



**UNIVERSIDAD NACIONAL DEL ALTIPLANO DE PUNO**  
**FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA**  
**ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E**  
**INFORMÁTICA**



**MODELO PREDICTIVO DE RIESGO DE MOROSIDAD PARA**  
**CRÉDITO BANCARIO A PARTIR DE DATOS SIMULADOS DE LA**  
**CAJA RURAL DE AHORRO Y CREDITO**

**LOS ANDES – PUNO, 2019**

**TESIS**

**PRESENTADA POR:**

**Bach. GROBER MAMANI CHOQUE**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**

**INGENIERO ESTADÍSTICO E INFORMÁTICO**

**PUNO – PERÚ**

**2020**



## DEDICATORIAS

### *A Dios*

*Por guiar mi camino durante toda mi vida y por su bendición para poder prosperar profesionalmente, por darme fortaleza para seguir adelante a pesar de las adversidades, por los momentos de felicidad compartidos con mi familia.*

### *A mis padres*

*Matías Sergio y Cirila Trifona por enseñarme que con perseverancia y disciplina puedo lograr mis metas, por ponerme retos que hicieron que me desarrolle como persona, por el apoyo incondicional en mi vida, por motivarme siempre a ser una persona mucho mejor, a pesar de los obstáculos en mi vida siempre están presente apoyándome, por el apoyo económico y moral.*

### *A mis hermanos*

*Por compartir experiencias vividas que me permitieron crecer como persona, por enseñarme que si hay un objetivo definido; no existen imposibles; por darme sus consejos y apoyos en el transcurso de mi formación profesional.*



## AGRADECIMIENTOS

*Agradezco a Dios, por estar conmigo en cada paso q doy en la vida, dándome fuerzas y ánimos para seguir adelante, por brindarme una vida llena de aprendizajes, experiencias; por permitir que todo esto fuese posible.*

*Agradezco a mi familia: apoyo fundamental de mi formación profesional y personal; quienes a la distancia supieron siempre transmitirme su confianza y aliento para seguir adelante manteniendo mi equilibrio como persona, a mis hermanos; por compartir vivencias increíbles.*

*Agradezco a la Universidad Nacional del Altiplano Puno, Escuela profesional de Ingeniería Estadística e Informática; por acogerme, formarme y permitir que culmine mis estudios.*

*A mis docentes; por brindarme su gama de experiencia profesional, quienes, con sus conocimientos, experiencia y su motivación han logrado que termine mi formación profesional exitosamente.*

*A mis jurados; quienes tuvieron la paciencia, visión crítica y firme, de dirigirme durante este proceso.*

*Finalmente, Agradezco a todos mis amigos(as), compañeros(ras) y personas que tuve la oportunidad de conocer, quienes ayudaron directa e indirectamente en mi crecimiento personal y profesional.*

Grober,  
Mamani Choque



# ÍNDICE GENERAL

Pág.

**DEDICATORIAS**

**AGRADECIMIENTOS**

**ÍNDICE GENERAL**

**ÍNDICE DE FIGURAS**

**ÍNDICE DE TABLAS**

**ÍNDICE DE ACRÓNIMOS**

**RESUMEN ..... 10**

**ABSTRACT..... 11**

## **CAPÍTULO I**

### **INTRODUCCIÓN**

**1.1. DESCRIPCIÓN DEL PROBLEMA ..... 13**

**1.2. FORMULACIÓN DEL PROBLEMA ..... 14**

**1.3. OBJETIVOS ..... 15**

**1.4. HIPÓTESIS DE LA INVESTIGACIÓN ..... 15**

**1.5. JUSTIFICACIÓN ..... 15**

## **CAPÍTULO II**

### **REVISIÓN DE LITERATURA**

**2.1. ANTECEDENTES ..... 18**

**2.2. MARCO TEÓRICO ..... 23**

2.2.1. Regresión..... 23

2.2.2. Distribución Binomial ..... 24

2.2.3. Regresión Logística ..... 25

2.2.4. Simulación..... 35

2.2.5. Factores de riesgo ..... 38



<b>2.3. DESCRIPCIÓN DE VARIABLES .....</b>	<b>40</b>
<b>CAPÍTULO III</b>	
<b>MATERIALES Y MÉTODOS</b>	
<b>3.1. UBICACIÓN GEOGRÁFICA DEL ESTUDIO.....</b>	<b>43</b>
<b>3.2. ÁMBITO DE ESTUDIO .....</b>	<b>43</b>
<b>3.3. TIPO Y DISEÑO DE INVESTIGACIÓN .....</b>	<b>43</b>
<b>3.4. POBLACIÓN Y MUESTRA.....</b>	<b>44</b>
3.4.1. Población .....	44
3.4.2. Muestra.....	44
<b>3.5. INSTRUMENTO DE RECOLECCIÓN Y PROCESAMIENTO DE DATOS</b> <b>.....</b>	<b>45</b>
3.5.1. Procedimiento de recolección de datos .....	45
3.5.2. Plan de procesamiento y análisis de datos.....	45
<b>CAPÍTULO IV</b>	
<b>RESULTADOS Y DISCUSIÓN</b>	
<b>4.1. MODELO DE RIESGO DE MOROSIDAD.....</b>	<b>47</b>
4.1.1. Tasa del riesgo de morosidad .....	48
4.1.2. Estimación de parámetros de los factores de capacidad de pago del cliente	
.....	48
<b>4.2. GENERACIÓN DE BASE DE DATOS MEDIANTE SIMULACIÓN .....</b>	<b>53</b>
<b>4.3. PREDICCIÓN DEL RIESGO DE MOROSIDAD A PARTIR DE LA BASE</b> <b>DE DATOS SIMULADA.....</b>	<b>58</b>
4.3.1. Selección de variables .....	58
4.3.2. Verosimilitud del modelo .....	65
4.3.3. Prueba de bondad de ajuste .....	66
4.3.4. Predicción del riesgo de morosidad.....	67
<b>V. CONCLUSIONES.....</b>	<b>69</b>
<b>VI. RECOMENDACIONES .....</b>	<b>71</b>



<b>VII. REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>72</b>
<b>ANEXOS.....</b>	<b>78</b>

**Área :** Estadística

**Tema :** Regresión Logística y Simulación

**FECHA DE SUSTENTACIÓN:** 09 de diciembre de 2020.



## ÍNDICE DE FIGURAS

	<b>Pág.</b>
<b>Figura 1.</b> Gráfica de la curva COR .....	65



## ÍNDICE DE TABLAS

	<b>Pág.</b>
<b>Tabla 1.</b> Calificación de días de morosidad .....	41
<b>Tabla 2.</b> Operacionalización de variables. ....	42
<b>Tabla 3.</b> Calidad de portafolio de créditos (%) de la Caja Los Andes - Puno .....	48
<b>Tabla 4.</b> Destino de crédito solicitado por los clientes de la Caja Los Andes - Puno, tercer trimestre 2019. ....	49
<b>Tabla 5.</b> Tipo de cliente que solicita el crédito de la Caja Los Andes - Puno, tercer trimestre 2019. ....	50
<b>Tabla 6.</b> Días de atraso (número de impagos) en el pago del crédito de la Caja Los Andes - Puno, tercer trimestre 2019. ....	50
<b>Tabla 7.</b> Estadísticos descriptivos del Ingreso total del cliente de la Caja Los Andes - Puno, tercer trimestre 2019. ....	51
<b>Tabla 8.</b> Tipo de empleo del cliente de la Caja Los Andes - Puno, tercer trimestre 2019. ....	52
<b>Tabla 9.</b> Sector de procedencia del cliente de la Caja Los Andes - Puno, tercer trimestre 2019. ....	53
<b>Tabla 10.</b> Estado civil del cliente de la Caja Los Andes - Puno, tercer trimestre 2019. ....	53
<b>Tabla 11.</b> Criterios de las variables a ser simuladas .....	56
<b>Tabla 12.</b> Entrenamiento y selección de tamaño de muestra. ....	57
<b>Tabla 13.</b> Variables que ingresan al modelo .....	59
<b>Tabla 14.</b> Variables que ingresan al nuevo modelo .....	62
<b>Tabla 15.</b> Prueba del coeficiente de determinación .....	63
<b>Tabla 16.</b> Tabla de clasificación .....	63
<b>Tabla 17.</b> Prueba del coeficiente de intercepción .....	66
<b>Tabla 18.</b> Prueba de bondad de ajuste.....	67
<b>Tabla 19.</b> Prueba de Ómnibus.....	67
<b>Tabla 20.</b> Predicción de riesgo de morosidad para dos casos. ....	68





## ÍNDICE DE ACRÓNIMOS

<b>SPSS</b>	: Producto de Estadística y Solución de Servicio
<b>VIH</b>	: Virus de Inmunodeficiencia Humana
<b>TIC</b>	: Tecnologías de la Información y Comunicación
<b>IMF</b>	: Instituciones Micro Financieras
<b>CMAC</b>	: Caja Municipal de Ahorro y Crédito
<b>BCP</b>	: Banco de Crédito del Perú
<b>RV</b>	: Razón de Verosimilitud
<b>FR</b>	: Factores de Riesgo



## RESUMEN

La morosidad se ha constituido en la principal causa de las dificultades que han sufrido algunos sistemas financieros y ciertas entidades de tamaño considerable. Así, una elevada cartera morosa es un serio problema que compromete tanto la viabilidad de la institución a largo plazo como la del propio sistema. En respuesta a lo mencionado, se estudió el riesgo de morosidad para crédito en los clientes de la Caja de Ahorro y Crédito “Los Andes” agencia Puno en el periodo 2019, haciendo uso de técnicas estadísticas como son los modelos de regresión logística y la simulación. El objetivo de esta investigación es predecir el riesgo de morosidad para crédito bancario, para lograr el objetivo se estimó un modelo de regresión logístico para la predicción de la probabilidad de no recuperar un crédito a partir de la generación de una Base de Datos mediante simulación con un proceso adecuado de entrenamiento y prueba del modelo usando para ello el programa SPSS 24.0. Los resultados fueron: El modelo de regresión logístico en general es significativo como modelo predictor, y este modelo planteado logra predecir una tasa de clasificación de verdaderos negativos (no morosos pronosticados como no morosos) que fue del 97.2%; mientras que la tasa más relevante en este caso, clasificar correctamente a los clientes morosos (verdaderos positivos) fue de 45%, demostrando gran capacidad predictiva del modelo, el modelo fue contrastado con los datos simulados, la metodología aquí presentada se puede ver como referencia para un trabajo futuro más realista, en tal caso, se debe tener cuidado en validar primero la base de datos mediante el estudio de las distribuciones adecuadas, así como validación final del modelo, o mejor aún, usar una base de datos reales.

**Palabras clave:** Morosidad, riesgo de crédito, regresión, modelo logístico.



## ABSTRACT

Late payment has become the main cause of the difficulties suffered by some financial systems and certain large entities. Thus, a high non-performing portfolio is a serious problem that compromises both the long-term viability of the institution and that of the system itself. In response to the aforementioned, the risk of delinquency for credit was studied in the clients of the Caja de Ahorro y Crédito "Los Andes" Puno agency in the period 2019, making use of statistical techniques such as logistic regression models and simulation. The objective of this research is to predict the risk of delinquency for bank credit, to achieve the objective, a logistic regression model was estimated to predict the probability of not recovering a loan from the generation of a Database by simulation with an adequate process of training and testing the model using the SPSS 24.0 program. The results were: The logistic regression model in general is significant as a predictor model, and this proposed model succeeds in predicting a rate of classification of true negatives (non-delinquent predicted as non-delinquent) that was 97.2%; While the most relevant rate in this case, correctly classifying delinquent clients (true positives) was 45%, demonstrating the model's great predictive capacity, the model was contrasted with the simulated data, the methodology presented here can be seen as a reference for a more realistic future work, in this case, care must be taken to first validate the database by studying the distributions. adequate, as well as final validation of the model, or better yet, use a real database.

**Keywords:** Late payment, credit risk, regression, logistic model.



# CAPÍTULO I

## INTRODUCCIÓN

Actualmente las Cajas Municipales están automatizando sus procesos operacionales en forma progresiva, existe aún una distancia considerable entre el proceso de automatización y la adecuación de los mecanismos de control, los cuales permitan minimizar los riesgos que están asociados a la sistematización de sus procesos. El sistema de Cajas Municipales de ahorro y crédito en el Perú en estos últimos años han presentado un crecimiento muy considerable, con tasas de crecimiento altas y aumentando su participación en el sistema financiero peruano; así mismo, las tasas de morosidad en estas instituciones también están en aumento, alcanzando niveles considerables en el sistema bancario, pues estos siempre registraron mayores tasas comparados a los bancos comerciales (Cermeño, León, & Mantilla, 2011).

Un problema grave en las instituciones de crédito, así como en la Caja de Ahorro y Crédito “Los Andes” agencia Puno son los préstamos no recuperados, por lo que, al momento de la solicitud, es deseable un modelo matemático para predecir el riesgo de tener un cliente potencialmente moroso, ya que su capacidad económica no garantiza por si sola el pago. Este trabajo de investigación muestra, a través de datos simulados, una técnica estadística conocida como regresión logística para predecir la probabilidad de no recuperar un crédito, con base al perfil socioeconómico del cliente, considerando el número de dependientes económicos, número de impagos previos, su empleo, destino del crédito y salario mensual. El modelo seleccionado predice la tasa aciertos de clientes morosos y una tasa global de clasificaciones correctas.

La regresión logística ha sido utilizada durante las últimas décadas como modelo de predicción y clasificación desde sus orígenes en investigaciones epidemiológicas y



clínicas (Hosmer, Lemeshow, & Sturdivant, 2013). La utilización de esta técnica en el presente trabajo, se justifica por que hoy en día se tiene múltiples aplicaciones de forma no limitada en investigaciones biomédicas, negocios y finanzas, criminología, ingeniería, salud pública, política, biología de la vida salvaje y psicología. Las aplicaciones van desde los análisis de riesgo crediticio, votación en elecciones políticas, hasta predicción de cáncer de próstata, riesgo de contagio de VIH, etcétera (Mures Quintana, García Gallego, & Vallejo Pascual, 2005).

La estructura del presente trabajo de investigación se dispone de la siguiente manera: en la siguiente sección, revisión general de la técnica estadística conocida como regresión logística, seguida de la generación de la base de datos mediante simulación y la estimación de los parámetros de las variables participantes en el modelo. En la segunda sección, realizamos una revisión de la literatura que fortalecerán el trabajo, la tercera describimos los materiales y métodos, en la cuarta parte, están los resultados donde realizamos la estimación de los parámetros del modelo logístico y la validación del modelo predictor. Finalmente, conclusiones, referencias bibliográficas y apéndices.

## **1.1. DESCRIPCIÓN DEL PROBLEMA**

El avance de las tecnologías de la información y comunicación (TIC) son en la actualidad uno de los más importantes temas de investigación en el campo de la Economía y Dirección de Empresas (García & Rialp, n.d.; Martín, 2016). Las instituciones micro financieras (IMF) son muy importantes en nuestra región de Puno, en la cual las Cajas Municipales, por su función prestadora de créditos al grupo de sectores que no cuentan con fácil acceso a la banca tradicional, sin embargo, la morosidad es un problema grave que enfrentan y constituyen una de las principales causas de la insolvencia y descapitalización.



Los clientes que solicitan crédito a las instituciones financieras cuentan con una calificación o récord crediticio, en sentido existen dos objetivos importantes en las técnicas para la calificación crediticia. Uno de ellos es la necesidad de identificar el riesgo de consumo, mientras que el otro es minimizar el porcentaje de clientes incumplidos, con lo cual las instituciones bancarias o crediticias optimizan sus carteras para un buen y mejor negocio, en ese sentido, el riesgo financiero en las cajas municipales es uno de los problemas científicos de gran importancia y preocupación de los directivos (Barzaga, Nevárez, Hidalgo Ávila, & Loor, 2018; Trejo, Martínez, & Venegas, 2017).

La morosidad se ha constituido en la principal causa de las dificultades que han sufrido algunos sistemas financieros y ciertas entidades de tamaño considerable. Así, una elevada cartera morosa es un serio problema que compromete tanto la viabilidad de la institución a largo plazo como la del propio sistema.

La Caja de Ahorro y Crédito Los Andes de Puno muestra una mejora de su cartera de créditos. La morosidad institucional alcanzó el 5.3% al finalizar el año 2016, sin embargo, hay probabilidad de que el índice de morosidad incremente en el próximo periodo. A la fecha no existe un estudio de impacto del riesgo crediticio en el incremento de la morosidad, de continuar esta situación podría afectar la estabilidad como agencia Puno, deteriorar su imagen financiera, no permitir la mejora de su calificación de riesgos y afectar seriamente el alcance de sus objetivos institucionales.

## **1.2. FORMULACIÓN DEL PROBLEMA**

¿Es posible generar un modelo predictivo que permite realizar la predicción del riesgo de morosidad para el crédito bancario a partir de datos simulados en



función de los factores identificados en los clientes de la Caja Rural de Ahorro y Crédito “Los Andes” de la ciudad de Puno?

### **1.3. OBJETIVOS**

#### **Objetivo general**

Predecir el riesgo de morosidad para crédito bancario a partir de datos simulados de la Caja Rural de ahorro y crédito “Los Andes”, 2019.

#### **Objetivos específicos**

- Estimar un modelo de regresión logístico para la predicción de la probabilidad de no recuperar un crédito de la Caja Rural de ahorro y crédito “Los Andes”, 2019.
- Generar una Base de Datos mediante simulación para entrenamiento y prueba del modelo.

### **1.4. HIPÓTESIS DE LA INVESTIGACIÓN**

El modelo de regresión logística para pronosticar la morosidad de un crédito bancario se desempeña adecuadamente como modelo clasificador sobre una base de datos simulada de la Caja Rural de ahorro y crédito “Los Andes”, 2019.

### **1.5. JUSTIFICACIÓN**

Durante los últimos años, el crecimiento económico sostenido que ha experimentado la economía peruana ha contribuido a expandir fuertemente las colocaciones de las instituciones bancarias. Ante un escenario de expansiones significativas en el nivel de colocaciones, resulta cada vez más difícil que las entidades bancarias obtengan información acerca de la capacidad de pago de los nuevos prestatarios. Por tanto, cuando los ciclos económicos son muy favorables y el ritmo de



colocaciones se acelera, el riesgo del portafolio crediticio tiende a aumentar ya que los créditos otorgados a deudores con dudosa capacidad de pago se incrementan bajo estos escenarios. En ese sentido, El riesgo de crédito es el tipo de riesgo más importante al que debe hacer frente cualquier entidad financiera. Un indicador de este riesgo es el nivel de morosidad de la entidad; es decir, la proporción de su cartera que se encuentra en calidad de incumplimiento.

En vista de la importancia de este indicador, es necesario identificar los determinantes de la tasa morosidad de las colocaciones de los bancos. Esta información permitiría que el regulador financiero implemente medidas de política para mantener o mejorar la calidad de las carteras de colocaciones. De esta manera, pueden anticipar y minimizar los efectos de las evoluciones desfavorables de la economía o de las políticas de gestión de cada una de las instituciones supervisadas en la tasa de morosidad que enfrentan.

Por tanto, recurrimos al uso de técnicas estadísticas para estudiar estos fenómenos económicos y sociales a través de la regresión logística como modelo de predicción; esta es una técnica de estadística multivariada basada en los principios de regresión lineal, pero a diferencia de ésta, la regresión logística utiliza una variable dependiente categórica binaria o dicotómica (en vez de una variable cuantitativa continua) cuyo resultado o valor de respuesta es 0 o 1. Debido a que la morosidad es una variable dicotómica que depende como variable de respuesta de otros indicadores tales como el salario y los gastos familiares, entre otros, es posible aplicar la regresión logística como modelo clasificador. Así mismo, el modelo puede utilizarse para estimar la probabilidad de que un cliente no devuelva el crédito; a dicha probabilidad se le conoce como riesgo de morosidad. La condición de morosidad se categorizará dependiendo si la probabilidad estimada,





evaluada con los datos del cliente, es mayor que un punto de corte, el cual debe prefijarse por cada institución particular con base en su experiencia.

En respuesta a lo antes mencionado, consideramos indispensable estudiar el riesgo de morosidad para crediticio en los clientes de la Caja de Ahorro y Crédito los Andes agencia Puno en el periodo 2019, haciendo uso de técnicas estadísticas como son los modelos logísticos y la simulación, de tal forma que permita a la institución financiera cumplir sus objetivos y la sostenibilidad en el negocio.



## CAPÍTULO II

### REVISIÓN DE LITERATURA

#### 2.1. ANTECEDENTES

Golman & Bekerman (2018), plantean como objetivo distinguir las variables determinantes sobre el mayor o menor grado de morosidad en el pago de los prestatarios micro-financieros en un estudio de caso correspondiente a la Asociación Civil Avanzar. Es a partir del análisis de la base de datos de dicha institución social Argentina, que se analizan las características más relevantes de los prestatarios, incluyendo las personales, del hogar y del mismo crédito otorgado, para explicar el comportamiento frente al pago de sus cuotas. Obtuvieron interesantes conclusiones que pueden permitir avanzar sobre los problemas de información existentes al momento de otorgar créditos para instituciones con similares características a la considerada en este estudio.

Meza (2017), en su trabajo desarrolló un modelo estadístico integrado para evaluar un préstamo otorgado por una entidad financiera, mediante el análisis de la información que se tiene de cada uno de los clientes, a través de un Modelo de Regresión Logística, para obtener las características más significativas y poder establecer una regla de aceptación. Concluye que, haciendo énfasis en que es igual de importante evaluar continuamente el modelo de credit scoring con el fin de revalidar su correcto ajuste con los valores reales, en conjunto con la contribución del conocimiento del experto para considerar todos los aspectos. El modelo de credit scoring depende únicamente de los datos con los que cuenta la entidad en cuestión, las variables que se incluyen en el modelo son propias para la institución por lo que



no serán las mismas por completo si se aplica a otra institución. Cuanto esté dispuesta a correr riesgos la institución dependerá de los objetivos de la misma, por lo que es fundamental considerarlo para aceptar o rechazar a un cliente dependiendo de su probabilidad de incumplimiento.

Leal, Aranguiz, & Gallegos (2017), se propusieron a fantasía un modelo de evaluación crediticia a sus clientes actuales y potenciales ajustado y ponderado a su realidad, que permite disminuir el riesgo de crédito o incobrables. El presente trabajo considera, una descripción de los modelos de evaluación de créditos y en específico de los modelos de credit scoring. A través de entrevistas a expertos, se definieron variables cuantitativas y cualitativas críticas a considerar en un proceso de gestión de créditos. Respecto de la calidad del modelo de evaluación crediticia propuesto, esta muestra que un 81,82% de los créditos otorgados a sus clientes han superado el nivel mínimo de evaluación o límite de aprobación por la empresa.

Matos Falvy (2017), se planteó como objetivo principal demostrar que el valor de la pérdida esperada de las Cajas Municipales de Ahorro y Crédito se reduce como resultado de una menor tasa de morosidad y mejora en la calidad crediticia, todo ello gracias al establecimiento de un punto de corte o cut off que optimiza la aceptación y rechazo de prestatarios, determinado por la curva ROC y regresiones logísticas, el mismo que permitió reducir la pérdida esperada medida por el Value at Risk. Estudiado para la población de prestatarios que han financiado su préstamo hipotecario en las Cajas Municipales de Ahorro y Crédito en el año 2016.

Pally (2016), tuvo como objetivos: explicar las políticas y reglamentos de créditos respecto al nivel de morosidad y determinar el nivel de cumplimiento de las funciones del sectorista de créditos y su influencia en la gestión del riesgo crediticio



de la institución. Uso el método analítico, deductivo y descriptivo haciéndose uso de la técnica e instrumento de encuesta a través de la aplicación del cuestionario, observación directa y revisión documentaria, los cuales nos permitieron obtener datos de manera confiable y directa. Concluyó que hubo una disminución de la cartera de créditos y colocaciones e incremento del nivel de morosidad en la caja municipal de ahorro y crédito de Piura agencia Juliaca en los periodos 2013 – 2014. En consecuencia, con su estudio se puede afirmar que existe una relación directa entre la gestión del riesgo crediticio y su incidencia en el nivel de morosidad de la caja municipal de ahorro y crédito de Piura agencia Juliaca.

Pantoja (2016), propuso un Modelo Logit para evaluar el Riesgo Crediticio en las Cajas Municipales de Ahorro y Crédito: Caso de la Caja Municipal de Huancayo, periodo 2011-2015 y se planteó como objetivo de investigación, el estimar las probabilidades de incumplimiento basado en el comportamiento de pago del cliente, para la CMAC de Huancayo. El resultado de la investigación muestra la estrategia y metodología, así como etapas necesarias para diseñar un modelo logit y la implementación de esta metodología crediticia avanzada en la CMAC de Huancayo, contribuyéndose a reducir los niveles de riesgo crediticio y el cálculo del ahorro en provisiones exigidas por las altas tasas de morosidad para ejemplo de la administración del riesgo de crédito para otras CMAC que busquen al igual que la CMAC Huancayo, mayor participación y rentabilidad ofreciendo asistencia técnica de calidad.

Ríos (2017), se planteó conocer los factores que influyen en la presencia de la enfermedad de Gastritis y para ello se utilizó la regresión logística como técnica estadística. Los factores de riesgo fueron: tomar antiinflamatorios para el dolor y/o inflamación, beber agua potable sin hervir, consumir alimentos en lugares que no



garantizan la calidad, consumir alimento hasta saciar el apetito, y la falta de tiempo para realizar tareas académicas universitarias. Los signos y síntomas característicos que influyeron en la presencia de la enfermedad de gastritis fueron: La acumulación de gases en el estómago (Aerofagia) y la presencia de vómitos con sangre (Hematemesis). Las características sociodemográficas que influyen en la presencia de la enfermedad de gastritis fueron: la edad, los hábitos alimenticios y toxicológicos, así como tomar desayuno todos los días, consumo de alimentos mirando televisión, consumo de alimentos fritos o dorados que son cocidos en aceites, consumo frecuente de alcohol y consumo frecuente de tabaco.

Guerra (2018), se planteó como objetivo analizar la gestión de riesgo crediticio para reducir el nivel de morosidad en el segmento de la pequeña y mediana empresa del Banco de Crédito del Perú. La gestión de riesgo crediticio del BCP se enfoca en la construcción de una gestión de riesgo sólido, empezando por la actualización de la estructura de gobierno para el apetito de riesgo, vinculando al BCP con sus unidades de negocio. Con el objetivo de consolidar una cultura de riesgos consistente con las mejores prácticas de la industria, el BCP dispuso programas y capacitaciones continuas de formación sobre gestión de riesgos para los colaboradores involucrados en la toma de riesgos. La gestión de riesgo crediticio está relacionada con los procesos y herramientas utilizadas para la aprobación de créditos, con estos nuevos procesos el BCP espera mejorar sus resultados. El Banco de Crédito del Perú cuenta con un proceso de otorgamiento de crédito compuesto por tres pilares: Otorgamiento, seguimiento y recuperación. El BCP realiza una forma de controlar el nivel de riesgo crediticio, que es a través de una alerta premura.

Cano (2018), se planteó como diseño de investigación el diseño descriptivo y como tipo de investigación, correlacional y explicativo, la población estuvo



conformada por todos los pacientes de la Red de Salud Puno con diagnóstico de tuberculosis en el año 2017, que son en total de 60, pacientes del programa de prevención de la TBC, como muestra se tomó 34 pacientes que no abandonaron y 26 pacientes que abandonaron el tratamiento de tuberculosis. La técnica estadística utilizada para el análisis fue la Regresión Logística, usando para ello el programa estadístico SPSS 24.0. Los resultados fueron: las variables que inducen al abandono del tratamiento de tuberculosis - Red de Salud Puno, 2017 son: X6: Dejó el tratamiento por molestias causadas por las pastillas, X17: Molestias durante el tratamiento, X23: Calidad de servicio al paciente, X24: Apoyo que reciben de parte del centro de salud. El Modelo Logístico es el siguiente:  $Y = 2,013 - 2,940X_6 + 0,415X_{17} + 1,310X_{23} - 1,870X_{24}$ . El cual explica en un 91.7% su fiabilidad.

Sanchez Nina (2017), su trabajo tuvo como propósito encontrar un modelo estadístico que determine la deserción estudiantil en las escuelas profesionales de la Universidad Nacional del Altiplano - Puno. para la determinación del modelo utilizó la estadística multivariada, concretamente la regresión logística binaria, los resultados son: Se determinó y codifico las variables para el modelo estadístico de deserción estudiantil en las escuelas profesionales de la UNA-PUNO. Utilizando la regresión logística binaria. El porcentaje de clasificación correcta utilizando el modelo es del 91.2%. Cuando las variables interactúan en conjunto utilizando la prueba chi-cuadrada o prueba ómnibus. Se determinó los coeficientes significativos del modelo estadístico adecuado para la deserción estudiantil utilizando la regresión logística binaria.



## 2.2. MARCO TEÓRICO

### 2.2.1. Regresión

Algunos profesionales, sin importar su especialidad, confían en su intuición para juzgar como se relacionan dos variables. Debido a ello, hacen pronósticos a tuestas e incluso temerarios; sin embargo, si dichos profesionales tienen la posibilidad de tomar datos y utilizar un procedimiento estadístico de análisis para determinar cómo lo conocido se relaciona con el evento futuro, podrían ayudar considerablemente en el mejoramiento de los procesos que administran o en la solución eficaz de los problemas que se les presentan. El procedimiento estadístico que se utiliza para este fin se conoce como análisis de regresión que permite establecer la relación funcional o ecuación matemática que relaciona las variables, así como la fuerza de esa relación (Cardona, Gonzáles, Rivera, & Cárdenas, 2013).

La regresión lineal es la técnica básica del análisis de variables. Mediante dicha técnica tratamos de determinar relaciones de dependencia de tipo lineal entre una variable dependiente o endógena, respecto de una o varias variables explicativas o exógenas. Gujarati (1975), define el análisis de regresión como el estudio de la dependencia de la variable dependiente, sobre una o más variables explicativas, con el objeto de estimar o predecir el valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en medias muestrales repetidas) de las últimas (Parra, 2016).

El análisis de regresión es ampliamente utilizado para la predicción y previsión, donde su uso tiene superposición sustancial en el campo de aprendizaje automático. El análisis de regresión se utiliza también para comprender cuales de

las variables independientes están relacionadas con la variable dependiente, y explorar las formas de estas relaciones. En circunstancias limitadas, el análisis de regresión puede utilizarse para inferir relaciones causales entre las variables independientes y dependientes. Sin embargo, esto puede llevar a ilusiones o relaciones falsas, por lo que se recomienda precaución (Armstrong, 2012).

Según (Tusell, 2011), el interés de señalar que el ajuste de un modelo de regresión no se limita a analizar la relación entre dos variables; en general, buscaremos relaciones del tipo

$$y = f(x_0, x_1, x_2, \dots, x_{p-1}) + \varepsilon$$

relacionando de manera aproximada los valores de  $y$  con los que toman otras variables,  $x_0, \dots, x_{p-1}$ . Por simplicidad, limitaremos por el momento nuestra atención a funciones  $f(x_0, \dots, x_{p-1})$  lineales; el modelo resultante

es el modelo de regresión lineal. Señalemos, finalmente, que el hecho de aislar una variable  $y$  al lado izquierdo y escribirla como función de otras más una perturbación aleatoria  $\varepsilon$  no prejuzga ninguna relación de causalidad en ningún sentido; sólo postulamos la existencia de una relación cuya forma y alcance queremos investigar.

### **2.2.2. Distribución Binomial**

Según Wayne, (2006), la distribución binomial es una de las distribuciones utilizadas más ampliamente en estadística aplicada. La distribución se deriva de un procedimiento conocido como ensayo de Bernoulli, nombrado así en honor del matemático suizo James Bernoulli (1654-1705), quien realizó contribuciones importantes en el campo de la probabilidad, incluyendo, particularmente, la distribución binomial. Cuando en un proceso aleatorio experimento, llamado





ensayo, puede ocurrir sólo uno de dos resultados mutuamente excluyentes, como vida o muerte, enfermo o sano, masculino o femenino, el ensayo se llama ensayo de Bernoulli.

### **Proceso de Bernoulli**

Con frecuencia un experimento consta de pruebas repetidas, cada una con dos resultados posibles que se pueden denominar éxito o fracaso. La aplicación más evidente tiene que ver con la prueba de artículos a medida que salen de una línea de ensamble, donde cada prueba o experimento puede indicar si un artículo está o no defectuoso. Podemos elegir definir cualquiera de los resultados como éxito. El proceso se conoce como proceso de Bernoulli y cada ensayo se denomina experimento de Bernoulli (Walpole, Myers, Myers, & Ye, 2012).

En términos estrictos el proceso de Bernoulli se caracteriza por lo siguiente:

- 1) El experimento consta de ensayos repetidos.
- 2) Cada ensayo produce un resultado que se puede clasificar como éxito o fracaso.
- 3) La probabilidad de un éxito, que se denota con  $p$ , permanece constante de un ensayo a otro.
- 4) Los ensayos repetidos son independientes.

### **2.2.3. Regresión Logística**

Dentro de los métodos de clasificación, una posibilidad es intentar construir un modelo que explique los valores de la variable de clasificación. Por ejemplo, si se desea discriminar entre créditos que se devuelven o que presentan problemas para su cobro, puede añadirse a la base de datos una nueva variable,  $y$ , que tome el valor cero cuando el crédito se devuelve sin problemas y el valor uno



en otro caso (riesgo crediticio). El problema de discriminación se convierte en prever el valor de la variable ficticia,  $y$ , en un nuevo elemento del que conocemos el vector de variables  $x$ . Si el valor previsto está más próximo a cero que a uno, clasificaremos al elemento en la primera población. En otro caso, lo haremos en la segunda. Para modelar este tipo de relaciones se utilizan los modelos de respuesta cualitativa (Peña, 2002).

La regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica o variable de respuesta binaria, para el cual el resultado de la respuesta para cada sujeto es un "éxito" o "fracaso" en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores. El análisis de regresión logística pertenece al conjunto de Modelos Lineales Generalizados (GLM) que usa como función de enlace la función logit. Las probabilidades que describen el posible resultado de un único ensayo se modelan, como una función de variables explicativas, utilizando una función logística (Agresti, 2007).

Según Visauta, (1998) la regresión logística es un modelo que, a partir de los coeficientes estimados para cada una de las variables independientes y fruto de la probabilidad de los individuos en la dependiente, va a permitir asignar los mismos a una u otra categoría y opción de respuesta.

Podemos deducir del concepto de regresión logística, que ésta guarda un gran parecido con otras técnicas de análisis multivariable como la regresión lineal múltiple, así como con los modelos loglineales. Sin embargo, la regresión logística se caracteriza por un conjunto de rasgos que la diferencian de dichas técnicas.

La regresión logística resulta útil para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de variables predictoras. Es similar a un modelo de regresión lineal pero está adaptado para modelos en los que la variable dependiente es dicotómica. Los coeficientes de regresión logística pueden utilizarse para estimar la razón de las ventajas (odds ratio) de cada variable independiente del modelo. La regresión logística se puede aplicar a un rango más amplio de situaciones de investigación que el análisis discriminante (Perez, 2004).

Esta técnica tiene por objetivo estimar la probabilidad de ocurrencia  $p(x)$  de un evento, dado un conjunto de covariables predictoras  $x = (x_1, x_2, \dots, x_k)$ , ajustadas o controladas a ciertos valores particulares. El modelo matemático de forma-S comúnmente utilizado es la conocida función logística:

$$p(x) = \frac{e^{f(x)}}{1 + e^{f(x)}} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}} = \frac{1}{1 + e^{-\beta_0 - \sum_{i=1}^k \beta_i x_i}}$$

Una transformación más práctica en términos de lo que se conoce como razón de ventajas, expresada como el cociente de la probabilidad  $p(x)$  de ser moroso contra la probabilidad de ser cumplido  $1-p(x)$ , nos lleva al modelo de regresión en unidad logarítmica o función logit (del inglés log-unit):

$$\log it \Rightarrow \ln = \frac{p(x)}{1-p(x)} = f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

El cual es ahora un modelo lineal para el que se deben estimar los parámetros  $\beta_s$  mediante la técnica de máxima verosimilitud, en vez de utilizar la

técnica de mínimos cuadrados de la regresión lineal. Aunque, cabe mencionar que en el caso particular de la regresión lineal ambas técnicas producen las mismas estimaciones.

Para ello, tomemos un ejemplo: supongamos que estamos estudiando la malnutrición en niños de 0 a 6 años y su relación con las pautas de crianza. Las categorías de nuestra variable dependiente (y) eutrófico (que tomará el valor de 0) y mal nutrido (que tomará el valor 1) y para la variable independiente (x) las categorías serán democrático e indiferente (Alderete, 2006).

### Estimación de parámetros

La generalización del modelo simple, relaciona la probabilidad de que ocurra un determinado suceso independiente denotado por el vector  $x' = (x_1, \dots, x_p)$  con probabilidad condicional  $P(y=1/x)$  en función de  $p$  variables independientes que pueden ser cuantitativas, cualitativas o ambos tipos según sea el tipo de diseño de estudio.

Cálculo de los parámetros del modelo de regresión logística (método de máxima verosimilitud).

El modelo de regresión logística asume que:

$$y_1 = 1, \text{ con probabilidad } p_i(x) = \frac{1}{1 + e^{-\sum_{j=1}^p \beta_j x_{ij}}}$$

donde  $x_{i0}$  y  $p$  es el número de variables explicativas. Los parámetros desconocidos del modelo son:  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ .

La función de verosimilitud de la muestra contendrá factores del tipo 1 si  $y_i = 1$ , y del tipo 2 si  $y_i = 0$ , con lo que la función de verosimilitud  $L$  será:

$$L_{\beta_0, \beta_1, \dots, \beta_p} = \frac{\prod_{i=1}^n e^{\left( y_i \sum_{j=1}^p \beta_j x_{ij} \right)}}{\prod_{i=1}^n \left( 1 + e^{\sum_{j=1}^p \beta_j x_{ij}} \right)} = \frac{e^{\sum_{j=1}^p \beta_j t_j}}{\prod_{i=1}^n \left( 1 + e^{\sum_{j=1}^p \beta_j x_{ij}} \right)}$$

$$\text{Con } t_j = \sum_{i=1}^n x_{ij} y_i$$

El ln con  $L$  queda como:

$$\ln \left[ L_{\beta_0, \beta_1, \dots, \beta_p} \right] = \sum_{j=0}^p \beta_j t_j - \sum_{j=1}^p \ln \left[ 1 + e^{\sum_{j=1}^p \beta_j x_{ij}} \right]$$

Los  $\beta_j$  que maximizan  $L(\cdot)$ , también maximizarán  $\ln[L(\cdot)]$  y se obtienen resolviendo el siguiente sistema de  $p+1$  ecuaciones no lineales:

$$I_{j_1 j_2} = - \sum \frac{x_{ij_1} - x_{ij_2} e^{\sum_{j=0}^p \beta_j x_{ij}}}{\left[ 1 + e^{\sum_{j=1}^p \beta_j x_{ij}} \right]} = 0, \quad j_1, j_2 = 0, 1, \dots, p$$

Llamamos  $I_{j_1 j_2}$  se tiene que  $I_{j_1 j_2}^{-1}$  es la matriz asintótica de varianzas y covarianzas de  $\beta_j$ 's.

Un IC  $(1 - \alpha)$  para los  $\beta_j$  se puede obtener usando:

$$\beta_j = Z \left( 1 - \frac{\alpha}{2} \right) \sqrt{I_{jj}^{-1}}$$

(Peñas, Muñoz, Juez, & Cortinas, 2003)

## Interpretación de los parámetros $\beta$

En el caso de una regresión lineal, el valor obtenido en la estimación de un parámetro  $p$  nos dice cuál va a ser la variación en la variable a explicar cuando la variable  $X_p$  asociada experimenta un incremento en una unidad, siempre y cuando las demás variables explicativas se mantengan constantes. En este contexto podemos definir el parámetro  $p$  como sigue:

$$\beta_p = f(x_p + 1) - f(x_p)$$

Siendo

$$f(x_p + 1) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + 1 + \dots + \beta_p x_p$$

$$f(x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \dots + \beta_p x_p$$

La interpretación de un parámetro  $\beta_p$  en un modelo LOGIT no es tan evidente. Teniendo en cuenta la relación que existe entre la variable a explicar y las variables explicativas, un parámetro  $\beta_p$  quedaría definido de la siguiente forma:

$$\beta_p = f(x_p + 1) - f(x_p)$$

Siendo

$$f(x_p + 1) = \ln\left(\frac{\prod x_p + 1}{1 - \prod x_p + 1}\right)$$

$$f(x_p) = \ln\left(\frac{\prod x_p}{1 - \prod x_p}\right)$$

$\beta_p$  representa, pues, el cambio que experimentaría la relación entre la probabilidad de que un fenómeno ocurra y la probabilidad de que no ocurra, cuando

la variable  $X_p$  aumente en una unidad y las demás variables explicativas se mantengan constantes.

La interpretación de un parámetro  $\beta_p$  no se efectuará directamente, sino que se hará a partir de un ratio de probabilidad derivado (odds ratio). Dicha interpretación también dependerá del tipo de variable explicativa de que se trate: dicotómica o de más de dos categorías.

### Intervalo de confianza

Consideremos la función logística multivariada con parámetros

$$\beta_0, \beta_1, \beta_2, \dots, \beta_k : p(t) = [1 + e^{u_t}]^{-1} \text{ donde } u_t = -\beta_0 - \sum_{i=1}^k t_i \beta_i ; t = t_1, t_2, \dots, t_k$$

Supongamos que  $\beta_0^*, \beta_1^*, \dots, \beta_k^*$  es el estimador máximo verosímil de  $\beta_0, \beta_1, \dots, \beta_k$  basado en una muestra aleatoria de tamaño suficientemente grande como para admitir que dicho vector de estimaciones sigue una distribución normal multivariada.

Se cumple entonces que, asintóticamente, para todo vector  $t$  fijo, la variable aleatoria  $u_t^* = -\beta_0^* - \sum_{i=1}^k t_i \beta_i^*$  tienen distribución normal con media  $m^*(t)$  y varianza igual a:

$$\sigma_t^2 = \text{Var}(\beta_0^*) + \sum_{i=1}^k t_i^2 \text{Var}(\beta_i^*) + 2 \sum_{i=1}^k t_i \text{Cov}(\beta_0^*, \beta_i^*) + 2 \sum_{1 \leq i < j} t_i t_j \text{Cov}(\beta_i^*, \beta_j^*)$$

así como que

$$S_t^2 = \hat{\text{Var}}(\beta_0^*) + \sum_{i=1}^k t_i^2 \hat{\text{Var}}(\beta_i^*) + 2 \sum_{i=1}^k t_i \hat{\text{Cov}}(\beta_0^*, \beta_i^*) + 2 \sum_{1 \leq i < j} t_i t_j \hat{\text{Cov}}(\beta_i^*, \beta_j^*)$$

es un estimador insesgado de  $\sigma^2$ .

Teniendo ahora en cuenta la expresión del intervalo de confianza para la media de la distribución normal:

$$P\left[u_t^* - Z_\alpha S_t \leq \mu_t \leq u_t^* + Z_\alpha S_t\right] = 1 - \alpha$$

### Prueba de significancia

Para contrastar la significatividad global en los modelos logit o probit no se pueden construir estadísticos  $F$  que tengan distribución  $F$  de Snedecor, debido a que son modelos no lineales. En su lugar usamos el estadístico de razón de verosimilitud (RV).

Como en el caso univariante, se prueba la significancia de las variables independientes del modelo mediante la prueba de verosimilitud con la significancia de los  $K+1$  parámetros, para determinar si las variables independientes influyen significativamente en la probabilidad del suceso del modelo relacionado a la variable en estudio del siguiente modo:

Concretamente, para contrastar la significatividad global del modelo las hipótesis nula y alternativa son las siguientes:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \beta_i \neq 0$$

mientras que la RV se calcula de la siguiente forma:

$$RV_0 = -2\ln(L_0) - \ln(L)$$

donde  $\ln(L)$  es el logaritmo de la función de verosimilitud que se ha obtenido al estimar el modelo completo, mientras que  $\ln(L_0)$  es el logaritmo de la función de verosimilitud que se ha obtenido al estimar el modelo con sólo el término



independiente. Se verifica que este estadístico se distribuye asintóticamente, es decir cuando el tamaño de la muestra es grande, aunque estrictamente es cuando el tamaño de la muestra tiende a  $\infty$  como una Chi cuadrado con  $k-1$  grados de libertad, siendo  $k$  el número de regresores del modelo incluido el término independiente. Por tanto,

$$RV_0 \square \chi_{k-1}^2$$

En general, cuando se desee contrastar la significatividad de un subconjunto de coeficientes, el estadístico a aplicar es el siguiente:

$$RV_0 = -2\ln(L_R) - \ln(L)$$

Donde  $\ln(L_R)$  es el logaritmo de la función de verosimilitud que se obtiene al estimar el modelo en el que están excluidos los  $r$  términos cuya significatividad se desea contrastar, Bajo la hipótesis nula el estadístico anterior se distribuye como:

$$RV_0 \square \chi_r^2 \text{ grados de libertad}$$

Una vez encontrado el mejor conjunto de variantes explicativas que predicen la variable  $Y$ , se debe evaluar mediante Wald cada coeficiente para determinar cuál o cuáles ingresan al modelo

### **Estadístico WALD**

Existe otra vía para probar, en la regresión logística, si una variable independiente hace un aporte estadísticamente significativo a la “explicación” de la variable de respuesta sin tener que usar la razón de verosimilitudes. Este método es frecuentemente llamado “prueba de Wald”. Esta prueba permite contrastar si un coeficiente  $\beta$  es significativamente diferente de 0.

El estadígrafo para realizar la prueba de Wald se obtiene dividiendo la estimación del coeficiente de interés por su error estándar:

$$Z_{Wald} = \frac{\hat{B}}{s_e(\hat{B})}$$

Donde  $s_e(\hat{B})$  es el error estándar de  $\hat{B}$ .

El estadígrafo  $Z_{Wald}$  sigue una distribución normal estándar. Por lo general los paquetes estadísticos proveen este último y su valor de  $p$  asociado.

La razón de verosimilitudes y la prueba de *Wald* brindan resultados similares en muestras grandes, por ello en estudios con suficiente tamaño de muestra no importa cuál de las dos pruebas sea la usada. Sin embargo, en muestras no muy grandes, dichas pruebas producen diferentes resultados. Por lo general se recomienda el uso de la razón de verosimilitudes.

Evalúa la significancia de los coeficientes, se define como el vector matriz de los coeficientes estimados, según hipótesis:

$$H_0: B_i = 0$$

$$H_1: \text{Para algún } B_i \neq 0, i = 1, \dots, k$$

Como se puede apreciar, la hipótesis viene a ser una extensión de la regresión logística simple, puesto que se trabajará con  $k$  parámetros debido a que nuestro trabajo comprende más de 1 variable con su respectivo coeficiente  $B$ .

La regla de decisión será con un nivel de significación fijado  $\alpha$ , se concluye que la variable independiente influye en la probabilidad del suceso.



## Prueba de bondad de ajuste

Para evaluar la bondad de ajuste del modelo se utiliza la probabilidad de Hosmer-Lemeshow, que consiste en calcular para cada observación del conjunto de datos las probabilidades esperadas y se compara con las frecuencias observadas mediante una prueba  $\chi^2$  con  $j-2$  grados de libertad, donde  $j$  es el número de grupos formados como se explicó en el modelo simple. El modelo se ajusta bien si no hay evidencia para rechazar la hipótesis nula

### 2.2.4. Simulación

La simulación es la técnica que consiste en realizar experimentos de muestreo sobre un modelo de un sistema. Un modelo no es más que un conjunto de variables junto con ecuaciones matemáticas que las relacionan y restricciones sobre dichas variables. La modelización es una etapa presente en la mayor parte de los trabajos de investigación (especialmente en las ciencias experimentales). En muchos casos la realidad es bastante compleja como para ser estudiada directamente y es preferible la formulación de un modelo que contenga las variables más relevantes que aparecen en el fenómeno en estudio y las relaciones más importantes entre ellas (Cao Abad, 2002).

La simulación se refiere a un gran conjunto de métodos y aplicaciones que buscan imitar el comportamiento de sistemas reales, generalmente en una computadora con un software apropiado. De hecho, la “simulación” puede ser un término extremadamente general dado que se utiliza en muchos campos, industrias, finanzas, etc. En estos días la simulación es más popular y poderosa que nunca, ya que las computadoras y el software son mejores de los que nunca han existido (Kelton, Sadowsky, & Sturrock, 2008).



La experimentación directa sobre la realidad puede tener muchos inconvenientes:

- Un coste muy alto
- Gran lentitud
- En ocasiones las pruebas son destructivas
- A veces no es ética (experimentación sobre seres humanos)
- Puede resultar imposible (un acontecimiento futuro)

Razones como estas pueden ser indicadores de la ventaja de trabajar con la simulación de un sistema real. A través de la estadística podremos estimar los parámetros y contrastar la validez de un modelo a partir de los datos observados del sistema real que se pretende simular.

### **Generación de números pseudoaleatorios**

Según Cao Abad, (2002), casi todos los métodos de simulación se basan en la posibilidad de generar números pseudoaleatorios con distribución uniforme  $U(0,1)$ . Hasta el gran desarrollo de los ordenadores los números aleatorios se obtenían por procedimientos experimentales y se almacenaban en tablas. En la actualidad estos números son generados por ordenadores y se denominan pseudoaleatorios ya que, en realidad, todos los números de la sucesión que se genera son predecibles a partir del primero, llamado semilla. En cualquier caso, todo generador de números pseudoaleatorios mínimamente aceptable debe comportarse como si se tratase de una muestra genuina de datos independientes de una  $U(0,1)$ .



## Generación de variables aleatorias

Hay una variedad de métodos para generar variables aleatorias. Cada método se aplica solo a un subconjunto de distribuciones y para una distribución en particular un método puede ser más eficiente que otro. Entre los métodos más conocidos se tiene: transformación inversa, método de rechazo, composición y convolución (Azarang & Garcia, 1996).

La facilidad de aplicación de dichos métodos, así como el coste computacional asociado a los mismos, varía mucho según la familia de variables aleatorias a las que se apliquen. Normalmente existen varios algoritmos que se pueden utilizar para generar valores de una determinada distribución, y diferentes factores que se pueden considerar para determinar qué algoritmo utilizar en un caso particular. Desafortunadamente dichos factores suelen entrar en conflicto unos con otros y a veces se ha de llegar a una solución de compromiso.

Algunos de estos factores son los siguientes:

- **Exactitud:** se han de obtener valores de una variable con una precisión dada. A veces se tiene suficiente con obtener una aproximación y otras no.
- **Eficiencia:** el algoritmo que implementa el método de generación tiene asociado un tiempo de ejecución y un gasto de memoria. Elegiremos un método que sea eficiente en cuanto al tiempo y a la cantidad de memoria requeridos.
- **Complejidad:** Buscamos métodos que tengan complejidad mínima, siempre y cuando se garantice cierta exactitud.



### 2.2.5. Factores de riesgo

Que se entienda por un agente causal y por un factor de riesgo es sin duda polémica. Cualquiera que sea la definición de actor de riesgo, la definición funcional de la causa de un efecto dado puede expresarse del modo siguiente: cualquier factor, condición o característica, cuya supresión elimina la posibilidad de que se produzca el efecto, es una causa del mismo. Por otra parte, están los elementos que, sin ser necesarios, son propiciatorios del efecto. Se trata de los llamados factores de riesgo (FR) para desarrollar una enfermedad o una condición. Se trata de factores susceptibles de ser modificados que, sin ser causas propiamente, pueden favorecer que el agente causal actúe. Ejemplos típicos de factores de riesgo en el área de salud son los de no usar anticonceptivos, circunstancia que favorece el embarazo y las enfermedades de transmisión sexual, el consumo de grasas saturadas a los efectos de las cardiopatías, o la obesidad como condición que propicia el surgimiento de dolencias vasculares y en el área financiera pueden ser, balance de una empresa ganancia o pérdida, solicitud de préstamo (aprobado y desaprobado), condición de pago, etc (Silva Ayçaguer, 2005).

### Índice Morosidad

Según la Real Academia Española (2006), enuncia que la morosidad es el pago después de los días pactados en el contrato de préstamo. A pesar que el deudor tiene conocimiento de su deuda, y se atrasa descuidadamente, ya sea por un hábito, la falta de un centro laboral o puntualidad.

Chavarrín (2015), refiere que en la literatura académica se acostumbra tomar como indicador del riesgo de crédito a la razón de provisiones preventivas para riesgos crediticios respecto a los préstamos totales. Pero, como señalan Louzis



et al. (2012), el riesgo de crédito ex post toma la forma del índice de morosidad, que se define como la proporción de cartera vencida respecto a la cartera total de un banco. La evolución del índice de morosidad es consecuencia de factores económicos, políticos y regulatorios, y varía entre bancos debido a las diferencias en sus políticas de gestión del crédito y de la tecnología y recursos con que operan éste.

### **Riesgo de morosidad para créditos bancarios**

El control de gestión es una herramienta en la cual se basa una institución financiera para medir su desempeño; los indicadores de riesgo son los instrumentos indispensables para medir dicho desempeño a través de fórmulas y cálculos matemáticos que se aplican a los estados financieros, resultados que nos ayudaran a medir la salud de instituciones financieras individuales (Valencia-Cárdenas & Restrepo-Morales, 2016).

Existen muchos sistemas que permiten medir el desempeño de las instituciones prestamistas, de su aplicación se crea las calificaciones crediticias que son combinaciones de letras que acompañarán al nombre de la entidad determinando su nivel de riesgo crediticio (Díaz Córdova, Coba Molina, & Navarrete, 2017).

Dada la dificultad de vinculación con el sector empresarial para utilizar datos reales, presentamos un caso de estudio simulado sobre créditos bancarios, el cual intenta ampliar los resultados y análisis reportados para regresión logística. Nuestro objetivo es determinar qué factores, caracterizan la morosidad en la Caja Rural de ahorro y crédito “Los Andes”, así como incluir variables adicionales de interés práctico como el ingreso.



## 2.3. DESCRIPCIÓN DE VARIABLES

**Variable Dependiente (VD)** = Riesgo de morosidad.

*Dimensión:* La **morosidad** de los créditos de la Oficina Puno de la Caja Rural de ahorro y crédito “Los Andes” del tercer trimestre 2019.

**Variables independientes (VI)** = Factores de capacidad de pago del cliente

El crédito debe otorgarse en función a la capacidad de pago del cliente, señala Garcia, (2019), seguidamente, refiere que la capacidad de pago del comprador se determina a través de su situación económico-financiera y de su flujo de caja, así como de sus antecedentes crediticios. Estos factores nos indicarán si el cliente cuenta y podrá seguir contando con capacidad y voluntad de pago.

Así mismo, estos factores son considerados del trabajo de investigación realizado por López Miranda, (2013) y Mures Quintana *et al.*, (2005), con base al perfil socioeconómico del cliente, considerando el número de dependientes económicos, número de impagos previos, su empleo, destino del crédito y salario mensual. Estos factores son adaptados a nuestro medio y considerando otros más, que investigaciones similares de nuestra región y país determinan importantes.

*Dimensiones:*

- **Destino del crédito:** Es el uso que se le dará al dinero prestado. De acuerdo a la información proporcionada por el analista de crédito (Anexo A), los clientes destinan a *Negocio* (capital de trabajo, activos fijos y compra deuda); *Vivienda* (Compra y construcción de vivienda) y *Otras inversiones* (el refinanciamiento y la libre disponibilidad).



- **Tipo cliente:** Corresponde a si el cliente es *Nuevo* o solicito una ampliación, prepago, paralelo, refinanciamiento o es recurrente, en este caso lo llamaremos en general *Recurrente*.
- **Días de atraso:** Un historial crediticio se daña cuando te retrasas en tus pagos, por eso, la Superintendencia de Banca, Seguros y AFP (SBS) presenta una calificación crediticia según la cantidad de días retrasados en el pago de la cuota mensual, como se muestra en la tabla 1.

**Tabla 1.**

*Calificación de días de morosidad*

<i>CATEGORÍA DE CLASIFICACIÓN</i>	<i>EN CREDITO DE CONSUMO*</i>
0: Normal	Pago puntual o atraso máximo de 8 días calendarios
1: Problemas potenciales	Atrasos en el pago de entre 9 a 30 días calendarios
2: Deficiente	Atrasos en el pago de entre 31 a 60 días calendarios
3: Dudoso	Atrasos en el pago de entre 61 a 120 días calendarios
4: Pérdida	Atrasos en el pago de más de 120 días calendarios

*(\*) Los créditos de consumo pueden ser de tipo revolvente (Ejm. tarjeta de crédito) y no revolvente (Ejm. Crédito vehicular)*

**Fuente:** SBS, (2019)

- **Total Ingresos:** Es el total de ingresos (en soles) que aporta el clientes a su familia y/o negocio durante un mes, pueden provenir del salario del clientes (en caso de ser empleado público o privado) y negocios independientes.
- **Empleo:** Es el ejercicio de funciones productivas. Se considera a todos los clientes de 18 años y más como población en edad activa o población en edad de

trabajar. Considerando las categorías de empleo independiente, estatal, privado, eventual y algunos estados combinados.

- **Sector procedencia:** Corresponde al origen y residencia del cliente, ya sea la sector urbano u rural.
- **Estado civil:** Es la situación de las personas físicas determinada por sus relaciones de familia, provenientes del matrimonio o del parentesco, que establece ciertos derechos y deberes.

**Tabla 2.**

*Operacionalización de variables.*

<i>Variable</i>	<i>Dimensión</i>	<i>Categoría</i>
VD: Riego de Morosidad	Morosidad	1: Moroso 0: No moroso
	Destino del crédito	0: Negocio 1: Vivienda 2: Otras inversiones
	Tipo Cliente	1: Nuevo 0: Recurrente
	Días de atraso	0: Normal 1: Problemas potenciales 2: Deficiente 3: Dudoso 4: Pérdida
VI: Factores de capacidad de pago del cliente	Total, ingresos	Soles (S/)
	Empleo	1: Independiente 2: Estatal 3: Privado 4: Eventual 5: Independiente y estatal 6: Independiente y privado 7: Jubilado 8: Desempleado
	Sector procedencia	0: Urbano 1: Rural
	Estado civil	0: Soltero 1: Casado 2: Divorciado 3: Conviviente 4: Viudo

*Fuente: Elaboración propia*



## CAPÍTULO III

### MATERIALES Y MÉTODOS

#### 3.1. UBICACIÓN GEOGRÁFICA DEL ESTUDIO

La ciudad de Puno, capital de distrito, provincia y del departamento de Puno, está ubicado a orillas del Lago Titicaca a 3827 m.s.n.m., lago navegable más alto del Mundo. Se encuentra en la región de la sierra a los 15° 50' 26" de latitud sur, 70° 01' 28" de longitud Oeste del meridiano de Greenwich. Tiene una extensión territorial de 460.75 km<sup>2</sup> y una población total de 141, 064 habitantes. Debido a su ubicación en la región sur andina de la cordillera predomina el clima frígido seco (otoño e invierno) a húmedo templado (primavera y verano). La investigación se realizó en local principal de la Caja Rural de ahorro y crédito “Los Andes” de la ciudad de Puno.

#### 3.2. ÁMBITO DE ESTUDIO

La Caja Rural de Ahorro y Crédito Los Andes S.A. (“la Caja”, “Los Andes”), es una institución financiera que inició operaciones en diciembre de 1997, con sede en Puno, bajo la denominación de “Caja Rural de Ahorro y Crédito Los Aymaras S.A.”, con dirección lega en el Jr. Junin N° 129, dedicándose desde entonces a actividades de intermediación microfinanciera, especializada en el sector rural de la sierra sur del Perú, en donde tiene el 100% de sus colocaciones.

#### 3.3. TIPO Y DISEÑO DE INVESTIGACIÓN

La investigación desarrollada corresponde al tipo de estudio correlacional y el diseño es no experimental transversal, porque se utilizó herramientas y evidencias



teóricas relacionadas con los temas de factores de riesgo y otorgamiento de créditos a clientes.

En la investigación, no se construyó ninguna situación, si no, que se observaron situaciones ya existentes como son el otorgamiento de créditos, el número de impagos, lo cual no fue provocada por el investigador.

Los diseños de investigación transeccional recolectan datos en un solo momento, en un tiempo único como lo es la presente investigación.

### **3.4. POBLACIÓN Y MUESTRA**

#### **3.4.1. Población**

Con el propósito obtener las variables con sus respectivas categorías y estimar los parámetros del modelo, la población está constituida por todos los clientes que solicitaron crédito en el segundo trimestre del 2019 de la Caja de ahorro y crédito “Los Andes” de Puno.

#### **3.4.2. Muestra**

Para desarrollar el presente trabajo de investigación se consideró al personal de la caja de ahorro y crédito “Los Andes” de Puno – agencia principal, en un número de 02 analistas de crédito y una muestra de 20 clientes de cada una, escogidos al azar y por conveniencia de la cartera de créditos de la institución, todos ellos hacen una sumatoria del 100% de la muestra total de los clientes de créditos.



### **3.5. INSTRUMENTO DE RECOLECCIÓN Y PROCESAMIENTO DE DATOS**

Los datos son proporcionados por los analistas de crédito de la caja de ahorro y crédito “Los Andes” de Puno; y estos fueron almacenados en libros de Excel, conteniendo información en general de variables financieras, sociodemográficas y de comportamiento según el código del deudor.

#### **3.5.1. Procedimiento de recolección de datos**

Las recolecciones de los historiales de crédito fueron conseguidas mediante una entrevista con los analistas de crédito y jefe del área de informática de la caja de ahorro y crédito “Los Andes” de Puno; y posterior entrega de fichas de recolección de datos.

#### **3.5.2. Plan de procesamiento y análisis de datos**

La información obtenida fue ordenada, seleccionada, analizada y presentada en tablas estadísticas de una y doble entrada - En este proceso se realizaron las siguientes actividades:

##### **Análisis de la información obtenida cantidades y porcentajes**

Para el procesamiento, se utilizó el programa estadístico IBM SPSS versión 24.0 ya que es un programa muy usado en las ciencias sociales, además de las empresas de investigación de mercado. El nombre originario correspondía al acrónimo de Statistical Package for the Social Sciences (SPSS), reflejando la orientación a su mercado original, aunque este programa es también muy utilizado en otros campos como las ciencias de la salud y la mercadotecnia.



## Diseño Estadístico

El presente trabajo presenta a través de la simulación, predecir la probabilidad de recuperar un crédito, con base a la información socioeconómica de los clientes de la caja de ahorro y crédito “Los Andes” de la ciudad de Puno en el periodo 2019-I. Para ello, es preciso aplicar una metodología determinada. De lo que se trata es de plantear un modelo de regresión logística que estime la probabilidad del comportamiento de pago de los clientes de las entidades financieras (Mures Quintana et al., 2005). Por tanto, el objetivo del trabajo se centra en explicar el comportamiento de una variable categórica con dos modalidades: ser un cliente moroso o bien un cliente no moroso.

En la técnica de regresión logística haciendo uso del paquete SPSS V.24 se analizó la significancia estadística mediante la prueba Wald como la prueba de Hosmer-Lemeshow, para determinar que tan bien se ajustan el modelo a los datos. Luego se estimó los parámetros del modelo, y se usó para medir su capacidad predictora, la interpretación de las ORs e intervalos de confianza.



## CAPÍTULO IV

### RESULTADOS Y DISCUSIÓN

A continuación, desarrolla el análisis empírico, que comprende las etapas de determinación de aquellas variables que permiten explicar la morosidad, obtenidas de investigaciones anteriores y la estimación de los parámetros de cada variable a partir de una muestra de clientes objeto de estudio, proporcionada por los analistas de crédito de la Caja Los Andes y finalizamos con la generación de la base de datos y a partir de la cual se estimó el modelo de regresión logística y la respectiva simulación de la variable de interés “Morosidad” con exposición de los principales resultados obtenidos.

#### 4.1. MODELO DE RIESGO DE MOROSIDAD

Mures Quintana et al., (2005), en su investigación determinan que tanto la técnica de análisis discriminante como la de regresión logística son adecuadas para el estudio y predicción de la morosidad, consiguiéndose mediante ambas técnicas una elevada eficacia predictiva, siendo esta última un método alternativo aplicado para el análisis de variables cualitativas. Tomando como referencia este trabajo, López Miranda, (2013) plantea un modelo de regresión logística para pronosticar la morosidad de un crédito bancario se desempeña adecuadamente como modelo clasificador sobre una base de datos simulada; en consecuencia, se asume factible que este tipo de modelos sea adecuado sobre casos reales o similares.

Basándonos especialmente en estos dos trabajos de investigación se toma como referencia el modelo de regresión logística y los factores que en ellas se describen adecuándolas a nuestro entorno de investigación, los clientes que solicitan un crédito en la Caja Los Andes de la ciudad de Puno. A continuación, se estiman

los parámetros de las variables independientes y dependiente para luego realizar la generación aleatoria de cada variable.

#### 4.1.1. Tasa del riesgo de morosidad

Dado que nuestro objetivo es caracterizar la morosidad o la insolvencia de las entidades financieras a través de la simulación, este parámetro es estimado a partir del informe que realiza la Caja Rural de ahorro y crédito “Los Andes” en su Memoria Anual 2018 e informes trimestrales del 2019, como se muestra a continuación:

**Tabla 3.**

*Calidad de portafolio de créditos (%) de la Caja Los Andes - Puno*

Rubro	2018	2019**		
		I	II	III
Índice de morosidad (%)	6.77*	6.10	10.02	12.94***

(\*) Al cierre del año 2018

(\*\*) Información trimestral del 2019

**Fuente:** Memoria Anual 2018, (Caja Los Andes, 2019; Gan@más, 2019)

En la tabla 3, el índice de morosidad de Caja Los Andes llegó a 12.94% en el tercer trimestre del 2019, registrando un aumento en 2.92 p.p. respecto al cierre del segundo trimestre del 2019 (10.02%) y se incrementó en 6.12 p.p. en comparación del tercer trimestre 2018 (6.82%).

#### 4.1.2. Estimación de parámetros de los factores de capacidad de pago del cliente

Como se indicó anteriormente, las variables o factores que se tomaron en cuenta para el modelo y posterior simulación fueron establecidos como significativos por López Miranda, (2013) y Mures Quintana *et al.*, (2005), con lo





cual, procedemos a estimar los parámetros de dichas variables a partir de la información proporcionada por los analistas de crédito.

**Tabla 4.**

*Destino de crédito solicitado por los clientes de la Caja Los Andes - Puno, tercer trimestre 2019.*

Destino de crédito	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Negocio	16	40 .0	40,0	40,0
Vivienda	1	2, 5	2,5	42,5
Otras inversiones	23	57 .5	57,5	100,0
Total	40	100,0	100,0	

**Fuente:** *Elaboración propia*

En la tabla 4, se observa que los clientes destinan el crédito en un 40% a los negocios (capital de trabajo, activos fijos y compra deuda), solo el 2.5% a vivienda (compra y construcción de vivienda) y el 57.5% lo destina a otras inversiones, es decir, libre disponibilidad. Estas estimaciones serán consideradas como parámetros para estimar el modelo de regresión logística.

**Tabla 5.**

*Tipo de cliente que solicita el crédito de la Caja Los Andes - Puno, tercer trimestre 2019.*

Tipo cliente	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Nuevo	18	45,0	45,0	45,0
Válido Recurrente	22	55,0	55,0	100,0
Total	40	100,0	100,0	

**Fuente:** *Elaboración propia*

En la tabla 5, se observa que el 45% son clientes nuevos, es decir que por primera vez han solicitado un crédito y estos serían potencialmente candidatos a ser morosos, dado que carecen de un historial crediticio y el 55% son clientes recurrentes con historial crediticio.

**Tabla 6.**

*Días de atraso (número de impagos) en el pago del crédito de la Caja Los Andes - Puno, tercer trimestre 2019.*

Días de atraso	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0 a 8 días	35	87,5	87,5	87,5
9 a 30 días	2	5,0	5,0	92,5
Válido 31 a 60 días	2	5,0	5,0	97,5
61 a 120 días	1	2,5	2,5	100,0
Total	40	100,0	100,0	

**Fuente:** *Elaboración propia*

La tabla 6, nos muestra que, el 87.5% no tienen retrasos o tienen 8 o menos días de retraso en sus pagos, de acuerdo a la calificación del SBS, estos clientes

están en la categoría de normal; tenemos con 5% a clientes entre 9 y 30 días de retraso, así como de 31 a 61 días con calificativos de clientes con problemas potenciales y deficientes respectivamente y finalmente clientes con 61 a 120 días de retraso solo con el 2.5% considerados clientes dudosos.

**Tabla 7.**

*Estadísticos descriptivos del Ingreso total del cliente de la Caja Los Andes - Puno, tercer trimestre 2019.*

	N°	Mínimo	Máximo	Media	Desviación estándar
Ingreso total	40	895,0	80700,0	15989,052	20415,5887
N válido (por lista)	40				

**Fuente:** *Elaboración propia*

En este caso la tabla 7 muestra el promedio y desviación del total de los ingresos (salario y otros ingresos) de los clientes, siendo un campo tomado muy en cuenta para acceder al crédito solicitado, entendemos que en promedio los clientes tienen ingresos que asciende a S/ 15,989.052 con una desviación de  $\pm$  S/. 20,415.59 indicando alta variabilidad en los ingresos de los clientes con un mínimo de S/ 895.00 y un máximo de S/ 80,700.00 considerando que hay más porcentaje de clientes que son del sector urbano, además no se tomó en cuenta las PYMES, considerando que corresponde a un análisis más profundo, debido a que tienen mayores ingresos.

**Tabla 8.***Tipo de empleo del cliente de la Caja Los Andes - Puno, tercer trimestre 2019.*

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Independiente	12	30,0	30,0	30,0
Estatad	10	25,0	25,0	55,0
Privado	8	20,0	20,0	75,0
Eventual	6	15,0	15,0	90,0
Independiente y estatal	1	2,5	2,5	92,5
Jubilado	1	2,5	2,5	95,0
Desempleado	2	5,0	5,0	100,0
Total	40	100,0	100,0	

**Fuente:** *Elaboración propia*

En la tabla 8, encontramos que el 30% son trabajadores independientes; el 25% son empleados del estado, es decir, tienen salario; el 20% trabajan en una empresa y/o institución privada, también cuenta con salario, un 15% son trabajadores eventuales; con un 2.5% trabajadores independientes y con trabajo estatal y Jubilados respectivamente; por último, están los desempleados pero que cuentan con ingresos con 5%. Es importante mencionar que en la muestra proporcionada por los analistas no se encuentra la categoría trabajador independiente y privado, pero será considerado en la generación de la variable aleatoria.

**Tabla 9.**

*Sector de procedencia del cliente de la Caja Los Andes - Puno, tercer trimestre 2019.*

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sec. Rural	13	32,5	32,5
	Sec. Urbano	27	67,5	100,0
	Total	40	100,0	100,0

**Fuente:** *Elaboración propia*

Con respecto al sector de procedencia, la tabla 9 muestra que el 67.5% de los clientes corresponden al sector urbano, este porcentaje es alto debido a que la muestra proviene de la oficina Puno y el 32,5% del sector rural.

**Tabla 10.**

*Estado civil del cliente de la Caja Los Andes - Puno, tercer trimestre 2019.*

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Soltero	23	57,5	57,5
	Casado	10	25,0	82,5
	Conviviente	6	15,0	97,5
	Viudo	1	2,5	100,0
	Total	40	100,0	100,0

**Fuente:** *Elaboración propia*

Por último, la tabla 10 muestra que el 57.5% de los clientes son solteros, el 25% casados, 15% convivientes y 2.5% viudo. La categoría divorciada no presenta casos en la muestra, pero será considerado para generación de la variable aleatoria con un porcentaje bajo.

#### **4.2. GENERACIÓN DE BASE DE DATOS MEDIANTE SIMULACIÓN**

El desarrollo de las técnicas de simulación está ligado primeramente al desarrollo de la tecnología de información y comunicación (TIC), y luego a la

necesidad de disminuir los riesgos adversos que se asocian a la actividad. Simultáneamente al desarrollo de esta tecnología se produce un importante avance en el estudio de los métodos de enseñanza y aprendizaje con simulación (Fraga-Sampedro, 2018).

La base de datos es implementada a través de la simulación mediante software (Excel). Cabe mencionar que dicha base sirve como conjunto de entrenamiento y prueba para generar el modelo, por lo que la variable morosidad es generado con los parámetros explicados anteriormente para luego contrastarla con el modelo clasificador-pronosticador, una vez que se estime las probabilidades o riesgo de morosidad. En ese sentido, es importante aclarar que la base de datos captura la supuesta relación intrínseca entre la variable de respuesta y los valores de las covariables, los cuales usamos para generar el modelo inmerso.

De acuerdo a la referencia teórica revisa, la variable Morosidad, Tipo cliente y Sector de procedencia son dicotómicas y tiene distribución de Beronulli y fueron generadas a través del método de transformada inversa, descrita en (Azarang & Garcia, 1996) como sigue:

$$\text{Si: } 0 \leq R_i < 1 - p \quad BE_i = 0$$

$$\text{Si: } 1 - p \leq R_i < 1 \quad BE_i = 1$$

Donde:

$p$  : probabilidad de ocurrencia del evento  $X = 1$ .

$1 - p$  : probabilidad de ocurrencia del evento  $X = 0$ .

$BE_i$  : número aleatorio con distribución de Beronulli.

Las variables Destino del crédito tiene distribución binomial y se generan a

través del método de convolución (García, García, & Cárdenas, 2013). El procedimiento es el siguiente:

$$BI_i = \sum_{j=1}^n BE_j$$

Donde:

$BE_j$  : número aleatorio con distribución de Bernoulli.

$N$  : número del evento máximo de la distribución binomial.

$p$  : probabilidad de la distribución Binomial que se involucra al generar los Bernoulli.

$BI_i$  : número aleatorio con distribución de Binomial.

Las variables Empleo, Días de atraso y Estado civil tienen distribución multinomial, entonces utilizando el método de la transformada inversa, para tomar una muestra de la variable  $X$  se procede del modo siguiente (García et al., 2013):

1. Tomar  $R$  de la uniforme  $[0,1]$ .
2. La variable toma el valor  $X = V$  que satisfaga:

$$\sum_{i=1}^{V-1} \pi_i < R \leq \sum_{i=1}^V \pi_i$$

Por acuerdo  $\sum_{i=1}^c \pi_i = 0$

3. Si desea obtener  $n$  muestras basta repetir el procedimiento anterior.

La variable Total ingreso tiene distribución normal con parámetros media  $\mu$  y desviación estándar  $\sigma$ , y por el método de convolución (Azarang & Garcia, 1996), se tiene:

$$N_i = \left[ \left( \sqrt{-2 \ln(1 - R_i)} \right) \cos(2\pi R_{i+1}) \right] \sigma + \mu$$

Donde:



- $\mu$  : media de la distribución normal.
- $\sigma$  : desviación estándar de la distribución normal.
- $N_i$  : número aleatorio con distribución normal.

**Tabla 11.**

*Criterios de las variables a ser simuladas*

<i>N</i>	<i>Descripción</i>	<i>Características</i>
<b>Variable dependiente (VD)</b>		
<i>Y</i>	Morosidad	Se generará una muestra de números aleatorios con distribución Bernoulli de parámetro $p = 0.1294$ (ver tabla 3).
<b>Variables independientes (VI)</b>		
$x_1$	Destino del crédito	Números aleatorios con distribución binomial de parámetros $n=2$ y $p=0.40$ . Donde 0 = Negocio, 1 = vivienda y 2 = otras inversiones.
$x_2$	Tipo cliente	Variable 0 - 1; muestra de números aleatorios con distribución Bernoulli con $p=0.55$ .
$x_3$	Días de atraso	Variable aleatoria discreta con valores enteros de categoría 0, 1, 2 y 3; más de 8 días de impagos suponemos rechazada automáticamente la solicitud de crédito. Se generará mediante una binomial ( $n=2$ , $p=0.125$ ).
$x_4$	Total ingreso	Variable aleatoria continua con distribución normal (en soles) con media S/15,989.052 y desviación estándar S/ 20,415.59
$x_5$	Empleo	Variable aleatoria con ocho categorías, generadas con una distribución multinomial de acuerdo a la tabla 8.
$x_6$	Sector de procedencia	La variable tiene distribución Bernoulli de parámetro $p = 0.325$ (ver tabla 9).
$x_7$	Estado civil	Variable aleatoria con 5 categorías, generadas con una distribución multinomial de acuerdo a la tabla 10.

**Fuente:** Elaboración propia.

El factor a considerar de importancia en un estudio de simulación es el tamaño de la muestra (número de corridas en el computador). La selección de un tamaño de muestra apropiado que sugiere un nivel deseado de precisión y a la vez minimice el costo de operación del modelo, es un problema algo difícil pero muy importante. Puesto que la información proporcionada por el experimento de simulación será la base para predecir con respecto a la operación del sistema real (Coss Bú, 2003).





En el ámbito epidemiológico la respuesta más común cuando se pregunta por el tamaño de la muestra en un estudio con regresión logística es la fórmula clásica de planteada por Freeman (1897) que indica en términos generales, el tamaño de muestra ha de ser por lo menos, unas diez veces el número de variables independientes a estimar más uno, es decir, considerando la variable dependiente (Ortega Calvo & Cayuela Domínguez, 2002), considerando las siete variables independientes del trabajo el tamaño de la muestra mínimo debería ser de 80 simulaciones.

El tamaño de una corrida de simulación depende principalmente del tipo de distribución que se intenta simular y, por decirlo de alguna forma, de la bondad del generador de números  $U(0,1)$  que se está utilizando y de las condiciones iniciales con que inició la simulación del sistema.

En el presente trabajo de investigación se toma un tamaño de muestra para la simulación de 200 (ver anexo B), de acuerdo a la fase de entrenamiento realizado para diferentes tamaños de muestra, en la cual se revisó rápidamente el estadístico de bondad de ajuste Prueba de Hosmer y Lemeshow y el R cuadrado de Nagelkerke eligiendo aquel con mayor probabilidad y  $R^2$ .

A continuación, en la tabla 12 se muestra un resumen del entrenamiento de la base de datos; en esta se observa que para una población simulación de 800 registros y una muestra de 200 con 10 corridas del modelo, se tienen el estadístico más elevado (0.994) y un  $R^2$  de 0.798 y se elige este.

**Tabla 12.***Entrenamiento y selección de tamaño de muestra.*

<i>Ensayo</i>	<i>Población simulada</i>	<i>Muestra</i>	<i>Corridas</i>	<i>Estadístico de Hosmer y Lemeshow para mejor modelo (sig.)</i>	<i>R cuadrado de Nagelkerke</i>
1	400	100	10	0.587	0.227
2	800	200	10	0.994	0.798
3	1200	400	10	0.824	0.064

**Fuente:** Elaboración propia.

Así mismo, una información proporcionada por los analistas de crédito, la Oficina de Puno accedió a un promedio de 200 préstamos de tipo Convenio, Pecuario y Personal, sin considerar los créditos a PYMES en el tercer trimestre 2019, por tanto, se trabajó con una muestra de 200 registros.

### **4.3. PREDICCIÓN DEL RIESGO DE MOROSIDAD A PARTIR DE LA BASE DE DATOS SIMULADA**

#### **4.3.1. Selección de variables**

Luego de la generación de la Base de Datos a través de la simulación y elegir el tamaño de muestra, se realizó la estimación inicial del modelo con siete variables predictoras, sin embargo, aplicando el método directo que brinda el SPSS por defecto, se determinó que no todas las variables son significativas.

**Tabla 13.**

*Variables que ingresan al modelo*

**Variables en la ecuación**

	B	Error estándar	Wald	Gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
Destino crédito			7,562	2	<b>,023</b>			
Destino crédito (1)	-5,267	2,286	5,306		,021	,005	,000	,456
Destino crédito (2)	-3,115	1,507	4,273		,039	,044	,002	,851
Tipo cliente	2,346	1,604	2,140		,144	10,445	,450	242,184
Días atraso			7,082		,132			
Días atraso (1)	-6,105	2,451	6,204		,013	,002	,000	,272
Días atraso (2)	-4,379	2,493	3,085		,079	,013	,000	1,661
Días atraso (3)	-8,176	4,019	4,138		,042	,000	,000	,742
Días atraso (4)	-25,090	15354,318	,000		,999	,000	,000	.
Total, ingreso	,000	,000	5,062		<b>,024</b>	1,000	1,000	1,000
Empleo			4,151		,656			
Empleo (1)	,877	2,308	,144		,704	2,404	,026	221,405
Empleo (2)	,270	1,898	,020		,887	1,310	,032	54,075
Empleo (3)	,787	1,942	,164		,685	2,197	,049	98,910
Empleo (4)	2,672	2,033	1,726		,189	14,466	,269	778,490
Empleo (5)	-10,176	40192,970	,000		1,000	,000	,000	.
Empleo (6)	4,433	3,147	1,984		,159	84,222	,176	40208,288
Sector procedencia	5,460	1,905	8,212		<b>,004</b>	235,149	5,616	9845,535
Estado civil			7,273		,122			



Estado civil (1)	-5,210	2,509	4,312	,038	,005	,000	,747
Estado civil (2)	-,976	2,183	,200	,655	,377	,005	27,186
Estado civil (3)	-,974	2,382	,167	,683	,378	,004	40,204
Estado civil (4)	24,595	19960,967	,000	,999	48003642310,0	,000	.
Constante	-2,589	3,707	,488	,485	,075		

a. Variables especificadas en el paso 1: Destino crédito, Tipo cliente, Dias atraso, Total ingreso, Empleo, Sector procedencia, Estado civil.

**Fuente:** Elaboración propia



En la tabla 13, se observa las variables significativas para el modelo:

- x1: Destino del crédito
- x4: Total ingreso
- x6: Sector de procedencia

Es importante mencionar que la variable x3: Días de atraso tiene dos categorías significativas, con lo que consideramos esta variable para el modelo.

En cuanto a las variables no significativas podemos decir que, el tipo de cliente no influye en el modelo, pues para la condición de moroso no es determinante que sea nuevo o recurrente el cliente; el tipo de empleo también no influye en el modelo, tal como lo muestra López Miranda, (2013), así como la variable estado civil.

En la tabla 14, muestra significancia de las variables x1, x3, x4 y x6 en el nuevo modelo y se observa claramente que son significativas, por tanto, influyen en el modelo. Cabe mencionar que se hicieron pruebas eliminando estas variables del modelo, sin embargo, su desempeño fue más pobre, por lo que se mantuvieron. También se puede incluir las estimaciones de interés de las ORs y cuya interpretación apoyaría al análisis; para una discusión más amplia ver (Álvarez, 1995).

**Tabla 14.**

*Variables que ingresan al nuevo modelo*

**Variables en la ecuación**

	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
						inferior	superior
Destino crédito		0,589		005			
Destino crédito (1)	1,827	855	,564	033	161	030	860
Destino crédito (2)	2,993	,106	,325	007	050	006	438
Días atraso		0,409		034			
Días atraso (1)	3,222	,121	,308	129	040	001	,546
Días atraso (2)	,631	,178	084	772	532	007	8,051
Días atraso (3)	3,364	,437	,905	167	035	000	,106
Días atraso (4)	21,069	7621,085	000	999	000	000	
Total, ingreso	000	000	,539	019	,000	,000	,000
Sector procedencia	,816	845	1,111	001	6,707	,190	7,487
Constante	1,436	,133	453	501	238		

a. Variables especificadas en el paso 1: Destino crédito, Días atraso, Total ingreso, Sector procedencia.

**Fuente:** Elaboración propia

En la tabla 15, se observa el coeficiente de determinación. Para nuestro modelo el coeficiente de determinación es de 54.2 por ciento. Es decir, un 54.2 por ciento de la variación de la variable morosidad (moroso o no moroso) es explicada por las variables incluidas en el modelo que son x1: Destino crédito, x3: Días de atraso de pago, x4: Total de ingreso y x6: Sector de procedencia.

**Tabla 15.**

*Prueba del coeficiente de determinación*

Resumen del modelo			
Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	75,341 <sup>a</sup>	,271	,542

a. La estimación ha terminado en el número de iteración 20 porque se ha alcanzado el máximo de iteraciones. La solución final no se puede encontrar.

**Fuente:** Elaboración propia

En la tabla 16, se observa la tabla de clasificación, para el modelo se obtiene una tasa de buena clasificación del 91.5 por ciento. En la tabla de clasificación se comprueba que el modelo tiene una especificidad alta (97.2%) y una sensibilidad de (45.5%). Con las variables predictoras: x1: Destino crédito, x3: Días de atraso de pago, x4: Total de ingreso y x6: Sector de procedencia, se comprueba que el modelo tiene una efectividad de 91.5 por ciento, es decir posee una buena capacidad predictora con las variables incluidas en el modelo. Hay que tomar en cuenta que el punto de corte es de 0.5.

**Tabla 16.**

*Tabla de clasificación*

	Observado	Pronosticado		Porcentaje correcto
		No moroso	Morosidad	
			Moroso	
Paso 1	Morosidad	173	5	97,2
		12	10	45,5
Porcentaje global				91,5

a. El valor de corte es ,500

**Fuente:** Elaboración propia

También de la tabla de clasificación podemos decir:

VP: La condición de que un cliente termine como moroso – Predicción correcta 45.5%.



VN: La condición de que un cliente termine como no moroso – Predicción correcta  
97.2%.

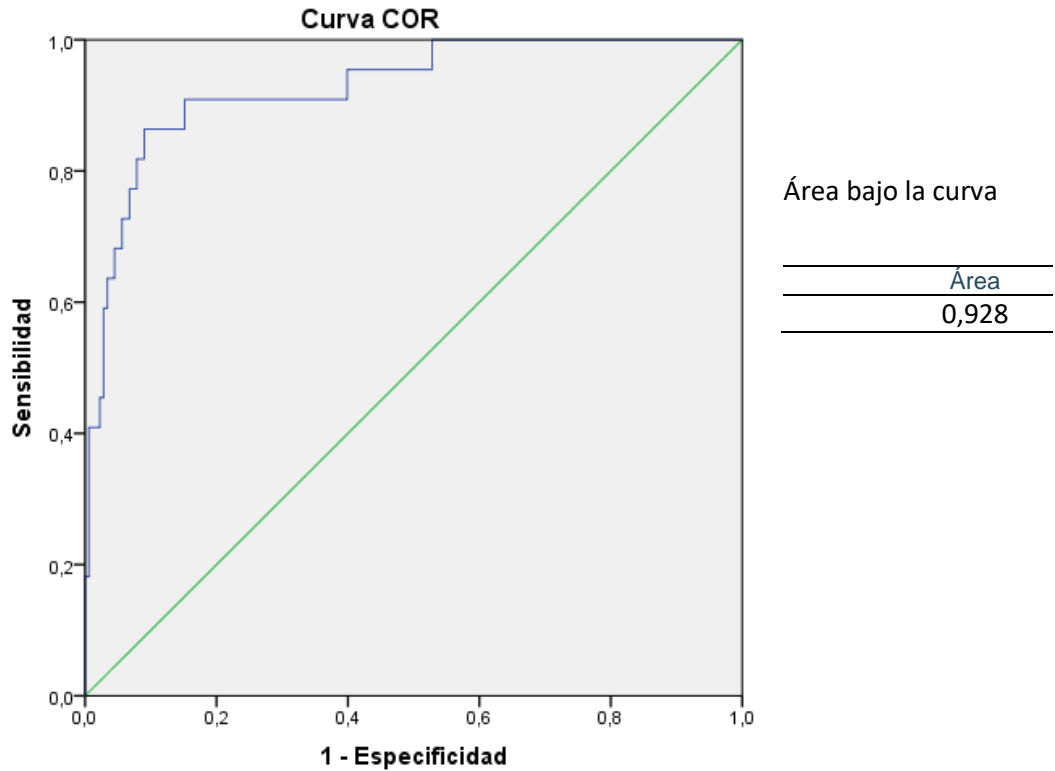
FP: La condición de que un cliente termine como no moroso – Predicción Incorrecta  
2.8%.

FN: La condición de que un cliente termine como moroso – Predicción Incorrecta  
54.5%.

Cabe mencionar que el Pc se eligió con base en la experiencia de la simulación, sin embargo, formalmente se puede seleccionar usando la llamada curva de características de operación o curva ROC (Receiver Operating Characteristic) curva ROC en español.

Para poder analizar el rendimiento (o la efectividad, en su más amplio sentido) de nuestro modelo clasificador, recurrimos a la curva COR, para lo cual tomamos los valores de la variable morosidad frente a los valores de probabilidad pronosticados. En la figura 1 se observa que nuestro modelo representado por la curva azul está por encima de la curva aleatoria cubriendo un área de 0,905.





**Figura 1.** Gráfica de la curva COR

**Fuente:** Rev. chil. infectol. vol.29

Si el área que encierra la Curva COR es 1, se tendría una clasificación perfecta y conforme nos alejamos de este valor la clasificación disminuye; entonces como la curva COR es 0.928 hace de este un buen modelo para clasificar a los clientes. Con punto de corte 0.5.

#### 4.3.2. Verosimilitud del modelo

La tabla 17, se observa la significación estadística con la prueba de Wald, la constante del modelo es significativa, por lo tanto, la constante se debe incluir en el modelo porque produce un cambio sobre la variable Morosidad (estado del cliente como no moroso o moroso).

**Tabla 17.***Prueba del coeficiente de intercepción*

		Variables en la ecuación					
		B	Error estándar	Wald	G1	Sig.	Exp(B)
Paso 0	Constante	-2,091	,226	85,588	1	,000	,124

**Fuente:** Elaboración propia.

Al evaluar la hipótesis nula ( $\beta_i = 0$ ) la significancia estadística asociada, y el valor de la OR( $\exp(B)$ ), las variables que contribuyen al modelo son x1: Destino crédito, x3: Días de atraso de pago, x4: Total de ingreso y x6: Sector de procedencia a un nivel de significación del 5 por ciento. Por tanto, el modelo estimado es el siguiente:

$$p(x) = \frac{1}{1 + e^{[1.436 + 1.827x_{11} + 2.993x_{12} + 3.222x_3 + 0.631x_{32} + 3.364x_{33} + 21.069x_{34} + 0.00x_4 - 2.816x_6]}}$$
$$\ln(odds) = -1.436 - 1.827x_{11} - 2.993x_{12} - 3.222x_3 - 0.631x_{32} - 3.364x_{33} - 21.069x_{34} - 0.00x_4 + 2.816x_6$$

#### 4.3.3. Prueba de bondad de ajuste

La tabla 18, se obtiene la prueba de bondad de ajuste del modelo, a un nivel de significación de 5 por ciento, el modelo es el adecuado, es decir el modelo de regresión logística se ajusta a los datos (p-valor=0.806).

**Tabla 18.***Prueba de bondad de ajuste***Prueba de Hosmer y Lemeshow**

Paso	Chi-cuadrado	gl	Sig.
	4,535	8	,806

*Fuente: Elaboración propia*

Así mismo, en la tabla 19, se tiene la prueba de Ómnibus:

**Tabla 19.***Prueba de Ómnibus*

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	63,265	8	,000
	Bloque	63,265	8	,000
	Modelo	63,265	8	,000

*Fuente: Elaboración propia.*

En la primera fila se contrasta la  $H_0$  de que los coeficientes de las variables añadidas son cero. Con la significancia de 0.000 se rechaza la  $H_0$ , es decir al menos uno de los coeficientes de las variables es diferente de cero.

Indica que el modelo con las variables pertenecientes al modelo tiene un ajuste significativo.

#### 4.3.4. Predicción del riesgo de morosidad

Si usamos los coeficientes estimados de la tabla 13, columna B, y los valores de las variables para cada caso en la base de datos simulada, podemos estimar la probabilidad o riesgo de morosidad de cada cliente mediante el modelo de regresión logística. Por ejemplo, en la tabla 20 se plantean dos casos (16 y 17 de la base de datos), cuyos valores de predicción de riesgo de morosidad son los siguientes:



**Tabla 20.**

*Predicción de riesgo de morosidad para dos casos.*

Variables	Caso	
	$Y_{16}$	$Y_{66}$
y: Morosidad	0	1
x1: Destino de crédito	0	2
x3: Días atraso	0	1
x4: Total ingreso	24236.59	36315.81
x6: Sector procedencia	0	1
Predicción	0.00676	0.95202

**Fuente:** Elaboración propia.

En el caso  $Y_{16}$  se representa datos de un cliente no moroso, cuyo destino del crédito fue para negocio, tiene días de retraso entre 0 y 8, tiene promedio de ingresos totales de S/ 24236.59 y proviene del sector urbano, el cual podríamos decir que es un perfil típico de un buen cliente. Observe que la probabilidad estimada por el modelo resultó menor que el  $P_c = 0.50$ , por lo que este primer caso es clasificado de manera correcta como cliente no moroso. Para el caso  $Y_{66}$ , que tiene antecedentes de morosidad y destino del crédito a vivienda es clasificado como moroso.

## V. CONCLUSIONES

### **PRIMERO:** DE OBJETIVO ESPECIFICO 1

Con el modelo de regresión logística se comprobó que existe relación causalidad entre  $x_1$ : Destino crédito,  $x_3$ : Días de atraso de pago,  $x_4$ : Total de ingreso y  $x_6$ : Sector de procedencia, para la predicción de la probabilidad de no recuperar un crédito de la Caja Rural de ahorro y crédito “Los Andes”, 2019, a un nivel de significación del 5 por ciento. Por tanto, el modelo estimado es el siguiente:

$$p(x) = \frac{1}{1 + e^{[1.436 + 1.827x_1 + 2.993x_2 + 3.222x_3 + 0.631x_4 + 3.364x_5 + 21.069x_6 + 0.00x_7 - 2.816x_8]}}$$

### **SEGUNDO:** DE OBJETIVO ESPECIFICO 2

La base de datos que se generó a través de la simulación permite realizar un adecuado entrenamiento y prueba del modelo, con un tamaño de muestra de 200 registros que presento una bondad de ajuste de Hosmer y Lemeshow con un valor de 0.994 y un R2 de 79.8% para la prueba del modelo.

### **TERCERO:** DE OBJETIVO GENERAL

Se mostró que el modelo de regresión logística para predecir la morosidad de un crédito de la Caja Rural de ahorro y crédito “Los Andes” se desempeña adecuadamente como un modelo clasificador a través de una base de datos simulada, cuya prueba de bondad de ajuste de Hosmer y Lemeshow, a un nivel de significación de 0.05, es de p-valor=0.806. La tasa de clasificación de verdaderos negativos (no morosos pronosticados como no morosos) fue del 97.2%; mientras que la tasa más relevante en este caso,



clasificar correctamente a los clientes morosos (verdaderos positivos) fue de 45%, demostrando gran capacidad predictiva del modelo.



## VI. RECOMENDACIONES

**PRIMERO:** Se recomienda hacer una investigación de tipo longitudinal de la situación moratoria de los préstamos financiados por las cajas de ahorro y crédito, con el fin de comparar los resultados históricos de la morosidad, vistos en esta investigación de tipo corte transversal que solo estudia a quienes son o no son morosos durante un momento dado en el tiempo.

**SEGUNDO:** Con referencia a los tipos de créditos otorgados (agrícola, grupal solidario, pecuario, Pyme, carta fianza, crédito convenio, etc.), se recomienda estimar modelos individuales, dado que los montos de crédito presentan alta variación por tipo de crédito y pueden variar respecto a las variables predictoras.



## VII. REFERENCIAS BIBLIOGRÁFICAS

- Agresti, A. (2007). *An introduction to categorical data analysis* (Second Edi). New Jersey: y John Wiley & Sons, Inc.
- Alderete, A. M. (2006). Fundamentos del Análisis de Regresión Logística en la Investigación Psicológica. *Online (Wilton, Connecticut)*, 6, 52–67. Retrieved from <https://doi.org/10.35670/1667-4545.v6.n1.534>
- Álvarez, R. (1995). *Estadística multivariante y no paramétrica con SPSS, Aplicaciones a las Ciencias de la Salud*. (D. de Santos, Ed.). Madrid - España.
- Armstrong, J. S. (2012). Illusions in regression analysis. *International Journal of Forecasting*, 28(3), 689–694. <https://doi.org/10.1016/j.ijforecast.2012.02.001>
- Azarang, M. R., & Garcia, E. (1996). *Simulación y análisis de modelos estocásticos*. México: McGRAW-HILL.
- Barzaga Sablón, O. S., Nevárez Barberán, J. V., Hidalgo Ávila, A. A., & Iloor Velez, D. L. (2018). Valoraciones entorno al riesgo financiero en las medianas empresas de la Provincia de Manabí. *Revista Lasallista de Investigación*, 15(2), 83–94. <https://doi.org/10.22507/rli.v15n2a6>
- Caja Los Andes. (2019). *Memoria Anual 2018*. Perú. Retrieved from <https://www.bvl.com.pe/hhii/J40630/20190311170201/MEMORIA32ANUAL32201832V5.1.PDF>
- Cano Calderon, M. M. (2018). *Modelo logístico para determinar las variables que inducen al abandono del tratamiento de Tuberculosis - RED de Salud Puno, 2017*. Universidad Nacional del Altiplano. Retrieved from





<http://repositorio.unap.edu.pe/handle/UNAP/7649>

- Cao Abad, R. (2002). *Introducción a la simulación y la teoría de colas* (Primera Ed).  
Coruña - España: NETBIBLO, S.L., A.
- Cardona, D. F., Gonzáles, J. L., Rivera, M., & Cárdenas, E. H. (2013). Aplicación de la  
regresión lineal en un problema de pobreza. *Revista Interacción*, 12, 73–84.
- Cermeño, R., León, J., & Mantilla, G. (2011). Determinantes de la morosidad: Un estudio  
panel para el caso de las cajas municipales de ahorro y crédito del Perú, 2003-2010.  
*Cide*, (NOVIEMBRE 2011), 2–41. Retrieved from <http://hdl.handle.net/11651/950>
- Chavarrín, R. (2015). Morosidad en el pago de créditos y rentabilidad de la banca  
comercial en México. *Banks, Latin America.*, 10(1), 73–85.
- Coss Bú, R. (2003). *Simulación: Un enfoque práctico*. (E. Noriega, Ed.) (Primera).  
México: EDITORIAL LIMUSA.
- Díaz Córdova, J. F., Coba Molina, E., & Navarrete, P. (2017). Lógica difusa y el riesgo  
financiero. Una propuesta de clasificación de riesgo financiero al sector cooperativo.  
*Contaduría y Administración*, 62(5), 1670–1686.  
<https://doi.org/10.1016/j.cya.2017.09.001>
- Fraga-Sampedro, M. L. (2018). Simulation as a learning tool for continuing education on  
cardiorespiratory arrest. *Enfermería Intensiva*, 29(2), 72–79.  
<https://doi.org/10.1016/j.enfi.2017.10.003>
- Gan@más. (2019). Microfinanzas. Retrieved September 20, 2011, from  
[https://revistaganamas.com.pe/morosidad-de-creditos-de-caja-los-andes-casi-se-  
duplico-en-un-ano/](https://revistaganamas.com.pe/morosidad-de-creditos-de-caja-los-andes-casi-se-duplico-en-un-ano/)



- García, A. (2019). Gestión de créditos y cobranzas. Retrieved from <https://www.esan.edu.pe/pee/areas/finanzas/gestion-de-creditos-y-cobranza/>
- García, E., García, H., & Cárdenas, L. E. (2013). *Simulación y análisis de sistemas con ProModel* (Segunda Ed). México: PEARSON.
- García, E., & Rialp, A. (n.d.). Tecnologías de la información y comunicación (TIC) y crecimiento de la empresa. *DAI*. Retrieved from <https://www.researchgate.net/publication/28202134>
- Guerra, J. A. (2018). *Análisis de gestión de riesgo para reducir el nivel de morosidad de los clientes del segmento pequeña y mediana empresa del Banco de Crédito del Perú de la ciudad de Juliaca periodo 2014-2015*. Universidad Nacional del Altiplano. Retrieved from <http://repositorio.unap.edu.pe/handle/UNAP/10628>
- Hosmer, J., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. (John Wiley, Ed.).
- Kelton, W. D., Sadowsky, R. P., & Sturrock, D. T. (2008). *Simulación con software Arena* (Cuarta Ed). México: MCGRAW-HILL.
- Leal Fica, A. L., Aranguiz Casanova, M. A., & Gallegos Mardones, J. (2017). Análisis de riesgo crediticio, propuesta del modelo credit scoring. *Revista Facultad de Ciencias Económicas*, 26(1), 181–207. <https://doi.org/10.18359/rfce.2666>
- López Miranda, C. A. (2013). Modelo predictivo de riesgo de morosidad para créditos bancarios usando datos simulados. *IEPISTEMUS: Investigación*, (2007–4530), 22–30. Retrieved from <https://s3.amazonaws.com/academia.edu.documents/35240840/14->



MODELO\_PREDICTIVO.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3  
A&Expires=1535831811&Signature=tbLt%2BXY7cwVAQ4d4czHfOq2yy18%3D  
&response-content-disposition=inline%3B  
filename%3DMODELO\_PREDICTIVO\_DE\_RI

Martín, S. de P. (2016). *El uso de las TICs en la Gestión Empresarial*. Universidad de Valladolid. Retrieved from <http://uvadoc.uva.es/bitstream/10324/23407/1/TFG-O967.pdf>

Matos Falvy, C. C. (2017). El riesgo crediticio hipotecario en las cajas municipales de ahorro y credito aplicando Credit Scoring y Var. *Universidad Nacional Agraria La Molina*. Retrieved from <http://repositorio.lamolina.edu.pe/handle/UNALM/3236>

Meza, E. (2017). *Evaluación del riesgo crediticio, a través de Credit Scoring mediante Regresión Logística: Un caso de estudio*. Benemérita Universidad Autónoma de Puebla.

Mures Quintana, M. J., García Gallego, A., & Vallejo Pascual, M. E. (2005). Aplicación del análisis discriminante y regresión logística en el estudio de la morosidad en las entidades financieras : comparación de resultados. *Pecunia : Revista de La Facultad de Ciencias Económicas y Empresariales, Universidad de León*, 1(1), 175. <https://doi.org/10.18002/pec.v0i1.746>

Ortega Calvo, M., & Cayuela Domínguez, A. (2002). Regresion logística no condicionada y tamaño de muestra: Revisión bibliográfica. *Revista Española de Salud Pública*, 76(2), 85–93. Retrieved from [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1135-57272002000200002&lng=es&nrm=iso&tlng=es](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1135-57272002000200002&lng=es&nrm=iso&tlng=es)



- Pally, U. (2016). *Análisis de la gestión del riesgo crediticio y su incidencia en el nivel de morosidad de la caja municipal de ahorro y crédito de Piura agencia Juliaca periodo 2013 – 2014*. Universidad Nacional del Altiplano. Retrieved from <http://repositorio.unap.edu.pe/handle/UNAP/1551>
- Pantoja, P. M. (2016). *Propuesta de un Modelo Logit para evaluar el Riesgo Crediticio en las Cajas Municipales de Ahorro y Crédito : Caso de la Caja Municipal de Huancayo, periodo 2011-2015*. Universidad San Ignacio de Loyola.
- Parra, F. (2016). *Curso de estadística con R*. Santander, Cantabria: ICANE.
- Peña, D. (2002). *Análisis de datos multivariantes*. Madrid - España: University Carlos III de Madrid.
- Peñas, J. S. A., Muñoz, A., Juez, P., & Cortinas, P. (2003). *Diseño de encuestas para estudios de mercado. Técnicas de Muestreo y Análisis Multivariante*. Centro de Estudios Ramón Areces.
- Perez, C. (2004). *Técnicas de Análisis Multivariantes de Datos, Aplicaciones con SPSS*. (P. Hall, Ed.). Madrid - España: PEARSON PRENTICE HALL.
- Ríos, K. D. (2017). *Regresión logística de los factores que influyen en la gastritis, hospital reional Manuel Nuñez Butrón - Puno*. Universidades Nacional del Altiplano.
- Sanchez Nina, A. (2017). *Modelo estadístico para determinar la deserción estudiantil de las escuelas profesionales de la UNA-Puno, 2017*. Universidad Nacional del Altiplano. Retrieved from <http://repositorio.unap.edu.pe/handle/UNAP/6297>
- SBS. (2019). Educación financiera. Retrieved from <https://www.sbs.gob.pe/educacion-financiera>



- Silva Ayçaguer, L. C. (2005). Una ceremonia estadística para identificar factores de riesgo. *Salud Colectiva*, 1(3), 309. <https://doi.org/10.18294/sc.2005.49>
- Trejo García, J. C., Martínez García, M. Á., & Venegas Martínez, F. (2017). Administración del riesgo crediticio al menudeo en México: una mejora econométrica en la selección de variables y cambios en sus características. *Contaduria y Administracion*, 62(2), 399–418. <https://doi.org/10.1016/j.cya.2017.02.006>
- Tusell, F. (2011). *Análisis de Regresión. Introducción Teórica y Práctica basada en R*. Bilbao.
- Valencia-Cárdenas, M., & Restrepo-Morales, J. A. (2016). Evaluación de la gestión financiera usando variables latentes en modelos estocásticos de fronteras eficientes. *DYNA (Colombia)*, 83(199), 35–40. <https://doi.org/10.15446/dyna.v83n199.54612>
- Visauta, B. (1998). *Análisis estadístico con SPSS para Windows. Estadística multivariante*. España: McGRAW-HILL.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias* (Novena Edi). México: PEARSON EDUCACIÓN.
- Wayne, D. (2006). *Bioestadística. Base para el análisis de las ciencias de la salud* (Cuarta Edi). España: LIMUSA Wiley.



## ANEXOS



## ANEXO A: Ficha de recolección de datos

	Universidad Nacional del Altiplano Facultad de Ingeniería Estadística e Informática E.P. de Ingeniería Estadística e Informática	<b>FICHA:</b> <div style="border: 1px solid black; padding: 2px; width: 150px; float: right;">N°</div>																
<b>FICHA DE RECOJO DE INFORMACIÓN</b>																		
<b>DATOS DEL SOLICITANTE</b>																		
<b>Nombre:</b> _____ <b>Estado Civil:</b> <input type="checkbox"/> Soltero(a) <input type="checkbox"/> Casado(a) <input type="checkbox"/> <input type="checkbox"/> Conviviente <input type="checkbox"/> Viudo(a)		<b>Fecha Nacimiento:</b> ___/___/___ <b>Lugar:</b> _____ <b>Vivienda:</b> <input type="checkbox"/> Propia <input type="checkbox"/> Alquilada <input type="checkbox"/> Otro <b>Carga familiar:</b> <input type="checkbox"/> hijos <b>Edades:</b> _____ <b>Residencia:</b> <input type="checkbox"/> Puno ciudad <input type="checkbox"/> Provincia <b>Labora desde:</b> _____																
<b>Empleo</b> <input type="checkbox"/> Independiente <input type="checkbox"/> Privado <input type="checkbox"/> Eventual <input type="checkbox"/> Estatal <input type="checkbox"/> Desempleado																		
<b>INFORMACIÓN PATRIMONIAL</b>																		
<b>Propiedades inmuebles</b>		<b>Dirección:</b> _____																
		<b>Valor:</b> _____																
1) _____		S/ _____																
2) _____		S/ _____																
3) _____		S/ _____																
<b>Vehículo - Marca</b>		<b>Modelo</b>																
		<b>Año</b>																
		<b>Placa</b>																
1) _____		S/ _____																
2) _____		S/ _____																
3) _____		S/ _____																
<input type="checkbox"/> TV <input type="checkbox"/> DVD <input type="checkbox"/> Equipo de sonido <input type="checkbox"/> Refrigeradora <input type="checkbox"/> Lavadora <input type="checkbox"/> Maq. cocer <input type="checkbox"/> Muebles de hogar <input type="checkbox"/> Otros																		
<b>PRESUPUESTO FAMILIAR</b>																		
<b>I. Ingresos</b>		<b>II. Egresos</b>																
1) Ingreso solicitante	S/ _____	1) Alimentación																
2) Ingreso conyuge	S/ _____	2) Educación																
3) Otros ingresos	S/ _____	3) Servicios																
<b>TOTAL INGRESOS</b>	S/ _____	4) Otros egresos																
		<b>TOTAL EGRESOS</b>																
		S/ _____																
<b>HISTORIAL DE CRÉDITOS</b>																		
Analista	N° de Pagaré	Fecha de desembolso	Monto de desembolso	Plazo y forma de pago	Cancelación de Cuotas												Calif. de crédito	
					1	2	3	4	5	6	7	8	9	10	11	12		
1																		
2																		
3																		
4																		
5																		
6																		
7																		
8																		
9																		
10																		
<b>OBSERVACIONES</b>																		
_____ Firma y sello del analista																		



**ANEXO B:** Base de datos para estimar los parámetros del modelo de riesgo de morosidad.

N	y	x1	x2	x3	x4	x5	x6	x7		N	y	x1	x2	x3	x4	x5	x6	x7
1	0	2	1	2	3860,26	1	1	1		51	0	2	1	0	15569,17	2	0	0
2	0	1	1	0	15205	1	0	1		52	0	0	1	0	10030,07	3	0	1
3	0	0	0	0	20152,99	2	0	0		53	0	1	1	0	8923,65	2	0	0
4	0	2	0	0	12715,15	1	1	0		54	0	0	1	0	11446,83	2	1	1
5	0	1	0	2	5429,68	1	0	0		55	0	2	1	0	27707,12	2	0	0
6	0	2	0	0	11089,94	2	1	1		56	0	0	0	0	14580,31	2	0	0
7	0	2	1	0	19952,67	1	0	4		57	0	0	0	0	31821,81	3	0	0
8	0	0	1	0	15424,55	3	1	2		58	0	1	1	1	1818,7	4	1	1
9	0	2	1	0	21632,25	1	0	0		59	0	0	0	0	16532,22	1	0	0
10	0	2	1	0	9609,87	2	1	0		60	0	1	1	0	30557,14	2	1	0
11	0	0	1	0	10837,64	5	0	1		61	0	1	1	0	12340,41	3	1	0
12	0	0	1	0	41080,68	3	0	2		62	0	1	0	0	12464,45	1	1	1
13	0	1	0	0	1243,57	1	1	1		63	0	2	1	0	1232,44	1	0	0
14	0	2	0	4	7420,92	1	1	0		64	0	2	0	0	1241,19	3	1	0
15	0	1	1	0	26470,3	3	0	0		65	0	1	0	0	12358,42	3	1	1
16	0	0	1	0	24236,59	3	0	0		66	1	2	0	1	36315,81	3	1	1
17	1	2	0	0	11416,78	4	1	0		67	0	0	1	0	21685,34	2	1	4
18	1	0	1	4	16870,4	7	0	1		68	0	0	1	0	25306,51	2	0	0
19	0	2	1	0	21083,18	4	1	0		69	0	2	1	0	11822,42	2	0	0
20	0	0	1	0	26792,41	4	0	1		70	0	1	0	0	1867,4	4	0	1
21	0	0	1	0	6627,63	3	0	2		71	1	2	1	0	23972,44	2	1	2
22	0	2	0	0	3188	2	0	0		72	0	2	1	0	14562,6	8	0	0
23	0	1	0	0	11129,88	3	1	0		73	0	2	0	0	11622,05	4	0	0
24	0	0	0	0	3996,6	4	1	1		74	0	0	1	0	4501,25	3	1	0
25	0	1	1	0	17818,52	1	0	0		75	0	0	1	0	7696,67	3	1	1
26	0	2	0	0	8549,6	2	0	0		76	0	1	0	0	14726,14	3	1	2
27	0	2	1	0	19383,99	4	0	1		77	0	1	0	0	17343,73	2	0	0
28	1	2	1	0	4993,57	1	0	3		78	0	2	0	0	2671,23	3	0	0
29	0	0	0	0	26810,83	1	0	1		79	0	1	0	3	3986,08	1	0	1
30	0	0	1	0	18250,93	2	0	2		80	0	0	0	0	15809,23	1	0	0
31	0	2	1	0	9971,82	3	0	2		81	0	1	1	0	8046,91	2	1	1
32	0	0	1	0	3204,24	2	0	1		82	0	0	1	0	3133,71	2	0	0
33	0	2	1	0	14859,2	1	1	0		83	0	1	0	0	15886,61	1	0	1
34	0	2	0	0	5756,07	8	0	0		84	0	0	1	0	20865,3	8	0	1
35	0	2	1	0	38570,43	1	0	0		85	0	0	1	0	7975,4	1	1	0
36	0	1	0	0	7472,07	1	0	0		86	0	1	0	0	3338,17	3	0	2
37	0	1	1	0	3860,65	1	0	4		87	0	1	1	0	5666,49	3	0	0
38	0	2	1	0	23013,27	1	1	0		88	0	0	0	0	15867,76	3	0	0
39	0	1	1	0	9646,5	2	1	2		89	0	2	0	0	17808,97	3	1	0
40	0	2	1	0	11137,16	3	0	0		90	0	0	1	4	1199,09	8	0	2
41	0	0	1	0	5451,47	4	0	2		91	0	1	1	2	41951,64	4	0	0
42	0	1	1	0	40249,07	3	1	0		92	1	2	1	1	30640,2	8	1	2
43	0	0	0	0	8972,3	3	0	2		93	1	2	1	0	22170,59	1	1	0
44	0	1	1	0	8229,99	4	1	0		94	0	0	0	0	3735,46	2	0	2
45	0	1	1	0	28797,44	2	1	0		95	0	0	0	0	17264,98	4	0	2
46	0	0	1	0	17318,06	1	1	1		96	0	1	1	0	19037,01	1	1	0
47	0	2	0	0	18013,41	1	0	0		97	0	0	1	0	16550,35	1	0	0
48	0	2	0	0	1665,89	3	1	2		98	0	0	0	0	22339,92	2	0	1
49	0	1	1	0	4051,25	1	0	2		99	0	0	1	0	22692,09	4	1	1
50	0	0	1	1	20995,11	1	0	0		100	0	1	0	3	42596,43	1	1	2





N	y	x1	x2	x3	x4	x5	x6	x7	N	y	x1	x2	x3	x4	x5	x6	x7
101	0	1	1	2	16764,2	1	1	0	151	0	1	0	3	19426,36	4	0	0
102	0	0	1	0	11565,54	3	0	1	152	0	2	0	0	6608,11	4	0	4
103	0	0	0	0	4152,5	2	0	1	153	0	0	0	0	22697,87	1	0	1
104	0	1	0	1	10980,41	3	1	1	154	1	2	1	1	20862,21	4	0	1
105	0	1	0	0	32189,22	3	1	0	155	0	1	0	0	12176,85	3	0	2
106	0	1	0	0	13565,79	1	1	2	156	0	0	1	0	8447,13	4	1	0
107	0	2	1	0	1328,12	2	0	0	157	0	0	1	0	27509,8	1	0	0
108	0	1	1	0	28393,55	8	0	1	158	1	1	1	0	14451,43	3	1	2
109	0	2	1	0	15502,78	8	1	2	159	0	1	1	0	12072,05	1	0	0
110	0	1	1	0	17846,88	3	1	0	160	0	1	0	0	22055,33	2	1	1
111	0	0	0	0	19945,36	7	0	1	161	0	1	0	0	8794,97	4	0	0
112	1	0	1	1	14186,49	3	1	3	162	0	0	1	0	19169,28	4	0	0
113	1	2	1	1	2351,89	2	1	1	163	1	2	1	2	25902,31	4	1	1
114	0	1	1	0	23287,67	2	0	0	164	0	2	0	0	10746,52	1	0	2
115	0	2	1	0	4935,26	4	0	4	165	1	2	1	0	47698,71	3	1	1
116	0	2	1	1	34737,85	7	0	0	166	0	1	0	0	34315,37	1	1	0
117	0	0	1	0	21813,78	1	1	1	167	0	2	0	2	23325,03	2	1	0
118	0	0	1	0	26008,23	1	0	0	168	0	0	1	0	36346,52	2	0	0
119	0	1	0	0	13581,79	8	1	0	169	0	1	0	0	8593,39	1	1	2
120	1	2	1	0	42206,01	4	1	0	170	0	2	0	0	14028,92	1	0	1
121	0	0	0	0	23407,33	2	1	0	171	1	2	1	1	28015,26	3	1	2
122	0	1	1	0	18352,7	3	0	2	172	0	0	1	0	12789,4	2	1	2
123	0	0	0	0	23289,66	1	1	0	173	0	2	1	0	2937,56	3	0	0
124	0	0	1	0	2045,89	3	0	0	174	0	2	1	0	7787,62	1	0	0
125	0	2	1	0	18780,95	4	0	1	175	0	1	0	0	9814,69	3	0	0
126	0	0	0	0	18243,71	2	0	1	176	0	1	1	0	18391,85	2	0	2
127	1	2	1	0	20560,81	7	1	2	177	0	1	0	0	892,89	3	0	1
128	0	0	0	0	26012,35	1	1	0	178	0	2	1	0	11090,36	4	1	0
129	0	2	0	0	20430,49	1	0	1	179	1	2	1	0	32266,16	3	1	2
130	0	0	0	0	27312,35	8	0	2	180	0	2	0	3	32309,4	4	0	0
131	1	2	1	1	23478,45	1	1	3	181	0	0	1	0	16883,76	3	0	1
132	0	1	0	0	8716,91	1	0	0	182	0	2	1	0	36014,96	2	1	0
133	1	2	1	0	10342,01	4	1	1	183	1	2	1	0	14533,19	2	1	4
134	0	0	1	0	2513,16	2	0	0	184	0	2	0	0	10586,15	2	0	2
135	1	2	0	0	39835,7	3	1	2	185	0	1	0	0	11162,29	3	0	2
136	0	0	0	0	3998,34	1	0	0	186	0	1	1	0	12101,9	1	0	0
137	0	0	1	0	20587,29	1	0	0	187	0	2	1	1	41015,32	2	0	0
138	0	2	0	0	20213,88	3	1	2	188	0	0	0	2	12868,78	1	1	0
139	0	1	1	0	19361,06	1	0	1	189	0	2	0	2	1525,65	3	0	1
140	0	2	1	0	20897,72	1	0	0	190	0	2	0	0	16606	1	0	0
141	0	1	1	1	23963,09	3	0	1	191	0	1	0	0	34642,75	1	0	0
142	0	1	0	0	24203,34	1	0	1	192	0	1	1	0	4866,55	4	1	1
143	0	2	1	1	5326,63	4	0	0	193	0	2	1	0	2397,85	1	0	0
144	0	2	1	0	12728,05	1	1	0	194	0	1	0	0	13450,69	1	0	2
145	0	2	1	0	10308,54	3	0	2	195	0	2	0	0	6887,45	2	1	2
146	0	1	0	0	7582,96	2	1	1	196	0	2	1	1	6855,43	8	0	1
147	0	1	1	0	11409,84	2	0	0	197	0	1	1	0	22547,73	1	1	0
148	0	0	1	0	22841,71	1	1	0	198	0	2	1	0	35110,52	3	1	0
149	0	2	1	0	9854,03	7	0	0	199	0	0	1	0	13025,33	3	0	0
150	0	2	0	0	9157,21	8	1	0	200	1	2	1	0	39592,88	8	1	0