

Análisis bivariado entre el monto de crédito y probabilidad de elegibilidad

Andrés González Fallas¹, Daniel Núñez Vargas¹

andres.gonzalezfallas@ucr.ac.cr, daniel.nunezvargas@ucr.ac.cr

RESUMEN

La investigación propuesta consiste en la elaboración de un modelo de clasificación para determinar probabilidad de elegibilidad de un individuo para un crédito considerando diversas variables de interés. Hay diversos estudios que siguen una línea similar, sin embargo, esta investigación se diferencia del resto por el hecho de que se va a realizar un análisis de dependencia de las variables y se marca como objetivo la creación de una función de distribución bivariada que contempla la probabilidad de elegibilidad. De esta forma, se distinguen dos principales etapas en el trabajo, la primera es la construcción de esas probabilidades de elegibilidad mediante una regresión logística. Sin embargo, en esta etapa se presenta un inconveniente con la naturaleza de los datos, ya que hay muchas variables y cada variable tiene múltiples categorías. Para solventar esto, se realiza una depuración para combinar categorías que tengan relación. La segunda etapa del trabajo consiste en la elaboración de una distribución bivariada para las probabilidades calculadas en el punto anterior; la otra variable de interés será el monto del crédito. Como se expondrá más adelante, hay muchos tipos diferentes de cópulas, por lo que primero se deberá encontrar la que más se ajuste a los datos por lo que primero se va a hacer un análisis entre diferentes estadísticos para determinar la mejor cópula. Una vez concluidas ambas etapas, se destina una sección para exponer los resultados alcanzados. Por un lado se tiene una regresión logística con buena sensibilidad y una precisión de aproximadamente el 76 %. Y con las distribuciones marginales ajustadas se prueban varias cópulas y se determina que el mejor ajuste se logra mediante una Frank Cópula. Como conclusión se logra obtener una función de distribución bivariada que presenta una correlación negativa, la cual es una Frank Cópula. Con esta función se puede realizar un análisis significativo para el banco y sus pérdidas crediticias.

PALABRAS CLAVE: cópula, elegibilidad crediticia, correlación, distribución bivariada, regresión logística

INTRODUCCIÓN

Día con día personas se presentan a entidades bancarias con el fin de solicitar créditos. A pesar de lo que se pueda pensar, esto es un negocio muy rentable para los bancos, sin embargo, el otorgamiento de los mismos debe realizarse de manera responsable y considerando múltiples criterios. Y dado el contexto actual en donde se está atravesando por una crisis económica importante en todo el mundo, la solicitud de créditos se ha vuelto una forma común de financiamiento para afrontar esta recesión económica. Este puede ser uno de los principales retos a lo que se enfrentan las entidades financieras, determinar a cuáles personas es rentable prestarles dinero.

¹Estudiantes de Ciencias Actariales de la Universidad de Costa Rica

Dicho esto, se distingue una relevancia social bastante más importante más de allá de aplicar algoritmos matemáticos para contestar una pregunta, ya que con estudios de esta índole se puede contribuir a mejorar la capacidad económica de la población.

Este proyecto se encarga de estudiar ese fenómeno con un modelo de clasificación como lo es la regresión logística, sin embargo, para dar una respuesta aún más completa hará falta hacer de otros métodos estadísticos. Esto debido a que este tipo de modelos por lo general no contemplan si existe relación entre las variables estudiadas. Mediante este proyecto se propone realizar un análisis de dependencia adicional para intentar relacionar esa probabilidad de elegibilidad del individuo con otras variables de interés como el monto del crédito. Se usarán datos de un banco alemán que mantiene su nombre en anonimato y hay muchas variables que este banco considera de interés para otorgar un préstamo dentro de las cuales se destaca la edad, el monto del crédito, la cantidad de créditos que tiene, entre otras.

La presente investigación consta de dos principales etapas, la primera es la construcción de modelo para la elegibilidad de los usuarios y la segunda es un análisis de dependencia de la probabilidad de elegibilidad. La primera etapa de investigación se basará en estudios ya existentes en las cuales se utiliza un modelo de clasificación logística para determinar la elegibilidad para un crédito. Por su contraparte, no se encontró literatura relacionada al análisis de dependencia propuesto entre la probabilidad de elegibilidad y variables de interés como el monto del crédito.

Con este proyecto, el principal objetivo es identificar la relación que existe entre la probabilidad de elegibilidad para un crédito con el monto de dicho crédito. Asimismo, se busca construir una función de distribución que explique dicha relación entre las variables y permita realizar cálculos como un Value at Risk. Para la construcción de la distribución bivariada se usarán cópulas, en donde primero se hará un análisis exploratorio para determinar la cópula que mejor explica la relación.

METODOLOGÍA

Antes de entrar de lleno con los métodos estadísticos propuesto para contestar la pregunta, se realiza un análisis de los datos por utilizar. La base fue recuperada de kaggle y originalmente fue obtenida de Penn State Eberly College of Science. Los datos son públicos y son de libre acceso, sin embargo, por motivos de confidencialidad el nombre del banco nunca se menciona. Asimismo, no se indica un contexto temporal.

La población de estudio se define como las personas que solicitaron un crédito en esta entidad bancaria ubicada en Alemania mientras que la unidad estadística se define como la persona solicitante de un crédito en el banco alemán estudiado. La muestra para el desarrollo de dicha investigación consta de 1000 individuos. Asimismo, la base cuenta con 21 variables de interés donde en su columna matriz se encuentra la variable binaria de elegibilidad, que toma el valor de 1 si fue elegible y el valor de 0 si no lo fue.

Además, es conveniente enfatizar que en un modelo de score crediticio es común estudiar muchas características que pueden ser relevantes para determinar la probabilidad de impago de un individuo, sin embargo, a nivel estadístico usualmente no es lo más apropiado. Debido a esto, se realiza un proceso de depuración de la base con el fin de reducir la cantidad de categorías presentes en cada variable categórica. Más adelante se detallan los pormenores de este proceso. A

continuación, una leve explicación de las variables que en primera instancia se pueden considerar relevantes para el estudio:

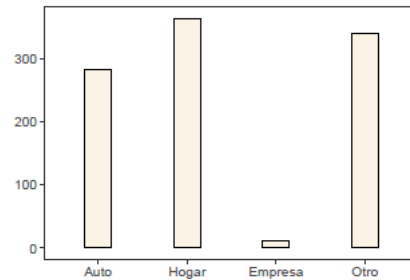
- Elegibilidad: toma el valor de 1 si fue elegible y el valor de 0 si no lo fue.
- Account Balance: variable categórica que toma el valor de 1 si la persona no cuenta con ninguna cuenta en el banco, el valor de 2 si no tiene un balance pendiente con el banco y el valor de 3 si sí tiene un balance pendiente.
- Payment Status of Previous Credit: variable categórica que toma el valor de 1 si el individuo presenta problema con el pago del crédito anterior, el valor de 2 si ya lo pagó y el valor de 3 si no tiene problemas con el crédito anterior
- Purpose: variable categórica que toma el valor de 1 si es para un auto, 2 si es préstamos relacionados a vivienda, 3 si es para un crédito empresarial y 4 si es para cualquier otra cosa.
- Credit Amount: el monto del crédito solicitado en "Deutsche Mark"(DM), que es la unidad monetaria usada en la base.
- Saving/Stock value: toma el valor de 1 si no tiene nada de ahorros o de stock, de 2 si el valor es menor a los 100 DM, 3 si se está en el intervalo [100, 500[DM , 4 si está en [500, 1000[DM y 5 si está arriba de los 1000 DM.
- Length of current employment: variable categórica que toma el valor de 1 si es desempleado, de 2 si tiene menos de año, de 3 si es de 1 a 4 años, de 4 si es de 4 a 7 años y de 5 si es mayor a 7 años.
- Most valuable asset: toma el valor de 1 si no tiene ninguno, el de 2 si es un carro, el de 3 si es un seguro de vida y el de 4 si son bienes raíces.
- Edad: edad en años
- Type of department: toma el valor de 1 si no paga renta/hipoteca, el de 2 en caso de pague renta o hipoteca y el de 3 si es dueño de la vivienda/apartamento.
- Foreign worker: variable binaria que toma el valor de 0 en caso de que sea un trabajador extranjero y 1 en caso de que no lo sea.

Como se pudo observar, la base cuenta con una cantidad considerable de variables de interés, y debido a ese exceso de variables, se tratará de reducirlas haciendo un proceso de depuración de los datos. Para ello, primeramente se reducirán la cantidad de categorías que existen en cada uno de las variables combinando categorías que compartan características o presenten muy pocas observaciones. Por ejemplo, en el caso de la variable del Propósito del Crédito, hay 11 categorías diferentes pero se simplificó de tal manera que solo hubiera 4 categorías. La primera son los préstamos relacionados a la compra de un Automóvil, ya sea nuevo o de segunda mano. La segunda a los préstamos realizados al Hogar, ya sea para compra de muebles o remodelaciones. Una tercera categoría relacionada a créditos Empresariales y una última categoría que incluyera todos los préstamos cuya razón de solicitud no entre en las categorías anteriores.

Una vez hechas estas modificaciones, se puede extraer información interesante como la cantidad de préstamos solicitados por propósito cuyo gráfico se muestra a continuación:

Figura 1

Distribución de la variable propósito del crédito



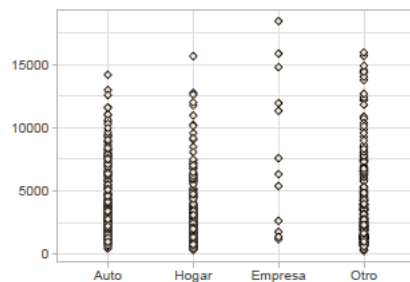
Fuente: Elaboración propia con los resultados obtenidos

Lo que revela que la mayoría de préstamos solicitados son con asuntos relacionados al hogar, mientras que hay una porción muy baja de préstamos destinados al área empresarial.

Asimismo, en el siguiente gráfico se muestra la relación que existe entre el monto del crédito según su propósito, donde cabe destacar que aunque los motivos empresariales es la razón menos frecuente en la base de datos, el crédito solicitado de mayor monto tiene como razón dicho motivo. Mediante este análisis fue que se descartó la posibilidad de combinar el propósito “Empresa” con alguna otra categoría, ya que hay información importante que pueda revelar.

Figura 2

Distribución del monto del crédito según su propósito



Fuente: Elaboración propia con los resultados obtenidos

La mayoría de variables que se encuentran en la base son de carácter categórico, por lo que se presenta el siguiente cuadro que muestra información sobre las variables de carácter numérico continuo:

Tabla 1*Resumen de 5 números de las variables cuantitativas de interés*

Variable	Mín	Q1	Mediana	Q3	Máx
Monto de crédito	250	1 365.5	2 319.5	3 972.25	18 424
Duración de crédito (mensual)	4	12	18	24	72
Edad (años)	19	27	33	42	75

Fuente: Elaboración propia con los resultados obtenidos

Dada la cantidad de variables categóricas con las que cuenta la base, se realizarán pruebas de tal manera que se puedan distinguir las variables de mayor relevancia y, a partir de las mismas, descartar las menos relevantes. Para ello se utilizará el la prueba de independencia Chi-Cuadrado. Se busca que los $p - values$ sean cercanos a cero con tal de afirmar que las variables son estadísticamente significantes en el estudio. Estas pruebas mostraron que las variables más significativas son: Account Balance, Payment Status, Purpose of the credit, Savings/Stock Value, Length of Current Employment, Type Apartment y Most Valuable Asset. A partir de estas variables se desarrolla un modelo de regresión logística en el que se toma un 60 % para entrenamiento y un 40 % para testing.

Con estos términos claros se puede continuar con el intento responder la pregunta del trabajo y el primer paso es el desarrollo de un modelo de clasificación de elegibilidad de crédito, para lo cual se implementó un modelo de regresión logística. Este tipo de regresión como menciona James, a diferencia de los métodos de regresión lineal, sirve para clasificar de manera binaria una variable. Entonces en vez de entrenar un modelo para determinar si hay una correlación entre la variable dependiente y las covariables, se entrena para clasificar en alguna de dos categorías a la variable dependiente de acuerdo a sus covariables. Esta definición del modelo es similar a la que hace Chitarroni en su artículo ya que lo define como un instrumento de análisis multivariado y dependiendo del enfoque se puede utilizar para realizar predicciones o inferencia. Se menciona que es muy útil cuando la variable dependiente es de carácter dicotómico (binario). Asimismo, se aclara que cuando las covariables son categóricas estas deberían de recibir una transformación y convertirlas en variables "dummy", es decir, variables simuladas.

Para el modelo de regresión logística se va a utilizar la siguiente forma:

$$p(\mathbf{X}) = \frac{e^{\beta_0 + X_1\beta_1 + \dots + X_p\beta_p}}{1 + e^{\beta_0 + X_1\beta_1 + \dots + X_p\beta_p}} \quad (1)$$

En donde se cumple $0 < p(X) < 1$ y X_k con $k = 1, \dots, p$ corresponden a las covariables con las que se entrena el modelo. (James, Witten, Hastie, y Tibshirani, 2021). James menciona diagnósticos para modelos de esta índole pero el que más destaca es el de la curva ROC (Receiver Operation Curve), la cual sirve para graficar la tasa de falsos positivos con la sensibilidad del modelo. En la sección de resultados se muestra dicha curva con las respectivas de sensibilidad y especificidad.

Para la segunda parte del trabajo se planea utilizar cópulas; estos modelos sirven para encontrar distribuciones conjuntas que generalmente tienen un alto grado de correlación entre sí, por lo que analizarlas por separado no es lo más recomendable. Como menciona Escarela, las

cóputas bivariadas son funciones que intentan correlacionar dos distribuciones univariadas por lo que estos modelos ayudan a obtener una distribución conjunta a partir de varias funciones de distribución asociadas a variables aleatorias con una cierta relación entre sí. Es decir, con este método se construye una distribución multivariada a partir de las distribuciones univariadas de las variables respectivas. Este tipo de modelo también resulta en una forma de estructurar la dependencia de estas parejas de variables aleatorias en distribuciones conjuntas. Es por esto que son tan populares, puesto que tienen una gran flexibilidad para encontrar distribuciones conjuntas a partir de cualquier pareja aleatoria, lo que es usual tener en muchas disciplinas. (Escarela y Hernández, 2009)

Por lo que es importante definir el concepto de cópula bidimensional, la cual es una función bivariada de un vector aleatorio $\mathbf{V} = (V_1, V_2)$ cuyas marginales V_1 y V_2 son uniformes en el intervalo $\mathbf{I} = (0, 1)$. Entonces, se puede sintetizar el concepto de cópula como una función de la forma: $C : \mathbf{I}^2 \rightarrow \mathbf{I}$. Sin embargo, hay muchos estilos diferentes de cópulas de dónde escoger, hay algunas especiales para modelar correlaciones negativas, otras que se usan para valores extremos, etc. Para efectos de la investigación, la que resulta de interés es la Frank Cópula que se detalla a continuación:

Las Frank Copulas son cópulas arquimedianas, donde cabe distinguir que son de las más usadas para resolver problemas empíricos. Las Frank Copulas son útiles para este problema debido a que pueden expresar la relación entre dependencias tanto positivas como negativas, asimismo, tienen una estructura de dependencia simétrica. (Handini, Maruddani, y Safitri, 2019)

Para entrar más en contexto con este tipo de cópulas, primero se procede a explicar lo que es un cópula arquimediana. Suponiendo una dimensión de d , una cópula se le llama arquimediana cuando es de la forma:

$$C(u_1, u_2, \dots, u_d) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_d)) \quad (2)$$

donde a la función φ se le conoce como función generadora, con el supuesto de que tiene solo un parámetro. Para el caso de las Frank Copulas, esta función generadora viene dada por:

$$\varphi(u) = \ln \left(\frac{e^{-\vartheta u} - 1}{e^{-\vartheta} - 1} \right) \quad (3)$$

con $\vartheta > 0$. Como se está trabajando en el caso bivariado, la función C se escribiría de la siguiente forma:

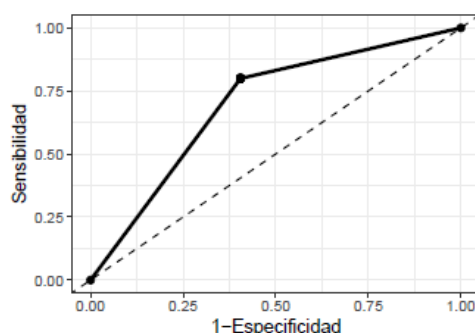
$$C(u_1, u_2) = -\frac{1}{\vartheta} \ln \left(1 + \frac{(e^{-\vartheta u_1} - 1)(e^{-\vartheta u_2} - 1)}{e^{-\vartheta} - 1} \right) \quad (4)$$

RESULTADOS

Para determinar si al solicitante se le dará el crédito, el umbral será del 0.5, por lo que si el resultado de la regresión es mayor a 0.5, se le da el crédito y en caso contrario no. El resultado de la regresión arroja una curva ROC como se muestra a continuación:

Figura 3

Curva ROC de la regresión logística



Fuente: Elaboración propia con los resultados obtenidos

Asimismo, este modelo arroja una precisión aproximada del 76 % con un intervalo de confianza de $]0,71, 0,80[$. Además se tiene una sensibilidad de 82 %, lo que indica que el modelo es bueno para clasificar a buenos deudores como buenos. Por otro lado, se tiene una especificidad de apenas el 61 % por lo que se concluye que el modelo no es óptimo para clasificar a malos deudores como malos.

El fijar el umbral para determinar la curva ROC se hizo meramente para mostrar que la regresión logística resulta ser buena, por lo que realmente este es un resultado secundario de la investigación. Cabe recordar que el propósito del trabajo es comparar la distribución marginal de las probabilidades de elegibilidad con variables como el monto del crédito, con el fin de calcular una distribución conjunta usando cópulas.

Dicho esto, si se calcula la correlación empírica entre las probabilidades de elegibilidad con el monto del crédito se llega a una correlación de -0.30. Esta correlación se calculó con el método del tau de Kendall ya que es el método más popular para comparar dependencia entre modelos de cópulas. Este resultado indica que, si el monto a solicitar por parte del cliente es muy alto, disminuyen las probabilidades de ser elegido. Lo mismo pasa de manera análoga cuando el monto del crédito es bajo, en donde las probabilidades de ser elegido aumentan.

Lo anterior expuesto se puede identificar como el primer hallazgo y resulta ser muy útil ya que esta asociación negativa funciona como una especie de filtro a la hora de escoger la cópula que mejor ajusta los datos. Esto sucede porque no todas las cópulas pueden modelar relaciones negativas.

Para poder ajustar un modelo de cópulas bivariadas, primero se tienen que saber las funciones de distribución marginales univariadas para cada variable de estudio. En este se necesitan encontrar dos de estas marginales pues el modelo de cópulas a implementar va a considerar como dos variables, el monto del crédito solicitado y el valor de la regresión logística asociado a la probabilidad de elegibilidad asignado con el modelo.

Con los histogramas en la figura 6 se pueden observar ciertas tendencias de como se distribuyen las variables, lo que permite buscar ajustar una distribución paramétrica de acuerdo a su forma. Bajo este hilo, se sigue que el monto del crédito es bastante asimétrica hacia la derecha,

mientras que la probabilidad de elegibilidad sigue una tendencia más simétrica.

Para el siguiente desarrollo se va a utilizar el lenguaje de programación R para realizar los modelos, cálculos y estimaciones. Primeramente, se va van a realizar distintos modelos para las distribuciones marginales de las variables de estudio donde por métodos como AIC y estimación de parámetros por máxima verosimilitud. Bajo esta metodología, se consigue que la mejor distribución que se ajusta para el monto del crédito es una distribución $\text{Gamma}(1.8115402, 1849)$ bajo una parametrización de forma y escala. Mientras que, al realizar este ajuste con la probabilidad de elegibilidad el modelo que mejor ajusta es una distribución $N(0.5625845, 0.1549248)$.

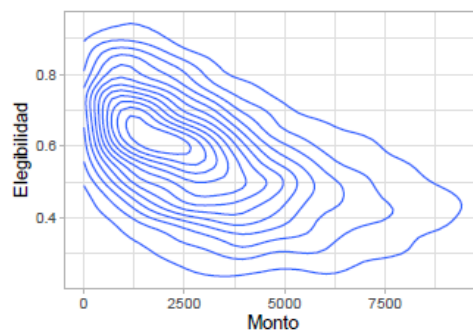
Una vez, con estas distribuciones marginales se procede a modelar la correlación entre estas variables mediante cópulas para lograr estimar una función de distribución bivariada considerando las variables de estudio.

Como se mencionó anteriormente los datos presentan una correlación negativa y más específicamente se cuenta con un tau de Kendall de -0.3. Con esta información, se procede a escoger una función de cópulas que más se ajuste a los datos observados para poder construir su función de distribución. Bajo métodos de escogencia de una función de cópulas como el AIC y la estimación de parámetros por medio de máxima verosimilitud se llega a un modelo con el mejor ajuste que sería una cópula de Frank con parámetro $\theta = -3,6$ y $\tau_K = -0,36$, lo cual es bastante cercano a la correlación empírica calculada anteriormente. Es decir, se obtiene una función de cópulas (Arquimediana) que mantiene la correlación de los datos, lo cual también es lo buscado.

Una vez con este modelo, y las marginales univariadas calculadas al principio, se procede a construir la función de distribución y densidad bivariada. Que es lo que se planea contestar con la pregunta de investigación planteada. Para su visualización se utilizan técnicas de graficación en tres dimensiones en dos dimensiones por lo que el siguiente gráfico muestra el contorno de la densidad bivariada.

Figura 4

Contorno de la función de densidad bivariada



Fuente: Elaboración propia con los resultados obtenidos

Como se mencionó anteriormente, la escogencia de la cópula fue mediante AIC sin embargo para poder determinar que tan bien se ajusta una cópula, se utiliza el método de *Goodness-of-fit test* o *Gof Test*. El cual se basa en calcular el estadístico de Cramér-von-Mises o S_n . Por lo que, en la siguiente tabla se muestran los resultados de las pruebas realizadas.

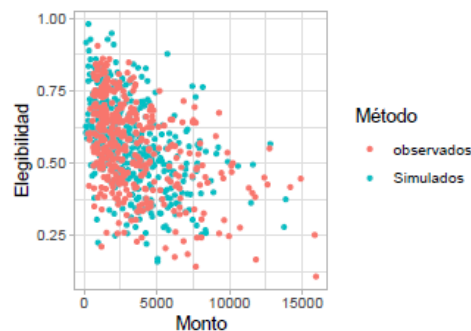
Tabla 2*Diagnósticos de las distintas cópulas*

Cópula	AIC	Estadístico Cramér-von-Mises
Frank	-63.4602	0.0267
Normal	-55.0481	0.0270
t	-53.0477	0.0426
Clayton	-10.9724	0.6173

Fuente: Elaboración propia con los resultados obtenidos

Como se puede observar, tanto el AIC como el S_n dan más bajos en el caso de una cópula de la familia Frank. Por lo que debido a estas pruebas se llega a la conclusión que el mejor modelo de cópulas para las variables de estudio es una cópula de Frank. También, vale la pena recalcar que las peores métricas se obtienen considerando un modelo de cópula de Clayton, lo cual corrobora el hecho que si se presenta una correlación negativa este modelo es peor y en algunos algoritmos ni se considera intentar ajustar con este modelo.

Finalmente, también se realizó una serie de datos simulados con la función de distribución bivariada contra los datos reales para de manera visual poder analizar tendencias y comportamientos entre el modelo y la realidad.

Figura 5*Datos simulados con la función de distribución ajustada vs. datos observados**Fuente: Elaboración propia con los resultados obtenidos*

Asimismo, la función de distribución bivariada ajustada se utilizó para determinar valores de significancia como lo son las medidas de riesgo asociadas a los datos. En este caso se encontró que el VaR al 95 % del monto de crédito se obtiene en el monto 8 233 mientras que el CVaR a ese mismo nivel de significancia es de 10 265,53. Similarmente, para la probabilidad de elegibilidad se encontró que el 95 % de la masa de su distribución se encuentra por debajo del 87 %. Este tipo de resultados resultan útiles a la hora de realizar análisis referentes a las pérdidas esperadas por el banco por impago crediticio.

CONCLUSIONES

La pregunta de investigación que se planteó en un principio era encontrar una distribución bivariada para el monto de crédito y la probabilidad de elegibilidad, sin embargo, antes de encontrar esta distribución fue necesario realizar una serie de procedimientos previos. Se optó por esta pregunta debido al peso tan importante que tienen los créditos en la actualidad, por lo que se consideró que valía la pena estudiar la relación entre la elegibilidad de un individuo.

El primero fue determinar la probabilidad de elegibilidad la cual se calculó mediante una regresión logística que tuvo una precisión aproximada del 76 %. El siguiente paso era determinar las funciones marginales de cada una de las dos variables estudiadas en donde para el caso del monto del crédito se llegó a una Gamma (1.8115, 1849). Por otro lado, para la probabilidad de elegibilidad se eligió una Normal(0.5625, 0.1549).

Una vez con las distribuciones marginales encontradas se sigue con el cálculo de la distribución bivariada para la cual se utilizan cópulas. Una de las ventajas que ofrecen las cópulas es que solo se necesitan las marginales para encontrar esa distribución bivariada. Dicho esto, a nivel metodológico fue bastante acertado usar este método pues ambas marginales ya se habían encontrado. Después de un exhaustivo proceso de selección de cópulas, se llegó que la más adecuada era la Frank.

Mediante esta cópula se llega a una correlación negativa entre la variable monto de crédito y la elegibilidad, lo que era lo esperado, ya que entre mayor sea el monto mayor serán los requerimientos para aplicar por el crédito y esto resulta en una probabilidad de elegibilidad menor.

Entre las limitaciones presentadas se distingue el uso de solo dos variables para crear la función de distribución bivariada ya que el uso de cópulas para más variables tiende a complicarse considerablemente, aunque podrían esperarse resultados mejores si se consideran más variables.

Otra limitación es que al usar una regresión logística para calcular las probabilidades de elegibilidad, por cada iteración de código que se corría, las probabilidades tendían a variar. Como consecuencia de ello, la distribución marginal de la variable podría variar. Para solventar esta limitante se optó por fijar las variables de elegibilidad y trabajar con ese set a lo largo de todo el código.

Entre las recomendaciones se destaca intentar mejorar los resultados de la regresión logística ya que para efectos de una entidad bancaria, una precisión del 76 % puede llegar a considerarse baja. Otra recomendación sería modelar la relación de la elegibilidad con otra variable de interés como la edad del solicitante por ejemplo.

El uso de cópulas es común cuando se quieren modelar problemas de distribuciones bivariadas y de hecho se pueden encontrar trabajos que tengan que ver con esto, sin embargo, no se ofrece un análisis que relacione la probabilidad de incumplimiento con los montos del crédito.

AGRADECIMIENTOS

Se agradece al profesor Maikol Solís Chacón por su guía durante el desarrollo de la investigación así como sus consejos a la hora de realizar la implementación de los modelos propuestos. También, se agradece al estudiante Eliécer Luna Jiménez puesto que sus indicaciones en las bitácoras fueron esclarecedoras para el éxito de esta investigación.

BIBLIOGRAFÍA

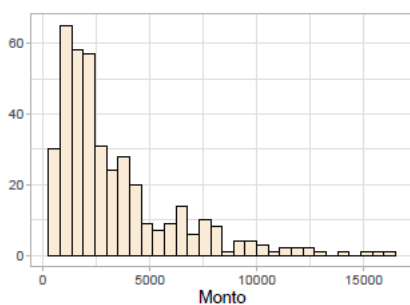
- Chitarroni, H. (2002). La regresión logística. *IDICSO*.
- Escarela, G., y Hernández, A. (2009). Modelado de parejas aleatorias usando cópulas. *Revista Colombiana de Estadística*, 32(1), 33–58.
- Funding Circle. (s.f.). Pérdida esperada (pe).
- Genest, C., Quessy, J.-F., y Remillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2), 337–366.
- Handini, J., Maruddani, D., y Safitri, D. (2019). Frank copula on value at risk (var) of the construction of bivariate portfolio (case study: stocks of companies awarded with the idx top ten blue with stock period of 20 october 2014 to 28 february 2018). En *Journal of physics: Conference series* (Vol. 1217, p. 012078).
- James, G., Witten, D., Hastie, T., y Tibshirani, R. (2021). *An introduction to statistical learning: with applications in r*. Springer.
- Klugman, S. A., Panjer, H. H., y Willmot, G. E. (2013). *Loss models: Further topics*. John Wiley & Sons.
- Montes de Oca, J. (2015). Crédito. *Economipedia*.
- Real Academia Española. (2014). *Elegibilidad* (23. ed.).

ANEXOS

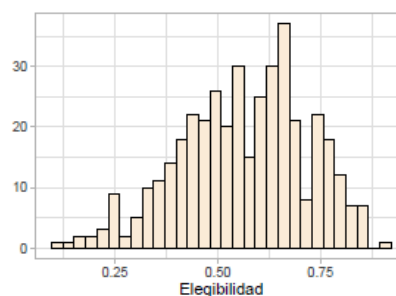
Figura 6

Histogramas de las variables de interés

(a) Histograma de los montos de crédito de la base de prueba



(b) Histograma de las probabilidades de elegibilidad



Fuente: Elaboración propia con los resultados obtenidos