumenta, el test de Hosmer-Lemeshow encontrará menos diferencias entre los valores esperados, esto porque la Chi- cuadrada aumenta su valor cuando crece el tamaño de la muestra lo cual no es buscado en un estimador de bondad de ajuste.

2.4. Eligiendo un mejor modelo

Un buen modelo debe satisfacer dos condiciones, la primera es que tenga una fuerte capacidad predictora y la segunda es que la estimación de los parámetros tenga una alta precisión. Una condición adicional es que el modelo sea lo más sencillo posible, esto es que contenga el mínimo de variables explicativas y que satisfaga las dos condiciones anteriores. En este sentido se pregunta uno si todas las variables explicativas son necesarias para construir el modelo, o si alguna de ellas puede ser excluida. Para determinar que variables podrán ser excluidas se realiza una prueba de hipótesis. Entre otros se pueden usar dos métodos automatizados de selección de

variables: el forward que consiste en ir incluyendo variable cada vez y probar en cada paso si hay mejoras en el modelo o el backward que consiste en iniciar con todas las variables y se va excluyendo una variable en cada paso; de igual manera se va probando si no existe desmejora en el modelo. Estos dos métodos automatizados tienen la desventaja de que una variable que es incluida (excluida) ya no vuelve a salir (entrar) del modelo, esto podría restringir las opciones de seleccionar un mejor modelo, en este sentido, una alternativa es permitir incluir o excluir, variables que ya entraron o salieron, usando criterios estadísticos como el de correlaciones parciales. En todo caso, se pueden obtener todos los posibles modelos y escoger entre ellos, el mejor, aunque este método exhaustivo consume más tiempo, (Nieto, 2010, p.37).

Capítulo 3

Caso aplicado de la regresión logística.

“Ninguna investigación humana puede ser llamada verdadera ciencia si no puede ser demostrada matemáticamente.”

Leonardo da Vinci (s.f.)

3.1. Diseño Estadístico

El propósito de este trabajo de investigación, como ya ha sido mencionado, es el de hacer uso de la regresión logística para identificar variables que afectan en el incumplimiento de pago.

3.1.1. Aspectos Generales

El presente trabajo es un estudio de tipo observacional, los datos corresponde a una entidad financiera. La información abarca el periodo que comprende desde Julio del 2016 hasta diciembre del 2017, los estados de la República Mexicana en los que se llevó acabo la recolección de estos datos fueron: Aguascalientes, Campeche, Guerrero, Puebla, Veracruz, Tamaulipas, Colima, Morelos, Sinaloa, Durango, Oaxaca, Chiapas, Nayarit, Tlaxcala, Zacatecas, Jalisco y el Estado de México; estados donde la financiera tiene sucursales. Durante este lapso de tiempo se obtuvo

información de una gran cantidad de clientes a los que se les realizó un cuestionario y dio seguimiento de sus pagos durante distintos periodos.

3.1.2. Criterios de Inclusión

La financiera tenía como objetivo obtener información para poder estudiarla y crear un programa piloto el cual consistía en crear un préstamo para la construcción a personas sin comprobante de ingresos, en su mayoría eran personas laborando dentro de la economía informal, la información sería obtenida por un cuestionario y el seguimiento a los solicitantes del crédito en un tiempo determinado; por falta de recursos y dar prioridad a otros objetivos no se culminó la investigación, sin embargo, se otorgó la información con la condición de no hacer mal uso de los datos o de ocuparlos con fines de lucro.

3.1.3. Criterios de Exclusión

El cuestionario contenía una gran cantidad de variables pero en su mayoría eran de respuestas abierta y poco reveladoras en cuanto a su condición real, un ejemplo de esto era cuando se preguntaba acerca de los tipos de bienes materiales que poseían en sus viviendas, las respuestas eran muy distintas, algunos se limitaban a mencionar artículos de línea blanca mientras otros daban información muy completa que incluía desde el número de celulares.

Dentro del seguimiento a los pagos realizados, algunos sujetos no contaban con información de sus retribuciones lo que nos daba información incompleta, esto se alude a un error a la hora de captar los datos dentro de Excel o a pérdida de información en la institución.

Algunos sujetos no contaron con el seguimiento del comportamiento de cada cliente durante el lapso establecido.

Haciendo los ajustes obtuvimos un total de 1182 sujetos dentro de nuestra nueva muestra con los cuales decidimos hacer la investigación. Las variables con las que se trabajó fueron: estado civil, tipo de vivienda, nivel máximo de estudios, tipo de economía y préstamo o valor del proyecto. En el cuestionario para estos datos en especial se daban opciones concretas para escoger, para realizar la investigación las respuestas fueron canalizadas a una relación de números-respuestas, como se muestra en la siguiente tabla.

Tabla 3.1 Variables de Estudio.

Variable	Descripción	Valores
Estado civil	Situación del estado de la persona física determinada por su situación familiar.	1.Soltero 2.Casado 3.Divorciado
Tipo de vivienda	Edificación que habitan las personas y denota su relación con ella.	1.Familiares 2.Propia 3.Rentado 4.Prestada
Nivel máximo de estudios	Formación académica a la cual se tuvo acceso.	1.Primaria 2.Secundaria 3.Preparatoria 4.Licenciatura 5.Prefesional Tec 6.Posgrado 7.ninguno
Tipo de economía	Actividad laboral que puede estar regulada o no por el gobierno.	1.Formal 2.Informal 3.Mixta
Préstamo o valor del proyecto	Cantidad que fue solicitada para un préstamo personal.	De 10,000 a 150,000 pesos.

Se llevará a cabo un análisis de regresión logística con el propósito de ver una posible relación entre las variables explicativas y la probabilidad de que cumplan con su préstamo, con el apoyo del software con enfoque estadístico, R: The R Project for Statistical Computing, y de la

paquetería de Microsoft, en particular Excel. En Microsoft Excel se capturaron los datos y separaron los que serían usados, teniendo como punto de partida los aspectos antes mencionados, se acomodaron de manera que su manipulación fuera sencilla. De la forma en que se presentan los datos en Excel es fácil observar que varios sujetos de la muestra no pagaban el monto acordado en las fechas establecidas, situación que se notaba más del séptimo mes en adelante; sin embargo, al final del plazo señalado para cubrir con su préstamo diversos individuos lograron ponerse al corriente con los pagos.

En R se logró establecer un vínculo el cual permitió tomar los datos de Excel y empezar la simulación sobre la regresión logística, lo cual se hizo para proceder al análisis estadístico el cual consiste en implementar el modelo de regresión logística, ajustarlo y apoyados en uno de diversos métodos verificar sea viable; a continuación se desglosará el procedimiento.

3.2. Análisis Estadístico.

Esto lo abordamos separando en dos pasos:

- **Análisis Preliminar:**

En primer lugar se llevara un análisis estadístico univariado con el objetivo de ver el comportamiento de cada una de las variables de estudio.

- **Análisis Definitivo:**

Como análisis definitivo se realizará un estudio de regresión logística con el objetivo de ver posible relación entre las variables de estudio y la probabilidad de incumplimiento de pago.

Capítulo 4

Resultados

“Si buscas resultados distintos, no hagas siempre lo mismo.”

Albert Einstein (s.f.)

En este capítulo se presentan los resultados obtenidos del análisis preliminar y del análisis definitivo.

4.1. Análisis Preliminar

En primer lugar se presentan los resultados del análisis univariado de cada una de las variables bajo estudio. Posteriormente en esta sección se presentan los resultados del análisis bivariado.

4.1.1. Análisis Univariado

La siguiente imagen muestra la distribución del estado civil de los individuos que contratan un préstamo.

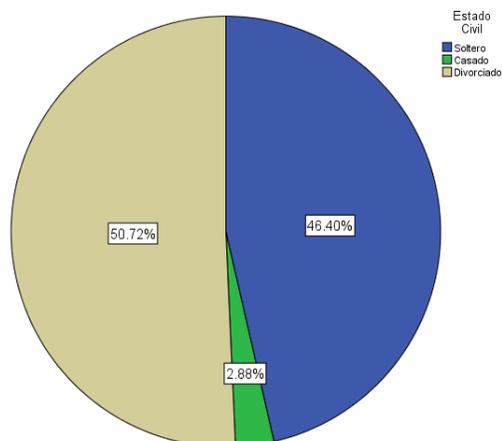


Figura 4.1 Distribución del Estado Civil.

Respecto al Estado Civil de los clientes, en la Figura 4.1 se observa que aproximadamente un 46% de los clientes son solteros, mientras que poco más de 50% mencionaron ser divorciados; observamos que estas dos categorías tienen números muy similares.

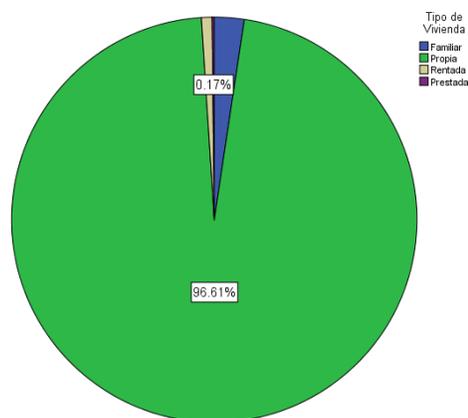


Figura 4.2 Distribución Tipo de Vivienda.

De la figura 4.2 se observa las cuatro categorías que formalizan la variable “tipo de vivienda”, es claro que la mayor parte de individuos que tienen la deuda están dentro de los que poseen una vivienda propia, las restantes tres categorías juntas apenas superan el 3%.

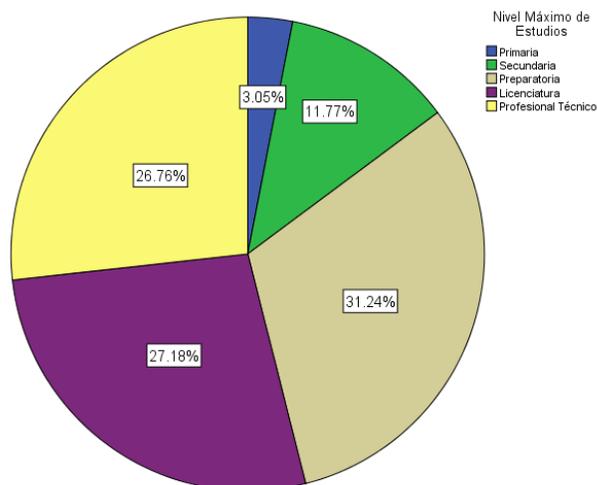


Figura 4.3. Distribución Nivel de Estudios.

La distribución del nivel de estudios se le puede observar en la Figura 4.3. Nótese que el nivel de estudios de educación primaria presenta aproximadamente por 3% de los clientes, Los niveles de estudio de preparatoria, licenciatura y profesional técnico son similares en proporción. Sin embargo, los que solo poseen preparatoria como máximo nivel son la mayoría con un 31.25%.

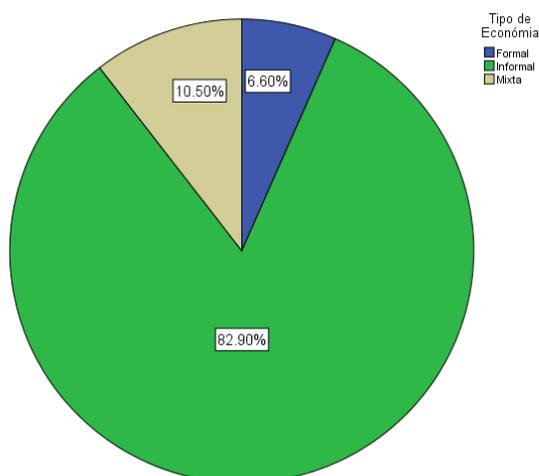


Figura 4.4. Distribución Tipo de Economía

El tipo de economía como se mencionó en la justificación de esta investigación, en su mayoría pertenecía a la economía informal del país, en la Figura 4.4 se encuentra dividido en 3 categorías que son: formal, informal, mixta. Se observa que la mayoría de los sujetos se encuentra en la economía informal con un 82.90%, ya que eran el grupo de interés.

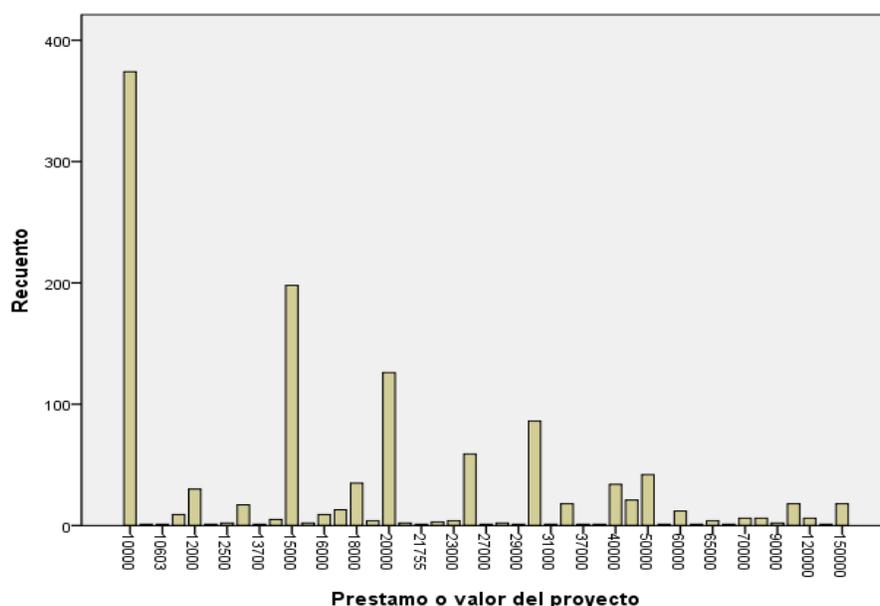


Figura 4.5. Distribución Valor del Proyecto.

Respecto al préstamo que se pidió a la institución financiera, en la Figura 4.5 se observa que el mayor número se concentra en \$10,000 con aproximadamente el 36%, se observa que la cantidad de \$15,000 también es ampliamente solicitada por los clientes.

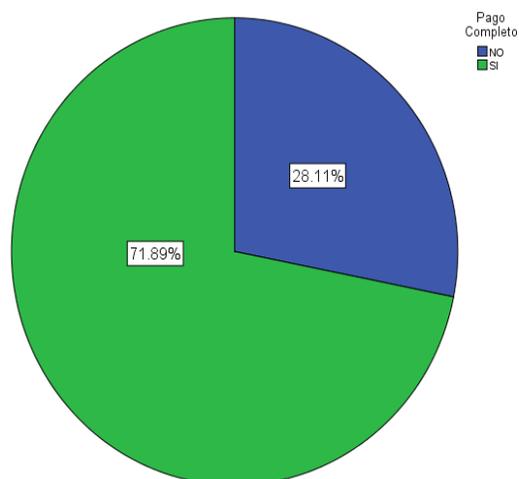


Figura 4.6. Pagos Completos

Luego observamos en la Figura 4.6 de manera simple la cantidad de personas que pagaron y las que no realizaron el pago por completo. El porcentaje de pago fue del 71.88%, lo cual es un cantidad muy buena.

4.1.2. Análisis Bivariado.

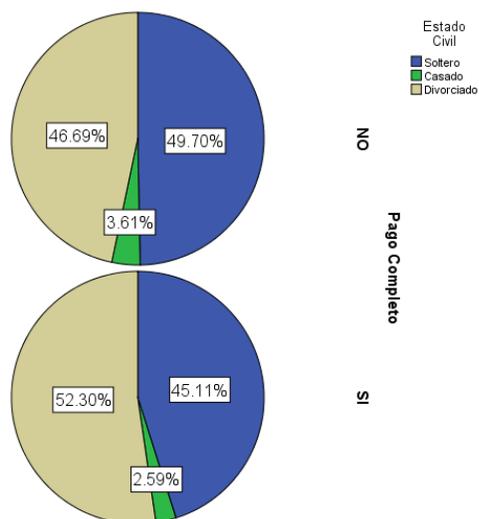


Figura 4.7 Distribución Pagos de Estado Civil General.

En la Figura 4.7 se observa que la distribución de estado civil es similar para los que cumplieron al final y de los deudores. En ambos casos aproximadamente el 46% de los clientes son solteros mientras que alrededor del 50% son divorciados.

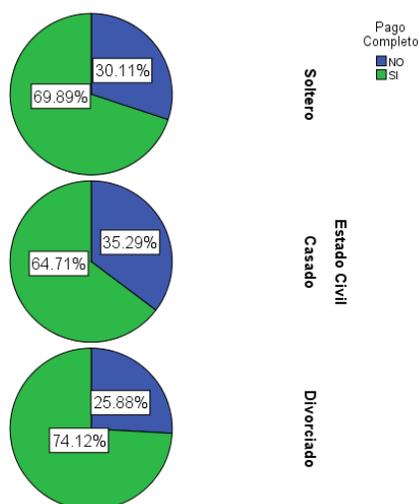


Figura 4.8 Distribución Pagos de Estado Civil, Particular.

Respecto a la distribución del cumplimiento de pago para cada una de las categorías del Estado Civil, se observa en la Figura 4.8 que para los divorciados y solteros el comportamiento es similar, siendo aproximadamente el 72 % de los clientes que cumplieron con el pago.

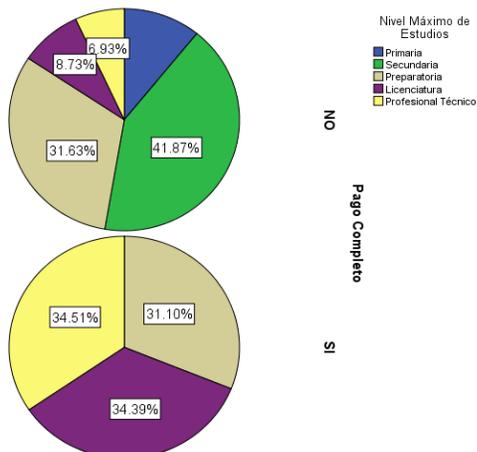


Figura 4.9 Estructura General de Nivel Máximo de Estudios.

En la Figura 4.9 se aprecia que en la distribución aquellos que pagaron es ligeramente diferente a los que no cumplieron, la diferencia más notoria se da en los de nivel secundaria que en mayor parte no cumplieron con la obligación financiera adquirida.

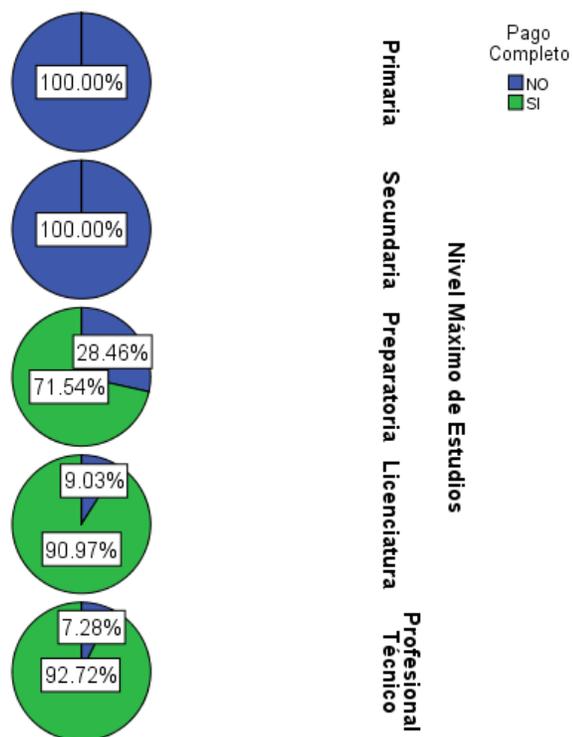


Figura 4.10 Pagos Nivel Máximo de Estudios.

Se observa en la Figura 4.10 algo interesante, secundaria y primaria son los que han mostrado comportamientos semejantes y ser los dos que más incumplimiento presentaron, aunque ligeramente comparado con licenciatura pero aquellos con estudios a nivel posgrado han cumplido más individuos en la categoría.

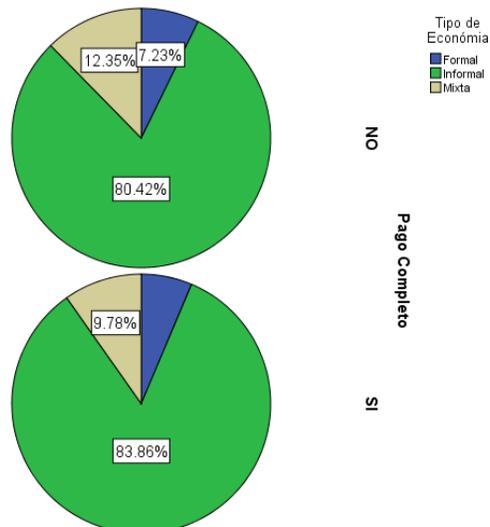


Figura 4.11 Distribución Pagos de Nivel Tipo de Economía.

En la variable tipo de economía para la categoría de informal se había mencionado que era la dominante por fines de estudio, la Figura 4.11 nos da una idea de que tan recurrente era sutilmente mayor el porcentaje de los que si pagan. Formal y mixta muestran tendencias parecidas aunque en menor porcentaje.

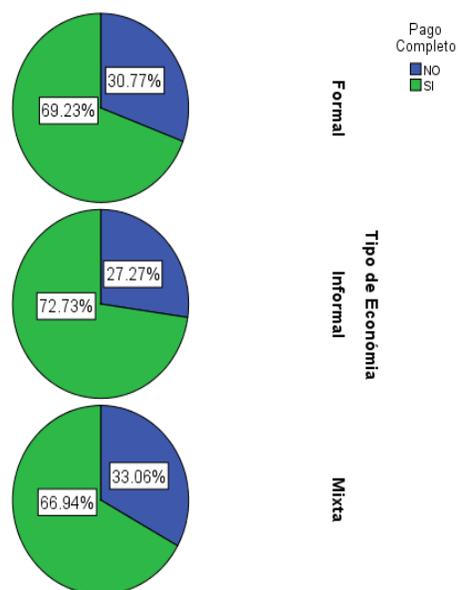


Figura 4.12 Pagos Completados de Nivel Tipo de Economía.

En la Figura 4.12 se presento un mayor porcentaje de cumplimiento que fue la categoría de informal.

Se mencionó antes en este analisis la Figura 4.12 nos da una mayor idea de que ha pasado con las restantes categorías, ahora, podemos decir que los de economía formal tienen mayor porcentaje de pagos completos lo cual solo se aprecia cuando estudiamos categoría por categoría. Ahora podemos asegurar que las tres cumplieron en su mayoría a pesar de ser diferentes categorías.

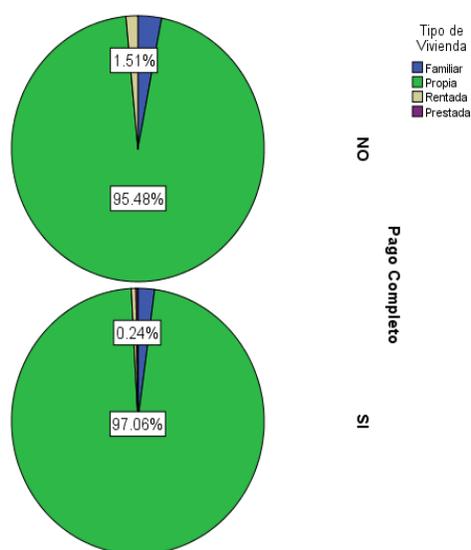


Figura 4.13 Distribución Pagos de Nivel Tipo de Vivienda, General

Para la Figura 4.13 se le asignó la variable tipo de vivienda, se puede observar que el grueso de la población se encuentra dentro de la categoría de propia, por eso aunque es un gran porcentaje de los que no pagaron no significa que en su mayoría no pagarán si no que una gran cantidad pertenece a este grupo, en el análisis individual se observa mejor. Ahora de manera similar a casos anteriores en lo particular podremos observar mejor como se comportan las otras categorías ya que por su cantidad baja de individuos no se puede ver bien.

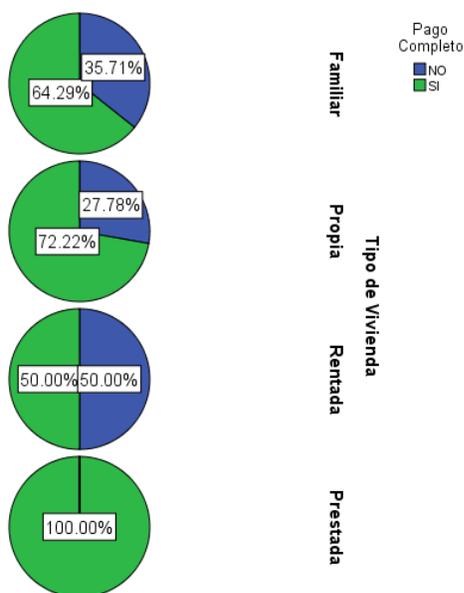


Figura 4.14 Distribución Pagos de Nivel Tipo de Vivienda, Particular

Como se recalco antes el análisis particular de la Figura 4.14 nos muestra facilmente y con mayor detalle el comportamiento de la categorías dentro de la variable. En propia vemos que aproximadamente el 72.22% fue el que cumplio y es bueno, representar la parte más grande de nuestro grupo; en rentada vemos se comportaron de manera completamente igual en porcentajes y al ser la vivienda de un familiar su mayor parte que en este caso fue de 64.29% cumplió con el pago. Ahora es bueno señalar que gracias al análisis individual podemos observar caso como el de una vivienda prestada donde se cumplió al 100% con el pago, esto no se notaba tan facil en la figura anterior.

4.2. Regresión logística.

Una vez revisada la estructura y conociendo las variables a estudiar pasamos a generar el modelo de regresión logística, apoyándonos en el modelo lineal generalizado que está relacionado con la variable de respuesta a través de una función de enlace, el modelo que ajustaremos será:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} \text{ con } i = 1, \dots, n \quad (1)$$

A continuación se presentan las estimaciones de los parámetros del modelo (1).

Tabla 4.1

Estimaciones de coeficiente del modelo (1)

Parámetro	Estimación	P-value
α	-5.717	< 0.0005
β_1	0.04512	0.603
β_2	0.7543	0.114
β_3	1.829	< 0.0005
β_4	-0.1717	0.401
β_5	-0.0000259	< 0.0005

De la Tabla 4.1 se tiene que varias posibles variables explicatorias no son significativas estadísticamente. La variable Estado Civil debe de ser eliminada del modelo, ya que es la que presenta el mayor p-value (0.603). Así que ahora se ajusta el modelo.

$$Y_i = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} \text{ con } i = 1, \dots, n \quad (2)$$

A continuación se presentan las estimaciones de los parámetros del modelo (2).

Tabla 4.2
Estimaciones de coeficiente del modelo (2)

Parámetro	Estimación	P-value
α	-5.585	< 0.0005
β_2	0.7309	0.124
β_3	1.831	< 0.0005
β_4	-0.1713	0.402
β_5	-0.0000259	< 0.0005

Ahora en la Tabla 4.2 procedemos a descartar la siguiente variable explicitaría que no es significativa la cuál es β_4 , tipo de economía, ya que presenta el p-value más alto.

$$Y_i = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_5 X_{i5} \text{ con } i = 1, \dots, n \quad (3)$$

A continuación se presentan las estimaciones de los parámetros del modelo (3).

Tabla 4.3
Estimaciones de coeficiente del modelo (3)

Parámetros	Estimación	P-value
α	-5.919	< 0.0005
β_2	0.7307	0.124
β_3	1.826	< 0.0005
β_5	-0.00002597	< 0.0005

Se observa dentro la Tabla 4.3 que la probabilidad del valor estadístico en β_2 es la de mayor valor, en el próximo modelo prescindimos de anterior parámetro.

$$Y_i = \beta_0 + \beta_3 X_{i3} + \beta_5 X_{i5} \text{ con } i = 1, \dots, n \quad (4)$$

A continuación se presentan las estimaciones de los parámetros del modelo (4).

Tabla 4.4

Estimaciones de coeficiente del modelo (4)

Parámetros	Estimación	P-value
α	-4.445	< 0.0005
β_3	1.829	< 0.0005
β_5	-0.00002562	< 0.0005

Congruente a nuestro modelo, el logaritmo de odds podemos observar que el nivel de estudios está positivamente relacionado con el hecho de cumplir con las obligaciones financieras ya que tiene un parámetro parcial de 1.816, y parece ser significativa la relación ya que el p-valor < 0.005. De la misma forma podemos ver una relación negativa (0.00002595) entre el logaritmo de odds del valor del proyecto (p-valor = 0.0000000000464).

Esto se puede ver desde la siguiente perspectiva, en concreto los odds que necesita un individuo de una categoría menor en nivel de estudios es de $e^{-(1.829)} = 6.22765589$ mayores que la próxima categoría.

$$\text{Logit} = \beta_0 + (1.816)X_3 + (.00002562)X_5$$

Después haciendo uso de la ecuación, previamente explicada en el capítulo anterior, se aplica a las categorías de los 1181 individuos en la muestra. La ecuación tomara la siguiente forma:

$$P(X = 1) = \frac{e^{-(\beta_0 + (1.816)X_3 + (.00002562)X_5)}}{1 + e^{-(\beta_0 + (1.816)X_3 + (.00002562)X_5)}}$$

En el modelo logístico se pueden obtener diversos resultados y entre ellos valores negativos, a continuación se muestra la instrucción que se da a nuestro software y su resultado.

```
> predict(modelo1)
      1      2      3      4      5      6      7
0.74659349 2.56251299 4.37843250 -2.88524552 0.74659349 0.74659349 2.56251299
      8      9     10     11     12     13     14
4.37843250 0.74659349 2.56251299 -1.06932602 2.56251299 4.37843250 -1.06932602
     15     16     17     18     19     20     21
-1.06932602 -2.88524552 4.37843250 -1.06932602 2.56251299 2.56251299 2.56251299
     22     23     24     25     26     27     28
4.37843250 2.56251299 0.74659349 0.74659349 -1.06932602 0.74659349 4.37843250
     29     30     31     32     33     34     35
4.37843250 2.56251299 2.56251299 0.74659349 2.56251299 2.56251299 2.56251299
     36     37     38     39     40     41     42
-1.06932602 -1.06932602 -2.88524552 0.74659349 0.74659349 0.74659349 4.37843250
     43     44     45     46     47     48     49
2.56251299 0.74659349 0.74659349 4.37843250 4.37843250 2.56251299 2.56251299
     50     51     52     53     54     55     56
2.56251299 4.37843250 0.74659349 2.56251299 -1.06932602 4.37843250 2.56251299
     57     58     59     60     61     62     63
2.56251299 2.56251299 0.74659349 0.74659349 0.74659349 4.37843250 -1.06932602
     64     65     66     67     68     69     70
0.74659349 0.74659349 4.37843250 4.37843250 2.56251299 0.74659349 -1.06932602
     71     72     73     74     75     76     77
0.74659349 2.56251299 4.37843250 0.74659349 4.37843250 -1.06932602 0.74659349
     78     79     80     81     82     83     84
4.37843250 -1.06932602 2.56251299 -1.06932602 2.56251299 2.56251299 -2.88524552
     85     86     87     88     89     90     91
```

```
    1093    1094    1095    1096    1097    1098    1099
1.53772139 -2.09411762 1.53772139 -2.09411762 -0.27819812 -0.27819812 -0.27819812
    1100    1101    1102    1103    1104    1105    1106
3.35364089 1.53772139 -2.09411762 -2.09411762 -0.27819812 -2.09411762 3.22554194
    1107    1108    1109    1110    1111    1112    1113
-4.16623503 -2.35031552 -0.53439602 3.09744299 -4.16623503 -4.16623503 1.28152349
    1114    1115    1116    1117    1118    1119    1120
1.28152349 3.09744299 1.28152349 -0.53439602 3.09744299 -2.40155510 -2.47841447
    1121    1122    1123    1124    1125    1126    1127
-0.66249497 -2.47841447 -0.66249497 -4.33701655 -0.79059392 -0.79059392 2.84124509
    1128    1129    1130    1131    1132    1133    1134
-2.60651342 2.84124509 -2.60651342 -2.86271133 2.58504719 -2.86271133 2.58504719
    1135    1136    1137    1138    1139    1140    1141
0.76912768 -1.04679182 0.51292978 -3.11890923 -1.55918762 -5.19102663 0.25673188
    1142    1143    1144    1145    1146    1147    1148
-3.37510713 -3.37510713 0.25673188 0.25673188 2.07265139 -3.37510713 2.07265139
    1149    1150    1151    1152    1153    1154    1155
-1.55918762 -3.37510713 -5.19102663 -3.37510713 0.25673188 -5.19102663 -3.37510713
    1156    1157    1158    1159    1160    1161    1162
-1.55918762 -5.70342243 -2.07158342 -3.88750293 1.56025558 -5.70342243 -5.70342243
    1163    1164    1165    1166    1167    1168    1169
-4.27179978 -1.02425762 -4.65609663 -4.65609663 -1.02425762 -6.47201614 -2.84017713
    1170    1171    1172    1173    1174    1175    1176
0.79166188 0.79166188 0.79166188 0.79166188 -4.65609663 -6.47201614 0.79166188
    1177    1178    1179    1180    1181
-4.65609663 -2.84017713 0.79166188 -4.65609663 -2.84017713
```

Como modelamos la probabilidad de la variable respuesta mediante el uso de LOG of ODDS, en los Odds se define como la razón de probabilidad verdadera, que es la proporción entre la probabilidad de un evento verdadero y la probabilidad de evento falso $\frac{p}{q}$.

La transformación de probabilidad a Odds conserva una relación, si los Odds aumentan también la probabilidad y recíprocamente, es decir, que la correspondencia es monotónica. El rango de valores que puede tomar las Odds es mayor al rango de la probabilidad, como está acotado entre $[0, 1]$ se requiere de una transformación logit, aunque se pueden usar otras, esta consiste en usar el logaritmo natural sobre las Odds, con este el rango de probabilidad se puede convertir entre $[-\infty, \infty]$.

Por dar un ejemplo si la probabilidad es igual a 0.001 su Odds será de 0.001001 y el $\text{Log}(\text{Odds})$ sería equivalente a -6.906755, ahora si nos vamos al otro extremo, una p con 0.9999 tendrá su Odds en 9999 y el $\text{Log}(\text{Odds})$ de 9.21024. Debemos tener en cuenta algunas características de estos datos:

- A diferencia de la probabilidad que no puede exceder de 1, los Odds no tienen límite superior.
- La transformación logit no existe para $p = 0$

Usaremos un comando que nos permitirá pasar nuestros valores del “predict” a valores de probabilidad acotados en $[0, 1]$ para los 1181 valores. Esto se consigue usando `type = “response”` el objeto de respuesta captura toda la información de una solicitud.

```
> predict(modelo1, type="response")
  1         2         3         4         5         6         7
0.678435984 0.928409664 0.987610420 0.052887768 0.678435984 0.678435984 0.928409664
  8         9        10        11        12        13        14
0.987610420 0.678435984 0.928409664 0.255531278 0.928409664 0.987610420 0.255531278
 15        16        17        18        19        20        21
0.255531278 0.052887768 0.987610420 0.255531278 0.928409664 0.928409664 0.928409664
 22        23        24        25        26        27        28
0.987610420 0.928409664 0.678435984 0.678435984 0.255531278 0.678435984 0.987610420
 29        30        31        32        33        34        35
0.987610420 0.928409664 0.928409664 0.678435984 0.928409664 0.928409664 0.928409664
 36        37        38        39        40        41        42
0.255531278 0.255531278 0.052887768 0.678435984 0.678435984 0.678435984 0.987610420
 43        44        45        46        47        48        49
0.928409664 0.678435984 0.678435984 0.987610420 0.987610420 0.928409664 0.928409664
 50        51        52        53        54        55        56
0.928409664 0.987610420 0.678435984 0.928409664 0.255531278 0.987610420 0.928409664
 57        58        59        60        61        62        63
0.928409664 0.928409664 0.678435984 0.678435984 0.678435984 0.987610420 0.255531278
 64        65        66        67        68        69        70
0.678435984 0.678435984 0.987610420 0.987610420 0.928409664 0.678435984 0.255531278
 71        72        73        74        75        76        77
0.678435984 0.928409664 0.987610420 0.678435984 0.987610420 0.255531278 0.678435984
 78        79        80        81        82        83        84
0.987610420 0.255531278 0.928409664 0.255531278 0.928409664 0.928409664 0.052887768
 85        86        87        88        89        90        91
```

```
 1093      1094      1095      1096      1097      1098      1099
0.823133238 0.109669873 0.823133238 0.109669873 0.430895587 0.430895587 0.430895587
 1100      1101      1102      1103      1104      1105      1106
0.966223859 0.823133238 0.109669873 0.109669873 0.430895587 0.109669873 0.961784232
 1107      1108      1109      1110      1111      1112      1113
0.015273645 0.087040696 0.369492171 0.956787148 0.015273645 0.015273645 0.782708996
 1114      1115      1116      1117      1118      1119      1120
0.782708996 0.956787148 0.782708996 0.369492171 0.956787148 0.083054189 0.077385328
 1121      1122      1123      1124      1125      1126      1127
0.340179374 0.077385328 0.340179374 0.012906719 0.312041158 0.312041158 0.944864362
 1128      1129      1130      1131      1132      1133      1134
0.068720402 0.944864362 0.068720402 0.054027960 0.929893020 0.054027960 0.929893020
 1135      1136      1137      1138      1139      1140      1141
0.683332164 0.259841635 0.625493032 0.042333972 0.173763249 0.005535477 0.563832747
 1142      1143      1144      1145      1146      1147      1148
0.033082551 0.033082551 0.563832747 0.563832747 0.888216483 0.033082551 0.888216483
 1149      1150      1151      1152      1153      1154      1155
0.173763249 0.033082551 0.005535477 0.033082551 0.563832747 0.005535477 0.033082551
 1156      1157      1158      1159      1160      1161      1162
0.173763249 0.003323452 0.111889596 0.020084796 0.826390025 0.003323452 0.003323452
 1163      1164      1165      1166      1167      1168      1169
0.013764535 0.264198897 0.009414020 0.009414020 0.264198897 0.001543719 0.055191301
 1170      1171      1172      1173      1174      1175      1176
0.688188057 0.688188057 0.688188057 0.688188057 0.009414020 0.001543719 0.688188057
 1177      1178      1179      1180      1181
0.009414020 0.055191301 0.688188057 0.009414020 0.055191301
```

4.2.1. Comparando Resultados.

El cut off de prueba que se sugiere es del 0.5 para que fuera la misma probabilidad y ver cómo se comporta el modelo, pero para este estudio empezaremos con un cut off de 0.7, observaremos el procedimiento de comparación.

Empezaremos con almacenar el modelo ya aplicado con los parámetros en rangos de probabilidad, para almacenarlo lo denotaremos con el nombre de “prob”, inmediatamente creamos una nueva variable a la cual se denominó como “clasif”, a diferencia de la anterior en esta variable estableceremos el cut off; después observamos la tabla de esta última variable.

```
> prob=predict(modelo1,type="response")
> clasif=ifelse(prob<0.7,0,1)
> table(clasif)
clasif
  0   1
558 623
```

En la tabla anterior si la probabilidad predicha es mayor a 0.7 se asigna a la categoría 1 (cumple con su obligación financiera), en cambio si esta es menor se mandará a la categoría 0 (no cumple con su obligación financiera). Después con la línea “table (pagos\$tipo)” obtenemos las observaciones reales de nuestro marco teórico.

```
> table(pagos$tipo)
  0   1
332 849
```

Se procede a cruzar las tablas de observaciones y la predicción; ahora en la coordenada (0, 1) nos dice que 40 individuos fueron clasificados como 1 por la predicción cuando en las observaciones eran 0, por otra parte la coordenada (1, 0) señala que 240 personas fueron señaladas con 1 cuando su verdadero valor es 0.

```
> table(pagos$tipo, clasif)
  clasif
    0    1
0 284  48
1 274 575
```

Con lo obtenido en la tabla cruzada y sus errores tenemos que calcular la tasa de mala clasificación, para hacerlo definimos “tmc” como la mala clasificación y multiplicar por 100 para obtener el porcentaje.

```
> tmc=1-sum(pagos$tipo==clasif)/1181
> tmc*100
[1] 27.26503
```

Con un cut off del 0.7 obtenemos un error del 27.26503% el cual es alto así que haremos otra comparación con un nuevo cut off.

Nuestro nuevo límite será de 0.6, procederemos repetir el procedimiento anterior y variando el nombre asignado a nuestra variable incrementando un 1 al igual que asignando el nuevo cutoff.

```
> clasif1=ifelse(prob<0.6,0,1)
> table(clasif1)
clasif1
  0    1
291 890
```

Volvemos revisar las observaciones originales ya que es necesario para poder hacer una comparación con nuestra nueva sentencia.

```
> table(pagos$tipo)
  0    1
332 849
```

Podemos observar que ahora la diferencia es muy grande, aunque en la coordenada (0, 1) aumento a más del doble el error de personas asignadas a 1 cuando eran fracasos en nuestro estudio,

claramente el error bajo drásticamente en la coordenada (1, 0) prácticamente bajo a más de 4 veces la mala colocación, anteriormente 274 que pertenecían a 1 se clasificaron como 0 en cambio ahora solo 68, gran diferencia entre ambos cut off

```
> table(pagos$tipo, clasif1)
      clasif1
      0      1
0 223 109
1  68 781
```

Para la tasa de mala clasificación se repiten los pasos previos del anterior cut off buscando el nuevo porcentaje de error que nos da esta predicción.

```
> tmc1=1-sum(pagos$tipo==clasif1)/1181
> tmc1*100
[1] 14.9873
```

Con esto podemos comparar que el cut off de 0.7 tiene un error del 27.26503% y el de 0.6 solo un 14.9873%, la forma en que se distribuyen los errores de clasificación están más equilibrados en esta última nos basaremos es el cut off de 0.6 para decir lo siguiente, el modelo es capaz de identificar correctamente 781 de los 849 individuos que completan sus pagos en el plazo final establecido. El porcentaje de falsos negativos es bajo aunque tal vez seleccionar otro cut off podría mejorar la exactitud del modelo.

Otra forma de comparar los resultados es evaluando el modelo, en este estudio se opta por la Razón de Verosimilitud (Likelihood Ratio), sabemos que los parámetros de medición más conocidos en las pruebas estadísticas son la sensibilidad, la especificidad y los valores predictivos; las razones de verosimilitud conforman otra herramienta que resume la precisión de una prueba diagnóstica. Pasaremos a definir las variables necesarias que nos darán la información de diferencia en residuos, grados de libertad y el p-valor. Todo esto aplicado al modelo1 el cual contiene solo a las variables X3 y X5 como variables explicativas.

```

> dif_res=modelo1$null.deviance-modelo1$deviance
>
>
>
> df=modelo1$df.null-modelo1$df.residual
>
>
>
> p_value=pchisq(q=dif_res,df=df,lower.tail=FALSE)

```

El siguiente paso es que nos muestre de una manera más sencilla los resultados, el cual nos arroja la diferencia de residuo, que no es otra cosa más que las diferencias entre los valores de la variable dependiente y los valores que predecimos a partir de nuestra recta de regresión, el estudio de residuos es una herramienta formidable en el análisis de las regresiones, nos ayudan a encontrar casos desviantes y puntos influyentes, en nuestro caso la diferencia de residuos que tenemos es de 531.56606, el cual está alejado de la recta de regresión.

Después vemos los grados de libertad los que podemos verlos como espacios mediante los cuales la unidad de medida resumen puede variar y exponer diferentes valores. Desde el punto de vista algebraico se entiende como la cantidad de ecuaciones establecidas usando los datos, aunque de manera más simple diremos que es la cantidad de información provista por los datos que se pueden usar para preciar los parámetros desconocidos de la población y calcular la variabilidad de las estimaciones. En este caso nuestros grados de libertad son de 2.

Como ya hemos mencionado el p-valor nos mide el nivel de significancia estadística, aunque nos presenta como 0 grados el verdadero valor que tomó fue de $4e-116$ y de esa forma lo dejamos debido a su extensión.

```

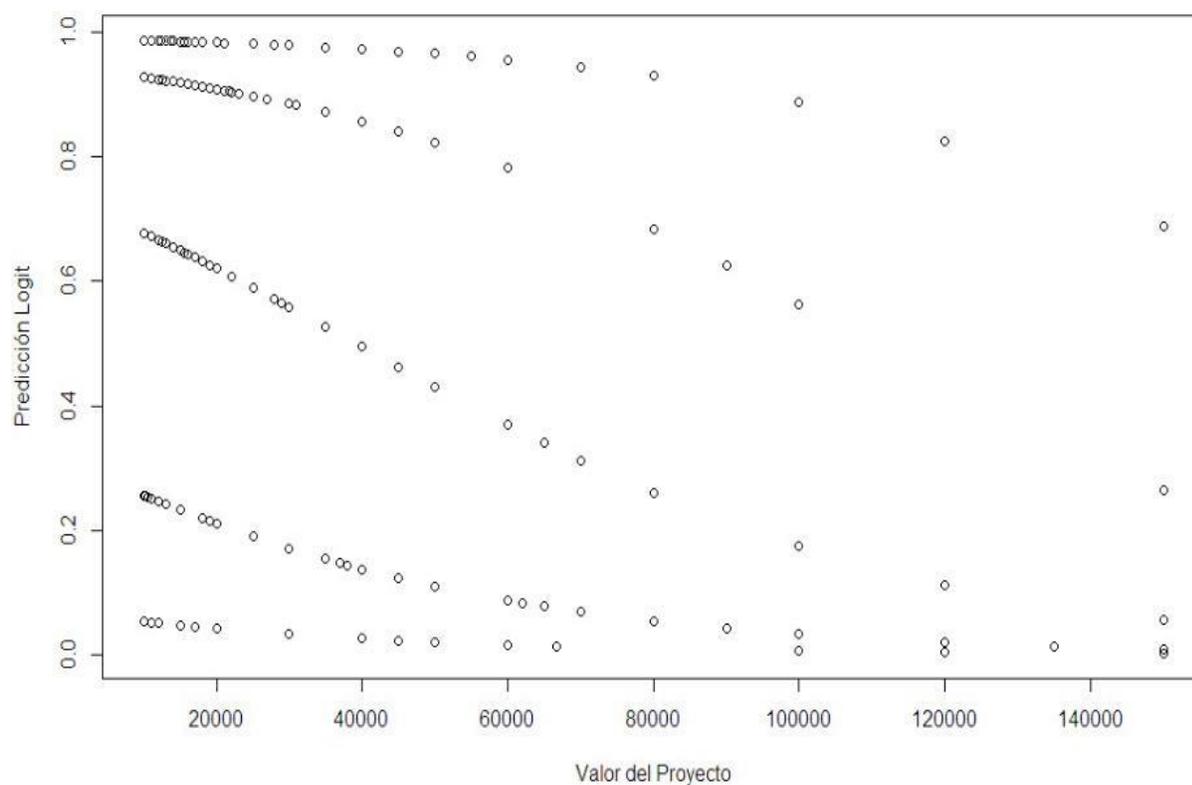
> paste("Diferencia Residuos:",round(dif_res,5))
[1] "Diferencia Residuos: 531.56606"
>
> paste("Grados de Libertad:",df)
[1] "Grados de Libertad: 2"
>
> paste("p-value:",round(p_value,115))
[1] "p-value: 0"

```

Después de esto podemos decir que el modelo en conjunto si es significativo y de acuerdo a nuestro modelo los predictores nivel de estudios (X3) y valor del proyecto (X5) tiene una significativa contribución en esta regresión logística.

3.2.2. Comportamiento de la regresión.

Para poder entender mejor el comportamiento graficaremos las variables, sin embargo como mencionamos antes, para poderlo hacer fijamos una de nuestras dos variables predictivas. Se eligió al valor del proyecto como la variable fija. Lo que haremos es que los predictores para el nivel de estudio se moverán con la finalidad de que podamos observar su comportamiento para cada una de sus 5 categorías en las que poseían valores.



Ahora si nos fijamos con detalle se puede observar en esta grafica de dispersión cómo se comportan las 5 categorías, primaria, secundaria, preparatoria, licenciatura y profesional técnico. Podemos notar algunos comportamiento de manera inmediata, la primera es que entre mayor nivel de estudios existe una relación directa con la probabilidad de pago, aquellos que poseen una licenciatura o técnica profesional poseen un aproximado de 65% más probabilidad de cumplir con sus obligaciones financieras que alguien que solo cuente con educación básica.

La segunda situación que se presenta está relacionada con la cantidad de préstamos, entre mayor sea la cantidad, la probabilidad de pago disminuye drásticamente cuando se sobrepasan los 60,000, en cualquiera que sea la categoría de la variable X3, más adelante concluiremos con detalle.

Con esto se da por terminada la aplicación de la regresión logística, su estructura y la interpretación; resumiendo es una técnica bastante robusta y ventajosa de usar sobre todo si nos apoyamos en algún software ya que al ser una regresión práctica se puede adaptar a un gran número de aplicaciones, también de su capacidad ahorrando costos y tiempo.

CONCLUSIONES.

La elaboración de esta tesis, no se hizo en equipo, pero si se ocuparon algunos trabajos previos hechos por otras personas para guiar la construcción de este trabajo de investigación las cuales se incluyen en la bibliografía.

De los capítulos en la tesis el tercero es de gran relevancia por mostrar la aplicación de un tema que se aborda durante la formación académica dentro de la facultad de ciencias físico matemáticas, siendo un tema de gran importancia por su aplicación en el ámbito laboral.

Recordando el objetivo principal el cual era determinar si las variables propuestas al inicio estado civil, tipo de vivienda, nivel máximo de estudios, tipo de economía y préstamo o valor del proyecto; afectaban la probabilidad de cumplir con el pago del préstamo. Podemos asegurar después de nuestro análisis que las variables nivel máximo de estudios y el valor del proyecto son las que más influyen en la probabilidad de éxito al efectuar el cumplimiento de la responsabilidad financiera.

El cut off se puede calcular siguiendo un proceso pero en algunos casos la empresa o la persona que requiere el estudio es quien lo solicita, en este caso no se dio alguno ya que no fue citado solo se expuso un ejemplo para poder revisar a detalle el modelo con fines de estudio.

Como se mencionó, de nuestro marco muestral un gran porcentaje pertenecía a la economía informal, un 82.89% razón por la que no fue estadísticamente significativa, y el 60% del total de individuos en la muestra se encontraban en esta categoría y cumplieron con el plazo acordado dicho de otra forma solo 267 fueron clientes morosos. Como el número de pago es elevado podría plantearse y analizar a fondo la posibilidad de crear un crédito especial que cuente con ciertas condiciones, un ejemplo sería que la cantidad no exceda los \$60,000 en un lapso de tiempo menor a un año porque en la gráfica se puede apreciar que al pasar esta cota cae de manera drástica el porcentaje de cumplimiento; el préstamo estaría dirigido a personas que no cuentan con un comprobante de ingresos o se encuentren en la economía informal. Sería un caso interesante aplicarlo ya que actualmente la economía informal sigue en crecimiento y la implementación de estos créditos sería beneficiosa para elevar la calidad de vida en este sector y generar una derrama económica sin correr riesgo de perder la inversión que hace la institución financiera.

Standard & Poor's Ratings prevé que la cartera vencida de clientes morosos de la banca en México oscile entre 2.3 y 2.5 por ciento en 2019 y 2020, aún con esto las personas dentro de la economía formal tiene menor probabilidad de caer en el incumplimiento financiero a comparación del sector informal, pero como antes se menciona podría considerarse un crédito especial y mejor estructurado que compitan con los que actualmente se trabajan en grandes instituciones bancarias y volver a comparar el porcentaje de éxito al cumplimiento.

En general podemos decir que los resultados obtenidos son útiles para ser aplicados en la práctica por diversas instituciones financieras, puede no ser el método más recomendable pero si es adecuado para la información y derivaciones obtenidas.

BIBLIOGRAFÍA.

- [1] Abraira, V. y Pérez de Vargas, A. (1996). *Métodos Multivariantes en Bioestadística*. Madrid: Centro de Estudios Ramón Areces.
- [2] Banco de México, *Definiciones básicas de Riesgos*, (2005).
- [3] Crédito al consumo: la estadística aplicada a un problema de riesgo crediticio. Soraida Nieto. 2010
- [4] Crédito y Cobranzas. Eva Elizabeth del Valle Córdoba. Universidad Autónoma de México. 2014
- [5] Crédito y Cobranza, José Antonio Morales Castro. 2014.
- [6] Cole, Robert H., *Administración del crédito a las empresas y al consumidor*, México, Diana, 1977.
- [7] Estefania Meza Saldaña, *Evaluación del Riesgo Crediticio, a través de Credit Scoring mediante Regresión logística: Un caso de estudio*, Tesis Benemérita Universidad Autónoma de Puebla. (2017).
- [8] <https://www.consumerfinance.gov/es/obtener-respuestas/que-es-un-puntaje-de-fico-es-1883/> Un sitio web oficial del gobierno federal de los Estados Unidos.
- [9] <https://economipedia.com/definiciones/morosidad-bancaria.html>
- [10] <https://www.condusef.gob.mx/Revista/index.php/usuario-inteligente/educacion-financiera/428-el-prestamo-en-el-mundo>
- [11] <http://imcp.org.mx/servicios/sintesis-informativa/en-mexico-54-millones-de-adultos-tienen-credito-formal/>
- [12] *Historia del Crédito al consumo* Rosa María Gelpí y Francois Jullien-Labruyère. 1998.
- [13] Hernández, M. (2007). *Epidemiología: diseño y análisis de estudios*. B.A.: Editorial Médica Panamericana.
- [14] Juan Picanyol i Tarrés, José Luis Gallizo Larraz. *Diagnóstico financiero en la concesión de crédito a la pequeña y mediana empresa familiar*. Tesis Doctoral, Universidad de Leida. (2013).
- [15] *Medición Integral del riesgo de crédito*. Alan Elizondo. Limusa Noriega Editores.2004.
- [16] Pilar Gómez y Antonio Partal Ureña. (2012). *Gestión de Riesgos financieros en la banca internacional*.
- [17] Reference (2019). *What Are the Functions of Financial Institutions?* Tomado de: reference.com.
- [18] *Riesgo de crédito: el enfoque actuarial*. Javier Gutierrez García, Elizondo flores Jesús Alan. S.f. http://www.actuaires.org/EVENTS/congresses/Cancun/ica2002_subject/credit_risk/credit_71_gutierrez_elizondo.pdf
- [19] Silva, L. C. y Barroso, I. M. (2004). *Regresión Logística*. Madrid: La Muralla.