

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



**CLASIFICACIÓN DEL TIPO DE CLIENTE EN FUNCIÓN A LAS
CARACTERÍSTICAS SOCIO ECONÓMICAS DEL BUEN
PAGADOR DE LA INSTITUCIÓN FINANCIERA
COMPARTAMOS – AGENCIA, PUNO 2017**

TESIS

PRESENTADA POR:

Bach. HORACIO AJAHUANA CONDORI

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO – PERÚ

2018

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

**CLASIFICACION DEL TIPO DE CLIENTE EN FUNCION A LAS
CARACTERISTICAS SOCIO ECONOMICAS DEL BUEN PAGADOR DE LA
INSTITUCION FINANCIERA COMPARTAMOS-AGENCIA PUNO 2017**

TESIS PRESENTADA POR:

Bach. HORACIO AJAHUANA CONDORI

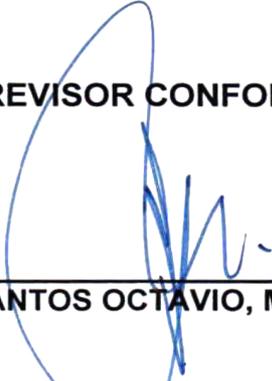


PARA OPTAR EL TITULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

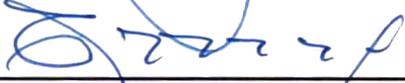
APROBADA POR EL JURADO REVISOR CONFORMADO POR:

PRESIDENTE:



M.C. SANTOS OCTAVIO, MORILLOS VALDERRAMA

PRIMER MIEMBRO:



M.C. CONFESOR MILAN, VARGAS VALVERDE

SEGUNDO MIEMBRO:



M.Sc. SAMUEL DONATO, PEREZ QUISPE

DIRECTOR / ASESOR:



Dr. EDGAR ELOY, CARPIO VARGAS

Área : Estadística

Tema : Modelos Predictivos

Fecha de sustentación: 19/10/2018

DEDICATORIA

A Dios por brindarme la vida, permitirme compartirla con las personas que más quiero y guiar mi camino.

A mi madre Ángela por su amor, fortaleza y confianza hacen la diferencia siempre exiges más de mi porque sabes que puedo hacerlo, a mi Padre Gregorio por sus palabras y aliento, su persona es el mejor ejemplo que un hijo podría tener.

A mis hijas Alison Milagros y Dania Valentina porque son el motor y motivo en vida que me inspiran a ser mejor persona.

Horacio Ajahuana Condori.

AGRADECIMIENTO

A la Facultad de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano.

A todos los docentes de la Facultad de Ingeniería Estadística e Informática por sus enseñanzas durante todos los años de formación universitaria.

A la Empresa Compartamos Financiera S.A. por brindarme todas las facilidades para la recopilación de información para lograr desarrollar el presente trabajo de Investigación.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS	7
ÍNDICE DE TABLAS	8
ÍNDICE DE ACRÓNIMOS	9
RESUMEN	10
ABSTRACT	11
CAPÍTULO I	
INTRODUCCIÓN	12
1.1 EL PROBLEMA DE INVESTIGACIÓN	12
1.2 DEFINICIÓN DEL PROBLEMA	14
1.3 OBJETIVO GENERAL	14
1.4 OBJETIVOS ESPECÍFICOS	14
CAPÍTULO II	
REVISIÓN DE LITERATURA	15
2.1 ANTECEDENTES INTERNACIONALES	15
2.2 MARCO TEÓRICO	22
2.2.1 El riesgo en las entidades financieras.....	22
2.2.2 Riesgo de crédito.....	25
2.2.3 Análisis Discriminante.....	28
2.2.4 Tipos de análisis discriminante	29
2.2.5 Utilidad del Análisis Discriminante.....	30
2.2.6 Asunciones (supuestos) del análisis discriminante,.....	31
2.2.7 El problema de la clasificación	33
2.2.8 Clasificación o discriminación con 2 grupos.	38
2.2.9 Discriminante k grupos	41
2.2.10 Cálculo de las funciones discriminantes (extracción de las funciones)	42
2.2.11 Correlación canónica y autovalores asociados a una función discriminante. 46	
2.2.12 Determinación del número de funciones discriminantes.....	48
2.2.13 Interpretación de los resultados	49
CAPÍTULO III	
MATERIALES Y MÉTODOS	50
3.1 ENFOQUE DE LA INVESTIGACIÓN:	50
3.2 DISEÑO DE INVESTIGACIÓN.	50
3.3 TIPO DE INVESTIGACIÓN.	50

3.4 POBLACIÓN.....	50
3.5 MUESTRA.	50
3.6 INSTRUMENTO DE INVESTIGACIÓN.....	51
3.7 METODOLOGÍA ESTADÍSTICA.	51

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN	55
RECOMENDACIONES	80
REFERENCIAS BIBLIOGRAFICAS	81
ANEXOS.....	84
Anexo A. Script en R para la normalidad.....	84
Anexo B. Matriz dentro de grupos combinados	85
Anexo C. Matriz de covarianzas.....	87
Anexo D. Estadísticas por casos	90
Anexo E. Comparación de resultados observados y predichos con el modelo.....	95

ÍNDICE DE FIGURAS

Figura 01. Utilización de un punto de corte equidistante de ambos centroides $n_1 = n_2$	34
Figura 02. Utilización de un punto de corte equidistante de ambos centroides $n_1 \neq n_2$	35
Figura 03. Probabilidad condicional de la puntuación discriminante d_i en el grupo 2...	36
Figura 04. Probabilidad condicional de la puntuación discriminante d_i en el grupo 2...	39
Figura 05. Histograma de cada grupo y centroides representados sobre la función discriminante.....	40
Figura 06. Mapa territorial.....	72
Figura. 07 Diagrama de dispersión de las funciones discriminantes canónicas del segmento A+.....	74
Figura 08. Diagrama de dispersión de las funciones discriminantes canónicas del segmento A.....	75
Figura 09. Diagrama de dispersión de las funciones discriminantes canónicas del segmento B.....	75
Figura 10. Diagrama de dispersión de las funciones discriminantes canónicas de los segmentos A+,A y B.....	76

ÍNDICE DE TABLAS

Tabla 01. Características socio económicas descriptivas del buen pagador de la Institución Financiera Compartamos - Agencia Puno 2017.	57
Tabla 02. características socio económicas descriptivas por segmento del buen pagador de la Institución Financiera Compartamos - Agencia Puno 2017.....	59
Tabla 03. Pruebas de igualdad de las medias de los grupos.....	61
Tabla 04. Logaritmos de los determinantes de los diferentes segmentos	62
Tabla 05. Tabla de resultados de la prueba M de Box.....	63
Tabla 06. Autovalores de las funciones canónicas discriminantes	64
Tabla 07. Lambda de Wilks. Contraste de las funciones del modelo.	64
Tabla 08. Valor de los centroides en las funciones discriminantes.....	65
Tabla 09. Coeficientes de la función discriminante canónica.....	66
Tabla 10. Coeficientes estandarizados de las funciones discriminantes canónicas.	67
Tabla 11. Matriz de estructuras.....	69
Tabla 12. Probabilidades previas (probabilidades a priori utilizadas en la clasificación).	70
Tabla 13. Cambio entre valores observados y reclasificados	76

ÍNDICE DE ACRÓNIMOS

SBS: Superintendencia de banca y seguros

IF: Institución Financiera

RC: Riesgo de Crédito

P: probabilidad

AD: Análisis discriminante

RESUMEN

Las entidades financieras ofrecen oportunidades de préstamos de inclusión al mayor número de personas en el menor tiempo posible y es de necesidad evaluar a los prestamistas para reducir el riesgo de moras y no pagos de cuotas que finalmente afectan a los clientes y a las empresas. El objetivo fue determinar la mejor relación para clasificar el tipo de cliente en función a las características socio económicas del buen pagador de la Institución Financiera Compartamos - Agencia Puno. El Diseño de investigación planteado fue el diseño descriptivo, y el tipo de investigación, descriptivo correlacional. La muestra estuvo conformada por 97 clientes y la técnica estadística para el análisis de datos fue el análisis discriminante, llegando a las siguientes conclusiones: Existen diferencias en las características socio económicas entre los segmentos de cliente muy bueno, cliente bueno y cliente regular, destacando que el atraso máximo de pago en días y atraso actual se encuentra en el segmento A+. Los clientes tienen préstamos en más de una entidad financiera, los ingresos y egresos se distribuyen de acuerdo al segmento La primera función explica el 75.2 % de la variabilidad disponible en los datos, mientras que la segunda función sólo explica el 24.8%.

$$D1 = -0.311X_1 + 0.549X_2 + 0.036X_3 - 0.146X_4 + 0.548X_5 - 1.459X_6 + 0.982X_7 \\ - 0.035X_8 + 1.064X_9 - 0.832X_{10} + 0.579X_{11}$$

$$D2 = 0.285X_1 - 0.005X_2 + 0.868X_3 + 0.272X_4 - 0.87X_5 - 0.072X_6 - 0.283X_7 \\ - 0.289X_8 + 0.310X_9 - 0.987X_{10} + 0.333X_{11}$$

El modelo ha permitido reclasificar el 44% de los clientes a otras categorías.

Palabras Clave: discriminante, buen pagador, riesgo, institución financiera

ABSTRACT

Financial institutions offer opportunities for inclusion loans as many people in the shortest possible time and prestige are translated into lenders to reduce the risk of default and non-payment of fees that are finally found in customers and companies. The objective was to determine the best relationship to classify the type of client based on the socioeconomic characteristics of the good payer of the Shared Financial Institution - Agencia Puno. The design of the research was descriptive design, and the type of research, descriptive correlational. The sample consisted of 97 clients and the statistical technique for data analysis and discriminant analysis, reaching the following questions: differences in socio-economic characteristics between the very good client, good customer and regular customer segments, highlighting that maximum delay of payment in days and current delay is in segment A +. The clients have to do with a financial entity, the income and the expenses are distributed in an agreement in the segment. The first function explains 75.2% of the variability available in the data, while the second function only explains 24.8%.

$$D1 = -0.311X_1 + 0.549X_2 + 0.036X_3 - 0.146X_4 + 0.548X_5 - 1.459X_6 + 0.982X_7 \\ - 0.035X_8 + 1.064X_9 - 0.832X_{10} + 0.579X_{11}$$

$$D2 = 0.285X_1 - 0.005X_2 + 0.868X_3 + 0.272X_4 - 0.87X_5 - 0.072X_6 - 0.283X_7 \\ - 0.289X_8 + 0.310X_9 - 0.987X_{10} + 0.333X_{11}$$

The model has allowed reclassifying 44% of customers to other categories.

Key Words: discriminant, good payer, risk, financial institution

CAPÍTULO I

INTRODUCCIÓN

1.1 EL PROBLEMA DE INVESTIGACIÓN

Todas las financieras como también la Financiera Compartamos tienen como propósito erradicar la exclusión financiera, ofreciendo oportunidades de inclusión al mayor número de personas en el menor tiempo posible, compartiendo beneficios, construyendo modelos comerciales innovadores, eficientes y rentables, beneficiando a todos los clientes, confiando en todas las personas, en su disposición para crecer y autorrealizarse. Poseer educación financiera, es otra forma de crecer y autorrealizarse, con la educación financiera es posible utilizar servicios financieros para su beneficio personal y familiar.

El riesgo que corren las Instituciones Financieras al otorgar un crédito, se ha vuelto una labor de suma importancia para su buen funcionamiento, este riesgo denominado riesgo de crédito, está asociado con la probabilidad de pago de un cliente al que se le otorga un crédito. El riesgo de crédito es importante en la medida en que gran parte de la solvencia y/o rentabilidad de las instituciones del sistema financiero dependen del pago de los créditos otorgados por los distintos clientes, muchos argumentan que este incremento es debido a la mala administración de este tipo de riesgo puesto que una gran cantidad de préstamos fueron otorgados a personas que no cumplían con los requerimientos mínimos de solvencia para el cumplimiento de sus obligaciones, por lo tanto, los

académicos han venido aplicando y desarrollando nuevas técnicas que permitan determinar el comportamiento de un deudor frente al pago de sus créditos. El problema de estimación del riesgo de crédito puede verse como un problema de clasificación de clientes entre los que pagan y los que no pagan, por lo tanto, se propone el uso del Análisis Discrimínate como técnica para analizar este problema. (Vela, 2012) por su parte indica que la calidad de la cartera está explicada por un conjunto de variables de carácter macroeconómico relacionados con el ciclo del producto y microeconómico tales como el nivel de solvencia de las entidades, la eficiencia y gestión de sus costos operativos, la tasa de crecimiento de los activos rentables.

La actividad económica se desarrolla en un ambiente de incertidumbre, convirtiendo el riesgo en un factor inherente a la misma. De este modo, surge aquél como la contingencia, probabilidad o proximidad de un daño o peligro, en concreto, de sufrir una pérdida. La incertidumbre, junto con la aleatoriedad, constituyen las características principales del riesgo, añadiéndose como tal el conflicto, ya que el riesgo se presenta ante situaciones diferenciadas.

Así pues, el riesgo se ha convertido en uno de los rasgos básicos del entorno económico actual al que se enfrentan las empresas. En el caso de las entidades financieras, esta característica es esencial a la actividad que desarrollan, consistente en la concesión de créditos, asumiendo un riesgo cuando prestan unos recursos financieros que otros clientes les han cedido, sin controlar posteriormente el destino y utilización de los mismos.

1.2 DEFINICIÓN DEL PROBLEMA

¿Cuál es la mejor relación para clasificar el tipo de cliente en función a las características socio económicas del buen pagador de la Institución Financiera Compartamos - Agencia Puno 2017?

1.3 OBJETIVO GENERAL

Determinar la mejor relación para clasificar el tipo de cliente en función a las características socio económicas del buen pagador de la Institución Financiera Compartamos - Agencia Puno 2017.

1.4 OBJETIVOS ESPECÍFICOS

1. Seleccionar las características socio económicas para clasificar el tipo de cliente buen pagador de la Institución Financiera Compartamos - Agencia Puno.
2. Seleccionar de la función discriminante que clasifique el tipo de cliente buen pagador de la Institución Financiera Compartamos - Agencia Puno.
3. Clasificar el grupo de pertenencia de los denominados como buen pagador de la Institución Financiera Compartamos - Agencia Puno.

CAPÍTULO II

REVISIÓN DE LITERATURA

2.1 ANTECEDENTES INTERNACIONALES

Fontalvo, Morelos y De la Hoz (2010), utiliza el Análisis Discriminante para evaluar el comportamiento de los indicadores financieros en empresas de sector carbón concluyendo que con el uso del Análisis Discriminante el modelo presentó una efectividad aceptable en la clasificación. Existe correlación entre las empresas estudiadas y el incremento en los índices de liquidez y rentabilidad de los años 2006, 2008 y 2010. De la función discriminante obtenida y de los estadísticos analizados se puede concluir que el indicador Razón Corriente no presentó diferencias significativas, lo que no ocurre para los indicadores Prueba Ácida que mejoró en los tres periodos analizados, Capital Neto, Margen Operacional, Margen Neto mejoraron en el 2006 y 2008, no evidencian el mismo resultado en el 2008 y 2010, el indicador Margen Bruto mejoró en los tres periodos y el indicador Utilidad Operacional se mantuvo del 2006 y 2008 pero mejoró del 2008 y 2010.

Mures, Garcia y Vallejo (2005), aplica la Regresión Logística y el Análisis Discriminante para estudiar la morosidad en entidades financieras concluyendo que los dos modelos estimados mediante Análisis Discriminante y Regresión Logística son significativos. Además, poseen una eleva capacidad predictiva, sobre todo entre los clientes no morosos, ya que la primera técnica clasifica

correctamente a la totalidad de tales individuos, alcanzando, la segunda, una tasa de aciertos del 98.1%. Dichos porcentajes son menores en cuanto a los individuos morosos, siendo los porcentajes de clasificación correcta del 88.9% para el Análisis Discriminante y del 94.4% para la Regresión Logística, lo que resulta en una tasa de aciertos del 97.1% para ambos modelos. En cuanto a las variables que determinan que un cliente sea moroso, han resultado significativas tanto aquéllas relativas a las características de la operación formalizada como las que reflejan el perfil del individuo. Así, dos variables, referidas al número de impagos anteriores y a la residencia del cliente y pertenecientes, por tanto, a cada uno de los dos bloques de variables explicativas seleccionadas, se incluyen en los dos modelos estimados. A éstas se añaden, para completar el modelo logit, otras dos variables, que recogen si la inversión financiada se ha destinado al traspaso de un negocio o a la adquisición de un vehículo. Por su parte, la Función Discriminante está formada por tres variables más, que hacen referencia a la duración de los retrasos que el cliente haya tenido, a la aportación de garantía hipotecaria y al estado civil del individuo. Todas las variables incluidas en el modelo de Regresión Logística tienen un efecto positivo en la variable dependiente, dado el signo de los coeficientes asociados, por lo que un cambio de una unidad en la variable provoca un aumento en el logaritmo de la odds de la probabilidad de que un cliente sea moroso, en el valor del coeficiente respectivo. Por lo que respecta a la Función Discriminante, las variables relativas al historial de impagos del cliente y la que recoge si es de nueva residencia en la zona provocan aumentos en la puntuación obtenida a partir de la función discriminante, mientras que la aportación de garantía real y el estado civil del

cliente, respecto a la situación de soltería, disminuyen el valor de la puntuación discriminante.

Choque (2014) estudia el riesgo crediticio utilizando la cartera bruta del banco, la mora bancaria, la tasa de crecimiento de la cartera bruta y la tasa de crecimiento de los clientes. Para realizar una lectura en términos porcentuales del modelo, se utilizan las elasticidades respecto a cada variable independiente, con la finalidad de explicar la ocurrencia del riesgo de crédito. De acuerdo a las elasticidades del modelo se indica que la cartera bruta muestra una relación inversa respecto al riesgo de crédito, si la cartera bruta se incrementara en 100%, entonces el riesgo de crédito se contrae hasta un 16%. La mora tiene una relación positiva, si la mora se incrementa en 100% esta afecta al riesgo, en un 41%. La tasa de crecimiento del número de clientes, reduce el riesgo de crédito, si el número de clientes se eleva en 100% el riesgo se reduce en un 11%.

Trejo, Rios y Martinez (2016), concluye que, en materia de riesgos financieros siempre es normal pensar en la administración del portafolio, instrumentos financieros y su valuación con diversas técnicas, tal es el caso del Movimiento Browniano y la fórmula de Black Scholes, o la valuación de bonos mediante la técnica de Merton. Aunque el riesgo mercado es la pérdida potencial más importante dentro de la administración de riesgos financiero, el riesgo crediticio no deja de ser importante para su administración en cualquier entidad financiera, aportando este último más de la mitad de los ingresos a la banca. Finalmente, propone un nuevo modelo de riesgo crediticio con su Análisis Discriminante respectivo en la identificación de créditos buenos y malos, para ello utiliza la regresión logística. La interpretación de los coeficientes, es

mediante la medición de la variación del modelo Logit estimado para una variación unitaria de la variable explicativa dada. Así, si ANT aumenta en un impago, el Logit estimado aumenta en 0.01 unidades, lo que sugiere una relación positiva respecto a la PIC. El mismo caso es para el Límite de Crédito y HIST; a mayor límite de crédito la PIC aumenta y, el aumento del historial de impagos provoca aumento de la PIC. En caso contrario, si % PAGO es mayor cada vez, entonces la PIC disminuye en 7.6 unidades. Así, la teoría de Crédito es aplicada correctamente en la regresión econométrica planteada mediante la estimación Logit.

García, Pérez y Flores (2013) utilizan un modelo discriminante para evaluar el riesgo bancario concluyendo que los ratios económico-financieros, individualmente considerados, no permiten distinguir entre empresas morosas o no morosas. El ratio con mayor poder discriminante fue $END6 = GF / (INCN + rOIE)$, con un 66% de clasificaciones correctas. Del análisis univariante de los ratios sí podemos, sin embargo, apuntar pautas de comportamiento en cada grupo, y de ahí elaborar un conjunto de acuerdo a esa teoría. Con la aplicación del análisis multivariante, llegamos a elaborar una función discriminante no sectorizada, con la que obtenemos un nivel de aciertos superior al 76%, siendo este modelo bastante fiable, ya que sus variables están libres de correlaciones y comportamientos anómalos. Los datos sobre la capacidad predictiva están en concordancia con las clasificaciones realizadas por los modelos, especialmente para el modelo general y para el modelo de empresas no constructoras, alcanzándose niveles de aciertos en la predicción del 69% y del 83% respectivamente.

Muñoz (1998), indica que las variables consideradas, la que más aporta a la discriminación entre bancos de altas y bajas utilidades es ingresos por concepto de intermediación financiera, seguido de la cartera de créditos vigente, variables que en el fondo están relacionadas una con otra, los ingresos por comisiones por servicios aportan un poco a la discriminación, pero en lo que respecta a las obligaciones que los bancos mantienen con el público no hay aporte importante a la discriminación

Espin y Rodriguez (2013), hace una introducción al desarrollo de los credit scoring, con el objetivo de introducir al lector en la utilización y finalidad de dichos modelos en la administración de riesgos. La metodología es sencilla con la que, primordialmente pequeñas empresas, pueden generar modelos confiables para originar clientes que no tengan experiencia crediticia.

Puertas & Marti (2012), indican que el problema de morosidad está cobrando una gran importancia en los países desarrollados. El estudio presenta un análisis comparativo entre dos técnicas paramétricas y una no paramétrica aplicadas a un problema real de clasificación crediticia. De manera que, dadas unas variables descriptivas del sujeto solicitante de un crédito, el modelo determinará con la mayor precisión posible si sería capaz o no de hacer frente a sus obligaciones crediticias. El problema del sobreaprendizaje que habitualmente se obvia en muchos estudios por el excesivo esfuerzo computacional que requiere. Los modelos de clasificación crediticia, que habitualmente utilizan las entidades financieras, se alimentan de la propia información que van generando, de manera que, si el modelo se equivoca muy frecuentemente, al cabo del tiempo el algoritmo deja de ser operativo porque los

resultados que genera no son, en absoluto, fiables. Por ello, y aunque dicha solución en principio pueda parecer un incremento de coste para la entidad, proponemos la distinción de las solicitudes en tres estados: conceder automáticamente (porque la probabilidad de que el cliente devuelva el crédito es superior al 70%); rechazar automáticamente (dado que su probabilidad será inferior al 30%); y distinguir un tercer estado que hemos denominado como «dudosos», en el cual se aconseja su estudio manual por parte del personal bancario. Lo que genera mayor quebranto a las entidades financieras es la insolvencia, es decir, que los clientes resulten fallidos. En el estudio se demuestra que se reduce considerablemente el error si, todas las solicitudes que no tengan una gran certeza de devolución o de incumplimiento, se procede a su análisis individual para determinar la conveniencia o no de rechazar/ conceder dicho crédito. Ahora bien, sabiendo que ello eleva el coste de tramitación de las entidades, consideramos que dicho grupo no debe ser demasiado elevado. Concluimos que el algoritmo CART obtiene mejores resultados que los modelos paramétricos, porque, aunque el error es algo superior, el de aciertos también lo es. Además, el porcentaje de calificados como dudosos no resultan excesivamente elevado, como sí ocurre con los otros métodos (AD y Logit).

Vigo (2010) indica que no existe diferencias en el porcentaje de error de entrenamiento en los métodos: regresión logística y árbol de clasificación (CART), utilizados para la clasificación de los clientes que solicitan un préstamo; sin embargo, con las redes neuronales se obtuvo un 84.18% de buena clasificación y un 74.32% de buena predicción. Con el modelo de Regresión Logística, se obtuvo mayor error en la clasificación y predicción debido a que

este método es sensible a los valores influyentes, al igual que los árboles de clasificación (CART), por otro lado, la red neuronal es insensible a valores influyentes. Para trabajos posteriores se recomienda utilizar meta modelos, que es la combinación de modelos diferentes, que subsanan el problema del sobreajuste de los datos de entrenamiento y que pueden mejorar la clasificación y la predicción.

Zapata (2009), Concluye “que el Análisis discriminante permite contrastar las variables en su conjunto y al observar sus interacciones reconocer algunas de sus dinámicas, que no son de fácil deducción. La regresión logística es una herramienta que puede ayudar en la toma de decisiones; reflejar las dinámicas de los componentes de las variables y realizar lecturas y aproximaciones fundamentadas a casos específicos, su valor en la calificación de clientes y evaluación de crédito puede ser confirmado en este trabajo.

Nieto (2010), Concluye que las variables directamente observables elegidas miden correctamente las variables latentes subyacentes de ellas; dado que el valor para el modelo en la tabla del Análisis de la Varianza es inferior a 0.01, hay una relación estadísticamente significativa entre las variables al 99 % de nivel de confianza. Además, el p-valor para los residuos es mayor o igual a 0.10, indicando que el modelo no es significativamente peor que el mejor modelo posible para estos datos al 90 % de nivel de confianza o superior. La ventana también muestra que el porcentaje de desviación en clase explicado por el modelo es igual a 3.78213. El porcentaje ajustado más adecuado para comparar modelos con diferentes números de variables independientes, es 3.7074 %.

Vela (2012), Indica que la calidad de la cartera está explicada por un conjunto de variables de carácter macroeconómico relacionados con el ciclo del producto y microeconómico tales como el nivel de solvencia de las entidades, la eficiencia y gestión de sus costos operativos, la tasa de crecimiento de los activos rentables. La relevancia global está corroborada con el coeficiente de determinación significativo encontrado en el modelo, sin embargo, la prueba de hipótesis “F”, demuestra que se deben incluir más variables en la especificación del modelo.

ANTECEDENTES LOCALES

Sarco (2015) hizo uso de la Regresión Logística para identificar factores a evaluar para no correr riesgos en el otorgamiento de crédito llegando a la conclusión que las variables intervinientes son: ingreso, gasto, hijos, edad, seguro, vivienda y material. El modelo explica el 78.7 de fiabilidad.

2.2 MARCO TEÓRICO

2.2.1 El riesgo en las entidades financieras.

Banegas y Garcia (2000). La actividad económica se desarrolla en un ambiente de incertidumbre, convirtiendo el riesgo en un factor inherente a la misma. De este modo, surge aquél como la contingencia, probabilidad o proximidad de un daño o peligro, en concreto, de sufrir una pérdida. La incertidumbre, junto con la aleatoriedad, constituyen las características

principales del riesgo, añadiéndose como tal el conflicto, ya que el riesgo se presenta ante situaciones diferenciadas entre las que elegir.

Así pues, el riesgo se ha convertido en uno de los rasgos básicos del entorno económico actual al que se enfrentan las empresas, que deben asumir riesgos en su toma de decisiones, en busca de la máxima rentabilidad, hablándose del binomio riesgo-rentabilidad como algo inseparable de la gestión de las mismas. En el caso de las entidades financieras, esta característica es esencial a la actividad que desarrollan, consistente en la concesión de créditos, asumiendo un riesgo cuando prestan unos recursos financieros que otros clientes les han cedido, sin controlar posteriormente el destino y utilización de los mismos.

El riesgo puede definirse, de forma más completa, como la exposición a la posibilidad de pérdidas económicas y financieras, de daños en las cosas y en las personas o de otros perjuicios como consecuencia de la incertidumbre que origina el llevar a cabo una determinada acción.

Esta incertidumbre que, como venimos comentando, es una de las características fundamentales del riesgo, se presenta en tres aspectos:

- En cuanto a la posibilidad de que se materialice una pérdida efectiva
- Sobre la cuantía de la pérdida
- Sobre el instante en que puede materializarse dicha pérdida.

El riesgo global en la actividad financiera resulta de la suma de distintos tipos de riesgos, que se describen, de forma breve, a continuación:

a) **Riesgo de mercado:** supone la probabilidad de pérdida ante movimientos adversos en los precios de los instrumentos financieros en los mercados donde se tengan posiciones. En función de su causa, se distingue, a su vez: riesgo de posiciones en divisas, riesgo de tipo de interés/precio, riesgo de operaciones con opciones y otros riesgos de mercado, relativos a correlación, pago anticipado y aseguramiento.

b) **Riesgo de crédito:** corresponde al riesgo de pérdida que se produce ante el incumplimiento de los pagos adeudados a la entidad, presentando como modalidades: riesgo de contraparte o sustitución, riesgo país, riesgo de liquidación o entrega y riesgo de emisor.

c) **Riesgo operacional y de tecnología:** comprende un conjunto de riesgos que surgen por la posible ocurrencia de sucesos inesperados, relacionados con la infraestructura operativa y tecnológica, tanto interna como externa, de la entidad.

d) **Riesgo de liquidez:** entendiéndose por liquidez la capacidad de transformar un activo en efectivo a los precios existentes en cada momento, sin incurrir en más pérdidas que las que, en su caso, imponga el mercado, puede definirse este riesgo como la posibilidad

de que una entidad financiera no disponga de liquidez y no pueda hacer frente a sus compromisos vencidos, ya sea por la imposibilidad de modificar una posición a tiempo en el mercado en un momento dado, por un desfase temporal en los flujos de caja o por necesidades de liquidez no previstas.

e) **Riesgo legal:** incluye una variedad diversa de riesgos que se generan por la imposibilidad de ejecutar un contrato en los términos previstos, y que comprenden el riesgo de documentación, el de legislación específica de cada país y el de capacidad de las contrapartes.

2.2.2 Riesgo de crédito

Beltran (1994). El riesgo de crédito es el tipo de riesgo más importante al que deben hacer frente las entidades financieras, por ser intrínseco a la actividad que desarrollan, y porque es la principal incertidumbre a la que estas entidades se enfrentan en las operaciones de activo que les vinculan a sus clientes, de ahí que este trabajo se centre en ese riesgo.

El riesgo de crédito podría definirse, de forma genérica, como el posible quebranto que originaría a la entidad una variación en las condiciones y características de una contrapartida, que pudiese alterar la capacidad de ésta para cumplir con los términos contractuales de una operación.

Una definición alternativa, que recoge las dos formas de manifestación del riesgo de crédito, es la que lo describe como la incertidumbre derivada de la posibilidad de sufrir un quebranto por el incumplimiento de alguna o de todas las obligaciones contractuales de la contraparte o contrapartida en una operación financiera, ya sea por entrada en mora del deudor, provocada por el retraso en dicho cumplimiento, o por el impago definitivo de sus obligaciones, lo que deviene en la insolvencia del mismo.

Esta distinción entre las manifestaciones del riesgo crediticio aparece, asimismo, recogida en diversas Circulares del Banco de España, que clasifican el crédito respecto a su situación, es decir, el estado de cumplimiento de las obligaciones de pago correspondientes. Atendiendo a esta normativa, se consideran como morosos aquellos créditos y demás operaciones vencidos y no cobrados, sin mediar novación o prórroga, cuando hayan transcurrido más de tres meses desde su vencimiento. Sin embargo, con carácter general, se emplea el calificativo de moroso para designar a la persona o deudor que se retrasa en el pago de sus deudas o en el cumplimiento de su obligación, al no efectuarlo en la fecha de vencimiento estipulado.

El riesgo de crédito se clasifica en dos grandes tipos de riesgos: el riesgo de contrapartida, cuyo origen está en el cambio de solvencia de una contrapartida o conjunto diferenciado de ellas, y el riesgo país, que se corresponde con la solvencia de contrapartidas que pertenecen a un área geográfica política y legalmente definida como estado. No obstante, dentro

de cada uno de estos riesgos pueden establecerse, a su vez, otras clasificaciones, según distintos criterios, como el evento o factor que genera el riesgo, la operación en que se manifiesta, el grado de concentración del riesgo o su estimación.

Por otro lado, el factor riesgo que acompaña a toda operación que las entidades formalizan con sus clientes, con el fin de concederles crédito, aparece en tres momentos clave de la relación con el cliente:

- En el estudio de la operación: la entidad financiera tiene que realizar un análisis riguroso del cliente antes de tomar la decisión, por lo que dedica un cierto tiempo a la evaluación de la operación, a partir de la documentación e información necesaria.
- En la formalización de la operación: se debe actuar de forma rigurosa, estableciendo los controles de seguridad adecuados.
- Entre la formalización y la cancelación: durante la vida de la operación, la entidad financiera debe realizar el oportuno seguimiento de la misma y del cliente.

Además, para medir el nivel de riesgo asociado a una operación de activo, es necesario considerar una serie de factores determinantes del mismo, que se refieren, en primer lugar, a las características de la operación. Dentro de éstas, se incluyen el plazo o vencimiento del término de la

operación, el importe o cuantía, la modalidad o tipo de instrumento de financiación solicitado, y el destino o uso de la inversión financiada.

Otro factor que influye en el nivel de riesgo, relacionado con la operación, son las garantías aportadas por el cliente como medio requerido por las entidades financieras para disminuir el riesgo que conllevan las operaciones de activo.

Por último, el nivel de riesgo se mide a través del conocimiento del solicitante de la operación. Para llegar a ese conocimiento, las entidades financieras realizan un análisis riguroso del cliente, a través del estudio de información interna de origen histórico, obtenida de bases de datos de la entidad. A su vez, estudian información externa.

2.2.3 Análisis Discriminante.

El Análisis Discriminante es una técnica de dependencia, en ocasiones conocida como análisis de clasificación, modelo de respuesta cualitativa o técnica de reducción de datos. Para realizar un Análisis Discriminante hay que tener en cuenta que la variable dependiente debe ser cualitativa y un conjunto de una o más variables independientes cuantitativas. Este método es factible aplicarlo incluso en aquella situación en la que la variable dependiente no es realmente nominal, pudiendo ser ordinal, de intervalos e incluso de razón. En este caso la solución al problema quedaría resuelto categorizando la variable no nominal (Pérez 2004).

El objetivo del análisis discriminante es encontrar una combinación lineal de las variables independientes que mejor permiten diferenciar (discriminar) a los grupos, una vez encontrada esa combinación (la función discriminante) podrá ser utilizada para clasificar nuevos casos. Dicho de otra manera, el Análisis Discriminante trata de establecer las relaciones óptimas existentes entre aquellas características de los individuos y sus grupos de pertenencia; lo que permitiría clasificar (identificar) nuevos individuos, a partir de sus características observadas, en uno de aquellos grupos y mediante una regla de decisión óptima que permitirá predecir la clasificación de los nuevos individuos de la forma más fiable posible con respecto a la realidad. La pertenencia de un individuo a un grupo se modeliza mediante una variable categórica que toma tantos valores como grupos haya y que también se conoce como variable grupo o variable independiente. Las características observadas a partir de las que se va a proceder a la identificación de los individuos se conocen como variables clasificadoras, variables criterios, variables predictoras o variables explicativas.

En la clasificación discriminante hay dos enfoques: el primero está basado en la obtención de funciones discriminantes de cálculo similar a regresión lineal múltiple y el segundo emplea técnicas de correlación canónica y de componentes principales y se denomina análisis discriminante canónico.

2.2.4 Tipos de análisis discriminante

Luque (2000). Se pueden clasificar atendiendo al número de

categorías de la variable dependiente

a) Análisis discriminante de dos grupos o simple, la variable dependiente tiene solo dos categorías de respuesta.

b) Análisis discriminante múltiple. la variable dependiente tiene más de dos categorías de respuesta.

2.2.5 Utilidad del Análisis Discriminante.

Según el objeto de investigación, se emplea con fines,

a) Explicativos, con la intención de cuantificar la contribución relativa de cada una de las variables independientes en la clasificación correcta de los individuos considerados dentro de los distintos grupos objeto de estudio, por tanto, se intenta probar el poder discriminante de cada una de las variables, en muchos casos con la finalidad de seleccionar el subconjunto que mejor discrimina los grupos.

b) Predictivos, es decir, encasillar a un individuo, del que no conocemos a qué grupo pertenece a priori, dentro de un grupo a partir de los valores de las variables independientes.

c) Reclasificadores, es decir, definidos los grupos, se desea recomponer esa partición. Este puede ser el caso cuando se desea

una clasificación orientada al reconocimiento o se busca una mejor interpretación de los grupos. Así, muchas veces se realiza un análisis clúster que posteriormente se intenta corroborar por medio de un análisis discriminante.

2.2.6 Asunciones (supuestos) del análisis discriminante,

Para poder aplicar el análisis discriminante, o por lo menos para que las conclusiones sean fiables, hay que tener en cuenta los siguientes supuestos fundamentales,

Homogeneidad de varianzas, es decir la matriz de covarianzas intra-grupo deben ser iguales o muy parecidos en todos los grupos objeto de estudio (dos matrices se dicen que son iguales si, y solo si, todos los elementos de las mismas coinciden). Si esto no es así, los resultados no son todo fiables, especialmente los test de significación y el proceso de clasificación. El problema es especialmente importante en el caso de que el tamaño de los diferentes grupos difiera en gran medida, ya que en el caso de que se incumpla esta restricción, se tiende a clasificar casos dentro de los grupos que tienen una matriz de covarianzas mayor. Por tanto, se debería optar por otra técnica alternativa. Para comprobar si se cumple o no esta restricción se suele recurrir al test de Box.

Cada uno de los grupos ha de ser una muestra procedente de una población que siga una distribución normal multivariante (normalidad de las

variables). En caso de que esto no se cumpla, se pueden producir problemas en la interpretación de las funciones discriminantes, sobre todo porque los test de significación que se aplican no son válidos, por lo que sería recomendable optar por otra técnica de análisis menos sensible a la violación de esta restricción como, por ejemplo, la regresión logística (Uriel, 1995), de forma tal que, si cada variable se distribuye normalmente, las variables conjuntamente se distribuirán como una normal multivariante. Si alguna de las variables no se distribuye normalmente, hay razones para suponer que la hipótesis de normalidad multivariante no se cumple.

La existencia de multicolinealidad entre las variables independientes implica que dos o más variables están altamente correlacionadas, por lo que una variable puede ser predicha o explicada por otras, es decir tiene escasa capacidad explicativa. La multicolinealidad no supone un problema si su presencia es similar en todas las posibles muestras. En caso contrario, multicolinealidad diferente según muestras, se presenta un problema como consecuencia de que los resultados dependerán de la muestra elegida para obtener la función discriminante (Sharma, 1996).

En la práctica, el análisis discriminante es una técnica robusta y funciona bien, aunque las dos restricciones anteriores no se cumplan, pero, también es aconsejable usar la regresión logística como una técnica alternativa.

2.2.7 El problema de la clasificación

La mayor utilidad de una función discriminante radica en su capacidad para clasificar nuevos casos. Ahora bien, la clasificación de casos es algo muy distinto de la estimación de la función discriminante. De hecho, una función perfectamente estimada puede no pasar de una pobre capacidad clasificatoria.

Una vez obtenida la función discriminante podemos utilizarla, en primer lugar, para efectuar una clasificación de los mismos casos, esto permitirá comprobar el grado de eficacia de la función desde el punto de vista de la clasificación. En segundo lugar, para clasificar futuros casos de los que, conociendo su puntuación en las variables independientes, se desconozca el grupo al que pertenecen.

Una manera de clasificar los casos consiste en calcular la distancia existente entre los centroides de ambos grupos y situar un punto de corte (d_c) equidistante de ambos centroides d_1 y d_2 (ver figura 01). A partir de ese momento, los casos cuyas puntuaciones discriminantes sean mayores que el punto de corte (d_c) serán asignados al grupo superior y los casos cuyas puntuaciones discriminantes sean menores que el punto de corte será asignado al grupo inferior.

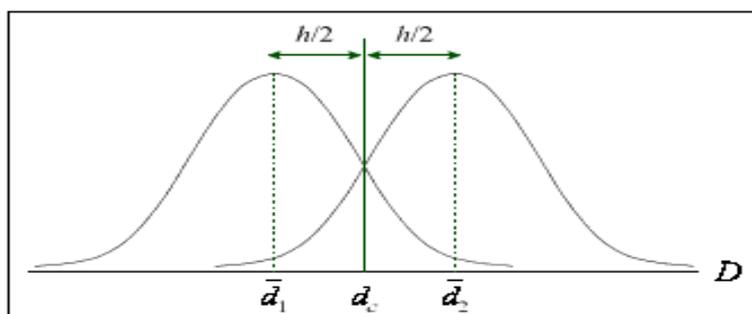


Figura 01. Utilización de un punto de corte equidistante de ambos centroides $n_1 = n_2$

Fuente: Luque, T. (2000). *Técnicas de análisis de datos en investigación de mercados*.

Esta regla de clasificación tiene un serio inconveniente: sólo permite distinguir entre dos grupos y es difícilmente aplicable al caso de más de dos grupos. Además, no tiene en cuenta que los grupos pueden tener distinto tamaño. Si ambos grupos son de igual tamaño, la situación real será muy similar a la descrita en la figura 01. Pero si, por el contrario, los tamaños muestrales son muy desiguales, la situación real será más parecida a la que muestra la figura 02. En esta figura puede verse con claridad que, si utilizamos el punto de corte como punto de clasificación, la proporción de casos mal clasificados en el grupo de menor tamaño (zona rayada horizontalmente) será mucho menor que en el grupo de mayor tamaño (zona rayada verticalmente). Por tanto, con tamaños desiguales es preferible utilizar una regla de clasificación que desplace el punto de corte hacia el centroide del grupo de menor tamaño buscando igualar los errores de clasificación. Para calcular este punto de corte podemos utilizar una distancia ponderada:

$$\bar{d}_c = \frac{n_1 \bar{d}_1 + n_2 \bar{d}_2}{n_1 + n_2}$$

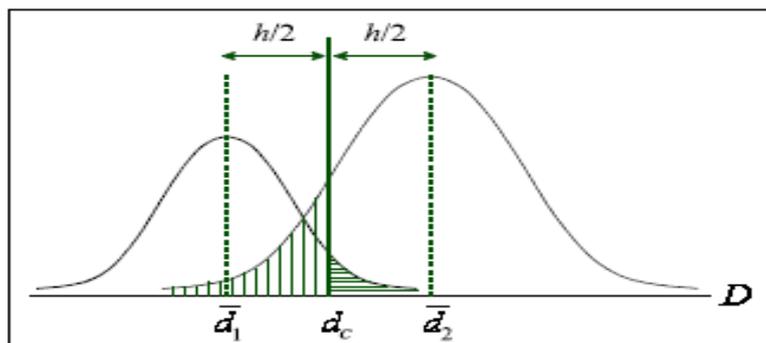


Figura 02. Utilización de un punto de corte equidistante de ambos centroides $n_1 \neq n_2$

Fuente: Luque, T. (2000). Técnicas de análisis de datos en investigación de mercados.

Fukunaga y Kessell (1972) han propuesto una regla de clasificación basada en la teoría bayesiana. Esta otra regla permite incorporar fácilmente la información relativa al tamaño de los grupos y, además, es extensible al caso de más de dos grupos.

Es frecuente que, aunque los tamaños de los grupos sean intrínsecamente diferentes, se desee compensar estadísticamente esa desigualdad a la hora de clasificar a los sujetos.

Las probabilidades a priori ofrecen alguna información sobre la representatividad de los casos, pero no ofrecen información concreta sobre un caso en particular. Además, las probabilidades a priori no tienen en cuenta que las probabilidades de aparición de las variables independientes en cada grupo pueden no ser simétricas. Por ejemplo, una sintomatología diagnóstica puede ser más frecuente en un grupo patológico que un grupo normal.

Por supuesto, siempre es posible aprovechar la información adicional que proporciona saber a qué grupo pertenece cada caso. Si asumimos que las puntuaciones discriminantes se distribuyen normalmente, podemos calcular la probabilidad asociada a un caso (es decir, la probabilidad que queda por encima o por debajo de ese caso) en cada uno de los grupos utilizados en el análisis. Esto es lo que se conoce como probabilidad condicional: $P(D > d_i | G = g_k)$ o, simplemente, $P(d_i|g_k)$. La probabilidad condicional de una puntuación discriminante puede calcularse mediante tablas de probabilidad asintótica o a partir de los cuantiles observados (ver Figura 03).

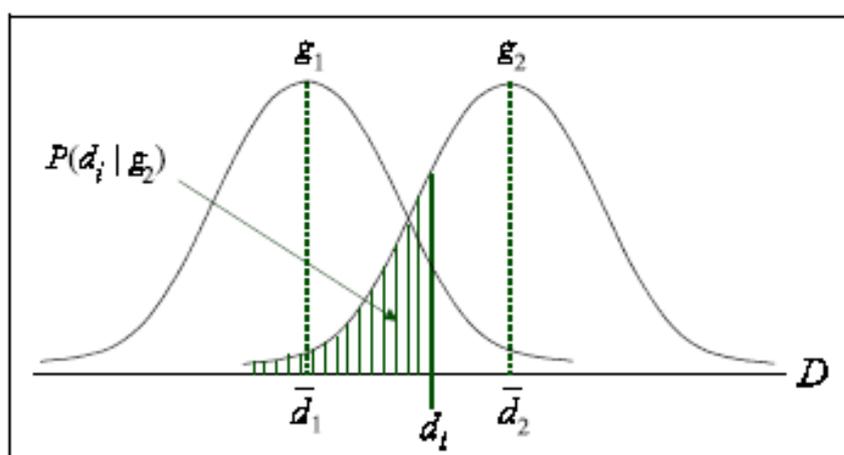


Figura 03. Probabilidad condicional de la puntuación discriminante d_i en el grupo 2.

Fuente: Luque, T. (2000). Técnicas de análisis de datos en investigación de mercados.

Una puntuación discriminante tiene asociadas tantas probabilidades condicionales como grupos hay en el análisis. Esas probabilidades condicionales indican cómo es de probable una puntuación concreta en cada uno de los grupos. Pero sólo son útiles cuando se conoce a qué grupo pertenece un caso. Cuando se desea clasificar un caso nuevo (del que,

obviamente se desconoce a qué grupo pertenece), es necesario comparar las probabilidades condicionales que le corresponden en cada uno de los grupos del análisis. Por ello, para clasificar un caso nuevo, es más apropiado utilizar las probabilidades a posteriori, es decir, las probabilidades de pertenecer a cada uno de los grupos, dado que a ese caso le corresponde una determinada puntuación discriminante, es decir:

$P(G=g_k|D = d_i)$ o, simplemente, $P(g_k | d_i)$. Estas probabilidades a posteriori se obtienen utilizando el teorema de Bayes:

$$p(g_k/d_i) = \frac{p(d_i/g_k) p(g_k)}{\sum_{k=1}^g p(d_i/g_k) p(g_k)}$$

El sumatorio del denominador posee tantos términos como grupos (no hay límite en el número de grupos). Con esta regla de clasificación, los casos nuevos son clasificados en el grupo al que corresponde mayor probabilidad a posteriori.

Aunque en la estimación de las probabilidades a priori es habitual utilizar los tamaños de los grupos, la aplicación del teorema de Bayes permite manipular esas probabilidades y asignarles un valor arbitrario (para reflejar mejor la composición de la población, para compensar el coste de una clasificación errónea, etc.). La manipulación de las probabilidades a priori hace que se desplace el punto de clasificación. Si se asigna igual probabilidad a priori a todos los grupos, el punto de corte para la clasificación

será equidistante de todos ellos; si se aumenta la probabilidad a priori de un grupo, el punto de corte para la clasificación se alejará de su centroide.

Ninguno de los procedimientos mencionados valora el coste de la clasificación errónea de los sujetos: todos ellos asumen igual coste para los aciertos y los errores en todos los grupos. Si existen costes diferenciales para cada tipo de acierto y para cada tipo de error, será necesario establecer el punto de corte mediante otro tipo de procedimientos más característicos de la Teoría de la toma de decisiones.

2.2.8 Clasificación o discriminación con 2 grupos.

El análisis discriminante permite diferenciar entre cualquier número de grupos. Sin embargo, comenzaremos con el caso de dos grupos, para luego ampliarlo a q grupos.

Partimos de que la población se divide en 2 grupos o subpoblaciones, G_1 y G_2 , están representados en el espacio bivalente definido por las variables X_1 y X_2 , la nube de puntos correspondientes a dos grupos hipotéticos. Los dos grupos representados se diferencian entre sí en ambas variables, pero no por completo, pues, de hecho, se solapan en una pequeña región situada entre ambos, Figura 04. Esta figura también muestra la función D , que es una combinación lineal de ambas variables. Sobre D se presenta la proyección de las dos nubes de puntos en forma de histograma, como si la función D cortara a las dos nubes de puntos en la dirección de su

eje. Las dos líneas punteadas de los histogramas representan la ubicación proyectada de los puntos medios de cada grupo (los centroides).

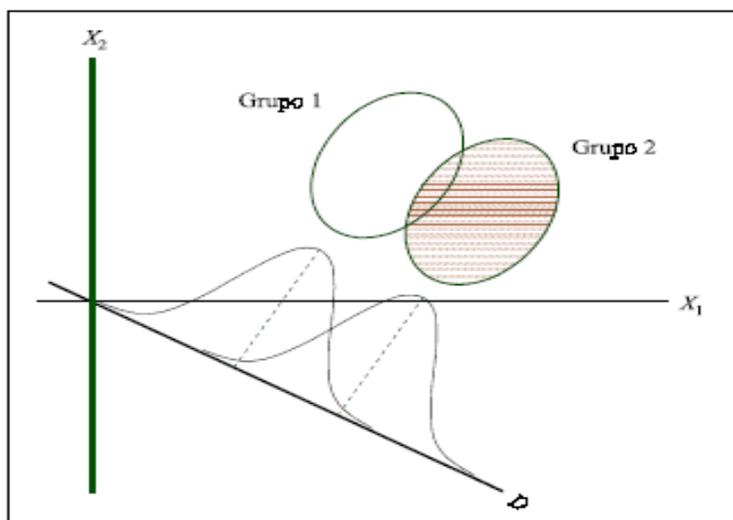


Figura 04. Probabilidad condicional de la puntuación discriminante d_i en el grupo 2.

Fuente: Luque, T. (2000). Técnicas de análisis de datos en investigación de mercados.

El propósito del análisis discriminante consiste en aprovechar la información contenida en las variables independientes para crear una función D combinación lineal de X_1 y X_2 capaz de diferenciar lo más posible a ambos grupos. La función discriminante es de la forma:

$$D = b_1X_1 + b_2X_2$$

Donde b_1 y b_2 son las ponderaciones de las variables independientes que consiguen hacer que los sujetos de uno de los grupos obtengan puntuaciones máximas en D , y los sujetos del otro grupo puntuaciones mínimas.

Una vez hallada la función Discriminante D , conviene centrar el interés en la representación de la función discriminante, que es unidimensional. La representación en p dimensiones resulta complicada cuando p es mayor de 2.

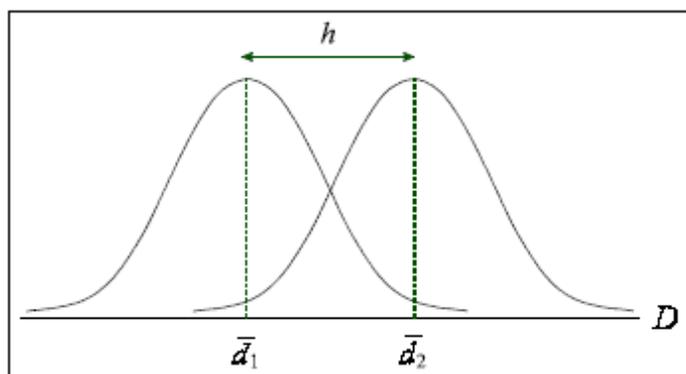


Figura 05. Histograma de cada grupo y centroides representados sobre la función discriminante.

Fuente: Luque, T. (2000). Técnicas de análisis de datos en investigación de mercados.

Sustituyendo en la función discriminante el valor de las medias del grupo 1 en las variables X_1 y X_2 , obtenemos el centroide del grupo 1.

$$\bar{d}_1 = b_1 \bar{x}_1^{(1)} + b_2 \bar{x}_2^{(1)}$$

De igual modo, sustituyendo las medias del grupo 2, obtenemos el centroide del grupo 2.

$$\bar{d}_2 = b_1 \bar{x}_1^{(2)} + b_2 \bar{x}_2^{(2)}$$

La función D debe ser tal que la distancia d entre los dos centroides sea máxima, consiguiendo que los grupos estén lo más distante posible. La distancia se expresa de la siguiente manera:

$$h = \bar{d}_1 - \bar{d}_2$$

Donde \bar{d}_1 y \bar{d}_2 son las medias del grupo 1 y del grupo 2 en la función D.

Se desea reducir la dimensionalidad de las p variables independientes en una sola dimensión (la de la combinación lineal D) en la que los grupos se diferencien lo más posible. Las puntuaciones de los sujetos en esa nueva dimensión (denominadas puntuaciones discriminantes) serán las que permitan llevar a cabo la clasificación de los sujetos.

2.2.9 Discriminante k grupos

Sea un conjunto de n objetos divididos en k grupos $\{G_i; i=1, \dots, k\}$ de tamaños $\{n_g; g=1, \dots, k\}$ que constituyen una partición de la población de la que dichos objetos proceden.

Sea $Y = (Y_1, \dots, Y_p)'$ un conjunto de variables numéricas observadas sobre dichos objetos con el fin de utilizar dicha información para discriminar entre los q grupos anteriores. (Mientras no se diga lo contrario, supondremos que las variables independientes son cuantitativas).

La expresión de una función discriminante, D, será:

$$D = b_1X_1 + b_2X_2 + \dots + b_pX_p + b_0$$

Y en particular, la puntuación discriminante para el i -ésimo individuo, $i = 1, \dots, n$ en la función discriminante será:

$$d_i = b_1x_1 + b_2x_2 + \dots + b_px_p + b_0$$

A partir de las puntuaciones discriminantes, un individuo i , para el que se conoce a cuál de los grupos pertenece, será clasificado en uno de ellos.

2.2.10 Cálculo de las funciones discriminantes (extracción de las funciones)

La discriminación entre los q grupos se realiza mediante el cálculo de unas funciones matemáticas denominadas funciones discriminantes. Existen varios procedimientos para calcularlas siendo el procedimiento de Fisher uno de los más utilizados.

Procedimiento Discriminante de Fisher.

El procedimiento de Fisher toma como funciones discriminantes, combinaciones lineales de las variables clasificadoras de la forma:

$$D = u_1X_1 + u_2X_2 + \dots + u_kX_k$$

Se trata de obtener los coeficientes de ponderación u_j . Si se considera que existe n observaciones, la función sería:

$$D_i = u_1 X_{1i} + u_2 X_{2i} + \dots + u_k X_{ki} \quad i = 1, 2, \dots, n$$

D_i es la puntuación discriminante correspondiente a la observación i -ésima.

Expresando las variables explicativas en desviaciones respecto a la media, D_i lo estará y la relación anterior se puede expresar en forma matricial como sigue:

Sean $\{d_{gk}; k = 1, \dots, n_g; g = 1, \dots, q\}$, los valores de la variable D en cada uno de los q grupos donde d_{gk} denota el valor de D en la k -ésima observación del g -ésimo grupo.

Sean,

$$\bar{d}_g = \frac{\sum_{k=1}^{n_g} d_{gk}}{n_g}; \quad g = 1, \dots, q$$

las medias muestrales de la variable D en cada uno de los q grupos y sea

$$\bar{d} = \frac{\sum_{g=1}^q \sum_{k=1}^{n_g} d_{gk}}{n}$$

la media de la variable D .

Observación

Si r es el número de funciones discriminantes se tiene que $\mathbf{WD} = \mathbf{I}_r$ y $\mathbf{BD} = \text{diag}(\lambda_1, \dots, \lambda_r)$ donde \mathbf{WD} y \mathbf{BD} son las matrices \mathbf{W} y \mathbf{B} calculadas utilizando las puntuaciones discriminantes. Se sigue que:

$$\lambda_i = \sum_{g=1}^q n_g (\bar{d}_g^i - \bar{d}^i)^2; \quad i = 1, \dots, r$$

donde $\{\bar{d}_g^i; g=1, \dots, q\}$ son las puntuaciones medias de la i -ésima función discriminante en los q grupos y \bar{d}^i es la puntuación media total.

Por lo tanto, los valores propios $\{\lambda_i; i=1, \dots, r\}$ miden el poder de discriminación de la i -ésima función discriminante de forma que si $\lambda_i = 0$ la función discriminante no tiene ningún poder discriminante. Dado que el rango de la matriz $\mathbf{W}^{-1}\mathbf{B}$ es a lo más $\min\{q-1, p\}$ el número máximo de funciones discriminantes que se podrán calcular será igual a $\min\{q-1, p\}$.

Lambda de Wilks para las funciones discriminantes.

(Evaluación de la significación de las funciones discriminantes)

Con la función ya desarrollada, el siguiente paso sería determinar si esta es verdaderamente significativa, es decir, si es capaz de discriminar entre grupos. Así pues, la hipótesis nula de que la media de todas las funciones discriminantes en todos los grupos es igual debe ser probada estadísticamente.

H_0 : la media de todas las funciones discriminantes en todos los

grupos es igual

$$H_0: \frac{Y}{G_i} \sim N_p(\mu_i, \Sigma); i = 1, \dots, q \text{ con } \mu_1 = \dots = \mu_q,$$

$$H_0: \lambda_1 = \dots = \lambda_{\min\{q-1, p\}} = 0$$

H_a : la media de todas las funciones discriminantes en todos los grupos no es iguales.

Usando los valores de la “chi cuadrado”. Si p es menor que el nivel de significancia se rechaza la hipótesis nula.

El estadístico viene dada por,

$$\Lambda = \frac{|W|}{|W + B|} = \frac{1}{\prod_{i=1}^{\min(q-1, p)} (1 + \lambda_i)}$$

Que indica la proporción total de la varianza en los resultados discriminantes no explicada por las diferencias entre grupos. Su valor oscila entre 0 y 1, cuanto más cerca de 0 esté, mayor es el poder discriminante de las variables consideradas y cuanto más cerca de 1, menor es dicho poder y no será adecuado para construir funciones discriminantes.

Este estadístico tiene una distribución lambda de Wilks con p , $q-1$ y $n-q$ grados de libertad.

M de Box. Prueba M de Box para el contraste de la hipótesis nula de igualdad de las matrices de varianzas-covarianzas poblacionales. Uno de los supuestos del análisis discriminante es que todos los grupos proceden de la misma población y, más concretamente, que las matrices de varianzas-covarianzas poblacionales correspondientes a cada grupo son iguales entre sí. El estadístico M de box toma la forma:

$$M = (n - g)\log|S| - \sum_{j=1}^g (n_j - 1)\log|S^{(j)}|$$

Donde S es la matriz de varianzas-covarianzas combinada, $S^{(j)}$ es la matriz de varianzas-covarianzas del j-ésimo grupo, n es el número total de casos, n_j es el número de casos en el j-ésimo grupo y g es el número de grupos. El estadístico M carece de distribución muestral conocida, pero puede transformarse en un estadístico F e interpretarse como tal (muchos analistas critican el uso de este estadístico por ser demasiado sensible a pequeñas desviaciones de la normalidad multivariante y a tamaños muestrales grandes, tendiendo a ser conservador).

2.2.11 Correlación canónica y autovalores asociados a una función discriminante.

La i-ésima correlación canónica viene dada por:

$$CR_i = \sqrt{\frac{\lambda_i}{1 + \lambda_i}} \quad i = 1, \dots, r$$

y mide, en términos relativos, el poder discriminante de la i -ésima función discriminante ya que es el porcentaje de la variación total en dicha función que es explicada por las diferencias entre los grupos.

Toma valores entre 0 y 1 de forma que, cuanto más cerca de 1 esté su valor, mayor es la potencia discriminante de la i -ésima función discriminante.

Mediante el estadístico lambda hemos comprobado que la información que aportará cada uno de las tres funciones es estadísticamente significativa, pero no conocemos que parte de la información será atribuible a cada una de ellas. La correlación canónica y el autovalor asociado a una función son dos medidas relacionadas con Lambda de Wilks, que permiten evaluar la información que aportara cada función discriminante en particular. La correlación canónica mide las desviaciones de las puntuaciones discriminantes entre grupos respecto a las desviaciones totales sin distinguir grupos. El autovalor mide las desviaciones de las puntuaciones discriminantes entre los grupos respecto a las desviaciones dentro de los grupos. En ambos casos, si el valor obtenido es grande (en el caso particular de la correlación canónica, si es próximo a 1) la dispersión será debida a las diferencias entre grupos y, en consecuencia, la función discriminará mucho los grupos.

2.2.12 Determinación del número de funciones discriminantes

El número de funciones discriminantes significativas se determina mediante un contraste de hipótesis secuencial.

Si denotamos por k = número de funciones discriminantes significativas el proceso comienza con $k = 0$. En el $(k+1)$ -ésimo paso del algoritmo la hipótesis nula a contrastar es

$$H_0: \lambda_{k+1} = \dots = \lambda_{\min\{G-1, p\}} = 0$$

Aunque Schatzoff (1966) obtuvo los puntos críticos exactos de la distribución de λ bajo ciertas condiciones, es más frecuente utilizar una transformación de λ que posee distribución aproximada conocida. Bartlett (1947) ha demostrado que el estadístico:

$$T = \left(n - 1 - \frac{p+q}{2} \right) \sum_{j=k+1}^{\min\{q-1, p\}} \log(1 + \lambda_j) = \left(n - 1 - \frac{p+q}{2} \right) \ln \lambda$$

se aproxima a la distribución chi-cuadrado con $(p-k)(q-k-1)$ grados de libertad: p es el número de variables independientes o discriminantes, q es el número de grupos, y k es el número funciones discriminantes obtenidas con anterioridad al contraste (cuando sólo existe una función –porque sólo hay dos grupos–, $k = 0$).

El p-valor asociado al contraste viene dado por:

$$P\left[\chi^2_{(p-k)(q-k-1)} \geq T_{\text{obs}}\right] \text{ donde } T_{\text{obs}} \text{ es el valor observado de } T.$$

El contraste para el primer valor de k para el cual la hipótesis nula H_0 se acepta.

La gran ventaja diagnóstica del estadístico lambda es que, puesto que se basa en las matrices de varianzas-covarianzas, puede calcularse antes de obtener las funciones discriminantes.

2.2.13 Interpretación de los resultados

Los resultados obtenidos se interpretan desde dos ópticas:

Significado de las dimensiones de discriminación entre los grupos proporcionadas por las funciones discriminantes mediante el análisis de la matriz de estructura y de la de los coeficientes estandarizados de las funciones discriminantes.

Análisis del sentido de la discriminación entre dichos grupos, es decir, averiguar qué grupos separa cada función discriminante y en qué sentido. Este análisis se lleva a cabo mediante representaciones gráficas del espacio de discriminación, así como de perfiles multivariantes correspondientes a cada grupo.

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 ENFOQUE DE LA INVESTIGACIÓN:

El enfoque de la presente investigación es cuantitativo debido a que se ha tomado el esquema y pasos de una investigación de este tipo.

3.2 DISEÑO DE INVESTIGACIÓN.

El diseño de investigación al que se ajusta la presente investigación es el descriptivo.

3.3 TIPO DE INVESTIGACIÓN.

El tipo de investigación es el descriptivo-correlacional

3.4 POBLACIÓN.

La población para el siguiente estudio lo conforman los datos registrados en todo el tiempo de funcionamiento de la financiera.

3.5 MUESTRA.

Por acceso a la información, la muestra estuvo conformada por 97

clientes registrados en el mes de octubre del año 2017.

3.6 INSTRUMENTO DE INVESTIGACIÓN.

Registro electrónico: Base de datos de la entidad financiera

3.7 METODOLOGÍA ESTADÍSTICA.

- 1) Seleccionar las características socio económicas para clasificar el tipo de cliente buen pagador de la Institución Financiera Compartamos - Agencia Puno.

Se depuro la base de datos determinando las características socio económicas utilizando el criterio de dependencia de las variables.

- 2) Seleccionar la función discriminante que clasifique el tipo de cliente buen pagador de la Institución Financiera Compartamos - Agencia Puno.
 - a) Test de normalidad. Se uso el test de Mardia que consiste en verificar la normalidad de las variables independientes.
 - b) Hipótesis Estadística

H_0 : Los datos tienden a la distribución normal.

H_1 : Los datos no tienden a la distribución normal

- c) Nivel de significancia: se establece en 0.05
- d) Prueba estadística: Test de Mardia
- e) Decisión. Si $p < \alpha$ no se rechaza la H_0 .

3) Prueba de igualdad de varianzas. El que se usara para observar la igualdad de varianzas dentro de los grupos es el test de Lambda de Wilks

4) Test de box. Esta prueba muestra si todos los grupos tienen la misma matriz de covarianzas o son muy parecidas.

a) Hipótesis Estadística

H_0 : Existe igualdad de matrices de varianzas-covarianzas (homoscedasticas). Homogeneidad de varianzas.

H_1 : Uno de los grupos es más variable que el otro. las matrices de covarianzas de poblaciones no son iguales

- b) Nivel de significancia: se establece en 0.05
- c) Prueba estadística: **M de Box**
- d) Decisión. Si $p < \alpha$ no se rechaza la H_0 .

5) Evaluación de la significación de la función.

- a) Obtención de los autovalores de las funciones canonicas discriminantes
 - b) Contraste del modelo. Lambda de Wilks
 - c) Obtener el valor de los centroides en las funciones discriminantes
 - d) Determinar los coeficientes de la función discriminante canónica
 - e) Determinar los coeficientes estandarizados de las funciones discriminantes cononicas
 - f) Obtener la matriz de estructuras para observar las correlaciones entre las variables
 - g) Determinar las probabilidades previas para los grupos
 - h) Determinar los coeficientes de clasificación
 - i) Observar el mapa de procesos
- 6) Clasificar el grupo de pertenencia de los denominados como buen pagador de la Institución Financiera Compartamos - Agencia Puno.

3.9 OPERACIONALIZACIÓN DE VARIABLES

Variable	Dimensión	Indicador	Índice
Tipo de cliente	Cliente	Clasificación	B: Cliente Regular A: Cliente Bueno A+: Cliente Muy Bueno
Características socio económicas	Sociales y Económicas	Mora promedio	Promedio simple en Días
		Atraso máximo del cliente	Días
		Atraso máximo del cliente en los últimos 3 meses	Días
		Atraso actual	Días
		créditos vigentes	Numero
		Máximo capital desembolsado vigente	Soles
		Deuda total en compartamos	Soles
		empresas en la SBS	Numero
		deuda directa en la SBS	Soles
		Ingresos	Soles
		Egresos	Soles

Fuente: Elaborado por el Ejecutor

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

Luego de la recolección de datos y el tratamiento de datos se llega a los siguientes resultados:

IDENTIFICACION DE VARIABLES

- ✓ Segmento cash: Crédito pre-aprobado. los clientes se estratifican según el comportamiento de pago (solo algunos clientes califican).

B: Cliente Regular

A: Cliente bueno

A+: Cliente Muy Bueno

Los clientes antes de desembolsar el préstamo pasan evaluación por un modelo predictivo por un área de la empresa (bajo la técnica de minería de Datos)

- ✓ Mora promedio del cliente (días): evaluación promedio de atraso del ultimo crédito en días (ejemplo: Si cliente saco préstamo para 12 meses) y pago de la siguiente manera:

	días de atraso
cuota 1	0
cuota 2	2
cuota 3	3
cuota 4	5
cuota 5	2
cuota 6	1
cuota 7	0
cuota 8	0
cuota 9	1
cuota 10	2
cuota 11	3
cuota 12	0
total	19

Promedio simple $19/12 = 1.58$ días de atraso en promedio

- ✓ Atraso máximo del cliente (en días): atraso máximo en el último crédito otorgado que en este caso (para el ejemplo, sería 5)
- ✓ Atraso máximo del cliente en los últimos 3 meses (días): de las tres últimas cuotas pagadas cual fue el atraso máximo.
- ✓ Atraso actual (días): si el crédito aun no fue cancelado cual es el atraso vigente que tiene.
- ✓ Número de créditos vigentes: a la fecha cuantos créditos vigentes tiene el cliente en la empresa.
- ✓ Máximo capital desembolsado vigente: cual fue el monto máximo que desembolsó el cliente en la empresa Compartamos Financiera

- ✓ Deuda total en compartamos: suma de sus préstamos saldo de capital del préstamo
- ✓ Número de empresas en la SBS: Numero de empresas (Bancos, Cajas, Financieras) en las que tiene prestamos según SBS (centrales de riesgo Infocorp)
- ✓ Deuda directa en la SBS: saldo capital que a la fecha debe a las empresas financieras.
- ✓ Ingresos: Ingresos del cliente según última evaluación Financiera
- ✓ Egresos: Gastos del cliente según última evaluación Financiera

Tabla 01. Características socio económicas descriptivas del buen pagador de la Institución Financiera Compartamos - Agencia Puno 2017.

	Informe										
	Atraso días	Atraso 3 meses	atraso actual	vigentes	SBS	Ingresos	egresos	mora	desembolso	deuda	deuda directa
Media	6,97	2,63	10,63	1,25	1,67	6434,05	4029,71	0,91	5865,69	4068,48	5971,34
Desviación estándar	7,252	2,881	7,539	0,43	0,718	4184,49	3701,92	0,90	6479,89	4978,43	6996,05
Mínimo	0	0	0	1	1	1600	140	,00	500	136,99	179,33
Máximo	54	14	29	2	3	22600	19190	4,29	30000	20899,22	34098,07

Fuente: Base de datos Financiera compartamos-agencia Puno.

Leyenda:

Atraso días: atraso máximo del cliente (días)

Atraso 3 meses: atraso máximo del cliente últimos 3 meses (días)

Atraso actual: atraso actual (días)

Vigentes: número de créditos vigentes

SBS: número de empresas en la SBS

Ingreso: Ingreso

Egreso: Egreso

Mora: mora promedio del cliente (días)

Desembolso: máximo capital desembolsado vigente

Deuda: deuda total en compartamos

Deuda directa: deuda directa

La tabla 01 muestra los estadísticos descriptivos de las variables en estudio donde podemos destacar que el promedio de desembolso de los clientes es S/. 5,866 con un monto máximo de S/. 30,000 y mínimo de S/. 500. El atraso promedio de pago en días es 7.97 días, el atraso promedio de los últimos 3 meses es 3 días es 2.63 días, el número promedio de créditos vigentes de los clientes es 1,25 indicando que adeudan a otras financieras. Los ingresos promedio se establecen en S/. 6,434 con ingresos mínimos de 1600 y máximos de S/. 22,600. Otro indicador es que la deuda de los clientes supera al desembolso, esto ratifica el hecho de que los clientes tiene más de un préstamo con las entidades financieras en su mayoría.

Tabla 02. características socio económicas descriptivas por segmento del buen pagador de la Institución Financiera Compartamos - Agencia Puno 2017.

Segmento	A+		A		B	
	Media	No ponderados	Media	No ponderados	Media	No ponderados
Atraso máximo del cliente (días)	8.25	24	6.67	55	6.17	18
Atraso máximo del cliente últimos 3 meses (días)	1.96	24	2.84	55	2.89	18
Atraso actual (días)	11.96	24	9.29	55	12.89	18
Número de créditos vigentes	1.25	24	1.25	55	1.22	18
Número de empresas en la SBS	1.33	24	1.80	55	1.72	18
Ingresos	7074.17	24	6490.96	55	5406.67	18
Egresos	4693.75	24	3996.13	55	3246.94	18
Mora promedio del cliente (días)	0.90	24	0.92	55	0.87	18
Máximo capital desembolsado vigente	4757.38	24	6778.09	55	4555.56	18
Deuda total en compartamos	3616.86	24	4685.22	55	2786.15	18
Deuda directa en la SBS	4186.27	24	7114.27	55	4859.14	18

Fuente: Base de datos Financiera compartamos-agencia Puno.

La tabla 02 muestra: Atraso máximo del cliente: el atraso de los clientes en días se da en el segmento A+ (Cliente muy bueno), seguido del segmento A (Cliente bueno) y finalmente los del segmento B (Cliente regular). Podría deberse a los montos de préstamo y el atraso debe considerarse dentro de los plazos establecidos, es decir que no están bajo mora.

Atraso máximo del cliente últimos 3 meses: Al observar los días promedio de atraso a los 3 meses pasa lo contrario al punto anterior, aquí se observa que los clientes del segmento B tienen mayor atraso en los pagos, aunque la diferencia no es muy marcada.

Atraso actual (días): observando el atraso actual, los clientes del segmento B son los más atrasados, seguidos por los del segmento A+.

Se observa igualmente que en promedio todos los clientes de los diferentes segmentos tienen más de una deuda con las entidades financieras

Ingresos: la tendencia de ingresos es de acuerdo a los segmentos (A+, A, B), coincidiendo con los egresos, es decir quienes más ganan tienen mayores egresos.

En promedio los que más desembolsaron fueron los clientes del segmento A.

Test de normalidad de Mardia.

Al no cumplir con el supuesto de normalidad de las variables observadas, se ha procedido a transformar las variables con la transformación del log en base 10.

Ejecutado el software R arroja los siguientes resultados.

Prueba	Estadístico	Valor p
sesgo	511.118915543936	6.20675661221441e-15
curtosis	3.37785626472252	0.0498247670362472

Observamos que no es significativa tanto en el sesgo como en la curtosis.

Tabla 03. Pruebas de igualdad de las medias de los grupos

	Lambda de Wilks	F	gl1	gl2	Sig.
Atraso máximo del cliente (días)	0,999	0,035	2	94	0,966
Atraso máximo del cliente últimos 3 meses (días)	0,976	1,180	2	94	0,312
Atraso actual (días)	0,997	0,130	2	94	0,879
Número de créditos vigentes	0,999	0,037	2	94	0,963
Número de empresas en la SBS	0,923	3,898	2	94	0,024
Ingresos	0,989	0,546	2	94	0,581
Egresos	0,992	0,381	2	94	0,684
Mora promedio del cliente (días)	0,996	0,193	2	94	0,824
Máximo capital Desembolsado vigente	0,974	1,279	2	94	0,283
deuda total en Compartamos	0,981	0,911	2	94	0,406
deuda directa en la SBS	0,953	2,322	2	94	0,104

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

ANOVAs univariados. Permite contrastar la hipótesis de igualdad de medias entre los grupos en cada variable independiente. Incluye también el estadístico Lambda de Wilks univariante. Suele utilizarse como prueba preliminar para detectar si los grupos difieren en las variables de clasificación seleccionadas; sin embargo, debe tenerse en cuenta que una variable no significativa a nivel univariante podría aportar información discriminativa a nivel multivariante.

Entre las diferentes variables independientes consideradas solo una tiene capacidad discriminante a un nivel de significancia del 5%. (número de empresas

en la SBS, es significativa $p(0.024 < \alpha(0.05))$. Debido a la robustez del método continuaremos con el análisis.

Asunciones relativas a la aplicación de la técnica.

Tabla 04. Logaritmos de los determinantes de los diferentes segmentos

segmento cash	Rango	Logaritmo del determinante
A+	11	-30,760
A	11	-30,121
B	11	-35,806
Dentro de grupos combinados	11	-29,595

Los logaritmos naturales y los rangos de determinantes impresos son los de las matrices de covarianzas de grupo.

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La tabla 4 muestra los logaritmos de los determinantes de todas las matrices utilizadas en el cálculo del estadístico M. Dado que el estadístico es multivariante, la tabla permite comprobar qué grupos (cuando hay más de dos) difieren más. Podemos observar que entre los grupos A+ y A no existe mucha diferencia, pero si estos frente al grupo B.

Tabla 05. Tabla de resultados de la prueba M de Box

M de Box		160,792
F	Aprox.	,943
	gl1	132
	gl2	8121,384
	Sig.	,666
Prueba la hipótesis nula de las matrices de covarianzas de población iguales.		

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

M de Box. Prueba la hipótesis nula de igualdad de las matrices de varianzas-covarianzas poblacionales. Uno de los supuestos del análisis discriminante es que todos los grupos proceden de la misma población y, más concretamente, que las matrices de varianzas-covarianzas poblacionales correspondientes a cada grupo sean iguales entre sí.

El resultado de la prueba permite no rechazar la hipótesis de igualdad de matrices de varianzas-covarianzas por tanto podemos concluir que existe igualdad de matrices de varianzas-covarianzas (homoscedasticas). Homogeneidad de varianzas y continuamos con el análisis.

Evaluación de la significación de la función

Tabla 06. Autovalores de las funciones canónicas discriminantes

Función	Autovalo r	% de varianza	% acumulado	Correlación canónica
1	,253 ^a	75,2	75,2	,449
2	,083 ^a	24,8	100,0	,277

a. Se utilizaron las primeras 2 funciones discriminantes canónicas en el análisis.

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La tabla 06 presenta los autovalores (eigenvalues) de las funciones canónicas discriminantes, que miden las desviaciones de las puntuaciones discriminantes entre grupos respecto a las desviaciones dentro de los grupos. En la tabla se observa las dos funciones que componen el modelo son muy desiguales. La primera función explica el 75.2 % de la variabilidad disponible en los datos, mientras que la segunda función sólo explica el 24.8%, la segunda aporta poca información. De manera similar, la correlación canónica de la primera función es más alta (0,449), mientras que la de la segunda función es más baja (0,277). Los valores de la correlación decrecen la primera función discrimina más que la segunda.

Tabla 07. Lambda de Wilks. Contraste de las funciones del modelo.

Prueba de funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a 2	,737	27,210	22	,203
2	,923	7,124	10	,714

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La lambda de Wilks de la tabla 07, contrasta de manera jerárquica la significación de las dos funciones obtenidas. En la primera línea (1 a 2) se contrasta la hipótesis nula de que el modelo completo (ambas funciones discriminantes tomadas juntas) no permite distinguir las medias de los grupos. $p(0.203) > \alpha(0.05)$.

En la segunda línea (2) se contrasta si las medias de los grupos son iguales en la segunda función discriminante. La lambda de Wilks toma un valor muy próximo a 1, pero el nivel crítico (Sig. = 0.714) es mayor que 0,05, por lo que podemos concluir que la segunda función tampoco permite discriminar entre, al menos, dos de los grupos.

Los p valores de lambda de Wilks certifican la significatividad de los dos ejes discriminantes, la capacidad explicativa será regular (no separan bien los grupos).

Tabla 08. Valor de los centroides en las funciones discriminantes.

Funciones en centroides de grupo		
segmento cash	Función	
	1	2
A+	-,850	,088
A	,222	-,213
B	,457	,535

Las funciones discriminantes
canónicas sin estandarizar se han
evaluado en medias de grupos

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La tabla 08, muestra la ubicación de los centroides en cada una de las funciones discriminantes.

Si las medias de los grupos en cada función son muy parecidas la función no discrimina los grupos. En nuestro caso, los del segmento A+ están ubicados opuestamente a los de los otros segmentos en la función 1. En la función 2 los del segmento A están ubicados al lado negativo de los del segmento A+ y B.

Tabla 09. Coeficientes de la función discriminante canónica

	Función	
	1	2
Atraso máximo del cliente (días)	-,762	,699
Atraso máximo del cliente últimos 3 meses (días)	1,652	-,014
Atraso actual (días)	,086	2,070
Número de créditos vigentes	-1,110	2,070
Número de empresas en la SBS	3,050	-,483
Ingresos	-5,409	-,265
Egresos	2,083	-,600
Mora promedio del cliente (días)	-,077	-,642
Máximo capital desembolsado vigente	2,373	,692
Deuda total en compartamos	-1,660	-1,970
Deuda directa en la SBS	1,134	,653
(Constante)	5,666	2,421

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La tabla 09 muestra los coeficientes de los modelos discriminantes, podemos indicar que existen coeficientes positivos y negativos dentro de los modelos 1 y 2. Desde luego entendiendo que son modelos adicionales.

Coeficientes no tipificados. Coeficientes brutos de la función canónica discriminante. Son los coeficientes utilizados por el programa para calcular las puntuaciones discriminantes y la ubicación de los centroides de

los grupos. La función discriminante incluye una constante correctora que consigue que las puntuaciones discriminantes tomen el valor 0 en algún punto entre los dos centroides.

Tabla 10. Coeficientes estandarizados de las funciones discriminantes canónicas.

	Función	
	1	2
Atraso máximo del cliente (días) (X_1)	-,311	,285
Atraso máximo del cliente últimos 3 meses (días) (X_2)	,549	-,005
Atraso actual (días) (X_3)	,036	,868
Número de créditos vigentes (X_4)	-,146	,272
Número de empresas en la SBS (X_5)	,548	-,087
Ingresos (X_6)	-1,459	-,072
Egresos (X_7)	,982	-,283
Mora promedio del cliente (días) (X_8)	-,035	-,292
Máximo capital desembolsado vigente (X_9)	1,064	,310
Deuda total en compartamos (X_{10})	-,832	-,987
Deuda directa en la SBS (X_{11})	,579	,333

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La matriz de coeficientes estandarizados (tabla 10) contiene dos columnas, una para cada función discriminante. Las funciones se encuentran siempre ordenadas en correspondencia con los autovalores, siendo la primera función la de mayor capacidad discriminativa. Los coeficientes estandarizados de la primera función no difieren sustancialmente de los obtenidos en la función estimada en el caso de dos grupos. Esta primera función discrimina, fundamentalmente a:

- Atraso máximo del cliente últimos 3 meses (días)
- Atraso actual (días)
- Número de empresas en la SBS

- Egresos
- Máximo capital desembolsado vigente
- Deuda directa en la SBS

Donde se atribuye mayor ponderación al desembolso del capital que estaría relacionado al segmento A+

Las ecuaciones de las funciones canónicas discriminantes tipificadas son:

$$D1 = -0.311X_1 + 0.549X_2 + 0.036X_3 - 0.146X_4 + 0.548X_5 - 1.459X_6 + 0.982X_7 \\ - 0.035X_8 + 1.064X_9 - 0.832X_{10} + 0.579X_{11}$$

$$D2 = 0.285X_1 - 0.005X_2 + 0.868X_3 + 0.272X_4 - 0.87X_5 - 0.072X_6 - 0.283X_7 \\ - 0.289X_8 + 0.310X_9 - 0.987X_{10} + 0.333X_{11}$$

Con las funciones estandarizadas podemos apreciar que variables influyen más en cada función discriminante.

Tabla 11. Matriz de estructuras.

	Función	
	1	2
Número de empresas en la SBS	0,567*	-0,141
Deuda directa en la SBS	0,418*	-0,247
Atraso máximo del cliente últimos 3 meses (días)	0,315*	-0,010
Máximo capital desembolsado vigente	0,298*	-0,237
Atraso actual (días)	-0,035	0,715*
Deuda total en compartamos	0,184	-0,361*
Ingresos	-0,128	-0,299*
Egresos	-0,076	-0,283*
Mora promedio del cliente (días)	0,001	-0,222*
Número de créditos vigentes	-0,017	-0,093*
Atraso máximo del cliente (días)	0,013	0,091*

Correlaciones dentro de grupos combinados entre las variables discriminantes y las funciones discriminantes canónicas estandarizadas

Variables ordenadas por el tamaño absoluto de la correlación dentro de la función.

*. La mayor correlación absoluta entre cada variable y cualquier función discriminante.

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La matriz de estructura (tabla 11) ofrece los coeficientes de correlación entre las variables independientes y las puntuaciones discriminantes de cada función. El coeficiente más alto de cada variable aparece marcado con un asterisco que indica cuál es la función con la que más correlaciona esa variable (lo que no significa que sea ésa la función en la que más discrimina la variable).

La primera función correlaciona con:

- número de empresas en la SBS

- deuda directa en la SBS
- atraso máximo del cliente últimos 3 meses (días)
- máximo capital desembolsado vigente

La segunda función correlaciona:

- Deuda total en compartamos
- Ingresos
- Egresos
- Mora promedio del cliente (días)
- Número de créditos vigentes
- Atraso máximo del cliente (días)

Para valorar la capacidad predictiva del modelo estimado debemos prestar atención a los resultados de la clasificación.

Tabla 12. Probabilidades previas (probabilidades a priori utilizadas en la clasificación).

segmento cash	Previa	Casos utilizados en análisis	
		No ponderados	Ponderados
A+	,333	24	24,000
A	,333	55	55,000
B	,333	18	18,000
Total	1,000	97	97,000

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

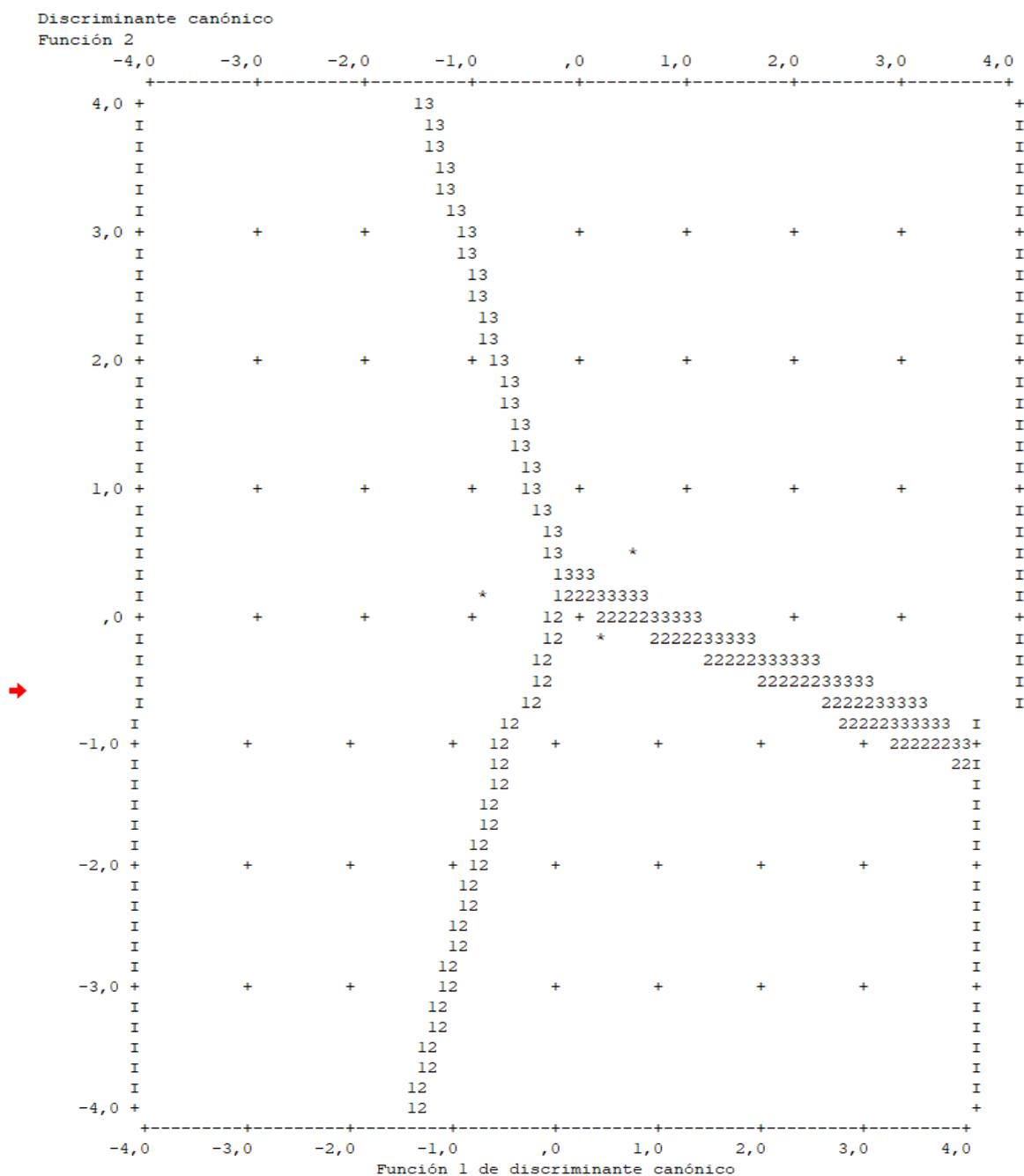
La tabla 12. ofrece las probabilidades previas o a priori que se utilizan también para clasificar los grupos. Estas probabilidades indican que se ha dado la misma importancia relativa a todos los grupos: 0,333 de pertenecer a cada grupo.

Tabla 13. Coeficientes de función de clasificación

	segmento cash		
	A+	A	B
Atraso máximo del cliente (días)	3,552	2,525	2,869
Atraso máximo del cliente últimos 3 meses (días)	7,820	9,595	9,973
Atraso actual (días)	1,634	1,103	2,672
Número de créditos vigentes	-27,559	-29,373	-28,085
Número de empresas en la SBS	-5,752	-2,338	-1,982
Ingresos	166,959	161,242	159,772
Egresos	-64,115	-61,701	-61,661
Mora promedio del cliente (días)	-5,520	-5,409	-5,908
Máximo capital desembolsado vigente	-17,480	-15,146	-14,071
Deuda total en compartamos	11,783	10,598	8,734
Deuda directa en la SBS	-9,641	-8,622	-7,867
(Constante)	-179,660	-173,999	-171,056
Funciones discriminantes lineales de Fisher			

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La tabla 13, presenta otra forma de clasificar casos que consiste en emplear funciones lineales de Fisher, de forma que se asignara cada individuo a aquel grupo para el que obtuviera una puntuación superior.



Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

Figura 06. Mapa territorial

Simbolos utilizados en mapa territorial

Simbolo	Grupo	Etiqueta
1	1	A+
2	2	A
3	3	B
*		Indica el centroide de un grupo

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La figura 6, muestra el mapa territorial. Un mapa territorial representa el territorio (espacio) que corresponde a cada uno de los grupos en el plano definido por las dos funciones discriminantes: la primera función en el eje de abscisas y la segunda función en el eje de ordenadas.

Los centroides de cada grupo están representados por asteriscos. Observando la ubicación de los centroides en la figura se aprecia claramente que la primera función posee mayor capacidad discriminativa que la segunda, pues los centroides se dispersan o alejan más en la dirección horizontal que en la vertical.

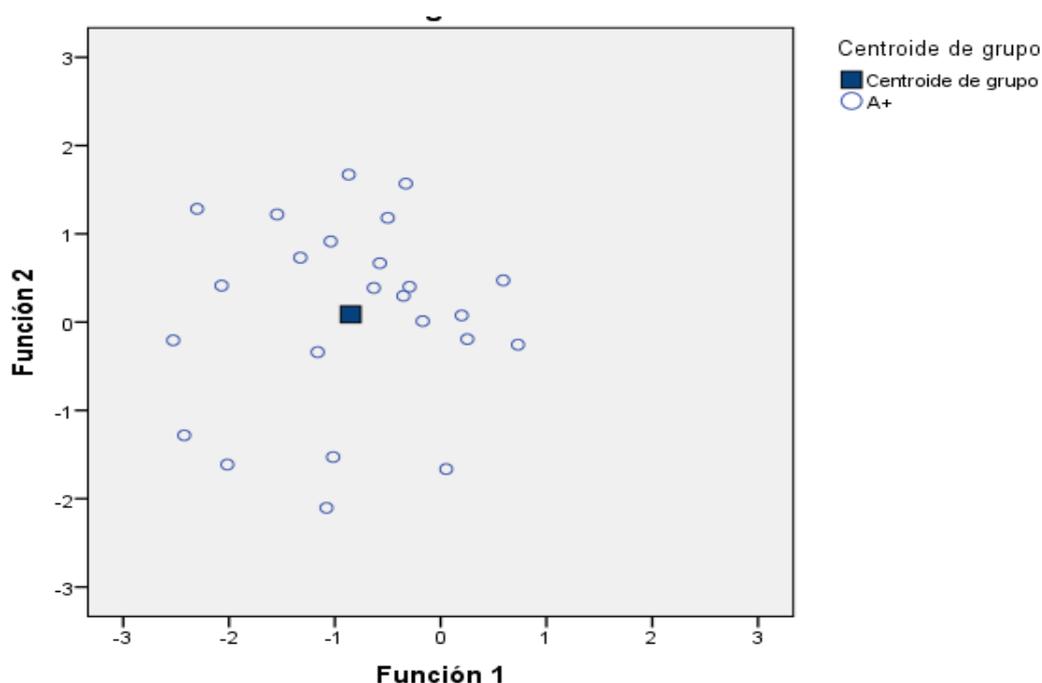


Figura. 07 Diagrama de dispersión de las funciones discriminantes canónicas del segmento A+.

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

La figura 07, muestra el diagrama de dispersión sobre el plano definido por las dos funciones discriminantes. Los casos están identificados por el segmento A+. La mayor utilidad de este gráfico radica en la posibilidad de identificar casos atípicos difíciles de clasificar. El grafico no muestra datos dispersos.

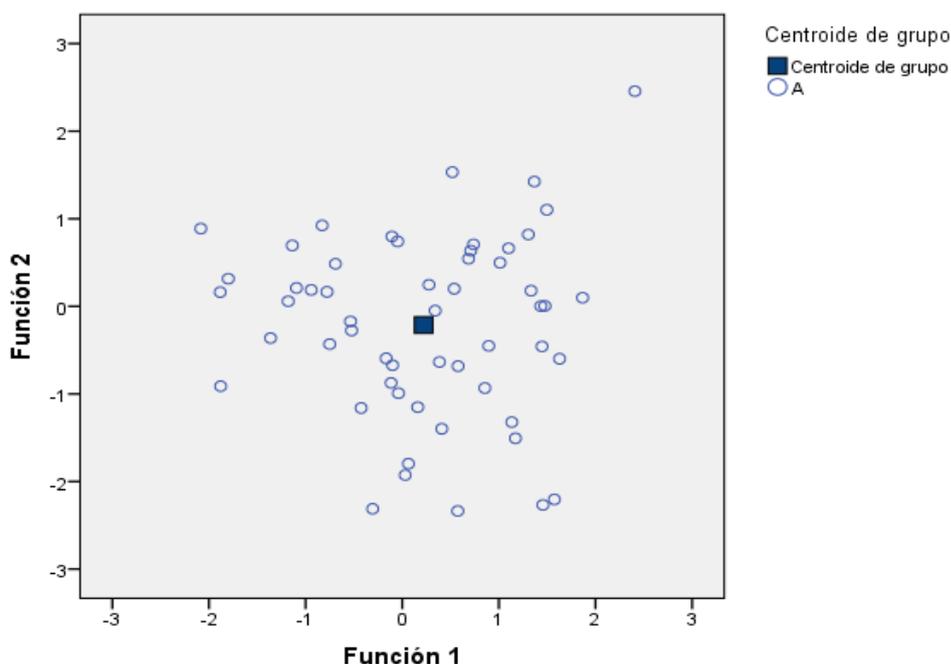


Figura 08. Diagrama de dispersión de las funciones discriminantes canónicas del segmento A.

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

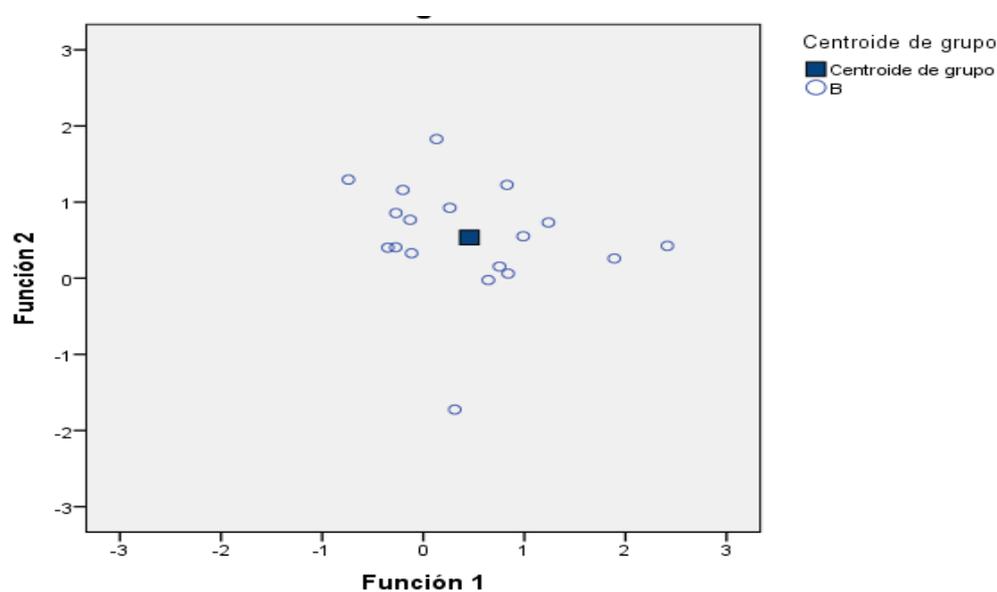


Figura 09. Diagrama de dispersión de las funciones discriminantes canónicas del segmento B.

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

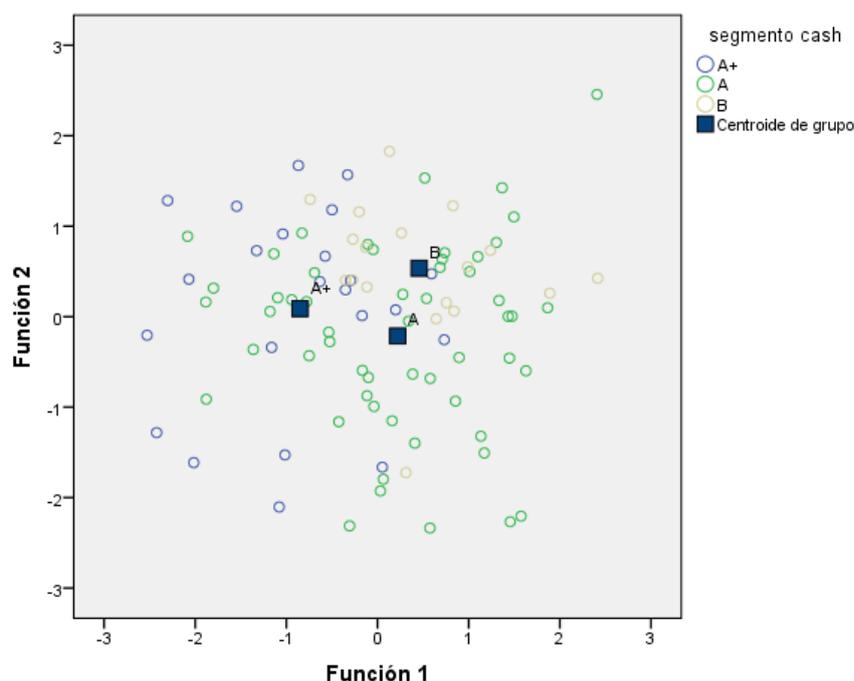


Figura 10. Diagrama de dispersión de las funciones discriminantes canónicas de los segmentos A+,A y B.

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

Excepto a que el segmento A tiene un valor alejado los demás segmentos discriminan en función al centroide.

Tabla 13. Cambio entre valores observados y reclasificados

Segmento	Cambios			Total
	A+	A	B	
A+	--	5	2	
A	13	--	18	
B	2	3	---	
Total				43

Fuente: elaborado por los ejecutores en base a los resultados.

El modelo ha permitido reclasificar el 44% de los clientes a otras categorías.

Según el estudio: (el cuadro resumen se muestra en el anexo)

- 18 (18.6%) clientes debieron ser clasificados como clientes regular mas no así, como clientes buenos

- 13 (13.4%) clientes debieron ser clasificados como clientes muy buenos mas no así, como Buenos clientes

- 5 clientes debieron ser clasificados como clientes como buenos clientes mas no así, como clientes muy buenos

CONCLUSIONES

PRIMERO:

Existen diferencias en las características socio económicas entre los segmentos de cliente muy bueno, cliente bueno y cliente regular, destacando que el atraso máximo de pago en días y atraso actual se encuentra en el segmento A+, los clientes tienen préstamos en más de una entidad financiera, los ingresos y egresos se distribuyen de acuerdo al segmento (los que más ganan, más se prestan), los del segmento A son los que tienen en promedio más cantidad de préstamo que los otros segmentos.

SEGUNDO:

Las funciones discriminantes son:

$$D1 = -0.311X_1 + 0.549X_2 + 0.036X_3 - 0.146X_4 + 0.548X_5 - 1.459X_6 \\ + 0.982X_7 - 0.035X_8 + 1.064X_9 - 0.832X_{10} + 0.579X_{11}$$

$$D2 = 0.285X_1 - 0.005X_2 + 0.868X_3 + 0.272X_4 - 0.87X_5 - 0.072X_6 - 0.283X_7 \\ - 0.289X_8 + 0.310X_9 - 0.987X_{10} + 0.333X_{11}$$

La primera función correlaciona con: Número de empresas en la SBS, Deuda directa en la SBS, Atraso máximo del cliente últimos 3 meses (días) y Máximo capital desembolsado vigente. La segunda función correlaciona: Deuda total en compartamos, Ingresos, Egresos, Mora promedio del cliente (días), Número de créditos vigentes y Atraso máximo del cliente (días)

TERCERO:

Teniendo en cuenta que el modelo de clasificación alcanza una predicción regular, el modelo ha permitido reclasificar el 44% de los clientes a otras categorías. El 18.6% de clientes debieron ser clasificados como clientes regular mas no así, como clientes buenos, el 13.4% clientes debieron ser clasificados como clientes muy buenos mas no así, como buenos clientes y 5 clientes debieron ser clasificados como buenos clientes mas no así, como clientes muy buenos.

CUARTO:

La mejor relación para clasificar el tipo de cliente en función a las características socio económicas del buen pagador de la Institución Financiera Compartamos - Agencia Puno 2017 se da mediante la primera ecuación D1.

RECOMENDACIONES

PRIMERO:

Utilizar otros modelos como el logístico y Machine Learning para comparar los resultados de la clasificación de los clientes.

SEGUNDO:

Continuar con el análisis incluyendo otras variables.

TERCERO:

Los futuros análisis deben incluir información que no se circunscriba a la que se obtiene de los clientes y del análisis de crédito, deben contemplarse la posibilidad de utilizar variables económicas de subsectores específicos vinculados al microcrédito como índices de precios, costos, productividad, así también variables macros como: tasas de intereses, tasa de cambio, desempleo, inflación.

CUARTO:

Utilizar Regresión con Maquinas de Soporte Vectorial (MSV) para clasificación

QUINTO:

Comparar con modelos predictivos.

REFERENCIAS BIBLIOGRAFICAS

Libros

Cea M.A. (2002). *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid: Ed. Síntesis

Cruz F.J. (1998). *Enfoques cuantitativos para la predicción del riesgo de crédito*

Hair J.F., Anderson R.E., Tatham, R.L. y Black; W.C. (1995): *Multivariate Data Analysis. 4ª edición*. Englewood Cliffs, NJ: Prentice Hall.

Luque, T. (2000). *Técnicas de análisis de datos en investigación de mercados*. Edit Ediciones PIRAMIDE. Madrid España.

Uriel, E. (1997): *Análisis de datos. Series temporales y análisis multivariante*. Madrid:

Uriel, E. y Aldas J. (2005). *Análisis multivariante aplicado*. Edit. Thomson. Madrid España

Revistas

Alonso, C. (2001). Creditscoring y consentimiento en la protección de datos, *Perspectivas del sistema financiero*, nº 72, pp. 85-91.

Banegas C. y Garcia F. (2000). El riesgo de crédito en la banca ante la nueva regulación del Banco de España, *Banca & Finanzas*, nº 60, pp. 6-9

Beltran, J. (1994). Control y gestión de los riesgos financieros, *Alta Dirección*, nº 173, pp. 65-72

Boal N. y Gonzales M. (2001). Estimación del riesgo de crédito mediante modelos internos, *Banca & Finanzas*, nº 66, pp. 40-45

Espin O. y Rodriguez C. (2013). Metodología para un scoring de clientes sin referencias crediticias, *Cuad. Econ. vol.32 no.59 Bogotá Jan./June.*

Fontalvo T., Mendoza A. y Vilbal D. (2016). Evaluación del comportamiento de los indicadores de productividad y rentabilidad en las empresas prestadores de salud del Régimen Contributivo en Colombia. *Revista científica Salud Uninorte, Vol 32, Nro. 3.*

Garcia D., Arquez A. y Clavo A. (2013) *Artículos doctrinales.* recuperado de https://www.aeca.es/old/refc_1972-2013/1995/82-8.pdf

Puertas R. y Martí M. Luisa recuperado de: <http://www.scielo.br/pdf/rae/v53n3/07.pdf...>

Mures M.J., García A. y Vallejo 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. Nro. 1.*

Trejo J.C., Rios H., Martinez, M.A. (2016). Análisis de la Administración del Riesgo Crediticio en México para Tarjetas de Crédito. *Revista mexicana de economía y finanzas. Vol. 11 Nro. 1. Mexico.*

Tesis

Choque C.Z. (2014). *factores determinantes del riesgo crediticio - caso banco solidario S.A.* Tesis Universidad mayor de San Andres. Bolivia

Muñoz E: (1998). La técnica de análisis discriminante: una aplicación para el área bancaria, banco central de costa rica división económica departamento de investigaciones económicas die-nt-03-98, recuperado de:

http://www.bccr.fi.cr/investigacioneseconomicas/sistemafinanciero/La_tecnica_analisis_discriminante,_aplicacion_area_bancaria.pdf

Sarco M.Y. (2017) Factores que determinan el otorgamiento de crédito de la financiera Credinka en la ciudad de Ayaviri, 2015. *Tesis UNA Puno.*

Zapata D.A. (2010). Caracterización de las variables determinantes del riesgo en el microcredito rural. *Tesis Universidad Nacional de Colombia.*
<Http://www.bdigital.unal.edu.co/2005/1/70192935.2010.pdf>

ANEXOS

Anexo A. Script en R para la normalidad

```
datos <- read.csv("ajahuanaMARDIA.csv",sep=";", header = T)
datos
# load MVN package
library(MVN)
## sROC 0.1-2 loaded
result <- mvn(data = datos, mvnTest = "mardia")
result$multivariateNormality
```

Anexo B. Matriz dentro de grupos combinados

Matrices dentro de grupos combinados^a

	atraso máximo del cliente (días)	atraso máximo del cliente últimos 3 meses (días)	atraso actual (días)	numero de créditos vigentes	número de empresas en la SBS	Ingresos	egresos	mora promedio del cliente (días)	máximo capital desembolsado vigente	deuda total en compartamos	deuda directa en la SBS
Covarianza	,166	,065	-,022	,002	-,002	,014	,028	,085	,014	-,012	,007
Atraso máximo del cliente últimos 3 meses (días)	,065	,111	-,016	,007	-,010	-,003	-,001	,047	,001	,002	,001
Atraso actual (días)	-,022	-,016	,176	-,012	-,006	,007	,013	-,012	,012	,017	-,003
numero de créditos vigentes	,002	,007	-,012	,017	-6,248E-5	,006	,004	-,007	,022	,029	,025
Número de empresas en la SBS	-,002	-,010	-,006	-6,248E-5	,032	,005	,001	,004	,010	,009	,031
Ingresos	,014	-,003	,007	,006	,005	,073	,112	-,003	,067	,062	,068
Egresos	,028	-,001	,013	,004	,001	,112	,222	-,009	,065	,059	,063
Mora promedio del cliente (días)	,085	,047	-,012	-,007	,004	-,003	-,009	,207	-,008	-,011	-,012
Máximo capital desembolsado vigente	,014	,001	,012	,022	,010	,067	,065	-,008	,201	,199	,188
Deuda total en compartamos	-,012	,002	,017	,029	,009	,062	,059	-,011	,199	,251	,211
Deuda directa en la SBS	,007	,001	-,003	,025	,031	,068	,063	-,012	,188	,211	,261
Correlación	1,000	,478	-,127	,038	-,025	,129	,146	,459	,074	-,057	,035
Atraso máximo del cliente últimos 3 meses (días)	,478	1,000	-,115	,154	-,163	-,033	-,004	,313	,005	,013	,004
Atraso actual (días)	-,127	-,115	1,000	-,212	-,086	,065	,066	-,065	,066	,080	-,012
Número de créditos vigentes	,038	,154	-,212	1,000	-,003	,168	,059	-,114	,374	,445	,379
Número de empresas en la SBS	-,025	-,163	-,086	-,003	1,000	,099	,008	,046	,124	,100	,333
Ingresos	,129	-,033	,065	,168	,099	1,000	,879	-,026	,551	,459	,497
Egresos	,146	-,004	,066	,059	,008	,879	1,000	-,041	,308	,249	,261
Mora promedio del cliente (días)	,459	,313	-,065	-,114	,046	-,026	-,041	1,000	-,039	-,049	-,053

Máximo capital desembolsado vigente	,074	,005	,066	,374	,124	,551	,308	-,039	1,000	,887	,823
Deuda total en compartamos	-,057	,013	,080	,445	,100	,459	,249	-,049	,887	1,000	,826
Deuda directa en la SBS	,035	,004	-,012	,379	,333	,497	,261	-,053	,823	,826	1,000

a. La matriz de covarianzas tiene 94 grados de libertad.

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

Anexo C. Matriz de covarianzas

Matrices de covarianzas^a

	atraso máximo del cliente (días)	atraso máximo del cliente últimos 3 meses (días)	atraso actual (días)	número de créditos vigentes	número de empresas en la SBS	Ingresos	egresos	mora promedio del cliente (días)	máximo capital desembolsado vigente	deuda total en compartamos	deuda directa en la SBS
A+											
Atraso máximo del cliente (días)	,219	,070	,010	-,016	,005	,019	,053	,112	-,020	-,069	-,049
Atraso máximo del cliente últimos 3 meses (días)	,070	,090	,011	-,004	-,001	,008	,009	,054	-,001	,000	-,005
Atraso actual (días)	,010	,011	,191	-,014	,007	,009	,009	,002	,070	,084	,080
Numero de créditos vigentes	-,016	-,004	-,014	,018	,000	-,003	-,009	-,011	,010	,021	,013
Número de Empresas en la SBS	,005	-,001	,007	,000	,024	,010	,017	,013	,009	,023	,024
Ingresos	,019	,008	,009	-,003	,010	,088	,146	,007	,044	,055	,063
Egresos	,053	,009	,009	-,009	,017	,146	,282	,033	,041	,060	,070
Mora promedio del cliente (días)	,112	,054	,002	-,011	,013	,007	,033	,205	-,007	-,025	-,028
Máximo capital desembolsado vigente	-,020	-,001	,070	,010	,009	,044	,041	-,007	,222	,247	,236
Deuda total en compartamos	-,069	,000	,084	,021	,023	,055	,060	-,025	,247	,341	,302
Deuda directa en la SBS	-,049	-,005	,080	,013	,024	,063	,070	-,028	,236	,302	,302
A											
Atraso máximo del cliente (días)	,169	,079	-,053	,013	-,007	,007	,009	,081	,027	,014	,043

Atraso máximo del cliente últimos 3 meses (días)	,079	,124	-,027	,008	-,015	-,003	,003	,053	,006	,004	,005
Atraso actual (días)	-,053	-,027	,188	-,012	-,010	,002	-,004	-,036	-,012	-,005	-,024
Número de créditos vigentes	,013	,008	-,012	,017	,000	,014	,013	-,002	,036	,041	,039
Número de empresas en la SBS	-,007	-,015	-,010	,000	,040	,006	,000	,003	,010	-,001	,033
Ingresos	,007	-,003	,002	,014	,006	,072	,102	-,010	,083	,080	,092
Egresos	,009	,003	-,004	,013	,000	,102	,196	-,022	,080	,074	,094
Mora promedio del cliente (días)	,081	,053	-,036	-,002	,003	-,010	-,022	,199	-,006	,000	-,006
Máximo capital desembolsado vigente	,027	,006	-,012	,036	,010	,083	,080	-,006	,221	,214	,205
Deuda total en compartamos	,014	,004	-,005	,041	-,001	,080	,074	,000	,214	,250	,203
Deuda directa en la SBS	,043	,005	-,024	,039	,033	,092	,094	-,006	,205	,203	,278
B											
Atraso máximo del cliente (días)	,087	,013	,035	-,007	,004	,032	,057	,061	,016	-,016	-,029
Atraso máximo del cliente últimos 3 meses (días)	,013	,096	-,018	,016	-,005	-,018	-,025	,020	-,015	,001	-,004
Atraso actual (días)	,035	-,018	,117	-,009	-,014	,021	,074	,043	,012	-,005	-,046
Número de créditos vigentes	-,007	,016	-,009	,016	,001	-,007	-,009	-,017	-,007	,003	,001
Número de empresas en la SBS	,004	-,005	-,014	,001	,019	-,006	-,019	-,008	,009	,023	,033
Ingresos	,032	-,018	,021	-,007	-,006	,055	,096	,004	,044	,014	,001
Egresos	,057	-,025	,074	-,009	-,019	,096	,225	-,022	,051	,008	-,045

Mora promedio del cliente (días)	,061	,020	,043	-,017	-,008	,004	-,022	,235	-,015	-,027	-,010
Máximo capital desembolsado vigente	,016	-,015	,012	-,007	,009	,044	,051	-,015	,110	,089	,070
Deuda total en compartamos	-,016	,001	-,005	,003	,023	,014	,008	-,027	,089	,134	,116
Deuda directa en la SBS	-,029	-,004	-,046	,001	,033	,001	-,045	-,010	,070	,116	,151
Total	,163	,064	-,020	,002	-,002	,014	,027	,083	,013	-,012	,007
Atraso máximo del cliente (días)	,064	,111	-,016	,007	-,007	-,004	-,001	,046	,004	,005	,006
Atraso máximo del cliente últimos 3 meses (días)	-,020	-,016	,180	-,012	-,007	,005	,010	-,015	,009	,012	-,006
Número de créditos vigentes	,002	,007	-,012	,017	-9,085E-5	,006	,004	-,007	,022	,029	,025
Número de empresas en la SBS	-,002	-,007	-,007	-9,085E-5	,034	,004	7,159E-5	,004	,013	,012	,036
Ingresos	,014	-,004	,005	,006	,004	,072	,111	-,002	,065	,061	,066
Egresos	,027	-,001	,010	,004	7,159E-5	,111	,219	-,007	,064	,059	,061
Mora promedio del cliente (días)	,083	,046	-,015	-,007	,004	-,002	-,007	,203	-,007	-,009	-,011
Máximo capital desembolsado vigente	,013	,004	,009	,022	,013	,065	,064	-,007	,202	,200	,193
Deuda total en compartamos	-,012	,005	,012	,029	,012	,061	,059	-,009	,200	,251	,214
Deuda directa en la SBS	,007	,006	-,006	,025	,036	,066	,061	-,011	,193	,214	,268

a. La matriz de covarianzas total tiene 96 grados de libertad.
Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

Anexo D. Estadísticas por casos

Estadísticas por casos

Número del caso	Grupo real	Grupo superior						Segundo grupo superior			Puntuaciones discriminantes	
		Grupo pronosticado	P(D>d G=g)		P(G=g D=d)	Distancia de Mahalanobis al cuadrado para centroide	Grupo	P(G=g D=d)	Distancia de Mahalanobis al cuadrado para centroide	Función 1	Función 2	
			p	gl								
Original 1	2	1**	,792	2	,659	,466	2	,234	2,533	-1,363	-,362	
2	2	2	,282	2	,620	2,532	1	,245	4,387	,064	-1,797	
3	1	1	,113	2	,849	4,352	2	,128	8,139	-2,424	-1,282	
4	2	2	,055	2	,771	5,796	3	,176	8,754	1,574	-2,205	
5	2	3**	,639	2	,520	,895	2	,406	1,390	1,333	,179	
6	2	1**	,339	2	,818	2,164	2	,092	6,531	-2,085	,888	
7	2	2	,487	2	,615	1,440	3	,195	3,739	,409	-1,399	
8	1	1	,814	2	,431	,412	3	,308	1,082	-,575	,668	
9	3	3	,820	2	,407	,398	1	,304	,980	-,130	,767	
10	2	3**	,856	2	,540	,310	2	,358	1,130	1,012	,498	
11	2	1**	,963	2	,590	,074	2	,236	1,910	-1,094	,210	
12	3	3	,345	2	,584	2,129	2	,377	3,005	1,889	,259	
13	1	2**	,903	2	,375	,203	1	,328	,469	-,170	,011	
14	1	1	,414	2	,695	1,765	3	,179	4,478	-1,545	1,220	
15	2	1**	,705	2	,499	,700	3	,287	1,804	-,829	,924	

16	2	3**	496	2	688	1,403	2	259	3,356	1,496	1,103
17	3	3	831	2	348	371	2	342	407	-116	328
18	2	2	057	2	776	5,743	3	164	8,850	1,456	-2,267
19	3	3	662	2	449	824	1	310	1,567	-202	1,159
20	3	1**	840	2	372	348	3	317	669	-351	402
21	2	1**	585	2	783	1,074	2	136	4,573	-1,884	162
22	2	2	854	2	466	315	1	308	1,140	-101	-672
23	2	1**	920	2	429	166	2	349	575	-535	-172
24	2	3**	538	2	498	1,238	2	434	1,517	1,434	003
25	2	1**	947	2	614	109	2	234	2,037	-1,180	058
26	2	2	276	2	675	2,576	3	236	4,680	1,171	-1,507
27	2	3**	964	2	524	074	2	337	958	709	635
28	2	3**	806	2	578	431	2	332	1,541	1,101	663
29	2	1**	869	2	488	280	2	342	992	-750	-432
30	2	1**	994	2	499	011	2	283	1,144	-779	166
31	3	2**	318	2	644	2,292	1	200	4,635	312	-1,725
32	3	2**	899	2	435	213	3	407	347	643	-024
33	2	2	980	2	425	041	3	363	354	341	-049
34	1	1	726	2	649	639	3	179	3,218	-1,326	730
35	3	1**	805	2	351	434	3	331	549	-273	406
36	3	3	910	2	502	190	2	289	1,296	262	924
37	1	2**	959	2	398	084	3	362	277	198	076
38	2	3**	517	2	503	1,320	2	433	1,621	1,476	004
39	1	1	934	2	453	137	2	282	1,088	-631	387
40	1	1	286	2	479	2,505	3	365	3,048	-869	1,671

41	1	1	267	2	531	2,640	2	389	3,260	-1,016	-1,528
42	1	1	234	2	881	2,903	2	086	7,562	-2,528	-206
43	2	1**	888	2	425	239	2	362	560	-524	-276
44	2	2	714	2	513	675	1	289	1,825	-039	-992
45	1	1	119	2	771	4,254	2	199	6,968	-2,016	-1,613
46	3	3	735	2	635	616	2	255	2,439	828	1,226
47	3	3	723	2	604	650	2	319	1,926	1,238	731
48	2	2	357	2	648	2,061	3	258	3,905	1,134	-1,322
49	1	1	816	2	357	405	3	326	584	-296	401
50	3	1**	480	2	458	1,469	3	349	2,010	-740	1,295
51	2	2	841	2	531	347	3	299	1,497	578	-683
52	2	2	518	2	472	1,316	1	382	1,741	-425	-1,161
53	2	3**	337	2	554	2,177	2	405	2,802	1,866	098
54	2	2	631	2	582	919	3	290	2,313	855	-933
55	3	3	890	2	450	233	2	410	418	754	154
56	3	3	411	2	612	1,777	1	203	3,989	131	1,828
57	1	2**	877	2	474	262	3	381	700	731	-255
58	2	3**	607	2	636	999	2	219	3,136	518	1,532
59	3	3	729	2	392	632	1	339	925	-271	856
60	1	1	518	2	391	1,317	3	388	1,332	-500	1,181
61	3	3	831	2	445	371	2	426	457	839	061
62	1	1	869	2	606	280	2	266	1,931	-1,162	-340
63	2	3**	824	2	416	387	1	297	1,058	-107	799
64	2	2	863	2	448	295	1	326	933	-166	-594
65	2	3**	024	2	890	7,501	2	098	11,908	2,408	2,457

91	2	1**	798	2	597	451	3	207	2,568	-1,138	695
92	2	1**	621	2	766	954	2	139	4,367	-1,800	316
93	2	3**	944	2	399	116	2	380	214	276	246
94	2	3**	445	2	722	1,619	2	221	3,993	1,367	1,424
95	2	2	903	2	512	205	3	285	1,374	385	-636
96	2	3**	947	2	540	109	2	327	1,114	738	707
97	2	3**	974	2	506	053	2	350	790	686	545

**.. Caso clasificado incorrectamente

Fuente: elaborado por los ejecutores en base a los de datos Financiera compartamos-agencia Puno.

1 **Anexo E.** Comparación de resultados observados y predichos con el modelo

observado	estimado
A	A+
A	A
A+	A+
A	A
A	B
A	A+
A	A
A+	A+
B	B
A	B
A	A+
B	B
A+	A
A+	A+
A	A+
A	B
B	B
A	A
B	B
B	A+
A	A+
A	A
A	A+
A	B
A	A+
A	A
A	B
A	B
A	A+
A	A+
B	A
B	A

3

A	A
A+	A+
B	A+
B	B
A+	A
A	B
A+	A+
A	A+
A	A
A+	A+
B	B
B	B
A	A
A+	A+
B	A+
A	A
A	A
A	B
A	A
B	B
B	B
A+	A
A	B
B	B
A+	A+
B	B
A+	A+
A	B
A	A
A	B

2

A	A
A	A
A	B
A+	B
A	A
A+	A+
A	B
B	B
A+	B
A+	A+
A+	A+
A+	A+
A	A
A+	A
A	A
A	B
A	A+
A+	A+
A+	A
A	A
B	B
A	A
A	A
A	A+
A	B
A	B
A	A
A	B
A	B